2016

Bounds and Algorithms for Subspace Estimation on Riemannian Quotient Submanifolds

Thomas A. Palka
University of Rhode Island, thomas.a.palka@gmail.com

Follow this and additional works at: https://digitalcommons.uri.edu/oa_diss

Recommended Citation

Palka, Thomas A., "Bounds and Algorithms for Subspace Estimation on Riemannian Quotient Submanifolds" (2016). *Open Access Dissertations*. Paper 519.
https://digitalcommons.uri.edu/oa_diss/519

This Dissertation is brought to you for free and open access by DigitalCommons@URI. It has been accepted for inclusion in Open Access Dissertations by an authorized administrator of DigitalCommons@URI. For more information, please contact digitalcommons-group@uri.edu.

BOUNDS AND ALGORITHMS FOR SUBSPACE ESTIMATION ON
RIEMANNIAN QUOTIENT SUBMANIFOLDS

BY

THOMAS A. PALKA

A DISSERTATION SUBMITTED IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY
IN
ELECTRICAL ENGINEERING

UNIVERSITY OF RHODE ISLAND

2016

DOCTOR OF PHILOSOPHY DISSERTATION
OF
THOMAS A. PALKA

APPROVED:

Dissertation Committee:

Major Professor Richard J. Vaccaro

Orlando Merino

Steven M. Kay

Nasser H. Zawia

DEAN OF THE GRADUATE SCHOOL

UNIVERSITY OF RHODE ISLAND

2016

ABSTRACT

Subspace estimation appears in a wide variety of signal processing applications such as radar, communications, and underwater acoustics, often as a prelude to high resolution parameter estimation. As with any estimation problem the availability of statistical benchmarks on estimator accuracy is key to developing and understanding algorithm performance. The parameter space in general subspace/basis estimation problems is naturally described as a Riemannian quotient manifold. The concept of a manifold is central to many parts of geometry and modern mathematical physics because it allows more complicated structures to be described and understood in terms of the properties of Euclidean spaces. This identification permits the well-developed tools of differential geometry to be brought to bear on the analysis of subspace/basis estimation problems. Classical Cramér-Rao bounds (CRB), originally formulated for standard Euclidean spaces, have recently been generalized to Riemannian quotient manifolds and the particular case of the standard real linear statistical model with the Grassmann manifold (set of subspaces) as the parameter space has been analyzed.

The present work applies this differential geometric approach to the analysis of the complex linear model estimation problem. We consider both the general unconstrained signal model and, most importantly, the parametrically constrained signal model. First, we show that the appropriate parameter space for the most general unconstrained signal model is a modified Stiefel manifold, termed here the Basis manifold. Elements of the Basis manifold are semi-unitary matrices grouped into equivalence classes dictated by the model invariances. Using this formulation we derive the full Fisher Information Matrix (FIM) and, from this, intrinsic CRBs on the both subspace and rotation estimation accuracy. Among the corollaries that flow from this Basis manifold formulation is a CRB on the individual columns of the eigen vector matrix, say Y (an $N \times p$ semi-unitary matrix), of the true signal covariance. The eigen decomposition of the sample covariance

matrix (SCM) yields estimate of Y ; the analytical expression for the corresponding estimation error covariance is well known from statistics. We show the relationship between this existing expression and the CRB developed here using the geometric formulation.

Secondly, we consider situation when the signal matrix is constrained, and in particular, the important special case when the constraint arises due to some underlying parametric model. Here the set of semi-unitary matrices representing the signal matrix are constrained to a submanifold of the Basis manifold. In this geometric approach these underlying parameters are viewed as a set of natural coordinates on the submanifold; the set of values that describe the related $N \times p$ matrix representation is an alternate (extrinsic) coordinate representation. Using this differential geometric framework, we derive the FIM with respect to a set of normal coordinates, and the resultant intrinsic CRB on subspace and rotational estimation accuracies for this constrained model. The relationship between this new bound on constrained subspace estimation accuracy and the existing bounds developed for system parameters (i.e., natural coordinates) is established and its relevance to the estimation problem is explored in detail.

The CRB development for the constrained model naturally suggests an asymptotically ML estimation approach that leverages the estimate given by the standard EVD of the SCM (ML for the unconstrained signal matrix assumption). This estimate, referenced to an algorithmically convenient extrinsic coordinate set, may be transformed to the natural coordinates defined by the underlying parametric model. By the maximum likelihood (ML) invariance principle, this estimation approach yields an asymptotically ML estimator of the desired parameters.

An efficient implementation of this general theoretical approach is developed for the important special case of the uniform multi-dimensional array and complex exponential waveform model. Utilizing the shift-invariant properties that define the constraint in this setting, we derive a closed-form estimator of the signal matrix Y . This

new estimation approach is applied to the challenging 2-D Direction-of-Arrival (DOA) estimation problem and a set specific scenarios drawn from the literature are evaluated to demonstrate performance.

TABLE OF CONTENTS

ABSTRACT	ii
TABLE OF CONTENTS	v
LIST OF TABLES	xi
LIST OF FIGURES	xii
CHAPTER	
1 Introduction	1
1.1 Differential Geometry and its Relevance to Subspace Estimation . .	2
1.1.1 Archetypal Example: 2-Sphere	2
1.1.2 Special Case Unitary Group	5
1.2 Estimation on Manifolds	7
1.2.1 Complex Linear Model: General	8
1.2.2 Complex Linear Model: Constrained	10
1.3 Organization	15
List of References	16
2 Differential Geometry for Signal Processing Estimation Applications .	18
2.1 Geometry of the Sphere	18
2.1.1 Estimation Example on the Unit Sphere	31
2.2 Geometry of Signal Processing Manifolds	33
2.2.1 Coordinates on Manifolds	34
2.2.2 Unitary Group as a Riemannian Manifold	35

	Page
2.2.3 Signal Processing Manifold Representations	42
2.2.4 Quotient Space Representation of the Complex Stiefel Manifold	43
2.2.5 Basis Manifold	52
2.2.6 Grassmann Manifold	54
2.2.7 Concrete Representation of the Stiefel Manifold	55
2.3 Subspace Angles, Subspace Distance and Geodesic Distance	60
2.3.1 Subspace Distance on Grassmann and Basis Manifolds	63
2.3.2 Alternate Non-Normal Coordinate System	68
List of References	75
3 Intrinsic Cramér Rao Bounds for Basis Manifold and Complex Linear Model	77
3.1 Complex Linear Model	77
3.2 Estimation on Riemannian (Quotient) Manifolds	81
3.3 Intrinsic Mean Square Error Measure	87
3.3.1 Subspace and Rotation Angles	87
3.3.2 Subspace Angles on Grassmann and Basis Manifolds	88
3.4 Intrinsic CRB	91
3.5 Nuisance Parameters	95
3.6 CRB for Basis Manifold formulation of Complex Linear Model . .	95
3.6.1 Intrinsic CRB Subspace Distance	100
3.6.2 Intrinsic CRB on Rotational Distance	101
3.7 Eigenvector CRB	102
3.8 Constrained Basis Estimation	104

	Page
3.8.1 Estimation on the Submanifold	107
3.8.2 Computing the constrained case FIM 1	107
3.8.3 Constrained CRB	109
3.9 Parametrically Constrained Problem	110
3.9.1 Defining the Submanifold Tangent Space	110
3.9.2 Computing the Natural Coordinate Basis	111
3.9.3 Generate an Orthonormal Basis on S_Y from the natural co- ordinate basis	113
3.9.4 Modal Forms	115
List of References	125
4 Maximum Likelihood Estimator	127
4.1 Approximating the compressed likelihood function near $\tilde{Y}$	130
4.1.1 Constrained MLE Criterion Function	133
4.2 Solution to the Constraint MLE criterion using an S-N Frame	135
4.3 Solution Methods	142
4.3.1 MLE for Submanifold Defined by Shift-Invariance: Uni- form Arrays	142
4.4 <i>MLESI</i> Algorithm Summary	151
List of References	154
5 Summary and Simulation Results	155
5.1 Unconstrained Problem: $Y \in \mathcal{B}$	155
5.1.1 Simulation Results	158
5.2 Constrained Problem: $Y \in \bar{\mathcal{B}} \subset \mathcal{B}$	160
5.2.1 Asymptotic ML Estimators for $Y \in \bar{\mathcal{B}}$	166

	Page
5.2.2 Application to the harmonic retrieval problem	172
5.2.3 Example: Correlated vs. Uncorrelated	175
5.2.4 Performance Metrics	178
5.2.5 Simulation Results	182
List of References	201

APPENDIX

A appendix A	204
A.1 Coordinate Representations	204
A.1.1 Coordinate System 1: The $N \times p$ matrix extrinsic coordinates	204
A.1.2 Coordinate System 2: Riemannian normal coordinate system (α, β)	205
A.1.3 Coordinate Transformations: Normal Coordinates and Extrinsic Coordinates	208
A.2 Curves on Manifolds	208
A.2.1 Geodesic Curves	209
A.2.2 Functions on Manifolds	211
A.2.3 Coordinate Curves	212
A.2.4 Calculus on Manifolds : Partial Derivative	213
A.3 Elements of Differential Geometry	214
A.3.1 Natural Coordinate Basis	214
A.3.2 Dual Basis, Tensors	214
A.3.3 Tensor Fields	217
A.3.4 Differentiation on Manifolds	218
A.4 Covariant Derivative	220

	Page
A.4.1 2nd Covariant Derivative Differential of a function	222
A.5 Computing Derivatives on $\mathcal{B}$	224
A.6 Inner Product Space	225
A.6.1 2 by 1 complex vectors over the Reals (C^2, R)	226
A.6.2 2 by 1 complex vectors over the Complex (C^2, C)	227
A.7 The space (C^N, R) for N even	228
A.8 Subspaces of (C^N, R) and (C^N, C)	229
A.8.1 Example: Subsets of C^4 that are vector spaces in (C^4, R) . .	230
A.9 Transforming a Basis for (W, R)	231
A.10 Orthogonalizing a Basis	232
A.11 Specifics of the Tangent Space Problem	232
A.11.1 N by p complex matrices over the Reals	232
A.11.2 Partitioning the space $C^{N \times p}$	233
A.11.3 Algebra of the B_Y space	234
A.12 Partitioned space and basis	237
A.12.1 Tensor Transform in partitioned frame	240
List of References	242
B appendix B	243
B.1 FIM appendix	243
B.1.1 Proof of Lemma 8	243
B.1.2 Proof of Lemma 9	244
B.1.3 Proof of Theorem 10 (Full FIM)	245
B.1.4 Proof of Corollary 13 (Eigenvector CRB)	255

	Page
B.1.5 Proof of Lemma 15	261
B.1.6 Proof of 18 (M-Dim Orthogonalizing Matrix Lemma) . . .	262
B.1.7 Proof of Lemma 16 (Equivalence of Forms)	264
B.1.8 Proof of Theorem 17	264
B.1.9 Proof of Theorem 19	266
B.1.10 Useful Matrix Identities	269
List of References	272
C appendix C	273
C.1 Proof of Lemma 20	273
C.2 Proof of Lemma 21	275
C.3 Proof of lemma 23	277
C.4 Proof of theorem 24	279
C.5 Proof of Lemma 25	281
C.6 MLE using Reference Solution and Geometric Newton's Method . .	281
C.6.1 Estimation for M-Dimensional Modal Signal Model	290
C.6.2 Geometric Newton's Method Refinement Algorithm Summary	294
BIBLIOGRAPHY	298

LIST OF TABLES

Table		Page
1	Horizontal $\mathcal{H}_{X_0}$ and Vertical $\mathcal{V}_{X_0}$ Space Dimension of Quotient Manifolds	56
2	Dimension of S_Y for $M = 1, 2$ dimensional Arrays, Undamped ($r = 1$) and Damped ($r = 2$) exponential signals	123
3	Dimension of subspace perturbation spaces for different signal assumptions (M =array dimension, p =number of modes)	174
4	Example 1: Spatial frequencies, 2 Sources	183
5	Example 5: Spatial frequencies, 4 widely spaced sources,	192
6	Example 6:Spatial frequencies, 4 Closely spaced sources	196

LIST OF FIGURES

Figure		Page
1	Subspace Based Estimation Form	3
2	Estimation on the Sphere	8
3	Coordinates on Product Manifold	12
4	Tangent and Normal Space	20
5	Horizontal and Vertical Spaces	45
6	Subspace Distance on Basis Manifold	67
7	S-N Basis for the Constraint Submanifold	137
8	Selection Matrices for 2-D Planar Arrays	148
9	Subspace Error: SNR	161
10	Subspace Error: Sample Support Test	162
11	Rotational Error SNR	163
12	Rotational Error Sample Support	164
13	Validation of (830)	180
14	Subspace Plot Set 1	184
15	Parameter Estimations Errors and CRB: 2 Sources	186
16	Sample Support	187
17	Performance as a function of Source Angular Separation	188
18	Correlated and Uncorrelated, Fixed Eigenvalues $\rho = 0.9999$	190
19	Two sources vary correlation, Fixed SNR	191
20	Subspace Estimator Performance : 4 Sources Widely Spaced Sources	193
21	Parameter Space Error: 4 Widely Spaced Sources	194

Figure		Page
22	Subspace Error vs Parameter Space Error	195
23	Subspace Error : 4 Sources very closely spaced	198
24	Parameter Space Error: very closely spaced	199
25	Scatter plot of Errors at fixed SNR	200
26	Histogram of Errors at fixed SNR	201
27	Subspace Error vs Parameter Space Error	202
C.1	Newton's Method1	287

CHAPTER 1

Introduction

Subspace and basis estimation appear in a wide variety of signal processing applications such as radar, communications, and underwater acoustics [1], either as an end in themselves or as a prelude to high resolution parameter estimation. As with any estimation problem the availability of statistical benchmarks on estimator accuracy is key to developing and understanding algorithm performance. The parameter space in general subspace/basis estimation problems is naturally described as a Riemannian quotient manifold. The concept of a manifold is central to many parts of geometry and modern mathematical physics because it allows more complicated structures to be analyzed in terms of the relatively well-understood properties of Euclidean spaces. This identification permits the well-developed tools of differential geometry to be brought to bear on the analysis of estimation problems on manifolds. Cramér-Rao bounds (CRB) have recently been generalized to Riemannian quotient manifolds and the particular case of the standard real linear statistical model with the Grassmann manifold (set of subspaces) as the problem parameter space has been analyzed [2]. Here we consider the complex linear model using a modified Stiefel manifold as the parameter space. In particular, we consider the case when the parameter space is constrained and develop intrinsic CRBs and estimation algorithms for this important model.

A vast number of algorithms for estimating unknown signal parameters from the measured output of an N -element sensor array have appeared in the literature. This problem is posed using the complex linear model with the $N \times p$ signal matrix ¹, $H(\theta)$, constrained by its parametric description. There are a variety of so-called subspace based estimation methods that employ a two-step procedure consisting of a subspace

¹ N sensors or array elements and p signals

estimation step followed by a parameter extraction step (see figure 1). An estimate in the form of a semi-unitary $N \times p$ matrix, say $\hat{Y}$, is formed from the K -snapshot measurement data matrix, say $Z = \begin{bmatrix} z_1 & \cdots & z_K \end{bmatrix}$, and the p -tuple $\hat{\theta}$, is then extracted from $\hat{Y}$ in the second step. Standard algorithms (e.g., R-D Unitary ESPRIT) for the important and commonly encountered special case of uniform arrays use the p -dominant eigenvectors of the sample covariance matrix (SCM) as the subspace estimate. This is the maximum likelihood (ML) estimate of Y for the unconstrained signal matrix model, but is not ML when constraints are present, and as a result, standard subspace-based methods are non-optimal. More recent algorithms such as structured least squares (SLS) [3] make partial use of the constraint information in forming the subspace estimate and therefore yield improved parameter estimators, but remain sub-optimal and perform poorly in certain challenging scenarios (e.g., highly correlated sources).

Formulating the problem with a constrained variation of the Steifel manifold as the parameter space we develop an intrinsic CRB on the subspace estimation step. Using the geometric insights that flow from this formulation we develop a closed-form, asymptotically ML, subspace estimator applicable to uniform arrays. Using this new constrained subspace estimator an improved subspace-based parameter estimation algorithm is achieved. This introductory chapter gives a brief outline of the general approach and summarizes some key results.

1.1 Differential Geometry and its Relevance to Subspace Estimation

1.1.1 Archetypal Example: 2-Sphere

The archetypal example of a Riemannian manifold is the surface of a ball, the 2-sphere denoted S^2 , embedded in the familiar 3-dimensional Euclidean space² denoted here as $\mathcal{E}$. The defining contribution to the theory of such surfaces was made by Gauss

²The two-dimensional surface of a three-dimensional ball is the 2-sphere, denoted S^2 , often simply called a sphere in the mathematical literature. A sphere may be defined geometrically as the set of all points at a fixed distance from a given point in space.

Figure 1. Subspace Based Estimation Form

[4] in two papers written in the early nineteenth century. The approach developed there “marked a new departure from tradition because for the first time Gauss considered the intrinsic geometry of a surface, the properties which are determined only by the geodesic distances between points on the surface independently of the particular way in which the surface is located in the ambient Euclidean space”[5]. This point of view was extended to higher-dimensional spaces by Riemann and led to what is known today as differential geometry. The most important general concepts within this broad subject relevant to our discussion are: manifolds, coordinates, tangent spaces, metrics, exponential map, geodesics, and covariant derivative.

While modern differential geometry generally prefers coordinate free approaches, calculations require numerical representations, rather than abstract ones, and coordinates on the manifold fulfill this requirement. A coordinate system assigns a real N -tuple of values to each point of manifold. A particular system, say x , is a mapping $x : \mathcal{M} \mapsto \mathbb{R}^N$ (perhaps defined only locally). The N -tuple given by

$$x = x(\mathcal{P}) \equiv (x^1(\mathcal{P}), \dots, x^N(\mathcal{P})); \quad \mathcal{P} \in \mathcal{M} \quad (1)$$

are the numerical coordinates of the point $\mathcal{P}$ with respect this particular coordinate sys-

tem³. This identification enables calculus on manifolds [2]. There are, of course, an infinite variety of ways to label points on manifolds. Points on the sphere, $S^2 \subset \mathcal{E}$, for example, may be labeled by their 3-space Cartesian coordinates (e.g.; Earth-Center-Earth Fixed (ECEF) coordinates), or by their latitude and longitude. The first set will be called an extrinsic set since it references points on S^2 to the ambient space and thereby uses 3 numbers to label a point on a 2-dimensional surface. Coordinates are chosen for their convenience to the particular task at hand. In an estimation application, a position estimator may chose the set of coordinates that is algorithmically convenient and results established in one coordinate set may be transformed to an alternate set, if desired, provided the transformation is known.

Tangent spaces, vector spaces notionally attached to each point $\mathcal{P}$ of a manifold $\mathcal{M}$, play a prominent role in the geometric development; at the base point $\mathcal{P} \in \mathcal{M}$ the tangent space is denoted $T_{\mathcal{P}}\mathcal{M}$. In the standard 3-dimensional space Euclidean space, for example, the set of all displacement vectors at the arbitrary point $\mathcal{P}$, defines a 3-dimensional vector space, denoted $T_{\mathcal{P}}\mathcal{E}$. The length of a displacement vector between two points, say $\Delta = \mathcal{P}_1 - \mathcal{P}_0$, defines the Euclidean distance between points, $dist_{\mathcal{E}}(\mathcal{P}_1, \mathcal{P}_0) = \|\Delta\|$. For the sphere, $\mathcal{M} \equiv S^2$, the plane tangent to the north pole, say $\mathcal{P}_0$, is a particular instance of a tangent space on the sphere denoted $T_{\mathcal{P}_0}S^2$; the plane tangent to a given point on the equator, say $\mathcal{P}_1$, is another denoted $T_{\mathcal{P}_1}S^2$. The space $T_{\mathcal{P}_0}S^2$, for example, is evidently a 2-dimensional vector subspace of the 3-dimensional space $T_{\mathcal{P}_0}\mathcal{E}$. The tangent spaces of the sphere inherit the Euclidean inner product of the ambient space via restriction and thereby become inner product spaces. This process yields a smoothly varying inner product on the tangent spaces, termed a Riemannian metric. In general, a manifold endowed with such a metric is a Riemannian manifold.

The distance between any two points on such manifolds is well defined and the dis-

³The common convention is to use the same symbol to denote both the mapping and the resultant set of values, that is $x = x(\mathcal{P})$.

tance minimizing curves between points are termed geodesics. The geodesic or Euler-Lagrange equation ⁴ is a vector differential equation generalizing the $\ddot{x} = 0$ that defines geodesics (straight lines) in a Euclidean space. The geodesic equation on S^2 referenced to a Cartesian coordinate system, y , with origin at the center of the sphere is given by $\ddot{y} + (\dot{y}^t \dot{y}) y = 0$. The geodesic emanating from the point $\mathcal{P}_0$, with coordinates y_0 , in the direction of the unit tangent vector $\Delta \in T_{y_0} S^2$, may be expressed in coordinates as the exponential mapping $y(t) = \exp_{y_0}(\Delta \cdot t)$ ⁵. The running length along this geodesic curve is $\text{dist}_{S^2}(\mathcal{P}(t), \mathcal{P}_0) = \|\Delta t\|$, so that, just as in the Euclidean case, the distance between points is given by the magnitude of this connecting tangent vector. The geodesic distance, an intrinsic measure or one that is independent of the choice of coordinates, is a useful estimation error measure allowing estimators that may use different coordinates to be evaluated with the same "yardstick".

1.1.2 Special Case Unitary Group

The sphere example, $S^2 \subset \mathcal{E}$, notionally generalizes to the orthogonal group $O(N)$, the set of all $N \times N$ orthogonal matrices viewed as a subset of an N^2 -dimensional Euclidean space, the set of all $N \times N$ real matrices. The tangents spaces of $O(N)$ inherit their metrics from the ambient Euclidean space in the same way as the sphere. The complex generalization of $O(N)$ is the unitary group $U(N)$, the set of all $N \times N$ unitary matrices.

The geodesic equation for $\mathcal{U}(N)$ is the matrix differential equation $\ddot{X} = X(\dot{X}^H \dot{X})$, where X is $N \times N$ unitary. Direct substitution verifies that this non-linear differential equation has the general solution

$$X(t) = X_0 \cdot \expm(Wt) \quad (2)$$

for W arbitrary anti-Hermitian, X_0 arbitrary unitary matrix, and $\expm$ the matrix expo-

⁴derived using the Calculus of Variations

⁵This is the solution to the geodesic equation with initial conditions $y(0) = y_0$ and $\dot{y}(0) = \Delta$ and is a great circle. The precise meaning of this notation, $\exp_{Y_0}(\Delta t)$ is given in section 2.1

nential operator. The short hand notation for this mapping (2) with $\Delta \equiv X_0 W$ is

$$\exp_{X_0}(t\Delta) \equiv X_0 \expm(tW) . \quad (3)$$

In this expression $\Delta \in T_{X_0}\mathcal{U}(N)$ is a unit tangent vector at X_0 represented as an $N \times N$ complex matrix. Thus, the exponential map describes geodesics and the curve (3) is said to be the geodesic emanating from X_0 in the direction of the tangent vector Δ .

The Steifel and Grassmann manifolds are particular examples of what are termed quotient manifolds and arise naturally in statistical signal processing applications. Such manifolds are defined via equivalence relations on $\mathcal{U}(N)$ in which subsets of equivalent elements are “points”. The collection of these subsets define the quotient manifold, denoted generically as $\mathcal{U}(N) \setminus \sim$, where $\sim$ stands for a particular equivalence relation.

The complex Stiefel manifold, $St_{(N,p)}$, is defined via the equivalence relation that identifies $X_1, X_2 \in \mathcal{U}(N)$ as equivalent if the first p columns of the $N \times N$ matrix X_1 are identical to those of X_2 . The equivalence subset denoted $[X_1]$, a fiber submanifold of $\mathcal{U}(N)$, is a “point” of the Stiefel manifold. Such quotient manifolds inherit their geometric properties from $\mathcal{U}(N)$ by restriction of geodesics to the quotient manifold tangent space (termed the horizontal space).

We introduce a closely related quotient manifold on $\mathcal{U}(N)$ termed here the Basis manifold and denoted as $\mathcal{B}_{N,p}$ (or simply $\mathcal{B}$ when N and p are understood). In estimation applications, the statistical model dictates the appropriate equivalence relation and, for the complex linear model considered here, this leads us to the Basis manifold. The tangent spaces of $\mathcal{B}$ naturally partition as $T_Y\mathcal{B} \equiv A_Y \oplus B_Y$. A geodesic emanating in the direction of a tangent vector in the B_Y space represents a subspace perturbation, while motion in the A_Y directions represents a rotation. This partitioning figures prominently in the formulation of intrinsic bounds on subspace and rotational estimation accuracy.

Since the tangent spaces of the Basis manifold are naturally partitioned, the

geodesic distance is likewise partitioned as $dist_B(\mathcal{P}_0, \mathcal{P}_1) = \sqrt{\|P_B \Delta\|^2 + \|P_A \Delta\|^2}$. In this expression $\Delta \equiv \exp_{\mathcal{P}_0}^{-1} \mathcal{P}_1$ is the connecting tangent vector; P_A and P_B are the projectors onto A_Y and B_Y , respectively. A standard measure of the difference between two subspaces, represented say by Y_1 and Y_0 , are the subspace angles, which generalize the notion of angles between vectors⁶. The root sum square of these angles is the subspace distance denoted $dist_{SS}(Y_1, Y_0)$. In the Grassman manifold, $\mathcal{G}$, formulation this corresponds to the geodesic distance between the two points, that is $dist_{SS}(Y_1, Y_0) \equiv dist_{\mathcal{G}}(Y_1, Y_0)$. For the Basis and Stiefel manifold formulations, the situation is more complicated. We show that for these formulations the subspace distance is equal to the minimum geodesic distance between the point Y_0 and an invariant submanifold over Y_1 . This relationship enables the development of an intrinsic CRB on subspace estimation accuracy for these parameter spaces.

1.2 Estimation on Manifolds

In estimation theory and statistics, the CRB relates the covariance matrix of estimators to the Fisher Information Matrix (FIM) of an estimation problem through matrix inequalities. In the classical case of a vector parameter $\theta \in R^p$ in the distribution $f(z|\theta)$ this inequality is

$$C \succeq F^{-1} \quad (4)$$

where C is the estimator error covariance $C = E[(\theta - \hat{\theta})(\theta - \hat{\theta})^T]$, F is the FIM given by $F_{i,j} \equiv E \left[\frac{\partial L}{\partial \theta^i}, \frac{\partial L}{\partial \theta^j} \right]; i, j = \dots N$, $L = \ln(f)$, and $\hat{\theta} = \hat{\theta}(z)$ is an unbiased estimator of θ based on the measurement z . The matrix inequality means that the difference $C - F^{-1}$ is positive semi-definite.

The derivation of this inequality implicitly relies on the assumption that the parameter space is a Euclidean space. For a more general Riemannian manifold as the parameter space, the role of the error vector $\hat{\theta} - \theta$ is played by the error tangent vector

⁶defined operationally as the $\cos^{-1}$ of the singular values of the matrix $Y_1^H Y_0$

Figure 2. Estimation on the Sphere

$\hat{\Delta} = \exp_{\theta}^{-1} \hat{\theta}$ and the relevant derivative is the covariant derivative. This derivative, a generalization of the directional derivative to curved spaces, results in additional terms appearing in the CRB inequality, so that in this setting (4) becomes [2]

$$C \succeq F^{-1} + \text{curvature terms} . \quad (5)$$

At high SNR, estimation errors are small and these additional terms become negligible. The CRB, therefore, depends asymptotically only on the FIM, just as with classical bounds. While the components of C and F depend on the particular choice of coordinates, the error variance $\text{tr}C$ and bound $\text{tr}F^{-1}$ do not, provided the metric is properly accounted for. When the parameter space is the unit sphere $\text{tr}C$ represents the 2-D mean square positional estimation error measured in radians and $\text{tr}F^{-1}$ bounds this error.

1.2.1 Complex Linear Model: General

The complex linear model has received a huge amount of attention due to its simplicity and relevance in a large number of applications. Consider the model

$$z = Hn_1 + n_0 \quad (6)$$

where n_0 is a normal random complex N -tuple $n_0 \sim \mathcal{CN}_p(0, \sigma^2 I)$, and n_1 is a normal random complex p -tuple, $n_1 \sim \mathcal{CN}_p(0, P)$. The observation z then is a normal random

N -tuple, $z \sim CN_n(0, R)$, where $R = E[zz^H] = R_s + \sigma^2 I$ with signal covariance R_s defined by

$$R_s(H, P) \equiv HPH^H. \quad (7)$$

In this expression P is an unknown deterministic full rank Hermitian matrix and represents the correlation matrix of the source symbols n_1 . The complex $N \times p$ signal matrix $H \in C^{N \times p}$ is a deterministic unknown of full rank. We refer to this case as unconstrained since there are no constraints other than rank of H and the positive-definite Hermitian form of P . This data model is most relevant when we are unable or unwilling to make more detailed assumptions regarding the signal matrix H .

Sampling of the distribution reveals information about the unknown parameters, P and H , and the observable information is subject to estimation. It is readily seen from (6) that this model precludes the simultaneous and unambiguous estimation of these unknowns. To see this, note that for any invertible T , substituting $H = HT^{-1}$ and $n_1 = Tn_1$ into (6) leaves the measurement unchanged. Thus, both the pairs (H, P) and (HT^{-1}, TPT^H) define the same measurement distribution. As noted in [6] “whenever two parameters give rise to the same measurement distribution they are indistinguishable in the sense that no argument based on the observed measurement can be used to favor one parameter over the other as estimator”.

An alternate parameterization of the problem is provided by the eigen decomposition of the signal covariance $R_s(H, P)$, (7). The eigen decomposition defines the mapping or transformation

$$eig : C^{N \times p} \times P_p \mapsto \mathcal{B} \times \mathcal{D}_p \quad (8)$$

from the (H, P) to (Y, Λ) parameterizations as

$$[Y, \Lambda] = eig(HPH^H). \quad (9)$$

In this expression $\mathcal{P}_p$ is the set all $p \times p$ positive-definite Hermitian matrices and $\mathcal{D}_p \subset$

$\mathcal{P}_p$ is the subset of positive-definite diagonal matrices.

The transformed problem in terms of this alternate parameterization is

$$z = Yn_1 + n_0, \quad Y \in \mathcal{B} \quad (10)$$

where $n_1 \sim N_p(0, \Lambda)$. The proper choice for the space $\mathcal{B}$ is dictated by the statistical model. As discussed in chapter 3, the Stiefel manifold is “too big” to serve as the proper parameter space for this model while the Grassmann manifold is “too small”. The given model dictates a quotient manifold formulation in which multiplication by unitary diagonal matrices have been “quotiented” out of the Stiefel manifold. The quotient manifold with this desired feature is called here the Basis manifold, $\mathcal{B}$, and defining the domain for Y in (10) as $\mathcal{B}$ yields a well-posed estimation problem.

We derive the full FIM for the estimation problem (10) on the product manifold, $\mathcal{B} \times D_p$, and, from this, intrinsic CRBs on subspace estimation accuracy. Among the corollaries that flow from this Basis manifold formulation is a CRB on eigenvector (interpreted as the columns on Y) estimation accuracy for the complex linear model. For the unconstrained case the maximum likelihood (ML) estimate of $Y \in \mathcal{B}$ is given by the p -dominant eigenvectors of the SCM. The corresponding estimation error covariance is well-known from statistics [7] and is often used to develop analytical expressions for estimator performance [8],[9],[10],[11]. The relationship between this error covariance expression and the CRB developed here is established in chapter 3.

1.2.2 Complex Linear Model: Constrained

When the set of allowable signal matrices, H in (6), is constrained in some way the set of allowable semi-unitary matrices Y in (10) is likewise constrained to submanifold denoted $\bar{\mathcal{B}} \subset \mathcal{B}$. We derive the general form for the intrinsic CRB on constrained subspace estimation accuracy on $\bar{\mathcal{B}}$ and then consider the special case when the constraint defining $\bar{\mathcal{B}}$ arises due to some underlying parameterization, that is when $H = H(\theta)$. In

this parametrically constrained case the eigen decomposition of the true signal covariance effects a transformation, denoted here as $[Y, \Lambda] = \Phi(\theta, P)$, between the (θ, P) and the $(Y \in \bar{\mathcal{B}}, \Lambda \in D_p)$ parameterizations. This transformation is

$$[Y, \Lambda] = \text{eig}(H(\theta)PH^H(\theta)) . \quad (11)$$

The parameters (θ, P) and (Y, Λ) are then two different sets of coordinates on the manifold $\bar{\mathcal{B}} \times \mathcal{D}_p$. This is analogous to using either latitude-longitude or 3-space Cartesian coordinates to label points on $S^2 \in \mathcal{E}$.

The tangents to the θ coordinate curves yield a natural coordinate basis for the tangent spaces at each point of $\bar{\mathcal{B}}$. From this an orthogonal basis may be formed and via the exponential map these new basis vectors yield a set of so-called normal coordinates, denoted here as (α, s, λ) about any given point. This set is a third parameterization of L , or equivalently, a third set of coordinates on the product manifold $\bar{\mathcal{B}} \times D_p$; the interrelationships between these coordinate choices is shown in figure 3.

We compute the general form for the constrained FIM in these partitioned normal coordinates and the corresponding intrinsic bounds on (constrained) subspace estimation accuracy. The coordinate Jacobian transforms the FIM in a tensorial way; bounds (and estimates) are likewise transformed. Quantities computed in one coordinate system, which may be more amenable to computation, may be transformed to an alternate system, which may have a more meaningful interpretation. In particular, the FIM with respect to the natural coordinates, $F_{\theta\theta}$ and the subspace (normal) coordinates, F_{ss} , are related via the Jacobean of the coordinate transformation as indicated in figure (3). As noted in [12] “because the Jacobian determines how coordinate transformations affect accuracy bounds, it is sometimes called a sensitivity matrix.”

In a variety of applications the signal matrix (6) has a multi-dimensional modal form, that is $H(\theta) = [h(\theta^1), \dots, h(\theta^N)]$ and the waveform $h(\cdot)$ is either a damped or undamped exponential. The FIM for this model, $F_{\theta\theta}$, was derived in [13] based on

Coordinates on the Product Manifold

Figure 3. Coordinates on Product Manifold

a maximum likelihood argument that ultimately relied on the eigenvector covariance result of [7]. A more direct “textbook” derivation of parameter space FIM, $F_{\theta\theta}$ and the associated stochastic CRB was later developed in [14]. Alternatively, $F_{\theta\theta}$ may be found from subspace block of the full FIM, F_{ss} , and the coordinate Jacobean. The geometric approach thus yields another derivation of parameter space FIM, $F_{\theta\theta}$.

Maximum Likelihood Estimator

Typically, estimates of the underlying parameters, θ , are required rather than the semi-unitary matrix that they define, that is, $Y(\theta)$ via (11). The parameter estimation step in two-step subspace based methods is related to the various submanifold coordinates as follows. The inverse of the eigen decomposition (11) defines a mapping from (Y, Λ) to (θ, P) coordinates and has the general form

$$(\theta, P) = \Phi^{-1}(Y, \Lambda) = (\varphi(Y), \varphi(Y, \Lambda)) . \quad (12)$$

In particular,

$$\theta = \varphi(Y); \quad Y \in \bar{\mathcal{B}} \subset \mathcal{B} . \quad (13)$$

that is, θ , depends only on $Y \in \bar{\mathcal{B}}$. Conceptually, a general parameter estimation/extraction algorithm is generated by extending the domain of the function φ (13), initially defined on $\bar{\mathcal{B}}$, to parent manifold $\mathcal{B}$. If $\tilde{Y} \equiv \tilde{Y}(Z)$ with the range space $\mathcal{B}$, is some estimator of $Y \in \bar{\mathcal{B}}$, then the corresponding parameter estimator in this generic subspace based approach is

$$\hat{\theta} = \varphi(\tilde{Y}) \quad (14)$$

where the mapping φ , (13), is here interpreted as having the extended domain $\mathcal{B}$. For harmonic signals on regular array geometries R-D Unitary ESPRIT [3] is an example implementation of such a general subspace-based estimation program. In this approach the subspace estimate $\tilde{Y}$ is the p -dominant eigenvector matrix of the SCM as computed by the EVD. This subspace estimate, $\tilde{Y}$, although an ML estimate for the unconstrained

assumption, $\mathcal{B}$, is not ML when constraints exist and $Y \in \bar{\mathcal{B}}$. The ML invariance principle ([15]) which states that the ML property is preserved under transformations, is therefore not applicable.

To improve subspace-based parameter estimation it is therefore desirable to find an ML estimator of $Y \in \bar{\mathcal{B}} \subset \mathcal{B}$, or failing that, one that more closely approximates to the constrained subspace CRB than does standard EVD estimate $\tilde{Y}$. For the one-dimensional uniform line array (ULA) problem the SVD of the forward-backward data matrix yields an improved subspace estimate. This approach, which makes partial use of the constraint information, is extended to the multidimensional case in the Higher-Order SVD (HO-SVD) [16] approach. More computationally intensive approaches like Structured Least Squares (SLS) [3] and Improved SLS (I-SLS) [17] make further implicit, but still incomplete, use of the $\bar{\mathcal{B}}$ submanifold constraint information. As a result, although they yield lower variance subspace estimators compared to the standard EVD, they are not ML estimators of $Y \in \bar{\mathcal{B}}$. The associated subspace-based parameter estimates are therefore degraded to varying degrees in stressful scenarios such as that of closely spaced sources and/or highly correlated sources [12].

In parameter estimation applications, a standard approach to achieving MLE performance is to use a sub-optimal estimator (e.g., R-D Unitary ESPRIT) to initialize a Newton's method approach that uses the full signal model. The success of this type of approach will depend, in part, on the quality of the initial estimate. As an alternative, we introduce here a subspace-based estimation algorithm for $Y \in \bar{\mathcal{B}}$ that is derived from the ML criterion for the damped exponential signal assumption. The approach leverages $\tilde{Y}$, the ML estimate for the *unconstrained* problem, and is notionally implemented using a set of normal coordinates on $\mathcal{B}$ with origin at $\tilde{Y}$. The overall approach may be viewed as an implementation of a geometric Newton's method on $\bar{\mathcal{B}}$ that avoids the need for explicit initialization by incorporating the submanifold constraint condition

(i.e., shift-invariance) in a more integrated way. The algorithm developed here is closely related to the approach first proposed in [18]. Given this ML estimator of $Y \in \bar{\mathcal{B}}_{damped}$ the corresponding ML parameter estimator (which here consists of angle and damping parameters) is found via the transformation (14).

Different signal waveform assumptions yield different submanifolds of $\mathcal{B}$. Denoting the submanifold corresponding to the damped and undamped signal assumptions as $\bar{\mathcal{B}}_{damped}$ and $\bar{\mathcal{B}}_{harmonic}$, respectively, we have the set relation

$$\bar{\mathcal{B}}_{harmonic} \in \bar{\mathcal{B}}_{damped} \in \mathcal{B} \quad (15)$$

As a practical application, we apply the asymptotically ML estimator developed here for the $\bar{\mathcal{B}}_{damped}$ assumption to problem of DOA-estimation for which the $\bar{\mathcal{B}}_{harmonic}$ submanifold applies. This new subspace-based estimation approach to the DOA problem is discussed in detail in chapter 5 and simulation results are shown for a variety of stressful scenarios, in which other methods are severely degraded, to demonstrate asymptotic efficiency.

1.3 Organization

Chapter 2 provides the required background in differential geometry and related topics required to develop the CRB for the Basis manifold formulation of the complex linear model. The chapter begins with brief discussion of the geometry of the sphere using notation consistent with later developments. The discussion of the Basis manifold is a straight forward extension of [19] to the complex case. Readers familiar with differential geometry and/or with [19], which gives a differential geometric foundation for numerical linear algebra, may wish to skim this chapter. In chapter 3 the complex linear statistical model is analyzed using this foundation and intrinsic CRBs are developed for rotational and subspace estimation accuracy for unconstrained and constrained signal models. A connection is made between the full inverse FIM for the unconstrained problem and the distribution of the eigenvectors of a sample covariance matrix given in [7].

Intrinsic bounds are developed for the constrained submanifold case and their relation to existing parameter bounds available for particular signal models is examined.

In chapter 4 a new subspace estimation algorithm is developed based on this differential geometric formulation for an important special case arising in array processing applications. In chapter 5 various estimation packages are compared to these intrinsic bounds for a variety of stressful scenarios. The highly correlated signal case often presented in the literature as an algorithm stress test is examined in detail.

List of References

- [1] H. Krim and M. Viberg, “Two decades of array signal processing research,” *IEEE Signal Processing Magazine*, pp. 67–94, July 1996.
- [2] S. Smith, “Covariance, subspace, and intrinsic cramér-rao bounds,” *IEEE Trans. Signal Process.*, vol. 53, no. 5, p. 1610 1630, 2005.
- [3] M. Haardt and G. D. Galdo, “Higher-order svd based subspace estimation to improve the parameter estimation accuracy in multi-dimensional harmonic retrieval problems,” *IEEE Trans. Signal Processing*, vol. 56, p. 3198 3213, July 2008.
- [4] C.F.Gauss, *General Investigations of Curved Surfaces*. New York: Raven Press, 1965.
- [5] “Differential geometry of surfaces.” [Online]. Available: https://en.wikipedia.org/wiki/Weingarten_map
- [6] A. Boumal, “On intrinsic cramér-rao bounds for riemannian submanifolds and quotient manifolds,” *IEEE Transactions on Signal Processin*, vol. 61, no. 7, April 2013.
- [7] D. J. Jeffries and D. R. Farrier, “Asymptotic results for eigenvector methods,” *Proc. Inst. Elec. Eng.*, vol. 132, no. 7, pp. 589 –594, June 1985.
- [8] B. D. Rao and K. V. S. Hari, “Performance analysis of esprit and tam in determining the direction of arrival of plane waves in noise,” *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 37, pp. 1990 –1995, December 1989.
- [9] Z. Xu, “Perturbation analysis for subspace decomposition with applications in subspace-based algorithms,” *IEEE Transactions on Signal Processing*, vol. 50, no. 11, p. 2820 2830, Nov 2002.
- [10] F. Roemer and M. Haardt, “Analytical performance assessment of 1-d structured least squares,” in *ICASSP Proceedings*, vol. 56, July 2011, p. 3198 3213.

- [11] F. Roemer and M. Haardt, "Performance analysis of esprit-type algorithms for strictly non-circular sources using structured least squares," in *Computational Advances in Multi-Sensor Adaptive Processing*, vol. 56, 2013.
- [12] L. L. Scharf and T. McWhorter, "Geometry of the cramér-rao bound," *Signal Process.*, vol. 31, p. 301 311, 1993.
- [13] P. Stoica and A. Nehorai, "Music, maximum likelihood, and cramér-rao bound," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, no. 5, p. 720 741, May 1989.
- [14] P. Stoica, E. Larson, and A. Gershman, "The stochastic crb for array processing: A textbook derivation," *IEEE Transactions on Signal Processing Letters*, vol. 8, no. 5, pp. 148–150, May 2001.
- [15] L. L. Scharf, *Statistical Signal Processing: Detection, Estimation, and Time Series Analysis*. Reading, MA: Addison-Wesley, 1991.
- [16] M. Haardt, R. S. Thoma, and A. Richter, *Multi-dimensional highresolution parameter estimation with applications to channel sounding*. New York, NY: Marcel Dekker, June 2004, ch. 5, p. 255 338.
- [17] R. Maurice, "Improved structured least squares for the application of unitary esprit to cross arrays," *IEEE Signal Processing Letters*, vol. 13, no. 6, pp. 349–352, 2006.
- [18] R. Vaccaro and Y. Ding, "A new state space approach for direction finding," *IEEE Transactions on Signal Processing*, pp. 3234–3237, November 1994.
- [19] A. Edelman, T. A. Arias, and S. Smith, "The geometry of algorithms with orthogonality constraints," *SIAM Matrix Anal. Appl.*, vol. 20, no. 2, p. 303353, 1998.

CHAPTER 2

Differential Geometry for Signal Processing Estimation Applications

2.1 Geometry of the Sphere

A differentiable manifold is one of the most fundamental concepts in mathematics and physics. We begin by considering the familiar 3-dimensional Euclidean space denoted as $\mathcal{E}$. We take the many properties of this space as given including its inner product space structure and related norm. Each point $\mathcal{P} \in \mathcal{E}$ may be assigned a real 3-tuple $Y = (y^1, y^2, y^3)$ that are said to be the Cartesian coordinates of $\mathcal{P}$. The 3-tuple Y is an element of the real 3-space $R^3 = R \times R \times R$. The assignment is a mapping, denoted here by the symbol y , from the manifold $\mathcal{E}$ to R^3

$$y : \mathcal{E} \rightarrow R^3 \quad (16)$$

or

$$Y = (y^1, y^2, y^3) = y(\mathcal{P}), \quad \mathcal{P} \in \mathcal{E} . \quad (17)$$

This assignment process, or labeling of the point $\mathcal{P}$, is not unique and there a number of common coordinates systems (e.g., spherical, affine) that are often used.

The space $\mathcal{E}$ is also an inner product space (a vector space with a inner product). If $\mathbf{v}$ and $\mathbf{w}$ are vectors represented in an orthonormal basis $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$ as $\mathbf{v} = \sum_{k=1}^3 \mathbf{e}_k v^k$ and $\mathbf{w} = \sum_{k=1}^3 \mathbf{e}_k w^k$ then the inner product is

$$\langle \mathbf{v}, \mathbf{w} \rangle = \mathbf{v}^t \mathbf{w} = \sum_{k=1}^3 v^k w^k . \quad (18)$$

The associated norm is

$$\|\mathbf{v}\|^2 = \langle \mathbf{v}, \mathbf{v} \rangle = \mathbf{v}^t \mathbf{v} = \sum_{k=1}^3 (v^k)^2 = \|v\|^2 . \quad (19)$$

The pair $(\mathcal{E}, \langle, \rangle)$ provides the simplest example of a Riemannian manifold. If $\mathcal{P}$ and $\mathcal{O}$ are two points in $\mathcal{E}$ and $\mathbf{v}$ is the vector with tail at $\mathcal{O}$ and head at $\mathcal{P}$, then the distance

between these points, denoted $d(\mathcal{P}, \mathcal{O})$, is given by $d(\mathcal{P}, \mathcal{O}) = \|\mathbf{v}\|$. In a different set of Cartesian coordinates obtained by translating or rotating the first the set, the 3-tuples representing both $\mathcal{P}$ and $\mathcal{O}$ will of course change. The distance $d(\mathcal{O}, \mathcal{P})$, however, will remain fixed independent of the choice of Cartesian coordinate system.

In general, a Riemannian manifold, $(\mathcal{M}, g_{\mathcal{P}})$, is roughly defined as a smooth manifold $\mathcal{M}$, equipped with an inner product or metric function $g_{\mathcal{P}}$ on the tangent space $T_{\mathcal{P}}\mathcal{M}$ at each point $\mathcal{P}$ that varies smoothly from point to point. For the Euclidean space example $g_{\mathcal{P}} \equiv \langle \cdot, \cdot \rangle$.

Smooth closed subsets of $\mathcal{E}$ inherit the metric function $\langle \cdot, \cdot \rangle$ and are thereby also Riemannian manifolds. The important example considered here is the usual 2-sphere. The approach in this section will be to discuss the geometry of the sphere “over notating” things and employing the notation used later when we consider signal processing manifolds. This approach makes things more cumbersome for the sphere, but makes the transition to signal processing applications more natural.

We consider a sphere as a set embedded in $\mathcal{E}$. The set of all points equi-distant from some fixed point, say $\mathcal{O}$, in this 3-dimensional point defines the 2 dimensional sphere S^2 . In particular, the unit sphere is the subset

$$S^2 = \{\mathcal{P} \in \mathcal{E} : d(\mathcal{P}, \mathcal{O}) = 1\} . \quad (20)$$

Let y denote a Cartesian coordinate system with origin at $\mathcal{O} \in \mathcal{E}$, that is $y(\mathcal{O}) = (0, 0, 0)$, and let the coordinates of $\mathcal{P}$ under y be denoted Y , that is

$$Y = y(\mathcal{P}); \quad Y \in R^3 . \quad (21)$$

Then, expressed in these coordinates, S^2 is defined as

$$y(S^2) = \{Y \in R^3 : Y^t Y = 1\} . \quad (22)$$

Figure 4. Tangent and Normal Space

Tangent Plane

The collection of all vectors with origin or “head” at the point $\mathcal{P}_0 \in \mathcal{E}$ is denoted $T_{\mathcal{P}_0}\mathcal{E}$. This is a 3-dimensional inner product space with the usual Euclidean space inner product (19). The tangent plane to the sphere at a point $\mathcal{P}_0$ is illustrated in figure 1. This 2 dimensional vector space is denoted as $T_{\mathcal{P}_0}(S^2)$ and is, evidently, a subspace of $T_{\mathcal{P}_0}\mathcal{E}$.

If $\mathcal{P}_0$ is the north pole, then the Cartesian coordinates, $Y_0 \equiv y(\mathcal{P}_0)$, are

$$Y_0 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}, \quad (23)$$

and the submanifold tangent space $T_{\mathcal{P}_0}(S^2)$ is spanned by the two vectors

$$\mathbf{b}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}; \mathbf{b}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}. \quad (24)$$

An arbitrary tangent vector $\Delta \in T_{\mathcal{P}_0}(S^2)$ expands as

$$\Delta = \mathbf{b}_1\beta^1 + \mathbf{b}_2\beta^2. \quad (25)$$

If we define $Y_{0\perp}$ as

$$Y_{0\perp} \equiv \begin{bmatrix} \mathbf{b}_1 & \mathbf{b}_2 \end{bmatrix} \equiv \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} \quad (26)$$

this Δ may alternatively be represented in numerical form as

$$\Delta = Y_{0\perp} B \quad (27)$$

where $B = [\beta^1 \beta^2]^t$ (this seemingly awkward representation of the vector Δ is used extensively in our later signal processing examples). Note that the defining requirement for $Y_{0\perp}$ is that $Y_0^t Y_{0\perp} = 0$, that is $Y_{0\perp}$ is the orthogonal to Y_0 , so that it may more generally be expressed as

$$Y_{0\perp} = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \\ 0 & 0 \end{bmatrix} \quad (28)$$

for θ arbitrary.

The dimension of the tangent space is equal to the number of linearly independent basis vectors required to span the space, or equivalently here the degrees of freedom in B , so that we see $\dim(T_{\mathcal{P}_0}(S^2)) = 2$. While trivial here, this approach is important in determining the dimension of the various tangent spaces encountered in later signal processing examples.

The orthogonal complement of $T_{\mathcal{P}_0}(S^2)$ is the normal space, denoted $N_{\mathcal{P}_0}$, yielding the decomposition

$$T_{\mathcal{P}_0}(\mathcal{E}) = T_{\mathcal{P}_0}(S^2) \oplus N_{\mathcal{P}_0} \quad (29)$$

For an arbitrary $\Delta \in T_{\mathcal{P}_0}(\mathcal{E})$ we have $P_\perp \Delta \in T_{\mathcal{P}_0}(S^2)$, where $P_\perp$ denotes the orthogonal projection operator unto $T_{\mathcal{P}_0}(S^2)$. The metric on the tangent space $T_{\mathcal{P}_0}(S^2)$ is simply the restriction of the ambient space metric to this tangent plane, that is

$$\langle \Delta_1, \Delta_2 \rangle|_{sphere} \equiv \Delta_1^t P_\perp \Delta_2 . \quad (30)$$

Given this metric the pair $\mathbf{b}_k = Y_{0\perp} B_k; k = 1, 2$ are orthogonal if

$$\langle \mathbf{b}_1, \mathbf{b}_2 \rangle|_{sphere} = B_1^t B_2 = 0 . \quad (31)$$

The standard basis for $T_{\mathcal{P}_0}(S^2)$ given an arbitrary valid choice for $Y_{0\perp}$ is defined by

$$B_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}; B_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \quad (32)$$

This choice of B_1, B_2 holds for every $\mathcal{P} \in S^2$ in the sense that for $Y = y(\mathcal{P})$ and $Y_{\perp}^t Y = 0$ the pair $Y_{\perp} B_1, Y_{\perp} B_2$ is an orthogonal basis for $T_{\mathcal{P}}(S^2)$

In such an orthonormal basis, the components of (25) are $\beta^k = \langle \mathbf{b}_k, \Delta \rangle$. This sum has the concrete numerical form

$$\Delta = Y_{0\perp} B_1 \beta^1 + Y_{0\perp} B_2 \beta^2 \quad (33)$$

so that $\Delta = Y_{0\perp} B$ where

$$B = B_1 \beta^1 + B_2 \beta^2 . \quad (34)$$

This tangent vector may also be expressed as

$$\Delta = U \sigma \quad (35)$$

where U and σ represent the pointing direction and length of Δ , respectively. The unit vector U is $U = Y_{0\perp} U_1$ for some 2-tuple U_1 . In later applications, the tangent vectors are represented in $N \times p$ matrix form and the above will correspond to the compact SVD of Δ (that is $\Delta = U \Sigma V^H$).

Tangents to curves

The tangent space at a point is alternately described by the collection of tangents to all curves passing through the point. If $Y(t)$ is a smooth curve on the sphere then at every point

$$Y(t)^t Y(t) - 1 = 0 . \quad (36)$$

Differentiating, and using the notation $\Delta = \dot{Y}$, yields

$$Y^t \Delta + \Delta^t Y = 0 . \quad (37)$$

At the point $Y_0 = Y(0)$ on this curve the above equation is satisfied by choosing

$$\Delta = Y_{0\perp} B \quad (38)$$

where B is arbitrary real 2-tuple and where $Y_0^t Y_{0\perp} = 0$.

Geodesics and Great Circles

A geodesic on a manifold is defined as the shortest path between two given points. For the Euclidean space geodesics are straightlines which may be specified by the familiar differential equation

$$\ddot{Y} = 0 . \quad (39)$$

This is the Euler-Lagrange or geodesic equation for the flat Euclidean space, represented in Cartesian coordinates, and has the general solution $Y(t) = Y_0 + Vt$.

The analogous geodesic equation in the general case is derived rigorously using the Calculus of Variations. For the sphere a simplified approach that leads to the geodesic equation is differentiating the defining relation (36) a second time. This approach yields

$$2\dot{Y}^t\dot{Y} + Y^t\ddot{Y} + \ddot{Y}^tY = 0 . \quad (40)$$

Direct substitution verifies this is equivalent to

$$\ddot{Y} = -Y(\dot{Y}^t\dot{Y}) . \quad (41)$$

This geodesic equation has the general solution

$$Y(t) = Y_0\cos(\sigma t) + U\sin(\sigma t) , \quad (42)$$

alternately expressed in matrix form as

$$Y(t) = \begin{bmatrix} Y_0 & U \end{bmatrix} \begin{bmatrix} \cos(\sigma t) \\ \sin(\sigma t) \end{bmatrix} \quad (43)$$

where Y_0 , U , σ specify initial location, launch direction and speed, respectively. This is referred to as the CS form of the geodesic. To verify this solution, note that the first derivative of $Y(t)$ is

$$\dot{Y}(t) = \sigma(-Y_0\sin(\sigma t) + U\cos(\sigma t)) \quad (44)$$

and that the 2nd derivative is

$$\ddot{Y}(t) = (\sigma^2(-Y_0\cos(\sigma t) - U\sin(\sigma t))) = -\sigma^2Y(t) . \quad (45)$$

The factor $\dot{Y}^t \dot{Y}$, the squared speed, is the constant

$$\dot{Y}^t(t) \dot{Y}(t) = \sigma^2(\sin^2(\sigma t) + \cos^2(\sigma t)) = \sigma^2 \quad (46)$$

Substituting these results into (40) verifies the solution.

Since Y is the radial vector from the origin to the surface of the sphere, the vector at the point Y extending in this direction is normal to the surface. From (45) we observe that the acceleration $\ddot{Y}$ is normal to the surface at every point on the geodesic and that the tangential acceleration is zero. In the case of the sphere, acceleration for uniform motion on a great circle is directed radially and therefore normal to the surface; therefore, great circles are geodesics on the sphere.

Initial Value Problem: The initial value problem (IVP) is specified by the differential equation (41) together with the initial conditions

$$Y(0) = Y_0; \quad \dot{Y}(0) = \sigma U. \quad (47)$$

In these expressions U is a unit tangent vector at Y_0 , that is $U \in T_{Y_0} S^2$, and σ is the scalar speed. This IVP has the solution (42) where the parameters U, Y_0 , and σ are specified as initial conditions. For unit speed (42) is said to be the geodesic emanating from the point $\mathcal{P}_0$ in the direction of $U \in T_{\mathcal{P}_0} S^2$.

Example: IVP at the North Pole : If $\mathcal{P}_0$ is the north pole in the given y coordinates ($Y_0 = [0, 0, 1]^t$), then an arbitrary tangent vector $\Delta = \sigma U$ in $T_{\mathcal{P}_0} S^2$ has the form

$$U = \begin{bmatrix} \cos(\gamma) \\ \sin(\gamma) \\ 0 \end{bmatrix} \quad (48)$$

where γ specifies direction in the tangent space; σ is the magnitude of the vector. Substituting this form into (42) yields

$$Y(t) = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \cos(\sigma t) + \begin{bmatrix} \cos(\gamma) \\ \sin(\gamma) \\ 0 \end{bmatrix} \sin(\sigma t), \quad (49)$$

and carrying out the sum gives

$$Y(t) = \begin{bmatrix} \cos(\gamma)\sin(\sigma t) \\ \sin(\gamma)\sin(\sigma t) \\ \cos(\sigma t) \end{bmatrix}. \quad (50)$$

For $t = 1$ and $Y_1 \equiv Y(1)$ we have

$$Y_1 = \begin{bmatrix} \cos(\gamma)\sin(\sigma) \\ \sin(\gamma)\sin(\sigma) \\ \cos(\sigma) \end{bmatrix} \quad (51)$$

Differentiating (42) we get the tangent to the geodesic curve at $Y(t)$ given by

$$\Delta(t) = \frac{dY(t)}{dt} = \sigma \begin{bmatrix} \cos(\gamma)\cos(\sigma t) \\ \sin(\gamma)\cos(\sigma t) \\ -\sin(\sigma t) \end{bmatrix}. \quad (52)$$

Exponential Map

The exponential map plays a prominent role in our later applications. The following lemma relates the geodesic specified by (42) to an equivalent exponential form.

Lemma 1. *The general solution to the geodesic equation given by (43)*

$$Y(t) = \begin{bmatrix} Y_0 & U \end{bmatrix} \begin{bmatrix} \cos(\sigma t) \\ \sin(\sigma t) \end{bmatrix} \quad (53)$$

is equal to the exponential map form defined by

$$Y(t) = \begin{bmatrix} Y_0 & Y_{0\perp} \end{bmatrix} \exp\left(t \begin{bmatrix} 0 & -B^H \\ B & 0 \end{bmatrix}\right) I_{3,1} \quad (54)$$

where $Y_{0\perp}$ is such that $Y_0^t Y_{0\perp} = 0$ and

$$B = Y_{0\perp}^t U \sigma \quad (55)$$

so that $Y_{0\perp} B = U \sigma$. In this expression $I_{3,1} = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix}^t$,

Proof: Consider the tangent vector $\Delta \in T_{Y_0} S^2$, at the point Y_0 , in the form $\Delta = \sigma U$. Let $B = Y_{0\perp}^t \Delta$ where $Y_{0\perp}$ is a 3×2 semi-orthogonal matrix such that $Y_{0\perp}^t Y_0 = 0$. Let B be represented by its SVD

$$B = U_1 \sigma = \begin{bmatrix} U_1 & U_2 \end{bmatrix} \begin{bmatrix} \sigma \\ 0_{2 \times 1} \end{bmatrix} \quad (56)$$

where U_1 is 2×1 and U_2 is 2×2 . Then

$$U = Y_{0\perp} U_1. \quad (57)$$

Using (56) the argument of the matrix exponential (54) is

$$\begin{bmatrix} 0 & -B^H \\ B & 0_{2 \times 2} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & U_1 & U_2 \end{bmatrix} \begin{bmatrix} 0 & -\sigma & 0 \\ \sigma & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & U_1^H \\ 0 & U_2^H \end{bmatrix}, \quad (58)$$

$$= \begin{bmatrix} 1 & 0 & 0 \\ 0 & U_1 & U_2 \end{bmatrix} X \begin{bmatrix} i\sigma & 0 & 0 \\ 0 & -i\sigma & 0 \\ 0 & 0 & 0 \end{bmatrix} X^{-1} \begin{bmatrix} 1^t & 0 \\ 0 & U_1^H \\ 0 & U_2^H \end{bmatrix}, \quad (59)$$

where

$$X = \begin{bmatrix} X_0 & 0 \\ 0 & 1 \end{bmatrix}; \quad X_0 = i \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}, \quad (60)$$

and where we have used the fact that

$$\begin{bmatrix} 0 & -\sigma \\ \sigma & 0 \end{bmatrix} = X_0 \begin{bmatrix} i\sigma & 0 \\ 0 & -i\sigma \end{bmatrix} X_0^H. \quad (61)$$

Using the general matrix exponential identity $\expm(U\Lambda V^H) = U\expm(\Lambda)V^H$ and (59) the matrix exponential in (54) expands as

$$= \begin{bmatrix} 1 & 0 & 0 \\ 0 & U_1 & U_2 \end{bmatrix} \begin{bmatrix} \cos(\sigma) & -\sin(\sigma) & 0 \\ \sin(\sigma) & \cos(\sigma) & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & U_1^H \\ 0 & U_2^H \end{bmatrix}, \quad (62)$$

where we have used the fact that

$$X \begin{bmatrix} \exp(i\sigma) & 0 & 0 \\ 0 & \exp(-i\sigma) & 0 \\ 0 & 0 & 0 \end{bmatrix} X^H = \begin{bmatrix} \cos(\sigma) & \sin(\sigma) & 0 \\ \sin(\sigma) & \cos(\sigma) & 0 \\ 0 & 0 & 0 \end{bmatrix}. \quad (63)$$

Post multiplying of (62) by the selection matrix $I_{3,1}$ yields

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & U_1 & U_2 \end{bmatrix} \begin{bmatrix} \cos(\sigma) \\ -\sin(\sigma) \\ 0 \end{bmatrix} = \begin{bmatrix} \cos(\sigma) \\ U_1 \sin(\sigma) \end{bmatrix}. \quad (64)$$

Premultiplication by $\begin{bmatrix} Y_0 & Y_{0\perp} \end{bmatrix}$, letting $\sigma = \sigma t$ and noting (57) yields

$$Y(t) = \begin{bmatrix} Y_0 & U \end{bmatrix} \begin{bmatrix} \cos(\sigma t) \\ \sin(\sigma t) \end{bmatrix}. \quad (65)$$

The desired form (53) is recovered and this establishes equivalence with the exponential map form (54).

The short hand notation for the exponential map (54) is

$$Y = \exp_{Y_0}(t\Delta) \quad (66)$$

where the tangent vector $\Delta = Y_{0\perp}B$. The exponential map defined in this way is seen to be a map from the tangent space at $\mathcal{P}_0$, $T_{\mathcal{P}_0}S^2$, to the manifold S^2 .

Inverse Exponential Map, Normal Coordinates and the Two-Point Boundary Value Problem

The principal inverse of the exponential map, termed the inverse exponential map, is therefore a mapping from S^2 to $T_{\mathcal{P}_0}S^2$. Given $\mathcal{P}_0$ and the tangent vector Δ , let $\mathcal{P}_1 = \exp_{\mathcal{P}_0}\Delta$. The inverse exponential map then is

$$\Delta = \exp_{\mathcal{P}_0}^{-1}\mathcal{P}_1 . \quad (67)$$

Given an arbitrarily chosen orthonormal basis $\mathbf{b}_1, \mathbf{b}_2$ for $T_{\mathcal{P}_0}S^2$ the tangent vector $\Delta \in T_{\mathcal{P}_0}S^2$ expands as

$$\Delta = \mathbf{b}_1\beta^1 + \mathbf{b}_2\beta^2 \quad (68)$$

where

$$\beta^1 = \langle \mathbf{b}_1, \Delta \rangle \quad \beta^2 = \langle \mathbf{b}_2, \Delta \rangle \quad (69)$$

Suppose that $Y_0 = y(P_0)$ and $Y_1 = y(Y_1)$ are the Cartesian coordinates of these points and that $\mathbf{b}_1 = Y_{0\perp}B_2$, $\mathbf{b}_2 = Y_{0\perp}B_1$. The above procedure that defines the 2-tuple β provides an alternate means of labeling points on the sphere. Let the transformation between these two set of coordinates, Y to β , which consists of computing the tangent vector Δ and its components β with respect to the given basis, be denoted as

$$\beta = \varphi(Y) = \varphi(y(\mathcal{P})) . \quad (70)$$

In general coordinates defined via the exponential map are termed normal coordinates.

Polar normal coordinates are the pair (σ, γ) , where

$$\sigma = \sqrt{(\beta^1)^2 + (\beta^2)^2}; \quad \gamma = \text{atan2}(\beta^2, \beta^1) , \quad (71)$$

or

$$\beta^1 = \sigma \cos(\gamma); \quad \beta^2 = \sigma \sin(\gamma) . \quad (72)$$

The coordinate transformation from the Y coordinates to these polar normal coordinates is denoted as

$$(\sigma, \gamma) = \psi(Y) = \psi(y(\mathcal{P})) . \quad (73)$$

The transformation from normal to the Cartesian coordinates is therefore $Y = y \circ \varphi^{-1}(\beta)$ and defined by exponential map (54). Using the equivalent CS form of the exponential mapping given by (42) with $t = 1$ we have

$$Y_1 = Y(1) = \begin{bmatrix} \cos(\gamma) \sin(\sigma) \\ \sin(\gamma) \sin(\sigma) \\ \cos(\sigma) \end{bmatrix} \quad (74)$$

An alternate definition of the sphere is given parameterically as

$$S^2 = \left[Y \in \mathcal{E} \text{ such that } Y = \psi^{-1}(\gamma, \sigma) \right] . \quad (75)$$

Two point boundary value problem

In the two point boundary value problem (BVP) the auxiliary conditions are $Y(0) = Y_0$ and $Y(1) = Y_1$. The objective here is to determine the initial velocity (i.e., speed σ and direction U), or tangent vector, required to move from Y_0 to Y_1 in unit time (in our later applications this procedure is analogous to determining the tangent vector Δ that connects Y_0 to Y_1 via a geodesic). The required initial velocity vector is determined from the boundary conditions. From (42) with $t = 1$

$$Y_1 = Y(1) = Y_0 \cos(\sigma) + U \sin(\sigma) \quad (76)$$

so that

$$Y_0^t Y_1 = \cos(\sigma) . \quad (77)$$

The speed σ then is

$$\sigma = \arccos(Y_0^t Y_1) \quad (78)$$

and the unit tangent vector direction is

$$U = \frac{Y_1 - Y_0 \cos(\sigma)}{\sin(\sigma)} = \frac{Y_1 - Y_0 Y_0^t Y_1}{\sin(\sigma)} \quad (79)$$

$$U = \frac{(I - Y_0 Y_0^t) Y_1}{\sin(\sigma)} . \quad (80)$$

The resultant tangent vector that satisfies this BVP is $\Delta = U\sigma$.

Example: 2 point BVP: Let Y_0 be the North Pole $Y_0 = \begin{bmatrix} 0 & 0 & 1 \end{bmatrix}^t$ and let Y_1 be an arbitrary point on the equator $Y_1 = \begin{bmatrix} \cos(\gamma) & \sin(\gamma) & 0 \end{bmatrix}^t$. Since $Y_0^t Y_1 = 0$ (78) yields $\sigma = \frac{\pi}{2}$ and from (80) $U = Y_1 \sin^{-1}(\sigma) = Y_1$ so that

$$\Delta = \frac{\pi}{2} Y_1 \quad (81)$$

The polar normal coordinates of Y_1 are therefore $(\frac{\pi}{2}, \gamma)$. The standard normal coordinates are $\beta^1 = \frac{\pi}{2} \cos(\gamma)$ and $\beta^2 = \frac{\pi}{2} \sin(\gamma)$ so that using (68)

$$\Delta = \mathbf{b}_1 \beta^1 + \mathbf{b}_2 \beta^2 = \frac{\pi}{2} Y_1 \quad (82)$$

as expected.

Coordinate Curves

Let $\mathbf{b}_1$ and $\mathbf{b}_2$ be an orthonormal basis for the tangent space at north pole represented in the extrinsic frame as $\mathbf{b}_1 = [1, 0, 0]^t$ and $\mathbf{b}_2 = [0, 1, 0]^t$. The extrinsic coordinate representation of the normal coordinate (geodesic) curve emanating from the Y_0 in the $\mathbf{b}_1$ direction is denoted $Y_{(1)}(t)$ and is given by

$$Y_{(1)}(t) = \begin{bmatrix} \sin(t) & 0 & \cos(t) \end{bmatrix}^t . \quad (83)$$

The tangent at an arbitrary time t is

$$\frac{dY_{(1)}(t)}{dt} = \begin{bmatrix} \cos(t) & 0 & -\sin(t) \end{bmatrix}^t . \quad (84)$$

At $t = \frac{\pi}{2}$ we have $Y_{(1)}(\frac{\pi}{2}) = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix}^t$ which corresponds to a point on the equator. The tangent to this coordinate curve at $t = \frac{\pi}{2}$ evaluates to $[0, 0, -1]^t$. This vector is attached to the point $Y_{(1)}$ and pointing in a downward direction.

Similarly, the normal coordinate (geodesic) curve emanating in the $\mathbf{b}_2 = \begin{bmatrix} 0 & 1 & 0 \end{bmatrix}^t$ direction is denoted $Y_{(2)}(t)$ and is given by

$$Y_{(2)}(t) = \begin{bmatrix} 0 & \sin(t) & \cos(t) \end{bmatrix}^t, \quad (85)$$

and the tangent vector is

$$\frac{dY_{(2)}(t)}{dt} = \begin{bmatrix} 0 & \cos(t) & -\sin(t) \end{bmatrix}^t. \quad (86)$$

At $t = \frac{\pi}{2}$ we have $Y_{(2)}(\frac{\pi}{2}) = \begin{bmatrix} 0 & 1 & 0 \end{bmatrix}^t$ which corresponds to a point on the equator. The tangent to this coordinate curve at $t = \frac{\pi}{2}$ evaluates to $[0, 0, -1]^t$.

Natural Geodesic Distance Between Points

The running length of the arbitrary smooth curve $Y(t)$ emanating from the point Y_0 is given by elementary calculus as

$$Length(Y(t), Y_0) = \int_{t_0}^{t_1} \sqrt{\langle \Delta(t), \Delta(t) \rangle} dt = \int_{t_0}^{t_1} \|\Delta(t)\| dt \quad (87)$$

where $\Delta(t) = \frac{dY(t)}{dt}$.

The geodesic distance between two points on the sphere is the length of the shortest curve connecting the two points. The distance between Y_1 and Y_0 is then defined as

$$dist_{S^2}(Y_1, Y_0) \equiv \min_{Y(t)} (Length(Y_1, Y_0)) \quad (88)$$

minimized over all curves $Y(t) \in S^2$ between Y_0 and Y_1 .

If $Y(t)$ is a geodesic then the tangent at each point on the curve is given by (52). The magnitude of $\Delta(t)$ is constant along the curve with $\|\Delta(t)\| = \|\Delta(0)\| = \sigma$. Using this expression in the integral (140) yields

$$Length(Y(t), Y_0) = \|\Delta(0)\| |t| = \sigma |t| \quad (89)$$

If $Y_1 = Y(t_1)$ then

$$dist_{S^2}(Y_1, Y_0) = \sigma |t_1| \quad (90)$$

On the unit sphere this arc is a great circle, and the arc length is the angle in radians separating the two end points ($\sigma = \frac{\text{radians}}{\text{sec}}$).

Remark: If Y_0 is the North Pole and $Y(t)$ is the geodesic given by (50), then

$$Y_0^T Y_1(t) = \cos(\sigma t) \quad (91)$$

and the geodesic distance between the points is evidently

$$\text{dist}_{S^2}(Y(t), Y_0) = \sigma |t| = \text{acos}((Y_0^T Y_1(t))) . \quad (92)$$

In the signal processing examples discussed later the subspace angles between the semi-unitary matrices Y_0 and Y_1 are defined as the *acos* of singular values of the $Y_0^H Y_1$.

2.1.1 Estimation Example on the Unit Sphere

Consider the linear model

$$z = Y n_1 + n_0; \quad Y \in S^2 \quad (93)$$

where n_0 is a 3×1 additive random noise vector $n_0 \sim N_p(0, \sigma_0^2 I)$, with σ_0 known, and n_1 is a scalar random variable

$$n_1 \sim N(0, \sigma^2) \quad (94)$$

where σ^2 is a deterministic unknown. The observation z then is a 3×1 normal random vector $z \sim N_n(0, R)$ where

$$R = E[zz^H] = R_s + \sigma_0^2 I \quad (95)$$

and where the signal covariance R_s is defined by

$$R_s \equiv \sigma^2 Y Y^t . \quad (96)$$

In this model Y is a deterministic unknown and represents the point on unit sphere. Let $Z = \begin{bmatrix} z_1, & \cdots, & z_K \end{bmatrix}$ be a $3 \times K$ data matrix whose columns are independent and

identically distributed (iid) random vectors defined by (93). The log-likelihood function is (ignoring constants)

$$L(Y, \sigma_p^2; \sigma_n^2) = -\frac{K}{2}(\text{tr}(R^{-1}\tilde{R}) + \log \det R) \quad (97)$$

where $\tilde{R}$ is the sample covariance matrix (SCM) $\tilde{R} = K^{-1}ZZ^t$. Representing Y in normal coordinates (74) yields an alternate parameterization of this function as $L(\beta) = L(Y)$.

Let $\hat{Y}$ denote an estimator of $Y \in S^2$ and let $\hat{\Delta} = \exp_Y^{-1}\hat{Y}$ be the estimation error tangent vector. For $\mathbf{b}_1, \mathbf{b}_2$ an orthonormal basis suppose that this stochastic error vector expands as

$$\hat{\Delta} = \mathbf{b}_1\hat{\beta}^1 + \mathbf{b}_2\hat{\beta}^2. \quad (98)$$

The geodesic distance error measure (92)

$$\text{dist}_{S^2}(\hat{Y}, Y) = \text{acos}(Y^t\hat{Y}) = \|\hat{\beta}\| \quad (99)$$

and

$$\epsilon = E \left[\text{dist}_{S^2}(\hat{Y}, Y)^2 \right]^{\frac{1}{2}} = E \left[\|\hat{\beta}\|^2 \right]^{\frac{1}{2}}. \quad (100)$$

Because this distance corresponds to the norm of coefficients β , the Fisher information, denoted $F_{\beta\beta}$, and inverse Fisher information matrix(FIM) are computed using this parameterization of the log-likelihood function. The FIM for this problem is developed in [1, equation 143]) and is given by

$$[F_{\beta\beta}]_{ij} = \frac{K}{\sigma^2} \text{Re} \left(\text{tr} \left(\sigma_{OPT}^2 B_i^t B_j \right) \right); \quad i, j = 1, 2 \quad (101)$$

where

$$\sigma_{OPT}^2 = \sigma^{22} \sigma_s^{-2} = (\sigma_s^2 - \sigma_0^2 I)^2 \sigma_s^{-2} \quad (102)$$

For the standard basis (32) this is expressed in the 2×2 matrix form

$$F_{\beta\beta} = \frac{K\sigma_{OPT}^2}{\sigma_0^2} I_2 \quad (103)$$

so that

$$\text{tr} F_{\beta\beta}^{-1} = \frac{2}{K} \frac{\sigma_{OPT}^2}{\sigma_0^2} . \quad (104)$$

Ignoring curvature, the variance of any unbiased estimator $\hat{Y}$ satisfies the inequality

$$\epsilon \geq \sqrt{\text{tr} F_{\beta\beta}^{-1}} = \sqrt{\frac{2}{K} \frac{\sigma_{OPT}^2}{\sigma_0^2}} \quad (105)$$

The ML estimator of $Y \in S^2$ for this linear model is the principal eigenvector of the SCM $\tilde{R}$.

2.2 Geometry of Signal Processing Manifolds

Many estimation problems can be formulated as the minimization of a function $F(Y)$ where the parameter Y is an orthonormal matrix, that is, Y is constrained to the set of $N \times p$ matrices such that $Y^t Y = I$. This set is known as the (real) Stiefel manifold. Often the objective function of interest has some type of homogeneity property so that $F(Y_1) = F(Y_0)$ whenever Y_1 and Y_0 satisfy some type of equivalence relation, $Y_1 \sim Y_0$. For example, it may be that $F(Y) = F(YQ)$, for every Y where Q is any $p \times p$ orthogonal matrix. In other words, the objective function depends only on the subspace spanned by the columns of Y ; it does not depend on the particular basis used to represent the subspace. The set of p -dimensional subspaces in R^n is called the Grassmann manifold and so this manifold is natural parameter space or domain for the functions with this homogeneity condition. A variation for the complex case is the homogeneity condition $F(Y) = F(YD)$ where D is a any $p \times p$ unitary diagonal matrix and Y is such that $Y^H Y = I$ (such matrices are termed semi-unitary). This situation naturally arises when considering the log-likelihood function for the standard complex linear model. The collection of semi-unitary matrices corresponding to this homogeneity condition is termed here the Basis Manifold. Manifolds defined through these types of equivalence relations are termed quotient manifolds. The geometrical descriptions of complex Basis manifold and related manifolds are developed in this chapter. These quotient manifolds

are derived from the Unitary Group, $U(N)$, which is the starting point for the discussion.

2.2.1 Coordinates on Manifolds

Modern differential geometry relies on coordinate-free methods, but because the focus is on estimation algorithms, a coordinate-based approach is used from the start. A coordinate system on a manifold $\mathcal{M}$ is a mapping, say y , from the abstract manifold $\mathcal{M}$ to R^N

$$y : \mathcal{M} \rightarrow R^N . \quad (106)$$

A point $\mathcal{Y} \in \mathcal{M}$ has the coordinates $Y \in R^N$, assigned by the coordinate mapping y

$$Y = y(\mathcal{Y}) . \quad (107)$$

In the usual theoretical development the dimensions of R^N and $\mathcal{M}$ are equal but this is not required, and in our signal processing applications it is more natural initially to work with extrinsic coordinate representations.

Extrinsic coordinates correspond to the natural numerical matrix representation of elements of the Stiefel manifold. The length of the extrinsic coordinate vector is greater than the dimension of the space. For example, the complex Stiefel manifold has dimension $2pN - p^2$ but its representation as $N \times p$ complex matrices requires $2Np$ values. The dimension of the Stiefel manifold reflects the orthonormality constraint on the columns of matrix representation. As discussed in the previous section, this is the same situation that occurs when a point on the usual 2-dimensional sphere is represented by the standard (x, y, z) Cartesian coordinates of the ambient 3-dimensional space.

The alternative to such extrinsic coordinates are intrinsic coordinates in which the number of required coordinate components equals the manifold dimension (normal coordinates are a special type of intrinsic coordinates). For the case of the sphere the u, v coordinates on the 2-sphere defined by the mapping to extrinsic coordinates, $x = \cos u \sin v, y = \sin u \sin v, z = \cos v$, provide an example of intrinsic coordinates.

A more general discussion of intrinsic coordinates and their application is given in appendix A.1.

2.2.2 Unitary Group as a Riemannian Manifold

As noted in [2], “given a manifold whose geometry is well understood (where there are closed form expressions for the geodesics and, perhaps also, parallel transport), there is a very natural, efficient, and convenient way to generate closed form formulas on quotient spaces of that manifold.” This is precisely the situation with the Stiefel, Basis, and Grassmann manifolds, which are quotient spaces within the Unitary Group $\mathcal{U}(N)$. We first consider the Unitary Group and review its geometric properties. Using this geometry the geometric properties of quotient manifolds on $\mathcal{U}(N)$ are developed in a straightforward way.

The set $C^{N^2} \equiv C^{N \times N}$, the collection of all $N \times N$ complex matrices, along with the standard operations of matrix addition and multiplication by real scalars, defines a real vector space of real dimension $2N^2$. We define an inner product on this real vector space as

$$\langle X, Z \rangle = \text{Re}(\text{tr}(X^H Z)); \quad Z, X \in C^{N^2}. \quad (108)$$

The set of all vectors with origin at the arbitrary point $X_0 \in C^{N^2}$ is a vector space, denoted by $T_{X_0}C^{N^2}$, with same dimension as C^{N^2} . This is also an inner product space with the inner product (108). This inner product is a bilinear form

$$g_e = \langle , \rangle : T_{X_0}C^{N^2} \times T_{X_0}C^{N^2} \rightarrow R \quad (109)$$

defined by

$$\langle \Delta_1, \Delta_2 \rangle = \text{Re}(\text{tr}(\Delta_1^H \Delta_2)) \quad (110)$$

for $\Delta_1, \Delta_2 \in T_{X_0}C^{N^2}$.

The set C^{N^2} along with this inner product defined on the tangent spaces, $T_{X_0}C^{N^2}$, defines a Riemannian manifold. The unitary group $\mathcal{U}(N)$ is defined here as a subman-

ifold of C^{N^2} as the set of all $X \in C^{N^2}$ such that $X^H X = I$. The set $\mathcal{U}(N)$ is also a group, where the group operator is standard matrix multiplication. $\mathcal{U}(N)$ is closed with respect to the group operation, matrix multiplication, so that for $X_0 \in \mathcal{U}(N)$, $Q \in \mathcal{U}(N)$ the product $X = X_0 \cdot Q$ is also in $\mathcal{U}(N)$. Note that any unitary matrix Q can be written as $Q = \expm(W)$, where W is $N \times N$ anti-Hermitian ($W + W^H = 0$), and $\expm$ is the matrix exponential function.

Example: Unitary Matrices

Consider the Hermitian matrix W and its eigen-decomposition

$$W = \begin{bmatrix} 0 & \theta \\ -\theta & 0 \end{bmatrix} = U \begin{bmatrix} i\theta & 0 \\ 0 & -i\theta \end{bmatrix} U^H \quad (111)$$

where

$$U = i \cdot \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}. \quad (112)$$

From the basic properties of matrix exponentials [3]

$$X = \expm(W) = U \begin{bmatrix} \exp(i\theta) & 0 \\ 0 & \exp(-i\theta) \end{bmatrix} U^H \quad (113)$$

Clearly $X^H X = I$ so that X is unitary. This is a general fact. If W is anti-Hermitian, then $\expm(W)$ is a unitary matrix. Furthermore, every unitary matrix has this form and therefore the exponential map from the set of anti-Hermitian matrices to the set of unitary matrices is surjective.¹

By virtue of the embedding in the Riemannian manifold C^{N^2} , the submanifold $\mathcal{U}(N)$ is endowed with the inherited metric (108) and as such constitutes a Riemannian manifold.

Tangent Space

Intuitively, the tangent space at a point is the plane tangent to the submanifold at that point (see Figure 2.1) in the ambient space. For d -dimensional manifolds, this plane

¹The function is surjective (onto) if every element of the codomain (range) is mapped to by at least one element of the domain. That is, the image and the codomain of the function are equal.

is a d -dimensional vector space with origin at the point of tangency. The normal space is the orthogonal complement in the ambient space. In the general case the tangent space at the point $\mathcal{P} \in \mathcal{M}$ is denoted $T_{\mathcal{P}}\mathcal{M}$; $\mathcal{P}$ is the base point or origin.

For $X_0 \in \mathcal{U}(N) \subset C^{N^2}$ the set of vectors tangent to the surface $\mathcal{U}(N)$ is a vector subspace, denoted $T_{X_0}\mathcal{U}(N)$, of $T_{X_0}C^{N^2}$. The full space $T_{X_0}C^{N^2}$ partitions as

$$T_{X_0}C^{N^2} = T_{X_0}\mathcal{U}(N) \oplus (T_{X_0}\mathcal{U}(N))^{\perp} . \quad (114)$$

All elements in $T_{X_0}\mathcal{U}(N)$ have the form $\Delta = X_0W$, W is anti-Hermitian, while all elements of $(T_{X_0}\mathcal{U}(N))^{\perp}$, the normal space, have the form $\Phi = X_0S$ where S is Hermitian.

The anti-Hermitian requirement on the matrix W above may be derived from the defining property unitary matrices

$$X^H(t)X(t) = I . \quad (115)$$

In this expression $X(t)$ represents an arbitrary smooth curve on $\mathcal{U}(N)$. Differentiating yields

$$\Delta^H X + X^H \Delta = 0 \quad (116)$$

where $\Delta = \dot{X}$ is the $N \times N$ matrix derivative.

Substituting $\Delta = XW$, the SVD representation of the arbitrary matrix Δ , into (116) yields

$$W + W^H = 0 ; \quad (117)$$

that is, W is constrained to be anti-Hermitian.

Since all elements of $T_{X_0}\mathcal{U}(N)$ must have the form X_0W the real dimension of this space is equal to the degrees of freedom in the $N \times N$ anti-Hermitian matrix W so that

$$\dim(\mathcal{U}(N)) = \dim(T_{X_0}\mathcal{U}(N)) = N^2 . \quad (118)$$

To see this, note that for an $N \times N$ matrix, there are $\frac{N(N-1)}{2}$ complex elements above the diagonal. If the matrix is anti-Hermitian, then the N diagonal elements are pure imaginary. The total number of independent elements in an anti-Hermitian matrix, therefore, is

$$2 \left(\frac{N(N-1)}{2} \right) + N = (N(N-1) + N) = N^2 . \quad (119)$$

Riemannian Manifold Inner Product

The restriction of the Euclidean metric (108) to $T_{X_0}\mathcal{U}(N)$ defines an inner product on the tangent space at each $X_0 \in \mathcal{U}(N)$, denoted

$$\langle , \rangle : T_{X_0}\mathcal{U}(N) \times T_{X_0}\mathcal{U}(N) \rightarrow R \quad (120)$$

by

$$\langle \Delta_1, \Delta_2 \rangle = \text{Re}(\text{tr}(\Delta_1^H \Delta_2)) . \quad (121)$$

Since all $\Delta \in T_{X_0}\mathcal{U}(N)$ have the form $\Delta = XW$ this becomes product becomes

$$\langle XW_1, XW_2 \rangle = \text{Re}(\text{tr}(W_1^H X^H X W_2)) = \text{Re}(\text{tr}(W_1^H W_2)) . \quad (122)$$

The square magnitude of the $\Delta = XW$ is

$$\|\Delta\|^2 = \langle \Delta, \Delta \rangle = \|W\|_F^2 . \quad (123)$$

Geodesics

On a Riemannian manifold a geodesic is the curve of shortest length between two points. From [1, page 309] “A straightforward exercise from the calculus of variations reveals that for the case of manifolds embedded in a Euclidean space the acceleration vector at each point along a geodesic is normal to the submanifold so long as the curve is traced with uniform speed (i.e., tangential acceleration is zero). This condition is necessary and sufficient. In the case of the sphere, acceleration for uniform motion on a great circle is directed radially and therefore normal to the surface; therefore, great circles are geodesics on the sphere.”

Thus for a curve $X(t)$ on $\mathcal{U}(N)$ to be a geodesic, the acceleration vector $\ddot{X}$ must be in the normal space $(T_{X(t)}\mathcal{U}(N))^\perp$ at every point along the curve $X(t)$. Since all elements of the normal space at $X(t)$ have the form $X(t)S$, S Hermitian, conclude that

$$\ddot{X} = XS . \quad (124)$$

An arbitrary curve $X(t)$ satisfies the manifold defining constraint (115); twice differentiating this equation yields

$$\ddot{X}^H X + 2\dot{X}^H \dot{X} + X^H \ddot{X} = 0 \quad (125)$$

If $X(t)$ is a geodesic then it must satisfy (124); substituting this into (125) results in

$$-SX^H X + 2\dot{X}^H \dot{X} - X^H XS = 0 \quad (126)$$

$$2S + 2\dot{X}^H \dot{X} = 0 \quad (127)$$

$$S = -\dot{X}^H \dot{X} \quad (128)$$

Back substituting this result in to (124) yields

$$\ddot{X} = -X(\dot{X}^H \dot{X}) \quad (129)$$

or equivalently

$$\ddot{X} + X(\dot{X}^H \dot{X}) = 0 . \quad (130)$$

This differential equation is referred the geodesic equation or the Euler-Lagrange equation of motion. Direct substitution verifies that this non-linear differential equation has the general solution

$$X(t) = C \exp(Wt) \quad (131)$$

for W anti-Hermitian and C a arbitrary scalar constant. From the basic properties of matrix exponentials [4] we have $\dot{X} = \exp(Wt)W$ and $\ddot{X} = \exp(Wt)W^2$. Substituting these into (130) yields

$$\exp(Wt)(WW + W^H W) = \exp(Wt)(WW - WW) = 0 \quad (132)$$

where we have used the anti-Hermitean condition $W^H = -W$.

The geodesic equation (130) and the arbitrary initial conditions, $X(0) = X_0$ and $\dot{X}(0) = X_0 W_0$, W_0 anti-Hermitian, define an initial value problem (IVP). This IVP has the solution

$$X(t) = X_0 \exp(W_0 t). \quad (133)$$

This curve in $\mathcal{U}(N)$ is said to be the geodesic emanating from X_0 in the direction of the tangent vector $\Delta = X_0 W_0$. The short hand notation for this geodesic mapping (133) is

$$X(t) = \exp_{X_0}(t\Delta) \equiv X_0 \expm(tW_0). \quad (134)$$

In this expression, Δ is a tangent vector at X_0 represented as a $N \times N$ complex matrix. The tangent at any point along the geodesic curve $X(t)$ is

$$\Delta(t) = \frac{dX(t)}{dt} = X_0 \expm(tW_0) W_0 = X(t) W_0, \quad (135)$$

and at $t = 0$ we have $\Delta(0) = X_0 W_0$.

Orthonormal Basis

An orthonormal basis for the tangent space $T_X \mathcal{U}$ is given by

$$\mathbf{e}_k = X W_k; \quad k = 1 \cdots N \quad (136)$$

where the anti-Hermitian matrices, W_k , are chosen to form an orthonormal set, that is

$$\langle \mathbf{e}_i, \mathbf{e}_j \rangle = \text{Re}(\text{tr}(W_i^H W_j)) = \delta_{ij}; \quad i, j = 1 \cdots N. \quad (137)$$

An arbitrary element of $\Delta \in T_X \mathcal{U}$ is then represented with w.r.t. this basis as $\Delta = \sum_{k=1}^N \mathbf{e}_k \alpha^k$ and $\alpha^k = \langle \mathbf{e}_k, \Delta \rangle$.

Example: For $N = 2$ we have $\dim(T_X \mathcal{U}) = N^2 = 4$. An orthonormal basis for $T_X \mathcal{U}$ is

$$\mathbf{e}_1 = \frac{1}{\sqrt{2}} \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}, \mathbf{e}_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} 0 & i \\ i & 0 \end{bmatrix}, \mathbf{e}_3 = \begin{bmatrix} i & 0 \\ 0 & 0 \end{bmatrix}, \mathbf{e}_4 = \begin{bmatrix} 0 & 0 \\ 0 & i \end{bmatrix}. \quad (138)$$

This choice of basis is not unique. Multiplying the above matrix representations of the basis vectors by an arbitrary 2×2 unitary matrix yields an alternate basis.

Example: The IVP defined by (130) and the initial conditions $X(0) = X_0$ and $\dot{X}(0) = \mathbf{e}_3$, defined above, has the solution

$$X(t) = X_0 \expm \left(t \begin{bmatrix} i & 0 \\ 0 & 0 \end{bmatrix} \right) = X_0 \begin{bmatrix} \exp(it) & 0 \\ 0 & 1 \end{bmatrix}. \quad (139)$$

Natural Geodesic Distance Between Points

The running length of the arbitrary smooth curve $X(t)$ emanating from the point X_0 is given by elementary calculus as

$$Length(X(t), X_0) = \int_{t_0}^{t_1} \sqrt{\langle \Delta(t), \Delta(t) \rangle} dt = \int_{t_0}^{t_1} \|\Delta(t)\| dt \quad (140)$$

where $\Delta(t) = \frac{dX(t)}{dt}$ is the tangent to the curve at $X(t)$.

The Riemannian or geodesic distance between two points is the length of the shortest curve connecting the two points. The distance between $X_1 \equiv X(t_1)$ and X_0 is then is defined as

$$dist(X_1, X_0) \equiv \min_{X(t)} (Length(X_1, X_0)) \quad (141)$$

minimized over all curves $X(t)$ between X_0 and X_1 . On the unit sphere this arc is a great circle and the arc length is the angle in radians separating two end points.

Let $X(t)$ be the geodesic curve emanating from the point X_0 in the direction $\Delta = X_0 W$ where W_0 is constant anti-Hermitian matrix. This curve is given by $X(t) = \exp_{X_0} t \Delta$ (133) and the tangent vector at any point along the this curve is

$$\Delta(t) \equiv \frac{dX(t)}{dt} = X(t) W_0. \quad (142)$$

Since $X^H(t) X(t) = I$, the magnitude square $\|\Delta(t)\|^2$ is constant at each point of the curve and is given by

$$\|\Delta(t)\|^2 = \text{Re}(\text{tr}(W_0^H X^H(t) X(t) W_0)) = \text{Re}(\text{tr}(W_0^H W_0)). \quad (143)$$

The right-hand side of this equation corresponds to $\|\Delta(0)\|^2$ so that magnitude to the running tangent vector is the constant $\|\Delta(t)\| = \|\Delta(0)\|$. Substituting this into (140) gives the running length of the curve as

$$Length(X(t), X_0) = \|\Delta(0)\| \int_{t_0}^t dt = \|\Delta(0)\|(t - t_0); \quad t \geq t_0 \quad (144)$$

or for $t_0 = 0$

$$dist(X(t), X_0) = \|\Delta(0)\|t. \quad (145)$$

If X_1 is an arbitrary point near X_0 , and $\Delta = exp_{X_0}^{-1}X_1$, then the geodesic distance between the points is

$$dist(X_1, X_0) = \|\Delta\|. \quad (146)$$

The equality of the distance between points and the magnitude of the corresponding tangent vector will be used later in analyzing the subspace estimation problem.

2.2.3 Signal Processing Manifold Representations

Having described the geometry of the Unitary group manifold we next consider the related spaces that naturally arise in signal processing applications. These spaces may be described in two different, but equivalent, ways.

In section 2.2.4 the spaces of interest are considered as quotient manifolds on $\mathcal{U}(N)$. This approach is useful for obtaining closed-form expressions for the geometrical objects of interest. The Stiefel manifold, denoted $St_{(N,p)}$, is represented as a quotient space within $\mathcal{U}(N)$ as

$$St_{(N,p)} = \mathcal{U}(N)/(I \times \mathcal{U}(N - p)). \quad (147)$$

Each point of $St_{(N,p)}$ represents a subset of $\mathcal{U}(N)$ defined by the equivalence relation

$$[X] \equiv \left\{ X \cdot \begin{bmatrix} I & 0 \\ 0 & Q \end{bmatrix} : Q \in \mathcal{U}(N - p) \right\}. \quad (148)$$

Thus two points in $\mathcal{U}(N)$ represented by matrices X_0, X_1 are in the same Stiefel equivalence class, $(X_1 \sim X_0)$, and therefore represent the same point on the Stiefel manifold

if

$$X_1 = X_0 \begin{bmatrix} I & 0 \\ 0 & Q \end{bmatrix} \quad (149)$$

for some unitary matrix $Q \in \mathcal{U}(N - p)$. This quotient manifold is sometimes written as $St_{(N,p)} = \mathcal{U}(N) / \sim$ to suggest that elements in the same equivalence class are “divided” or “quotiented” out, where $\sim$ is understood. Variations on this equivalence relation lead to the Basis and Grassmann manifolds discussed in the following sections.

In section 2.2.7 we consider more concrete manifold representations that are useful in numerical applications. For numerical signal processing applications, a single representative of an equivalence class must necessarily be chosen. In this alternate concrete representation, a point on the Stiefel manifold $St_{(N,p)}$ is represented by an $N \times p$ complex matrix and the manifold is defined as the collection

$$St_{(N,p)} \equiv \{Y \in C^{N \times p} : Y^H Y = I_p\} . \quad (150)$$

2.2.4 Quotient Space Representation of the Complex Stiefel Manifold

Representing $St_{(N,p)}$ as a quotient space of $\mathcal{U}(N)$ (147) allows the geometric properties of $St_{(N,p)}$ to be derived from the known geometry of $\mathcal{U}(N)$ (e.g, geodesics defined by the exponential map). The key idea here is the partitioning of unitary group tangent spaces into horizontal and vertical spaces, the metric, and their relationship to geodesics.

Referring to (147) the dimension of the $St_{(N,p)}$ is given by the difference

$$\dim(St_{(N,p)}) = \dim(\mathcal{U}(N)) - \dim(\mathcal{U}(N - p)) \quad (151)$$

$$\dim(St_{(N,p)}) = N^2 - (N - p)^2 = 2Np - p^2 \quad (152)$$

A point in the Stiefel manifold is a particular subset of the unitary matrices. In particular, the equivalence class $[X]$ (148) is the set of all $N \times N$ unitary matrices with the same first p columns as X . The set $[X]$ is a Riemannian submanifold of $\mathcal{U}(N)$, termed a fiber submanifold. To introduce the decomposition of the tangent space in the quotient space

setting, we write the $N \times N$ unitary matrix X_0 unitary matrix in partitioned form as

$$X_0 = \begin{bmatrix} Y_0 & Y_{0\perp} \end{bmatrix} \quad (153)$$

where Y_0 and $Y_{0\perp}$ correspond to the first p and the remaining $(N - p)$ columns of X_0 , respectively. Then, for any $X_1 \in \mathcal{U}(N)$ partitioned as $X_1 = \begin{bmatrix} Y_1 & Y_{1\perp} \end{bmatrix}$, $X_1 \sim X_0$ in the Stiefel sense if the first p columns of X_1 and X_0 are equal: that is, if $Y_1 = Y_0$. Although we introduce the partitioning notation in this section, we are working with equivalence classes and an $N \times N$ class representative rather than an $N \times p$ matrix representation. This later concrete method of representation, applicable to the Stiefel manifold, is consider in section 2.2.7.

Quotient Tangent Space

The general form of a tangent vector $\Delta \in T_{X_0}\mathcal{U}(N)$ is $\Delta = X_0W$ where W is $N \times N$ anti-Hermitian matrix (see (135)). Writing W in partitioned form as

$$W = \begin{bmatrix} A & -B^H \\ B & C \end{bmatrix} \quad (154)$$

where A is $p \times p$ anti-Hermitian, C is $(N - p) \times (N - p)$ anti-Hermitian, and B is $(N - p) \times p$ arbitrary complex, and using (153), the tangent vector $\Delta = X_0W$ takes the form

$$\Delta = \begin{bmatrix} Y_0 & Y_{0\perp} \end{bmatrix} \begin{bmatrix} A & -B^H \\ B & C \end{bmatrix}. \quad (155)$$

The tangent space $T_{X_0}\mathcal{U}(N)$ can be partitioned into two complementary subspaces based on this tangent vector form and the properties of the exponential map. These spaces, described below, are called the vertical and horizontal spaces and denoted respectively as $\mathcal{V}_{X_0}$ and $\mathcal{H}_{X_0}$. The tangent space $T_{X_0}\mathcal{U}(N)$ then may be expressed as

$$T_{X_0}\mathcal{U}(N) = \mathcal{V}_{X_0} \oplus \mathcal{H}_{X_0}. \quad (156)$$

Figure 5. The vertical fiber submanifold is the set $[X] = \exp_X \Phi$. All points on this fiber belong to the same equivalence class and therefore represent the same point of the quotient manifold. The metric and geodesics must all be restricted to the horizontal space.

Vertical Space $\mathcal{V}_{X_0}$

The vertical space, denoted $\mathcal{V}_{X_0}$, is a subspace of $T_{X_0}\mathcal{U}(N)$ consisting of tangent vectors that leave the first p -columns of X_0 unchanged under the action of the exponential map. Using the notation (153), vectors in this space have the form

$$\Phi = \begin{bmatrix} Y_0 & Y_{0\perp} \end{bmatrix} \begin{bmatrix} 0 & 0 \\ 0 & C \end{bmatrix} = Y_{0\perp} C . \quad (157)$$

Movement along a geodesic in the direction of any vector in this space stays within the equivalence class (148). To verify this, compute the exponential map $\exp_{X_0}\Phi$

$$\begin{bmatrix} Y_1 & Y_{1\perp} \end{bmatrix} = \begin{bmatrix} Y_0 & Y_{0\perp} \end{bmatrix} \expm \left(\begin{bmatrix} 0 & 0 \\ 0 & C \end{bmatrix} \right) , \quad (158)$$

$$\begin{bmatrix} Y_1 & Y_{1\perp} \end{bmatrix} = \begin{bmatrix} Y_0 & Y_{0\perp} \end{bmatrix} \begin{bmatrix} I_p & 0 \\ 0 & \expm(C) \end{bmatrix} , \quad (159)$$

so that

$$Y_1 = Y_0 , \quad (160)$$

and

$$Y_{1\perp} = Y_{0\perp} Q_C \quad (161)$$

where $Q_C \equiv \expm(C)$ is unitary since C is anti-Hermitian. Observe that the first p columns of $X_1 \equiv \begin{bmatrix} Y_1 & Y_{1\perp} \end{bmatrix}$ are left unchanged, while the remaining columns undergo a rotation. The vertical space and the resulting fiber manifold (the set $[X] = \exp_X \Phi$) are illustrated in figure 5. The dimension of $\mathcal{V}_{X_0}$ is equal to degrees of freedom of $C \in U(N-p)$ appearing in (157), that is

$$\dim(\mathcal{V}_{X_0}) = \dim(\mathcal{U}(N-p)) = (N-p)^2 . \quad (162)$$

Horizontal Space

The orthogonal complement of $\mathcal{V}_{X_0}$ in $T_{X_0}\mathcal{U}(N)$ is defined as the horizontal space, denoted $\mathcal{H}_{X_0}$. The significance of the horizontal space is that it provides a representation

of tangents to the quotient manifold space, which in the current discussion is $St_{(N,p)}$, that is

$$T_{X_0}St_{(N,p)} \equiv \mathcal{H}_{X_0} . \quad (163)$$

Movement along a geodesic in the $\mathcal{H}_{X_0}$ direction changes the first p -columns resulting in a different point in the quotient space. As discussed previously, movements in the vertical direction make no change in the quotient space. Therefore, the metric and geodesics must all be restricted to the horizontal space.

Referring to (157) and (155), tangent vectors in $\mathcal{H}_{X_0}$ have the general partitioned matrix form as

$$\Delta = \begin{bmatrix} Y_0 & Y_{0\perp} \end{bmatrix} \begin{bmatrix} A & -B^H \\ B & 0 \end{bmatrix} \quad (164)$$

or equivalently

$$\Delta = \begin{bmatrix} Y_0 A + Y_{0\perp} B, & -Y_0 B^H \end{bmatrix} . \quad (165)$$

The unitary group geodesic corresponding to an arbitrary tangent vector in $\mathcal{H}_{X_0}$ is evidently

$$X(t) = \exp_{X_0}(\Delta t) \equiv X_0 \exp \left(t \begin{bmatrix} A & -B^H \\ B & 0 \end{bmatrix} \right) \quad (166)$$

or in partitioned form

$$\begin{bmatrix} Y(t) & Y_{\perp}(t) \end{bmatrix} = \begin{bmatrix} Y_0 & Y_{0\perp} \end{bmatrix} \exp \left(t \begin{bmatrix} A & -B^H \\ B & 0 \end{bmatrix} \right) . \quad (167)$$

The tangent vector $\Delta(t) \in \mathcal{H}_{X(t)}$ at every point along this curve is

$$\Delta(t) = \frac{dX(t)}{dt} = X(t) \begin{bmatrix} A & -B^H \\ B & 0 \end{bmatrix} . \quad (168)$$

The curve defined by (167), a geodesic in $\mathcal{U}(N)$, is a geodesic in the quotient space $St_{(N,p)}$ as well. We will return to this formula when deriving an expression for geodesics with respect to the concrete representation of the Stiefel manifold in section 2.2.7.

The dimension of $\mathcal{H}_{X_0}$ is equal to degrees of freedom of $A \in U(p)$ plus degrees of freedom of $B \in C^{(N-p) \times p}$ or

$$\dim(\mathcal{H}_{X_0}) = p^2 + 2p(N-p) = 2Np - p^2 . \quad (169)$$

Note that

$$\dim(\mathcal{U}_{X_0}) = \dim(\mathcal{V}_{X_0}) + \dim(\mathcal{H}_{X_0}) = (N - p)^2 + 2Np - p^2 = N^2 . \quad (170)$$

The partitioning of $T_{X_0}\mathcal{U}(N)$ into vertical and horizontal subspaces (156) is dependent on the choice of equivalence relation. Different equivalence relations result in different partitions. This will be evident when we consider the related Basis and Grassmann manifold formulations (see Table 2).

Partitioning the Horizontal space

In later signal processing applications we wish to distinguish between pure rotations and motions which change the subspace. It is therefore useful to partition the horizontal space $\mathcal{H}_{X_0}$ into a rotation space and its complement. Referring to (154), an arbitrary element of $\mathcal{H}_{X_0}$ can be expressed as the sum

$$\Delta = X_0 \begin{bmatrix} A & -B^H \\ B & 0 \end{bmatrix} = X_0 \begin{bmatrix} A & 0 \\ 0 & 0 \end{bmatrix} + X_0 \begin{bmatrix} 0 & -B^H \\ B & 0 \end{bmatrix} . \quad (171)$$

The space A_{X_0} is defined as the subspace of $\mathcal{H}_{X_0}$

$$A_{X_0} = \{ \Delta \in \mathcal{H}_{X_0} \text{ such that } \Delta = X_0 \begin{bmatrix} A & 0 \\ 0 & 0 \end{bmatrix}, A \text{ anti-Hermitian} \} . \quad (172)$$

The dimension of this space, denoted N_a , is equal to the degrees of freedom in a $p \times p$ skew-Hermitian matrix so that

$$N_a = \dim(A_{X_0}) = p^2 . \quad (173)$$

The complement of A_{X_0} in $\mathcal{H}_{X_0}$, denoted B_{X_0} , is given by

$$B_{X_0} = \{ \Delta \in \mathcal{H}_{X_0} \text{ such that } \Delta = X_0 \begin{bmatrix} 0 & -B^H \\ B & 0 \end{bmatrix}, B \text{ arbitrary} \} . \quad (174)$$

Since B is an arbitrary $(N - p) \times p$ complex matrix, the real dimension of the vector space B_{X_0} , denoted N_b , is

$$N_b = \dim(B_{X_0}) = 2p(N - p) . \quad (175)$$

By construction

$$T_{X_0}St_{(N,p)} \equiv \mathcal{H}_{X_0} = A_{X_0} \oplus B_{X_0} . \quad (176)$$

Referring to (169), (173) and (175), we note that the spaces are dimensionally correct since

$$\dim(\mathcal{H}_{X_0}) = p^2 + 2p(N - p) = 2Np - p^2 . \quad (177)$$

The exponential map corresponding to tangent vectors in A_{X_0} (172) is

$$\begin{bmatrix} Y & Y_{\perp} \end{bmatrix} = \begin{bmatrix} Y_0 & Y_{0\perp} \end{bmatrix} \expm \left(\begin{bmatrix} A & 0 \\ 0 & 0 \end{bmatrix} \right) , \quad (178)$$

and in partitioned form the first p -columns are

$$Y = Y_0 \expm(A) = Y_0 Q . \quad (179)$$

The last $N - p$ columns are unchanged. Motions along A_{X_0} directions (178) represent pure rotations of the p columns of Y_0 ; there is no change in the subspace.

The significance of the B_{X_0} space is that it corresponds to motions that alter the subspace spanned by first p -columns of X_0 . The exponential map corresponding to tangent vectors in B_{X_0} (174) is given in partitioned form as

$$\begin{bmatrix} Y & Y_{\perp} \end{bmatrix} = \begin{bmatrix} Y_0 & Y_{0\perp} \end{bmatrix} \exp \left(\begin{bmatrix} 0 & -B^H \\ B & 0 \end{bmatrix} \right) . \quad (180)$$

The first-order approximation of this map is

$$Y \approx Y_0 + Y_{0\perp} B \quad (181)$$

$$Y_{\perp} \approx Y_{0\perp} - Y_0 B^H . \quad (182)$$

Basis for A_{X_0} and B_{X_0}

An orthonormal basis for A_{X_0} has the form

$$\mathbf{a}_k = X_0 \begin{bmatrix} A_k & 0 \\ 0 & 0 \end{bmatrix} ; \quad k = 1 \cdots N_a \quad (183)$$

where the indexed set of $p \times p$ matrices A_k , are anti-Hermitian and furthermore satisfy the orthonormality condition

$$\text{Re}(\text{tr}(A_i^H A_j)) = \delta_{ij}; \quad i, j = 1 \cdots N_a \quad (184)$$

Similarly, an orthonormal basis for B_{X_0} is

$$\mathbf{b}_k = X_0 \begin{bmatrix} 0 & -B_k^H \\ B_k & 0 \end{bmatrix}; \quad k = 1 \cdots N_b \quad (185)$$

where the indexed set of $(N - p) \times p$ matrices B_k are arbitrary complex and satisfy the orthonormality condition

$$\text{Re}(\text{tr}(B_i^H B_j)) = \delta_{ij}; \quad i, j = 1 \cdots N_b. \quad (186)$$

Any tangent vector $\Delta \in \mathcal{H}_{X_0}$ can be represented with respect to the above basis vectors as

$$\Delta = \sum_1^{N_a} \mathbf{a}_k \alpha^k + \sum_1^{N_b} \mathbf{b}_k \beta^k \quad (187)$$

for some real N_a -tuple α and some real N_b -tuple β . For a given Δ the coefficients α and β are computed in the standard way

$$\alpha^k = \langle \mathbf{a}_k, \Delta \rangle, \quad k = 1 \cdots N_a, \quad (188)$$

$$\beta^k = \langle \mathbf{b}_k, \Delta \rangle, \quad k = 1 \cdots N_b. \quad (189)$$

If we define

$$A = \sum_{k=1}^{N_a} A_k \alpha^k, \quad B = \sum_{k=1}^{N_b} B_k \beta^k \quad (190)$$

then we recover the A and B matrices appearing in tangent vector form (165).

Example: Basis Vectors for Stiefel Manifold Tangent Space

For $p = 2$, N arbitrary, we have $\dim(A_{X_0}) = N_a = p^2 = 4$ and $\dim(B_{X_0}) = N_b = 4(N - 2)$. For this case a set of basis matrices A_k appearing in (183) are

$$A_1 = \frac{1}{\sqrt{2}} \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}, A_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} 0 & i \\ i & 0 \end{bmatrix}, A_3 = \begin{bmatrix} i & 0 \\ 0 & 0 \end{bmatrix}, A_4 = \begin{bmatrix} 0 & 0 \\ 0 & i \end{bmatrix}. \quad (191)$$

The standard set of B_k matrices appearing in (185) are

$$B_k = \frac{1}{\sqrt{2}} \text{reshape}(i_k, N - p, p); \quad k = 1, \dots, \frac{1}{2}N_b \quad (192)$$

$$B_{k+M} = j \cdot B_k; \quad M = \frac{1}{2}N_b \quad (193)$$

where i_k is a real $p(N - p) \times 1$ unit vector of all zeros and 1 in the k th location. For example, if $N = 5, p = 2$, then $\dim(B_{X_0}) = N_b = 2(N - p)p = 12$. In this case the $k = 2$ basis matrix is

$$B_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 0 \end{bmatrix} \quad (194)$$

and for $k = 2 + \frac{N_b}{2} = 8$

$$B_8 = \frac{1}{\sqrt{2}} \begin{bmatrix} 0 & 0 \\ i & 0 \\ 0 & 0 \end{bmatrix}. \quad (195)$$

Relation between inner product and norm and metric The canonical metric on the Stiefel manifold is the restriction of the unitary group metric (110) to the horizontal space, sH_{X_0} . Any element $\Delta \in \mathcal{H}_{X_0}$ can be represented as (187) or equivalently (165). The squared norms corresponding to these two representations are

$$\langle \Delta, \Delta \rangle = \sum_{k=1}^{N_a} (\alpha^k)^2 + \sum_{k=1}^{N_b} (\beta^k)^2 = \|\alpha\|^2 + \|\beta\|^2 \quad (196)$$

and

$$\langle \Delta, \Delta \rangle = \text{Re} \left(\text{tr} \left(\begin{bmatrix} A & -B^H \\ B & 0 \end{bmatrix}^H \begin{bmatrix} A & -B^H \\ B & 0 \end{bmatrix} \right) \right) \quad (197)$$

$$= \text{Re} \left(\text{tr} \begin{bmatrix} A^H A + B^H B & -A^H B \\ -BA & BB^H \end{bmatrix} \right) \quad (198)$$

$$= \text{Re} \left(\text{tr}(A^H A) + 2 \cdot \text{tr}(B^H B) \right) \quad (199)$$

The components α^k and β^k are related to the A and B matrices appearing in the (187) representation by

$$\alpha^k = \langle \mathbf{a}_k, \Delta \rangle = \text{Re} \left(\text{tr} \left(\begin{bmatrix} A_k & 0 \\ 0 & 0 \end{bmatrix}^H \begin{bmatrix} A & -B^H \\ B & 0 \end{bmatrix} \right) \right) \quad (200)$$

$$= \text{Re} \left(\text{tr} \left(\begin{bmatrix} A_k^H A & A_k^H B^H \\ 0 & 0 \end{bmatrix} \right) \right) = \text{Re} \left(\text{tr}(A_k^H A) \right) , \quad (201)$$

and

$$\beta^k = \langle \mathbf{b}_k, \Delta \rangle = \text{Re} \left(\text{tr} \left(\begin{bmatrix} 0 & -B_k^H \\ B_k & 0 \end{bmatrix}^H \begin{bmatrix} A & -B^H \\ B & 0 \end{bmatrix} \right) \right) \quad (202)$$

$$= \text{Re} \left(\text{tr} \left(\begin{bmatrix} B_k^H B & 0 \\ B_k A & B_k^H B \end{bmatrix} \right) \right) = 2 \cdot \text{Re} \left(\text{tr}(B_k^H B) \right) . \quad (203)$$

Note that

$$\|\beta\|^2 = \sum_{k=1}^{N_b} (\beta^k)^2 = \sum_{k=1}^{N_b} (\beta^k 2 \text{Re}(\text{tr}(B_k^H B))) \quad (204)$$

$$= \sum_{k=1}^{N_b} (2 \text{Re}(\text{tr}(\beta^k B_k^H B))) = 2 \text{Re}(\text{tr}(B^H B)) \quad (205)$$

so that $\|\beta\|^2 = 2 \text{Re}(\text{tr}(B^H B))$ and the expressions for the squared norm given by (199) and (196) are equivalent. The canonical metric on the Stiefel manifold is then simply the restriction of the unitary group metric to the horizontal space, that is

$$\langle \Delta_1, \Delta_2 \rangle|_{St(N,p)} = \langle P_{\mathcal{H}} \Delta_1, P_{\mathcal{H}} \Delta_2 \rangle \quad (206)$$

where $P_{\mathcal{H}}$ is the projector onto $\mathcal{H}$.

2.2.5 Basis Manifold

The Basis manifold is defined as the quotient $\mathcal{U}(N)/(D_p \times U(N-p))$ with each point being the equivalence class given by

$$[X] = \left\{ X \begin{bmatrix} M_p & 0 \\ 0 & Q \end{bmatrix} : M_p \in D_p, Q \in \mathcal{U}(N-p) \right\} \quad (207)$$

where D_p is the space of $p \times p$ unitary diagonal matrices. Elements of D_p have the form $\text{diag}(\exp(i\theta_1), \dots, \exp(i\theta_p))$ for θ_k real so that $\dim(D_p) = p$. A point in the Basis manifold is a particular subset of the unitary matrices, and the Basis manifold itself is the collection of all these subsets. The dimension of $\mathcal{B}_{(N,p)}$ is given by the difference

$$\dim(\mathcal{B}_{(N,p)}) = \dim(\mathcal{U}(N)) - \dim(\mathcal{U}(N-p)) - \dim(D_p) = 2Np - p^2 - p \quad (208)$$

or equivalently $\dim(\mathcal{B}_{(N,p)}) = 2p(N-p) + p^2 - p$.

The vertical space, $\mathcal{V}_{X_0}$, consist of vectors of the form

$$\Phi = \begin{bmatrix} Y_0 & Y_{0\perp} \end{bmatrix} \begin{bmatrix} M & 0 \\ 0 & C \end{bmatrix} \quad (209)$$

where M is $p \times p$ unitary diagonal, and C is $(N-p) \times (N-p)$ anti-Hermitian. The corresponding exponential map is

$$\begin{bmatrix} Y & Y_\perp \end{bmatrix} = \begin{bmatrix} Y_0 & Y_{0\perp} \end{bmatrix} \expm \begin{bmatrix} M & 0 \\ 0 & C \end{bmatrix} \quad (210)$$

so that

$$Y = Y_0 \cdot \expm(M) = Y_0 \cdot \begin{bmatrix} \exp(M_{11}) & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \exp(M_{pp}) \end{bmatrix} \quad (211)$$

where the elements M_{kk} are pure imaginary. The dimension of $\mathcal{V}_{X_0}$ is equal to the degrees of freedom of $C \in U(N-p)$ plus the degrees of freedom of $M \in D_p$, that is

$$\dim(\mathcal{V}_{X_0}) = \dim(\mathcal{U}(N-p)) + \dim(D_p) = (N-p)^2 + p \quad (212)$$

The dimension of the vertical space for the basis manifold formulation is *increased* by p relative to the Stiefel manifold case (compare eqns (209) and (157)). Accordingly, the dimension of the horizontal space $\mathcal{H}_{X_0}$ is *decreased* by p relative to the Stiefel manifold case.

As in the Steifel manifold case, it is useful to partition $T_{X_0}\mathcal{B}_{(N,p)} \equiv \mathcal{H}_{X_0}$, into two component spaces as in (176). The space B_{X_0} is defined in exactly the same way as Stiefel manifold case (174). The space A_{X_0} is defined in a similar way as Stiefel manifold case, with the exception that unitary diagonal matrices have been removed (these are exactly those that appear in vertical space $\mathcal{V}_{X_0}$ for Basis manifold case (209)).

Example: Basis Vectors for Basis Manifold Tangent Space

Referring to example 2.2.4, the basis elements for the A_{X_0} space are

$$A_1 = \frac{1}{2} \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}, A_2 = \frac{1}{2} \begin{bmatrix} 0 & i \\ i & 0 \end{bmatrix}. \quad (213)$$

The basis elements A_3 and A_4 , appearing in the complex Stiefel manifold example, do not appear in the Basis manifold since these are unitary diagonal matrices and, as such, belong to the Basis manifold vertical space $\mathcal{V}_{X_0}$.

In general, the standard basis for the A_{X_0} basis manifold space is found by removing the p diagonal unitary basis elements that appear in the Stiefel manifold A_{X_0} space (note that we are using the same symbol A_{X_0} to denote two different spaces as they appear in the Stiefel and Basis manifold formulations). The dimension of Basis manifold A_{X_0} space is therefore reduced by p relative to the Stiefel formulation (173), so that $\dim(A_{X_0}) = p^2 - p = p(p - 1)$. The basis elements for B_{X_0} are identical to those in the Stiefel case (192) as this space is unchanged.

The canonical metric on the Basis manifold is the restriction of the unitary group metric to the horizontal space $\mathcal{H}_{X_0} = A_{X_0} \oplus B_{X_0}$.

2.2.6 Grassmann Manifold

The Grassmann manifold is defined as the quotient $\mathcal{U}(N)/(\mathcal{U}(p) \times \mathcal{U}(N-p))$ with each point an equivalence class defined by

$$[X] = \left\{ X \begin{bmatrix} Q_p & 0 \\ 0 & Q_{N-p} \end{bmatrix} : Q_p \in \mathcal{U}(p), \quad Q_{N-p} \in \mathcal{U}(N-p) \right\}. \quad (214)$$

A point in the Grassmann manifold is a particular subset of the unitary matrices, and the Grassmann manifold itself is the collection of all these subsets. The dimension of the $G_{(N,p)}$ is given by the difference

$$\dim(G_{(N,p)}) = \dim(\mathcal{U}(N)) - \dim(\mathcal{U}(N-p)) - \dim(\mathcal{U}(p)) = N^2 - (N-p)^2 - p^2 \quad (215)$$

$$\dim(G_{(N,p)}) = 2Np - p^2 - p^2 = 2p(N-p). \quad (216)$$

The vertical space at a point X_0 consists of vectors of the form

$$\Phi = \begin{bmatrix} Y_0 & Y_{0\perp} \end{bmatrix} \begin{bmatrix} A & 0 \\ 0 & C \end{bmatrix} \quad (217)$$

where A is $p \times p$ anti-Hermitian and C is $(N - p) \times (N - p)$ anti-Hermitian. The corresponding exponential map $\Phi \in \mathcal{V}_{X_0}$ is

$$\begin{bmatrix} Y & Y_\perp \end{bmatrix} = \begin{bmatrix} Y_0 & Y_{0\perp} \end{bmatrix} \expm \left(\begin{bmatrix} A & 0 \\ 0 & C \end{bmatrix} \right) \quad (218)$$

or equivalently

$$Y = Y_0 \cdot \expm(A) = Y_0 Q_A, \quad (219)$$

$$Y_\perp = Y_{0\perp} \cdot \expm(C) = Y_{0\perp} Q_C. \quad (220)$$

We see that the first p columns and the last $(N - p)$ columns undergo independent rotations. The $N \times p$ matrices Y_0 and Y are in the same Grassmann equivalence class so that $Y \sim Y_0$. Note the dimension of the vertical space for the basis manifold formulation is increased by p^2 relative to the Stiefel manifold case (compare eqns (209) and (157)).

The horizontal space $\mathcal{H}_{X_0}$ at X_0 consists of the tangent vectors of the form

$$\Delta = \begin{bmatrix} Y_0 & Y_{0\perp} \end{bmatrix} \begin{bmatrix} 0 & -B^H \\ B & 0 \end{bmatrix}. \quad (221)$$

The A_{X_0} subspace of $\mathcal{H}_{X_0}$ appearing in different forms in the Stiefel and Basis manifold cases is empty here since these rotations, (219), now belong to the vertical space $\mathcal{V}_{X_0}$.

The canonical metric on the Grassmann manifold is the restriction of the unitary group metric (121) to the horizontal space $\mathcal{H}_{X_0} = T_{X_0}\mathcal{G}$, that is

$$\langle \Delta_1, \Delta_2 \rangle|_{\mathcal{G}} \equiv \langle P_{\mathcal{H}_{X_0}} \Delta_1, P_{\mathcal{H}_{X_0}} \Delta_2 \rangle \quad (222)$$

where $P_{\mathcal{H}_{X_0}}$ is the projection onto $\mathcal{H}_{X_0}$; for the Grassmann manifold $\mathcal{H}_{X_0} \equiv B_{X_0}$. The table below summarizes the dimensions of the various spaces encountered in the Quotient manifold discussion. We note that $\dim(\mathcal{H}_Y) + \dim(\mathcal{V}_Y) = N^2 = \dim(U(N))$ for each of the manifolds as expected.

2.2.7 Concrete Representation of the Stiefel Manifold

In this section the Stiefel manifold is represented as the collection of all $N \times p$ semi-unitary matrices, viewed as a subset of $C^{N \times p}$. This is in contrast to the quotient space

Table 1. Horizontal $\mathcal{H}_{X_0}$ and Vertical $\mathcal{V}_{X_0}$ Space Dimension of Quotient Manifolds

Manifold	B_Y	A_Y	$\mathcal{H}_Y$	$\mathcal{V}_Y$
$St_{(N,p)}$	$2p(N-p)$	p^2	$2pN - p^2$	$(N-p)^2$
$\mathcal{B}_{(N,p)}$	$2p(N-p)$	$p^2 - p$	$2pN - p^2 - p$	$(N-p)^2 + p$
$G_{(N,p)}$	$2p(N-p)$	0	$2p(N-p)$	$(N-p)^2 + p^2$

representation where points on the manifold were identified with subsets of $\mathcal{U}(N) \subset C^{N \times N}$. For this complex case the Stiefel manifold is defined in concrete form as

$$St_{(N,p)} \equiv \{Y \in C^{N \times p} : Y^H Y = I\} . \quad (223)$$

When $p = N$, $St_{(N,p)}$ is the unitary group $\mathcal{U}(N)$.

Tangent Space

In the quotient manifold approach the ambient space was N^2 -dimensional and tangent vectors were represented as $N \times N$ matrices of the form (171) or (165). Since, in this section, we are working in an Np -dimensional ambient space, here tangent vectors are represented as $N \times p$ matrices. For the arbitrary curve $Y(t)$ on $St_{(N,p)}$ differentiating the defining equation (223) yields

$$\frac{d}{dt}(Y^H(t)Y(t)) = \dot{Y}^H Y + Y^H \dot{Y} = 0 . \quad (224)$$

Introducing the notation $\Delta \equiv \dot{Y}$ this becomes

$$\Delta^H Y + Y^H \Delta = 0 . \quad (225)$$

An arbitrary $N \times p$ matrix $\Delta \in T_Y C^{N \times p}$ may be decomposed as

$$\Delta = Y(A + S) + Y_\perp B \quad (226)$$

for A , $p \times p$ anti-Hermitian, S , $p \times p$ Hermitian, B , arbitrary $(N-p) \times p$ complex matrix and $Y^H Y_\perp = 0$. The tangent space $T_Y St_{(N,p)} \subset T_Y C^{N \times p}$ is constrained by tangency

condition (225). Substituting (226) into (225) shows that the symmetric part of (226) must vanish, that is $S \equiv 0$. Elements of $T_Y St_{(N,p)}$ are therefore

$$\Delta = Y A + Y_{\perp} B . \quad (227)$$

This form is the same as found by selecting the first p -columns of the quotient space tangent vector representation (165). The space $T_Y St_{(N,p)}$ partitions as before (176), that is

$$T_Y St_{(N,p)} = A_Y \oplus B_Y \quad (228)$$

where A_Y and B_Y are defined by selecting the first p -columns in the definitions (172) and (174). respectively. The symmetric component of (226), $Y S$, is normal to the embedded manifold in the ambient space. The collection of all such vectors is the normal space $(T_Y St_{(N,p)})^{\perp}$. The total space decomposes as

$$T_Y C^{N \times p} = T_Y St_{(N,p)} \oplus (T_Y St_{(N,p)})^{\perp} \quad (229)$$

Note that for $p = 1$, Y is $N \times 1$ unit vector and S is a complex scalar.

Canonical Metric

A metric on manifold is a smoothly varying mapping

$$g : T_Y \mathcal{M} \times T_Y \mathcal{M} \mapsto R \quad (230)$$

that at each point satisfies the standard inner product conditions (see section A.3.2). The standard Euclidean metric on $\mathcal{M} = C^{N \times p}$, viewed as a vector space over the reals, is

$$g_e(\Delta, \Delta) = 2 \cdot \text{Re}(\text{tr}(\Delta^H \Delta)) \quad (231)$$

The scale factor 2 has been included for reasons of simplicity that will become evident in the following. The subscript “ e ” stands for Euclidean. Substituting the tangent vector $\Delta \in T_Y St_{(N,p)}$ form given by (227) into (231) yields

$$g_e(\Delta, \Delta) = 2 \cdot \text{Re}(\text{tr}(B^H B) + \text{tr}(A^H A)) \quad (232)$$

This is *not* the metric selected for $St_{(N,p)}$ since it counts independent elements of A of two times. To see this consider the anti-Hermitian matrix

$$A = \begin{bmatrix} 0 & -a^* \\ a & 0 \end{bmatrix} . \quad (233)$$

Then

$$\text{tr}(A^H A) = \text{tr}\left(\begin{bmatrix} |a|^2 & 0 \\ 0 & |a|^2 \end{bmatrix}\right) = 2|a|^2 \quad (234)$$

We see that the independent element a is counted twice. To correct this, the canonical metric on $St_{(N,p)}$, denoted as g_c , is instead selected as

$$g_c(\Delta, \Delta) = 2 \cdot \text{Re}\left(\text{tr}(\Delta^H (I - \frac{1}{2} Y Y^H) \Delta)\right) . \quad (235)$$

or equivalently as

$$g_c(\Delta, \Delta) = 2 \cdot \text{Re}\left(\text{tr}(\Delta^H (Y_\perp Y_\perp^H + \frac{1}{2} Y Y^H) \Delta)\right) . \quad (236)$$

For an arbitrary $\Delta \in T_Y St_{(N,p)}$ given by (227) this yields

$$g_c(\Delta, \Delta) = 2 \cdot \text{Re}\left(\text{tr}(B^H B) + \text{tr}(A^H A)\right) . \quad (237)$$

Comparing the above result (236) to (199), the metric for the quotient manifold formulation, we see that the two forms they are identical. Therefore, the formulas for geodesics for the Stiefel manifold given in section 2.2.4 are correct if we view the Stiefel manifold as the set of semi-unitary $N \times p$ matrices with the metric g_c (235). The canonical metric g_c is thus not simply the restriction of the g_e to the submanifold.

For simplicity we sometimes will use the alternate angle bracket notation for the metric

$$\langle \Delta, \Delta \rangle \equiv g_c(\Delta, \Delta) \quad (238)$$

where any dependence on the base point Y has been suppressed.

Geodesics (Stiefel)

Let $Y(t)$ denote an arbitrary smooth curve in $St_{(N,p)}$ between the points Y_0 and Y_1 , such $Y_0 = Y(0)$ and $Y_1 = Y(t_1)$. The length of this curve is given by the integral

$$L = \int_0^{t_1} \sqrt{g_c(\dot{Y}, \dot{Y})} dt \quad (239)$$

where $\dot{Y} = \frac{dY}{dt}$ is a tangent to the curve at $Y(t)$.

The Euler-Lagrange equations of motion (geodesic equation) are derived from this path length integral using the Calculus of Variations. The path length defined above is minimized over the set of all smooth curves between Y_0 and Y_1 . This minimization leads a 2nd order nonlinear differential equation called the geodesic equation. For the choice of the canonical metric (236) this equation is derived in [?] as

$$\ddot{Y} + \dot{Y}(\dot{Y}^H Y) + Y(Y^H \dot{Y})^2 + \dot{Y}^H \dot{Y} = 0 . \quad (240)$$

Direct substitution verifies that general solution is given by

$$Y(t) = \begin{bmatrix} Y_0 & Y_{0\perp} \end{bmatrix} \exp \left(t \begin{bmatrix} A & -B^H \\ B & 0 \end{bmatrix} \right) I_{N,p} . \quad (241)$$

As anticipated, this is identical to the geodesic derived from the quotient manifold formula and given by (167).

Basis Manifold as a Quotient manifold of the Stiefel Manifold

The Basis manifold can be naturally defined as a quotient of the above concrete Stiefel manifold formulation. Recalling (207), the Basis manifold is represented as the quotient $St_{(N,p)}/D_p$ with each point being an equivalence class given by

$$[Y] = \{YM : M \in D_p\} . \quad (242)$$

A point in the Basis manifold is thus a particular subset of the $N \times p$ semi-unitary matrices, and the Basis manifold itself is the collection of all these subsets. When

performing computations on the Basis manifold we will use an $N \times p$ matrix Y to represent an entire equivalence class; any representative of the class will do.

The vertical space, a subspace of $T_{Y_0}St_{(N,p)}$, corresponds to motions that remain in the set (242) and consists of tangent vectors of the form

$$\Phi = Y_0 M; \quad M \in D_p \quad (243)$$

where D is unitary-diagonal. This collection defines the vertical space $\mathcal{V}_0$ at Y_0 . The orthogonal complement is the horizontal space, $\mathcal{H}_{Y_0}$, which here consists of vectors of the form

$$\Delta = Y_0 A + Y_{0\perp} B. \quad (244)$$

The significance of the horizontal space is that it provides a representation of tangents to the quotient manifold space which in the current discussion is $\mathcal{B}_{N,p}$, that is

$$T_{Y_0}\mathcal{B}_{N,p} \equiv \mathcal{H}_{Y_0} \quad (245)$$

Movements in the vertical direction make no change in the quotient space. Therefore, the metric and geodesics must all be restricted to the horizontal space. Viewed as a quotient manifold of $St_{(N,p)}$ we have

$$\dim(\mathcal{B}_{(N,p)}) = \dim(St_{(N,p)}) - \dim(D_p) = 2Np - p^2 - p = 2p(N - p) + p^2 - p \quad (246)$$

The Grassmann manifold can be defined as a quotient manifold of $St_{(N,p)}$ in a similar way.

2.3 Subspace Angles, Subspace Distance and Geodesic Distance

The principal angles between subspaces (PABS) (also called canonical angles) provide a measure of the difference between two subspaces. We begin by stating two alternate and equivalent definitions of PABS. Next we relate the root sum squared (RSS) PABS, defined as the subspace distance, to the natural (geodesic) distance between points on the signal processing manifolds of interest (i.e., Grassmann, Basis and Stiefel).

The acute angle γ between two real unit vectors x and y is defined implicitly as $\cos \gamma = \langle x, y \rangle$. This definition can be recursively extended to PABS; see, e.g., [2, 10, 11].

Definition 1: Recursive PABS. Let $X \in C^{N \times p}$ and $Y \in C^{N \times p}$. The principal angles, γ^k , between X and Y are recursively defined by

$$\cos(\gamma^k) = \max_{x \in \text{col}(X)} \max_{y \in \text{col}(Y)} \langle x, y \rangle \quad (247)$$

subject to $|x| = 1, |y| = 1$ and $\langle x_i, x \rangle = 0$ and $\langle y_i, y \rangle = 0$ for all $i < k$. The notation $x \in \text{col}(X)$ indicates x is an $N \times 1$ vector in the column space of X , $\text{col}(X)$. The inner product is $\langle x, y \rangle = \text{Re}(x^H y)$.

Remark: Denote the pair that maximizes the first iteration as (x_1, y_1) . The second iteration, maximizes over the set of all $x \in X, y \in Y$ with the constraint that $\langle x_1, x \rangle = 0$ and $\langle y_1, y \rangle = 0$.

An alternate definition of PABS, from [10, 11], that is useful for computations is based on the SVD of the product $X^H Y$. It is stated below in the form of a theorem.

Theorem 2 ((PABS from SVD of $X^H Y$)). *Let the SVD of $X^H Y$ be*

$$UCV^H := X^H Y \quad (248)$$

where U and V are unitary matrices and C is a $p \times p$ diagonal matrix with the real diagonal elements C_{kk} in non increasing order. Then the subspace angles γ^k between $\text{col}(X)$ and $\text{col}(Y)$ from definition 1 are equal to

$$\gamma^k = \arccos(C_{kk}); k = 1 \cdots p, \quad 0 < \gamma^k < \pi \quad (249)$$

Note that for the special case of $p = 1$, and X and Y real $N \times 1$ unit vectors

$$\gamma = \arccos(X^t Y) \quad (250)$$

which agrees with the standard definition of the angle between two vectors.

The subspace distance between X and Y is defined in terms of this p -tuple of angles, γ , as

$$\text{dist}SS(Y_1, Y_0) \equiv \|\gamma\| . \quad (251)$$

This p -tuple may also be used to define several different “distances” between two subspaces [?, page 337].

For the Grassmann manifold formulation this distance is exactly equal to the geodesic distance. For the Stiefel and Basis manifold formulations the situation is more complicated. In the next section we show that for these formulations, $\text{dist}SS(Y_1, Y_0)$ corresponds to the geodesic distance between Y_0 and an appropriately defined fiber submanifold over Y_1 . This submanifold corresponds to the equivalence class containing Y_1 in the Grassmann formulation.

The rotational distance defined below for the Stiefel and Basis manifolds recognizes that on these manifolds X and XQ , where Q an arbitrary unitary matrix, are different points.

Definition: Steifel and Basis manifold Rotational Angles: Let the SVD of $X^H Y$ be given by (248) and let $Q = U^H V$. The rotational angles, denoted ν_k , between X and Y are

$$\exp(i\nu_k) = \rho_k; \quad k = 1 \cdots p \quad (252)$$

where ρ_k is the k -th singular value of Q for the Steifel manifold and k -th singular value of $(Q - \text{diag}(Q))$ for the Basis manifold. The rotational distance is defined as

$$\text{dist}R(X, Y) = \|\nu\| . \quad (253)$$

Note that in the Basis manifold formulation $Y \sim Y\Lambda_Q$ for Λ_Q unitary diagonal. The definition of rotational angles using $(Q - \text{diag}(Q))$ for the Basis manifold accounts for this equivalence.

2.3.1 Subspace Distance on Grassmann and Basis Manifolds

For the Grassmann manifold, the subspace distance (251) is identical to the geodesic distance between points [1]; that is, for any pair Y_1 and Y_0 , $dist_{SS}(Y_1, Y_0) = dist_G(Y_1, Y_0)$. This simple relation reflects the quotient structure of this manifold in which $N \times p$ complex matrices that have the same span are represented by the same point. For the Basis and Stiefel manifold cases this simple relationship does not hold since here matrices with different rotations are represented by different points (see section 2.3.2).

In all the manifold formulations the geodesic distance between two points, Y_1 and Y_0 , is given by the magnitude of tangent vector $\Delta = \exp_{Y_0}^{-1} Y_1$ (146). Note that the meaning of this inverse operation depends on the manifold formulation. Suppose $Y_1 = Y_0 Q$; in the Grassmann formulation $\exp_{Y_0}^{-1} Y_1 = 0$ (zero vector) while in the Stiefel manifold $\exp_{Y_0}^{-1} Y_1 = Y_0 \log m(Q)$. We first establish that the subspace distance corresponds to the geodesic distance in the Grassmann manifold formulation of the problem. This is done most easily using the CS decomposition form of the exponential map discussed below. This form was encountered earlier in the discussion of the sphere in section 2.1.

Theorem 3 (Equivalence of Forms CS decomposition). *The exponential (geodesic) map corresponding to the tangent vector $\Delta = Y_{0\perp} B$ at Y_0 is*

$$Y = \begin{bmatrix} Y_0 & Y_{0\perp} \end{bmatrix} \exp \left(\begin{bmatrix} 0 & -B^H \\ B & 0 \end{bmatrix} \right) I_{N,p} \quad (254)$$

This is equal to

$$Y = Y_0 V \cos(\Sigma) V^H + U \sin(\Sigma) V^H \quad (255)$$

or equivalently

$$Y = \begin{bmatrix} Y_0 V & U \end{bmatrix} \begin{bmatrix} \cos(\Sigma) \\ \sin(\Sigma) \end{bmatrix} V^H \quad (256)$$

where U, V and Σ are defined by the compact SVD of Δ as

$$\Delta := U\Sigma V^H \quad (257)$$

Proof:

Represent the $(N - p) \times p$ matrix B by its SVD

$$B = U_1 \Sigma V^H = \begin{bmatrix} U_1 & U_2 \end{bmatrix} \begin{bmatrix} \Sigma \\ 0 \end{bmatrix} V^H \quad (258)$$

Then

$$\begin{bmatrix} 0 & -B^H \\ B & 0 \end{bmatrix} = \begin{bmatrix} V & 0 & 0 \\ 0 & U_1 & U_2 \end{bmatrix} \begin{bmatrix} 0 & -\Sigma & 0 \\ \Sigma & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} V^H & 0 \\ 0 & U_1^H \\ 0 & U_2^H \end{bmatrix} \quad (259)$$

Using the standard matrix exponential identity $\expm(U\Lambda U^H) \equiv U\expm(\Lambda)U^H$ and the fact that

$$\begin{bmatrix} 0 & -\Sigma \\ \Sigma & 0 \end{bmatrix} = X\Lambda X^H \quad (260)$$

the matrix exponential in (254) is

$$\expm\left(\begin{bmatrix} 0 & -B^H \\ B & 0 \end{bmatrix}\right) I_{N,p} = \quad (261)$$

$$= \begin{bmatrix} V & 0 & 0 \\ 0 & U_1 & U_2 \end{bmatrix} \begin{bmatrix} \cos(\Sigma) & -\sin(\Sigma) & 0 \\ \sin(\Sigma) & \cos(\Sigma) & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} V^H \\ 0 \\ 0 \end{bmatrix} \quad (262)$$

Premultiplying (262) by $\begin{bmatrix} Y_0 & Y_{0\perp} \end{bmatrix}$ yields

$$Y = \begin{bmatrix} Y_0 V & Y_{0\perp} U_1 & 0 \end{bmatrix} \begin{bmatrix} \cos(\Sigma) \\ \sin(\Sigma) \\ 0 \end{bmatrix} V^H \quad (263)$$

Simplifying yields the desired result (256)

$$Y = \begin{bmatrix} Y_0 V & U \end{bmatrix} \begin{bmatrix} \cos(\Sigma) \\ \sin(\Sigma) \end{bmatrix} V^H \quad (264)$$

where we have used the fact that $U = Y_{0\perp} U_1$ (verified by premultiplying both sides of (258) by $Y_{0\perp}$ and noting (257)).

Using the CS form (255), with the left hand side as $Y_1 = Y$, the product $Y_0^H Y_1$ is

$$Y_0^H Y_1 = V \cos(\Sigma) V^H . \quad (265)$$

Since the subspace angles, γ_k , are defined as the arccos of the singular values of $Y_0^H Y_1$, see (249), we conclude that

$$\gamma_k = \Sigma_{kk}; k = 1 \cdots p . \quad (266)$$

The subspace distance (251), therefore, can equivalently be written in terms of these singular values as

$$dist_{SS}(Y_1, Y_0) = \sqrt{\sum_{k=1}^p \Sigma_{kk}^2} . \quad (267)$$

The geodesic distance between Y_0 and Y_1 is equal to magnitude of the tangent vector Δ , which expressed in terms of its' SVD (257) is

$$dist_{\mathcal{G}}(Y_0, Y_1) = \|\Delta\| = \|U \Sigma V^H\| = \sqrt{\sum_{k=1}^p \Sigma_{kk}^2} . \quad (268)$$

Therefore the subspace distance defined by (251) is equal to the geodesic distance between Y_0 and Y_1 in the Grassmann manifold formulation. That is

$$dist_{SS}(Y_0, Y_1) \equiv dist_{\mathcal{G}}(Y_0, Y_1) . \quad (269)$$

Finally, for an arbitrary orthonormal basis $\mathbf{b}_k$ for the tangent space $T_{Y_0} \mathcal{G} \equiv B_{Y_0}$, the tangent vector Δ expands as $\Delta = \sum_{k=1}^{N_b} \mathbf{b}_k \beta^k$. The geodesic distance is, therefore, equivalently expressed in terms of the real N_b -tuple β as

$$dist_{\mathcal{G}}(Y_0, Y_1) = \|\beta\| . \quad (270)$$

Basis Manifold

In contrast to the Grassmann manifold case, the Basis manifold formulation does not “quotient out” rotations so that the matrices Y_1 and $Y_1 Q$ are represented by different

points on the manifold. While $dist_{\mathcal{G}}(Y_1, Y_1 Q) = 0$ this is not true for the Basis manifold; that is, $dist_{\mathcal{B}}(Y_1, Y_1 Q) \neq 0$ and as a result, in general,

$$dist_{\mathcal{B}}(Y_0, Y_1) \neq dist_{SS}(Y_0, Y_1) . \quad (271)$$

For $Y_0, Y_1 \in \mathcal{B}$ and $\Delta = exp_{Y_0}^{-1} Y_1$ the geodesic distance is

$$dist_{\mathcal{B}}(Y_0, Y_1) = \|\Delta\| = \sqrt{\|P_A \Delta\|^2 + \|P_B \Delta\|^2} \quad (272)$$

where P_A and P_B are the orthogonal projectors

$$P_A : T_Y \mathcal{B} \mapsto A_Y \quad (273)$$

and

$$P_B : T_Y \mathcal{B} \mapsto B_Y . \quad (274)$$

Note that this distance (272) includes a measure of the rotation between the points and so does not correspond to the subspace distance measure desired.

Rather, the subspace distance is equal to the minimum geodesic distance between the point Y_0 and an appropriately defined fiber submanifold over the point Y_1 . This fiber submanifold, denoted $\mathcal{A}_{Y_1}$, consists of points reached via rotations of Y_1 ; that is, all points of the form $Y_1 Q$, for Q an arbitrary $p \times p$ unitary matrix, or

$$\mathcal{A}_{Y_1} \equiv \{Y \in \mathcal{B} : Y = Y_1 \cdot Q, \quad Q \in U(p) / UD(p)\} . \quad (275)$$

All $Y \in \mathcal{A}_{Y_1}$ have column identical column spans and therefore represent the same subspace as Y_1 . In the Grassman manifold formulation $\mathcal{A}_{Y_1}$ is the equivalence class containing Y_1 . In the Basis manifold formulation elements in $\mathcal{A}_{Y_1}$ are regarded as different points.

Using this definition the subspace distance for the Basis manifold case is equal to

$$dist_{SS}(Y_1, Y_0) = \min_{Y \in \mathcal{A}_{Y_1}} dist_{\mathcal{B}}(Y, Y_0) . \quad (276)$$

Figure 6. Subspace Distance on Basis Manifold

Let Y_{g0} denote the point of $\mathcal{A}_{Y_1}$ that is minimum geodesic distance from Y_0

$$Y_{g0} = \min_{Y \in \mathcal{A}_{Y_1}} (dist_B(Y, Y_0)) , \quad (277)$$

then

$$dist_{SS}(Y_1, Y_0) = dist_B(Y_{g0}, Y_0) = \|\Delta_{g0}\| \quad (278)$$

where $\Delta_{g0} \equiv exp_{Y_0}^{-1} Y_{g0}$.

The above formulation also yields the rotational distance as the geodesic distance between the intermediate point Y_{g0} , and the end point Y_1 . The rotational distance then is

$$dist_R(Y_1, Y_0) = dist_B(Y_{g0}, Y_1) = \|\Delta_{g1}\| \quad (279)$$

where $\Delta_{g1} = exp_{Y_1}^{-1} Y_{g0}$. Since both Y_1 and Y_{g0} , are points on $\mathcal{A}_{Y_1}$ it follows that $\Delta_{g1} \in \mathcal{A}_{Y_1}$.

In the following section we introduce an alternate coordinate system and show that

if $\Delta = \exp_{Y_0}^{-1} Y_1$, then to first order

$$\text{dist}SS(Y_1, Y_0) \approx \|P_B \Delta\| = \|\beta\| \quad (280)$$

and

$$\text{dist}R(Y_1, Y_0) \approx \|P_A \Delta\| = \|\alpha\|. \quad (281)$$

2.3.2 Alternate Non-Normal Coordinate System

Geodesic distances on manifolds are invariants, independent of any particular coordinate system. Although a coordinate independent quantity, subspace distance for the Stiefel/Basis manifold case is most easily analyzed using the non-normal coordinate system introduced in this section. This system approximates a normal coordinate system to first order in the sense discussed below.

These coordinates are developed as a two-leg path between Y_0 and Y_1 by way of an intermediate point $Y_g \in \mathcal{A}_{Y_1}$ with each path being a geodesic. The set of parameters which define this path become the coordinates of Y_1 . Choose the point $Y_g \in \mathcal{A}_{Y_1}$ such that $\Delta_g = \exp_{Y_0}^{-1} Y_g$ is $\Delta_g \in B_{Y_0}$ and therefore has the form $\Delta_g = Y_{0\perp} \check{B}$. Then in the full space $\mathcal{U}(N)$ the exponential map between $X_0 = \begin{bmatrix} Y_0 & Y_{0\perp} \end{bmatrix}$ and $X_g = \begin{bmatrix} Y_g & Y_{g\perp} \end{bmatrix}$ is given by

$$\begin{bmatrix} Y_g & Y_{g\perp} \end{bmatrix} = \begin{bmatrix} Y_0 & Y_{0\perp} \end{bmatrix} \exp \left(\begin{bmatrix} 0 & -\check{B}^H \\ \check{B} & 0 \end{bmatrix} \right). \quad (282)$$

Since by definition $Y_g, Y_1 \in \mathcal{A}_{Y_1}$ we have $Y_1 = Y_g Q$ for some unitary Q . Let $\check{A}$ be the anti-Hermitian matrix such that $Q = \expm(\check{A})$. Then Y_1 is related to Y_g by

$$Y_1 = Y_g Q = \begin{bmatrix} Y_g & Y_{g\perp} \end{bmatrix} \expm \left(\begin{bmatrix} \check{A} & 0 \\ 0 & 0 \end{bmatrix} \right) I_{N,p} \quad (283)$$

Substituting (282) into the above yields

$$Y_1 = \begin{bmatrix} Y_0 & Y_{0\perp} \end{bmatrix} \expm \left(\begin{bmatrix} 0 & -\check{B}^H \\ \check{B} & 0 \end{bmatrix} \right) \expm \left(\begin{bmatrix} \check{A} & 0 \\ 0 & 0 \end{bmatrix} \right) I_{N,p}. \quad (284)$$

where we have used the fact that $Q = \expm(\check{A})$ can be expressed as

$$\expm\left(\begin{bmatrix} \check{A} & 0 \\ 0 & 0 \end{bmatrix}\right) I_{N,p} = \begin{bmatrix} \expm(\check{A}) & 0 \\ 0 & I \end{bmatrix} I_{N,p} = \expm(\check{A}) . \quad (285)$$

Using the standard basis matrices A_k, B_k , (191) and (192), define the real parameters

$$\check{\alpha}^k = \text{Re}\left(\text{tr}(A_k^H \check{A})\right); \quad k = 1 \cdots N_a \quad (286)$$

$$\check{\beta}^k = \text{Re}\left(\text{tr}(B_k^H \check{B})\right); \quad k = 1 \cdots N_b . \quad (287)$$

We thus have a parameterization, or coordinates, for arbitrary points $Y_1 \in \mathcal{B}$ near Y_0 .

Denote the implied coordinate mapping as

$$(\check{\alpha}, \check{\beta}) = \check{\varphi}(Y_1) . \quad (288)$$

Equation (284) defines a two-leg path from Y_0 to Y_1 in which each of the legs is a geodesic path. The first leg travels in the direction of the tangent vector $\Delta_g \in B_{Y_0}$ at Y_0 to the intermediate point Y_g (282). The second leg from Y_g to the destination Y_1 is a pure rotation (283).

These $(\check{\alpha}, \check{\beta})$ coordinates may be related to a Riemannian normal coordinate representation of Y_1 with origin at Y_0 . Let $\Delta = \exp_{Y_0}^{-1} Y_1$ be represented with respect to the basis $\mathbf{a}_k, \mathbf{b}_k$ for the tangent space at Y_0 as

$$\Delta(\alpha, \beta) \equiv \sum_{k=1}^{N_a} \mathbf{a}_k \alpha^k + \sum_{k=1}^{N_b} \mathbf{b}_k \beta^k , \quad (289)$$

then by definition

$$Y_1 = \exp_{Y_0} \Delta(\alpha, \beta) . \quad (290)$$

This exponential map has the numerical form

$$Y_1 = \begin{bmatrix} Y_0 & Y_{0\perp} \end{bmatrix} \expm\left(\begin{bmatrix} A & -B^H \\ B & 0 \end{bmatrix}\right) I_{N,p} \quad (291)$$

where A and B are defined by (190). The Riemannian normal coordinates of Y_1 are (α, β) and the implied coordinate mapping is denoted

$$(\alpha, \beta) = \varphi(Y_1) . \quad (292)$$

The (α, β) and $(\check{\alpha}, \check{\beta})$ coordinates are related as

$$(\alpha, \beta) = h(\check{\alpha}, \check{\beta}) \quad (293)$$

where from (288) and (292)

$$h = \varphi \circ \check{\varphi}^{-1} . \quad (294)$$

Using a fixed set of basis matrices, A_k and B_k , the transformation between the matrix forms is notionally expressed in a similar way

$$(A, B) = h(\check{A}, \check{B}) \quad (295)$$

In terms of the non-normal coordinates, all points on $\mathcal{A}_{Y_1}$ have the form $(\check{A}, \check{B})$ where $\check{B}$ is constant for all points on $\mathcal{A}_{Y_1}$.

To develop the explicit functional relationship for (293), note that since Y_1 is arbitrary, from (291) and (284) we have

$$\exp \left(\begin{bmatrix} A & -B^H \\ B & 0 \end{bmatrix} \right) I_{N,p} = \exp \left(\begin{bmatrix} 0 & -\check{B}^H \\ \check{B} & 0 \end{bmatrix} \right) \exp \left(\begin{bmatrix} \check{A} & 0 \\ 0 & 0 \end{bmatrix} \right) I_{N,p} . \quad (296)$$

Recall that for arbitrary X and Y

$$\expm(X + Y) \neq \expm(X) \exp(Y) \quad (297)$$

so in general $A \neq \check{A}$ and $B \neq \check{B}$; equality holds, however, for the special case of $\check{A} = 0$ and for the special case of $\check{B} = 0$.

The following lemma shows that the Jacobian of the transformation (293) between these two non-linearly related coordinate systems is the identity matrix.

Theorem 4. Jacobean Transformation

Let (α, β) and $(\check{\alpha}, \check{\beta})$ be the coordinates defined above. Then the Jacobian transformation matrix is

$$J(\alpha, \beta; \check{\alpha}, \check{\beta}) = \begin{bmatrix} I_{N_a} & 0 \\ 0 & I_{N_b} \end{bmatrix}. \quad (298)$$

Proof: The Baker-Campbell-Hausdorff (BCH) formula for matrix exponentials is

$$\expm(U)\expm(V) = \expm\left(U + V + \frac{1}{2}[U, V] + \frac{1}{12}[[U, V], V] - \frac{1}{12}[[U, V], U] - h.o.t\right) \quad (299)$$

where the Lie Bracket of two matrices is defined as

$$[U, V] \equiv UV - VU \quad (300)$$

Identifying U and V as

$$U = \begin{bmatrix} 0 & -\check{B}^H \\ \check{B} & 0 \end{bmatrix} \quad V = \begin{bmatrix} \check{A} & 0 \\ 0 & 0 \end{bmatrix} \quad (301)$$

we have

$$[U, V] = \begin{bmatrix} 0 & -\check{B}^H \\ \check{B} & 0 \end{bmatrix} \begin{bmatrix} \check{A} & 0 \\ 0 & 0 \end{bmatrix} - \begin{bmatrix} \check{A} & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & -\check{B}^H \\ \check{B} & 0 \end{bmatrix} = \begin{bmatrix} 0 & \check{A}\check{B}^H \\ \check{B}\check{A} & 0 \end{bmatrix}. \quad (302)$$

Using this expression in (299) the matrix argument of two-leg path form (284) becomes

$$U + V + \frac{1}{2}[U, V] + o(3) = \begin{bmatrix} \check{A} + o(3) & -\check{B}^H + \frac{1}{2}\check{A}\check{B}^H + o(3) \\ \check{B} + \frac{1}{2}\check{B}\check{A} + o(3) & 0 \end{bmatrix}. \quad (303)$$

Equating the matrix arguments of (291) and (284), and using (303), yields

$$\begin{bmatrix} A & -B^H \\ B & 0 \end{bmatrix} = \begin{bmatrix} \check{A} + o(3) & -\check{B}^H + \frac{1}{2}\check{A}\check{B}^H + o(3) \\ \check{B} + \frac{1}{2}\check{B}\check{A} + o(3) & 0 \end{bmatrix} \quad (304)$$

and therefore

$$A = \check{A} + o(3), \quad (305)$$

$$B = \check{B} - \frac{1}{2}\check{B}\check{A} + o(3). \quad (306)$$

Expressing the above in expanded form with respect to the same basis matrix set A_k yields

$$\sum_{k=1}^{N_a} A_k \alpha^k = \sum_{k=1}^{N_a} A_k \check{\alpha}^k + o(3) \quad (307)$$

so that

$$\alpha^k = \check{\alpha}^k + o(3) \quad (308)$$

and therefore

$$\frac{\partial \alpha^k}{\partial \check{\alpha}^i} = \delta_i^k, \quad \frac{\partial \alpha^k}{\partial \check{\beta}^i} = 0. \quad (309)$$

An analogous result holds for the $\check{\beta}$ coordinate. Writing this in matrix form yields the desired result, that is

$$J(\alpha, \beta; \check{\alpha}, \check{\beta}) = \begin{bmatrix} I_{N_a} & 0 \\ 0 & I_{N_b} \end{bmatrix}. \quad (310)$$

The point Y_g used in the coordinate construction was defined as a point on $\mathcal{A}_{Y_1}$ such that the tangent vector $\Delta_g = \exp_{Y_0}^{-1} Y_g$ was in the space B_{Y_0} . We next show that this intermediate point Y_g is, in fact, the point on $\mathcal{A}_{Y_1}$ that is nearest to Y_0 . That is, noting (277), $Y_g = Y_{g0}$.

Lemma 5 (Lemma: The tangent vector to closest point on B_{Y_0}). *Let Y_{g0} denote the point of $\mathcal{A}_{Y_1}$ that is minimum geodesic distance from Y_1*

$$Y_{g0} = \min_{Y \in \mathcal{A}_{Y_1}} (\text{dist}_{\mathcal{B}}(Y, Y_0)). \quad (311)$$

Then $\Delta_{g0} = \exp_{Y_0}^{-1} Y_{g0}$ is an element of B_{Y_0} , that is $\Delta_{g0} \in B_{Y_0}$.

Proof: Let $Y \in \mathcal{A}_{Y_1}$ be in the neighborhood of minimum distance point Y_{g0} and let $\Delta = \exp_{Y_0}^{-1} Y$ be represented in matrix form as

$$\Delta = Y_0 A + Y_{0\perp} B. \quad (312)$$

The coefficients A and B define the normal coordinates of Y_1 . Since $\|P_B \Delta\|^2 = \|B\|_F^2 = \text{tr}(B^H B)$ and $\|P_A \Delta\|^2 = \|A\|_F^2 = \text{tr}(A^H A)$ it follows from (272) and (312) that

$$\text{dist}_{\mathcal{B}}(Y, Y_0)^2 = \|B\|_F^2 + \|A\|_F^2. \quad (313)$$

In terms of the non-normal coordinates, all points on $\mathcal{A}_{Y_1}$ have the form $(\check{A}, \check{B})$ where $\check{B}$ is constant on $\mathcal{A}_{Y_1}$. Using coordinate relation (306) the factor $B^H B$ appearing in the (313) is

$$B^H B = \check{B}^H \check{B} - \frac{1}{2} \check{B}^H \check{B} \check{A} - \frac{1}{2} \check{A}^H \check{B}^H \check{B} + \frac{1}{4} \check{A}^H \check{B}^H \check{B} \check{A} + h.o.t \quad (314)$$

$$= \check{B}^H \check{B} - \text{Re}(\check{B}^H \check{B} \check{A}) + \frac{1}{4} \check{A}^H \check{B}^H \check{B} \check{A} + h.o.t. \quad (315)$$

so that $\text{tr}(B^H B) = \|B\|_F^2$ is

$$\|B\|_F^2 = \|\check{B}\|_F^2 + \frac{1}{4} \|\check{B} \check{A}\|_F^2 + h.o.t. \quad (316)$$

In developing the above we have used the fact that

$$\text{tr}(\text{Re}(\check{B}^H \check{B} \check{A})) = 0 \quad (317)$$

since in general $\text{tr}(SA) = 0$ for S Hermitian and A anti-Hermitian. Substituting (316) into the distance formula (313) yields

$$\text{dist}_{\mathcal{B}}(Y_1, Y_0) = \|\check{B}\|_F^2 + \frac{1}{4} \|\check{B} \check{A}\|_F^2 + \|\check{A}\|_F^2 + h.o.t. \quad (318)$$

Since $\check{B}$ is fixed for all $Y \in \mathcal{A}$ we conclude that the distance is minimized when $\check{A} = 0$ and therefore, from (305), the normal coordinate parameter $A = 0$. Therefore $P_A \Delta_{g0} = 0$ and so $P_B \Delta_{g0} = \Delta_{g0}$, that is, $\Delta_{g0} \in B_{Y_0}$.

Computing the Tangent Vector /Inverse Exponential Map

Given any two points $Y_1, Y_0 \in \mathcal{B}$ there is a two-leg geodesic path between them of the form

$$Y_g = \exp_{Y_0} \Delta; \quad \Delta \in B_{Y_0} \quad (319)$$

$$Y_1 = Y_g Q. \quad (320)$$

The tangent vector $\Delta \in B_{Y_0}$ and the rotation matrix Q that define this path are computed as follows.

Expressing the tangent vector in terms of its SVD, $\Delta = U\Sigma V^H$, the above exponential map from Y_0 to Y_1 (319) can be expressed in the CS decomposition form (329) as

$$Y_1 = Y_0 V \cos(\Sigma) V^H Q^H + U \sin(\Sigma) V^H Q^H . \quad (321)$$

Since $\Delta \in B_{Y_0}$, we have $Y_0^H \Delta = 0$, so that $Y_0^H U = 0$. Right multiplying (321) by Y_0^H therefore yields

$$Y_0^H Y_1 = V \cos(\Sigma) V^H Q^H . \quad (322)$$

Let the SVD of the complex $p \times p$ matrix $Y_0^H Y_1$ be given by

$$U_1 C_1 V_1^H := Y_0^H Y_1 \quad (323)$$

with singular values of C_1 in ascending order. Identifying the factors on the r.h.s. of (322) with l.h.s. of (323) yields

$$V \equiv U_1 , \quad (324)$$

$$\cos(\Sigma) \equiv C_1 , \quad (325)$$

$$V^H Q^H \equiv V_1^H . \quad (326)$$

The last of these is equivalent to

$$Q^H \equiv V V_1^H; \quad Q = V_1 V^H . \quad (327)$$

The diagonal Σ matrix is

$$\Sigma_{kk} = \arccos([C_1]_{kk}); \quad k = 1 \cdots p . \quad (328)$$

This completes the specification of the factors V, Σ, Q in the first term of (321) in terms of the SVD of $Y_1^H Y_0$.

The remaining factor U is found by solving (321) for U in terms of the given Y_1 and Y_0 and the factors computed above.

$$US_1V^H = Y_1Q - Y_0VC_1V^H \quad (329)$$

$$U = (Y_1QV - Y_0VC_1)S_1^{-1} \quad (330)$$

where $[S_1]_{kk} = \sin(\Sigma_{kk})$. Finally, for small Σ , $C_1 \approx I$ and $S_1 \approx \Sigma$ so that

$$U \approx Y_1QV\Sigma^{-1} - Y_0V\Sigma^{-1}. \quad (331)$$

Two-Leg form Algorithm

The algorithm for computing the two-leg path, (319) and (320), between the given points Y_0 and Y_1 is summarized below.

Input: Y_0 and Y_1

Compute SVD of $Y_0^HY_1$

$$VC_1V_1^H = Y_0^HY_1. \quad (332)$$

Compute Q (327)

$$Q^H \equiv VV_1^H. \quad (333)$$

Compute Σ and S_1

$$\Sigma_{kk} = \arccos([C_1]_{kk}), \quad (334)$$

$$[S_1]_{kk} = \sin(\Sigma_{kk}). \quad (335)$$

Compute U

$$U = (Y_1QV - Y_0VC_1)S_1^{-1}. \quad (336)$$

Compute $\Delta = U\Sigma V^H$.

List of References

- [1] S. Smith, "Covariance, subspace, and intrinsic cramer-rao bounds," *IEEE Trans. Signal Process.*, vol. 53, no. 5, p. 1610 1630, 2005.
- [2] A.Edelman, T. A.Arias, and S. Smith, "The geometry of algorithms with orthogonality constraints," *SIAM Matrix Anal. Appl.*, vol. 20, no. 2, p. 303353, 1998.

- [3] R. J. Vaccaro, "Weighted subspaces fitting using subspace perturbation expansion," *Proc. ICASSP*, pp. 1973–1976, May 1998.
- [4] R. Vaccaro and Y. Ding, "Optimal subspace-based parameter estimation," *IEEE Conf. Acoustics Speech, Signal Processing*, vol. 4, no. 2, pp. 368–371, 1993.

CHAPTER 3

Intrinsic Cramér Rao Bounds for Basis Manifold and Complex Linear Model

The likelihood function that arises in the standard complex linear model may be expressed as a function of an $N \times p$ semi-unitary matrix Y . This function has the homogeneity property that $L(Y) = L(YM_p)$ for M_p an arbitrary unitary diagonal matrix. In the Stiefel manifold formulation the matrices Y and YM_p are considered different points and so the Stiefel manifold is “too big” to serve as the proper parameter space for this function. The form of the likelihood function dictates a different quotient manifold formulation in which multiplication by such unitary diagonal matrices have been removed or “quotiented” out. The quotient manifold with this desired feature is specified below and called here the Basis manifold, denoted as $\mathcal{B}_{(N,p)}$.

3.1 Complex Linear Model

Consider the complex linear model

$$z = Hn_1 + n_0 \tag{337}$$

where $n_0 \sim CN_N(0, \sigma^2 I)$; that is, n_0 , is a normal random complex N -tuple where σ^2 is known. Similarly, $n_1 \sim CN_p(0, P)$ is a normal random complex p -tuple. In this expression P is an unknown deterministic full rank Hermitian matrix and represents the correlation matrix of the source symbols n_1 . The complex $N \times p$ signal matrix $H \in \mathbb{C}^{N \times p}$ is a deterministic unknown of full rank. We refer to this case as unconstrained since there are no constraints other than rank of H and P . In section 3.8 we consider the case when H is constrained in some way.

The observation z , then, is the normal random N -tuple

$$z \sim CN_n(0, R) \tag{338}$$

where

$$R = E[zz^H] = R_s + \sigma^2 I \quad (339)$$

with signal covariance R_s given by

$$R_s(H, P) \equiv HPH^H. \quad (340)$$

Let $Z = [z_1, \dots, z_K]$ be a random sampling of (337), so that Z is a complex $N \times K$ data matrix whose columns are independent and identically distributed (iid) random vectors. The data matrix Z is an element of the measurement sample space $\mathcal{M} = C^N \times \dots \times C^N$. The pdf of Z is

$$f(Z|R) = \left(\left(\pi^N \det R \right)^{-1} \exp \left(-\text{tr} \tilde{R} R^{-1} \right) \right)^K \quad (341)$$

where $\tilde{R}$ is the sample covariance matrix (SCM) given by $\tilde{R} = K^{-1} Z Z^H$. The log-likelihood function is (ignoring constants)

$$L(H, P; \tilde{R}) = -K(\text{tr}(R^{-1} \tilde{R}) + \log \det R). \quad (342)$$

Sampling of the distribution reveals information about the unknown parameters, P and H , and the observable information is subject to estimation. It is readily seen from (339) and (340) that this model precludes the unambiguous and simultaneous estimation of the unknowns H and P . To see this, note that for any invertible T , substituting $H = HT^{-1}$ and $n_1 = Tn_1$ into (339) leaves the measurement unchanged. Thus, both the pairs (H, P) and (HT^{-1}, TPT^H) define the same measurement distribution. As noted in [1] “whenever two parameters give rise to the same measurement distribution, they are indistinguishable, in the sense that no argument based on the observed measurement can be used to favor one parameter over the other as estimator”. We also observe that, in general, the pairs (H, P) and (HT^{-1}, P) define different measurement distributions. Since H and HT^{-1} define the same subspace, $\text{col}(H)$, it is evident for fixed P that the

subspace is not sufficient to uniquely specify the distribution for this model.^{1 2}

An alternate parameterization of the problem is provided by the eigen decomposition of the signal covariance (340). The eigen decomposition defines a mapping from the (H, P) parameter space to (Y, Λ) parameter space by

$$[Y, \Lambda] = \text{eig}(HPH^H) . \quad (343)$$

The “eig” function is viewed here as the transformation or mapping

$$\text{eig} : C^{N \times p} \times \mathcal{P}_p \mapsto \mathcal{B} \times \mathcal{D}_p . \quad (344)$$

In this expression $\mathcal{P}_p$ is the set all positive-definite Hermitian matrices and $\mathcal{D}_p \subset \mathcal{P}_p$ is the subset of positive-definite diagonal matrices. Under this transformation the complex linear model (337) is restated as

$$z = Yn_1 + n_0, \quad Y \in \mathcal{B} \quad (345)$$

where $n_1 \sim N_p(0, \Lambda)$ and $n_0 \sim N_N(0, \sigma^2 I)$ as in the original formulation. The set $\mathcal{B}$ appearing in (344) and (345) is discussed in the following.

The eigenvector matrix Y and the signal matrix H are related as $Y = HT$ for some T nonsingular and so $\text{col}(H) = \text{col}(Y)$. In addition to yielding the subspace defined by $\text{col}(H)$, the eigenvector matrix Y also reveals information about the source correlation matrix P . In terms of this alternate parameterization the signal covariance R_s (340) is

$$R_s(Y, \Lambda) = Y\Lambda Y^H \quad (346)$$

and the full covariance (337) is therefore

$$R = R(Y, \Lambda, \sigma^2) = R_s + \sigma^2 I = Y\Lambda_s Y^H + \sigma^2 Y_\perp Y_\perp^H \quad (347)$$

¹Observe that $R_s = HPH^H = HT^{-1}TPT^H T^{-H}H^H = HPH^H$

²[2] “Therefore, only the column span of H and the covariance matrix of z may be measured, and we ask how accurately we are able to estimate this subspace, i.e., the column span of H , in the presence of the unknown covariance matrix P .”

where $Y_{\perp}^H Y = 0$ and $\Lambda_s \equiv \Lambda + \sigma^2 I$.

The estimation problem is thus recast in terms of the semi-unitary matrix Y and the diagonal covariance Λ . The collections of semi-unitary matrices found in signal processing applications can be analyzed using several different manifold formulations, with the choice dictated by the given statistical model. The collection of all $N \times p$ semi-unitary matrices constitutes the complex Stiefel manifold (see section 2.2.7) defined here in concrete form as a subset of $C^{N \times p}$ by

$$St_{(N,p)} \equiv \{Y \in C^{N \times p} : Y^H Y = I\} . \quad (348)$$

A well-posed estimation problem requires enough parameters to completely specify the distribution but not too many so that the problem is ambiguous (by ambiguous we mean different parameters values that yield identical measurements). This was the situation in the initial formulation of the problem, where it was noted that (H, P) and $(HT, T^{-1}PT^{-H})$ defined the same distribution. For the given log likelihood function (342) involving complex data, the complex Stiefel manifold is “too big ” for problem in the sense that we can choose different elements of $St_{(N,p)}$ that yield identical measurement distributions. Suppose that the parameter Λ is fixed. Then both Y and $Y\Lambda_Q$, Λ_Q an arbitrary unitary diagonal matrix, define the same signal covariance R_s and so give rise to the same measurement distribution (same set of numbers). This follows since diagonal matrices commute $\Lambda_Q \Lambda \Lambda_Q^* = \Lambda \Lambda_Q \Lambda_Q^* = \Lambda$ since $\Lambda_Q \Lambda_Q^* = I$. Since Y and $Y\Lambda_Q$ are different points on the complex Stiefel manifold, we are led to an alternate quotient manifold defined by the equivalence relation $Y \sim Y\Lambda_Q$. The set of elements in $St_{(N,p)}$ that are so related defines an equivalence class, denoted by $[Y]$, and given by

$$[Y] = \{Y\Lambda_Q : \Lambda_Q \in UD(p)\} . \quad (349)$$

All elements in the equivalence class define the same distribution and so the quotient space $\mathcal{B}_{N,p} \equiv St_{(N,p)}/UD(p)$ is the natural parameter space for the estimation

problem. The quotient manifold $\mathcal{B}_{N,p}$ is formed from the complex Steifel manifold by “quotienting out” multiplications by unitary diagonal matrices (denoted here as $UD(p)$). Said another way, a point in the Basis manifold is a particular subset of the semi-unitary matrices defined (349), and the Basis manifold itself is the collection of all these subsets. Note that in the Grassmann manifold formulation the defining equivalence relation is $Y \sim YQ$, Q arbitrary unitary; in this formulation Y and YQ represent the same point. Since, in general, $R_s(YQ, \Lambda) \neq R_s(Y, \Lambda)$, see (346), we conclude that the Grassmann manifold is “too small” to properly parameterize L . While the Steifel manifold is “too big” and the Grassmann manifold is “too small” the basis manifold $\mathcal{B}_{N,p}$ is just right for the problem. Since we assume that N and p are known throughout the remainder we drop the subscripts and denote the basis manifold as simply $\mathcal{B}$.

The parameters of the log-likelihood are then $Y \in \mathcal{B}$ and the real p -tuple of positive eigenvalues specified in the diagonal matrix $\Lambda \in \mathcal{D}_p$. The complete parameter space of interest is then the product manifold

$$\mathcal{B} \times \mathcal{D}_p. \quad (350)$$

This has real dimension $\dim(\mathcal{B} \times \mathcal{D}_p) = 2p(N - p) + p^2$ (see section 2.2.5).

3.2 Estimation on Riemannian (Quotient) Manifolds

The unitary group $\mathcal{U}(N)$, the collection of all $N \times N$ unitary matrices, may be viewed a submanifold of the complex space $C^{N \times N}$, the set of all $N \times N$ complex matrices. Manifolds such as $\mathcal{U}(N)$ and the usual 2-sphere are examples of Riemannian manifolds, general curved spaces with smoothly verifying metric functions. The submanifold inherits an inner product or metric from the parent manifold and this establishes the intrinsic geometry of the space, most importantly geodesics (see chapter ?? and [3]). The Steifel and Basis manifolds likewise inherit their metric functions from their formulations as a quotient manifolds on $U(N)$.

Estimation problems where the parameter space is a Riemannian manifold differ from the more usual parameter estimation in two important ways. In the typical estimation problem (e.g; angle and Doppler) a fixed set of coordinates for the parameter space is naturally provided. In contrast, for the general manifold case, no set of intrinsic or preferred coordinates are given. This is the situation with position estimation on the Euclidean flat plane, for example, where no particular set of coordinates are naturally provided and Cartesian and polar are equally valid choices. Estimators and evaluators are free to choose among these or any others. The second difference with the usual estimation problem is that more general curved Riemannian manifolds such as the sphere are not themselves vector spaces and so the vector space concepts implicitly used in classical Cramér-Rao bounds (CRB) approaches are not immediately applicable.

In these situations, a universally meaningful estimation error measure is the distance between the estimate and the true location. Distance is an invariant or intrinsic quantity dependent on the given metric but independent of any particular choice of coordinates (e.g., the 2-D mean square positional error on the sphere).

In [2] a methodology is established for computing the CRB when the parameter space is an arbitrary Riemannian Quotient manifolds. The frequently encountered example of estimating an unknown subspace is analyzed and the author gives intrinsic CRBs in the form of matrix inequalities relating the covariance of estimators and the Fisher information. The classical bounds developed for Euclidean spaces are generalized to Riemannian Quotient manifolds, such as $\mathcal{B}$, via the exponential map, i.e., geodesics emanating from the estimate to points in the parameter space³. In section 10 the approach detailed in [2] is used to compute the full FIM for the Basis manifold formulation of the unconstrained problem (342). The subspace submatrix of the FIM for this case is shown to be identical in form to that given in [2] for the Grassmann manifold formulation. The

³ “Just as with classical bounds, the unbiased intrinsic CRBs depend asymptotically only on the Fisher information and do not depend in any nontrivial way on the choice of measurement units (e.g., degree versus radians) [2].

rotation submatrix of the FIM which appears in the Basis manifold formulation (but absent in the Grassmann case) is also computed here.

In section 3.8 the FIM for the constrained version of (342) is developed and closed form results are provided for the important special case when the signal matrix $H(\theta)$ that defines the constraint has a modal form.

Exponential Map and Normal Coordinates

Geodesics on general curved Riemannian manifolds are defined as distance minimizing curves between points. For the special case of the standard Euclidean (flat) manifold $\mathcal{E}$, a geodesic emanating from the point Y_0 in the direction of the vector Δ is the straight line $Y(t) = Y_0 + t \cdot \Delta$. The vector Δ is an element of a vector space with origin at Y_0 . This vector space is denoted $T_{Y_0}\mathcal{E}$ where “ T ” is read as “tangent” and Y_0 is the origin or base which corresponds to the “head” of the vector.

For the unitary group manifold $U(N)$, a geodesic emanating from the point X_0 in the direction $\Delta = X_0 W$ (W anti-Hermitian) is the curve given by exponential map

$$X(t) = X_0 \expm(tW) . \quad (351)$$

Note that

$$\Delta \equiv \left. \frac{dX(t)}{dt} \right|_{t=0} = X_0 W \quad (352)$$

so that the description of Δ as a tangent is consistent with intuition. The collection of all such tangents through the point X_0 defines a vector space termed the tangent space. This space is denoted as $T_{X_0}U(N)$ and the tangent vector Δ is an element of this space.

The short-hand notation for the above exponential map at the point X_0 with tangent vector $\Delta = X_0 \cdot W$ (351) is ⁴

$$X(t) = \exp_{X_0}(\Delta \cdot t) . \quad (353)$$

⁴ Note that for W anti-Hermitian $Q(t) = \expm(tW)$ is a unitary matrix so that $X(t) = X_0 Q(t)$ is also in $U(N)$ (group property: closure under multiplication).

The exponential or geodesic map plays a central role in the development of the intrinsic CRB for $\mathcal{U}(N)$ and related quotient manifolds. For the $\mathcal{B}$ manifold, a quotient manifold of $U(N)$, geodesics are naturally defined by constraining the set of tangent vectors to a subspace of $T_{X_0}U(N)$, termed the horizontal space, that corresponds to $T_{X_0}\mathcal{B}$. Elements of this subspace have the same general form as with (352), with the constraint expressed through the matrix W in a partitioned form as

$$W = \begin{bmatrix} A & -B^H \\ B & 0 \end{bmatrix}. \quad (354)$$

The exponential map can be expressed in a more concrete form if we select only the first p columns of (351) by right multiplying by the selection matrix $I_{N,p}$. Writing the general unitary matrix X in partitioned form as $X = \begin{bmatrix} Y & Y_\perp \end{bmatrix}$, where Y denotes the first p columns we have

$$Y(t) = \begin{bmatrix} Y_0 & Y_{0\perp} \end{bmatrix} \exp(tW) I_{N,p} \quad (355)$$

where $Y(t) = X(t)I_{N,p}$. The tangent vector

$$\Delta \equiv \left. \frac{dY(t)}{dt} \right|_{t=0} = \begin{bmatrix} Y_0 & Y_{0\perp} \end{bmatrix} W \cdot I_{N,p} = Y_0 A + Y_{0\perp} B \quad (356)$$

is as an element of the tangent space $T_Y\mathcal{B}$. This space is evidently partitioned as the direct sum

$$T_Y\mathcal{B} = A_Y \oplus B_Y \quad (357)$$

where elements of A_Y and B_Y have the general form $Y A$ and $Y_\perp B$ respectively. The exponential map on $\mathcal{B}$ defined by (355) and (354) is denoted in short hand as $Y(t) = \exp_{Y_0} t \Delta$ where $\Delta \in T_{Y_0}\mathcal{B}$ has the form (356).

Denoting the basis vectors for A_Y and B_Y as $\mathbf{a}_k$ and $\mathbf{b}_k$, respectively, an arbitrary Δ expands as

$$\Delta = \sum_{k=1}^{N_a} \mathbf{a}_k \alpha^k + \sum_{k=1}^{N_b} \mathbf{b}_k \beta^k \quad (358)$$

where α and β are the real components of Δ with respect to the given basis.⁵

The exponential map $Y = \exp_{Y_0} \Delta$ which maps Y_0 to Y , based on Δ , is invertible (for small Δ this map is well-approximated numerically by $Y_0 + \Delta$). Given an arbitrary point Y near Y_0 this inverse exponential map, which assigns a tangent vector $\Delta \in T_{Y_0} \mathcal{B}$ to the pair Y_0, Y , is denoted as $\Delta = \exp_{Y_0}^{-1} Y$. This defines the difference vector between the points, and is the manifold generalization of the difference vector $\Delta = Y - Y_0$ in the usual flat space. If $\Delta = \exp_{Y_0}^{-1} Y$ is expressed with respect to a given basis as (358) then we see the parameters α, β provide an alternate means of labeling the point Y . These parameters are the so-called normal coordinates of Y . The exponential map and its inverse provide the transformation between extrinsic Y coordinates and intrinsic (α, β) normal coordinates. Since the choice of basis vectors is not unique so neither are the normal coordinates.

Functions defined on the product manifold (350) may be expressed in terms of any convenient set of coordinates. Thus the log-likelihood function $L(Y, \Lambda)$ may alternately be written as $L(\alpha, \beta, \Lambda)$. This situation is analogous to the alternate coordinate representations of points on the usual sphere S^2 in 3-space. An arbitrary point $\mathcal{P} \in S^2$ may be represented in extrinsic (x, y, z) Cartesian coordinates or by latitude-longitude (u, v) coordinates with the systems related by the invertible transformation $(x, y, z) = h(u, v)$. A real function on S^2 is the mapping $L: S^2 \rightarrow R$ and its value at the point $\mathcal{P}$ is $L(\mathcal{P})$. For the two above coordinate systems the value is expressed as $L(\mathcal{P}) = L_Y(x, y, z) = L_\varphi(u, v)$ (see section A.1).

The distance between two points on a Riemannian Manifold is the path length of geodesic curve between them. It was shown in section 2.2.7 that this length is equal to the magnitude of the tangent vector $\Delta = \exp_{Y_0}^{-1} Y$ so that

⁵We use the word vector in the abstract sense, not as a $N \times 1$ column matrix; this is sometimes confusing.

$$dist_{\mathcal{B}}(Y_1, Y_0) \equiv \|\Delta\| = \sqrt{\|\alpha\|^2 + \|\beta\|^2} . \quad (359)$$

Estimation Problem

We consider an estimation problem on the product manifold $\mathcal{B} \times \mathcal{D}$ based on measurements in $\mathcal{M} = C^N \times \cdots \times C^N$, with the probability density function of the measurement given a parameter $(Y \in \mathcal{B}, \lambda \in \mathcal{D})$ denoted as

$$f(:, Y, \lambda) : \mathcal{M} \rightarrow R . \quad (360)$$

Let

$$L(:, Z) : \mathcal{B} \times \mathcal{D} \rightarrow R \quad (361)$$

be the associated log likelihood function $L(Y, \lambda; Z) = \log f(Z; Y, \lambda)$ where $Z \in \mathcal{M}$.

The measurement data matrix Z is a random sample whose probability density function is shaped by the unknown parameters Y and Λ . An estimator of Y , denoted $\hat{Y}$, is a mapping from the probability space $\mathcal{M}$ to the parameter space $\mathcal{B}$, that is, $\hat{Y} : \mathcal{M} \rightarrow \mathcal{B}$ or $\hat{Y}(Z)$. Similarly a generic estimator of Λ is denoted $\hat{\Lambda}$ and is given by an analogous mapping $\hat{\Lambda} : \mathcal{M} \rightarrow \mathcal{D}$.

The estimation error in the general manifold case is the tangent vector $\hat{\Delta}(Z) \equiv \exp_Y^{-1} \hat{Y}(Z)$. Going forward for conciseness we write $\hat{Y}$ to mean $\hat{Y}(Z)$ and $\hat{\Delta}$ to mean $\hat{\Delta}(Z)$. By definition, $\hat{\Delta} \in T_Y \mathcal{B}$, and given a partitioned orthonormal basis for this space the estimation error vector expands as

$$\hat{\Delta} = \sum_{k=1}^{N_a} \mathbf{a}_k \hat{\alpha}^k + \sum_{k=1}^{N_b} \mathbf{b}_k \hat{\beta}^k . \quad (362)$$

For conciseness we again write $\hat{\alpha}, \hat{\beta}$ to mean the stochastic quantities $\hat{\alpha}(Z), \hat{\beta}(Z)$.

For unbiased estimators the error covariance is given by the tensor product $E[\hat{\Delta} \otimes \hat{\Delta}]$. Viewed as a matrix defined by the coefficients with respect to the given partitioned

basis, this covariance matrix is

$$C = \begin{bmatrix} C_{\alpha\alpha} & C_{\beta\alpha} \\ C_{\alpha\beta} & C_{\beta\beta} \end{bmatrix} = E \left[\begin{bmatrix} \hat{\alpha} \\ \hat{\beta} \end{bmatrix} \begin{bmatrix} \hat{\alpha} \\ \hat{\beta} \end{bmatrix}^T \right] = E \begin{bmatrix} \hat{\alpha}\hat{\alpha}^T & \hat{\beta}\hat{\alpha}^T \\ \hat{\alpha}\hat{\beta}^T & \hat{\beta}\hat{\beta}^T \end{bmatrix} \quad (363)$$

where for convenience $C_{\alpha\alpha} \equiv E[\hat{\alpha}\hat{\alpha}^t]$, etc. (see section A.1 in particular (A.89)).

Denoting the orthogonal projector onto $B_Y < T_Y\mathcal{B}$ as P_B we have $P_B\hat{\Delta} = \sum_{k=1}^{N_b} \mathbf{b}_k \hat{\beta}^k$ and so

$$E[\|P_B\hat{\Delta}\|^2] = E[\|\hat{\beta}\|^2] = \text{tr}C_{\beta\beta} \quad (364)$$

where we have used the standard vector identity $\|\beta\|^2 = \text{tr}(\beta\beta^t)$.

3.3 Intrinsic Mean Square Error Measure

In this section we define the mean square error measures used to analyze estimator performance. Because there are no intrinsic or preferred coordinates in the general manifold case, different system evaluators are free to chose different coordinate systems. The error measure chosen to evaluate estimator performance should therefore be an invariant quantity, or one that is independent of the arbitrarily chosen coordinate system. Given such an error measure, all evaluators will record the same “error” irrespective of their choice of coordinates.

3.3.1 Subspace and Rotation Angles

The difference between two elements of the basis manifold can be represented in terms of two sets of angles (subspace and rotational angles) that correspond to the natural partitioning of the tangent space (357). We first define these angles independently of the geometric discussion.

Let Y_0 and Y_1 be two $N \times p$ semi-unitary matrices and denote the SVD of the product $Y_1^H Y_0$ as

$$U\Sigma V^H = Y_1^H Y_0. \quad (365)$$

The p -tuple of subspace angles, γ , between Y_0 and Y_1 is defined by

$$\gamma_k = \arccos(\Sigma_{kk}); k = 1, \dots, p. \quad (366)$$

The root sum square (RSS) of these angles is referred to as the subspace distance and denoted $distSS(Y_1, Y_0)$,

$$distSS(Y_1, Y_0) \equiv \|\gamma\| \equiv \sqrt{\sum_{k=1}^p \gamma_k^2}. \quad (367)$$

The p -tuple of rotational angles denoted ν is defined implicitly by

$$\exp(i\nu_k) \equiv \rho_k; \quad k = 1 \dots p \quad (368)$$

where ρ_k is the k -th singular value of $p \times p$ matrix $(Q - \text{diag}(Q))$ where $Q = U^H V$. The subtraction of $\text{diag}(Q)$ term above is in recognition of the Basis manifold quotient structure for which Y , and $Y\Lambda_Q$, Λ_Q unitary diagonal, represent the same point. The RSS of these angles is referred to as the rotation distance and denoted $distR(Y_1, Y_0)$,

$$distR(Y_1, Y_0) = \sqrt{\sum_{k=1}^p \log(\rho_k)^2} = \sqrt{\sum_{k=1}^p \nu_k^2}. \quad (369)$$

3.3.2 Subspace Angles on Grassmann and Basis Manifolds

In this section (see also section 2.3.1) we relate the subspace distance defined above to geometric distances on the Grassmann and Basis manifolds. In the Grassmann manifold case this subspace distance is equal [2] to the geodesic distance between the points on the manifold, represented as the semi-unitary matrices, say Y_1 and Y_0 ; that is

$$distSS(Y_1, Y_0) = dist_G(Y_1, Y_0). \quad (370)$$

For the Basis manifold formulation this is not the case. Rather, the subspace distance here is equal to the minimum geodesic distance between Y_0 and a fiber submanifold containing Y_1 . This submanifold, denoted $\mathcal{A}_{Y_1}$, consists of all points related to Y_1 via a rotation; that is, matrices of the form $Y_1 Q$ for Q unitary. For the Basis manifold we have

$$distSS(Y_1, Y_0) = \min_{Y \in \mathcal{A}_{Y_1}} (dist_B(Y, Y_0)). \quad (371)$$

In the Grassmann formulation $\mathcal{A}_{Y_1}$ is recognized as the equivalence class containing Y_1 . Denoting the point of $\mathcal{A}_{Y_1}$ that is minimum geodesic distance from Y_0 as Y_{g0} we have

$$dist_{SS}(Y_1, Y_0) = dist_{\mathcal{B}}(Y_{g0}, Y_0) . \quad (372)$$

The rotational distance (369) then is equivalent to geodesic distance between Y_{g0} and Y_1

$$dist_R(Y_1, Y_0) = dist_{\mathcal{B}}(Y_{g0}, Y_1) . \quad (373)$$

Defining $\Delta_{g0} = exp_{Y_0}^{-1} Y_{g0}$ and $\Delta_{g1} = exp_{Y_1}^{-1} Y_{g0}$ the above distances are $dist_{SS}(Y_1, Y_0) = \|\Delta_{g0}\|$ and $dist_R(Y_1, Y_0) = \|\Delta_{g1}\|$. It can be shown that Δ_{g0} corresponds to a pure subspace motion (no rotation) so that $\Delta_{g0} \in B_{Y_0}$ (see figure 6).

The geodesic distance between points on Riemannian manifolds such as $\mathcal{B}$ is independent of the choice of coordinates. For numerical calculations some set of coordinates are required. Normal coordinates are a standard choice but the analysis here is facilitated by the alternate non-normal $\check{\phi}$ coordinates defined below (see also section 2.3.2). Since $\Delta_{g0} \in B_{Y_0}$ we have $\Delta_{g0} = Y_{0\perp} \check{B}$ for some matrix $\check{B}$ and so the point Y_{g0} can be expressed as

$$Y_{g0} = \begin{bmatrix} Y_0 & Y_{0\perp} \end{bmatrix} exp(W_B) I_{N,p} \quad (374)$$

where

$$W_B = \begin{bmatrix} 0 & -\check{B}^H \\ \check{B} & 0 \end{bmatrix} . \quad (375)$$

Since by construction, $Y_1 = Y_{g0} Q$ for some unitary Q , combining this with (374) yields the relation between Y_1 and Y_0

$$Y_1 = \begin{bmatrix} Y_0 & Y_{0\perp} \end{bmatrix} exp(W_B) exp(W_A) I_{N,p} \quad (376)$$

where

$$W_A = \begin{bmatrix} \check{A} & 0 \\ 0 & 0 \end{bmatrix} \quad (377)$$

and where we have used $Q = \expm(\check{A})$. The desired alternate non-normal coordinates are determined using inner product (110) as

$$\check{\alpha}_k = \text{Re} \left(\text{tr}(A_k^H \check{A}) \right); \quad k = 1 \cdots N_a \quad (378)$$

$$\check{\beta}_k = \text{Re} \left(\text{tr}(B_k^H \check{B}) \right); \quad k = 1 \cdots N_b \quad (379)$$

where A_k, B_k are the standard set of orthogonal basis matrices (see section A.1).

The above defines an algorithm for assigning a unique pair $\check{\alpha}, \check{\beta}$ to each point Y in the neighborhood of Y_0 . The result of this process is the coordinate mapping

$$\check{\varphi} : \mathcal{B} \mapsto R^{N_a} \times R^{N_b}. \quad (380)$$

These coordinates are closely related but not identical to normal coordinates. The fact that they are not identical follows from the fact that in general $\exp(W_A + W_B) \neq \exp(W_A)\exp(W_B)$. The advantage of this choice is that once the $\check{\varphi}$ coordinates of Y_1 are calculated as $(\check{\alpha}, \check{\beta})$ those of Y_{g0} are readily revealed as $(0, \check{\beta})$. The normal coordinates of Y_{g0} are likewise $(0, \check{\beta})$ but those of Y_1 are only approximately equally to $(\check{\alpha}, \check{\beta})$. From (359) we see that $\text{dist}_{\mathcal{B}}(Y_0, Y_{g0}) = \|\check{\beta}\|$ and so from (371) we have

$$\text{dist}_{SS}(Y_1, Y_0) = \|\check{\beta}\|. \quad (381)$$

We seek a bound on the subspace distance error measure

$$\epsilon_{SS} \equiv E[\text{dist}_{SS}^2(\hat{Y}_0, Y_0)]^{\frac{1}{2}} \quad (382)$$

where $\hat{Y}_0$ is an unbiased estimator of Y_0 . Since this distance is given by $\|\check{\beta}\|$ we develop the bound for problem in terms of this N_b -tuple. Using the Baker-Campbell-Hausdorff formula (299) it can be shown the Jacobian of the transformation between the φ normal coordinates and the non-normal $\check{\varphi}$ coordinates is the identity matrix. Because of this relationship, we are able to develop the intrinsic CRB on (382) by considering the FIM for the log likelihood function expressed in normal coordinates as $L(\alpha, \beta)$.

If P_A and P_B are the projection operators from $T_{Y_0}\mathcal{B}$ to A_{Y_0} and B_{Y_0} respectively then

$$\text{dist}SS(\hat{Y}_0, Y_0) = \|\Delta_{g0}\| = \|\check{\beta}\| \approx \|P_B\Delta\| = \|\beta\| \quad (383)$$

and

$$\text{dist}R(\hat{Y}, Y) = \|\Delta_{g1}\| = \|\check{\alpha}\| \approx \|P_A\Delta\| = \|\alpha\| . \quad (384)$$

3.4 Intrinsic CRB

Cramér-Rao bounds relate the covariance matrix of estimators to the Fisher information matrix (FIM) of an estimation problem through matrix inequalities. In the general manifold case the Fisher information metric tensor $\mathbf{F}$ (see section A.3.2) is defined with respect to a given coordinate basis by

$$F_{i,j} \equiv \mathbf{F}[\mathbf{e}_i, \mathbf{e}_j] = E[\nabla L[\mathbf{e}_i], \nabla L[\mathbf{e}_j]] ; \quad i, j = \cdots N . \quad (385)$$

In this expression $\nabla L[\mathbf{e}_i]$ is the directional derivative of L in the direction of the coordinate basis vector $\mathbf{e}_i$ at the point of interest. This coordinate basis is formed from the tangents to the coordinate functions $\mathbf{e}_k = \frac{d\theta_{(k)}(t)}{dt}$. To emphasize this relation the coordinate basis vectors are often denoted symbolically as $\mathbf{e}_k \equiv \frac{\partial}{\partial\theta^k}$, while the dual coordinate basis is denoted $\mathbf{e}^k \equiv d\theta^k$. Alternate notations for this directional derivative are

$$\nabla L[\mathbf{e}_i] = \nabla L \left[\frac{\partial}{\partial\theta^i} \right] = \frac{\partial L}{\partial\theta^i} . \quad (386)$$

Using the last of these, the component matrix takes the more familiar form

$$F_{i,j} \equiv E \left[\frac{\partial L}{\partial\theta^i}, \frac{\partial L}{\partial\theta^j} \right] ; \quad i, j = \cdots N \quad (387)$$

In the classical estimation problem of a vector parameter $\theta \in R^p$ in the distribution $f(\theta)$, the CRB is given by the inverse of the Fisher Information matrix F and the lower bound is expressed as the matrix inequality

$$C \succeq F^{-1} \quad (388)$$

where C is the estimator error covariance, $C = E[(\theta - \hat{\theta})(\theta - \hat{\theta})^T]$. The classical proof of the inequality (388) relies on the fact [4] that the partial derivative of the estimation error vector is

$$\frac{\partial}{\partial \theta}(\theta - \hat{\theta}) = -I . \quad (389)$$

In the general Riemannian manifold case, the role of the error vector $\hat{\Delta} = \theta - \hat{\theta}$ is played by the error tangent vector $\hat{\Delta} = \exp_{\hat{\theta}}^{-1}\hat{\theta}$. The usual partial derivative $\frac{\partial}{\partial \theta}$ becomes the covariant derivative denoted by the symbol ∇ which pulls in the manifold's geometry (curvature) into the derivative definition. The term $\frac{\partial}{\partial \theta}(\hat{\theta} - \theta)$ (389) is thus replaced in the intrinsic CRB by $\nabla \exp_{\hat{\theta}}^{-1}\hat{\theta}$ which evaluates to

$$\nabla \exp_{\hat{\theta}}^{-1}\hat{\theta} = -I + \text{curvature terms} \quad (390)$$

where “curvature terms” represent second-order and higher terms involving the manifold's sectional and Riemannian curvatures. This key difference results in additional terms appearing in the intrinsic CRB.

The intrinsic generalization of the CRB for the general Riemannian manifold case is

Theorem 6 (Cramer-Rao [2]). *Let $f(z|Y)$ be a family of pdfs parameterized by $Y \in \mathcal{M}$, let $L(Y)$ be the log-likelihood function, $\mathbf{F} = E[\nabla L \otimes \nabla L]$ be the Fisher information metric, and ∇ be a Riemannian connection on $\mathcal{M}$. For any an unbiased estimator $\hat{Y}$ of Y the C covariance of the tangent vector $\Delta = \exp_Y^{-1}\hat{Y}$ satisfies the matrix inequality*

$$C \succeq F^{-1} - \frac{1}{3}(R_m(F^{-1})F^{-1} - F^{-1}R_m(F^{-1})) . \quad (391)$$

In this expression C is the matrix form of $E[\hat{\Delta} \otimes \hat{\Delta}]$ and F is the component matrix corresponding the tensor $\mathbf{F}$ in a arbitrary coordinates (see section 3.2). 2) Neglecting the curvature terms at small errors

$$C \succeq F^{-1} . \quad (392)$$

At high SNR estimation errors are small and these additional terms become negligible. Thus just as with classical bounds, the unbiased intrinsic CRBs depend asymptotically only on the Fisher information. The error variance is then bounded by

$$\text{tr}C \succeq \text{tr}F^{-1} . \quad (393)$$

Alternate form of FIM

Reference [2] gives an alternate form for the “Fisher information that employs the Hessian of the log-likelihood function for arbitrary affine connections - a useful tool because in the great majority of applications the second-order terms are much easier to compute”.

The Hessian tensor of L is denoted as $\nabla^2 L$, or in coordinates as $L_{;ij} = \nabla^2 L [\mathbf{e}_i, \mathbf{e}_j]$, where $\mathbf{e}_i$ is a coordinate basis for the tangent space. In arbitrary curvilinear coordinates θ , at the point θ_0 the components $L_{;ij}$ evaluate to

$$L_{;ij} = \frac{\partial^2 L}{\partial \theta^i \partial \theta^j} - \sum \Gamma_{ij}^k(\theta_0) \frac{\partial L}{\partial \theta^k} \quad (394)$$

where the derivatives are evaluated at θ_0 and Γ_{ij}^k are the Christoffel symbols which reflect the bending and twisting of the coordinate curves (see section A.4).

Theorem 1 from [2], restated below, allows us to compute the Fisher information on an arbitrary manifold using the same approach as is done in classical Cramér-Rao analysis.

Theorem 7 ([2]). *Let $f(Z|Y)$ be a family of pdfs parameterized by $\theta \in \mathcal{M}$ $L = \log f$ be the log-likelihood function, and (385) be the Fisher information tensor. Suppose that the pdf satisfies the regularity condition*

$$\int \nabla L(z|\alpha) f = E [\nabla L] = 0 \quad (395)$$

or equivalently in the arbitrary coordinates θ

$$E \left[\frac{\partial L}{\partial \theta^k} \right] = \int \frac{\partial L}{\partial \theta^k} f = 0; \quad k = 1 \cdots N. \quad (396)$$

Then for any affine connection ∇ defined on $\mathcal{M}$

$$F_{i,j} = -E [L_{;ij}] = -E [\nabla^2 L [\mathbf{e}_i, \mathbf{e}_j]]; \quad i, j = \cdots N. \quad (397)$$

Proof:

$$E [L_{;ij}] = \int L_{;ij} f = \int \frac{\partial^2 L}{\partial \theta^i \partial \theta^j} f - \sum \Gamma_{ij}^k \int \frac{\partial L}{\partial \theta^k} f. \quad (398)$$

The second term in above expression vanishes as a result of the regularity condition.

Integration by parts of the first term in the sum as in classical CRB analysis yields

$$E [L_{;ij}] = \int \frac{\partial L}{\partial \alpha^i} \frac{\partial L}{\partial \alpha^j} = - \int \frac{\partial^2 L}{\partial \theta^i \partial \theta^j} \quad (399)$$

Computing the Hessian Tensor

As discussed in section A.4.1 the Hessian, $\nabla^2 L$, of the function L is defined by the quadratic form

$$\nabla^2 L[\Delta, \Delta] \equiv \frac{d^2}{dt^2} L_y(Y(t))|_{t=0} \quad (400)$$

where $Y(t)$ is the geodesic emanating in the direction Δ . Let $\mathbf{a}_k, \mathbf{b}_k$ be an orthonormal basis for $T_{Y_0} \mathcal{B} = A_{Y_0} \oplus B_{Y_0}$. The geodesic emanating in the direction of the $\Delta = \mathbf{b}_k \in B_{Y_0}$ corresponds to the β^k coordinate curve. In extrinsic coordinates this geodesic is

$$Y_{(k)}(t) \equiv \exp_{Y_0}(\mathbf{b}_k t). \quad (401)$$

Using (400) we have

$$\frac{\partial^2 L}{\partial \beta^k \partial \beta^k} \equiv \nabla^2 L[\mathbf{b}_k, \mathbf{b}_k] = \frac{d^2 L_y(Y_{(k)}(t))}{dt^2}|_{t=0}; \quad k = 1 \cdots N_b. \quad (402)$$

The second derivatives of the left-hand side of (402) represent the ordinary Hessian matrix of L interpreted to be a scalar function defined on R^{N_b} . To compute the above substitute the expanded form of $Y_{(k)}(t)$ given by (A.39) into the right hand side of (402) and compute the derivative (see section A.5). The cross terms of the full Hessian are found via polarization (A.126) based on (400) and vectors of the form $\Delta = \mathbf{b}_i + \mathbf{b}_j$, $\Delta = \mathbf{a}_i + \mathbf{b}_j$, etc. as required. This approach is used in the next section to compute the full FIM for log likelihood function (342).

3.5 Nuisance Parameters

The form of the statistical model (337) dictated the choice of the Basis manifold rather than the Grassmann manifold as the proper parameter space for both analysis and algorithm development. In estimation applications, a nuisance parameter is “any parameter which is not of immediate interest but which must be accounted for in the analysis of those parameters which are of interest”. In many signal processing applications it is the subspace that is of primary estimation interest, rather than the precise set of basis vectors that parameterize the model. In DOA estimation problems, for example, the subspace is sufficient to unambiguously determine the DOA parameters. In the (α, β, λ) normal coordinates, the parameter β is the subspace parameter; the rotation parameter α and eigenvalues λ would be viewed as nuisance parameters in the DOA application.

3.6 CRB for Basis Manifold formulation of Complex Linear Model

Theorem 7 gave the FIM as the expected value of the Hessian, $\nabla^2 L$, of the likelihood function. The likelihood function of interest has as its domain the product manifold $\mathcal{B} \times \mathcal{D}$ and so the derivative computation requires a basis for $T_Y \mathcal{B}$ and $T_\Lambda \mathcal{D}$ at each point (Y, Λ) . In addition, for the Basis manifold formulation, the $T_Y \mathcal{B}$ tangent space is naturally partitioned as $T_Y \mathcal{B} = A_Y \oplus B_Y$ (357). Given a basis $\mathbf{a}_k, \mathbf{b}_k$ for A_Y and B_Y ,

respectively, the coefficient matrices of $\nabla^2 L$ with respect to this set are

$$[L_{\alpha\alpha}]_{ij} \equiv \frac{\partial^2 L}{\partial \alpha^i \partial \alpha^j} \equiv \nabla^2 L[\mathbf{a}_i, \mathbf{a}_j]; \quad i, j = 1 \cdots N_a, \quad (403)$$

$$[L_{\beta\beta}]_{ij} = \frac{\partial^2 L}{\partial \beta^i \partial \beta^j} \equiv \nabla^2 L[\mathbf{b}_i, \mathbf{b}_j]; \quad i, j = 1 \cdots N_b. \quad (404)$$

In addition, for $\mathbf{d}_k$ a basis for $T_\Lambda \mathcal{D}$ at an arbitrary point of $\mathcal{D}$

$$[L_{\lambda\lambda}]_{i,j} = \frac{\partial^2 L}{\partial \lambda^i \partial \lambda^j} \equiv \nabla^2 L[\mathbf{d}_i, \mathbf{d}_j]; \quad k = 1 \cdots p. \quad (405)$$

All derivatives are evaluated at the base point (Y, Λ) with normal coordinates $(0, 0, \lambda)$. The approach outlined in section 3.4 is used to compute the full FIM components. The required ancillary calculations are provided in the following two lemmas and the final FIM submatrices are given in the follow-on theorem. Since the parameters of interest, Y and Λ , make their appearance in the log likelihood (342) through the covariance matrix R (347) we first record the expansion of L in terms of R .

Lemma 8 (Log Likelihood Expansion as a function of the model Covariance R). *The given log likelihood function (342) viewed as a function of R can be expressed as*

$$L(R + \delta R) = L(R) + \delta L(\delta R, R) \quad (406)$$

where the first order and higher order Taylor series terms are

$$\delta L(\delta R, R) \equiv -K \left(\text{tr}(\delta R R^{-1})(I - \tilde{R} R^{-1}) + \text{tr}(\delta R R^{-1})^2 (\tilde{R} R^{-1} - \frac{1}{2} I) + \cdots \right). \quad (407)$$

The expectation of δL is

$$E[\delta L] = -\frac{K}{2} \left(\text{tr}(\delta R R^{-1})^2 + \cdots \right). \quad (408)$$

Proof: (see appendix 8)

The following lemma records the value of $\mathbf{F}[\Delta, \Delta]$ for an arbitrary tangent vector choice in $T_Y \mathcal{B}$.

Lemma 9. *Let $Y(t)$ be the geodesic emanating from $Y_0 = Y(0)$ in the direction $\Delta = Y_0 A + Y_{0\perp} B$. The term δR in the expansion (406) is then given by*

$$\delta R = t\delta R_1(A, B) + o(t^2) \quad (409)$$

where

$$\delta R_1(A, B) = Y_0 A \Lambda Y_0^H + Y_{0\perp} B \Lambda Y_0^H + Y_0 \Lambda B^H Y_{0\perp}^H. \quad (410)$$

The Hessian $\mathbf{F}[\Delta, \Delta]$ is

$$\mathbf{F}[\Delta, \Delta] = K \cdot \text{tr}(\delta R_1(A, B) R^{-1})^2. \quad (411)$$

Proof: (see appendix 9)

The component submatrices $F_{\alpha\alpha}$, $F_{\beta\beta}$ and $F_{\alpha\beta}$ of the full FIM are computed from (411) and (410) using the appropriate coordinate basis vectors. For the coordinate basis vector $\Delta = \mathbf{a}_k = Y A_k$ we have

$$[F_{\alpha\alpha}]_{kk} \equiv \mathbf{F}[\mathbf{a}_k, \mathbf{a}_k] = K \cdot \text{tr}(\delta R_1(A_k, 0) R^{-1})^2; \quad k = 1 \cdots N_a. \quad (412)$$

Similarly for $\Delta = \mathbf{b}_k = Y_{\perp} B_k$

$$[F_{\beta\beta}]_{kk} \equiv \mathbf{F}[\mathbf{b}_k, \mathbf{b}_k] = K \cdot \text{tr}(\delta R_1(0, B_k) R^{-1})^2; \quad k = 1 \cdots N_b. \quad (413)$$

The submatrix of cross terms $F_{\alpha\beta}$ and the off diagonal terms of $F_{\beta\beta}$ and $F_{\alpha\alpha}$ are computed via polarization.

Theorem 10 (Full FIM for the Complex Linear Model (337)). *The full Fisher Information matrix for the log-likelihood function given by (342) in the (α, β, λ) normal coordinates (defined by the basis $\mathbf{a}_k$, $\mathbf{b}_k$ and $\mathbf{d}_k$) at the arbitrary point (Y, Λ) (normal coordinates $(0, 0, \lambda)$) has the block matrix form*

$$F = \begin{bmatrix} F_{\alpha\alpha} & F_{\alpha\beta} & F_{\alpha\lambda} \\ F_{\beta\alpha} & F_{\beta\beta} & F_{\beta\lambda} \\ F_{\lambda\alpha} & F_{\lambda\beta} & F_{\lambda\lambda} \end{bmatrix}.$$

All the off-diagonal submatrices vanish

$$F_{\alpha\beta} = 0, F_{\alpha\lambda} = 0, F_{\beta\lambda} = 0 \quad (414)$$

and the square matrices $F_{\alpha\alpha}$, $F_{\beta\beta}$ and $F_{\lambda\lambda}$ are defined below with respect to an arbitrary choice of basis for the tangent space. The basis set is $\mathbf{a}_k = Y A_k$, $\mathbf{b}_k = Y_{\perp} B_k$ for A_Y and for B_Y respectively. The matrices A_k, B_k are chosen so as to define an orthonormal basis. In the following Λ_s and Λ are as given in (347) and (343), respectively and K is the number of independent snapshots.

$F_{\alpha\alpha}$ **Block**

$$[F_{\alpha\alpha}]_{kk} = K \cdot \text{tr}(A_k A_k - \Lambda A_k \Lambda_s^{-1} A_k - \sigma^2 A_k A_k \Lambda_s^{-1}); \quad k = 1 \cdots N_a \quad (415)$$

$$[F_{\alpha\alpha}]_{i,j} = 0; i \neq j \quad (416)$$

$F_{\beta\beta}$ **Block** [2, equation 143])

$$[F_{\beta\beta}]_{ij} = \frac{2K}{\sigma^2} \text{Re} \left(\text{tr}(\Lambda_{OPT} B_i^H B_j) \right); \quad i, j = 1 \cdots N_b \quad (417)$$

where

$$\Lambda_{OPT} = \Lambda^2 \Lambda_s^{-1} = (\Lambda_s - \sigma^2 I)^2 \Lambda_s^{-1} \quad (418)$$

or

$$\Lambda_{OPT} = \Lambda^2 (\Lambda + \sigma^2 I)^{-1} \quad (419)$$

This result is expressed alternatively in matrix form as

$$F_{\beta\beta} = \frac{2K}{\sigma^2} \text{Re}(E_b^H \Lambda_{bb} E_b) \quad (420)$$

where Λ_{bb} is the $\bar{N}_b \times \bar{N}_b$ real diagonal matrix

$$\Lambda_{bb} \equiv (\Lambda_{OPT} \otimes I_{N-p}) \quad (421)$$

and where E_b is a $\bar{N}_b \times N_b$ complex matrix defined in terms of the basis matrices B_k as

$$E_b = \begin{bmatrix} \text{vec}(B_1) \cdots \text{vec}(B_{N_b}) \end{bmatrix} \quad (422)$$

where $\bar{N}_b = \frac{N_b}{2}$ (recall $N_b = \dim(B_Y)$).

$F_{\lambda\lambda}$ **Block** [2, equation 80]

$$F_{\lambda\lambda} = K[\Lambda^2 \Lambda_s^{-2}] \quad (423)$$

$$[F_{\lambda\lambda}]_{k,k} = \frac{K\lambda_k^2}{(\lambda_s)_k^2} = \frac{K\lambda_k^2}{\lambda_k^2 + \sigma^2} = \frac{K}{1 + SNR_k^{-1}} \quad (424)$$

where $SNR_k = \frac{\lambda_k}{\sigma^2}$

Proof: (see appendix B.1)

Since the cross terms $F_{\beta\alpha}$ and $F_{\beta\lambda}$ vanish there is no estimation loss on the subspace in this example.

Remark: The standard basis for B_Y is defined by the set $\mathbf{b}_k = Y_\perp B_k$ where the matrix B_k has 1 or i in the k th location and zeros elsewhere (see section 2.2.4). For this case E_b in (422) is

$$E_b = \begin{bmatrix} I_{\bar{N}_b} & iI_{\bar{N}_b} \end{bmatrix} . \quad (425)$$

Substituting this into (420) yields

$$F_{\beta\beta} = \frac{2K}{\sigma^2} \begin{bmatrix} (\Lambda_{OPT} \otimes I_{N-p}) & 0 \\ 0 & (\Lambda_{OPT} \otimes I_{N-p}) \end{bmatrix} = \frac{2K}{\sigma^2} (I_2 \otimes \Lambda_{bb}) . \quad (426)$$

The inverse of $F_{\beta\beta}$ (426) is

$$F_{\beta\beta}^{-1} = \frac{\sigma^2}{2K} \begin{bmatrix} (\Lambda_{OPT}^{-1} \otimes I_{N-p}) & 0 \\ 0 & (\Lambda_{OPT}^{-1} \otimes I_{N-p}) \end{bmatrix} . \quad (427)$$

3.6.1 Intrinsic CRB Subspace Distance

Ignoring curvature terms the covariance matrix $C_{\beta\beta}$ (363) and FIM submatrix $F_{\beta\beta}$ are related by the CRB matrix inequality

$$C_{\beta\beta} \succeq F_{\beta\beta}^{-1}. \quad (428)$$

As consequence $\text{tr}C_{\beta\beta} > \text{tr}F_{\beta\beta}^{-1}$. Note that both sides of the inequality are independent of the orthonormal basis selected for B_Y .

The RMS subspace distance between the true subspace represented by the matrix Y and any unbiased estimate $\hat{Y}$ was defined independently of the geometric formulation as

$$\epsilon_{ss} \equiv E[\text{dist}SS^2(Y, \hat{Y})]^{1/2}. \quad (429)$$

Following the discussion in section 382 the intrinsic CRB on the subspace estimation accuracy is given by the formula

$$\epsilon_{ss} \geq \sqrt{\text{tr}F_{\beta\beta}^{-1}}. \quad (430)$$

Using (427) $\text{tr}F_{\beta\beta}^{-1}$ evaluates to

$$\text{tr}F_{\beta\beta}^{-1} = \frac{\sigma^2}{K} \text{tr}(\Lambda_{OPT}^{-1} \otimes I_{N-p}) = (N-p) \frac{\sigma^2}{K} \text{tr}(\Lambda_{OPT}^{-1}) \quad (431)$$

$$= \frac{(N-p)\sigma^2}{K} \sum_{k=1}^p (\lambda_{OPT})_k^{-1} = \frac{(N-p)\sigma^2}{K} \sum_{k=1}^p \frac{\lambda_k + \sigma^2}{\lambda_k^2} \quad (432)$$

where we have used (419) $(\lambda_{OPT})_k \equiv \frac{\lambda_k^2}{\lambda_k + \sigma^2}$.

Remark 1: For high SNR $\lambda_k \gg \sigma^2$ and (432) simplifies to

$$\text{tr}F_{\beta\beta}^{-1} \approx \frac{(N-p)\sigma^2}{K} \sum_{k=1}^p \lambda_k^{-1} = \frac{(N-p)}{K} \sum_{k=1}^p \text{SNR}_k^{-1} \quad (433)$$

where $\text{SNR}_k \equiv \frac{\lambda_k}{\sigma^2}$

Remark 2: Equi-Power Sources:

If $\lambda_k \equiv \lambda$ for all $k = 1 \cdots p$, we have $SNR = \frac{\lambda}{\sigma^2}$ and (432) becomes

$$\text{tr} F_{\beta\beta}^{-1} = \frac{p(N-p)(1+SNR^{-1})}{K \cdot SNR} = \frac{p(N-p)(SNR+1)}{K \cdot SNR^2} \quad (434)$$

at high SNR (434) and (433) further simplifies to

$$\text{tr} F_{\beta\beta}^{-1} \approx \frac{p(N-p)}{K \cdot SNR}. \quad (435)$$

Recognizing the numerator, $N_b = 2p(N-p)$, as the dimension of the tangent space B_Y this high SNR equi-powered case can be expressed as

$$\text{tr} F_{\beta\beta}^{-1} \approx \frac{\dim(B_Y)}{2K \cdot SNR}. \quad (436)$$

Remark 3: Note that neither the basis represented by the matrix Y nor its tangent space represented by the matrix E_b appear in the bound (431).

3.6.2 Intrinsic CRB on Rotational Distance

The RMS rotational distance between the true subspace represented by the matrix Y and any unbiased estimate $\hat{Y}$ was defined as

$$\epsilon_{RR} \equiv E[\text{dist} R^2(Y, \hat{Y})]^{1/2}. \quad (437)$$

Following the discussion in section 382 the intrinsic CRB on the rotational estimation accuracy is given by the formula

$$\epsilon_{RR} \geq \sqrt{\text{tr} F_{\alpha\alpha}^{-1}}. \quad (438)$$

Corrollary 11 (Simplified form of $F_{\alpha\alpha}$ and $\text{tr} F_{\alpha\alpha}^{-1}$). *For the standard basis on A_Y , $\mathbf{a}_k = Y A_k$; $k = 1 \cdots N_a$ the FIM subblock $F_{\alpha\alpha}$ (415) is given by the $N_a \times N_a$ real matrix*

$$F_{\alpha\alpha} = K \begin{bmatrix} \gamma_1 I_2 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \gamma_{\bar{N}_a} I_2 \end{bmatrix} \quad (439)$$

or equivalently

$$F_{\alpha\alpha} = K(\text{diag}(\gamma) \otimes I_2) \quad (440)$$

where the elements of the $\bar{N}_a$ -tuple γ are given by

$$\gamma_{(i,j)} = \frac{\lambda_i(\lambda_i + \sigma^2) - 2(\lambda_i + \sigma^2)(\lambda_j + \sigma^2) + \lambda_j(\lambda_j + \sigma^2)}{(\lambda_i + \sigma^2)(\lambda_j + \sigma^2)}; \quad j > i \quad (441)$$

Proof: Appendix B.1.3

At high SNR, $\lambda_j \gg \sigma^2; j = 1 \cdots \bar{N}_a$ we have

$$\frac{1}{\gamma_{(i,j)}} = \frac{\lambda_i \lambda_j}{(\lambda_i - \lambda_j)^2}; \quad j > i \quad (442)$$

so that

$$\text{tr} F_{\alpha\alpha}^{-1} \approx 2K \sum_{i=1}^p \sum_{j=2; j>i}^p \frac{\lambda_i \lambda_j}{(\lambda_i - \lambda_j)^2} \quad (443)$$

or equivalently

$$\text{tr} F_{\alpha\alpha}^{-1} \approx K \sum_{i=1}^p \sum_{j=1; i \neq j}^p \frac{\lambda_i \lambda_j}{(\lambda_i - \lambda_j)^2}. \quad (444)$$

Example: For $p = 3, \bar{N}_a = \frac{4(4-1)}{2} = 6$ and $\lambda_j \gg \sigma^2$ the 6-tuple

$$\gamma = \left(\gamma_{(1,2)}, \gamma_{(1,3)}, \gamma_{(1,4)}, \gamma_{(2,3)}, \gamma_{(2,4)}, \gamma_{(3,4)} \right). \quad (445)$$

3.7 EigenVector CRB

Using the mathematical framework established above we develop a CRB, in the form of a matrix inequality, on estimation of the ordered columns of Y in the statistical model (337). Recall that Y is defined as the eigenvector vector matrix of the true signal covariance (343). The standard estimator of Y is the eigenvector matrix of the SCM and, as is well known (and shown in chapter 4), this is the ML estimator of Y for the given model.

For convenience and consistency with [5] in this section we denote the columns of the $N \times p$ semi-unitary matrix Y_0 as

$$Y_0 = \begin{bmatrix} s_1 & \cdots & s_p \end{bmatrix} \quad (446)$$

and those of the orthogonal space as

$$Y_{0\perp} = \begin{bmatrix} g_1 & \cdots & g_{N-p} \end{bmatrix}. \quad (447)$$

Similarly we denote columns of the estimate $\hat{Y}$ as $\hat{Y} = \begin{bmatrix} \hat{s}_1 & \cdots & \hat{s}_p \end{bmatrix}$. The classical error covariance matrix result [5] when the estimator $\hat{Y}$ is the EVD (p -dominant part) of the SCM, is stated below.

Lemma 12 (Covariance of Eigenvectors [6]). *The estimation errors $\hat{s}_i - s_i$ are asymptotically (for large K) jointly Gaussian distributed with zero means and covariance matrix given*

$$E[(\hat{s}_i - s_i)(\hat{s}_j - s_j)^H] = \frac{\lambda_i}{K} \left[\sum_{k=1; k \neq i}^p \frac{\lambda_k}{(\lambda_k - \lambda_i)^2} s_k s_k^H + \sum_{k=1}^{N-p} \frac{\sigma^2}{(\sigma^2 - \lambda_i)^2} g_k g_k^H \right] \delta_{ij}. \quad (448)$$

The CRB for this estimation covariance is developed as follows. In the geometric approach the estimation error is given by the tangent vector $\hat{\Delta} = \exp_{Y_0}^{-1} \hat{Y}$. Given a fixed basis $\mathbf{a}_k, \mathbf{b}_k$ for the partitioned tangent space this vector expands as (362)

$$\hat{\Delta} = \sum_{k=1}^{N_a} \alpha^k \mathbf{a}_k + \sum_{m=1}^{N_b} \beta^m \mathbf{b}_m. \quad (449)$$

The covariance $\text{cov}(\hat{\Delta})$ satisfies the matrix inequality

$$\begin{bmatrix} C_{\alpha\alpha} & C_{\beta\alpha} \\ C_{\alpha\beta} & C_{\beta\beta} \end{bmatrix} \succeq \begin{bmatrix} F_{\alpha\alpha} & 0 \\ 0 & F_{\beta\beta} \end{bmatrix}^{-1} = \begin{bmatrix} F_{\alpha\alpha}^{-1} & 0 \\ 0 & F_{\beta\beta}^{-1} \end{bmatrix} \quad (450)$$

where the FIM sub-block component matrices $F_{\alpha\alpha}$ and $F_{\beta\beta}$ are given by (439) and (426) (and $F_{\alpha\beta} = 0$). In this formulation the error in estimating the k -th column of Y is $\hat{\Delta} \cdot i_k$, where i_k is a unit vector with 1 at the k -th location. This representation follows since $\hat{\Delta} \approx \hat{Y} - Y$ and so $\hat{\Delta} \cdot i_k \approx \hat{s}_k - s_k$. We therefore seek a lower bound on error covariance $\text{cov}(\hat{\Delta} \cdot i_k)$. This bound is developed from (450) and is given in the following theorem.

Corollary 13 (CRB on Eigenvector Covariance). *For $\hat{Y}$ an unbiased estimator of $Y \in \mathcal{B}$ in the statistical model (337) the covariance $\text{cov}(\hat{\Delta} \cdot i_k)$, where $\hat{\Delta} = \exp_Y^{-1} \hat{Y}$, satisfies the matrix inequality*

$$\text{cov}(\hat{\Delta} \cdot i_k) \succeq \begin{bmatrix} Y T_k & Y_{\perp} M_k \end{bmatrix} \begin{bmatrix} F_{\alpha\alpha}^{-1} & 0 \\ 0 & F_{\beta\beta}^{-1} \end{bmatrix} \begin{bmatrix} Y T_k & Y_{\perp} M_k \end{bmatrix}^H \quad (451)$$

where T_k and M_k are selection matrices defined below (B.132) and (455)) and $F_{\alpha\alpha}$ and $F_{\beta\beta}$ are as defined in theorem 10. At high SNR the r.h.s. simplifies resulting in the inequality

$$\text{cov}(\hat{\Delta} \cdot i_i) \succeq \frac{\lambda_i}{K} \left[\sum_{k=1; k \neq i}^p \frac{\lambda_k}{(\lambda_k - \lambda_i)^2} s_k s_k^H + \sum_{k=1}^{N-p} \frac{\sigma^2}{(\sigma^2 - \lambda_i)^2} g_k g_k^H \right] \quad (452)$$

Proof: (see B.1.4, outline of proof follows) The k -th column of $\hat{\Delta}$ can be expressed as

$$\hat{\Delta} \cdot i_k = Y T_k \alpha + Y_{\perp} M_k \beta; \quad k = 1 \cdots p \quad (453)$$

where T_k ($\bar{N}_a \times N_a$) and M_k ($(N-p) \times N_b$) are selection matrices. For example for $p = 3$, the T_1 matrix is

$$T_1 = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ -1 & i & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & i & 0 & 0 \end{bmatrix} \quad (454)$$

and M_2 is $(N-p) \times N_b$

$$M_2 = \begin{bmatrix} X_2 & iX_2 \end{bmatrix} \quad (455)$$

where $X_2 = \begin{bmatrix} 0 & I_{N-p} & 0 \end{bmatrix}$.

The corresponding $N \times N$ covariance matrix is

$$\text{cov}(\Delta \cdot i_k) = \begin{bmatrix} Y T_k & Y_{\perp} M_k \end{bmatrix} \begin{bmatrix} C_{\alpha\alpha} & C_{\beta\alpha} \\ C_{\alpha\beta} & C_{\beta\beta} \end{bmatrix} \begin{bmatrix} Y T_k & Y_{\perp} M_k \end{bmatrix}^H. \quad (456)$$

Using the main result of theorem 10 expressed as (450) we have

$$\text{cov}(\Delta \cdot i_k) \succeq \begin{bmatrix} Y T_k & Y_{\perp} M_k \end{bmatrix} \begin{bmatrix} F_{\alpha\alpha}^{-1} & 0 \\ 0 & F_{\beta\beta}^{-1} \end{bmatrix} \begin{bmatrix} Y T_k & Y_{\perp} M_k \end{bmatrix}^H. \quad (457)$$

Substituting (439), (442) (439) for $F_{\alpha\alpha}^{-1}$ and using (446) and (447) yields the high SNR asymptotic result (452).

3.8 Constrained Basis Estimation

In many problems of interest the signal matrix H is constrained to some subset $\mathcal{M}$ of $C^{N \times p}$, that is $H \in \mathcal{M} \subset C^{N \times p}$. When this is the case the set of $N \times p$ semi-unitary matrices defined by the transformation (343), $[Y, \Lambda] = \text{eig}(H P H^H)$, will likewise be

constrained to a submanifold of the basis manifold $\mathcal{B}$. We assume this set to be a smooth Riemannian totally geodesic submanifold which we denote as $\bar{\mathcal{B}} \subset \mathcal{B}$. The likelihood function for the constrained problem, denoted $\bar{L}$, has the same functional form as the likelihood function for the unconstrained problem with a restriction on the functions domain, that is

$$\bar{L}(Y, \Lambda) \equiv L(Y, \Lambda); \quad Y \in \bar{\mathcal{B}}, \quad \Lambda \in D. \quad (458)$$

The eigendecomposition of the true covariance HPH^H defines a mapping between the (H, P) and (Y, Λ) parameterizations of likelihood function. Restricting the domain of the general case mapping (344) we have

$$eig : \mathcal{M} \times P_p \rightarrow \bar{\mathcal{B}} \times D. \quad (459)$$

The source correlation matrix $P \in P_p$ is not constrained and because of this the submanifold $\bar{\mathcal{B}}$ has the property that if $Y \in \bar{\mathcal{B}}$ then $YQ \in \bar{\mathcal{B}}$ for any unitary Q . This property is important in defining the tangent space for $\bar{\mathcal{B}}$ and is established in the following lemma.

Lemma 14. *The submanifold $\bar{\mathcal{B}}$ defined as the image of mapping (459) has the property that if $Y \in \bar{\mathcal{B}}$ then $YQ \in \bar{\mathcal{B}}$ for any $p \times p$ unitary matrix Q .*

Proof:

Let Y_0 denote the matrix of left singular vectors of $H = Y_0 D_0 V_0^H$ for arbitrary $H \in \mathcal{M}$. Let Q and Λ be defined by the eigendecomposition below

$$Q \cdot \Lambda \cdot Q^H := (D_0 V_0^H P V_0 D_0) \quad (460)$$

The signal covariance $R_s \equiv HPH^H$ therefore can be expressed as

$$H \cdot P \cdot H^H = Y_0 Q \cdot \Lambda \cdot Q^H Y_0^H. \quad (461)$$

The eigenvector matrix of the signal covariance is therefore $Y_0 Q$. Note that Y_0 depends only on H while Q depends on both H and P through (460). Since the correlation matrix $P \in P_p$ is not constrained the $q \times q$ unitary matrix Q is likewise unconstrained.

Since $\bar{\mathcal{B}} \subset \mathcal{B}$ the submanifold tangent space at any point $Y \in \bar{\mathcal{B}}$ is a vector subspace of full tangent space $T_Y \mathcal{B}$, that is $T_Y \bar{\mathcal{B}} < T_Y \mathcal{B}$, so that

$$T_Y \mathcal{B} = T_Y \bar{\mathcal{B}} \oplus N_Y \quad (462)$$

where $N_Y \equiv (T_Y \bar{\mathcal{B}})^\perp$.

As a result of the lemma the tangent space $T_Y \bar{\mathcal{B}}$ can be partitioned into a rotational space and its complement in the same way as $T_Y \mathcal{B}$. The rotational space of $T_Y \bar{\mathcal{B}}$ is identical to that of $T_Y \mathcal{B}$ and is again denoted A_Y . The orthogonal complement of A_Y in $T_Y \bar{\mathcal{B}}$ is denoted S_Y so that

$$T_Y \bar{\mathcal{B}} = S_Y \oplus A_Y . \quad (463)$$

Using this expression in (462) the full tangent space of $\mathcal{B}$ can be expressed as $T_Y \mathcal{B} = (S_Y \oplus N_Y) \oplus A_Y$. Comparing with (357) we see that the space of subspace perturbations, B_Y , is decomposed as

$$B_Y = S_Y \oplus N_Y . \quad (464)$$

Motion in the N_Y direction corresponds to subspace perturbations that leave the submanifold, while motions in the S_Y and A_Y directions remain on the constraint submanifold $\bar{\mathcal{B}}$. Motion in the A_Y corresponds to a rotation and motion in the S_Y direction corresponds to a subspace perturbation. In remainder of this chapter we denote dimension of S_Y as q , $\dim(S_Y) = q$ so that $\dim(N_Y) = N_b - q$.

Let $\mathbf{s}_k$ denote an orthonormal basis for S_Y and, as in the unconstrained case, let $\mathbf{a}_k$ denote an orthonormal basis for A_Y . Then with respect to this basis any tangent vector $\Delta \in T_Y \bar{\mathcal{B}}$ expands as

$$\Delta = \sum_{k=1}^{N_a} \mathbf{a}_k \alpha^k + \sum_{k=1}^q \mathbf{s}_k s^k \quad (465)$$

where s and α are the real components of the representation. The numerical form of the basis vectors for S_Y is

$$\mathbf{s}_k = Y_\perp S_k; \quad k = 1 \cdots q \quad (466)$$

where the S_k 's are appropriately chosen $(N - p) \times p$ complex matrices. The matrices that define the spanning basis for S_Y vary from point to point on the submanifold and are defined by the problem specifics (i.e., the submanifold definition).

3.8.1 Estimation on the Submanifold

Let $\hat{Y}$ be an estimator for $Y \in \bar{\mathcal{B}}$. The estimation error vector is defined by $\hat{\Delta} = \exp_Y^{-1} \hat{Y}$. This vector $\hat{\Delta} \in T_Y \bar{\mathcal{B}} = A_Y \oplus S_Y$ may be expressed with respect to the basis $\mathbf{a}_k, \mathbf{s}_k$ at Y as

$$\hat{\Delta} = \sum_{k=1}^{N_a} \mathbf{a}_k \hat{\alpha}^k + \sum_{k=1}^q \mathbf{s}_k \hat{s}^k. \quad (467)$$

In this expression the basis vectors are fixed and the coefficients are the stochastic error components $\hat{\alpha} = \hat{\alpha}(Z)$ and $\hat{s} = \hat{s}(Z)$. The error covariance, denoted C , is defined in terms of these components as

$$C \equiv \begin{bmatrix} C_{\alpha\alpha} & C_{s\alpha} \\ C_{\alpha s} & C_{ss} \end{bmatrix} \quad (468)$$

where

$$C_{ss} = E[\hat{s}\hat{s}^t], \quad C_{\alpha\alpha} = E[\hat{\alpha}\hat{\alpha}^t], \quad C_{s\alpha} = E[\hat{s}\hat{\alpha}^t] \quad (469)$$

(see (363)).

3.8.2 Computing the constrained case FIM 1

Given an orthonormal basis for the partitioned space (463) the computation of the FIM proceeds exactly as in the unconstrained case. Since the source of the constraint in this problem is the signal matrix H , while the source correlation matrix P is unconstrained, there are no additional constraints on the rotations or the eigenvalues. The associated submatrices of $\bar{F}$ are therefore identical to those of the unconstrained problem, that is $\bar{F}_{\alpha\alpha} = F_{\alpha\alpha}$, $\bar{F}_{\lambda\lambda} = F_{\lambda\lambda}$, as given in theorem 10. Also since $F_{\beta\alpha} = 0$ and $F_{\beta\lambda} = 0$ in the unconstrained case it follows that $F_{s\alpha} = 0$, $F_{s\lambda} = 0$. The full FIM for the constrained case is denoted $\bar{F}$ and has the general block form

$$\bar{F} = \begin{bmatrix} F_{\alpha\alpha} & 0 & 0 \\ 0 & F_{ss} & 0 \\ 0 & 0 & F_{\lambda\lambda} \end{bmatrix} \quad (470)$$

The $q \times q$ submatrix F_{ss} with respect to the basis $\mathbf{s}$ (466) on S_Y is

$$F_{ss} = \frac{2K}{\sigma^2} \text{Re}(E_s^H \Lambda_{bb} E_s) \quad (471)$$

where E_s is the $\bar{N}_b \times q$ complex matrix defined by

$$E_s \equiv \begin{bmatrix} \text{vec}(S_1) & \cdots & \text{vec}(S_q) \end{bmatrix} . \quad (472)$$

In the parametrically defined submanifolds discussed in a following section, the parameterization defines $\bar{\mathcal{B}}$ and are the natural coordinates on $\bar{\mathcal{B}}$. The tangents to these coordinate curves naturally define a basis for the tangent spaces S_Y . A basis defined in this way is termed a natural coordinate basis.

Remark: The approach here uses an arbitrary but explicit basis (472) for the submanifold tangent space. An intrinsic CRB on the geodesic distance can be alternately computed using a projection onto the submanifold tangent space $P_S : B_Y \rightarrow S_Y$ [1, Theorem 2]. In this situation

$$\text{tr}(C_{ss}) \geq \text{tr}((P_S F_{\beta\beta} P_S)^+) . \quad (473)$$

Of course, if an orthogonal basis for S_Y is available defined by the transformation T_s then

$$C_{\beta\beta} = T_s C_{ss} T_s^t \quad (474)$$

$$\text{tr}(C_{\beta\beta}) = \text{tr}(T_s C_{ss} T_s^t) = \text{tr}(C_{ss} T_s^t T_s) = \text{tr}(C_{ss}) . \quad (475)$$

Note that in this expression $C_{\beta\beta}$ is the submanifold error as measured in the ambient frame. See section A.12 for additional details on transformations.

3.8.3 Constrained CRB

Ignoring curvature terms, as in the unconstrained case, the lower bound on the subspace estimation accuracy is the given by the formula

$$\epsilon_{ss} \geq \sqrt{\text{tr} F_{ss}^{-1}}. \quad (476)$$

Example: Processing Gain due to Reduced Dimension The spaces B_Y and $S_Y \subset B_Y$ represent the subspace perturbation tangent spaces corresponding to $\mathcal{B}$ and $\bar{\mathcal{B}} \subset \mathcal{B}$ respectively. In the array processing estimation problems, the dimension of S_Y is $\dim(S_Y) = q = rMp$, where M is physical array dimension (i.e. is $M = 1$ for ULA, $M = 2$ for URA), p is the number of modes and $r = 1$ for harmonic signals (undamped exponentials) and $r = 2$ for damped exponentials. Denoting the S_Y space for the damped and harmonic signal assumptions as $S_{Y;damped}$ and $S_{Y;harmonic}$ respectively, and recalling that $\dim(B_Y) = N_b = 2p(N - p)$, we have the dimension ratios

$$\frac{\dim(B_Y)}{\dim(S_{Y;harmonic})} = \frac{N_b}{q} = \frac{2(N - p)}{M} \quad (477)$$

and

$$\frac{\dim(B_Y)}{\dim(S_{Y;damped})} = \frac{(N - p)}{M}. \quad (478)$$

The dimension ratio between the two different constraint assumption is

$$\frac{\dim(S_{Y;damped})}{\dim(S_{Y;harmonic})} = \frac{2Mp}{Mp} = 2. \quad (479)$$

The ratio of the unconstrained CRB to the constrained CRB gives the processing gain realized by an optimal subspace estimator on $\bar{\mathcal{B}} \subset \mathcal{B}$ relative to one that is optimal subspace estimator on $\mathcal{B}$, that is

$$Gain = \frac{\text{tr} F_{\beta\beta}^{-1}}{\text{tr} F_{ss}^{-1}}. \quad (480)$$

For benign signal conditions (e.g., widely spaced uncorrelated sources in DOA estimation problems) this ratio is found to be equal to the dimension ratio given in (478), that

is

$$Gain = \frac{\text{tr} F_{\beta\beta}^{-1}}{\text{tr} F_{ss}^{-1}} \approx \frac{2(N-p)}{rM} . \quad (481)$$

As is shown later, an estimator that is optimal for damped exponential modal model ($r = 2$) will suffer 3 dB loss when applied to the harmonic signal model problem ($r = 1$) consistent with (479).

3.9 Parametrically Constrained Problem

The computation of F_{ss} (471) requires a basis for the submanifold tangent space S_Y . In the important special case when the constraint defining the submanifold is due to some underlying parameterization of the signal matrix $H = H(\theta)$, $\theta \in R^p$, this basis can be readily computed. The signal function $H(\theta)$ together with the eigendecomposition of $H(\theta)PH(\theta)^H$ defines a mapping between the parameters (θ, P) and (Y, Λ) , given explicitly by

$$[Y, \Lambda] = \text{eig}(H(\theta) \cdot P \cdot H^H(\theta)) . \quad (482)$$

As in the general constrained case if $Y \in \bar{B}$, then $YQ \in \bar{B}$. The submanifold $\bar{B}$ is then the collection of all $N \times p$ semi-unitary matrices of the form $Y(\theta)Q$ where $Y(\theta)$ is left singular vector matrix of $H(\theta)$ for some $\theta \in R^p$ and Q is arbitrary $p \times p$ unitary.

3.9.1 Defining the Submanifold Tangent Space

Let $\theta_0 = \begin{bmatrix} \theta_0^1 & \cdots & \theta_0^q \end{bmatrix}$ be arbitrary and let Y_0 be the corresponding eigenvector matrix of $H(\theta_0)PH(\theta_0)^H$. Denote the k th coordinate curve in parameter space as

$$\theta_{(k)}(t) = \theta_0 + t \cdot e_k \quad (483)$$

where e_k is a $q \times 1$ unit vector with 1 at the k -th location. The corresponding representation of this curve in the extrinsic Y coordinates on $\bar{B}$ with origin at Y_0 is given by $Y_{(k)}(t) \equiv Y(\theta_{(k)}(t))$ where the function Y is defined by (482). This curve is in general *not* a geodesic. The tangent vector to $Y_{(k)}(t)$ at Y_0 is denoted $\Delta_{(k)} \in T_Y \bar{B}$ and is

defined by

$$\Delta_{(k)} = \frac{dY_{(k)}(t)}{dt} \Big|_{t=0} . \quad (484)$$

The component of this vector corresponding to a subspace change is

$$\mathbf{s}'_k \equiv P_B \Delta_{(k)} = P_B \frac{dY_{(k)}(t)}{dt} \Big|_{t=0} \quad (485)$$

where P_B is the orthogonal projector onto B_Y . The set of q vectors generated from the q coordinate curves in this way defines S_Y (and therefore is a basis for S_Y). Since this basis is derived from a set of natural coordinate curves it is called a natural coordinate basis. In general this is not an orthonormal basis. An arbitrary vector $\Delta \in S_Y$ expands in this basis as

$$\Delta = \sum_{k=1}^p \mathbf{s}'_k \delta\theta^k . \quad (486)$$

We have used the suggestive notation $\delta\theta^k$ to denote the components of Δ with respect to this natural coordinate basis.

3.9.2 Computing the Natural Coordinate Basis

To calculate the derivative (484) $\frac{dY_{(k)}(t)}{dt}$ recall that the matrix $Y(\theta)$ is related to the signal matrix $H(\theta)$ as $Y(\theta) = H(\theta) \cdot T(\theta)$ where $T(\theta)$ is a non-singular $p \times p$ matrix ($T(\theta) = H(\theta)^\# Y(\theta)$). Suppressing the index k , the curve $Y(\theta_{(k)}(t))$ can therefore be written as

$$Y(\theta(t)) = H(\theta(t)) \cdot T(\theta(t)) . \quad (487)$$

The tangent vector to this curve at $Y(\theta_0)$, $\theta_0 = \theta(0)$ is

$$\Delta = \frac{dH(\theta(t))}{dt} \cdot T(\theta(t)) + H(\theta(t)) \cdot \frac{dT(\theta(t))}{dt} \quad (488)$$

where the derivatives are evaluated at $t = 0$.

A change in the parameter θ results in motion in both the S_Y (change in subspace) and A_Y (pure rotation) directions. The component in S_Y is

$$P_B \Delta = P_B \frac{dH(\theta(t))}{dt} \cdot T(\theta(t)) \quad (489)$$

which follows since $P_B H = 0$ (274). If $\theta(t)$ is a coordinate curve (483) then equation (489) defines the k th natural coordinate basis vector in S_Y at Y . The set of vectors generated in this way yields a basis for S_Y which have the general form

$$\mathbf{s}'_k = Y_{\perp} \dot{S}_k \quad k = 1 \cdots q \quad (490)$$

where the $(N - p) \times p$ matrices $\dot{S}_k$ are defined using (489) with $P_B = Y_{\perp} Y_{\perp}^H$ as

$$\dot{S}_k \equiv Y_{\perp}^H \frac{dH(\theta_{(k)}(t))}{dt} \Big|_{t=0} \cdot T(\theta_{(k)}(0)) . \quad (491)$$

Remark A straight line in R^p , the θ parameter space, emanating from the fixed point θ_0 in the direction of the vector $\delta\theta$ is given by the following equivalent forms

$$\theta(t) = \theta_0 + t\delta\theta \quad (492)$$

$$\theta(t) = [\theta_0^1 + t\delta\theta^1, \dots, \theta_0^q + t\delta\theta^q] . \quad (493)$$

The value of $\theta(t)$ at $t=1$ is $\theta_0 + \delta\theta$ and $\delta\theta = [\delta\theta^1 \quad \dots \quad \delta\theta^q]$. The tangent to the corresponding curve $Y(\theta(t))$ is

$$\frac{dY(\theta(t))}{dt} = \sum_{k=1}^p \frac{\partial Y}{\partial \theta^k} \delta\theta^k \quad (494)$$

where we have used

$$\delta\theta = \frac{d\theta(t)}{dt} \Big|_{t=0} . \quad (495)$$

Recalling from elementary calculus (see section A.2.4) that the partial derivative of $Y(\theta)$ with respect to θ^k is defined as

$$\frac{\partial Y}{\partial \theta^k} \equiv \frac{dY(\theta_{(k)}(t))}{dt} \quad (496)$$

where $\theta_{(k)}(t)$ is the coordinate curve (483), an alternate expression for the coordinate basis vectors (485) is given by

$$\mathbf{s}'_k = P_B \frac{\partial Y}{\partial \theta^k} . \quad (497)$$

Let $Y_0 \in \bar{\mathcal{B}}$ be the matrix corresponding to $\theta = \theta_0$, let $Y_1 \in \bar{\mathcal{B}}$ be an arbitrary point near Y_0 , and let the difference vector be $\Delta = \exp_{Y_0}^{-1} Y_1$. The first order approximation of $\exp_{Y_0}(\Delta)$ is

$$\exp_{Y_0}(\Delta) \approx Y_0 + \Delta = Y_0 + P_B \Delta + P_A \Delta . \quad (498)$$

The term $P_B \Delta$, which represents the change in the subspace, is an element of S_{Y_0} and expands as

$$P_B \Delta = \sum_{k=1}^p \acute{s}_k \acute{s}^k . \quad (499)$$

The components $\acute{s}^k$ are found using a Gram-Schmidt procedure (with respect to the inner product) and to first order $\acute{s}^k \approx \delta\theta^k$. The value of the parameter θ corresponding to Y_1 is

$$\theta_1 \approx \theta_0 + \delta\theta . \quad (500)$$

Note that for $Y_1 = Y_0 Q$, that is Y_1 is a pure rotation of Y_0 , the tangent vector $\Delta \equiv \exp_{Y_0}^{-1} Y_1$ is such that $P_B \Delta = 0$ so that $\delta\theta = 0$.

3.9.3 Generate an Orthonormal Basis on S_Y from the natural coordinate basis

The computation of F_{ss} is facilitated by using an orthonormal basis while the natural coordinate basis $\acute{s}_k$ developed above is useful when analyzing the underlying parameterization. An orthonormal basis may be formed from an appropriate linear combination of natural coordinate basis vectors $\acute{s}_k$ in the usual way. Let $\mathbf{s}_k$ be defined by

$$\mathbf{s}_k = \sum_{m=1}^q \mathbf{J}_k^m \acute{s}_m, \quad k = 1 \cdots q \quad (501)$$

or in formal matrix form as

$$\begin{bmatrix} \mathbf{s}_1 & \cdots & \mathbf{s}_q \end{bmatrix} = \begin{bmatrix} \acute{s}_1 & \cdots & \acute{s}_q \end{bmatrix} \mathbf{J} \quad (502)$$

where $\mathbf{J}$ is a $q \times q$ real matrix.

An arbitrary element $\Delta \in S_Y$ may be represented with respect to these two sets of basis vectors as

$$\Delta = \sum_{k=1}^q \mathbf{s}_k s^k = \sum_{k=1}^q \acute{\mathbf{s}}_k \acute{s}^k \quad (503)$$

or in formal matrix form as

$$\Delta = \begin{bmatrix} \mathbf{s}_1 & \cdots & \mathbf{s}_q \end{bmatrix} s = \begin{bmatrix} \acute{\mathbf{s}}_1 & \cdots & \acute{\mathbf{s}}_q \end{bmatrix} \acute{s} . \quad (504)$$

Noting (502) the components transform as

$$\acute{s} = \mathbf{J} s \quad (505)$$

or equivalently $s = \mathbf{J}^{-1} \acute{s}$.

The transformation matrix $\mathbf{J}$ (Jacobian matrix) may be selected so that $\mathbf{s}_k$ is an orthonormal basis. The real orthogonalizing matrix $\mathbf{J}$ is computed from the basis representations in the usual vector space way. Let the numerical representation of the basis vectors $\mathbf{s}_k$ and $\acute{\mathbf{s}}_k$ be expressed as

$$\mathbf{s}_k = Y_{\perp} S_k; \quad k = 1 \cdots q \quad (506)$$

$$\acute{\mathbf{s}}_k = Y_{\perp} \acute{S}_k; \quad k = 1 \cdots q \quad (507)$$

where $\acute{S}_k$ is given and S_k is to be determined. The transformation (501) may then be expressed as

$$S_k = \sum_{m=1}^q \mathbf{J}_m^k \acute{S}_m; \quad k = 1 \cdots q \quad (508)$$

or alternatively in matrix form as

$$E_s = E_{\acute{s}} \mathbf{J} \quad (509)$$

where E_s and $E_{\acute{s}}$ are formed by vectorizing the set of matrices S_k and $\acute{S}_m$, that is

$$E_{\acute{s}} = \begin{bmatrix} \text{vec}(\acute{S}_1) & \cdots & \text{vec}(\acute{S}_q) \end{bmatrix} . \quad (510)$$

The real $q \times q$ orthogonalizing matrix $\mathbf{J}$ is then (see section A.10)

$$\mathbf{J} = \text{Re}(E_{\acute{s}}^H E_s)^{-\frac{1}{2}} . \quad (511)$$

Using this transformation, estimates and bounds computed with respect to one coordinate frame may be transformed to the alternate frame, which may have a more meaningful interpretation. Since F_{ss} (471) is a tensor quantity it obeys the tensor transformation rules, that is

$$F_{\acute{s}\acute{s}} = \mathbf{J}^{-1} \cdot F_{\acute{s}\acute{s}} \cdot \mathbf{J}^{-t} \quad (512)$$

or equivalently

$$F_{ss} = \mathbf{J} \cdot F_{\acute{s}\acute{s}} \cdot \mathbf{J}^t . \quad (513)$$

As noted in [2] “because the Jacobian determines how coordinate transformations effect accuracy bounds, it is sometimes called a sensitivity matrix.” This effect on error measures is explored in section 5.2.4.

3.9.4 Modal Forms

The above expressions (491) for the basis vectors on S_Y take a simple form when the signal matrix $H(\theta)$ has a modal form. We first develop the 1-parameter per mode case ($r = 1, M = 1$ so that $q \equiv r \cdot M \cdot p = p$),

$$H(\theta) = \begin{bmatrix} h(\theta^1) & \cdots & h(\theta^p) \end{bmatrix} \quad (514)$$

before addressing the general case. For convenience we define the $N \times p$ matrix of derivatives as

$$D(\theta) = \begin{bmatrix} d_1 & \cdots & d_p \end{bmatrix} \quad (515)$$

where $d_k = d(\theta^k) \equiv \frac{dh(\theta)}{d\theta} \big|_{\theta=\theta^k}$.

The k -th column of the derivative matrix $D(\theta)$ corresponds to $\frac{dH(\theta_{(k)}(t))}{dt}$. Noting (483) and the modal form (514) this derivative is

$$\frac{dH(\theta_{(k)}(t))}{dt} = [0, \dots, 0, d_k, 0, \dots, 0] = D(\theta) \cdot J_{kk} . \quad (516)$$

In this expression J_{kk} is the $p \times p$ matrix selection matrix defined by

$$J_{kk} = (e_k \otimes e_k^t) \quad (517)$$

where e_k is a $p \times 1$ unit vector with 1 at the k -th location. J_{kk} thus has 1 at the k -th diagonal location and zeros elsewhere.

Substituting (516) into (491) yields the modal form for the $\dot{S}_k$ matrix, appearing in the basis vector $\dot{s}_k = Y_\perp \dot{S}_k$, as

$$\dot{S}_k = Y_\perp^H D(\theta) J_{kk} \cdot T(\theta) . \quad (518)$$

Suppressing the arguments, using the matrix identity $\text{vec}(ABC) = (C^t \otimes A)\text{vec}(B)$ and noting (510), yields

$$E_{\dot{s}}(:, k) = \text{vec}(\dot{S}_k) = (T^t \otimes Y_\perp^H \cdot D) j_{kk} \quad (519)$$

where $j_{kk} = \text{vec}(J_{kk})$. The complete $E_{\dot{s}}$ matrix (510) may then be written as

$$E_{\dot{s}} = (T^t \otimes Y_\perp^H \cdot D) G \quad (520)$$

where G is a $p^2 \times p$ real matrix whose columns are the vectorized $p \times p$ selection matrices J_{kk} , that is

$$G \equiv \begin{bmatrix} j_{11} & \cdots & j_{pp} \end{bmatrix} \equiv \begin{bmatrix} \text{vec}(J_{11}) & \cdots & \text{vec}(J_{pp}) \end{bmatrix} . \quad (521)$$

The factor $(T^t \otimes Y_\perp^H \cdot D)$ is $\bar{N}_b \times p^2$ so that $E_{\dot{s}}$ is $\bar{N}_b \times p$, where $\bar{N}_b = p(N - p)$.

Example: For $p = 2$ the individual selection matrices are

$$J_{11} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \quad J_{22} = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \quad (522)$$

so that

$$G = \begin{bmatrix} \text{vec}(J_{11}) & \text{vec}(J_{22}) \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 1 \end{bmatrix} . \quad (523)$$

Lemma 15 (Orthogonalizing Matrix **J**, for 1-Dimensional case, $q = p$). *For the modal case the $p \times p$ orthogonalizing matrix **J** (511) is*

$$\mathbf{J} = \text{Re}((D^H P_\perp D) \odot (H^H H)^{-1})^{-\frac{1}{2}} \quad (524)$$

where $P_{\perp} \equiv Y_{\perp} Y_{\perp}^H$

Proof: (see B.1.5)

To bring the F_{ss} matrix into the desired form we use the following identity which relates the diagonal matrix Λ_{OPT} (418) appearing in general FIM definition (471) to the matrix $P_s \equiv (P \cdot H^H \cdot R^{-1} \cdot H \cdot P)$ which appears later in the parameter space FIM definition (533) (see also [6]).

Lemma 16. *Equivalence of Forms [5, eqn. 75]*

Let $Y \Lambda_s Y^H = H P H^H$ (340) and let be T be the $p \times p$ non-singular matrix defined by

$$Y = HT. \quad (525)$$

Define the $p \times p$ matrix

$$P_s \equiv (P \cdot H^H \cdot R^{-1} \cdot H \cdot P) \quad (526)$$

where $R = H P H^H + \sigma^2 I$. Then

$$P_s = T \Lambda_{OPT} T^H \quad (527)$$

where Λ_{OPT} is given by (419).

Proof: (Appendix B.1.7)

The FIM subspace block matrix F_{ss} for the modal case is found using (527) and (520) in the general case definition, (471), and is recorded in the following theorem. The proof is a straightforward algebraic manipulation that makes use of (509), (520), and (511).

Theorem 17 (FIM subspace block matrix F_{ss} Modal Case). *Let $E_s = E_s \mathbf{J}$ (509) where E_s is defined by (520). Then the F_{ss} matrix (471) with respect to this basis is*

$$F_{ss} = \frac{2K}{\sigma^2} \cdot \mathbf{J} \text{Re} \left((D^H P_{\perp} D) \odot P_s \right) \mathbf{J}^t \quad (528)$$

where $\mathbf{J}$ and P_s are defined by (524) and (526) respectively.

Proof:(Appendix B.1.8)

The inverse of (528) is

$$F_{ss}^{-1} = \mathbf{J}^{-1} (\text{Re}(\frac{\sigma^2}{2K} (D^H P_{\perp} D) \odot P_s))^{-1} \mathbf{J}^{-t} . \quad (529)$$

Substituting (529) into (476) the lower bound on subspace estimation accuracy is the given by the formula

$$\epsilon_{ss} \geq \sqrt{\frac{\sigma^2}{2K} \text{tr}(((D^H P_{\perp} D) \odot P_s)^{-1} \cdot \mathbf{J}^{-2})} \quad (530)$$

Using (524) $\mathbf{J}^{-2}$ evaluates to

$$\mathbf{J}^{-2} = (D^H P_{\perp} D) \odot (H^H H)^{-1} . \quad (531)$$

Coordinate Transformation of the subspace FIM

The subspace FIM with respect to the natural coordinate basis, denoted F'_{ss} , may be found via the transformation of F_{ss} as (513). Using (528) and (513) yields

$$F'_{ss} = \frac{2K}{\sigma^2} \text{Re}((D^H P_{\perp} D) \odot P_s) . \quad (532)$$

The inverse FIM is

$$F'^{-1}_{ss} = \frac{\sigma^2}{2K} \text{Re}((D^H P_{\perp} D) \odot P_s)^{-1} \quad (533)$$

and the bound on the estimation error covariance is

$$C'_{ss} \succeq F'^{-1}_{ss} . \quad (534)$$

The bound on the error variance is

$$\text{tr} C'_{ss} > \text{tr} F'^{-1}_{ss} . \quad (535)$$

Recalling that $\Delta = \sum_{k=1}^q \dot{s}_k \delta \theta^k$, that is, the components of the tangent vector represent the error in the parameter estimates, we have

$$C'_{ss} = C_{\theta\theta} = E[(\hat{\theta} - \theta)(\hat{\theta} - \theta)^t] \quad (536)$$

and

$$\text{tr}C_{\hat{s}\hat{s}} = \text{tr}C_{\theta\theta} = \sum_{k=1}^p E[(\hat{\theta}^k - \theta^k)^2] . \quad (537)$$

The inverse FIM in (533) is recognized as the asymptotic (for $K \gg 0$) covariance matrix of the unconditional maximum likelihood (UML) estimate as given in [6]. Of course, the FIM could first be computed with respect to the natural coordinate basis (520) directly from (471) and then transformed into the orthonormal coordinates.

Generalization to Higher Dimensions: Modal Model

The above results were developed for the one parameter per mode case. These result can be generalized to the $\bar{M} = r \cdot M$ parameters per mode case in a straightforward way. In array processing applications M denotes the geometric dimension of the array and $r = 1$ for undamped exponentials and $r = 2$ damped exponentials.

In this general case the dimension of the S_Y tangent space is $q = r \cdot M \cdot p$ where p is the number of modes. In the 1-dimensional array problem $M = 1$ and for the damped exponential signal model (shift invariant model) there are 2-parameters per mode, $r = 2$ so that $q = 2p$. In the 2-dimensional array case, $M = 2$, for the undamped exponential model $r = 1$ so that $q = 2p$, In the damped exponential model, $r = 2$, so that for the 2-dimensional array case $q = 4p$.

In the general case we denote the collection of $q = rMp$ parameters as

$$\theta = \begin{bmatrix} \theta_1 & \cdots & \theta_{rM} \end{bmatrix} \quad (538)$$

where θ_k is the p -tuple for the k th parameter type

$$\theta_k = \begin{bmatrix} \theta_k^1 & \cdots & \theta_k^p \end{bmatrix}; \quad k = 1 \cdots rM . \quad (539)$$

The corresponding signal function for p modes is the $N \times p$ complex matrix

$$H(\theta) = H(\theta_1, \cdots, \theta_{rM}) = \begin{bmatrix} h(\theta_1^1, \cdots, \theta_{rM}^1), & \cdots & h(\theta_1^p, \cdots, \theta_{rM}^p) \end{bmatrix} . \quad (540)$$

Thus each mode $h(\cdot)$ is defined by $r \cdot M$ parameters. The derivative matrix is the $N \times q$ complex matrix

$$D = \begin{bmatrix} D_1 \cdots D_{rM} \end{bmatrix} \quad (541)$$

where D_m is an $N \times p$ complex matrix given by

$$D_m = \begin{bmatrix} d_{m,1} & \cdots & d_{m,p} \end{bmatrix}, \quad (542)$$

and

$$d_{m,k} \equiv \frac{\partial h}{\partial \theta_m^k}. \quad (543)$$

The $m; k$ -th column of the derivative matrix $D(\theta)$ corresponds to $\frac{dH(\theta_{(m;k)}(t))}{dt}$ and is

$$\frac{dH(\theta_{(m;k)}(t))}{dt} = [0, \dots, 0, d_{m,k}, 0, \dots, 0] = D(\theta) \cdot J_{mkk} \quad (544)$$

where J_{mkk} is a $q \times p$ selection matrix given by

$$J_{mkk} = i_m \otimes J_{kk}; \quad m = 1 \cdots rM \quad k = 1 \cdots p. \quad (545)$$

In this expression, J_{kk} is the $p \times p$ selection matrix given by (517) and i_m is $rM \times 1$ unit vector with 1 at the m -th location (zeros elsewhere). Then the multi-dimensional extension of (518) is

$$\dot{S}_l = \dot{S}_{m;k} = Y_{\perp}^H D J_{mkk} T \quad (546)$$

and so

$$E_{\dot{s}}(:, l) = \text{vec}(\dot{S}_l) = (T^t \otimes Y_{\perp}^H D) \text{vec}(J_{mkk}) \quad (547)$$

where the index l is given by $l = m + k$.

The p -columns of the matrix $E_{\dot{s}}$ corresponding to the m th parameter vector θ_m (539) are

$$\begin{bmatrix} \text{vec}(\dot{S}_{m;1}) & \cdots & \text{vec}(\dot{S}_{m;p}) \end{bmatrix} = (T^t \otimes Y_{\perp}^H D) G_m \quad (548)$$

where G_m is the $qp \times p$

$$G_m = \begin{bmatrix} \text{vec}(J_{m11}) & \cdots & \text{vec}(J_{mpp}) \end{bmatrix}. \quad (549)$$

Concatenating rM such matrices of this form gives the complete $E_{\dot{s}}$ matrix of size $(\bar{N}_b \times q)$ as

$$E_{\dot{s}} = \left[(T^t \otimes Y_{\perp}^H D)G_1, \dots, (T^t \otimes Y_{\perp}^H D)G_{rM} \right] \quad (550)$$

or equivalently

$$E_{\dot{s}} = (T^t \otimes Y_{\perp}^H D)G \quad (551)$$

where

$$G = \left[G_1 \dots G_{rM} \right]. \quad (552)$$

Note that $E_{\dot{s}}$ is $\bar{N}_b \times q$ since the factor $(T^t \otimes Y_{\perp}^H D)$ is $(\bar{N}_b \times qp)$ and $\left[G_1 \dots G_M \right]$ is $qp \times q$.

Example:2-Dimensional Array Harmonic Signals In this case $r = 1, M = 2$ so that there are 2-parameters per mode and for p modes $q = 2p$. The modal signal matrix has the functional form

$$H(\theta) = H(\theta_h, \theta_v) = \left[h(\theta_h^1, \theta_v^1), \dots, h(\theta_h^p, \theta_v^p) \right]. \quad (553)$$

In this expression we used the suggestive notation $m = h, v$ for the parameter index (i.e., h =horizontal, v =vertical). The derivative matrix (541) is $N \times 2p$

$$D = \left[D_h, D_v \right] \quad (554)$$

where

$$D_h = \left[d_{h,1} \dots d_{h,p} \right] \quad (555)$$

$$D_v = \left[d_{v,1} \dots d_{v,p} \right] \quad (556)$$

and $d_{h,k} \equiv \frac{\partial h}{\partial \theta_h^k}, d_{v,k} \equiv \frac{\partial h}{\partial \theta_v^k}$ are derivatives evaluated at (θ_h^k, θ_v^k) . The two sets of selection matrices corresponding to the $D_1 = D_h$ and $D_2 = D_v$ matrices are each $N \times p$. For this $\bar{M} = rM = 2$ parameter per mode examples the G matrix (552) has the form

$$G = \left[G_1 \ G_2 \right] \quad (557)$$

where

$$G_m = \begin{bmatrix} \text{vec}(J_{m11}) & \text{vec}(J_{m22}) \end{bmatrix}; m = 1, 2 \quad (558)$$

and the $2p \times p$ selection matrices (545) are

$$J_{1kk} = i_1 \otimes J_{kk} = \begin{bmatrix} J_{kk} \\ 0_{p \times p} \end{bmatrix}; \quad k = 1 \cdots p \quad (559)$$

and

$$J_{2kk} = i_2 \otimes J_{kk} = \begin{bmatrix} 0_{p \times p} \\ J_{kk} \end{bmatrix}; \quad k = 1 \cdots p. \quad (560)$$

For 2 source signals, that is $p = 2$ we have $\bar{N}_b = 2(N - 2)$ and $q = 4$. The D_h, D_v in (564) are each $N \times 2$

$$D_h = \begin{bmatrix} D_{h;1} & D_{h;2} \end{bmatrix}, \quad D_v = \begin{bmatrix} D_{v;1} & D_{v;2} \end{bmatrix} \quad (561)$$

so that D is $N \times 4$.

The two J_{ikk} selection matrices are each 4×2 so that $\text{vec}(J_{ikk})$ is 8×1 . The matrix G (557) formed from 4 such selection matrices is therefore 8×4 and given by

$$G = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \quad (562)$$

The $\bar{N}_b \times q$ matrix E_s for this choice of M, p and r case is $(2(N - 2) \times 4)$. The factor $(T^t \otimes Y_{\perp}^H D)$ is $(2(N - 2) \times 8)$ so that the product $(T^t \otimes Y_{\perp}^H D)G$ which defines E_s (551) is $(2(N - 2) \times 4)$ and so is dimensionally correct.

Example: 2-Dimensional Damped Exponentials Signals. In this example we add horizontal and vertical damping parameters to the previous case so that $r = 2$ and there are 4-parameters per mode, ($\bar{M} = rM = 4$). (This signal model is of interest

since on a uniformly spaced 2-D array such signals are shift-invariant). In this situation the modal signal matrix has the functional form

$$H(\theta) = H(\theta_h, \theta_v, \rho_h, \rho_v) = \begin{bmatrix} h(\theta_h^1, \theta_v^1, \rho_h^1, \rho_v^1), & \cdots & h(\theta_h^p, \theta_v^p, \rho_h^p, \rho_v^p) \end{bmatrix} \quad (563)$$

and the $N \times 4p$ derivative matrix D has the form

$$D = \begin{bmatrix} D_{\theta_h}, D_{\theta_v}, D_{\rho_h}, D_{\rho_v} \end{bmatrix} \quad (564)$$

where each of the 4 submatrices appearing above are $N \times p$.

Pertinent Signal Models Summary The parameters for some pertinent signal model cases are listed in the table below. The last column in the table gives the dimension of the space S_Y , the space of subspace perturbations corresponding to the constraint submanifold. Recall that the $S_Y < B_Y$, where B_Y is the space of subspace perturbations for the parent manifold; $\dim(B_Y) = N_b = 2p(N - p)$.

Table 2. Dimension of S_Y for $M = 1, 2$ dimensional Arrays, Undamped ($r = 1$) and Damped ($r = 2$) exponential signals

Data Model	Derivative Matrix Form	parameters per mode ($\bar{M} = r \cdot M$)	$\dim(S_Y)$
1 dim undamped Exp	$[D_\theta]$	1	p
1 dim damped Exp	$\begin{bmatrix} D_\theta & D_\rho \end{bmatrix}$	2	$2p$
2 dim undamped Exp	$\begin{bmatrix} D_{\theta_h} & D_{\theta_v} \end{bmatrix}$	2	$2p$
2 dim damped Exp	$[D_{\theta_h}, D_{\rho_h}, D_{\theta_v}, D_{\rho_v}]$	4	$4p$

Generalization to Higher Dimensions: FIM and CRB

Lemma 18 (Orthogonalizing Matrix $\mathbf{J}$ for $\bar{M} > 1$ case). *For the matrix E_s given by (551) the following equality holds*

$$\text{Re}(E_s^H E_s) = \text{Re}((D^H P_\perp D) \odot G^{(M)}) \quad (565)$$

where

$$G^{(M)} \equiv (1_M \otimes 1_M^t) \otimes (H^H H)^{-1}; \quad k = 1 \cdot q \quad (566)$$

and where 1_M is a $M \times 1$ column vector of all 1s.

Proof:

The real symmetric $q \times q$ orthogonalizing matrix $\mathbf{J}$ is

$$\mathbf{J} = (\text{Re}(D^H P_\perp D) \odot X^{(\bar{M})})^{\frac{1}{2}} \quad (567)$$

where $X^{\bar{M}}$ is $q \times q$ given by

$$X^{\bar{M}} = (1_{\bar{M}} \otimes 1_M^t) \otimes (H^H H)^{-1} \quad (568)$$

and where $1_{\bar{M}}$ is an $\bar{M} \times 1$ column vector of all ones ($\bar{M} = rM$).

Proof: Appendix B.1.6

Remark: For the case of the harmonic signals ($r = 1$) on a 1 dimensional array $M = 1$, $\bar{M} = 1$ and $(1_{\bar{M}} \otimes 1_M^t)$ reduces to the scalar 1 so the previous result for the orthogonalizing matrix given by (511) is recovered.

The multidimensional $\bar{M} > 1$ generalizations of formulas for F_{ss} (528) and $F_{\dot{s}\dot{s}}$ (532) are given in the following theorem.

Theorem 19 (FIM subspace block matrix F_{ss} Modal Case, $\bar{M} > 1$). *The subspace block of the FIM with respect to the natural coordinate basis is*

$$F_{\dot{s}\dot{s}} = \text{Re}\left(\frac{2K}{\sigma^2}(D^H P_\perp D) \odot P_s^{(\bar{M})}\right) \quad (569)$$

where

$$P_s^{(\bar{M})} = P_s \otimes (1_{\bar{M}} \otimes 1_M^t) \quad (570)$$

and P_s is given by (526) or (527). The subspace block of the FIM w.r.t the orthonormal basis is

$$F_{ss} = \mathbf{J} \text{Re}\left(\frac{2K}{\sigma^2}(D^H P_\perp D) \odot P_s^{(\bar{M})}\right) \mathbf{J}^t. \quad (571)$$

where $\mathbf{J}$ is given by (567)

Proof: B.1.9

Intrinsic CRB on subspace estimation accuracy

The lower bound on subspace estimation accuracy is the given by the formula

$$\epsilon_{SS} > \sqrt{\text{tr} F_{ss}^{-1}} \quad (572)$$

where

$$\text{tr} F_{ss}^{-1} = \frac{\sigma^2}{2K} \text{tr}(((D^H P_{\perp} D) \odot P_s^{(M)})^{-1} \cdot \mathbf{J}^{-2}) \quad (573)$$

and from (567)

$$\mathbf{J}^{-2} = ((D^H P_{\perp} D) \odot X^{(\bar{M})})^{-1} . \quad (574)$$

CRB on the parameter θ

Following the discussion for the 1-dimensional case, the estimation error covariance $C_{\theta\theta}$ satisfies the CRB matrix inequality

$$C_{\theta\theta} \succeq F_{\dot{s}\dot{s}}^{-1} = \frac{\sigma^2}{2K} \text{Re}((D^H P_{\perp} D) \odot P_s^{(\bar{M})})^{-1} \quad (575)$$

where the right hand side is inverse of $F_{\dot{s}\dot{s}}$ (569) and

$$C_{\dot{s}\dot{s}} = C_{\theta\theta} = E[(\hat{\theta} - \theta)(\hat{\theta} - \theta)^t] . \quad (576)$$

This is recognized as the asymptotic (for $K \gg 0$) covariance matrix of the unconditional maximum likelihood (UML) estimate as given in [6]. The error variance is therefore bounded as

$$\text{tr} C_{\dot{s}\dot{s}} = \text{tr} C_{\theta\theta} > \text{tr} \left(\frac{\sigma^2}{2K} \text{Re}((D^H P_{\perp} D) \odot P_s^{(\bar{M})})^{-1} \right) . \quad (577)$$

The variance $\text{tr} C_{\theta\theta}$ is equivalent to

$$\text{tr} C_{\theta\theta} = \sum_{k=1}^p \sum_{j=1}^{\bar{M}} E[(\hat{\theta}_{(j)}^k - \theta_{(j)}^k)^2] . \quad (578)$$

List of References

- [1] A. Boumal, “On intrinsic cramér-rao bounds for riemannian submanifolds and quotient manifolds,” *IEEE Transactions on Signal Processin*, vol. 61, no. 7, April 2013.

- [2] S. Smith, "Covariance, subspace, and intrinsic cramér-rao bounds," *IEEE Trans. Signal Process.*, vol. 53, no. 5, p. 1610 1630, 2005.
- [3] A. Edelman, T. A. Arias, and S. Smith, "The geometry of algorithms with orthogonality constraints," *SIAM Matrix Anal. Appl.*, vol. 20, no. 2, p. 303353, 1998.
- [4] L. L. Scharf, *Statistical Signal Processing: Detection, Estimation, and Time Series Analysis*. Reading, MA: Addison-Wesley, 1991.
- [5] D. J. Jeffries and D. R. Farrier, "Asymptotic results for eigenvector methods," *Proc. Inst. Elec. Eng.*, vol. 132, no. 7, pp. 589 –594, June 1985.
- [6] P. Stoica and A. Nehorai, "Music, maximum likelihood, and cramér-rao bound," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, no. 5, p. 720 741, May 1989.

CHAPTER 4

Maximum Likelihood Estimator

The log likelihood function (ignoring constants) for the stochastic complex linear model introduced in chapter 3 is

$$L(Y, \Lambda, \sigma^2; \tilde{R}) = -K(\text{tr}(R^{-1}\tilde{R}) + \log \det R) \quad (579)$$

where the parameters Y, Λ, σ^2 appear in the function as the model covariance

$$R = R(Y, \Lambda, \sigma^2) = Y\Lambda_s Y^H + \sigma^2 I = Y\Lambda Y^H + \sigma^2 Y_\perp Y_\perp^H \quad (580)$$

where

$$\Lambda \equiv \Lambda_s + \sigma^2 I. \quad (581)$$

The data $Z \equiv [z_1, \dots, z_k]$ is compressed in the SCM, denoted by $\tilde{R} = K^{-1} Z Z^H$, which has the eigen decomposition

$$\tilde{R} = \tilde{Y}_s \tilde{\Lambda}_s \tilde{Y}_s^H + \tilde{Y}_\perp \tilde{\Lambda}_o \tilde{Y}_\perp^H \quad (582)$$

where $\tilde{\Lambda}_s$ is $p \times p$ and $\tilde{\Lambda}_o$ is $(N - p) \times (N - p)$. The same symbol “ R ” is used for both the SCM as “ $\tilde{R}$ ” and for a model covariance (580) expressed as a function of the unknown parameters Y, Λ, σ^2 .

The MLE estimates of σ^2 and Λ are given in terms of the components of the SCM (582) as

$$\hat{\sigma}^2 = \frac{1}{N - p} \text{tr}(\tilde{\Lambda}_o) \quad (583)$$

$$\hat{\Lambda} = \tilde{\Lambda}_s - \hat{\sigma}^2 I \quad (584)$$

In the remainder of the discussion we consider the compressed likelihood function defined as $L(Y; \tilde{R}) \equiv L(Y, \Lambda = \hat{\Lambda}, \sigma^2 = \hat{\sigma}^2; \tilde{R})$.

Let α, β denote a normal coordinate system generated by the partitioned basis $(\mathbf{a}_k, \mathbf{b}_k)$ with its origin at an arbitrary point $Y \in \mathcal{B}$. The arbitrary tangent vector Δ expands as

$$\Delta(\alpha, \beta) \equiv \sum_{k=1}^{N_a} \mathbf{a}_k \alpha^k + \sum_{k=1}^{N_b} \mathbf{b}_k \beta^k . \quad (585)$$

The transformation to extrinsic coordinates is $Y(\alpha, \beta) = \exp_Y \Delta(\alpha, \beta)$. Under this coordinate transformation, or change of variables, a function L defined on $\mathcal{B}$ in the extrinsic Y coordinates as $L(Y)$ becomes

$$L(\alpha, \beta) \equiv L(Y(\alpha, \beta)) \quad (586)$$

where $Y = Y(\alpha, \beta)$ expresses the coordinate transformation.

The partial derivative $\frac{\partial L}{\partial(\alpha^k)}$ is defined as the directional derivative in the direction specified by the coordinate basis vector $\mathbf{a}_k = Y_{\perp} A_k$. The partial derivative $\frac{\partial L}{\partial(\beta^k)}$ is defined in a similar way. The following lemma first records the form of the directional derivative for an arbitrary direction specified by the tangent vector

$$\Delta = Y A + Y_{\perp} B \quad (587)$$

and then computes the desired coordinate partials using the appropriate choices for A and B .

Lemma 20 (First Directional Derivative). *The directional derivative of the function L (579) in direction of the arbitrary tangent vector $\Delta = Y A + Y_{\perp} B$ is*

$$\nabla L[\Delta] = \frac{dL(Y(t))}{dt} \Big|_{t=0} = -K \cdot \text{tr}(\delta R_1(A, B) \tilde{R} R^{-1} (I - \tilde{R} R^{-1})) \quad (588)$$

where

$$\delta R_1(A, B) = Y A \hat{\Lambda} Y^H + Y \hat{\Lambda} A^H Y^H + Y_{\perp} B \hat{\Lambda} Y^H + Y \hat{\Lambda} B^H Y_{\perp}^H . \quad (589)$$

For $\Delta = \mathbf{b}_k = Y_{\perp} B_k$, a basis vector for B_Y , this evaluates to

$$\frac{\partial L}{\partial \beta^k} = \nabla L[\mathbf{b}_k] = \sigma^{-2} 2 \text{tr}(\text{Re}(\hat{\Lambda} \tilde{\Lambda}_s^{-1} Y^H \tilde{R} Y_{\perp} B_k)); \quad k = 1 \cdots N_b . \quad (590)$$

For $\Delta = \mathbf{a}_k = Y A_k$, a basis vector for A_Y , this evaluates to

$$\frac{\partial L}{\partial(\alpha^k)} \equiv \nabla L[\mathbf{a}_k] = \text{tr}(Y(\tilde{\Lambda}_s^{-1} A \hat{\Lambda} \tilde{\Lambda}_s^{-1} + \tilde{\Lambda}_s^{-1} \hat{\Lambda} A^H \tilde{\Lambda}_s^{-1}) Y^H \tilde{R}) \quad k = 1 \cdots N_a. \quad (591)$$

Proof: (see appendix C.1)

The ML estimate of $Y \in \mathcal{B}$ is defined as

$$\hat{Y} = \max_{Y \in \mathcal{B}} L(Y; \tilde{R}). \quad (592)$$

At this value the partials (590) and (591) vanish; the $Y \in \mathcal{B}$ satisfying this simultaneous set of equations is recorded in the following theorem.

Theorem 21 (Unconstrained Maximum Likelihood Basis Estimate). *A necessary condition for the point Y to be an extremal of L is that partial derivatives w.r.t. (α, β) vanish, that is*

$$\frac{\partial L}{\partial(\alpha^k)} = 0; \quad k = 1 \cdots N_a \quad (593)$$

and

$$\frac{\partial L}{\partial(\beta^k)} = 0; \quad k = 1 \cdots N_b. \quad (594)$$

The solution to this set of equations that maximizes the value of the L is given by

$$\hat{Y} = \tilde{Y}_s. \quad (595)$$

Proof: (see appendix C.2)

Recall that the equivalence relation on the basis manifold is $Y \sim Y \Lambda_Q$ where Λ_Q is an arbitrary unitary diagonal matrix. Therefore the numerical equivalence class of ML estimates is the set $\hat{Y} \Lambda_Q$, Λ_Q arbitrary unitary diagonal. In numerical applications one representative of the class must be selected and the choice used in what follows is $\hat{Y} = \tilde{Y}_s$, the principal eigenvectors given by EVD of the SCM (582).

The model covariance function R (580) evaluated at the MLE parameter values is denoted $\hat{R}$ and is given by

$$\hat{R} = R(Y = \tilde{Y}, \Lambda = \hat{\Lambda}, \sigma^2 = \hat{\sigma}^2) = \tilde{Y} \tilde{\Lambda}_s \tilde{Y}^H + \hat{\sigma}^2(I - \tilde{Y} \tilde{Y}^H) \quad (596)$$

and the inverse is

$$\hat{R}^{-1} = \tilde{Y} \tilde{\Lambda}_s^{-1} \tilde{Y}^H + \hat{\sigma}^{-2}(I - \tilde{Y} \tilde{Y}^H). \quad (597)$$

Using this form the product $\hat{R}^{-1} \tilde{R}$ which appears in the ML function definition when evaluated at the MLE is

$$\hat{R}^{-1} \tilde{R} = \tilde{Y} \tilde{Y}^H + \hat{\sigma}^{-2}(I - \tilde{Y} \tilde{Y}^H)(\tilde{Y}_\perp \tilde{\Lambda}_o \tilde{Y}_\perp^H) \quad (598)$$

$$= \tilde{Y} \tilde{Y}^H + \hat{\sigma}^{-2} \tilde{Y}_\perp \tilde{\Lambda}_o \tilde{Y}_\perp^H \quad (599)$$

$$= I - \tilde{Y}_\perp \tilde{Y}_\perp^H + \hat{\sigma}^{-2} \tilde{Y}_\perp \tilde{\Lambda}_o \tilde{Y}_\perp^H \quad (600)$$

or

$$\hat{R}^{-1} \tilde{R} = I + (\tilde{Y}_\perp \delta \tilde{\Lambda}_o \tilde{Y}_\perp^H) \quad (601)$$

where

$$\delta \tilde{\Lambda}_o = (\tilde{\Lambda}_o - \hat{\sigma}^2 I_{N-p}) / \hat{\sigma}^2. \quad (602)$$

Using this expression for $\hat{R}^{-1} \tilde{R}$ the value of the function at the ML estimates is

$$L(\tilde{R}; \hat{Y}, \hat{\Lambda}, \hat{\sigma}^2) = -K(\text{tr}(I + \tilde{Y}_\perp \delta \tilde{\Lambda}_o \tilde{Y}_\perp^H) + \log \prod_{k=1} \tilde{\Lambda}_{s;k} - (N-p) \log(\hat{\sigma}^2)) \quad (603)$$

$$= -K((N + \text{tr}(\tilde{Y}_\perp \delta \tilde{\Lambda}_o \tilde{Y}_\perp^H) + \log \prod_{k=1} \tilde{\Lambda}_{s;k} - (N-p) \log(\hat{\sigma}^2)). \quad (604)$$

4.1 Approximating the compressed likelihood function near $\tilde{Y}$

In this section we consider the compressed log-likelihood function as $L(Y) \equiv L(Y; \hat{\Lambda}, \hat{R})$. We develop a 2nd order approximation to $L(Y)$ in a neighborhood of the unconstrained MLE $\tilde{Y} \in \mathcal{B}$. Let $\Delta \in T_{\tilde{Y}} \mathcal{B}$ be an arbitrary unit length tangent vector and let $Y(t)$ be the corresponding geodesic curve emanating from $\tilde{Y}$ given by $Y(t) =$

$\exp_{\tilde{Y}}(t\Delta)$. The value of the log likelihood function on this curve is denoted $L(Y(t))$.

The Taylor series expansion of $L(Y(t))$ is

$$L(Y(t)) = L(Y(0)) + t \frac{dL(Y(0))}{dt} + \frac{1}{2} t^2 \frac{d^2 L(Y(0))}{dt^2} + h.o.t \quad (605)$$

where $Y(0) = \tilde{Y}$. At the MLE the first derivative $\frac{dL(Y(0))}{dt}$ vanishes by definition.

The component sub-matrices of the 2nd covariant differential $\nabla^2 L$ with respect to this basis are given by the ordinary Hessian of $L(\alpha, \beta)$ in block form by

$$[\nabla^2 L] = \begin{bmatrix} L_{\alpha\alpha} & L_{\alpha\beta} \\ L_{\beta\alpha} & L_{\beta\beta} \end{bmatrix} \quad (606)$$

(see section 3.6). For the log likelihood function (579) the component matrices $L_{\beta\beta}$ and $L_{\alpha\beta}$ are as given in the following lemma.

Lemma 22. *The component sub-matrices $L_{\beta\beta}$ and $L_{\alpha\beta}$ in (606) are given by*

$$L_{\beta\beta} = \text{Re}(E_b^H L_{bb} E_b) \quad (607)$$

where

$$L_{bb} \equiv \frac{2K}{\tilde{\sigma}^2} \tilde{\Lambda}_{bb} \quad (608)$$

and where $\tilde{\Lambda}_{bb}$ is real the $\bar{N}_b \times \bar{N}_b$ diagonal matrix

$$\tilde{\Lambda}_{bb} \equiv (\tilde{\Lambda}_{OPT} \otimes (I_{N-p} + \delta \tilde{\Lambda}_o)) \quad (609)$$

and $\tilde{\Lambda}_{OPT} = \tilde{\Lambda}^2(\tilde{\Lambda} - \hat{\sigma}^2 I)^{-1}$ with $\delta \tilde{\Lambda}_o$ given by (602).

The mixed derivative submatrix is identically zero

$$L_{\alpha\beta} = 0. \quad (610)$$

For the standard basis on $B_{\tilde{Y}}$, the basis matrix given by (425), the matrix $L_{\beta\beta}$ (607) simplifies to

$$L_{\beta\beta} = I_2 \otimes L_{bb} = \frac{2K}{\tilde{\sigma}^2} \begin{bmatrix} \tilde{\Lambda}_{bb} & 0 \\ 0 & \tilde{\Lambda}_{bb} \end{bmatrix}. \quad (611)$$

Proof: Appendix

The second directional derivative in the direction $\Delta = Y A + Y_{\perp} B$ is

$$\frac{d^2 L(t)}{dt^2} = -K \text{tr}((\delta R_1(A, B) R^{-1})^2 (\tilde{R} R^{-1} - \frac{1}{2} I)) \quad (612)$$

and using (601)

$$(\tilde{R} R^{-1} - \frac{1}{2} I) = \frac{1}{2} I + (\tilde{Y}_{\perp} \delta \tilde{\Lambda}_o \tilde{Y}_{\perp}^H) . \quad (613)$$

(612) is

$$\frac{d^2 L(t)}{dt^2} = -K \left(\frac{1}{2} \text{tr}((\delta R_1(A, B) R^{-1})^2) + \text{tr}((\delta R_1(A, B) R^{-1})^2 (\tilde{Y}_{\perp} \delta \tilde{\Lambda}_o \tilde{Y}_{\perp}^H)) \right) . \quad (614)$$

To compute the desired derivative recall (410) and (B.81), that is

$$(\delta R_1(0, B) R^{-1})^2 = \sigma^{-2} Y_{\perp} B \Lambda \Lambda_s^{-1} \Lambda B^H Y_{\perp}^H + \sigma^{-2} Y \Lambda B^H B \Lambda \Lambda_s^{-1} Y^H . \quad (615)$$

Using this, the second term in (614) is

$$\text{tr}((\delta R_1(0, B) R^{-1})^2 (\tilde{Y}_{\perp} \delta \tilde{\Lambda}_o \tilde{Y}_{\perp}^H)) = \sigma^{-2} \text{tr}(Y_{\perp} B \Lambda_{OPT} B^H \delta \tilde{\Lambda}_o \tilde{Y}_{\perp}^H) \quad (616)$$

$$= \sigma^{-2} \text{tr}(\Lambda_{OPT} B^H \delta \tilde{\Lambda}_o B) \quad (617)$$

where $\Lambda_{OPT} \equiv \Lambda \Lambda_s^{-1} \Lambda = \Lambda_s^{-1} \Lambda^2$. Evaluating the first term in a similar way and summing the two terms yields

$$[L_{\beta\beta}]_{kk} = \frac{2K}{\sigma^2} \text{Re} \left(\text{tr}(\Lambda_{OPT} B_k^H B_k) + \text{tr}(\Lambda_{OPT} B_k^H \delta \tilde{\Lambda}_o B_k) \right); \quad k = 1 \cdots N_b . \quad (618)$$

Applying the matrix identity B.1.10 the second term is

$$\text{tr}(\Lambda_{OPT} B^H \delta \tilde{\Lambda}_o B) = b^H (\Lambda_{OPT} \otimes I_{N-p}) (I_p \otimes \delta \tilde{\Lambda}_o) b = b^H (\Lambda_{OPT} \otimes \delta \tilde{\Lambda}_o) b \quad (619)$$

and the first term is

$$\text{tr}(\Lambda_{OPT} B^H B) = b^H (\Lambda_{OPT} \otimes I_{N-p}) b . \quad (620)$$

Replacing the B by B_k and substituting (619) and (620) into (618) yields the desired result

$$L_{\beta\beta} = \frac{2K}{\sigma^2} \text{Re} \left(E_b \left(\Lambda_{OPT} \otimes (I_{N-p} + \delta \tilde{\Lambda}_o) \right) E_b \right) \quad (621)$$

The second order term in the Taylor expansion (605) may be written as

$$t^2 \frac{d^2 L(Y(0))}{dt^2} = t^2 [\alpha^t \beta^t] \begin{bmatrix} L_{\alpha\alpha} & L_{\alpha\beta} \\ L_{\beta\alpha} & L_{\beta\beta} \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} \quad (622)$$

Using the above quadratic expression and $L_{\alpha\beta} = 0$ the form (605) is

$$L(Y(t)) = L(t\alpha, t\beta) = L(0, 0) + t^2(\beta^t L_{\beta\beta} \beta + \alpha^t L_{\alpha\alpha} \alpha) + h.o.t. \quad (623)$$

where we have used the notation $L(t\alpha, t\beta) = L(Y(t))$ to represent $L(Y(t))$ in the (α, β) normal coordinates. Absorbing the parameter t into the components (α, β) a 2nd order expression for the likelihood function in the neighborhood of $\tilde{Y}$ is

$$L(\alpha, \beta) = L(0, 0) + \beta^t L_{\beta\beta} \beta + \alpha^t L_{\alpha\alpha} \alpha. \quad (624)$$

Note that the origin of (α, β) coordinate system corresponds to $\tilde{Y}$.

Partitioning the real N_b -tuple β as $\beta = \begin{bmatrix} \beta_R \\ \beta_I \end{bmatrix}$ and using (607), the quadratic form $\beta^t L_{\beta\beta} \beta$ is

$$\beta^t L_{\beta\beta} \beta = \text{Re} \left(\left(\beta_R^t L_{bb} \beta_R + \beta_I^t L_{bb} \beta_I \right) + i \cdot \left(\beta_I^t L_{bb} \beta_R + \beta_R^t L_{bb} \beta_I \right) \right). \quad (625)$$

Defining the complex vector $b \equiv \beta_R + i \cdot \beta_I$ this becomes

$$\beta^t L_{\beta\beta} \beta = \text{Re} \left(b^H L_{bb} b \right) \quad (626)$$

or

$$\beta^t L_{\beta\beta} \beta = \frac{2K}{\tilde{\sigma}^2} \text{Re} \left(b^H \tilde{\Lambda}_{bb} b \right). \quad (627)$$

4.1.1 Constrained MLE Criterion Function

In the constrained problem the log likelihood function has the same form as in the unconstrained case but with a restriction on the domain. In this problem the set of $N \times p$ semi-unitary matrices Y is restricted to a lower dimensional submanifold, denoted $\bar{\mathcal{B}}$, of the full basis manifold $\mathcal{B}$.

The constrained MLE of Y is then defined by

$$\hat{Y} = \max_{Y \in \bar{\mathcal{B}}} L(Y) . \quad (628)$$

The membership constraint condition $Y \in \bar{\mathcal{B}}$ can be stated in terms of the exponential map based at the unconstrained MLE, $\tilde{Y}$, as $\exp_{\tilde{Y}} \Delta \in \bar{\mathcal{B}}$ where $\Delta \in T_{\tilde{Y}} \mathcal{B}$. The set of tangent vectors at $\tilde{Y}$ that satisfy this constraint is denoted $\mathcal{V}_{\tilde{Y}}$

$$\mathcal{V}_{\tilde{Y}} = \{ \Delta \in T_{\tilde{Y}} \mathcal{B} \text{ such that } \exp_{\tilde{Y}} \Delta \in \bar{\mathcal{B}} \} . \quad (629)$$

Using the quadratic approximation of the likelihood function (624) and representing $\Delta = \Delta(\alpha, \beta)$ as (585) the quadratic ML criterion is

$$(\hat{\alpha}, \hat{\beta}) = \min_{\alpha, \beta} (\beta^t L_{\beta\beta} \beta + \alpha^t L_{\alpha\alpha} \alpha) \text{ subject to } \Delta(\alpha, \beta) \in \mathcal{V}_{\tilde{Y}} . \quad (630)$$

Let $Y \in \bar{\mathcal{B}}$ be an arbitrary point near $\tilde{Y}$. Suppose that the corresponding tangent vector is $\Delta = \exp_{\tilde{Y}}^{-1} Y$ and that the components of Δ are (α, β) so that $Y = \exp_{\tilde{Y}} \Delta(\alpha, \beta)$. Following the discussion in section 2.3.2 we have

$$Y \equiv \exp_{\tilde{Y}} \Delta(\alpha, \beta) \approx \exp_{\tilde{Y}} \Delta(0, \beta) Q_1 \quad (631)$$

where the matrix Q_1 can be expressed as $Q_1 = \exp \left(\sum \mathbf{a}_k \alpha^k \right)$. Recall from lemma 14 that if $Y \in \bar{\mathcal{B}}$, it is also the case that $YQ \in \bar{\mathcal{B}}$ for Q an arbitrary $p \times p$ unitary matrix. As a result

$$\Delta(\alpha, \beta) \in \mathcal{V}_{\tilde{Y}}, \text{ for all } \alpha. \quad (632)$$

Since both terms in (630) are positive definite this expression is minimized by choosing the free parameter $\alpha = 0$. Using this result the quadratic criterion (630) can be expressed in terms of the parameter β alone as

$$\hat{\beta} = \min_{\beta} \left(\beta^t L_{\beta\beta} \beta \right) \text{ subject to } \Delta(0, \beta) \in \bar{\mathcal{V}}_{\tilde{Y}} \quad (633)$$

where the set of tangent vectors $\bar{\mathcal{V}}_{\tilde{Y}} \subset \mathcal{V}_{\tilde{Y}}$ is defined as

$$\bar{\mathcal{V}}_{\tilde{Y}} = \{\Delta \in B_{\tilde{Y}} \text{ such that } \exp_{\tilde{Y}} \Delta \in \bar{\mathcal{B}}\} . \quad (634)$$

Since elements of the space $B_{\tilde{Y}}$ have the general numerical representation

$$\Delta = \tilde{Y}_{\perp} B \quad (635)$$

where B is an $(N - p) \times p$ complex matrix, the set $\bar{\mathcal{V}}_{\tilde{Y}}$ (641) can be defined in terms of complex $\bar{N}_b$ -tuple $b \equiv \text{vec}(B)$ as

$$\bar{\mathcal{V}}_{\tilde{Y}} \equiv \{b \in C^{\bar{N}_b} \text{ such that } \exp_{\tilde{Y}} \Delta \in \bar{\mathcal{B}} \text{ where } \Delta = \tilde{Y}_{\perp} \text{mat}(b)\} . \quad (636)$$

The constrained MLE form (633) can then be restated as in terms of complex $\bar{N}_b$ -tuple b using (A.265) as

$$\hat{b} = \min_b \text{Re} \left(b^H \tilde{\Lambda}_{bb} b \right) \text{ subject to } b \in \bar{\mathcal{V}}_{\tilde{Y}} . \quad (637)$$

4.2 Solution to the Constraint MLE criterion using an S-N Frame

For any $Y_0 \in \bar{\mathcal{B}}$ the full tangent space B_{Y_0} , the space corresponding to subspace perturbations, is naturally partitioned as $B_{Y_0} = S_{Y_0} \oplus N_{Y_0}$, where S_{Y_0} is the space tangent to $\bar{\mathcal{B}}$ and N_{Y_0} is the orthogonal complement in B_{Y_0} (see figure 7). Using parallel translation, a similarly partitioned basis for $B_{\tilde{Y}}$ at the unconstrained MLE $\tilde{Y}$ may be formed. Using this S-N frame an alternate form of the MLE criterion (637) is developed. After developing this form we consider a new solution method when the submanifold is defined by a shift-invariant constraint condition.

Constraint Submanifold using an S-N Frame: Examples

Before considering the full problem on the Basis manifold we discuss an example set in a 3-dimensional Euclidean space with the primary objective of illuminating the submanifold membership constraint condition (636). In order to make the ideas more

easily transferable we denote this Euclidan space by $\mathcal{B}$ and generic points in the space by the symbol Y . The submanifold of interest, denoted $\bar{\mathcal{B}}$, is a two-dimensional plane surface.

Consider a point $Y_0 \in \bar{\mathcal{B}}$ and let $\mathbf{n}$ be normal to $\bar{\mathcal{B}}$ and let $\mathbf{a}$ and $\mathbf{s}$ be an orthogonal pair tangent to $\bar{\mathcal{B}}$. Define the 2-dimensional space spanned by $\mathbf{n}$ and $\mathbf{s}$ as B_{Y_0} and the 1-dimensional space spanned by $\mathbf{a}$ as A_{Y_0} (the full 3-dimensional tangent space, denoted $T_{Y_0}\mathcal{B}$ is then $T_{Y_0}\mathcal{B} = A_{Y_0} \oplus B_{Y_0}$).

Let $\tilde{Y}$ denote a point not on submanifold $\bar{\mathcal{B}}$ that lies in plane defined by B_{Y_0} . The submanifold membership condition $Y \in \bar{\mathcal{B}}$ can be stated in terms of $\tilde{Y}$ and the displacement vector Δ as $(\tilde{Y} + \Delta) \in \bar{\mathcal{B}}$. The set of vectors Δ that satisfy this condition is denoted $\bar{\mathcal{V}}$ and is defined as

$$\bar{\mathcal{V}}_{\tilde{Y}} = \{\Delta \in B_{\tilde{Y}} \text{ such that } (\tilde{Y} + \Delta) \in \bar{\mathcal{B}}\} \quad (638)$$

(compare to (634)).

By translating the set of basis vectors $\mathbf{a}, \mathbf{s}, \mathbf{n}$ from Y_0 to $\tilde{Y}$ we establish a basis for the vector space $B_{\tilde{Y}}$. Since the space is flat these vectors can be translated to the point $\tilde{Y}$ in the usual rigid way (i.e.; no rotations). Representing Δ with respect to this basis creates a new coordinate system with origin at $\tilde{Y}$.

The vector with head at $\tilde{Y}$ and tail at Y_0 is given by $\tilde{\Delta} = (Y_0 - \tilde{Y})$. Since by construction Y_0 and $\tilde{Y}$ lie in the same B_{Y_0} plane this vector can be represented in the S-N basis for $B_{\tilde{Y}}$ as $\Delta = \mathbf{s}s_0 + \mathbf{n}\hat{n}$, where

$$\hat{n} = \langle \tilde{\Delta}, \mathbf{n} \rangle, \quad (639)$$

$$s_0 = \langle \tilde{\Delta}, \mathbf{s} \rangle. \quad (640)$$

The 3-tuple $(\alpha = 0, s = s_0, n = \hat{n})$ are the normal coordinates of Y_0 with respect to the given basis with origin at $\tilde{Y}$.

Figure 7. Geometry of the Constraint Submanifold in this example at 2-D plane.

Since, by construction, s at $\tilde{Y}$ is parallel to the plane $\bar{\mathcal{B}}$, elements in the set $\bar{\mathcal{V}}_{\tilde{Y}}$ are given by

$$\Delta = s s + n \hat{n} \quad (641)$$

where $\hat{n}$ is given by (639) and s is a free parameter.

If $\hat{Y} \in \bar{\mathcal{B}}$ then $\hat{Y} = \tilde{Y} + (s\hat{s} + n\hat{n})$ for $\hat{n}$ fixed and s chosen appropriately. The components $(0, \hat{s}, \hat{n})$ are the normal coordinates of $\hat{Y}$ with respect to the s, n basis at $\tilde{Y}$. The point $\hat{Y}$ can, of course, also be expressed in terms of a coordinate system at with origin Y_0 . Referenced to Y_0 the coordinates of $\hat{Y}$ are $(0, \hat{s} - s_0, 0)$.

The distance between $\tilde{Y}$ and Y is equal to the length of the tangent vector Δ and, for the s, n basis, we have

$$dist_{\mathcal{B}}(\tilde{Y}, Y) = \|\Delta\| = \sqrt{\|s\|^2 + \|\hat{n}\|^2}. \quad (642)$$

From this expression it is evident that the point of $\bar{\mathcal{B}}$ nearest to $\tilde{Y}$, denoted Y_{MD} where the subscript stands for minimum distance, has normal coordinates $(0, \hat{b})$ so that

$$Y_{MD} = \tilde{Y} + (n\hat{n}). \quad (643)$$

Constraint Submanifold using an S-N Frame: Basis Manifold

The development of the submanifold membership condition for the manifold case proceeds in same way outlined above with a few modifications. Recall that for any point $Y_0 \in \bar{\mathcal{B}}$, the tangent space B_{Y_0} of the ambient manifold $\mathcal{B}$ partitions as

$$B_{Y_0} = S_{Y_0} \oplus N_{Y_0} \quad (644)$$

where N_{Y_0} is normal to the submanifold tangent plane $T_{Y_0}\bar{\mathcal{B}} = A_{Y_0} \oplus S_{Y_0}$. Let s_k and n_k be an orthonormal basis set for S_{Y_0} and N_{Y_0} , respectively.

In the Euclidean space example this basis defined at Y_0 was translated in the usual parallel way to form a basis for the tangent space $B_{\tilde{Y}}$ at the point $\tilde{Y}$. In the general

case this simple translation is replaced by its generalization, termed parallel translation. Denoting the parallel translation operation as τ we can define a set of q vectors at $\tilde{Y}$ by

$$(\mathbf{s}_k)_{\tilde{Y}} \equiv \tau((\mathbf{s}_k)_{Y_0}), k = 1 \cdots q. \quad (645)$$

Denote the subspace of $B_{\tilde{Y}}$ spanned by this set as $S_{\tilde{Y}}$ and its orthogonal complete as $N_{\tilde{Y}}$ so that

$$B_{\tilde{Y}} = S_{\tilde{Y}} \oplus N_{\tilde{Y}}. \quad (646)$$

Let $(\mathbf{n}_k)_{\tilde{Y}}$ be an orthonormal basis for $N_{\tilde{Y}}$. When it is clear from the context the subscript indicating the base point of the vector is suppressed for convenience.

Using the orthonormal S-N basis for $B_{\tilde{Y}}$ a vector in this space expands as

$$\Delta = \sum_{k=1}^q \mathbf{s}_k s^k + \sum_{k=1}^{N_b-q} \mathbf{n}_k n^k. \quad (647)$$

Given the two points $\tilde{Y}$ and Y_0 the tangent vector between them is defined by $\tilde{\Delta} = \exp_{\tilde{Y}}^{-1} Y_0$ and the components, denoted $(s_0, \hat{n})$ are found using (639) and (640). These components $(0, s_0, \hat{n})$ are the normal coordinates of Y_0 with respect to the given basis for $B_{\tilde{Y}}$.

The set of tangent vectors $\Delta \in B_{\tilde{Y}}$ that satisfy the membership condition $\exp_{\tilde{Y}} \Delta \in \bar{\mathcal{B}}$ can be expressed in terms of this S-N basis and the $\hat{n}$ component defined above. Since $S_{\tilde{Y}}$ is parallel to S_{Y_0} the set of tangent vectors $\bar{\mathcal{V}}_{\tilde{Y}}$ (641) have the form

$$\Delta_{\tilde{Y}} = \sum_{k=1}^q \mathbf{s}_k s^k + \sum_{k=1}^{N_b-q} \mathbf{n}_k \hat{n}^k \quad (648)$$

where $\hat{n}$ is fixed by (639) and where s is a real free q -tuple.

The geodesic distance between the unconstrained MLE $\tilde{Y}$ and nearby elements of $Y \in \bar{\mathcal{B}}$ is given by the magnitude of the tangent vector between them. Using (647) this magnitude is given in terms of the constraint-satisfying pair $(s, \hat{n})$ as

$$dist_{\mathcal{B}}(\tilde{Y}, Y) = \|\Delta\| = \sqrt{\|s\|^2 + \|\hat{n}\|^2}. \quad (649)$$

The point on $\bar{B}$ that is the minimum distance from $\tilde{Y}$, denoted Y_{MD} , occurs for $s = 0$ and is $Y_{MD} = \exp_{\tilde{Y}} \Delta$ where

$$\Delta = \sum_{k=1}^{Nn} \mathbf{n}_k \hat{n}^k . \quad (650)$$

Alternate Real Representations and the set $\bar{\mathcal{V}}_{\tilde{Y}}$

Suppose that a partitioned set of orthonormal S-N basis vectors for $B_{\tilde{Y}}$ are given by $\mathbf{s}_k = \tilde{Y}_{\perp} S_k$ and $\mathbf{n}_k = \tilde{Y}_{\perp} N_k$. Using these basis vectors to represent the vector $\Delta = \tilde{Y}_{\perp} B$ as in (647) yields

$$B \equiv \sum_{k=1}^q S_k s^k + \sum_{k=1}^{N_b-q} N_k n^k . \quad (651)$$

Vectorizing both sides of this equation yields

$$b = E_s s + E_n n \quad (652)$$

where $b = \text{vec}(B)$ and

$$E_s = \begin{bmatrix} \text{vec}(S_1) & \cdots & \text{vec}(S_q) \end{bmatrix} , \quad (653)$$

$$E_n = \begin{bmatrix} \text{vec}(N_1) & \cdots & \text{vec}(N_{(N_b-q)}) \end{bmatrix} . \quad (654)$$

From (648) the set of complex $\bar{N}_b$ -tuples that satisfy the membership constraint condition is given by

$$b = b_n + E_s s \quad (655)$$

where

$$b_n = E_n \hat{n} \quad (656)$$

and s is a free real q -tuple.

Constrained MLE in terms of b

Substituting the constraint-satisfying form for b (655) into the constrained MLE criterion (637) yields an unconstrained minimization in s

$$\hat{s} = \min_s \text{Re} \left((b_n + E_s s)^H \tilde{\Lambda}_{bb} (b_n + E_s s) \right) . \quad (657)$$

Expanding the quadratic form on the right hand side yields

$$\hat{s} = \min_s \left(L_s|_{Y_{MD}} + s^t L_{ss} s + \text{Re}(b_n^H \tilde{\Lambda}_{bb} b_n) \right) \quad (658)$$

where the submatrix F_{ss} (see (471)) is

$$L_{ss} \equiv \text{Re}(E_s^H \tilde{\Lambda}_{bb} E_s) \quad (659)$$

and

$$L_s|_{Y_{MD}} \equiv 2\text{Re}(E_s^H \tilde{\Lambda}_{bb} b_n) . \quad (660)$$

This term (660) corresponds to the first derivative of L with respect to the subspace coordinate variable s evaluated at the point Y_{MD} (643). This is discussed more fully in appendix C.6. Since the last term on the right hand side of is $\text{Re}(b_n^H \tilde{\Lambda}_{bb} b_n) \geq 0$ and independent of s the solution to (657) is given by

$$\hat{s} = \min_s (s^t (L_s|_{Y_{MD}} + L_{ss} s)) \quad (661)$$

which is

$$\hat{s} = L_{ss}^{-1} L_s|_{Y_{MD}} . \quad (662)$$

Substituting this result for $\hat{s}$ into (655) yields

$$\hat{b} = b_n + E_s \hat{s} . \quad (663)$$

The MLE for $Y \in \tilde{\mathcal{B}}$ denoted $\hat{Y}$ then is given by

$$\hat{Y} = \exp_{\tilde{Y}} \hat{\Delta} \quad (664)$$

where the MLE tangent vector is

$$\hat{\Delta} = \tilde{Y}_\perp \hat{B}; \quad \hat{B} = \text{mat}(\hat{b}) . \quad (665)$$

4.3 Solution Methods

The estimation method outlined above required a reference point Y_0 on the submanifold $\bar{\mathcal{B}}$ that was near to the unconstrained MLE $\tilde{Y}$. When the submanifold $\bar{\mathcal{B}}$ is defined by some underlying parametric model, $H = H(\theta)$, a standard approach to implementing the above program is to leverage a suboptimal estimate of the parameter, say θ_0 , to establish this reference point. Using this point and the differentiable signal model, a basis for S_{Y_0} is readily computed. Parallel translations then yields a basis for $S_{\tilde{Y}}$, if desired, although it is generally preferable to work directly with S_{Y_0} when available. In this situation the overall approach reduces to a 1-step Newton's method (662) on the likelihood function in the parameter space. The overall performance of this approach depends in large part on the quality of the initial suboptimal estimate θ_0 . An example of this type of MLE approach using Newton's method is described in [1]. This Newton's method approach is discussed in the context of the established geometric framework in appendix C.6.

In the special case of complex damped exponential signal model on uniformly sampled multidimensional arrays, the constrained submanifold is denoted $\bar{\mathcal{B}}_{damped}$. In this case, matrix representation of any $Y \in \bar{\mathcal{B}}_{damped}$ satisfies a shift invariant property defined below. By using this shift-invariant structure in a more integrated way, we develop a closed-form ML estimator of $Y \in \bar{\mathcal{B}}_{damped}$.

4.3.1 MLE for Submanifold Defined by Shift-Invariance: Uniform Arrays

In this section we consider the submanifold $\bar{\mathcal{B}}_{damped}$; for notational convenience we drop the subscript and denote this submanifold simply as $\bar{\mathcal{B}}$. The set $\bar{\mathcal{V}}_{\tilde{Y}}$ (641) of submanifold constraint satisfying tangent vectors is specified in (655) by the matrix E_s and the complex $\bar{N}_b$ -tuple b_n . The matrix E_s is a representation of a basis for $S_{\tilde{Y}}$. The $\bar{N}_b$ -tuple b_n corresponds to the tangent vector from $\tilde{Y}$ to the nearest point on $\bar{\mathcal{B}}$. Together $\tilde{Y}, E_s$ and b_n define the set of constraint-satisfying tangent vectors (655). Once this set

is known, the ML estimate $\hat{Y} \in \bar{\mathcal{B}}$ is computed from (664), (665) and (663).

Complex Representation of Constraint

In the previous development we treated the submanifold tangent spaces S_Y as real vector spaces. This was convenient for analysis purposes, since in the $H(\theta)$ model the parameters of interest are themselves real (and potentially odd in number). For the special case of the damped exponential signal model, these spaces may be represented as complex vector spaces of lower dimension. This key feature follows from the form of the derivative matrix that defines the space S_Y and the related fact that the signal parameters occur in pairs, so that the real dimension of S_Y is even.

Lemma 23 (Basis for the subspace S_Y on $\bar{\mathcal{B}}_{damped}$). *The basis matrix E_s (653) for the $\bar{\mathcal{B}}_{damped}$ assumption has the form as*

$$E_s = \begin{bmatrix} \bar{E}_s & i \cdot \bar{E}_s \end{bmatrix} \quad (666)$$

where $\bar{E}_s$ is $N_b \times \bar{q}$ and $\bar{q} = \frac{q}{2}$.

Proof: (see appendix C.3)

The columns of $\bar{E}_s$ span S_Y as a complex vector space of dimension $\bar{q} = \frac{q}{2}$. To see this, let $\Delta \in S_Y$ be arbitrary, where S_Y is spanned by E_s (see (A.161)). By definition, this vector has the form $\Delta = Y_{\perp} B$ for some matrix B . Let s be the real q -tuple such that $b = E_s s$, where $b = \text{vec}(B)$. Referring to (666) this can be expressed as

$$b = E_s s = \bar{E}_s s_R + i \bar{E}_s s_I \quad (667)$$

where the real q -tuple s has been partitioned into two $\bar{q}$ -tuples. (i.e.; s_R and s_I are defined by the first $\bar{q}$ and last $\bar{q}$ elements of s , respectively). Define the complex $\bar{q}$ -tuple $\bar{s} = s_R + i \cdot s_I$, so that

$$s = \begin{bmatrix} \text{Re}(\bar{s}) \\ \text{Im}(\bar{s}) \end{bmatrix} = \begin{bmatrix} s_R \\ s_I \end{bmatrix}. \quad (668)$$

Using this, (667) becomes

$$b = E_s s = \bar{E}_s \bar{s} . \quad (669)$$

The following theorem records the complex form of the derivative matrices L_s and L_{ss} that appear in the ML estimate of s using this complex vector space formulation of the space S_Y .

Theorem 24. *If the basis matrix E_s has the form (666) then $s = L_{ss}^{-1} L_s$ defined by (659),(660) is equivalently expressed in complex form as*

$$\hat{\bar{s}} \equiv \bar{L}_{ss}^{-1} \bar{L}_s \quad (670)$$

where $\bar{L}_{ss} \equiv \bar{E}_s^H \tilde{\Lambda}_{bb} \bar{E}_s$ and $\bar{L}_s \equiv \bar{E}_s^H \tilde{\Lambda}_{bb} b_n$. The real q -tuple s and the complex $\bar{q}$ -tuple $\bar{s}$ are related by (668).

Proof: see C.4

As a consequence the ML estimate (663) is given by

$$\hat{b} = b_n + \bar{E}_s \hat{\bar{s}} . \quad (671)$$

Solution

Elements of $\bar{\mathcal{B}}_{damped}$ satisfy a shift invariance condition, defined later, which can be expressed in terms of a vector valued function $r(\cdot)$ defined on $\mathcal{B}$ such that $Y \in \bar{\mathcal{B}}$ if and only if $r(Y) = 0$. In this case the submanifold membership condition is

$$\exp_{\bar{Y}} \Delta \in \bar{\mathcal{B}} \iff r(\exp_{\bar{Y}} \Delta) = 0 \quad (672)$$

and the set of constraint-satisfying tangent vectors $\bar{\mathcal{V}}_{\bar{Y}}$ (641) is

$$\bar{\mathcal{V}}_{\bar{Y}} = \{\Delta \in B_{\bar{Y}} \text{ such that } r(\exp_{\bar{Y}} \tilde{\Delta}) = 0\} . \quad (673)$$

Using the fact that elements of $B_{\bar{Y}}$ have the general form $\Delta = \tilde{Y}_{\perp} B$, we express the constraint as a function of $b = \text{vec}(B)$ for a fixed $\tilde{Y}_{\perp}$

$$r(b; \tilde{Y}) \equiv r(\exp_{\bar{Y}} \Delta) . \quad (674)$$

Consider the Taylor series expansion of the constraint function $r(b; \tilde{Y})$

$$r(b; Y) = r(0; \tilde{Y}) + \sum_{k=1}^{\bar{N}_b} \frac{\partial r}{\partial b^k} b^k + h.o.t. \quad (675)$$

and for convenience define

$$[\widehat{H}]_k^m \equiv -\frac{\partial r^m}{\partial b^k} . \quad (676)$$

($\widehat{H}$ is a function of $\tilde{Y}_\perp$). The first order approximation of the constraint then is

$$r(b; Y) \approx r(\tilde{Y}) - \widehat{H}b . \quad (677)$$

A critical feature of this approximation is that $\widehat{H}$ is constructed in such a way that it has a null space $null(\widehat{H})$ of complex dimension $\bar{q} = \dim S_{\tilde{Y}}$, and this null space well approximates $S_{\tilde{Y}}$. Let $\hat{E}_s$ denote the matrix whose columns are a basis for $null(\widehat{H})$, then

$$col(\hat{E}_s) \approx col(\bar{E}_s) \quad (678)$$

where $\bar{E}_s$ is a basis for $S_{\tilde{Y}}$.

The set of constraint-satisfying tangent vectors $\bar{\mathcal{V}}_{\tilde{Y}}$ (636), (673) is approximated by

$$\bar{\mathcal{V}}_{\tilde{Y}} \approx \{b \in C^{\bar{N}_b} \text{ such that } b = \min \|r(\tilde{Y}) - \widehat{H}b\|^2\} . \quad (679)$$

Elements in this set are given by least squares solution of $\|r(\tilde{Y}) - \widehat{H}b\|^2$ with respect to b , which is

$$b = \hat{b}_n + \hat{E}_s \bar{s} \quad (680)$$

where

$$b_n = \hat{H}^\# r(\tilde{Y}) \quad (681)$$

and $\bar{s}$ is a complex free $\bar{q}$ -tuple.

Remark: If $\tilde{Y} \in \bar{\mathcal{B}}$ then $r(\tilde{Y}) = 0$ and therefore $\hat{b}_n = 0$ (681) so that the solution (680) simplifies to

$$b = \hat{E}_s \bar{s} . \quad (682)$$

Example:1-Dimensional ULA On a N -element uniform line array (ULA) a single damped exponential signal has the form

$$h = z^{(0:N-1)} \quad (683)$$

where $z = \rho \exp(i\theta)$. The parameters θ and ρ are reals and denote the spatial frequency and damping parameter, respectively, of the complex exponential waveform. If J_0 and J_1 represent selection matrices that select the elements $(0 : N - 2)$ and $(1 : N - 1)$ respectively, then $J_0 \cdot h \cdot z = J_1 \cdot h$.

In the case of multiple signals we use subscripts so that $h_k = z_k^{(0:N-1)}$ and $z_k = \rho_k \exp(i\theta_k)$. For p signals the signal matrix is $H = \begin{bmatrix} h_1 & \cdots & h_p \end{bmatrix}$ and has the shift-invariant property

$$J_0 H \Psi = J_1 H \quad (684)$$

where $\Psi = \text{diag}(z_1, \dots, z_p)$.

Let Y be the semi-unitary matrix defined by left singular vector matrix of H so that $H = YT$ where for T is non-singular. Using this (684) becomes $J_0 YT \Psi = J_1 YT$ or equivalently

$$J_0 Y \cdot \Phi = J_1 Y \quad (685)$$

where $\Phi = T \Psi T^{-1}$. Evidently

$$\Phi = (J_0 Y)^\# J_1 Y. \quad (686)$$

By construction, $\text{diag}(\Psi) = \text{eig}(\Phi)$ so that $\text{eig}(\Phi) = \begin{bmatrix} z_1 & \cdots & z_p \end{bmatrix}$. In the pure harmonic case ($\rho_k = 0$) the p -tuple θ is evidently $\theta = \angle(\text{diag}(\Psi)) = \angle(\text{eig}(\Phi))$.

Because a single damped exponential is specified by two real parameters all sub-manifolds here are of even dimension, that is $\dim(S_{Y_0}) = q = 2p$. $p = \bar{q}$.

Definition: Shift Invariance for Semi-Unitary Matrices: Define the function

$$X_k(Y) \equiv (J_k Y \Phi_k - J_0 Y) \quad (687)$$

where

$$\Phi_k(Y) \equiv (J_k Y)^\# J_0 Y \quad (688)$$

and $J_k, J_0 \in R^{N(sel) \times N}$ are the selection matrices that select the $N(sel)$ elements from the N elements. A $N \times p$ semi-unitary matrix Y is said to be shift invariant with respect to the pair of selection matrices J_k and J_0 if

$$X_k(Y) = 0 . \quad (689)$$

The relation (689) is often referred to in the literature as the invariance equation [1].

Lemma 25. *If Y is shift-invariant with respect to J_0 and J_k then YQ is also shift invariant with respect to this same pair.*

Proof: (appendix C.5)

The following lemma expresses the invariance condition $X_k(Y) = 0$ in a reduced form.

Lemma 26. *Let $Z = \text{orth}(J_0 Y)$, and $Z_\perp$ its orthogonal complement, and define*

$$R_k(Y) \equiv Z_\perp^H X_k(Y) . \quad (690)$$

Then

$$R_k(Y) = 0 \iff X_k(Y) = 0 . \quad (691)$$

Proof: The columns of $[ZZ_\perp]$ span the range space of X_k so that

$$X_k(Y) = P_Z X_k(Y) + P_{Z_\perp} X_k(Y) . \quad (692)$$

It follows that

$$X_k(Y) = 0 \iff \begin{bmatrix} Z^H X_k(Y) \\ R_k(Y) \end{bmatrix} = 0 . \quad (693)$$

By definition all $Y \in \bar{\mathcal{B}}$ satisfies $X_k(Y) = 0$. Furthermore if $Y \in \bar{\mathcal{B}}$ then $YQ \in \bar{\mathcal{B}}$ and therefore if Y satisfies $X_k(Y) = 0$ then $X_k(YQ) = 0$ for Q unitary. If $R_k(Y) \neq 0$ then $Y \notin \bar{\mathcal{B}}$. $\dot{Z} = \text{orth}(J_0 YQ)$.

Figure 8. Selection Matrices for 2-D Planar Arrays

For the 1-dimensional ULA case a single pair of selection matrices are sufficient to specify $\bar{\mathcal{B}}$. For M -dimensional array case, $M > 1$, additional pairs are required and a single constraint $R_k(Y) = 0$ is a necessary, but not sufficient, condition for $Y \in \bar{\mathcal{B}}$. The required set of selection matrices depends on the array geometry (e.g.; rectangular array versus corner clipped array). In what follows we assume a base array defined by the selection matrix J_0 and consider shift pairs (J_0, J_1) , (J_0, J_2) , etc. on rectangular arrays (see figure 8). Defining the vectorized form of the component constraint as $r_k(Y) \equiv \text{vec}(R_k(Y))$, the full constraint takes the form

$$r(Y) \equiv \begin{bmatrix} r_1(Y) \\ \vdots \\ r_{\bar{M}}(Y) \end{bmatrix}. \quad (694)$$

Note that each $r_k(Y)$ is a complex $\bar{N}_b$ -tuple so that $r(Y)$ is $\bar{M} \cdot \bar{N}_b \times 1$. For an appropriately chosen a set of $\bar{M}+1$ selection matrices the elements of $\bar{\mathcal{B}}_{damped}$ satisfy

$$Y \in \mathcal{B}_{damped} \iff r(Y) = 0 \quad (695)$$

or equivalently, in terms of the exponential map at $\tilde{Y}$,

$$\exp_{\tilde{Y}} \Delta \in \mathcal{B}_{damped} \iff r(\exp_{\tilde{Y}} \Delta) = 0. \quad (696)$$

Approximating the constraint

A key idea in developing the approximation for (694) is incorporating the complex dimension of the submanifold tangent space S_Y . For p signals on an M -dimensional array, the complex dimension is $\dim(S_Y) = \bar{q} = Mp$. Earlier we showed that in the general case, the set of constraint-satisfying tangent vectors had the form $b = b_n + \bar{E}_s \bar{s}$ (655) where $\bar{s}$ was a free complex $\bar{q}$ -tuple and $\bar{E}_s$ was an $N_b \times \bar{q}$ complex matrix which represented a basis for S_Y .

We next develop an approximation for the submanifold component membership function (690). As a first step in the linearization, we record the expanded form of the function $R_k(\cdot)$ (690) when the argument is $Y = \exp_{\tilde{Y}} \tilde{\Delta}$. Substituting the expanded form of the exponential map $\exp_{\tilde{Y}} \tilde{\Delta} = \tilde{Y} + \tilde{\Delta} + h.o.t$ into (690) yields

$$R_k(\exp_{\tilde{Y}} \tilde{\Delta}) \equiv R_k(\tilde{Y}) - M_{0k}(\tilde{Y}; \tilde{\Delta}) - M_{1k}(\tilde{Y}; \tilde{\Delta}) \quad (697)$$

where

$$M_{0k}(\tilde{Y}; \tilde{\Delta}) = Z_{\perp}^H (J_k \tilde{\Delta} \tilde{\Phi}_k - J_0 \tilde{\Delta}) \quad (698)$$

and

$$M_{1k}(\tilde{Y}; \tilde{\Delta}) = Z_{\perp}^H J_k Y \delta \Phi_k(\tilde{Y}, \tilde{\Delta}) + h.o.t. \quad (699)$$

In the above we have used

$$\Phi_k(\exp_{\tilde{Y}} \Delta) = \tilde{\Phi}_k + \delta \Phi_k(\tilde{Y}, \Delta) \quad (700)$$

and where $\tilde{\Phi}_k \equiv \Phi_k(\tilde{Y})$ and

$$\delta \Phi_k = (Y^H J_k^t J_k Y)^{-1} J_0 \Delta = (\tilde{Y}^H J_k^t J_k \tilde{Y})^{-1} J_0 \tilde{Y}_{\perp} B + h.o.t. \quad (701)$$

The tangent vectors of interest are elements of $B_{\tilde{Y}} < T_{\tilde{Y}}\tilde{\mathcal{B}}_{damped}$, which have the general form $\tilde{\Delta} = \tilde{Y}_{\perp} B$. Substituting this form into (698) yields

$$M_{0k}(\tilde{Y}; \tilde{\Delta}) = (Z_{\perp}^H J_k \cdot \tilde{Y}_{\perp} B \cdot \tilde{\Phi}_k - Z_{\perp}^H J_0 \cdot \tilde{Y}_{\perp} B) . \quad (702)$$

Vectorizing (702) and (699) using the $vec(ABC) = (C^t \otimes A)b$ identity yields the vectorized form for (697) as

$$r_k(exp_{\tilde{Y}} \Delta) = r_k(\tilde{Y}) - \tilde{H}_k b - C_k vec(\delta \tilde{\Phi}_k(b)) + h.o.t \quad (703)$$

where

$$\tilde{H}_k = (\tilde{\Phi}_k^t \otimes Z_{\perp}^H J_k \tilde{Y}_{\perp}) - (I \otimes Z_{\perp}^H J_0 \tilde{Y}_{\perp}) \quad (704)$$

and

$$C_k = (I_p \otimes Z_{\perp}^H J_k \tilde{Y}) \quad (705)$$

and where $r_k(\cdot) \equiv vec(R_k(\cdot))$ and $\delta \tilde{\Phi}_k(b) \equiv \delta \tilde{\Phi}_k(Y_{\perp} B)$. Note that H_k and C_k are $\bar{N}_b \times \bar{N}_b$.

To arrive at the desired form we note that, at high SNR, small $\|b\|$, that

$$\|C_k vec(\delta \tilde{\Phi}_k(b))\|_F \ll \|H_k b\|_F . \quad (706)$$

Dropping this term and the higher order terms in (703) results in the approximation

$$r_k(exp_{\tilde{Y}} \Delta) \approx r_k(\tilde{Y}) - \tilde{H}_k b . \quad (707)$$

From these component function linearizations (707) we approximate the full constraint (694) as

$$r(Y) \approx r(\tilde{Y}) - \tilde{H}(\tilde{Y})b \quad (708)$$

where $\tilde{H}(\tilde{Y})$ is

$$\tilde{H}(\tilde{Y}) = \begin{bmatrix} \tilde{H}_1(\tilde{Y}) \\ \vdots \\ \tilde{H}_{\bar{M}}(\tilde{Y}) \end{bmatrix} . \quad (709)$$

The matrix $\tilde{H}(\tilde{Y})$ is approximately rank $N - \bar{q}$. Motivated by the dimension of the constraint submanifold tangent space S_Y with complex dimension $\bar{q}$, we replace $\tilde{H}(\tilde{Y})$ by its rank $N - \bar{q}$ approximation, denoted $\widehat{H}$. This defines the matrix $\widehat{H}$ in (677) and (679) and using (659) through (671) completes the subspace estimation algorithm, which is summarized in the next section.

4.4 *MLESI* Algorithm Summary

1 Compute EVD of the SCM in partitioned form (p -dominant part)

$$\tilde{R} = \tilde{Y}_s \tilde{\Lambda}_s \tilde{Y}_s^H + \tilde{Y}_o \tilde{\Lambda}_o \tilde{Y}_o^H \quad (710)$$

where $\tilde{Y} = \tilde{Y}_s$ is $N \times p$ and

$$\hat{\sigma}^2 \equiv \frac{1}{N-p} \text{tr}(\tilde{\Lambda}_o) \quad (711)$$

$$\tilde{\Lambda}_{OPT} \equiv (\tilde{\Lambda}_s - \hat{\sigma}^2 I)^2 \tilde{\Lambda}_s^{-1} \quad (712)$$

$$\tilde{\Lambda}_{bb} \equiv \tilde{\Lambda}_{OPT} \otimes I_{N-p} . \quad (713)$$

2 Compute Component functions

For $k = 1$ to M compute

$$\tilde{\phi}_{k0} = (J_k \tilde{Y})^\# J_0 \tilde{Y} \quad (714)$$

$$r_{k0}(\tilde{Y}) = \text{vec}(J_k \tilde{Y} \tilde{\phi}_{k0} - J_0 \tilde{Y}) \quad (715)$$

$$\tilde{H}_{k0} = (\tilde{\phi}_{k0}^t \otimes J_k \tilde{Y}_\perp) - (I \otimes J_0 \tilde{Y}_\perp) . \quad (716)$$

3 Form $r(\tilde{Y})$ and the matrix $\tilde{H}$

$$r(\tilde{Y}) = \begin{bmatrix} r_{10}(\tilde{Y}) \\ \vdots \\ r_{m0}(\tilde{Y}) \end{bmatrix} , \quad \tilde{H} = \begin{bmatrix} H_{10}(\tilde{Y}) \\ \vdots \\ H_{m0}(\tilde{Y}) \end{bmatrix} . \quad (717)$$

4 Compute the rank $\rho = p(N-p) - 2Mp$ approximation of $\tilde{H}$

Compute the full SVD of $\tilde{H}$ as $\tilde{H} = U_1 D_1 V_1^H + U_0 D_0 V_0^H$ and its rank ρ approximation

$$\hat{H} \equiv U_1 D_1 V_1^H . \quad (718)$$

The $\bar{E}_s$ matrix is $\bar{E}_s \equiv V_0$.

4 Compute the least squares solution

$$\hat{b}_n = \hat{H}^\# r(\tilde{Y}) , \quad (719)$$

$$\bar{L}_{ss} \equiv \bar{E}_s^H \tilde{\Lambda}_{bb} \bar{E}_s , \quad (720)$$

$$\bar{L}_s \equiv \bar{E}_s^H \tilde{\Lambda}_{bb} b_n , \quad (721)$$

$$\bar{s} \equiv \bar{L}_{ss}^{-1} \bar{L}_s , \quad (722)$$

$$\hat{b} = b_n + \bar{E}_s \hat{s} . \quad (723)$$

5 Compute Final Estimate

$$\hat{Y} = \tilde{Y} + \tilde{Y}_\perp \text{mat}(\hat{b}) . \quad (724)$$

Example:2-Dimensional Regular Grid

For the 2-dimensional uniform rectangular (URA) let N_1 and N_2 denote the number of elements in the two array dimensions (row and column) so that $N = N_1 \cdot N_2$, is the total number of elements. A signal may be represented by the Kronecker product $h^{(h)} \otimes \left(h^{(v)}\right)^t$. For the multiple signal case the k th signal is

$$h_k^{(h)} \otimes \left(h_k^{(v)}\right)^t \quad (725)$$

where $h_k^{(r)}$ is the $N_k \times 1$ complex vector

$$h_k^{(r)} = (\rho_k^{(r)} \exp(i\mu_k^{(r)}))^{(0:N_k-1)} , \quad r = h, v \quad (726)$$

Let $J_{h;0}$, $J_{h;1}$ and $J_{v;0}$, $J_{v;1}$ by the selection matrices for the two array dimensions; that is,

$$J_{h;0}h^{(h)}z_h = J_{h;1}h^{(h)} \quad (727)$$

and

$$J_{v;0}h^{(v)}z_v = J_{v;1}h^{(v)} . \quad (728)$$

This holds for the different permutations. First consider

$$J_{h;0}h^{(h)}z_h \otimes (J_{v;0}h^{(v)}z_v)^t = J_{h;1}h^{(h)} \otimes (J_{v;1}h^{(v)})^t \quad (729)$$

The factor on the right hand side is

$$= J_{h;0}(h^{(1)} \otimes (h^{(2)})^t)J_{h;0}^t \quad (730)$$

¹ Vectorizing this equation using the identity $vec(ABC) = (C^t \otimes A)vec(B)$, observe that the right hand side becomes

$$= (J_{v;0} \otimes J_{h;0})vec(h^{(1)} \otimes (h^{(2)})^t) \quad (732)$$

$$= J_1h \quad (733)$$

where $h \equiv vec(h^{(1)} \otimes (h^{(2)})^t)$ and $J_1 \equiv (J_{h;1} \otimes J_{v;1})$. The left hand side of (730) factors in the same way yielding

$$J_0hz_h \cdot z_v = J_3h \quad (734)$$

where $J_3 \equiv (J_{h;1} \otimes J_{v;1})$. In a similar way $J_0hz_h = J_1h$ and $J_0hz_v = J_2h$ where the complete set of selection matrices is

$$\begin{bmatrix} J_0 = J_{h;0} \otimes J_{v;0} \\ J_1 = J_{h;1} \otimes J_{v;0} \\ J_2 = J_{h;0} \otimes J_{v;1} \\ J_3 = J_{h;1} \otimes J_{v;1} \end{bmatrix} . \quad (735)$$

¹

$$(J_{h;0} \otimes 1)(h^{(1)} \otimes (h^{(2)})^t)(1 \otimes J_{v;0}^t) = J_{h;0}(h^{(1)} \otimes (h^{(2)})^t)J_{h;0}^t \quad (731)$$

Organizing these into matrix form yields

$$J_0 \begin{bmatrix} h_1 \cdots h_p \end{bmatrix} \cdot \Psi_h = J_1 \begin{bmatrix} h_1 \cdots h_p \end{bmatrix}, \quad (736)$$

$$J_0 \begin{bmatrix} h_1 \cdots h_p \end{bmatrix} \cdot \Psi_v = J_2 \begin{bmatrix} h_1 \cdots h_p \end{bmatrix}, \quad (737)$$

$$J_0 \begin{bmatrix} h_1 \cdots h_p \end{bmatrix} \cdot \Psi_h \Psi_v = J_3 \begin{bmatrix} h_1 \cdots h_p \end{bmatrix}. \quad (738)$$

where $\Psi_v = \text{diag}(z_{v;l}, \dots, z_{v;p})$, $\Psi_h = \text{diag}(z_{h;l}, \dots, z_{h;p})$ and where $h_k \equiv \text{vec}(h_k^{(1)} \otimes (h_k^{(2)})^t)$.

EXAMPLE: 3×3 regular planar array

Consider a 2-two dimensional regular planar array $M = 2$ with $N_1 = 3, N_2 = 3, N = 9$. For $p = 2$, complex dimension $\dim(S_Y) = M \cdot p = 4$. The row/column selection matrices are

$$J_{h;0} = [110]; J_{h;1} = [011]; J_{v;0} = [110]; J_{v;1} = [011]; \quad (739)$$

so that $J_1 = J_{h;0} \otimes J_{h;0}^t$ is 4×9 (the full selection matrices for $N_1 = 5, N_2 = 5$ case are as indicated in figure 8).

The complex dimension of the full space is $\bar{N}_b = p(N - p) = 14$, so that each H_k (716) is 14×14 and the overall H matrix (717) is 52×14 (where $p \cdot \bar{N}_b = 52$). Since $\dim(S_Y) = 4$ we require a rank $52 - 4 = 48$ approximation $\hat{H}$ of $\tilde{H}$ (717).

List of References

- [1] S. Smith, "Covariance, subspace, and intrinsic cramer-rao bounds," *IEEE Trans. Signal Process.*, vol. 53, no. 5, p. 1610 1630, 2005.

CHAPTER 5

Summary and Simulation Results

We briefly summarize the previous developments before presenting numerical simulation results demonstrating the various bounds and estimators. In chapter 3 we considered the standard stochastic complex linear model in the form

$$z = Yn_1 + n_0; \quad Y \in \mathcal{B} \quad (740)$$

where Y was an unknown deterministic $N \times p$ semi-unitary matrix, viewed as a point in the Basis manifold $\mathcal{B}$, (a quotient manifold of the complex Stiefel manifold).

The data appeared in the log likelihood function $L(Y, \Lambda; \tilde{R})$ as the SCM, $\tilde{R} = K^{-1}ZZ^H$ where Z is the $N \times K$ data matrix $Z = \begin{bmatrix} z_1 & \cdots & z_N \end{bmatrix}$. The eigen decomposition of the SCM is

$$\tilde{R} = \tilde{Y}\tilde{\Lambda}_s\tilde{Y}^H + \tilde{Y}_\perp\tilde{\Lambda}_o\tilde{Y}_\perp^H \quad (741)$$

where $\tilde{Y}$ is the $N \times p$ semi-unitary matrix corresponding to the p largest eigenvalues. The parameter $Y \in \mathcal{B}$ appears in the compressed model covariance given by

$$R = R(Y; \tilde{\Lambda}, \tilde{\sigma}^2) = Y\tilde{\Lambda}Y^H + \tilde{\sigma}^2I \quad (742)$$

where $\tilde{\sigma}^2$, defined in equation (583), and $\tilde{\Lambda} = \tilde{\Lambda}_s - \tilde{\sigma}^2I$ correspond to the ML estimates of Λ and σ , respectively.

In this geometric formulation our objective was to establish intrinsic CRBs for the accuracy of unbiased estimators of $Y \in \mathcal{B}$, or equivalently, bounds on subspace and rotational estimation accuracy.

5.1 Unconstrained Problem: $Y \in \mathcal{B}$

The subspace distance error measure was defined, independently of the differential geometric formulation, as the square root of the sum of squares of the subspace angles

γ , between the generic estimator $\hat{Y}$ and the true Y

$$\text{distSS}(\hat{Y}, Y) \equiv \|\gamma\| . \quad (743)$$

In MATLAB γ is computed in radians as

$$\text{gamma} = (\text{acos}(\text{svd}(\text{orth}(Y1)' \text{orth}(Y2)))) . \quad (744)$$

The mean square error used to evaluate subspace estimation accuracy was

$$\epsilon_{ss}^2 \equiv E[\text{distSS}^2(\hat{Y}, Y)] . \quad (745)$$

We showed that in the geometric Basis manifold formulation, this subspace distance corresponds to a minimum point-to-fiber submanifold geodesic distance (see figure 6, section 2.3.1). The tangent spaces of $\mathcal{B}$ are naturally partitioned as $T_Y \mathcal{B} = A_Y \oplus B_Y$, with B_Y the space of subspace perturbations and A_Y the space of rotations. We showed that, if P_A and P_B are the orthogonal projection operators from $T_Y \mathcal{B}$ to A_Y and B_Y respectively, then, to first order, $\text{distSS}(\hat{Y}, Y) = \|P_B \hat{\Delta}\|$, where $\hat{\Delta}$ is the estimation error vector $\hat{\Delta} = \exp_{Y_0}^{-1} \hat{Y}$. Given an orthonormal basis $\mathbf{a}_k, \mathbf{b}_k$, for A_{Y_0} and B_{Y_0} respectively, an arbitrary tangent vector Δ expands as

$$\Delta(\alpha, \beta) = \sum_{k=1}^{N_a} \mathbf{a}_k \alpha^k + \sum_{k=1}^{N_b} \mathbf{b}_k \beta^k \quad (746)$$

and the parameters (α, β) are said to be the normal coordinates of the point $\hat{Y} \in \mathcal{B}$. We used this change of variables to express the log likelihood function as $L(\alpha, \beta) \equiv L(Y(\alpha, \beta))$ and to compute the full FIM for this coordinate choice in theorem 10. Since $\|P_B \Delta\| = \|\beta\|$, and therefore

$$\epsilon_{ss}^2 \equiv E[\|\hat{\beta}\|^2] , \quad (747)$$

we developed the CRB on subspace estimation accuracy by computing FIM in this normal coordinate parameterization (see section A.1). In particular, we showed that the subspace block of the full FIM was

$$F_{\beta\beta} = \frac{2K}{\sigma^2} \begin{bmatrix} (\Lambda_{OPT} \otimes I_{N-p}) & 0 \\ 0 & (\Lambda_{OPT} \otimes I_{N-p}) \end{bmatrix} \quad (748)$$

where

$$\Lambda_{OPT} = \Lambda^2 \Lambda_s^{-1} = \Lambda^2 (\Lambda + \sigma^2 I)^{-1} = (\Lambda_s - \sigma^2 I)^2 \Lambda_s^{-1} \quad (749)$$

and that the off diagonal submatrices, $(F_{\alpha\beta}$, etc.), vanished. From this we showed that the intrinsic lower bound on p -dimensional subspace estimation accuracy, CRB_{SS} , was

$$CRB_{SS} \equiv \sqrt{\text{tr}(F_{\beta\beta}^{-1})} = \sqrt{\frac{2(N-p)\sigma^2}{K} \sum_{k=1}^p \frac{\lambda_k + \sigma^2}{(\lambda_k)^2}} \quad (750)$$

so that $\epsilon_{ss} \geq CRB_{SS}$. In the simplest case of orthogonal signals with equal signal powers, we showed that $CRB_{SS} \approx \sqrt{\frac{2p(N-p)}{K \cdot SNR}}$ rad where $SNR = \frac{\sigma_s^2}{\sigma^2}$ and σ_s^2 is the power of each source.

The ML estimate $\hat{Y}$ of $Y \in \mathcal{B}$ is given by the truncated EVD of the SCM (741), that is $\hat{Y} = \tilde{Y}$. Denoting the error for this particular estimator as ϵ_{ss}^{svd} , asymptotic efficiency in this context implies that

$$\epsilon_{ss}^{svd} \rightarrow CRB_{SS} \quad (751)$$

as the number of independent snapshots $K \rightarrow \infty$. This is confirmed in the numerical simulations that follow in section 5.1.1.

The mean square error used to evaluate rotational estimation accuracy was

$$\epsilon_R^2 \equiv E[\text{dist}R^2(\hat{Y}, Y)] \quad (752)$$

where $\text{dist}R^2(\hat{Y}, Y) = \|\mu\|^2$ and μ is the p -tuple of rotational angles between $\hat{Y}$ and Y . Using the same normal coordinate parameterization, we showed that the intrinsic lower bound on rotational estimation accuracy was

$$CRB_R \equiv \sqrt{\text{tr}(F_{\alpha\alpha}^{-1})} = \sqrt{2 \sum_{i=1}^p \sum_{j=1; j \neq i}^p \frac{(\lambda_i - \lambda_j)^2}{\lambda_i \lambda_j}}, \quad (753)$$

so that $\epsilon_R \geq CRB_R$. Denoting the rotational error in the EVD estimate as ϵ_R^{svd} we have

$$\epsilon_R^{svd} \rightarrow CRB_R \quad (754)$$

as $K \rightarrow \infty$; this is confirmed in the numerical simulations that follow. These rotational parameters are considered nuisance parameter in the usual subspace signal processing application, but are required to properly parameterize the likelihood function.

5.1.1 Simulation Results

In chapter 4 we showed that the EVD of the SCM was the ML estimator of $Y \in \mathcal{B}$ for the statistical model (740). In this section the performance of the EVD estimator is compared to the intrinsic CRBs for subspace and rotational accuracy established in chapter 3. A Monte Carlo simulation based on the statistical model (740) for $N = 5$ and $p = 2$ is considered. We note that for this unconstrained case, both CRB_{SS} (750) and CRB_R (753) are independent of Y . In the standard simulation approach Y would be fixed with n_1 and n_0 drawn from random distributions. Because the relevant CRBs are independent of Y , in the simulation approach used here, Y is also chosen randomly from the uniform distribution on $\mathcal{B}_{5,2}$. In MATLAB this random semi-unitary matrix is generated as

$$Y = \text{orth}(\text{randn}(N, p) + i \text{randn}(N, p)); \quad (755)$$

The random source vector, n_1 , is drawn from $CN(0, \sigma_s^2 I_p)$ and the additive noise, n_0 , is drawn from $CN(0, \sigma^2 I_N)$. The input SNR is then defined as $SNR = \frac{\sigma_s^2}{\sigma^2}$. Once σ^2, σ_s^2, Y , and the number of independent snapshots K are specified, CRB_{SS} is computed using (750). Note that in the constrained cases considered in the next sections, that is $Y \in \bar{\mathcal{B}} \subset \mathcal{B}$, the subspace accuracy bound *does* depend on the choice of Y , and so Y must be fixed in the later simulation examples.

One thousand ($N_{trials} = 1000$) Monte Carlo trials are performed, each of which consists of generating a normal $N \times K$ data matrix $Z = \begin{bmatrix} z_1 & \cdots & z_N \end{bmatrix}$ and computing the SCM $\tilde{R} = K^{-1} Z Z^H$. Using the MATLAB “eig” function, the estimate $\hat{Y}$ is formed as the p dominant eigenvectors vectors, $\tilde{Y}$, of $\tilde{R}$ (741). The sample RMSE subspace distance, denoted $\hat{\epsilon}_{SS}$, is then computed over the N_{trials} simulation trials as

$$\hat{\epsilon}_{SS} \equiv \sqrt{\frac{1}{N_{trials}} \sum_{n=1}^{N_{trials}} (dist_{SS}(Y, \hat{Y}_n)^2)} \quad (756)$$

where $\hat{Y}_n$ represents the estimate for the n -th trial. Likewise, the sample RMSE rotational distance is

$$\hat{\epsilon}_R \equiv \sqrt{\frac{1}{N_{trials}} \sum_{n=1}^{N_{trials}} (dist_R(Y, \hat{Y}_n)^2)} . \quad (757)$$

The results comparing the accuracy of the EVD estimator to the CRB_{SS} are shown in Figs. 9 and 10 as functions of the SNR and sample support. In figure 9, the number of samples is fixed at $K=10$, and the SNR is varied. For this small sample support case, the EVD-based estimator achieves an accuracy that is a small constant fraction above CRB_{SS} (750). In figure 10 the SNR is fixed and the number of snapshots is varied. As anticipated, EVD-based estimator closely approaches CRB_{SS} as the number of snapshots, K , becomes large.

The intrinsic lower bound on rotational estimation accuracy, CRB_R (753), is a function of the difference between eigenvalues, but does *not* depend on the particular choice of $Y \in \mathcal{B}$. The EVD rotational estimation performance, measured as (757), is assessed using the same general Monte Carlo simulation approach used to assess subspace estimation performance. In the first test case, the additive noise power σ^2 and the larger signal eigenvalue λ_1 are fixed and smaller of the two signal eigenvalues is varied as $\lambda_2 = c \cdot \lambda_1$, $c \leq 1$. Results for this test case are shown in figure 11 as a family of curves, with each curve representing a different SNR (i.e., noise power). The curves in the lower portion of the figure are high SNR, while the curves in the upper portion are low SNR. As the figure indicates, for fixed SNR the larger the power difference between the two sources (power ratio, c , is small), the better we are able to estimate the rotation. In the second test case λ_1 , λ_2 and σ^2 are fixed and the sample support is varied. Results for this case, shown in figure 12, demonstrate that the EVD-based estimate of rotation asymptotically approaches the lower bound, i.e., the EVD is an asymptotically efficient

estimator of rotation.

5.2 Constrained Problem: $Y \in \bar{\mathcal{B}} \subset \mathcal{B}$

In many problems of interest the set of $N \times p$ semi-unitary matrices, appearing in the model (740), are constrained to a submanifold $\bar{\mathcal{B}} \subset \mathcal{B}$. In this constrained case the matrix of p -dominant eigenvectors $\tilde{Y}$ of the SCM is no longer the MLE of Y , and $\text{col}(\tilde{Y})$ is no longer the subspace MLE.

The log likelihood function for this constrained model is equal to the log likelihood function of the unconstrained model with a restriction on the domain. At every point $Y \in \bar{\mathcal{B}} \subset \mathcal{B}$ we have $T_Y \bar{\mathcal{B}} \subset T_Y \mathcal{B}$, and for the submanifolds of interest $T_Y \bar{\mathcal{B}}$ decomposes into a rotational space and subspace perturbation space in a similar way as full space ($T_Y \mathcal{B} = A_Y \oplus B_Y$), that is

$$T_Y \bar{\mathcal{B}} = A_Y \oplus S_Y . \quad (758)$$

The rotational space A_Y is identical in both decompositions and $S_Y \subset B_Y$, so that B_Y , the space of subspace perturbations for the full tangent space $T_Y \mathcal{B}$, can be decomposed as

$$B_Y = S_Y \oplus (S_Y)^\perp = S_Y \oplus N_Y . \quad (759)$$

Motion in the S_Y direction corresponds to subspace perturbations that remain on the submanifold, while motion in N_Y corresponds to subspace perturbations that exit the submanifold in a normal direction. A basis $\{\mathbf{s}_k\}$ and $\{\mathbf{n}_k\}$ for S_Y and N_Y , respectively, forms a basis for B_Y termed here an S-N basis.

The FIM for the constrained case is computed in same way as in unconstrained case. The subspace block of full FIM (471) for the general constrained problem is

$$F_{ss} = \frac{2K}{\sigma^2} \text{Re}(E_s^H (\Lambda_{OPT} \otimes I_{N-p}) E_s) \quad (760)$$

where E_s is defined in terms of a basis on S_Y , $\mathbf{s}_k = Y_\perp S_k$, as

$$E_s = \begin{bmatrix} \text{vec}(S_1) & \cdots & \text{vec}(S_q) \end{bmatrix} \quad (761)$$

Figure 9. RMSEs of the EVD-based subspace estimator and the CRB of (435) versus SNR for the estimation problem of (740) on $\mathcal{B}$. The RMSE of the EVD estimate is the the redline; CRB_{SS} is the solid (gray) curve. The RMSEs of the EVD-based subspace estimators are a small constant fraction above the subspace CRB over all SNRs shown. Results are shown for $K = 10$ and $K = 128$.

Figure 10. RMSEs of EVD-based subspace estimator and CRB_{SS} (750) versus sample support for the estimation problem of (740) on $\mathcal{B}_{5,2}$. CRB_{SS} is the solid (gray) curve; the subspace RMSE values appear as red dots. Results are shown for three SNRs, 1 dB, 7 dB and 13 dB. As expected, the RMSE of the EVD estimate approaches the subspace CRB as the sample support becomes large, i.e., this maximum likelihood estimator is asymptotically efficient.

Figure 11. RMSE of the EVD-based rotation estimator and the rotational CRB (753) versus the source power ratio c for the estimation problem of (740) on $\mathcal{B}_{5,2}$. The additive noise power σ^2 and the larger signal eigenvalue λ_1 are fixed and smaller of the two signal eigenvalues is varied as $\lambda_2 = c \cdot \lambda_1$. The family of curves with each curve represent a different noise power value. The curves in the lower portion of the figure are high SNR, while the curves in the upper portion are low SNR. The number of snapshots is fixed at $K = 10$.

Figure 12. RMSEs of EVD-based subspace estimator and CRB_{SS} (750) versus sample support for the estimation problem of (740) on $\mathcal{B}_{5,2}$. CRB_{SS} is the solid (gray) curve; subspace RMSE values appear as red dots. The signal eigenvalues are fixed as $\lambda_2 = 0.5\lambda_1$. As expected, the RMSE of the EVD estimate approaches the rotational CRB as the sample support becomes large, i.e., this maximum likelihood estimator is asymptotically efficient.

for an appropriately chosen set $(N - p) \times p$ matrices S_k . We use the notation crb_{SS} to denote the subspace CRB for a general constrained submanifold assumption (the subspace CRB for the unconstrained assumption is denoted CRB_{SS}). Ignoring curvature, the bound on subspace estimation accuracy is

$$crb_{SS} \equiv \sqrt{\text{tr}(F_{ss}^{-1})} . \quad (762)$$

Parametric Constraint General and Modal

When the signal matrix is defined by some underlying parameterization, $H = H(\theta)$, the eigen decomposition of the true signal covariance effects a transformation between the (θ, P) and (Y, Λ) parametrizations

$$[Y, \Lambda] = \text{eig}(H(\theta)PH^H(\theta)) . \quad (763)$$

By definition, the eigenvector matrix Y is an element of $Y \in \bar{\mathcal{B}} \subset \mathcal{B}$. Both parameterization serve as coordinate systems on the product manifold $\bar{\mathcal{B}} \times D$, the domain of the log likelihood function.

The (θ, P) parameterization of the problem has a total of $q + p^2$ independent parameters. The dimension of $\bar{\mathcal{B}}$ is equal to the dimension of the tangent space $T_Y \bar{\mathcal{B}} = A_Y \oplus S_Y$, and therefore the dimension of product $\bar{\mathcal{B}} \times D$ is

$$\dim(S_Y) + \dim(A_Y) + \dim(\Lambda) = q + p(p - 1) + p = q + p^2 . \quad (764)$$

Note this value is consistent with the (θ, P) parametrization.

Changes in the source correlation P induce rotations corresponding to motion in A_Y direction, as well as changes in the eigenvalue space, but no motion in S_Y and therefore no change in subspace. Changes in θ result in motion confined to the subspace S_Y direction along with changes in the eigenvalue space.

As discussed in section 3.9.4, the differentiable signal model $H(\theta)$ defines a natural coordinate basis for S_{Y_0} , as $\acute{s}_k = Y_{\perp} S_k$ where

$$S_k = Y_{\perp}^H D_k T; \quad k = 1 \cdots q \quad (765)$$

and where D_k is the derivative matrix of H with respect to θ_k , evaluated at the point of interest. In the modal case, $H = [h(\theta_1), \cdots h(\theta_q)]$ and the E_s basis matrix (761) has the form (520)

$$E_s = (T^t \otimes Y_{\perp}^H \cdot D)G \quad (766)$$

where D is the $N \times q$ derivative matrix and G is a selection matrix.

We showed that substituting this form into general form (760) and manipulating terms yielded (513), repeated here:

$$F_{\dot{s}\dot{s}} = \frac{\sigma^2}{2K} \cdot \text{Re}((D^H P_{\perp} D) \odot P_s^{(M)}) . \quad (767)$$

The right hand side is recognized as the parameter space FIM for the multi-dimensional modal problem [1],[2]. The natural coordinate basis, in general non-orthogonal, is brought to orthonormal form by the linear transformation

$$\begin{bmatrix} \mathbf{s}_1 & \cdots & \mathbf{s}_q \end{bmatrix} = \begin{bmatrix} \dot{\mathbf{s}}_1 & \cdots & \dot{\mathbf{s}}_q \end{bmatrix} \mathbf{J} \quad (768)$$

where $\mathbf{J}$ is computed in the standard vector space way (see (511)). The components of the subspace FIM with respect to this orthonormal basis are (513)

$$F_{ss} = \mathbf{J} \cdot F_{\dot{s}\dot{s}} \cdot \mathbf{J}^t . \quad (769)$$

The CRB on ϵ_{ss} for this parametrically constrained problem, then, is

$$crb_{ss} = \sqrt{\frac{\sigma^2}{2K} \text{tr}(\text{Re}((D^H P_{\perp} D) \odot P_s^{(M)})^{-1} \cdot \mathbf{J}^{-2})} . \quad (770)$$

5.2.1 Asymptotic ML Estimators for $Y \in \bar{\mathcal{B}}$

In this geometric approach, our objective is to estimate the $N \times p$ semi-unitary matrix Y representing a point on the submanifold $\bar{\mathcal{B}}$ defined through the signal covariance

R_s by (763). From the geometric perspective, the underlying parameterization provides a particular coordinate system on the submanifold. Since the parameterization also defines the submanifold, these parameters are termed here the natural coordinates.

For the unconstrained case, the p principal eigenvectors and eigenvalues of the SCM as computed by EVD and denoted here as $(\tilde{Y}, \tilde{\Lambda})$, respectively, are maximum likelihood estimates of $Y \in \mathcal{B}$ and $\Lambda \in D$. We consider the compressed likelihood function (CLF), $L(Y) \equiv L(Y; \tilde{\Lambda})$ when Y is constrained to a submanifold $\bar{\mathcal{B}} \subset \mathcal{B}$. In this case the ML criterion is

$$\hat{Y} = \max_{Y \in \bar{\mathcal{B}}} L(Y) . \quad (771)$$

Using a partitioned (α, β) normal coordinate system with origin at the unconstrained MLE $\tilde{Y}$, nearby points may be represented as $Y = \exp_{\tilde{Y}} \Delta(\alpha, \beta)$ (see (746)). The constrained ML criterion (771) in terms of these normal coordinates is

$$(\hat{\alpha}, \hat{\beta}) = \max(L(\alpha, \beta)) \quad \text{subject to } \exp_{\tilde{Y}} \Delta(\alpha, \beta) \in \bar{\mathcal{B}} . \quad (772)$$

In this expression $\exp_{\tilde{Y}} \Delta(\alpha, \beta) \in \bar{\mathcal{B}}$ represents the submanifold membership constraint condition, $Y \in \bar{\mathcal{B}}$, expressed in terms of the exponential map based at $\tilde{Y}$ and the tangent vector $\Delta(\alpha, \beta)$.

The Taylor series the expansion of $L(\alpha, \beta)$ is given by

$$L(\alpha, \beta) = L(0, 0) + \alpha^t L_{\alpha\alpha} \alpha + \beta^t L_{\beta\beta} \beta + h.o.t \quad (773)$$

where we have used the fact that $L_{\alpha\beta} = 0$ and that the first derivatives in α, β vanish at the unconstrained MLE, $\tilde{Y}$.

As a consequence of the assumption that source correlation matrix P is unconstrained, the set $\bar{\mathcal{B}}$ is such that if $Y \in \bar{\mathcal{B}}$ then $YQ \in \bar{\mathcal{B}}$, Q unitary and arbitrary. Thus the parameter α , which defines Q through $Q = \expm \left(\sum \mathbf{a}_k \alpha^k \right) I_{N,p}$, is likewise unconstrained. Since in (773) both $L_{\alpha\alpha}$ and $L_{\beta\beta}$ are positive definite we conclude that $\hat{\alpha} = 0$.

The ML criterion (772) thus is reduced to one in β alone. Here, β is the real component N_b -tuple of the vector $\Delta \in B_Y$ with respect to the standard basis $\mathbf{b}_k$. Such tangent vectors have the general form $\Delta = \sum \mathbf{b}_k \beta^k = \tilde{Y}_\perp B$. The parameter β can alternatively be expressed in complex form $b = \begin{bmatrix} I & iI \end{bmatrix} \beta$ so that $B = \text{mat}(b)$. Using this, the quadratic form of the constrained MLE criterion (772) reduces to

$$\hat{b} = \min_b \text{Re}(b^H \tilde{\Lambda}_{bb} b) \text{ subject to } b \in \bar{\mathcal{V}}_{\tilde{Y}} \quad (774)$$

(see (637)) where $\bar{\mathcal{V}}_{\tilde{Y}}$ is the set of complex $\bar{N}_b$ -tuples that satisfy the submanifold membership condition, that is

$$\bar{\mathcal{V}}_{\tilde{Y}} \equiv \{b \in C^{\bar{N}_b} \text{ such that } \exp_{\tilde{Y}} \Delta \in \bar{\mathcal{B}} \text{ where } \Delta = \tilde{Y}_\perp \text{mat}(b)\} . \quad (775)$$

Using an S-N basis for space $B_{\tilde{Y}}$, we showed that these constraint-satisfying complex $\bar{N}_b$ -tuples have the form

$$b = b_n + E_s s . \quad (776)$$

In this expression, E_s is the basis matrix for $S_{\tilde{Y}}$, s is a free real q -tuple, and b_n defines the tangent vector from $\tilde{Y}$ to the closet point on $\bar{\mathcal{B}}$. Denoting this closet point, or minimum subspace distance point, as Y_{MD} , we have

$$Y_{MD} = \exp_{\tilde{Y}} \Delta \in \bar{\mathcal{B}} , \quad \Delta = \tilde{Y}_\perp \text{mat}(b_n) . \quad (777)$$

The second term in the sum (776), $E_s s$, corresponds to free subspace motion along the submanifold.

Substituting (776) into (774) results in an unconstrained minimization in s which has the solution

$$\hat{s} = L_{ss}^{-1} L_s |_{Y_{MD}} \quad (778)$$

where L_{ss} and $L_s |_{Y_{MD}}$ are given by (659) and (660) respectively. The solution to this quadratic approximation of (771) is then expressed as an $N \times p$ semi-unitary matrix as

$$\hat{Y} = \exp_{\tilde{Y}} \hat{\Delta}; \quad \hat{\Delta} = \tilde{Y}_\perp \text{mat}(\hat{b}) \quad (779)$$

where

$$\hat{b} = b_n + E_s \hat{s} . \quad (780)$$

This estimation approach requires that the submanifold membership condition, expressed as (776), be known. This, in turn, requires that the S-N partitioning of the space $B_{\hat{Y}} = S_{\hat{Y}} \oplus N_{\hat{Y}}$ is known and that the minimum distance point Y_{MD} is also known.

Relation to Newton's Method

When the submanifold $\bar{\mathcal{B}}$ is defined by some known underlying parameterization, a practical approach to solving the above constrained ML formulation is to leverage a suboptimal parameter estimate, say $\hat{\theta}_0$, to establish a reference point $Y_0 \in \bar{\mathcal{B}}$. Using the differentiable signal model, a basis is constructed for S_{Y_0} expressed as the matrix $E_{\hat{s}}$.

The ML tangent vector based at Y_0 then is given by

$$\Delta = Y_{0;\perp} \hat{B}; \quad \hat{B} = \text{mat}(\hat{b}) \quad (781)$$

where

$$\hat{b} = E_{\hat{s}} \widehat{\delta\theta} \quad (782)$$

and

$$\widehat{\delta\theta} = L_{\hat{s}\hat{s}}^{-1} L_{\hat{s}}|_{Y_0} . \quad (783)$$

The resultant ML estimate of $Y \in \bar{\mathcal{B}}$ is $\hat{Y} = \exp_{Y_0} \hat{\Delta}$. Comparing (782) with (780), note that the term $b_n = E_n \hat{n}$ is absent from (782). This corresponds to fact that here, the reference point Y_0 lies on the desired submanifold $\bar{\mathcal{B}}$, so no movement in the normal direction, represented by the term b_n in (780), is required.

The ML tangent vector based at Y_0 is alternately expressed as

$$\hat{\Delta} = \sum_{k=1}^{N_s} \hat{s}_k \widehat{\delta\theta}^k . \quad (784)$$

The component estimate $\widehat{\delta\theta}$ represents a change in the underlying θ and the equivalent parameter estimate is

$$\hat{\theta} = \hat{\theta}_0 + \widehat{\delta\theta} . \quad (785)$$

This approach (783) is recognized as the a 1-step Newton's refinement if we make the following correspondences

$$[L_{\dot{s}}]_i = \frac{\partial L}{\partial \theta^i} |_{\hat{\theta}_0} , \quad (786)$$

$$[L_{\dot{s}\dot{s}}]_{ij} \approx \frac{\partial^2 L}{\partial \theta^i \partial \theta^j} |_{\hat{\theta}_0} . \quad (787)$$

The overall performance of this approach relies on the performance of the unspecified estimator that provides the initial parameter estimate $\hat{\theta}_0$. For large initial errors, multiple iterations may be required and convergence is not guaranteed. For harmonic estimation problems on regular arrays, R-D Unitary ESPRIT may be used to generate the initial $\hat{\theta}_0$ estimate.

ML Estimation on $\bar{\mathcal{B}}_{damped}$

The above algorithm for the constrained optimization problem relies on a suboptimal parameter estimator to establish the reference point Y_0 that initializes Newton's method. In section 4.3.1 we introduced a new algorithm for the important special case of the damped exponential signal model on regular arrays. The submanifold corresponding to this signal model is denoted $\bar{\mathcal{B}}_{damped} \subset \mathcal{B}$. This new algorithm used the structure of $\bar{\mathcal{B}}_{damped}$ in a more integrated and optimal way and did not require a suboptimal reference solution. Given this model, we showed that, for an appropriately defined vector valued function $r(\cdot)$ on $\bar{\mathcal{B}}_{damped}$, $Y \in \bar{\mathcal{B}}_{damped} \iff r(Y) = \mathbf{0}$. The submanifold membership condition in terms of the tangent vector $\Delta \in B_{\tilde{Y}}$ based at $\tilde{Y}$ then is

$$\exp_{\tilde{Y}} \Delta \in \bar{\mathcal{B}}_{damped} \iff r(b; \tilde{Y}) = \mathbf{0} \quad (788)$$

where $r(b; \tilde{Y}) \equiv r(\exp_{\tilde{Y}} \Delta)$. The function notation $r(b; \tilde{Y})$ is based on the fact that elements of $B_{\tilde{Y}}$ have the general form $\Delta = \tilde{Y}_{\perp} B$ for a fixed $\tilde{Y}_{\perp}$ and $b = \text{vec}(B)$.

The Taylor series expansion of the constraint function $r(b; \tilde{Y})$ is

$$r(b; Y) = r(0; \tilde{Y}) + \sum_{k=1}^{\bar{N}_b} \frac{\partial r}{\partial b^k} b^k + h.o.t. \quad (789)$$

The first order approximation of the constraint function, then, is expressed in matrix form as

$$r(b; Y) \approx r(\tilde{Y}) - \widehat{H}b, \quad (790)$$

where for convenience we define the matrix $\widehat{H}$ as

$$[\widehat{H}]_k^m \equiv -\frac{\partial r^m}{\partial b^k} \quad (791)$$

($\widehat{H}$ is a function of $\tilde{Y}_\perp$). A critical feature of this approximation is that $\widehat{H}$ is constructed in such a way that its null space well approximates the submanifold tangent space $S_{\tilde{Y}}$

$$\text{null}(\widehat{H}(\tilde{Y})) \approx S_{\tilde{Y}}. \quad (792)$$

If $\hat{E}_s$ denotes the matrix whose columns are a basis for $\text{null}(\widehat{H})$, then notionally

$$\text{col}(\hat{E}_s) \approx \text{col}(\bar{E}_s), \quad (793)$$

where $\bar{E}_s$ is a basis for $S_{\tilde{Y}}$. Since the space $S_{\tilde{Y}}$ is $\bar{q}$ dimensional, this requires that the null space of the approximation $\widehat{H}$ be $\bar{q}$ dimensional as well, or $\text{rank}(\widehat{H}) = \bar{N}_b - \bar{q}$.

Using this approximation for $\widehat{H}$, and (677) the set of constraint-satisfying tangent vectors $\bar{\mathcal{V}}_{\tilde{Y}}$ (775) is approximated as

$$\bar{\mathcal{V}}_{\tilde{Y}} \approx \{b \in C^{N_b} \text{ such that } b = \min_b \|r(\tilde{Y}) - \widehat{H}b\|^2\} \quad (794)$$

Elements in this set are given by the $\bar{q}$ -parameter family of least squares solutions

$$b = \hat{b}_n + \hat{E}_s \bar{s} \quad (795)$$

where

$$b_n = \hat{H}^\# r(\tilde{Y}) \quad (796)$$

and $\bar{s}$ is a complex free $\bar{q}$ -tuple. This solution is in the same form as (776), the result developed using a reference point Y_0 .

Given the constraint-satisfying tangent vector specified by (680), the final estimate $\hat{Y}$ of $Y \in \bar{\mathcal{B}}_{damped}$ is found using the complex form of (778) (see theorem 24)

$$\hat{s} \equiv \bar{L}_{ss}^{-1} \bar{L}_s \quad (797)$$

where $\bar{L}_{ss} \equiv c \cdot \bar{E}_s^H \tilde{\Lambda}_{bb} \bar{E}_s$ and $\bar{L}_s \equiv c \cdot \bar{E}_s^H \tilde{\Lambda}_{bb} b_n$. The final estimate of $Y \in \bar{\mathcal{B}}$ then is

$$\hat{Y} = \exp_{\tilde{Y}} \hat{\Delta}; \quad \hat{\Delta} = \tilde{Y}_{\perp} \text{mat}(\hat{b}) \quad (798)$$

where

$$\hat{b} = \hat{b}_n + \hat{E}_s \cdot \hat{s} \quad (799)$$

The estimate (798) may be approximated to first order as $\tilde{Y} + \hat{\Delta}$.

5.2.2 Application to the harmonic retrieval problem

Multi-dimensional harmonic retrieval on regular arrays appears in a number of different applications. In the geometric approach, this estimation problem is framed in terms of the submanifold corresponding to harmonic signal model, denoted $\bar{\mathcal{B}}_{harmonic}$, with the set relationship

$$\bar{\mathcal{B}}_{harmonic} \subset \bar{\mathcal{B}}_{damped} \subset \mathcal{B} . \quad (800)$$

The eigen decomposition of the true signal covariance effects a transformation between the (θ, P) and (Y, Λ) parametrizations, that is $[Y, \Lambda] = \text{eig}(H(\theta)PH^H(\theta))$. The inverse transformation has the general form

$$(\theta, P) = \Phi^{-1}(Y, \Lambda) = (\varphi(Y), \varphi(Y, \Lambda)) \quad (801)$$

and, in particular,

$$\theta = \varphi(Y); \quad Y \in \bar{\mathcal{B}}_{harmonic} \subset \mathcal{B} . \quad (802)$$

For harmonic signals on regular array geometries R-D Unitary ESPRIT [3], restricted to $\bar{\mathcal{B}}_{harmonic}$, is the mapping φ .

If $\hat{Y}$ is an ML estimator of $Y \in \bar{\mathcal{B}} \subset \bar{\mathcal{B}}_{\text{harmonic}}$, then by the invariance principle [4], $\hat{\theta} = \varphi(\hat{Y})$ is an ML estimator of $\theta \in R^p$. In the standard approach, the subspace estimate $\tilde{Y}$ is the p -dominant singular vector matrix of the SCM as computed by the EVD. This is the ML estimate for the unconstrained signal model corresponding to the $\mathcal{B}$ manifold. It is *not* ML with respect to the $\bar{\mathcal{B}}_{\text{harmonic}} \subset \mathcal{B}$ assumption, and so the invariance principle is not applicable here.

A subspace-based parameter estimation algorithm is formed by extending the domain of the mapping φ (802) to $\mathcal{B}$; as noted, R-D Unitary ESPRIT with $\tilde{Y}$ as input is the prime example of such an approach. Since $\tilde{Y}$ is not ML with respect to the $\bar{\mathcal{B}}_{\text{harmonic}}$ assumption R-D Unitary ESPRIT is not an ML parameter estimator. To improve subspace-based parameter estimation, it is therefore desirable to find an ML estimator of $Y \in \bar{\mathcal{B}} \subset \bar{\mathcal{B}}_{\text{harmonic}}$, or failing that, one that is closer to the constrained subspace CRB than standard EVD estimate $\tilde{Y}$. In chapter 4 we introduced a subspace estimation algorithm for the damped exponential signal assumption, $\bar{\mathcal{B}}_{\text{damped}}$ that was derived from the ML criterion. This estimation algorithm, termed *MLESI*, may be viewed as an implementation of a geometric Newton's method on $\mathcal{B}$ that avoids the need for explicit initialization by incorporating the submanifold constraint condition (i.e., shift-invariance) in a more integrated way. In the most benign scenario the subspace accuracy bound for the $\bar{\mathcal{B}}_{\text{damped}}$ assumption is $10 \log 10((N - p))$ below that of the unconstrained $\mathcal{B}$ assumption. For the highly correlated signal examples considered in later, this difference is much larger. We show via simulation in section 5.2.5, that this subspace estimator is, as expected, asymptotically efficient with respect to the intrinsic subspace CRB on $\bar{\mathcal{B}}_{\text{damped}}$ for a variety of stressful scenarios in which other methods are severely degraded.

This ML estimator of $Y \in \bar{\mathcal{B}}_{\text{damped}}$ is applied to the problem of estimating the θ parameter for the $\bar{\mathcal{B}}_{\text{harmonic}}$ assumption. A parameter estimation package is generated

by using the *MLESI* estimate $\hat{Y}$ to replace the standard EVD estimate, $\tilde{Y}$, as input to R-D Unitary ESPRIT. Since the subspace CRB for $\bar{\mathcal{B}}_{damped}$, CRB_{damped} , is only 3 dB above the subspace CRB of $\bar{\mathcal{B}}_{harmonic}$, $CRB_{harmonic}$, the *MLESI* estimate, $\hat{Y}$, has a much smaller subspace error variance compared to the SVD estimate $\tilde{Y}$. Improved parameter estimation results from this reduced subspace error.

Ratio of the Bounds

The ratio of bounds for the different signal or submanifold assumptions indicate the processing gain available to a subspace estimator using the full constraint information, relative to the standard EVD. The potential gain processing is defined as

$$\text{Gain:} = 20 \log 10 \left(\frac{CRB_{SS}}{crb_{SS}} \right) \quad (803)$$

where CRB_{SS} is the bound based on the unconstrained assumption and crb_{SS} is the bound computed based on the constraint information (e.g., for harmonic signal model $crb_{SS} = CRB_{harmonic}$).

For benign scenarios consisting of widely spaced uncorrelated signals, the relative difference between the bounds is the square root of the dimension ratio of the space of subspace perturbations tangent space for different manifold assumptions. The dimension of these spaces for the different assumptions and the resultant ratios are summarized in the table below.

Table 3. Dimension of subspace perturbation spaces for different signal assumptions (M =array dimension, p =number of modes)

Manifold	Subspace	dim(Subspace)	Ratio expressed in dB
$\mathcal{B}$	B_Y	$2p(N - p)$	NA
$\bar{\mathcal{B}}_{damped}$	$S_{damped;Y}$	$2Mp$	$10 \log 10 \left(\frac{N-p}{M} \right)$
$\bar{\mathcal{B}}_{harmonic}$	$S_{harmonic;Y}$	Mp	$10 \log 10 \left(\frac{2(N-p)}{M} \right)$

Thus, for benign scenarios on a 2-dimensional array, ($M = 2$), potential gains for the $\bar{\mathcal{B}}_{damped}$ and $\bar{\mathcal{B}}_{harmonic}$ assumptions are respectively, $10 \log 10 \left(\frac{N-p}{2} \right)$ and

$10 \log 10(N-p)$. The *MLESI* estimator, which is asymptotically efficient with respect to the $\bar{\mathcal{B}}_{damped}$ assumption, realizes a gain of $10 \log 10(\frac{N-p}{2})$ in this benign example.

5.2.3 Example: Correlated vs. Uncorrelated

The case of two highly correlated signals is often used to compare the performance of different estimators, since in this stressful situation, standard approaches like ESPRIT are severely degraded [3]. To better understand the mechanisms at play, we consider two orthonormal signals h_1 , and h_2 and the signal matrix $H = [h_1 \ h_2]$, where

$$h_1^H h_2 = 0, \quad h_1^H h_1 = h_2^H h_2 = 1. \quad (804)$$

The SVD of the signal matrix in this case is simply $H := \sqrt{N}Y_0$ (i.e., singular values equal), so that $R_s = HPH^H = N(Y_0PY_0^H)$ and

$$R = R_s + \sigma^2 I = N(Y_0PY_0^H) + \sigma^2 I. \quad (805)$$

Consider the two source correlation matrices corresponding to correlated and uncorrelated sources:

$$P_{corr} \equiv \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}, \quad P_{uncorr} \equiv \begin{bmatrix} 1+\rho & 0 \\ 0 & 1-\rho \end{bmatrix}. \quad (806)$$

P_{corr} has the eigen decomposition $P_{corr} = Q\Lambda Q^H$ where

$$\Lambda = \begin{bmatrix} 1+\rho & 0 \\ 0 & 1-\rho \end{bmatrix} \quad (807)$$

and

$$Q = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + \frac{1}{\sqrt{2}} \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}. \quad (808)$$

The signal covariance matrices for correlated and uncorrelated cases are given, respectively, by

$$R_{s,corr} \equiv (Y_0Q)\Lambda(Y_0Q)^H \quad (809)$$

and

$$R_{s,uncorr} \equiv Y_0\Lambda Y_0^H. \quad (810)$$

Thus, for the correlated case (806) the parameter ρ represents the correlation coefficient, while in the uncorrelated case $1 + \rho$ and $1 - \rho$ are the powers of the two sources. By construction, the signal covariance matrices have identical eigenvalues, that is $\sigma(R_{uncorr}) = \sigma(R_{corr})$. In a Grassmann manifold formulation, the eigenvector matrices Y_0 and $Y_0 Q$, that define the same subspace belong to the same equivalence class, and are therefore represented by the same point (as such, the (Grassmann) geodesic distance between them is zero). In the Basis manifold formulation used here, these two matrices belong to different equivalence classes and are therefore represented by distinct points; the geodesic distance between them is non-zero.

The bounds for the subspace estimation problem depend on constraint assumptions of the signal matrix, H which transform into constraints on $Y \in \mathcal{B}$. For the unconstrained case, the $\mathcal{B}$ manifold applies and the subspace CRB is given by (750). For the two signal case, $p = 2$, and high SNR this bound is

$$CRB_{SS} \approx \sqrt{\frac{(N-p)\sigma^2}{K} \sum_{k=1}^p \lambda_k^{-1}} = \sqrt{\frac{2(N-2)}{K} \left(\frac{1}{1-\rho^2} \right) \sigma^2}. \quad (811)$$

This CRB depends on the problem's eigenvalues but not on Y and so CRB_{SS} for both the $R_{s,corr}$ and $R_{s,uncorr}$ cases, is identical.

If, on the other hand, H is constrained, then the $\bar{\mathcal{B}} \subset \mathcal{B}$ submanifold formulation applies and $crb_{SS} = \sqrt{\text{tr}(F_{ss}^{-1})}$ where the subspace FIM, F_{ss} , has the general form (760). If $E_{s;u}$ is the basis matrix for the uncorrelated case, then the corresponding matrix for the correlated case is $E_{s;c} = (Q \otimes I_{N-p})E_{s;u}$ where Q is given by (808). The ratio of unconstrained bound to the constrained bound measures the potential processing gain available to an estimator using the full constraint information so that

$$\text{Gain:} = \frac{CRB_{ss}}{crb_{ss}} = \sqrt{\left(\frac{\text{tr}(F_{\beta\beta}^{-1})}{\text{tr}(F_{ss}^{-1})} \right)}. \quad (812)$$

Example: For harmonic signals the $\bar{\mathcal{B}}_{harmonic}$ submanifold assumption applies. The

subspace CRB for the $p = 2$ uncorrelated case, P_{uncorr} (806), is

$$crb_{ss}^{(u)} \approx \sqrt{\frac{8}{K} \left(\frac{1}{1 - \rho^2} \right)} \sigma^2 \quad (813)$$

and for the correlated case, P_{corr} (806),

$$crb_{ss}^{(c)} \approx \sqrt{\frac{8}{K}} \sigma^2. \quad (814)$$

The potential processing gains (812) for these two cases are

$$\text{Gain Correlated:} = \frac{CRB_{ss}}{crb_{ss}^{(c)}} = \sqrt{\frac{N - 2}{2(1 - \rho^2)}}, \quad (815)$$

and

$$\text{Gain Uncorrelated:} = \frac{CRB_{ss}}{crb_{ss}^{(u)}} = \sqrt{\frac{N - 2}{2}}. \quad (816)$$

For the uncorrelated case, we see that the gain is equal to the ratio of dimensions of the spaces B_Y and $S_Y < B_Y$

$$\frac{\dim(B_Y)}{\dim(S_Y)} = \frac{2p(N - p)}{4p} = \frac{N - 2}{2}. \quad (817)$$

For the example considered later of a the 2-D 5×5 square array with $\rho = 0.9999$, the potential gains in the two cases are: $47.1 \text{ dB} = 10 \log_{10} \left(\frac{25-2}{2(1-0.9999^2)} \right)$ for the correlated case and $13 \text{ dB} = 10 \log_{10} \left(\frac{25-2}{2} \right)$ for the corresponding uncorrelated case.

Because both Y_0 and $Y_0 Q$ map to the same parameter value θ , the distinction is not relevant in the harmonic parameter extraction step (802). The rotation Q (or equivalently the parameter α) is a nuisance parameter with respect to the θ parameter estimation objective. However, as the difference between the CRBs for the correlated and uncorrelated problems indicate, the distinction between Y_0 and $Y_0 Q$ is relevant to the overall estimation problem.

The ratio between the intrinsic bounds for the different constraints, uncorrelated and correlated, is

$$\frac{crb_{ss}^{(u)}}{crb_{ss}^{(c)}} = \sqrt{\frac{1}{1 - \rho^2}} \quad (818)$$

For $\rho = 0.9999$ as in the example above this ratio is 37 dB. This dependence on the parameter ρ is reflected in the θ parameter space CRB. The parameter space FIM at high SNR is given by

$$F_{\theta\theta} = \frac{2K}{\sigma^2} \text{Re}(D^H P_{\perp} D \odot P) \quad (819)$$

and the bound on the parameter error variance is

$$CRB_{\theta\theta} = \sqrt{\frac{\sigma^2}{2K} \text{tr}(\text{Re}(D^H P_{\perp} D \odot P))^{-1}}. \quad (820)$$

For the orthogonal signal example under consideration $D^H P_{\perp} D = I_p$, so that the CRB on the parameter total RMS estimation error is reduces to

$$CRB_{\theta\theta} = \sqrt{\frac{\sigma^2}{2K} \text{tr}(I_p \odot P)^{-1}}. \quad (821)$$

The parameter CRBs for the uncorrelated and correlated cases (806) are, respectively,

$$CRB_{\theta\theta}^{(u)} = \sqrt{\frac{\sigma^2}{2K} \left(\frac{1}{1-\rho} + \frac{1}{1+\rho} \right)} = \sqrt{\frac{\sigma^2}{K} \frac{1}{1-\rho^2}} \quad (822)$$

and

$$CRB_{\theta\theta}^{(c)} = \sqrt{\frac{\sigma^2}{K}}. \quad (823)$$

The ratio between the constrained bounds for the uncorrelated and correlated cases is

$$\frac{CRB_{\theta\theta}^{(u)}}{CRB_{\theta\theta}^{(c)}} = \sqrt{\frac{1}{1-\rho^2}}. \quad (824)$$

As expected, this result is consistent with the subspace bound ratio given by (818).

5.2.4 Performance Metrics

The performance evaluation of the various estimators is based on two related criteria. First, for the estimation packages that provide estimates of the parameter θ directly, we use the total RMS parameter estimation error. For the harmonic signal estimation problem on M -dimensional arrays (e.g., DOA estimation), the total RMSE in the spatial frequency domain, denoted $\hat{\epsilon}_{\theta}$, is

$$\hat{\epsilon}_{\theta} \equiv \sqrt{\frac{1}{N_{\text{trials}}} \sum_{n=1}^{N_{\text{trials}}} \delta\theta(n)^t \delta\theta(n)} = \sqrt{\frac{1}{N_{\text{trials}}} \sum_{n=1}^{N_{\text{trials}}} \|\delta\theta(n)\|^2} \quad (825)$$

where $\delta\theta(n) = \hat{\theta}(n) - \theta$ denotes the error vector in the estimate for the n -th trial. For the M -dimensional harmonic case with p modes, $\delta\theta(n)$ is $Mp \times 1$.

The second criterion is the subspace distance between the true signal subspace, represented as $Y(\theta)$, and the estimate $\hat{Y} \in \mathcal{B}$. The RMSE is

$$\hat{\epsilon}_{SS} \equiv \sqrt{\frac{1}{N_{trials}} \sum_{n=1}^{N_{trials}} distSS(Y, \hat{Y}(n))^2}. \quad (826)$$

In this expression $\hat{Y}(n)$ is an $N \times p$ matrix that represents the estimate for the n -th trial.

The EVD/SVD and *MLESI* estimator provide estimates in the form of $N \times p$ semi-unitary matrices that define $\hat{Y}$ appearing above. For the parameter estimators such as Unitary ESPRIT, the reconstructed subspace estimate is defined by column space of the left singular vector matrix of $H(\hat{\theta})$, that is $col(\hat{Y})$ where $\hat{Y} = Y(\hat{\theta})$.

Relationship between parameter errors and subspace angles

The relationship between these two error measures, $distSS^2(\hat{Y}, Y(\theta))$ and $\delta\theta^t \delta\theta$, is developed in this section. The basic result we wish to establish is

$$distSS^2(\hat{Y}, Y(\theta)) \approx \delta\theta^t \mathbf{J}^{-2} \delta\theta; \quad \hat{Y} = Y(\hat{\theta}). \quad (827)$$

The error tangent vector $\Delta = exp_{\hat{Y}}^{-1} Y(\hat{\theta})$ is in $T_{\hat{Y}} \bar{\mathcal{B}} < T_{\hat{Y}} \mathcal{B}$ and the subspace component is $\bar{\Delta} \equiv P_B \Delta \in S_Y$. This vector expands as

$$\bar{\Delta} = \sum_{k=1}^q \mathbf{s}_k s^k = \sum_{k=1}^q \acute{\mathbf{s}}_k \acute{s}^k. \quad (828)$$

In this expression $\acute{\mathbf{s}}_k$ denotes the natural coordinate (non-orthogonal) basis and $\mathbf{s}_k$ is an orthonormal basis. If $\mathbf{J}$ denotes the transformation matrix between $\mathbf{s}_k$ and $\acute{\mathbf{s}}_k$, ($\mathbf{J}$ orthogonalizes $\mathbf{s}_k$) then $s = \mathbf{J}^{-1} \acute{s}$ (see (505)) and as a result

$$\|\bar{\Delta}\| = \sqrt{s^t s} = \sqrt{\acute{s}^t \mathbf{J}^{-2} \acute{s}}. \quad (829)$$

The development of the subspace estimation accuracy bound for the Basis manifold formulation relied on the fact that, to first order

$$distSS(Y(\theta_0), \hat{Y})^2 = \|\bar{\Delta}\|^2. \quad (830)$$

Figure 13. Scatter plot of the subspace distance error computed directly via (826) and indirectly using (831). The scatter lies along a line with unit slope confirming equivalence of the two formulas.

In addition, since to first order $\delta\theta \approx \acute{\theta}$, substituting this result into (829) and using (830) yields

$$\text{dist}SS^2(Y(\hat{\theta}), Y(\theta)) \approx \delta\theta^t \mathbf{J}^{-2} \delta\theta . \quad (831)$$

To demonstrate this relation we show a scatter plot, figure 13, of the left and right hand sides of (831) over 1000 trials for a stressful example. The scatter lies along a straight line confirming equivalence of the two formulas.

From (831) we see that the relationship of the RSS subspace error to the RSS parameter space error $(\delta\theta^t \delta\theta)$ is determined by $\mathbf{J}$. The relationship between $\hat{\epsilon}_{SS}$ and

$\hat{\epsilon}_\theta$ (825) is illustrated with scatter plots of these two measures in section 5.2.5. In the benign scenarios considered the eigenvalues of $\mathbf{J}$ are all approximately equal, ($\sigma(\mathbf{J}) \approx c$), so that the two measures differ by a scale factor, that is

$$distSS(Y(\theta_0), \hat{Y}) \approx \cdot c^{-1} \sqrt{p} \|\delta\theta\| \quad (832)$$

In contrast, for closely spaced sources the eigenvalues of $\mathbf{J}$ vary widely and this simple relation does not hold. This situation is illustrated in examples 5 and 6 considered in section 5.2.5.

LIST OF ESTIMATORS

In this section we review variants of the estimation packages that are evaluated in the simulation test cases of section 5.2.5.

UE Unitary ESPRIT/Standalone, ($\hat{\theta}_{ESP}$) [5]: 2-D Unitary ESPRIT, and its extension to higher dimensions (R-D Unitary ESPRIT), has the highly desirable feature of automatically associating the distinct DOA parameters that define a mode (e.g., automatic paring of bearing and elevation estimates in the 2-D array case). This estimator does not satisfy any ML criterion and is highly biased for large spatial angles at low SNR. As discussed in section 5.2.2, Unitary ESPRIT may also be viewed as coordinate transformation (802) when its domain is restricted to $\bar{\mathcal{B}}_{harmonic}$. An important practical advantage of UE estimator is that it has low computational cost and minimal storage requirements.

UEN Unitary ESPRIT with Newton's Refinement, ($\hat{\theta}_{ESPN}$): This estimation package consists 2-D Unitary ESPRIT followed by the a step Newton's method with respect to the log likelihood function. The Newton's step uses the EVD estimates, $(\tilde{Y}, \tilde{\Lambda})$, as inputs. Performance depends on the quality of UE parameter estimate; for large initial errors multiple iterations may be required and convergence is not guaranteed. Since it is developed to an ML criterion, it is asymptotically efficient contingent on convergence to the global maximum.

MLESI: This estimator is asymptotically efficient with respect to the damped exponential signal assumption, $\bar{\mathcal{B}}_{damped}$. The estimate $\hat{Y} \in \bar{\mathcal{B}}_{damped}$ corresponds to a set of damping and angle parameters, which in this context are viewed as an alternate coordinate system on $\bar{\mathcal{B}}_{damped}$. See section 4.4 for the algorithm summary.

MLESIESP: This estimation package uses the *MLESI* estimate $\hat{Y}$ as input to the UE parameter estimator in place of the EVD estimate $\tilde{Y}$. The improved parameter estimation performance stems from smaller subspace error variance in the *MLESI* estimate, $\hat{Y}$, relative to that of the EVD estimate, $\tilde{Y}$. This fact is important when $\mathcal{B}$ subspace CRB for the unconstrained case (achieved by the EVD) is much above $\mathcal{B}_{harmonic}$ bound. This is the situation that arises when considering highly correlated signals. The proposed estimator *MLESIESP* is observed to be unbiased facilitating the analysis.

MLESIESPN This estimator extends the previous estimation package with the addition of a Newton's refinement step.

Other Estimators: The standard version of ESPRIT (SE) is intended for the $\mathcal{B}_{damped}$ case while Unitary ESPRIT (UE) is intended for the $\mathcal{B}_{harmonic}$ case. The tensor version of these algorithms, Standard Tensor ESPRIT (STE) and Unitary Tensor ESRPIRT (UE), are likewise intended for these different signal assumptions [3],[3].

5.2.5 Simulation Results

In this section we compare the performance of the above estimation packages to the intrinsic CRBs for the 3 different signal assumptions (800), and show numerical results to demonstrate the improved performance of the proposed estimators (*MLESI* variants). For relevance, we consider the standard narrowband direction of arrival (DOA) estimation problem; in the geometric formulation this corresponds to $\bar{\mathcal{B}}_{harmonic}$ submanifold assumption. The first example introduces the estimator variants in a benign test scenario to fix the performance analysis approach and to illustrate the relationships between the different subspace bounds. These benign examples are followed

by more stressful scenarios that show the features of the different estimators and the dependence of the bounds on system parameters (e.g., sample support). We consider a 2-dimensional, 8×8 , uniform rectangular array (URA) with $\frac{1}{2}$ wavelength element spacing in both directions in all the examples.

Example 1: 2 widely spaced, equal power, uncorrelated sources

This case consists of two equal power uncorrelated sources with the spatial frequencies specified in table 4. The $N \times p$ matrix $Y(\theta_0)$, ($N = 8^2 = 64, p = 2$) is defined

Table 4. Example 1: Spatial frequencies, 2 Sources

ID	horizontal	vertical
Source 1	1.0	-1.0
Source 2	-1.0	1.0

by the eigen decomposition of the signal covariance $R_s = H(\theta)PH(\theta)^H$ and represents a point on $\mathcal{B}_{harmonic}$. The source correlation matrix is $P = I$ and the number of independent snapshots is $K = 10$. The eigenvalues of signal covariance R_s and the Jacobian $\mathbf{J}$ are, respectively,

$$\sigma(R_s) = [62.62, 65.38] \quad (833)$$

$$\sigma(\mathbf{J}) = [2.2, 2.2, 2.2, 2.2] . \quad (834)$$

Note that for $P = I$, $\sum(\sigma(R_s)) = Np = 64 \cdot 2 = 128$.

Using a Monte-Carlo simulation, a $N \times K$ data matrix is generated according to (740) and provided as input to each of the estimators for each trial. We measure the performance of the various estimators using the RMS subspace distance $\hat{\epsilon}_{SS}$ (826) and the total RMS parameter error $\hat{\epsilon}_\theta$ (825), where appropriate, over a set of 1000 realizations for fixed SNR. The subspace distance CRBs for the 3 signal models (800): harmonic, damped exponential, and unconstrained are plotted versus SNR in figure 14 along with the $\hat{\epsilon}_{SS}$ for the *EVD*, *MLESI* and *MLESIESP* subspace estimators. This set of plots is referred to as the subspace error plot set. Following the discussion

Figure 14. The RMSEs $\hat{e}_{SS}$ for the *EVD*, *MLESI* and *MLESIESP* subspace estimators and the CRBs for the 3 manifold assumptions vs SNR for the estimation problem on the complex linear model (740). The spatial frequencies of the two, equal power, uncorrelated sources are specified in table:examp1 and the number of snapshots is ten ($K=10$). The *EVD* subspace estimate remains a small increment above $CRB_{SS}S$ for this small snapshot case. The *MLESI* and *MLESIESP* estimators are observed to approach their respective bounds, CRB_{damped} and $CRB_{harmonic}$, at sufficient SNR.

in section 5.2.2, for this benign scenario the difference in the bounds may be attributed to the difference in the dimension of the space of subspace perturbations for each of the 3 manifold assumptions. Referring to table 3, with $N = 64$ and $p = 2$, we compute these differences (relative to the unconstrained bound CRB_{SS}) as 14 dB and 17 dB for the damped and harmonic assumptions, respectively. The *EVD* subspace estimate remains a small increment above the bound for the unconstrained signal (full $\mathcal{B}$ manifold) assumption, $CRB_{SS}S$, for this 10 snapshot ($K = 10$) example (the effect of varying the number of snapshots is considered in example 2). The *MLESI* estimator closely approaches the CRB for the $\mathcal{B}_{damped}$ manifold assumption, CRB_{damped} , as SNR

increases. Likewise, the *MLESIESP* estimator closely approaches the CRB for the $\mathcal{B}_{harmonic}$ manifold assumption, $CRB_{harmonic}$, as SNR increases. For the parameter estimator *MLESIESP*, an estimate in the form of an $N \times p$ semi-unitary matrix element of $\bar{\mathcal{B}}_{harmonic}$ is computed from the parameter estimates $\hat{\theta}$ as the left singular vector matrix of $H(\hat{\theta})$. This same procedure is used to generate subspace estimates for each of the direct parameter estimation packages.

The parameter estimation packages are also evaluated directly using the total RMS parameter error measure $\hat{\epsilon}_{\theta}$ (825). The parameter space bound $CRB_{\theta\theta}$ (820) is plotted versus SNR in figure 15 along with $\hat{\epsilon}_{\theta}$ for a set of parameter estimation packages. The *MLESIESP* estimator achieves the parameter space bound at sufficient SNR. The UE algorithm is approximately 2 dB above the bound while the UEN estimator achieves the bound as expected.

Example 2: Vary Sample Support (2 Sources, Equal Power)

In this scenario the source locations are fixed as in example 1, the SNR is fixed at 21 dB, and the sample support size (i.e., number of snapshots, K) is varied. The subspace distance CRBs for the 3 signal models are plotted versus sample support in figure 16 along with the $\hat{\epsilon}_{SS}$ for the *EVD*, *MLESI* and *MLESIESP* subspace estimators. As the sample support increases the EVD estimate and *MLESI* meet their respective bounds, CRB_{SS} and CRB_{damped} . This is consistent with the asymptotically efficient property of ML estimators. The estimator *MLESIESP* similarly approaches $CRB_{harmonic}$.

Example 3: Varying Angular Separation, 2 Sources, Equal Power

In this test scenario the sample size and SNR are fixed and the angular separation between the two sources is varied. The estimator performance and the relevant bounds are shown in figure 17. If the spatial frequencies given in table 4 are denoted as μ_0 , the

Figure 15. RMSE $\hat{\epsilon}_\theta$ (825) for the *MLESIESP* and *MLESIESPN*, *UE* and *UEN* estimators and the parameter space bound $CRB_{\theta\theta}$ (820) versus SNR. The *MLESIESP* estimator achieves the parameter space bound at sufficient SNR. The *UE* algorithm is approximately 3 dB above the bound while the *UEN* estimator achieves the bound as expected.

Figure 16. RMSEs $\hat{\epsilon}_{SS}$ of the EVD, $MLESI$, $MLESIESP$ subspace estimators and the CRBs for the 3 manifold assumptions versus sample support for the estimation problem of (740). The spatial frequencies are as given in table 4. A constant SNR of 21 dB and 1000 Monte Carlo trials are used. Note that, as expected, the RMSE of the EVD estimate (green line, open squares) approaches the unconstrained subspace bound, CRB_{SS} (dashed gray line) as the sample support becomes large, i.e., this maximum likelihood estimator is asymptotically efficient. Similarly, the $MLESI$ estimator (red line, open circles) approaches the $CRBD$ (dashed black line) as the sample support becomes large.

Figure 17. RMSEs $\hat{\epsilon}_{SS}$ of the EVD, $MLESI$, $MLESI_{ESP}$ subspace estimators and the CRBs for the 3 manifold assumptions versus source separation factor for the estimation problem of (740). If the spatial frequencies given in table 4 are denoted as μ_0 , the spatial frequencies in this example are $s \cdot \mu_0$; the scale factor s is the horizontal axis labeled, "sep factor". Observe that the proposed estimator $MLESI$ shows good agreement with the bound CRB_{damped} down to vary small angular source separation.

spatial frequencies in this example are $s \cdot \mu_0$; the scale factor s is the horizontal axis in the figure . Observe that the estimator $MLESI$ shows good agreement with the bound CRB_{damped} down to vary small angular source separation.

Example 4 Correlated vs Uncorrelated

This scenario examines the highly correlated signal case discussed in section 5.2.3. Examples like this one are often used in the literature to evaluate proposed estimators since the performance of standard algorithms like UE is poor in this situation. This particular example with $\rho = 0.9999$ (correlation coefficient), corresponds to test case considered in [3][figure 2]. A pair of widely spaced signals are considered as in the first

example (see table 4). The number of snapshots is $K = 10$. The two source correlation matrices that correspond to the correlated and uncorrelated test cases (806) discussed in 5.2.3 are considered. The eigenvalues of both P_{uncorr} and P_{corr} are $(1 + \rho, 1 - \rho)$ and for the choice $\rho = 0.9999$ these eigenvalues are $(1.9999, 0.0001)$. The unconstrained subspace CRB, CRB_{SS} (813), is identical for both cases since the eigenvalues are identical. The constrained CRB, crb_{SS} , differs for the two different source correlation matrices. For the $P = P_{corr}$ (816) case the available gain is

$$\text{Gain: } \approx \sqrt{\frac{N-2}{2(1-\rho^2)}} = 47.5\text{dB} \quad (835)$$

while for $P = P_{uncorr}$, the available gain is

$$\text{Gain: } \approx \sqrt{\frac{N-2}{2}} = 10.5 \text{ dB} . \quad (836)$$

Simulation results for the $P = P_{corr}$ and $P = P_{uncorr}$ test cases are overlaid in figure 18. Both the EVD performance and CRB_{SS} are identical for the two cases, while the bounds for the constrained model, $\bar{\mathcal{B}}_{damped}$, differ as does estimator performance. At sufficiently high SNR the gain values, (836) and (835), are confirmed. For the highly correlated case, higher SNR is required for the constrained bounds to be achieved. Figure 19 shows UE and UEN along with $MLESIESP$ and $MLESIESPN$ for the correlated example. As evident, UE is severely degraded in the highly uncorrelated case; the one step Newton's method (UEN) estimator shows dramatic improvement and achieves the bound at high SNR. Performance is close to that of the $MLESIESPN$ result.

Example 5 : 4 Very Widely spaced at Large Angles: Uncorrelated, Equal-Power

This benign example consists of four widely spaced sources with locations specified in table 5. The number of independent snapshots is $K = 10$. The eigenvalues of signal covariance R_s and the Jacobian $\mathbf{J}$ are, respectively,

$$\sigma(R_s) = [81.48, 65.75, 60.20, 48.57] \quad (837)$$

Figure 18. RMSEs $\hat{\epsilon}_{SS}$ of the EVD, *MLESI*, *MLESIESP* subspace estimators and CRB_{SS} , CRB_{damped} and $CRB_{harmonic}$ versus SNR for two different source correlation matrices as given in (806). The eigenvalues of both P_{uncorr} and P_{corr} are $(1+\rho, 1-\rho)$ and for the choice $\rho = 0.9999$ these eigenvalues are $(1.9999, 0.0001)$. For the unconstrained assumption, CRB_{SS} is identical for both cases since the eigenvalues are identical. The bounds for the constrained assumption (i.e., CRB_{damped} and $CRB_{harmonic}$) differ for the two choices by the amount expected.

Figure 19. RMSE $\hat{e}_\theta$ (825) for the *MLESIESP* and *MLESIESPN*, *UE* and *UEN* estimators and the parameter space bound $CRB_{\theta\theta}$ (820) versus SNR for two different source correlation matrices as given in (806). Since the system eigenvalues are the same for both cases $CRBS$ is the same for both. The constrained bounds (e.g.; $CRBH$) differ significantly between the two cases.

Table 5. Example 5: Spatial frequencies, 4 widely spaced sources,

ID	horizontal	vertical
Source 1	1.84	-0.92
Source 2	-1.85	-0.921
Source 3	1.82	1.000
Source 4	-1.86	1.001

$$10\log_{10}(\sigma(\mathbf{J})) = (3.84, 3.79, 3.76, 3.66, 3.61, 3.54, 3.51, 3.49) \quad (838)$$

Figure 20 shows the subspace performance plot for this scenario. We observe that *MLESI* achieves the CRB for the $\bar{\mathcal{B}}_{damped}$ assumption (CRB_{damped}), while both *MLESIESP* and *MLESIESPN* achieve the CRB for the $\bar{\mathcal{B}}_{harmonic}$ assumption ($CRB_{harmonic}$). Following the discussion in section 5.2.2, and referring to table 3 with $N = 64$ and $p = 4$, we compute gains of 14 dB and 17 dB for the damped and harmonic assumptions. The error variance performance measure of three estimators is ordered in the expected way as

$$MLESIESPN < MLESIESP < MLESI. \quad (839)$$

Since the *MLESIESP* estimator very closely approaches $CRB_{harmonic}$ at moderate SNR, the Newton's refinement in the *MLESIESPN* estimation package has little room to show improvement in this example except at the lowest SNRs. Figure 21 shows the parameter space error plot set which shows the same performance ordering. In the more stressful example that follows (example 6) this performance ordering (839) is maintained only at the highest SNRs.

Following the discussion in section 5.2.4 we plot in figure 22 the RSS subspace error, $distSS(Y(\theta_0), \hat{Y})$, versus the RSS parameter space error, $\sqrt{\delta\theta^t \delta\theta}$, for each of 1000 realizations at fixed SNR of 21 dB. Recall that these errors are related

$$distSS(Y(\theta_0), \hat{Y}) \approx \sqrt{\delta\theta^t \mathbf{J}^{-2} \delta\theta}. \quad (840)$$

Because the eigenvalues of $\mathbf{J}$ are all approximately equal (838), so that $\mathbf{J}^{-2} \approx c^2 \mathbf{I}$, we

Figure 20. Subspace Error plots for example 5 consisting of 4 widely spaced sources as specified in table 5. The error performance of three estimators is ordered in the expected way as $MLESIESPN < MLESIESP < MLESI$. The relative gains for the damped and harmonic model assumptions are 14 dB and 17 dB respectively. Since the $MLESIESP$ estimator achieves bound at sufficient SNR the Newton's method refinement in the $MLESIESPN$ estimation package has no room to show improvement in this example except at the lowest SNRs.

Figure 21. RMSE $\hat{\epsilon}_{\theta}$ (825) for the *MLESIESP* and *MLESIESPN*, *UE* and *UEN* estimators and the parameter space bound $CRB_{\theta\theta}$ (820) versus SNR for 4 widely spaced, equal power, uncorrelated sources as specified in specified in table 5.

Figure 22. Scatter plot of the error the RSS error $\delta\theta^t\delta\theta$ versus the RSS $distSS(\hat{Y}, Y_0)$ for 1000 realization at an SNR of 21 dB the for closely spaced source example specified in table 5. Each gray dot represents the error measured for a single trial for the *MLESIESP* estimator. The green dot represents the corresponding RMSE. The black dot represents the CRB in for the two different measures, ($CRB_{\theta\theta}$ and $CRB_{harmonic}$). The red dot represents the RMSE for UE w.r.t the two measures.

have

$$distSS(Y(\theta_0), \hat{Y}) \approx c \cdot \sqrt{\delta\theta^t\delta\theta} . \quad (841)$$

This fact gives the scatter plot the appearance of a straightline. In example 6, consisting for four very closely spaced sources, the eigenvalues of the Jacobian have a large spread (see (843)) so this simple relationship is not maintained.

Example 6 : 4 very closely spaced, equal power sources at large angles

This example consists of four very closely spaced sources with locations specified in table 6. The number of independent snapshots is $K = 512$. The eigenvalues of signal

Table 6. Example 6: Spatial frequencies, 4 Closely spaced sources

ID	horizontal	vertical
Source 1	1.84	-0.92
Source 2	1.85	-0.921
Source 3	1.82	-1.000
Source 4	1.86	-1.001

covariance R_s and the Jacobian $\mathbf{J}$ are, respectively

$$\sigma(R_s) = (253.56, 2.14, 0.28, 0.0005) \quad (842)$$

and

$$10\log_{10}(\sigma(\mathbf{J})) = (18.09, 10.45, 4.51, 1.52, 1.51, -4.13, -8.74, -16.39) . \quad (843)$$

Figure 23 shows the subspace performance plot for this scenario. Once again $MLESI$ asymptotically (in SNR) closely approaches CRB_{damped} and maintains near optimal performance across the range of SNR shown.

The performance ordering of the 3 estimators, $MLESI$, $MLESIESP$ and $MLESIESPN$, depends on SNR and 3 distinct SNR regions are evident. At very high SNR, above 90 dB, we see the standard situation as occurred in the previous example in which the estimator performance is ordered as

$$MLESIESPN < MLESIESP < MLESI . \quad (844)$$

In the SNR range approximately 65 to 90 dB we see that $MLESI < MLESIESP$. The implication is that UE parameter extraction step of the $MLESIESP$ package has increased, rather than decreased, the subspace error. The

refinement step ($MLESIESPN$) reduces the error to a level below that of $MLESI$ so that in this SNR range we have performance ordering

$$MLESIESPN < MLESI < MLESIESP . \quad (845)$$

As SNR is decreased a point is reached at which the UE parameter estimation step of $MLESIESP$ is poor enough that the 1-step Newton's method yields a error larger than $MLESI$. In this SNR regime

$$MLESI < MLESIESPN < MLESIESP . \quad (846)$$

We observe that $MLESI$ estimate maintains good performance, as measured by the subspace error, well below the SNR threshold (approximately 80 dB) at which the UE parameter extraction step of the $MLESIESP$ estimator begins to degrade. In contrast, while the UEN achieves the $CRB_{harmonic}$ bound at the highest SNRs shown it begins to deviate from the bound at SNRs less than 90 dB. This compares to a deviation point for $MLESIESPN$ of 68 dB. For reference we include standard UE which serves as initialization for UEN. As is evident UE is far from bound even at the highest SNRs.

A scatter plot the parameter estimates (1000 trials) at an SNR just above the breakdown level is shown in figure 25. To better illustrate performance differences a histogram of the estimation errors associated with one of the two DOA parameters, for one of the sources, is shown in figure 26. The plot on the left in this figure shows the $MLESIESP$ estimator and the one on the right shows the UE estimator.

Figure 24 shows the parameter performance for this scenario. Although the relative ordering of the different estimators is preserved with respect to the two measures, subspace error and parameter error, the two performance plots, appear quite different. This apparent difference is explained by the error relation (841).

Figure 23. Subspace Error plots for example 6 consisting of 4 closely spaced sources as specified in specified in table 6. The performance ordering of the 3 estimators, *MLESI* (red line, open circles), *MLESIESP* (red line, *) and *MLESIESP*N (red line, diamond), depends on SNR and 3 distinct SNR regions are evident. For reference purposes UE (black line, *) and UEN (black line, diamond) are also plotted.

Figure 24. Parameter Error plots for example 6 consisting of 4 closely spaced sources as specified in specified in table 6. The performance ordering of the 3 estimators, $MLESI$, $MLESI_{ESP}$ and $MLESI_{ESPN}$, depends on SNR and 3 distinct SNR regions are evident.

Figure 25. A scatter plot of the *MLESIESP* and *UE* parameter estimates (1000 trials) for an SNR of 63 dB. Source locations are given in table 6. The plot is a closeup centered on two of the four sources. The UE estimates appear as the black dots and the *MLESIESP* estimates as the red dots; the true source locations are given by the white dots.

Figure 26. A histogram of the *MLESIESP* and *UE* parameter estimation errors (1000 trials) for an SNR just above the breakdown level. The figure shows the horizontal spatial frequency estimates for one of the four sources.

Interpretation of Errors as Measured in Different Coordinates

In example 5 Jacobian was $\mathbf{J} \approx c^{-1} \cdot \mathbf{I}$ so that $\text{dist}SS^2(\hat{\mathbf{Y}}, Y_0) \approx c^{-2} \delta\theta^t \delta\theta$. This fact gave the error scatter plot the appearance of a straight line (see figure 22). In the present example the eigenvalues of $\mathbf{J}$ (843) have a large spread so the relation between the two error measures is no longer so simple. Two parameter error vectors, say $\delta\theta_1$ and $\delta\theta_2$ of equal magnitudes $\delta\theta_1^t \delta\theta_1 = \delta\theta_2^t \delta\theta_2$ can result in subspace errors of very different magnitudes because of the weighting matrix form, $\mathbf{J}^{-2}$. This fact accounts for the appearance of the error scatter plot for this example given in figure 27 (compare this with 22). The horizontal axis is $\delta\theta^t \delta\theta$ while the vertical axis is $\text{dist}SS^2(\hat{\mathbf{Y}}, Y_0)$ and each gray dot represents the error measured for 1 trial. The bold black dot in the plot represents the CRBs for the subspace and parameter space as the point $(CRB_{\text{harmonic}}, CRB_{\theta})$ for the fixed set of system parameters. The green dot gives the MS errors for the *MLESIESP* estimator for the two measures as the point $(\hat{\epsilon}_{ss}, \hat{\epsilon}_{\theta})$.

List of References

- [1] P. Stoica and A. Nehorai, “Music, maximum likelihood, and cramér-rao bound,” *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, no. 5, p. 720 741, May 1989.

Figure 27. Scatter plot of the error the RSS error $\delta\theta^t\delta\theta$ versus the RSS $\text{dist}SS(\hat{Y}, Y_0)$ for 1000 realization at an SNR of 21 dB the for closely spaced source example specified in table 6.

- [2] P. Stoica, E. Larson, and A. Gershman, "The stochastic crb for array processing: A textbook derivation," *IEEE Transactions on Signal Processing Letters*, vol. 8, no. 5, pp. 148–150, May 2001.
- [3] M. Haardt and G. D. Galdo, "Higher-order svd based subspace estimation to improve the parameter estimation accuracy in multi-dimensional harmonic retrieval problems," *IEEE Trans. Signal Processing*, vol. 56, p. 3198 3213, July 2008.
- [4] L. L. Scharf, *Statistical Signal Processing: Detection, Estimation, and Time Series Analysis*. Reading, MA: Addison-Wesley, 1991.
- [5] Z. Xu, "Perturbation analysis for subspace decomposition with applications in subspace-based algorithms," *IEEE Transactions on Signal Processing*, vol. 50, no. 11, p. 2820 2830, Nov 2002.

APPENDIX A

appendix A

A.1 Coordinate Representations

The notion of coordinates plays an important role in manifold theory [1]. For the case of a sphere embedded in a 3-dimensional Euclidean space, two different coordinate systems were discussed in chapter 1. The first, called extrinsic coordinates, represented points on the sphere by their (x, y, z) Cartesian coordinates on the ambient 3-D space. The second set, called normal coordinates, was defined in terms of the tangent space at a given point and the exponential (geodesic) map (see section 2.1). The transformation from the (u, v) polar normal coordinates to the (x, y, z) was given by

$$x = \cos u \sin v, \quad y = \sin u \sin v, \quad z = \cos v. \quad (\text{A.1})$$

In this section we consider coordinates for the basis manifold $\mathcal{B}$ that are analogous to these extrinsic and normal coordinates defined for the sphere example. Normal coordinates are a special type of intrinsic coordinates where the coordinate lines are themselves geodesics generated by the exponential map.

A.1.1 Coordinate System 1: The $N \times p$ matrix extrinsic coordinates

A point on the basis manifold $\mathcal{B}$ may be represented by an $N \times p$ semi-unitary matrix. This is usual representation used in signal processing applications where the focus is on numerical calculation and where $\mathcal{B}$ is viewed as embedded in $C^{N \times p}$.

The more abstract coordinate-free representation of points on manifolds was not utilized in the main text and coordinates were used at the outset. In this section we proceed more formally and introduce the $N \times p$ matrix representation of points as the result of a coordinate mapping. This coordinate function, denoted y , is defined as mapping from $St_{(N,p)}$ to $C^{N \times p}$

$$y : St_{(N,p)} \rightarrow C^{N \times p}. \quad (\text{A.2})$$

The coordinates of the abstract point $\mathcal{Y} \in St_{(N,p)}$ are then represented as

$$Y = y(\mathcal{Y}) . \quad (\text{A.3})$$

where Y is a $N \times p$ complex matrix. This coordinate mapping is invertible (on its image) so that

$$\mathcal{Y} = y^{-1}(Y) . \quad (\text{A.4})$$

The representations Y , which reference the ambient space, are called extrinsic coordinates on $\mathcal{B}$ and are distinguished by the fact that the number of components in the coordinate N -tuple is greater than the dimension of the manifold (this is analogous to using 3 (x, y, z) Cartesian coordinates to specify a point on a 2-dimensional sphere). The number of values required to specify the complex $N \times p$ matrix Y is $2Np$; this compares to the dimension of $St_{(N,p)}$ as $2Np - p^2$. As was the case with the sphere, the Stiefel manifold, $St_{(N,p)}$, may be specified in terms of these extrinsic coordinates as the set

$$y(St_{(N,p)}) = \{Y \in C^{N \times p} : Y^H Y = I\} . \quad (\text{A.5})$$

A.1.2 Coordinate System 2: Riemannian normal coordinate system (α, β)

Riemannian normal coordinates are a local coordinate system in a neighborhood of a point $\mathcal{Y}_0$ obtained by applying the exponential map to the tangent space at $\mathcal{Y}_0$. By the exponential map, any arbitrary choice of basis for $T_{\mathcal{Y}_0}\mathcal{B}$ yields coordinates on $\mathcal{B}$ near $\mathcal{Y}_0$. These coordinates are formed as follows. Let $\Delta \in T_{\mathcal{Y}_0}\mathcal{B}$ denote the tangent vector between $\mathcal{Y}_0$ (thought of as the origin of the coordinate system) and an arbitrary $\mathcal{Y} \in \mathcal{B}$, that is

$$\Delta = \exp_{\mathcal{Y}_0}^{-1}\mathcal{Y} . \quad (\text{A.6})$$

Given an arbitrary, abstract, partitioned orthonormal basis, $\mathbf{a}_k, \mathbf{b}_k$, for $T_{\mathcal{Y}_0}\mathcal{B} = A_{\mathcal{Y}_0} \oplus B_{\mathcal{Y}_0}$ the vector Δ expands as

$$\Delta(\alpha, \beta) = \sum_{k=1}^{N_a} \mathbf{a}_k \alpha^k + \sum_{k=1}^{N_b} \mathbf{b}_k \beta^k . \quad (\text{A.7})$$

For a fixed orthonormal basis, any such tangent vector Δ is specified by the pair (α, β) . We will sometimes write $\Delta(\alpha, \beta)$ to denote the vector whose components are (α, β) when the basis is understood.

Define the invertible mapping, $\mathcal{X}$, which, for the given basis, assigns to the tangent vector Δ its components. Since the tangent space for the problem of interest is naturally partitioned this mapping has the form

$$\mathcal{X}: T_{\mathcal{Y}}\mathcal{B} \rightarrow R^{N_a} \times R^{N_b} \quad (\text{A.8})$$

and, referring to (A.7), is given by

$$\mathcal{X}(\Delta) = (\alpha, \beta) . \quad (\text{A.9})$$

This mapping is invertible

$$\mathcal{X}^{-1}: R^{N_a} \times R^{N_b} \rightarrow T_{\mathcal{Y}}\mathcal{B} \quad (\text{A.10})$$

and the inverse map constructs the tangent vector Δ from the pair (α, β) according to

$$\mathcal{X}^{-1}(\alpha, \beta) = \Delta = \sum_{k=1}^{N_a} \mathbf{a}_k \alpha^k + \sum_{k=1}^{N_b} \mathbf{b}_k \beta^k . \quad (\text{A.11})$$

The normal coordinate function denoted φ

$$\varphi: \mathcal{B} \rightarrow R^{N_a} \times R^{N_b} , \quad (\text{A.12})$$

is then defined by the composite mapping

$$\varphi = \mathcal{X} \circ \exp_{\mathcal{Y}_0}^{-1} . \quad (\text{A.13})$$

This function assigns to each point $\mathcal{Y} \in \mathcal{B}$ the real tuples $(\alpha, \beta) \in R^{N_a} \times R^{N_b}$ by

$$(\alpha, \beta) = \varphi(\mathcal{Y}) = \mathcal{X} \circ \exp_{\mathcal{Y}_0}^{-1} \mathcal{Y} . \quad (\text{A.14})$$

The composite coordinate function φ first computes $\Delta = \exp_{\mathcal{Y}_0}^{-1} \mathcal{Y}$, the tangent vector Δ from $\mathcal{Y}_0$ to the point $\mathcal{Y}$. The function $\mathcal{X}$ then represents Δ with respect to an arbitrary

fixed basis for the tangent space. The coefficients of Δ , with respect to this basis, are (α, β) . The real tuples (α, β) are said to be the Riemannian Normal coordinates of $\mathcal{Y}$.

The coordinate mapping φ is invertible

$$\varphi^{-1}: R^{N_a} \times R^{N_b} \rightarrow \mathcal{B}, \quad (\text{A.15})$$

and the point in $\mathcal{B}$ corresponding the coordinates (α, β) is given by the inverse coordinate map

$$\varphi^{-1} = (\mathcal{X} \circ \exp_{\mathcal{Y}_0}^{-1})^{-1} = \exp_{\mathcal{Y}_0} \circ \mathcal{X}^{-1}, \quad (\text{A.16})$$

so that $\mathcal{Y} = \varphi^{-1}(\alpha, \beta)$.

The exponential mapping at the point $\mathcal{Y}_0$ is a mapping from the tangent space $T_{\mathcal{Y}_0}\mathcal{B}$ to $\mathcal{B}$,

$$\exp_{\mathcal{Y}_0}: T_{\mathcal{Y}_0}\mathcal{B} \rightarrow \mathcal{B}. \quad (\text{A.17})$$

Noting (A.11), the inverse coordinate mapping φ^{-1} then is

$$\mathcal{Y} = \varphi^{-1}(\alpha, \beta) = \exp_{\mathcal{Y}_0} \Delta(\alpha, \beta). \quad (\text{A.18})$$

Normal coordinates are not unique since the choice of the orthonormal basis (A.7) used to represent the tangent vector was not unique. Note that the total number of the normal coordinate components is equal to the dimension of the manifold (i.e., $N_b = 2p(N - p)$) and $N_b = p^2$ so that $N_a + N_b = 2Np - p^2$ which is equal to the dimension of $St_{(N,p)}$.

In a normal coordinate system, the Christoffel symbols (see section A.4.1) of the connection vanish at the point $\mathcal{Y}_0$, thus often simplifying local calculations. In such coordinates, the covariant derivative reduces to a partial derivative (at $\mathcal{Y}_0$ only), and the geodesics through $\mathcal{Y}_0$ are locally linear functions. This feature comes into play when we later compute the 2nd covariant derivative of the likelihood function.

A.1.3 Coordinate Transformations: Normal Coordinates and Extrinsic Coordinates

The transformation between these normal coordinates and extrinsic coordinates is given by the composite map

$$(y \circ \varphi^{-1}): R^{N_a} \times R^{N_b} \rightarrow C^{N \times p} \quad (\text{A.19})$$

or

$$Y = y \circ \varphi^{-1}(\alpha, \beta) . \quad (\text{A.20})$$

From (A.16) we have

$$\mathcal{Y} = \varphi^{-1}(\alpha, \beta) = \exp \circ \mathcal{X}^{-1}(\alpha, \beta) \quad (\text{A.21})$$

and using (A.11)

$$Y = y(\mathcal{Y}) = \exp_{y_0} \Delta(\alpha, \beta) \quad (\text{A.22})$$

(see figure 3).

A.2 Curves on Manifolds

A curve on $\mathcal{B}$ is defined by a smooth map from reals R to $\mathcal{B}$, $c: R \rightarrow \mathcal{B}$. This curve is expressed in extrinsic coordinates via the composite mapping

$$y \circ c: R \rightarrow C^{N \times p} . \quad (\text{A.23})$$

When the curve “ c ” is understood, this is composite mapping is often written as

$$Y(t) = y(c(t)) . \quad (\text{A.24})$$

This same curve may be represented in normal coordinates by

$$\varphi \circ c: R \rightarrow R^{N_a} \times R^{N_b} . \quad (\text{A.25})$$

Similarly, when the curve “ c ” is understood, this is composite mapping is written as

$$(\alpha(t), \beta(t)) = \varphi(t) = \varphi(c(t)) \quad (\text{A.26})$$

The transformation from the normal coordinate representation of $c(t)$ to extrinsic coordinate representation is given by

$$Y(t) = \exp_{\mathcal{Y}_0} \Delta(\alpha(t), \beta(t)) . \quad (\text{A.27})$$

This corresponds to the computational form

$$Y(t) = \begin{bmatrix} Y_0 & Y_{0\perp} \end{bmatrix} \expm \left(\begin{bmatrix} A(t) & -B^H(t) \\ B(t) & 0 \end{bmatrix} \right) \cdot I_{N,p} \quad (\text{A.28})$$

where the matrices $A(t)$ and $B(t)$ are defined with respect to the basis vectors $\{\mathbf{a}_k = Y_0 \cdot A_k\}$ and $\{\mathbf{b}_k = Y_{0\perp} \cdot B_k\}$ as

$$A(t) = \sum_{k=1}^p A_k \alpha^k(t) , \quad (\text{A.29})$$

$$B(t) = \sum_{k=1}^p B_k \beta^k(t) . \quad (\text{A.30})$$

A.2.1 Geodesic Curves

A geodesic curve emanating from the point $\mathcal{Y}_0$ in the fixed direction Δ is given by

$$c(t) \equiv \exp_{\mathcal{Y}_0}(t\Delta) . \quad (\text{A.31})$$

Representing the tangent vector $\Delta(t) \equiv t\Delta$ with respect to an orthonormal basis as

$$\Delta(t) = t\Delta = t \left(\sum_{k=1}^{N_a} \mathbf{a}_k \alpha^k + \sum_{k=1}^{N_b} \mathbf{b}_k \beta^k \right) \quad (\text{A.32})$$

the corresponding normal coordinate representation of a geodesic curve $c(t)$ is then given by the straight line in $R^{N_a} \times R^{N_b}$ as

$$\mathcal{X}(\Delta(t)) = (\alpha(t), \beta(t)) = (t\alpha, t\beta) . \quad (\text{A.33})$$

The corresponding extrinsic coordinate representation of the geodesic curve $c(t)$ is $y(c(t))$ and is given by

$$Y(t) = \begin{bmatrix} Y_0 & Y_{0\perp} \end{bmatrix} \exp \left(t \begin{bmatrix} A & -B^H \\ B & 0 \end{bmatrix} \right) \cdot I_{N,p} \quad (\text{A.34})$$

where we have used (A.29), (A.30) and (A.33).

Geodesic Coordinate Curves

A geodesic coordinate curve corresponding to the α^k coordinate is given in normal coordinates by

$$(t \cdot i_k, 0_{N_b}) \quad (\text{A.35})$$

where i_k is an $N_\alpha \times 1$ unit vector with 1 in the k -th location and zeros elsewhere and 0_{N_b} is the $N_b \times 1$ zero vector. If the unit tangent

$$\mathbf{a}_k \equiv \mathcal{X}^{-1}(i_k, 0_{N_b}) , \quad (\text{A.36})$$

is represented as $\mathbf{a}_k = Y_0 A_k$ then the geodesic (A.35) is represented in extrinsic coordinates as

$$Y_{(k)}(t) \equiv Y(t \cdot i_k, 0) = [Y_0 \ Y_{0\perp}] \expm \left(t \begin{bmatrix} A_k & 0 \\ 0 & 0 \end{bmatrix} \right) \cdot I_{N,p} . \quad (\text{A.37})$$

Likewise, a geodesic coordinate curve corresponding to the β^k coordinate is given in normal coordinates by $(0_{N_a}, t \cdot i_k)$. If the unit tangent

$$\mathbf{b}_k \equiv \mathcal{X}^{-1}(0_{N_a}, i_k) , \quad (\text{A.38})$$

is represented as $\mathbf{b}_k = Y_{0\perp} B_k$ then the geodesic coordinate curve is represented in extrinsic coordinates as

$$Y_{(k)}(t) = Y(0, t \cdot i_k) = [Y_0 \ Y_{0\perp}] \expm \left(t \begin{bmatrix} 0 & -B_k^H \\ B_k & 0 \end{bmatrix} \right) \cdot I_{N,p} \quad (\text{A.39})$$

For later reference the expanded form of this curve is

$$Y_{(k)}(t) = Y(0, t \cdot i_k) = Y_0 \left(I_p - \frac{1}{2} (B_k^H B_k) + o(t^4) \right) + Y_{0\perp} \left(t \cdot B_k - t^3 \frac{1}{2} (B_k B_k^H B_k) + o(t^5) \right) \quad (\text{A.40})$$

The first order approximation is $Y_{(k)}(t) = Y(0, t \cdot i_k) \approx Y_0 + t \cdot Y_{0\perp} B_k$. The form (A.40), and a corresponding one for the α geodesic coordinate curve set, will be important when we compute the covariant 2nd differential of the likelihood function defined on $\mathcal{B}$.

Referring (A.37) to (A.39) and the extrinsic representation of tangents to these geodesic coordinate curves are

$$\mathbf{a}_k = \frac{dY(t \cdot i_k, 0)}{dt} = Y_0 A_k , \quad (\text{A.41})$$

$$\mathbf{b}_k = \frac{dY(0, t \cdot i_k)}{dt} = Y_{0\perp} B_k . \quad (\text{A.42})$$

The right hand side of both expressions are $N \times p$ complex matrices.

A.2.2 Functions on Manifolds

A real function on $\mathcal{B}$ is the mapping $L: \mathcal{B} \rightarrow R$, and the value of the function at $\mathcal{Y} \in \mathcal{B}$ is denoted $L(\mathcal{Y})$. The extrinsic coordinate representation of L is denoted L_y and is defined by

$$L_y = L \circ y^{-1} \quad (\text{A.43})$$

so that

$$L_y: C^{N \times p} \rightarrow R . \quad (\text{A.44})$$

The value of the function at Y is $L_y(Y)$ and this equals $L(\mathcal{Y})$ for $\mathcal{Y} = y^{-1}(Y)$. Likewise, in the (α, β) normal coordinates, the representation of L is denoted L_φ and is defined by

$$L_\varphi = L \circ \varphi^{-1} \quad (\text{A.45})$$

so that

$$L_\varphi: R^{N_a} \times R^{N_b} \rightarrow R . \quad (\text{A.46})$$

The value the function at (α, β) is $L_\varphi(\alpha, \beta)$ and this equals $L(\mathcal{Y})$ for

$$\mathcal{Y} = \varphi^{-1}(\alpha, \beta) . \quad (\text{A.47})$$

The functional forms $L_\varphi(\alpha, \beta)$ and $L_y(Y)$ are related by

$$L_\varphi = L \circ y^{-1} \circ y \circ \varphi^{-1} = L_y \circ y \circ \varphi^{-1} \quad (\text{A.48})$$

or

$$L_\varphi(\alpha, \beta) = L_y(Y) \quad (\text{A.49})$$

where

$$Y = y \circ \varphi^{-1}(\alpha, \beta) . \quad (\text{A.50})$$

In later sections we will denote L_φ , L_y and L by the same symbol “ L ” since the meaning will generally be clear from the choice of arguments. That is

$$L(\mathcal{Y}) = L(Y) = L(\alpha, \beta) \quad (\text{A.51})$$

when $Y = y(\mathcal{Y})$ and $(\alpha, \beta) = \varphi(\mathcal{Y})$. Henceforth we deal with the extrinsic coordinate representation $Y = y(\mathcal{Y})$.

A.2.3 Coordinate Curves

Let $c_{(i)}(t)$ denote the geodesic coordinate curve emanating from $\mathcal{Y}_0$ in the $\mathbf{a}_i$ direction. The normal coordinate representation of this curve is $(t \cdot i_k, 0)$, that is

$$\varphi \circ c_{(i)}(t) = (t \cdot i_k, 0) \quad (\text{A.52})$$

then

$$L(c_{(i)}(t)) = L_\varphi(t \cdot i_k, 0) . \quad (\text{A.53})$$

Therefore, we can compute the directional derivative in the direction $\mathbf{a}_k$ by

$$\nabla L[\mathbf{a}_k] = \frac{dL_\varphi(t \cdot i_k, 0)}{dt} . \quad (\text{A.54})$$

For computational purposes we can write this in terms of the extrinsic coordinates

$$L_y(Y_{(k)}(t)) = L_\varphi(t \cdot i_k, 0)|_{t=0}; \quad k = 1 \cdots N_a . \quad (\text{A.55})$$

where we have used the notation

$$Y_{(k)}(t) \equiv Y(t \cdot i_k, 0) , \quad (\text{A.56})$$

A.2.4 Calculus on Manifolds : Partial Derivative

Let x be a coordinate system on the manifold $\mathcal{M}$, that is

$$x : \mathcal{M} \mapsto R^N \quad (\text{A.57})$$

then $L_x \equiv L \circ x^{-1}$ is a function defined on R^N . The standard practice of using the same symbol for both the independent variable and a function is used here. In elementary calculus, the partial derivative of $L(x)$ with respect to x^k is defined as

$$\frac{\partial L}{\partial x^k} \Big|_{x=x_0} \equiv \lim_{t \rightarrow 0} \frac{L(x_0^1, \dots, x_0^k + t, \dots, x_0^q)}{dt} \quad (\text{A.58})$$

This is equivalent to

$$\frac{dL(x_{(k)}(t))}{dt} \equiv \lim_{t \rightarrow 0} \frac{L(x_0^1, \dots, x_0^k + t, \dots, x_0^q)}{dt} \quad (\text{A.59})$$

where $x_{(k)}(t)$ is the coordinate curve

$$x_{(k)}(t) = (x_0^1, \dots, x_0^k + t, \dots, x_0^q) . \quad (\text{A.60})$$

This curve is also expressed in the alternate form

$$x_{(k)}(t) = x_0 + t \cdot e_k \quad (\text{A.61})$$

where e_k is a $q \times 1$ unit vector with 1 at the k -th location.

Applied to the coordinate function $Y = (Y^1, \dots, Y^M)$ we have

$$\frac{\partial Y^i}{\partial x^k} \equiv \frac{dY^i(x_{(k)}(t))}{dt}, \quad m = 1 \dots M . \quad (\text{A.62})$$

In matrix form this is

$$\frac{\partial Y}{\partial x^k} \equiv \frac{dY(x_{(k)}(t))}{dt} . \quad (\text{A.63})$$

A.3 Elements of Differential Geometry

A.3.1 Natural Coordinate Basis

Let x be coordinate system with origin at the point Y , that is $x(\mathcal{Y}) = (0, \dots, 0)$. The k -th coordinate curve is $c_{(k)}(t) = x^{-1} \circ x_{(k)}(t)$ where $x_{(k)}(t) \equiv (0, \dots, t, \dots, 0)$ (that is, t in the k -th location and zeros elsewhere). The set of tangents

$$\mathbf{e}_k = \frac{dc_{(k)}(t)}{dt}; \quad k = 1, \dots, N \quad (\text{A.64})$$

define a basis for $T_Y \mathcal{B}$ termed a coordinate basis. Such coordinate basis vectors are often denoted using the notation $\frac{\partial}{\partial x^k}$ rather than $\mathbf{e}_k$.

$$\mathbf{e}_k \leftrightarrow \frac{\partial}{\partial x^k} . \quad (\text{A.65})$$

If $\acute{x}$ is an alternate coordinate system then we write $\acute{\mathbf{e}}_k \leftrightarrow \frac{\partial}{\partial \acute{x}^k}$.

An arbitrary vector $\mathbf{b}$ expands as

$$\mathbf{b} = \sum \mathbf{e}_k \beta^k = \sum \acute{\mathbf{e}}_k \acute{\beta}^k . \quad (\text{A.66})$$

If T is the transformation between $\{\mathbf{e}_i\}$ and $\{\acute{\mathbf{e}}_i\}$, that is

$$\acute{\mathbf{e}}_k = \sum T_k^j \mathbf{e}_j \quad (\text{A.67})$$

then, in matrix form, the coefficients in (A.66) are transformed as

$$\acute{\beta} = T^{-1} \beta . \quad (\text{A.68})$$

A.3.2 Dual Basis, Tensors

The dual space of an N -dimensional vector space V , denoted V^* , is defined by the collection of linear functions on V , that is

$$V^* \equiv \{f : V \rightarrow R\} . \quad (\text{A.69})$$

The dual space of $T_Y \mathcal{B}$, denoted $T_Y^* \mathcal{B}$ and called the cotangent space, is defined by the collection of linear functions on $T_Y \mathcal{B}$

$$T_Y^* \mathcal{B} \equiv \{\omega : T_Y \mathcal{B} \rightarrow R\} . \quad (\text{A.70})$$

If $\mathbf{e}_i$ (subscript index) is a basis for $T_Y\mathcal{B}$, denote the corresponding dual-basis for $T_Y^*\mathcal{B}$ by $\mathbf{e}^i$ (superscript index). This dual basis is defined by its action

$$\mathbf{e}^j[\mathbf{e}_i] = \delta_j^i. \quad (\text{A.71})$$

In the previous section we introduced the notation $\frac{\partial}{\partial x^k}$ to stand for $\mathbf{e}_k$ when referring to coordinate basis vectors. Similarly here dx^k stands for $\mathbf{e}^k$, that is

$$\mathbf{e}^k \leftrightarrow dx^k. \quad (\text{A.72})$$

Using this notation the relation (A.71) becomes

$$dx^i \left[\frac{\partial}{\partial x^j} \right] = \delta_j^i \quad (\text{A.73})$$

An arbitrary $\omega \in T_Y^*\mathcal{B}$ expands as

$$\omega = \sum_{k=1}^n w_k \mathbf{e}^k \quad (\text{A.74})$$

for some set of coefficients w_k . For a given ω and basis these components are

$$w_k = \omega(\mathbf{e}_k). \quad (\text{A.75})$$

The above objects are particular examples of what are more generally called tensors. Elements of the tangent space $T_Y\mathcal{B}$ are called tangent vectors or contravariant (1,0) tensors. Elements of the cotangent space $T_Y^*\mathcal{B}$ are called cotangent vectors, 1-forms, or covariant (0,1) tensors. Geometric objects in higher dimensions are constructed in an analogous way. A covariant (0,2)-tensor, denoted generically as $\mathbf{T}$, is a bilinear function

$$\mathbf{T} : T_Y\mathcal{B} \times T_Y\mathcal{B} \rightarrow R \quad (\text{A.76})$$

and the collection of all such functions is denoted as $T_Y^{*2}\mathcal{B}$.

A basis for $T_Y^{*2}\mathcal{B}$ is formed from the tensor products of the basis vectors for $T_Y^*\mathcal{B}$; these are $(\mathbf{e}^i \otimes \mathbf{e}^j)$. This dual basis is defined by its action in the same way as (A.71)

$$(\mathbf{e}^i \otimes \mathbf{e}^j)[\mathbf{e}_k, \mathbf{e}_m] = \delta_k^i \delta_m^j. \quad (\text{A.77})$$

An arbitrary $\mathbf{F} \in T_Y^{*2}\mathcal{B}$ expands as

$$\mathbf{F} = \sum \sum F_{ij} \mathbf{e}^i \otimes \mathbf{e}^j \quad (\text{A.78})$$

for some set of coefficients F_{ij} . In an operation analogous to (A.75) that makes use of (A.77), the components of $\mathbf{F}$ with respect to this basis are given by

$$F_{km} \equiv \mathbf{F}[\mathbf{e}_k, \mathbf{e}_m] . \quad (\text{A.79})$$

The metric tensor g (230) is a tensor of this type. The Fisher Information form is also a tensor of this type. Just as vectors can be represented in different basis sets so can tensors. Let $\mathbf{e}_i$ be an alternate basis for $T_Y\mathcal{B}$ and let $\mathbf{e}^i$ be the corresponding dual basis for $T_Y^*\mathcal{B}$. Then the component matrix of $\mathbf{F}$ with respect to this dual basis is

$$\acute{F}_{ij} \equiv \mathbf{F}[\mathbf{e}_i, \mathbf{e}_j] . \quad (\text{A.80})$$

An alternate representation of the tensor $\mathbf{F}$ (A.78) is

$$\mathbf{F} = \sum \sum \acute{F}_{ij} \mathbf{e}^i \otimes \mathbf{e}^j . \quad (\text{A.81})$$

Substituting (A.67) into (A.80) yields

$$\acute{F}_{ij} \equiv \mathbf{F}[\sum T_i^k j \mathbf{e}_k, \sum T_j^m \mathbf{e}_m] = \sum \sum T_i^k T_j^m \mathbf{F}[\mathbf{e}_k, \mathbf{e}_m] \quad (\text{A.82})$$

and noting (A.79) we have

$$\acute{F}_{km} = \sum \sum T_k^i T_m^j F_{ij} \quad (\text{A.83})$$

or in matrix form

$$\acute{F} = T^t F T . \quad (\text{A.84})$$

This equation represents the change of basis for a (0,2) tensor (covariant 2-tensor) (see [2, eqn 28]).

The expression for a quadratic form is as follows. Let $\mathbf{b} = \sum \mathbf{e}_i \beta^i$, then

$$\mathbf{F}[\mathbf{b}, \mathbf{b}] = \mathbf{F}[\sum \mathbf{e}_i \beta^i, \sum \mathbf{e}_j \beta^j] \quad (\text{A.85})$$

$$= \sum \sum \beta^i \mathbf{F}[\mathbf{e}_i, \mathbf{e}_j] \beta^j = \sum \sum \beta^i F_{ij} \beta^j . \quad (\text{A.86})$$

Expressing the summation on the right hand side in matrix form the result is

$$\mathbf{F}[\mathbf{b}, \mathbf{b}] = \beta^t F \beta . \quad (\text{A.87})$$

Using (A.84) and (A.87) the quadratic form in the alternate basis is

$$\mathbf{F}[\mathbf{b}, \mathbf{b}] = \dot{\beta}^t \dot{F} \dot{\beta} = \beta^t T^{-t} \cdot T^t \cdot F \cdot T \cdot T^{-1} \beta = \beta^t F \beta . \quad (\text{A.88})$$

As noted in [3] ”...the tensor product of a tangent vector with itself is equivalent to the outer product given a particular choice of coordinates.” For the tangent vector $\Delta = \sum \mathbf{e}_k \beta^k$ the tensor product is

$$\Delta \otimes \Delta = \sum \sum (\beta^i \beta^j) \mathbf{e}_i \otimes \mathbf{e}_j . \quad (\text{A.89})$$

The components of the tensor $C \equiv \Delta \otimes \Delta$ are

$$C^{km} \equiv (\Delta \otimes \Delta) [\mathbf{e}^k, \mathbf{e}^m] = \beta^k \beta^m , \quad (\text{A.90})$$

or in matrix form

$$C = \beta \beta^t . \quad (\text{A.91})$$

A.3.3 Tensor Fields

Vectors defined at all points on a manifold are called vector fields. “Intuitively, a vector field is best visualized as an “arrow” attached to each point of a region, with variable length and direction. One example of a vector field on a curved space is a weather map showing horizontal wind velocity at each point of the Earth’s surface.” Tensor fields are defined in an analogous way. The 9 element stress tensor appearing in fluid and solid mechanics is standard example (1 normal and 2 shear stresses on each of the 3 faces of the cube).

Generic tensors fields are denoted here as $\mathbf{T}$, and the value of the field at a particular point $\mathcal{P}$ is $\mathbf{T}(\mathcal{P})$. The most relevant types for our discussion are listed below.

(1,0)-tensor, tangent vector field

$$\mathbf{T}(\mathcal{P}) = \sum T^i \mathbf{e}_i = \sum T^i \frac{\partial}{\partial x^i} \quad (\text{A.92})$$

(0,1)-tensor, cotangent vector field, 1-form

$$\mathbf{T}(\mathcal{P}) = \sum T_k \mathbf{e}^k = \sum T^i dx^i \quad (\text{A.93})$$

(0,2)-tensor, covariant 2-tensor field

$$\mathbf{T}(\mathcal{P}) = \sum T^{ij} \mathbf{e}^i \otimes \mathbf{e}^j = \sum T_{ij} dx^i \otimes dx^j \quad (\text{A.94})$$

(2,0)-tensor, contravariant 2 tensor field

$$\mathbf{T}(\mathcal{P}) = \sum T^{ij} \mathbf{e}_i \otimes \mathbf{e}_j = \sum T_{ij} \frac{\partial}{\partial x^i} \otimes \frac{\partial}{\partial x^j} \quad (\text{A.95})$$

where $T^k \equiv T^k(\mathcal{P})$ and $\mathbf{e}^k \equiv \mathbf{e}^k(\mathcal{P})$, etc.

A.3.4 Differentiation on Manifolds

Quoting from [3], “Differentiating real-valued functions defined on manifolds is straight forward because we may subtract the value of the function at one point on the manifold from its value at a different point. Differentiating first derivatives again or tangent vectors, however, is not well defined (without specifying additional information) because a vector tangent at one point is not (necessarily) tangent at another. Subtraction between two different vector spaces is not intrinsically defined. The additional structure required to differentiate vectors is called the affine connection, so-called because it allows one to connect different tangent spaces and compare objects defined separately at each point”. The connection defines and is defined by the geodesics. The 2nd covariant differential computation described below relies on the geodesic definition.

1st Covariant Derivative

The covariant differential of a function L along tangent vector $\mathbf{v}$ is denoted variously as $dL[\mathbf{v}]$, $DL[\mathbf{v}]$, $\nabla L[\mathbf{v}]$, or $\nabla_{\mathbf{v}}L$ and simply represents the standard directional

derivative, that is

$$\nabla_{\mathbf{v}} L = \frac{dL(t)}{dt} = \lim_{t \rightarrow 0} \frac{L(t) - L(0)}{t}, \quad (\text{A.96})$$

where $L(t) \equiv L(x(t))$, and $x(t)$ is an arbitrary smooth curve with tangent $\mathbf{v}$ at $t = 0$,

$$\mathbf{v} = \frac{dx(t)}{dt}, \quad v^k = \frac{dx^k(t)}{dt}, \quad k = 1, \dots, N. \quad (\text{A.97})$$

On the curve, we have $L(x(t)) = L(x^1(t), x^2(t), \dots, x^N(t))$ and by the chain rule

$$\frac{dL(t)}{dt} = \sum \frac{\partial L}{\partial x^k} \frac{dx^k(t)}{dt} = \sum \frac{\partial L}{\partial x^k} v^k. \quad (\text{A.98})$$

The directional derivative of the real valued function L in the direction of the coordinate tangent vector $\mathbf{e}_i$ is defined by

$$\nabla_{\mathbf{e}_i} L = \frac{\partial L}{\partial x^i} = \frac{dL(x_{(i)}(t))}{dt}. \quad (\text{A.99})$$

For convenience we use the notation $\nabla_i = \nabla_{\mathbf{e}_i}$ when the basis is understood. Define the operator in coordinates

$$\nabla_{\mathbf{v}} \equiv \sum v^k \nabla_{\mathbf{e}_k} \quad (\text{A.100})$$

so that

$$\nabla_{\mathbf{v}} L = (\sum v^k \nabla_{\mathbf{e}_k}) L = \sum \frac{\partial L}{\partial x^k} v^k. \quad (\text{A.101})$$

Define ∇L to be the linear operator, $\nabla L : T_Y \mathcal{B} \mapsto R$ that gives $\nabla_{\mathbf{v}} L$ when it acts on $\mathbf{v}$

$$\nabla L [\mathbf{v}] \equiv \nabla_{\mathbf{v}} L. \quad (\text{A.102})$$

If dx^k is a basis for $T_Y^* \mathcal{B}$ (dual to the $\frac{\partial}{\partial x^k}$ basis for $T_Y \mathcal{B}$) then

$$\nabla L \left[\frac{\partial}{\partial x^i} \right] = \sum \frac{\partial L}{\partial x^k} dx^k \left[\frac{\partial}{\partial x^i} \right] = \frac{\partial L}{\partial x^i}. \quad (\text{A.103})$$

The covariant differential of a function is an element of the cotangent space $T_Y^* \mathcal{B}$, that is $\nabla L \in T_Y^* \mathcal{B}$, and is represented in coordinates as

$$\nabla L = \sum \frac{\partial L}{\partial x^k} dx^k. \quad (\text{A.104})$$

A.4 Covariant Derivative

The covariant derivative generalizes the notion of the a directional derivative in an invariant, coordinate free way.

Definition Covariant Derivative:

The “covariant derivative”, $\nabla_{\mathbf{v}}\mathbf{T}$, of tensor field $\mathbf{T}$ along the arbitrary curve $\mathcal{P}(t)$ is defined as

$$\nabla_{\mathbf{v}}\mathbf{T} = \frac{d\mathbf{T}(t)}{dt} \equiv \lim_{t \rightarrow 0} \frac{\mathbf{T}(t) \text{ ll translated to } 0 - \mathbf{T}(0)}{t} \quad (\text{A.105})$$

where $\mathbf{T}(t) = \mathbf{T}(\mathcal{P}(t))$ and $\mathbf{v}$ is given by (A.97) .

The “ll translated to 0” subscript stands for the operation of parallel translating the tensor $\mathbf{T}$ at $\mathcal{P}(t)$ to the point $\mathcal{P}(0)$. Once referenced to the same vector space, that is the one at $\mathcal{P}(0)$, the subtraction operation is well defined. The geodesic gives meaning to parallel translation operation.

As before, define $\nabla\mathbf{T}$ to be the linear operator that gives $\nabla_{\mathbf{v}}\mathbf{T}$ when it acts on $\mathbf{v}$

$$\nabla\mathbf{T} [\mathbf{v}] \equiv \nabla_{\mathbf{v}}\mathbf{T} . \quad (\text{A.106})$$

For a given basis, $\mathbf{e}_k$, this is

$$\nabla_{\mathbf{v}}\mathbf{T} = \sum v^k \nabla_{\mathbf{e}_k} \mathbf{T} = \sum v^k \nabla_k \mathbf{T} \quad (\text{A.107})$$

where $\mathbf{v} = \sum v^k \mathbf{e}_k$.

Let $\mathbf{T}$ be a tangent vector ((1,0)-tensor) field (A.92). Then using the Leibniz rule

$$\nabla_i \mathbf{T} = \nabla_i \left(\sum T^j \mathbf{e}_j \right) = \nabla_i T^j \mathbf{e}_j + T^j \nabla_i \mathbf{e}_j . \quad (\text{A.108})$$

The factor $\nabla_j T^i$ represents the derivative of a set of scalar functions, so that

$$\nabla_j T^i = \frac{\partial T^i}{\partial x^j} . \quad (\text{A.109})$$

The second term in (A.108) is the derivative of a vector (covariant basis, $\mathbf{e}_i$) field. This is

$$\nabla_j \mathbf{e}_i = \frac{d\mathbf{e}_i(t)}{dt} \equiv \lim_{t \rightarrow 0} \frac{\mathbf{e}_i(t) \text{ ll translated to } 0 - \mathbf{e}_i(0)}{t} . \quad (\text{A.110})$$

The “parallel translated to 0” denotes the operation of bringing the vector at $T_{\mathcal{P}(t)}\mathcal{M}$ to the vector space $T_{\mathcal{P}(0)}\mathcal{M}$, enabling the subtraction operation in the same vector space.

This translation in the direction $\mathbf{v} = \sum \mathbf{e}_i v^i$ is given by

$$\mathbf{e}_i(t) \parallel \text{translated to } 0 = \mathbf{e}_i(0) + t \cdot \sum \Gamma_{ij}^m \mathbf{e}_m(0) v^j + o(t^2) \quad (\text{A.111})$$

and for the special case $\mathbf{v} = \mathbf{e}_i$

$$\mathbf{e}_i(t) \parallel \text{translated to } 0 = \mathbf{e}_i(0) + t \cdot \sum \Gamma_{ij}^m \mathbf{e}_m(0) + o(t^2) . \quad (\text{A.112})$$

The term $\Gamma_{ij}^m \mathbf{e}_m$ represents the “twisting, turning, expansion, and contraction of the basis vectors” in terms of the connection coefficients Γ_{ij}^m . Substituting this expression into (A.110) yields

$$\nabla_i \mathbf{e}_j = \sum \Gamma_{ij}^k \cdot \mathbf{e}_k . \quad (\text{A.113})$$

Substituting (A.113) into (A.108) yields covariant derivative of $\mathbf{T}$ in the direction $\mathbf{e}_i$, in a particular coordinate system, as

$$\nabla_i \mathbf{T} = \frac{\partial T^j}{\partial x^i} \mathbf{e}_j + T^j \Gamma_{ij}^k \mathbf{e}_k = \left(\frac{\partial T^k}{\partial x^i} + \Gamma_{ij}^k T^j \right) \mathbf{e}_k . \quad (\text{A.114})$$

The factor in parenthesis on the right hand side represents the matrix components of the covariant derivative, and is denoted variously as $T^k{}_{;i}$ or $\nabla_i T^k$

$$T^k{}_{;i} = \nabla_i T^k \equiv \frac{\partial T^j}{\partial x^i} + \Gamma_{ij}^k T^j . \quad (\text{A.115})$$

Geodesics parallel translate their own tangent vectors, so that in a normal coordinate system we have $\Gamma_{ij}^m(0) \equiv 0$. If x is a set of normal coordinates then

$$T^k{}_{;i} = \nabla_i T^k \equiv \frac{\partial T^j}{\partial x^i} . \quad (\text{A.116})$$

The covariant derivative of an element of the cotangent space is defined in a same way.

In particular, we have

$$\nabla_j \mathbf{e}^i = -\Gamma_{jk}^i \mathbf{e}^k \quad (\text{A.117})$$

so that the covariant derivative of the $(0, 1)$ tensor (A.93) is

$$T_{i;j} = \nabla_j T_i \equiv \frac{\partial T_i}{\partial x^j} - \Gamma_{ij}^k T_k . \quad (\text{A.118})$$

Example: $\mathbf{T} = \sum \frac{\partial L}{\partial x^i} \mathbf{e}^i$

Substituting this $T_i \equiv \frac{\partial L}{\partial x^i}$ into (A.118) yields

$$L_{;ij} \equiv \nabla_i \left(\frac{\partial L}{\partial x^j} \right) = \frac{\partial^2 L}{\partial x^i \partial x^j} - \Gamma_{ij}^k \frac{\partial L}{\partial x^k} . \quad (\text{A.119})$$

These are the components of covariant (0,2) tensor (A.94) expressed as

$$\nabla^2 L = \sum L_{;km} \mathbf{e}^k \otimes \mathbf{e}^m = \sum L_{;km} dx^k \otimes dx^m . \quad (\text{A.120})$$

The components are

$$\nabla^2 L [\mathbf{e}_i, \mathbf{e}_j] = \nabla^2 L \left[\frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^j} \right] = L_{;ij} . \quad (\text{A.121})$$

Going forward, we use the notation $[\nabla^2 L]$ to denote the component matrix of the Hessian tensor when the coordinates are understood.

A.4.1 2nd Covariant Derivative Differential of a function

“For arbitrary manifolds, the Hessian is the second derivative along geodesics. In differential geometry it is the second covariant differential of L ”. The Hessian of a function L is defined as the quadratic form

$$\nabla^2 L[\Delta, \Delta] \equiv Hess L[\Delta, \Delta] = \frac{d^2}{dt^2} \Big|_{t=0} L(c(t)) \quad (\text{A.122})$$

where $c(t)$ is the geodesic emanating in the direction of Δ (i.e., $\dot{c}(0) = \Delta$) [4, equation 2.54].

Let x be a normal coordinate system with origin at the point $\mathcal{P}$. In such a coordinate system the Christofel symbols vanish at $\mathcal{P}$ so that

$$\nabla^2 L [\mathbf{e}_k, \mathbf{e}_k] = \frac{\partial^2 L}{\partial x^k \partial x^k} = \frac{d}{dt} \left(\frac{\partial L(x_{(k)}(t))}{\partial x^k} \right) = \frac{d}{dt} \left(\frac{dL(x_{(k)}(t))}{dt} \right) \quad (\text{A.123})$$

or

$$\frac{\partial^2 L}{\partial x^k \partial x^k} = \frac{d^2 L(x_{(k)}(t))}{dt^2} \Big|_{t=0}; \quad k = 1 \cdots N . \quad (\text{A.124})$$

As with any bilinear form, the off-diagonal terms may be computed using the standard process of polarization

$$L_{ij} \equiv \nabla^2 L[\mathbf{e}_i, \mathbf{e}_j] = \frac{1}{4} \left(\nabla^2 L[\mathbf{e}_i + \mathbf{e}_j, \mathbf{e}_i + \mathbf{e}_j] - \nabla^2 L[\mathbf{e}_i - \mathbf{e}_j, \mathbf{e}_i - \mathbf{e}_j] \right) . \quad (\text{A.125})$$

The first term on the left hand side is given by

$$\nabla^2 L[\mathbf{e}_i + \mathbf{e}_j, \mathbf{e}_i + \mathbf{e}_j] = \frac{d^2 L(\phi(t))}{dt^2} \Big|_{t=0} \quad (\text{A.126})$$

where $\phi(t)$ is the geodesic emanating in the direction $\mathbf{e}_i + \mathbf{e}_j$. The second term is similarly computed, with $\phi(t)$ the geodesic emanating in the direction $\mathbf{e}_i - \mathbf{e}_j$.

Example: Cartesian Coordinates on the Plane

The usual Cartesian coordinate lines are geodesics on the flat plane. The components of the second covariant derivative $\nabla^2 L$ are

$$\nabla^2 L[\mathbf{e}_x, \mathbf{e}_x] = \frac{\partial^2 L}{\partial x^2} = \frac{d^2 L(\varphi_x(t))}{dt^2} \Big|_{t=0} \quad (\text{A.127})$$

where $\varphi_x(t)$ is the geodesic emanating in the direction $\mathbf{e}_x$

$$\varphi_x(t) \equiv (x + t, y) . \quad (\text{A.128})$$

Similarly

$$\nabla^2 L[\mathbf{e}_y, \mathbf{e}_y] = \frac{\partial^2 L}{\partial y^2} = \frac{d^2 L(\varphi_y(t))}{dt^2} \Big|_{t=0} \quad (\text{A.129})$$

where

$$\varphi_y(t) \equiv (x, y + t) . \quad (\text{A.130})$$

To compute the polarization form (A.125) we require the geodesic curve in the $\mathbf{e}_x + \mathbf{e}_y$ and $\mathbf{e}_x - \mathbf{e}_y$ directions. Denoting the geodesic in this direction as $\varphi_{x+y}(t) \equiv (x + t, y + t)$ we have

$$\nabla^2 L[\mathbf{e}_x + \mathbf{e}_y, \mathbf{e}_x + \mathbf{e}_y] = \frac{d^2 L(\varphi_{x+y}(t))}{dt^2} \Big|_{t=0} \quad (\text{A.131})$$

and similarly

$$\nabla^2 L[\mathbf{e}_x - \mathbf{e}_y, \mathbf{e}_x - \mathbf{e}_y] = \frac{d^2 L(\varphi_{x-y}(t))}{dt^2} \Big|_{t=0} \quad (\text{A.132})$$

where $\varphi_{x-y}(t)$ denotes the geodesic tangent to the vector $\mathbf{e}_x - \mathbf{e}_y$. The standard Hessian matrix is define as

$$[\nabla^2 L] = HessL = \begin{bmatrix} \frac{\partial^2 L}{\partial y^2} & \frac{\partial^2 L}{\partial x \partial y} \\ \frac{\partial^2 L}{\partial y \partial x} & \frac{\partial^2 L}{\partial x^2} \end{bmatrix}. \quad (\text{A.133})$$

A.5 Computing Derivatives on $\mathcal{B}$

In this section we restate the derivative definitions for functions defined on the $\mathcal{B}$ manifold using the notation just established. We recall that tangent spaces for this formulation naturally partition as $T_{Y_0} \mathcal{B} = A_{Y_0} \oplus B_{Y_0}$; we denote the partitioned basis as $\mathbf{a}_k, \mathbf{b}_k$.

The geodesic coordinate curve emanating in the $\mathbf{b}_k$ direction is given in extrinsic coordinates as $Y_{(k)}(t) \equiv \exp_{Y_0}(\mathbf{b}_k t)$ (see (A.40)). The components of the first differential with respect to this basis are

$$\frac{\partial L}{\partial \beta^k} \equiv \nabla L[\mathbf{b}_k] = \frac{dL_y(Y_{(k)}(t))}{dt} \Big|_{t=0}; \quad k = 1 \cdots N_b. \quad (\text{A.134})$$

The components of the Hessian are computed using (A.122) and the diagonal elements corresponding to the $\mathbf{b}_k$ basis vectors, for example, are

$$\frac{\partial^2 L}{\partial \beta^k \partial \beta^k} \equiv \nabla^2 L[\mathbf{b}_k, \mathbf{b}_k] = \frac{d^2 L_y(Y_{(k)}(t))}{dt^2} \Big|_{t=0}; \quad k = 1 \cdots N_b. \quad (\text{A.135})$$

To compute these quantities, substitute the expanded form of $Y_{(k)}(t)$, given by (A.39), into the given function and compute. This Taylor series expansion has the form

$$L_y(Y_{(k)}(t)) = L_0(B_k) + L_1(B_k)t + \frac{1}{2!}L_2(B_k)t^2 + \frac{1}{3!}L_3(B_k)t^3 \quad (\text{A.136})$$

so that

$$\frac{\partial L}{\partial \beta^k} = L_1(B_k) \quad (\text{A.137})$$

and

$$\frac{\partial^2 L}{\partial \beta^k \partial \beta^k} = L_2(B_k) . \quad (\text{A.138})$$

The cross terms of the full Hessian are found using polarization (A.125). The computation of the cross terms

$$[L_{\alpha\beta}]_{ij} = \frac{\partial^2 L}{\partial \alpha^i \partial \beta^j} \quad i = 1 \cdots N_a, N_b = 1 \cdots N_b \quad (\text{A.139})$$

require vectors of the form $\Delta = \mathbf{a}_i + \mathbf{b}_j$; these have the explicit form $\Delta = Y_0 A_i + Y_{0\perp} B_j$. The $\nabla^2 L[\Delta, \Delta]$ is computed in the same way as shown above for the particular case of the vector $\mathbf{b}_k$. The required derivatives are found by substituting particular choices for A and B (that is for $\mathbf{b}_1$, set $A = 0, B = B_1$). In 2.2.4 we use this approach to compute the Hessian of the likelihood function that arises in the complex linear model. These second derivatives represent the ordinary Hessian matrix of $L(\alpha, \beta)$, interpreted to be a scalar function defined on $R^{N_a} \times R^{N_b}$.

A.6 Inner Product Space

A vector space is a collection V of objects called vectors, which may be added together and multiplied (“scaled”) by numbers, called scalars in this context. The set of scalars is denoted $\mathcal{F}$; these are often taken to be real numbers, but there are also vector spaces with scalar multiplication by complex numbers, rational numbers, or generally any field. The operations of vector addition and scalar multiplication must satisfy certain requirements, called axioms. The vector space is then the pair $(V, \mathcal{F})$

When the scalar field $\mathcal{F}$ is the real numbers R , the vector space is called a real vector space. When the scalar field is the complex numbers C , it is called a complex vector space. These two cases are the ones used most often in engineering.

An inner product space is a vector space with an additional structure called an inner product. This additional structure associates each pair of vectors in the space with a scalar quantity known as the inner product of the vectors. Inner products allow the

rigorous introduction of intuitive geometrical notions such as the length of a vector, the angle between two vectors, and orthogonality between vectors (zero inner product).

Formally, an inner product space is a vector space $(V, \mathcal{F})$ with an inner product given by the mapping

$$\langle \cdot, \cdot \rangle : V \times V \rightarrow \mathcal{F} \quad (\text{A.140})$$

that satisfies certain axioms.

A.6.1 2 by 1 complex vectors over the Reals (C^2, R)

Let V be the collection of all complex n -tuples with addition defined in the usual “vector” way. We can define this as a vector space over either the real or complex field. We first consider the real case, $\mathcal{F} = R$. For $n = 2$ a set of basis vectors for this space (real dimension = 4) is

$$\mathbf{e}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}; \mathbf{e}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}; \mathbf{e}_3 = \begin{bmatrix} i \\ 0 \end{bmatrix}; \mathbf{e}_4 = \begin{bmatrix} 0 \\ i \end{bmatrix}. \quad (\text{A.141})$$

An arbitrary element $\mathbf{v} \in V$ given by

$$\mathbf{v} = \begin{bmatrix} a + ib \\ c + id \end{bmatrix} \quad (\text{A.142})$$

has the representation

$$\mathbf{v} = v^1 \cdot \mathbf{e}_1 + v^2 \cdot \mathbf{e}_2 + v^3 \cdot \mathbf{e}_3 + v^4 \cdot \mathbf{e}_4 \quad (\text{A.143})$$

where $v^1 = a; v^2 = b; v^3 = c; v^4 = d$ or

$$v = \begin{bmatrix} a & b & c & d \end{bmatrix}^t. \quad (\text{A.144})$$

For convenience we arrange the basis vectors in the matrix form

$$E \equiv \begin{bmatrix} \mathbf{e}_1 & \mathbf{e}_2 & \mathbf{e}_3 & \mathbf{e}_4 \end{bmatrix} = \begin{bmatrix} I & i \cdot I \end{bmatrix} \quad (\text{A.145})$$

so that sum (A.143) can be expressed as

$$\mathbf{v} = Ev. \quad (\text{A.146})$$

Note that $\mathbf{e}_1$ and $\mathbf{e}_3$ are linearly independent since $\mathbf{e}_1 \neq s \cdot \mathbf{e}_3$ for any $s \in \mathcal{F} \equiv R$. The standard inner product (A.140) for this space is

$$\langle \mathbf{v}, \mathbf{w} \rangle \equiv \text{Re}(\mathbf{v}^H \mathbf{w}) = v^t \text{Re}(E^H E) w . \quad (\text{A.147})$$

The idea of orthogonality depends on the choice of inner product; the basis defined above is an orthonormal one since $\text{Re}(E^H E) = I$.

For an arbitrary basis $\mathbf{e}_k$, $k = 1, \dots, 4$ the components of $\mathbf{v}$ with respect to this basis are

$$v = \text{Re}(E^H E)^{-1} \text{Re}(E^H \mathbf{v}) \quad (\text{A.148})$$

where $E = [\mathbf{e}_1 \ \mathbf{e}_2 \ \mathbf{e}_3 \ \mathbf{e}_4]$.

A.6.2 2 by 1 complex vectors over the Complex (C^2, C)

We now consider the same set V with scalar field given by the set of complex numbers, that is $\mathcal{F} \equiv C$. For this choice $\mathbf{e}_1$ and $\mathbf{e}_3$ defined above (A.141) *are* linearly dependent since

$$\mathbf{e}_1 = -i \cdot \mathbf{e}_3, \quad i \in C . \quad (\text{A.149})$$

The elements $\mathbf{e}_2$ and $\mathbf{e}_4$ are dependent in the same manner. It is easily verified that a basis for this complex vector space is given by the pair $\mathbf{e}_1$ and $\mathbf{e}_2$; the space said to have complex dimension of 2. The arbitrary element $\mathbf{v}$ (A.142) has the representation

$$\mathbf{v} = \bar{v}^1 \cdot \mathbf{e}_1 + \bar{v}^2 \cdot \mathbf{e}_2 \quad (\text{A.150})$$

where the scalar components (elements of $\mathcal{F} \equiv C$) are

$$\bar{v}^1 = (-ia + b), \quad \bar{v}^2 = (-ia + b) . \quad (\text{A.151})$$

Again, for convenience, we arrange the basis vectors in the matrix form

$$\bar{E} \equiv \begin{bmatrix} \mathbf{e}_1 & \mathbf{e}_2 \end{bmatrix} = I \quad (\text{A.152})$$

so that sum (A.150) can be expressed as

$$\mathbf{v} = \bar{E} \bar{v} . \quad (\text{A.153})$$

The standard inner product (A.140) for this complex space is

$$\langle \mathbf{w}, \mathbf{v} \rangle \equiv \mathbf{w}^H \mathbf{v} = \bar{w}^H \bar{E}^H \bar{E} \bar{v} . \quad (\text{A.154})$$

For an arbitrary basis $\mathbf{e}_k$, $k = 1, 2$ the components of $\mathbf{v}$ with respect to this basis are

$$\bar{v} = (\bar{E}^H \bar{E})^{-1} (\bar{E}^H \mathbf{v}) \quad (\text{A.155})$$

where $\bar{E} = [\mathbf{e}_1 \ \mathbf{e}_2]$.

A.7 The space (C^N, R) for N even

The set of all complex N -tuples, C^N for N even maybe represented as the real vector space (C^N, R) , of dimension N , or the complex vector space (C^N, C) , of dimension N . The standard basis for $V \equiv (C^N, R)$, N even, can be expressed in the matrix form as

$$E = \begin{bmatrix} \bar{E} & i \cdot \bar{E} \end{bmatrix} \quad (\text{A.156})$$

where $\bar{E} = I_N$. An arbitrary element $\mathbf{v}$ of V can be represented with respect to this basis as

$$\mathbf{v} = E v \quad (\text{A.157})$$

for some real $2N$ -tuple v . Anticipating the discussion to follow, this real tuple is expressed in partitioned form as

$$v = \begin{bmatrix} v_R \\ v_I \end{bmatrix} . \quad (\text{A.158})$$

The standard basis for (C^N, C) in matrix form is $\bar{E} = I_N$. The same element $\mathbf{v}$ can be represented with respect to this basis as

$$\mathbf{v} = \bar{E} \bar{v} \quad (\text{A.159})$$

where $\bar{v}$ is the complex N -tuple given by

$$\bar{v} = v_R + i \cdot v_I . \quad (\text{A.160})$$

A.8 Subspaces of (C^N, R) and (C^N, C)

In this section we consider subsets of C^N , N even, which are particular vector subspaces of (C^N, R) . Let V be a subspace of (C^N, R) of dimension $2q \leq N$ defined by the span of the set E of the form

$$E = \begin{bmatrix} \bar{E} & i \cdot \bar{E} \end{bmatrix} \quad (\text{A.161})$$

where $\bar{E} = [\mathbf{e}_1, \dots, \mathbf{e}_q]$ is a linearly independent set. This same set V may be represented as a q dimensional subspace of (C^N, C) . In this formulation V is represented by the span of the basis $\bar{E}$, where $\bar{E}$ is related the basis E appearing in the real formulation by (A.161).

An arbitrary vector $\mathbf{v} \in V$ can be expressed in the (C^N, R) formulation as

$$E \cdot v = \begin{bmatrix} \bar{E} & i \cdot \bar{E} \end{bmatrix} \begin{bmatrix} v_R \\ v_I \end{bmatrix}. \quad (\text{A.162})$$

On the right hand side of this expression the real $2q$ -tuple v has been partitioned into two q -tuples as

$$v = \begin{bmatrix} v_R \\ v_I \end{bmatrix}. \quad (\text{A.163})$$

Carrying out the multiplication yields

$$E \cdot v = \bar{E}v_R + i \cdot \bar{E}v_I = \bar{E}(v_R + i \cdot v_I) = \bar{E} \cdot \bar{v} \quad (\text{A.164})$$

where

$$\bar{v} = v_R + i \cdot v_I. \quad (\text{A.165})$$

Thus, in the complex formulation $\mathbf{v} = \bar{E} \cdot \bar{v}$, and so V is complex dimension 2. Given complex components $\bar{v}$, the real component v are found as

$$v \equiv \text{Re}(E^H \bar{v}) = \begin{bmatrix} \text{Re}(\bar{v}) \\ \text{Im}(\bar{v}) \end{bmatrix}. \quad (\text{A.166})$$

A.8.1 Example: Subsets of C^4 that are vector spaces in (C^4, R)

Let $V \subset C^4$ be defined by the span of the following set of vectors in (C^4, R)

$$\mathbf{e}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}; \mathbf{e}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}; \mathbf{e}_3 = \begin{bmatrix} i \\ 0 \\ 0 \\ 0 \end{bmatrix}; \mathbf{e}_4 = \begin{bmatrix} 0 \\ i \\ 0 \\ 0 \end{bmatrix}; \quad (\text{A.167})$$

For convenience we arrange the basis vectors in the matrix form

$$E \equiv \begin{bmatrix} \mathbf{e}_1 & \mathbf{e}_2 & \mathbf{e}_3 & \mathbf{e}_4 \end{bmatrix} = \begin{bmatrix} I_{4 \times 2} & i \cdot I_{4 \times 2} \end{bmatrix} \quad (\text{A.168})$$

By construction, this is a 4-dimensional subspace of the 8-dimensional space (C^4, R) .

Observe that in the (C^4, C) formulation, $\mathbf{e}_1 = i \cdot \mathbf{e}_3$ and $\mathbf{e}_2 = i \cdot \mathbf{e}_4$. This same subset V can therefore be represented as a 2-dimensional space in (C^4, C) formulation defined by the span of the basis

$$\bar{E} \equiv \begin{bmatrix} \mathbf{e}_1 & \mathbf{e}_2 \end{bmatrix}. \quad (\text{A.169})$$

To verify this, we need to show that an arbitrary $\mathbf{v} \in V$ can be expressed as a linear combination of $\mathbf{e}_1$ and $\mathbf{e}_2$ when the scalar field is complex, $\mathcal{F} = C$. Suppose that $\mathbf{v}$ is represented in (C^4, R) as

$$\mathbf{v} = E \cdot v \quad (\text{A.170})$$

where $v = \begin{bmatrix} a & b & c & d \end{bmatrix}^t$. Then, choosing complex scalar components

$$\bar{v}^1 = (-ia + c), \quad \bar{v}^2 = (-ib + d) \quad (\text{A.171})$$

we have the representation of $\mathbf{v}$ in (V, C) as

$$\mathbf{v} = \bar{E} \cdot \bar{v}. \quad (\text{A.172})$$

This confirms V is a 2-dimensional subspace of the 4-dimensional space (C^4, C) .

This relation between the dimension of the real and complex vector space representations does not hold for arbitrarily defined subspaces. To see this, consider the subset

$W \subset C^4$ be defined by the span of

$$\mathbf{e}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}; \mathbf{e}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}; \mathbf{e}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}; \mathbf{e}_4 = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}. \quad (\text{A.173})$$

By construction W is a vector subspace of (C^4, R) . Unlike the previous case (A.167), this set basis vectors is also linearly independent in the complex vector space formulation.

A.9 Transforming a Basis for (W, R)

Let $\mathbf{w}_k$ and $\mathbf{v}_k$ be two basis sets for (W, R) and suppose that they are related by the real transformation T

$$\mathbf{w}_j = \sum_{k=1}^n \mathbf{v}_k T_k^j. \quad (\text{A.174})$$

Using the matrix form with

$$E_v \equiv \begin{bmatrix} \mathbf{v}_1 & \cdots & \mathbf{v}_n \end{bmatrix} \quad (\text{A.175})$$

and

$$E_w \equiv \begin{bmatrix} \mathbf{w}_1 & \cdots & \mathbf{w}_n \end{bmatrix} \quad (\text{A.176})$$

we have

$$E_w = E_v \cdot T. \quad (\text{A.177})$$

The real coefficient matrix T then is

$$T = \text{Re}(E_v^H E_v)^{-1} \text{Re}(E_v^H E_w). \quad (\text{A.178})$$

Note that

$$\langle \mathbf{w}_k, \mathbf{v}_m \rangle = \text{Re}(E_w^H(:, k) E_v(:, m)). \quad (\text{A.179})$$

Let $\mathbf{x} \in (V, R)$ have the representations given below with respect to the basis sets $\mathbf{v}_k$ and $\mathbf{w}_k$,

$$\mathbf{x} = \sum_{k=1}^n \mathbf{v}_k v^k = \sum_{k=1}^n \mathbf{w}_k w^k. \quad (\text{A.180})$$

The transformation of the basis vectors (A.177) defines a transformation of the components. Using (A.177) we have

$$\mathbf{x} = E_v T T^{-1} v = E_w w \quad (\text{A.181})$$

so that

$$w = T^{-1} v . \quad (\text{A.182})$$

A.10 Orthogonalizing a Basis

A basis $\mathbf{v}_i$ is orthonormal with respect to the inner product $\langle, \rangle$ (A.147) if

$$\langle \mathbf{v}_i, \mathbf{v}_j \rangle = \delta_{ij}; \quad i = 1, \dots, N; j = 1, \dots, N . \quad (\text{A.183})$$

Using (A.175) this relation is expressed in matrix form as

$$\text{Re}(E_v^H E_v) = I . \quad (\text{A.184})$$

Given a non-orthogonal basis $\mathbf{w}_k$ the real matrix $\mathbf{J}$ that orthogonalizes this set is the real symmetric matrix

$$\mathbf{J} = (\text{Re}(E_w^H E_w))^{-\frac{1}{2}} . \quad (\text{A.185})$$

To verify this let $E_v = E_w \mathbf{J}$, then

$$\text{Re}(E_v^H E_v) = \mathbf{J}^t \text{Re}(E_w^H E_w) \mathbf{J} = I . \quad (\text{A.186})$$

Note that $\text{Re}(E_w^H(:, i) E_w(:, j)) = \langle \mathbf{w}_i, \mathbf{w}_j \rangle$ so that in matrix form

$$\text{Re}(E_w^H E_w) = \begin{bmatrix} \langle \mathbf{w}_1, \mathbf{w}_1 \rangle & \cdots & \langle \mathbf{w}_1, \mathbf{w}_N \rangle \\ \vdots & \ddots & \vdots \\ \langle \mathbf{w}_N, \mathbf{w}_1 \rangle & \cdots & \langle \mathbf{w}_N, \mathbf{w}_N \rangle \end{bmatrix} . \quad (\text{A.187})$$

A.11 Specifics of the Tangent Space Problem

A.11.1 N by p complex matrices over the Reals

The set of $N \times p$ complex matrices denoted $C^{N \times p}$, along with the usual matrix addition and scalar multiplication operations defines a vector space. Typically, this space

is considered as a vector space with the scalar field as the set complex numbers, $\mathcal{F} = \mathbb{C}$. Instead, here we consider the scalar field to be the set of reals, $\mathcal{F} = \mathbb{R}$. The vector space $(\mathbb{C}^{N \times p}, \mathbb{R})$ has real dimension $2Np$. The standard inner product space on this space is

$$\langle \mathbf{v}, \mathbf{w} \rangle = \text{Re}(\text{tr}(V^H W)) \quad (\text{A.188})$$

where V and W are the $N \times p$ complex matrices corresponding to the symbols $\mathbf{v}$ and $\mathbf{w}$.

The standard basis for this space is given by the set

$$\mathbf{e}_k = \text{reshape}(e_k, N, p); \quad k = 1, \dots, Np \quad (\text{A.189})$$

$$\mathbf{e}_k = i \cdot \text{reshape}(e_k, N, p); \quad k = Np + 1, \dots, 2Np \quad (\text{A.190})$$

where e_k is a unit $2Np$ -tuple with 1 at the k -th location.

An arbitrary complex matrix $\mathbf{w} \in \mathbb{C}^{N \times p}$ is represented with respect to this basis as

$$\mathbf{w} = \sum_k^{2Np} \mathbf{e}_k w^k \quad (\text{A.191})$$

where

$$w^k = \langle \mathbf{e}_k, \mathbf{w} \rangle. \quad (\text{A.192})$$

If $\mathbf{w}$ represents the matrix W then (A.192) evaluates to

$$w = \begin{bmatrix} \text{Re}(\text{vec}(W)) \\ \text{Im}(\text{vec}(W)) \end{bmatrix}. \quad (\text{A.193})$$

A.11.2 Partitioning the space $\mathbb{C}^{N \times p}$

The SVD of an arbitrary matrix $X \in \mathbb{C}^{N \times p}$ may be expressed in the general partitioned form as

$$X = \begin{bmatrix} Y & Y_\perp \end{bmatrix} \begin{bmatrix} D & 0 \\ 0 & D_\perp \end{bmatrix} \begin{bmatrix} V^H \\ V_\perp^H \end{bmatrix}. \quad (\text{A.194})$$

In this expression Y and $Y_\perp$ are full rank semi-unitary matrices of size $N \times p$ and $(N - p) \times p$, respectively, and $Y^H Y_\perp = 0$. Expanding the above yields

$$X = \begin{bmatrix} Y & Y_\perp \end{bmatrix} \begin{bmatrix} C \\ B \end{bmatrix} = YC + Y_\perp B \quad (\text{A.195})$$

where C is complex $(p \times p)$ given by $C = DV^H$ and B is complex $((N - p) \times p)$ given by $B = D_\perp V^H$. Expressing the $(p \times p)$ matrix C as the sum of Hermitian (S) and anti-Hermitian(A) components as $C = A + S$, yields

$$X = YS + YA + Y_\perp B . \quad (\text{A.196})$$

This partitions the space into three mutually orthogonal spaces since

$$\langle YA, YS \rangle = \text{tr}(A^H Y^H YS) = \text{tr}(A^H S) = 0 \quad (\text{A.197})$$

for any for S Hermitian and A anti-Hermitian.

The three spaces are defined by collections of the form YA, YS and $Y_\perp B$ are denoted A_Y, N_Y and B_Y , respectively. The space $(C^{N \times p}, R)$ then is decomposed as

$$C^{N \times p} = A_Y \oplus B_Y \oplus N_Y . \quad (\text{A.198})$$

In particular, the B_Y space is the collection of all $N \times p$ matrices of the form $Y_\perp B$, where B is an $(N - p) \times p$ arbitrary complex matrix. This space has real dimension $N_b \equiv 2p(N - p)$ and is a subspace of the full space, that is $B_Y < C^{N \times p}$.

A.11.3 Algebra of the B_Y space

The inner product on B_Y is defined by the restriction of the inner product (A.188) defined on $C^{N \times p}$. It is denoted $\langle : , : \rangle_{B_Y}$ and is given by

$$\langle \mathbf{v}, \mathbf{w} \rangle_{B_Y} \equiv \langle P_B \mathbf{v}, P_B \mathbf{w} \rangle \quad (\text{A.199})$$

where P_B is the orthogonal projection $P_B : C^{N \times p} \mapsto B_Y$. For $\mathbf{v}, \mathbf{w} \in B_Y$ given by $\mathbf{v} = Y_\perp V$ and $\mathbf{w} = Y_\perp W$, the inner product is

$$\langle \mathbf{v}, \mathbf{w} \rangle_{B_Y} = \text{Re}(\text{tr}(V^H Y_\perp^H P_B Y_\perp W)) = \text{Re}(\text{tr}(V^H W)) \quad (\text{A.200})$$

where $P_B \equiv Y_\perp Y_\perp^H$.

Let $\mathbf{b}_j = Y_\perp B_j$, $j = 1 \cdots N_b$, be an orthonormal basis for B_Y . Then any $\Delta \in B_Y$ expands as

$$\Delta = \sum_{k=1}^{N_b} \mathbf{b}_k \beta^k = \sum_{k=1}^{N_b} Y_\perp B_k \beta^k = Y_\perp \sum_{k=1}^{N_b} B_k \beta^k. \quad (\text{A.201})$$

The components of Δ with respect to this orthonormal basis are computed in the usual way

$$\beta^k = \langle \mathbf{b}_k, \Delta \rangle \quad (\text{A.202})$$

or

$$\beta^k = \text{Re} \left(\text{tr} \left(B^k Y_\perp^H \Delta \right) \right). \quad (\text{A.203})$$

Defining the matrix B as

$$B \equiv \sum_{k=1}^{N_b} B_k \beta^k \quad (\text{A.204})$$

yields the representation $\Delta = Y_\perp B$. Vectorizing both sides of this sum yields the alternate matrix form

$$b = E_b \beta \quad (\text{A.205})$$

where $b = \text{vec}(B)$ and

$$E_b = \begin{bmatrix} \text{vec}(B_1) & \cdots & \text{vec}(B_{N_b}) \end{bmatrix}. \quad (\text{A.206})$$

Using this expression, (A.203) is expressed in matrix form as

$$\beta = \text{Re}(E_b^H b). \quad (\text{A.207})$$

The standard basis for B_Y is defined by the set $\mathbf{b}_k = Y_\perp B_k$; $k = 1 \cdots N_b$, $N_b = 2p(N - p)$, where the each matrix B_k has 1 or i in the k th location and zeros elsewhere.

For the case $N = 5, p = 2, N_b = 12$ and the standard basis is defined by the set

$$B_1 = \begin{bmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}, \quad \dots, B_6 = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 1 \end{bmatrix}, \quad B_7 = i \cdot B_1, \quad \dots, B_{12} = i \cdot B_1, \quad (\text{A.208})$$

In this case, the matrix E_b appearing in (A.206) is

$$E_b = \begin{bmatrix} I_{N_b} & iI_{N_b} \end{bmatrix} \quad (\text{A.209})$$

where $\bar{N}_b = \frac{N_b}{2}$. Then with respect to the standard basis (A.207) evaluates to

$$\beta = \text{Re} \left(\begin{bmatrix} I_{\bar{N}_b} \\ -iI_{\bar{N}_b} \end{bmatrix} b \right) = \begin{bmatrix} \text{Re}(b) \\ \text{Im}(b) \end{bmatrix}. \quad (\text{A.210})$$

The inverse relation (A.205) evaluates to

$$b = \begin{bmatrix} I_{\bar{N}_b} & iI_{\bar{N}_b} \end{bmatrix} \beta = \beta_R + i \cdot \beta_I \quad (\text{A.211})$$

where β_R and β_I represent the first $\bar{N}_b$ and last $\bar{N}_b$ elements of β , that is

$$\beta = \begin{bmatrix} \beta_R \\ \beta_I \end{bmatrix}. \quad (\text{A.212})$$

Change of basis

Let $\acute{\mathbf{b}}_j$ be an alternate basis for B_Y . This alternate basis expands in the standard orthonormal basis $\mathbf{b}_k$ as

$$\acute{\mathbf{b}}_k = \sum T_k^j \mathbf{b}_j; \quad k = 1 \cdots N_b. \quad (\text{A.213})$$

If these basis vectors are represented in numerical form as $\mathbf{b}_k = Y_{\perp} B_k$ and $\acute{\mathbf{b}}_k = Y_{\perp} \acute{B}_k$ then (A.213) is

$$Y_{\perp} \acute{B}_k = \sum T_k^j Y_{\perp} B_j. \quad (\text{A.214})$$

Therefore

$$\acute{B}_k = \sum T_k^j B_j. \quad (\text{A.215})$$

Vectorizing both sides of (A.215) yields

$$\acute{E}_b = E_b T \quad (\text{A.216})$$

where E_b is given by (A.206) and where $\acute{E}_b = \begin{bmatrix} \text{vec}(\acute{B}_1) & \cdots & \text{vec}(\acute{B}_{N_b}) \end{bmatrix}$.

Let $\Delta = Y_{\perp} B$ be an arbitrary vector that is represented with respect to the $\mathbf{b}_k$ basis as $\Delta = \sum \mathbf{b}_k \beta^k$. Then, referring to (A.205)

$$b = E_b \beta = E_b T T^{-1} \beta = \acute{E}_b \acute{\beta} \quad (\text{A.217})$$

where $b = \text{vec}(B)$, so that the components of Δ transform as

$$\beta' = T^{-1}\beta. \quad (\text{A.218})$$

The real transformation matrix appearing in (A.213) is defined by

$$T_j^i = \langle \mathbf{b}_i, \mathbf{b}'_j \rangle \quad (\text{A.219})$$

and using the concrete form of the vectors $\mathbf{b}_i$ and $\mathbf{b}'_j$

$$T_j^i = \text{Re}(\text{tr}(B_i^H Y_\perp^H Y_\perp \dot{B}_j)) = \text{Re}(\text{tr}(B_i^H \dot{B}_j)). \quad (\text{A.220})$$

In matrix form, the above is

$$T = \text{Re}(E_b^H \dot{E}_b) \quad (\text{A.221})$$

which is equivalent to $T_j^i = \text{Re}(E_b(:, i)^H \dot{E}_b(:, j))$.

A.12 Partitioned space and basis

In the signal processing applications encountered later, the space B_Y is partitioned into two orthogonal subspaces

$$B_Y = S_Y \oplus N_Y \quad (\text{A.222})$$

where $\dim(S_Y) = q$ and $\dim(N_Y) = \dim(B_Y) - \dim(S_Y) = N_b - q$. Let $\mathbf{s}_k = Y_\perp S_k$ and $\mathbf{n}_k = Y_\perp N_k$ be given orthonormal basis vectors for S_Y and N_Y , respectively. These basis elements are related to the basis $\mathbf{b}_k$ by

$$\mathbf{s}_k = \sum_{m=1}^{N_b} \mathbf{b}_m [T_s]_k^m; \quad k = 1 \cdots N_s, \quad (\text{A.223})$$

and

$$\mathbf{n}_k = \sum_{m=1}^{N_b} \mathbf{b}_m [T_n]_k^m \quad k = 1 \cdots N_n. \quad (\text{A.224})$$

The real $N_b \times q$ matrix T_s is

$$[T_s]_j^i = \langle \mathbf{b}_i, \mathbf{s}_j \rangle; \quad i = 1 \cdots N_b, \quad j = 1 \cdots q \quad (\text{A.225})$$

and the real $N_b \times (N_b - q)$ matrix T_n is

$$[T_n]_j^i = \langle \mathbf{b}_i, \mathbf{n}_j \rangle; \quad i = 1 \cdots N_b, \quad j = 1 \cdots N_b - q. \quad (\text{A.226})$$

In matrix form (A.223) is

$$Y_\perp S_k = Y_\perp \sum_{m=1}^{N_s} B_k [T_s]_k^m; \quad k = 1 \cdots N_s \quad (\text{A.227})$$

and therefore

$$S_k = \sum_{m=1}^{N_s} B_k [T_s]_k^m; \quad k = 1 \cdots N_s. \quad (\text{A.228})$$

Vectorizing both sides of the above yields the matrix form

$$E_s = E_b T_s \quad (\text{A.229})$$

where we have used (A.206) and

$$E_s \equiv \begin{bmatrix} \text{vec}(S_1) & \cdots & \text{vec}(S_q) \end{bmatrix}. \quad (\text{A.230})$$

Likewise, $E_n = E_b T_n$, so that in full matrix form

$$\begin{bmatrix} E_s & E_n \end{bmatrix} = E_b \begin{bmatrix} T_s & T_n \end{bmatrix} \quad (\text{A.231})$$

or

$$\begin{bmatrix} E_s & E_n \end{bmatrix} = E_b T \quad (\text{A.232})$$

where $T \equiv \begin{bmatrix} T_s & T_n \end{bmatrix}$.

The matrix T_s (A.229) may be expressed in terms of the matrices E_b and E_s as

$$T_s = \text{Re}(E_b^H E_s). \quad (\text{A.233})$$

Similarly, T_n is given by

$$T_n = \text{Re}(E_b^H E_n). \quad (\text{A.234})$$

Let $\Delta = Y_\perp B$ be an arbitrary vector that expands as $\Delta = \sum_{k=1}^{N_b} \mathbf{b}_k \beta^k$. Then with respect to this partitioned basis the form (A.205) becomes

$$b = E_b T T^{-1} \beta = \begin{bmatrix} E_s & E_n \end{bmatrix} \begin{bmatrix} s \\ n \end{bmatrix} \quad (\text{A.235})$$

where $b = \text{vec}(B)$. The real N_b -tuple β transform as

$$\begin{bmatrix} s \\ n \end{bmatrix} = T^{-1}\beta = \begin{bmatrix} X_s^t \\ X_n^t \end{bmatrix} \beta \quad (\text{A.236})$$

or

$$s = X_s^t \beta \quad (\text{A.237})$$

$$n = X_n^t \beta \quad (\text{A.238})$$

where for convenience we have defined

$$\begin{bmatrix} X_s^t \\ X_n^t \end{bmatrix} \equiv \begin{bmatrix} T_s & T_n \end{bmatrix}^{-1}. \quad (\text{A.239})$$

Note that

$$P = TT^{-1} = \begin{bmatrix} T_s & T_n \end{bmatrix} \begin{bmatrix} X_s^t \\ X_n^t \end{bmatrix} = T_s X_s^t + T_n X_n^t = I \quad (\text{A.240})$$

$$X_s^t = (T_s^t T_s)^{-1} T_s^t \quad (\text{A.241})$$

$$X_n^t = (T_n^t T_n)^{-1} T_n^t. \quad (\text{A.242})$$

The matrix representation of orthogonal projectors $P_S : B_Y \mapsto S_Y$ and $P_N : B_Y \mapsto S_Y$ with respect to this set of basis vectors is

$$P_S = T_s X_s^t = T_s (T_s^t T_s)^{-1} T_s^t, \quad (\text{A.243})$$

$$P_N = T_n X_n^t = T_n (T_n^t T_n)^{-1} T_n^t. \quad (\text{A.244})$$

Note that if $\mathbf{s}_k, \mathbf{n}_k$ and $\mathbf{b}_k$ are all orthonormal sets then $(T_s^t T_s) = I_q$ and $(T_n^t T_n) = I_{N_b-q}$ and hence $X_s = T_s$ and $X_n = T_n$.

For completeness, note that from (A.236)

$$\beta = \begin{bmatrix} T_s & T_n \end{bmatrix} \begin{bmatrix} s \\ n \end{bmatrix}. \quad (\text{A.245})$$

A.12.1 Tensor Transform in partitioned frame

In this section, and throughout the remainder, the subscripts β, α, s and n are not indices, but are used to denote sub-matrices of a larger matrix as shown below. We use i, j, k, l, m as indices unless otherwise noted.

Let $\mathbf{L}$ be a covariant 2-tensor (see section A.3.2), that is, $\mathbf{L}$ is a function defined on

$$\mathbf{L} : B_Y \times B_Y \rightarrow R . \quad (\text{A.246})$$

Denote the coefficient matrix with respect to the basis $\mathbf{b}_i$ for B_Y as $L_{\beta\beta}$, so that

$$[L_{\beta\beta}]_{ij} \equiv \mathbf{F}[\mathbf{b}_i, \mathbf{b}_j] . \quad (\text{A.247})$$

Recall that $B_Y = S_Y \oplus N_Y$ and the transformation (A.84) to the partitioned form is

$$\begin{bmatrix} L_{ss} & L_{sn} \\ L_{ns} & L_{nn} \end{bmatrix} \equiv \begin{bmatrix} T_s^t \\ T_n^t \end{bmatrix} L_{\beta\beta} \begin{bmatrix} T_s & T_n \end{bmatrix} , \quad (\text{A.248})$$

$$L_{ss} \equiv T_s^t L_{\beta\beta} T_s , \quad (\text{A.249})$$

$$L_{nn} \equiv T_n^t L_{\beta\beta} T_n , \quad (\text{A.250})$$

$$L_{sn} \equiv T_s^t L_{\beta\beta} T_n . \quad (\text{A.251})$$

The quadratic form $\beta^t L_{\beta\beta} \beta$ is expressed in terms of the partitioned basis as

$$\beta^t L_{\beta\beta} \beta = \beta^t T^{-t} T^t L_{\beta\beta} T T^{-1} \beta = \begin{bmatrix} s^t & n^t \end{bmatrix} \begin{bmatrix} L_{ss} & L_{sn} \\ L_{ns} & L_{nn} \end{bmatrix} \begin{bmatrix} s \\ n \end{bmatrix} \quad (\text{A.252})$$

where we have used (A.236).

Special Form

If the real coefficient matrix $L_{\beta\beta}$ has the special form

$$L_{\beta\beta} = \text{Re}(E_b^H \tilde{\Lambda}_{bb} E_b) \quad (\text{A.253})$$

where $\tilde{\Lambda}_{bb}$ is a $\bar{N}_b \times \bar{N}_b$ real diagonal matrix, then the above sub-matrices evaluate to

$$L_{ss} \equiv T_s^t L_{\beta\beta} T_s = T_s^t \text{Re}(E_b^H \tilde{\Lambda}_{bb} E_b) T_s = \text{Re}(E_s^H \tilde{\Lambda}_{bb} E_s) , \quad (\text{A.254})$$

$$L_{nn} \equiv T_n^t L_{\beta\beta} T_n = \text{Re}(E_n^H \tilde{\Lambda}_{bb} E_n) , \quad (\text{A.255})$$

$$L_{sn} \equiv T_s^t L_{\beta\beta} T_n = \text{Re}(E_s^H \tilde{\Lambda}_{bb} E_n) . \quad (\text{A.256})$$

In the above expressions we have used $E_s = E_b T_s$ and $E_n = E_b T_n$ (A.232).

A sum of the form $L_s(s, n) \equiv L_{ss}s + L_{sn}n$ appears later in Newton's method. In this expression s and n are real tuples. Using (A.254) and (A.256) this sum, which we denote $L_s(s, n)$, is given by

$$L_s(s, n) = \text{Re}(E_s^H \tilde{\Lambda}_{bb} E_s)s + \text{Re}(E_s^H \tilde{\Lambda}_{bb} E_n)n , \quad (\text{A.257})$$

$$= \text{Re}(E_s^H \tilde{\Lambda}_{bb} (E_s s + E_n n)) , \quad (\text{A.258})$$

$$= \text{Re}(E_s^H \tilde{\Lambda}_{bb} b) , \quad (\text{A.259})$$

where b is the complex $\bar{N}_b$ -tuple defined by $b = E_s s + E_n n$ (A.235).

Alternate Complex form of Quadratic

Furthermore, since in the complex case under consideration $N_b = \dim(B_Y) = 2p(N - p)$ the basis matrix has the form

$$E_b = \begin{bmatrix} \bar{E}_b & i \cdot \bar{E}_b \end{bmatrix} . \quad (\text{A.260})$$

Using this (A.253) becomes

$$L_{\beta\beta} = \text{Re} \left(\begin{bmatrix} \bar{E}_b^H \\ -i \bar{E}_b^H \end{bmatrix} \tilde{\Lambda}_{bb} \begin{bmatrix} \bar{E}_b & i \bar{E}_b \end{bmatrix} \right) , \quad (\text{A.261})$$

$$= \text{Re} \left(\begin{bmatrix} L_{bb} & i \cdot L_{bb} \\ -i \cdot L_{bb} & L_{bb} \end{bmatrix} \right) , \quad (\text{A.262})$$

where

$$L_{bb} \equiv (\bar{E}_b^H \tilde{\Lambda}_{bb} \bar{E}_b) . \quad (\text{A.263})$$

For the real N_b -tuple β partitioned as $\beta = \begin{bmatrix} \beta_R \\ \beta_I \end{bmatrix}$ the quadratic form $\beta^t L_{\beta\beta} \beta$ then is

$$\beta^t L_{\beta\beta} \beta = \text{Re} \left((\beta_R^t L_{bb} \beta_R + \beta_I^t L_{bb} \beta_I) + i \cdot (\beta_I^t L_{bb} \beta_R + \beta_R^t L_{bb} \beta_I) \right) . \quad (\text{A.264})$$

Using $b = E_0\beta = \beta_R + i \cdot \beta_I$ this can be expressed equivalently as

$$\beta^t L_{\beta\beta} \beta = \text{Re} \left(b^H L_{bb} b \right) . \quad (\text{A.265})$$

Remark: For the standard basis $\bar{E}_b = I_{\bar{N}_b}$ so that (A.263) simplifies to the real diagonal matrix

$$L_{bb} \equiv \left(\tilde{\Lambda}_{bb} \right) , \quad (\text{A.266})$$

from which it follows that (607)

$$L_{\beta\beta} = I_2 \otimes L_{bb} = \begin{bmatrix} \tilde{\Lambda}_{bb} & 0 \\ 0 & \tilde{\Lambda}_{bb} \end{bmatrix} . \quad (\text{A.267})$$

List of References

- [1] M. Spivak, *A Comprehensive Introduction to Differential Geometry*, 3rd ed. Houston, TX: Publish or Perish, 1999, vol. 1.
- [2] A. Boumal, “On intrinsic cramér-rao bounds for riemannian submanifolds and quotient manifolds,” *IEEE Transactions on Signal Processin*, vol. 61, no. 7, April 2013.
- [3] S. Smith, “Covariance, subspace, and intrinsic cramér-rao bounds,” *IEEE Trans. Signal Process.*, vol. 53, no. 5, p. 1610 1630, 2005.
- [4] A.Edelman, T. A.Arias, and S. Smith, “The geometry of algorithms with orthogonality constraints,” *SIAM Matrix Anal. Appl*, vol. 20, no. 2, p. 303353, 1998.

APPENDIX B

appendix B

B.1 FIM appendix

B.1.1 Proof of Lemma 8

Using the binomial expansion identity $\frac{1}{1+x} = 1 - x + x^2 + h.o.t.$ in the form

$$(R + \delta R)^{-1} = R^{-1}(I - (\delta R R^{-1}) + (\delta R R^{-1})^2 - \dots) \quad (\text{B.1})$$

the first term on the right hand side of log likelihood function (342) is

$$\text{tr}((R + \delta R)^{-1} \tilde{R}) = \text{tr}(R^{-1} \tilde{R}) - \text{tr}(\delta R R^{-1} \tilde{R} R^{-1}) + \text{tr}((\delta R R^{-1})^2 \tilde{R} R^{-1}) + h.o.t. \quad (\text{B.2})$$

Using the identity $\det(I + A) = 1 + \text{tr}(A) + h.o.t.$, the log $\det(R + \delta R)$ term in the log likelihood function (342) is expressed as

$$\det(R + \delta R) = \det\left(R(I + R^{-1}\delta R)\right) \quad (\text{B.3})$$

$$= \det(R)(1 + \text{tr}(R^{-1}\delta R) + h.o.t.) . \quad (\text{B.4})$$

Taking the natural log yields

$$= \log(\det(R)) + \log\left((1 + \text{tr}(R^{-1}\delta R) + h.o.t.)\right) \quad (\text{B.5})$$

Using the identity $\log(1 + x) = x - \frac{x^2}{2} + \frac{x^3}{3} + h.o.t.$ the above is

$$\log(\det(R + \delta R)) = \log(\det(R)) + \text{tr}(R^{-1}\delta R) - \frac{1}{2}\text{tr}(R^{-1}\delta R)^2 + h.o.t. \quad (\text{B.6})$$

Adding (B.2) and (B.6) yields the desired result (407)

$$\delta L(\delta R, R) \equiv -K \left(\text{tr}(\delta R R^{-1}(I - \tilde{R} R^{-1})) + \text{tr}((\delta R R^{-1})^2(\tilde{R} R^{-1} - \frac{1}{2}I)) + \dots \right) . \quad (\text{B.7})$$

In addition, since $E[\tilde{R}] = R$ we have $E[I - \tilde{R} R^{-1}] = 0$ so that

$$E\left[\tilde{R} R^{-1} - \frac{1}{2}I\right] = \frac{1}{2}I . \quad (\text{B.8})$$

Using these two expectations along with (B.7) yields the second desired result (408).

$$E[\delta L] = -\frac{K}{2} \left(\text{tr}(\delta R R^{-1})^2 + \dots \right) \quad (\text{B.9})$$

B.1.2 Proof of Lemma 9

The model covariance R evaluated along the geodesic curve $Y(t)$ is

$$R(Y(t)) = Y(t)\Lambda Y^H(t) + \sigma^2 I . \quad (\text{B.10})$$

Substituting the geodesic in the expanded form given by

$$Y(t) = \exp_Y(t\Delta) = Y + \delta Y(t) \quad (\text{B.11})$$

into (B.10) yields

$$R(Y + \delta Y(t)) = R(0) + \delta R(A, B; t) \quad (\text{B.12})$$

where

$$\delta R(A, B; t) = \delta Y(t)\Lambda Y^H + Y\Lambda\delta Y^H(t) + \delta Y(t)\Lambda\delta Y^H(t) . \quad (\text{B.13})$$

In this form $\delta R(A, B; t)$ represents the change in R along the geodesic $Y(t)$. Since $Y(t)$ is determined by its tangent vector $\Delta = Y A + Y_\perp B$ and the path parameter t , we have used the notation $\delta R(A, B; t)$. The term $\delta Y(t)$ in (B.11) is given by (A.40) as

$$\delta Y(t) = t(Y A + Y_\perp B) + \frac{1}{2}t^2(Y(A^2 - B^H B) + Y_\perp B A) + h.o.t \quad (\text{B.14})$$

Substituting this into (B.13)

$$\delta R(A, B; t) = t\delta R_1(A, B) + \frac{t^2}{2}(\delta R_{2,1}(A, B) + \delta R_{2,2}(A, B)) + h.o.t \quad (\text{B.15})$$

where the first-order term is

$$\delta R_1(A, B) = Y A \Lambda Y^H + Y \Lambda A^H Y^H + Y_\perp B \Lambda Y^H + Y \Lambda B^H Y_\perp^H \quad (\text{B.16})$$

and the second-order terms are

$$\delta R_{2,1}(A, B) = Y A \Lambda A^H Y^H + Y_\perp B \Lambda A^H Y^H + Y_\perp B \Lambda B^H Y_\perp^H \quad (\text{B.17})$$

$$\delta R_{2,2}(A, B) = Y A^2 \Lambda Y^H + Y \Lambda A^2 Y^H - \frac{1}{2}Y B^H B \Lambda Y^H + Y_\perp B A Y^H + Y \Lambda Y^H B^H Y_\perp^H \quad (\text{B.18})$$

Since differentiation and expectation commute using the form (406) we have

$$E \left[\frac{d^2 L(Y(t))}{dt^2} \Big|_{t=0} \right] = E \left[\frac{d^2 \delta L(t)}{dt^2} \right] = \frac{d^2}{dt^2} E [\delta L(t)] . \quad (\text{B.19})$$

Substituting (409) for $\delta L(t)$, using (408) and taking derivatives yields

$$E \left[\frac{d^2 L(Y(t))}{dt^2} \Big|_{t=0} \right] = -K \cdot \text{tr}(\delta R_1(A, B) R^{-1})^2 . \quad (\text{B.20})$$

Noting theorem 7 and (400) yields the desired result (411), that is

$$\mathbf{F}[\Delta, \Delta] = K \cdot \text{tr}(\delta R_1(A, B) R^{-1})^2 . \quad (\text{B.21})$$

B.1.3 Proof of Theorem 10 (Full FIM)

Compute $F_{\alpha\alpha}$

For notational convenience in the calculations below we drop the k subscript on the factor A_k .

For $\Delta = Y A$, with the coefficient matrix A anti-Hermitian,

$$\delta R_1(A, 0) R^{-1} \equiv Y C_1 Y^H + Y C_2 Y^H \quad (\text{B.22})$$

where

$$C_1 = A \Lambda \Lambda_s^{-1} , \quad (\text{B.23})$$

$$C_2 = \Lambda A^H \Lambda_s^{-1} . \quad (\text{B.24})$$

Using this we have

$$(\delta R_1(A, 0) R^{-1})^2 = Y C_1^2 Y^H + Y C_1 C_2 Y^H + Y C_2 C_1 Y^H + Y C_2^2 Y^H . \quad (\text{B.25})$$

Using $\text{tr}(Y C Y^H) = \text{tr}(C Y^H Y) = \text{tr}(C)$ yields

$$\text{tr}((\delta R_1(A, 0) R^{-1})^2) = \text{tr}(C_1^2 + C_1 C_2 + C_2 C_1 + C_2^2) . \quad (\text{B.26})$$

Substituting for C_1 and C_2 , the terms on the right hand side evaluate to

$$\text{tr}(C_1^2) = \text{tr}(A \Lambda \Lambda_s^{-1} A \Lambda \Lambda_s^{-1}); \quad (\text{B.27})$$

$$\text{tr}(C_1 C_2) = \text{tr}(C_2 C_1) = \text{tr}(A \Lambda \Lambda_s^{-1} \Lambda A^H \Lambda_s^{-1}); \quad (\text{B.28})$$

$$\text{tr}(C_2^2) = \text{tr}(\Lambda A^H \Lambda_s^{-1} \Lambda A^H \Lambda_s^{-1}) \quad (\text{B.29})$$

$$\text{tr}(C_2^2) = \text{tr}(A \Lambda_s^{-1} \Lambda A \Lambda_s^{-1}) = \text{tr}(A \Lambda \Lambda_s^{-1} A \Lambda \Lambda_s^{-1}) . \quad (\text{B.30})$$

Noting that $\text{tr}(C_1^2) = \text{tr}(C_2^2)$ (B.26) is

$$\text{tr}((\delta R_1(A, 0) R^{-1})^2) = 2\text{tr}(C_1^2 + C_1 C_2) . \quad (\text{B.31})$$

Next we evaluate the terms on the right hand side of the above. Using the fact that

$$\Lambda \equiv (\Lambda_s - \sigma^2 I)$$

$$\Lambda \Lambda_s^{-1} = (\Lambda_s - \sigma^2 I) \Lambda_s^{-1} = (I - \sigma^2 \Lambda_s^{-1}) \quad (\text{B.32})$$

we have

$$\text{tr}(C_1^2) = \text{tr}(A(I - \sigma^2 \Lambda_s^{-1}) A(I - \sigma^2 \Lambda_s^{-1})) , \quad (\text{B.33})$$

$$\text{tr}(C_1^2) = \text{tr}(A A - \sigma^2 A \Lambda_s^{-1} A - \sigma^2 A A \Lambda_s^{-1} + \sigma^4 A \Lambda_s^{-1} A \Lambda_s^{-1}) , \quad (\text{B.34})$$

$$\text{tr}(C_1^2) = \text{tr}(A A - 2\sigma^2 A A \Lambda_s^{-1} + \sigma^4 A \Lambda_s^{-1} A \Lambda_s^{-1}) . \quad (\text{B.35})$$

Substituting $A = -A^H$ and (B.32) into (B.28) yields

$$\text{tr}(C_1 C_2) = -\text{tr}(A \Lambda \Lambda_s^{-1} \Lambda A \Lambda_s^{-1}) , \quad (\text{B.36})$$

$$= -\text{tr}(A(I - \sigma^2 \Lambda_s^{-1}) \Lambda A \Lambda_s^{-1}) , \quad (\text{B.37})$$

$$= -\text{tr}(A \Lambda A \Lambda_s^{-1} - \sigma^2 A \Lambda_s^{-1} \Lambda A \Lambda_s^{-1}) , \quad (\text{B.38})$$

$$= -\text{tr}(A \Lambda A \Lambda_s^{-1} - \sigma^2 A(I - \sigma^2 \Lambda_s^{-1}) A \Lambda_s^{-1}) , \quad (\text{B.39})$$

$$= -\text{tr}(A \Lambda A \Lambda_s^{-1} - \sigma^2 A A \Lambda_s^{-1} + \sigma^4 A \Lambda_s^{-1} A \Lambda_s^{-1}) , \quad (\text{B.40})$$

$$= \text{tr}(-A \Lambda A \Lambda_s^{-1} + \sigma^2 A A \Lambda_s^{-1} - \sigma^4 A \Lambda_s^{-1} A \Lambda_s^{-1}) . \quad (\text{B.41})$$

Substituting (B.41) and (B.35) into (B.31) and noting (412) yields

$$\mathbf{F}[\mathbf{a}, \mathbf{a}] = K \cdot \text{tr}(A A - \Lambda A \Lambda_s^{-1} A - \sigma^2 A A \Lambda_s^{-1}) . \quad (\text{B.42})$$

(Note that the terms $\sigma^4 A \Lambda_s^{-1} A \Lambda_s^{-1}$ in (B.41) and (B.35) cancel to yield the above sum.)

Let $\mathbf{a}_k = Y A_k$ be an orthonormal basis for A_Y , where the A'_k s are an anti-Hermitian orthogonal set. The diagonal elements of the rotation FIM submatrix $F_{\alpha\alpha}$ are then

$$[F_{\alpha\alpha}]_{kk} = K \text{tr}(A_k A_k - \Lambda A_k \Lambda_s^{-1} A_k - \sigma^2 A_k A_k \Lambda_s^{-1}); \quad k = 1 \cdots N_a \quad (\text{B.43})$$

The off diagonal terms of are computed via the polarization approach as

$$[F_{\alpha\alpha}]_{ij} = \frac{1}{4}(\mathbf{F}[\mathbf{a}_i + \mathbf{a}_j, \mathbf{a}_i + \mathbf{a}_j] - \mathbf{F}[\mathbf{a}_i - \mathbf{a}_j, \mathbf{a}_i - \mathbf{a}_j]) . \quad (\text{B.44})$$

Substituting $A = A_i + A_j$ into (B.42) and expanding the first term on the right hand side can be expressed as

$$\mathbf{F}[\mathbf{a}_i + \mathbf{a}_j, \mathbf{a}_i + \mathbf{a}_j] = [f(A_i, A_i) + f(A_j, A_j) + f(A_i, A_j) + f(A_j, A_i)] \quad (\text{B.45})$$

where

$$f(A_i, A_j) \equiv K \text{tr}(A_i A_j - \Lambda A_i \Lambda_s^{-1} A_j - \sigma^2 A_i A_j \Lambda_s^{-1}) \quad (\text{B.46})$$

This follows since (B.27) has the form

$$\text{tr}(C_1^2) = \text{tr}((A_i + A_j)X(A_i + A_j)X) , \quad (\text{B.47})$$

which evaluates to

$$= \text{tr}(A_i X A_i) + \text{tr}(A_i X A_j) + \text{tr}(A_j X A_i) + \text{tr}(A_j X A_j) . \quad (\text{B.48})$$

The second term on the right hand side of (B.44) then is

$$\mathbf{F}[\mathbf{a}_i - \mathbf{a}_j, \mathbf{a}_i - \mathbf{a}_j] = [f(A_i, A_i) + f(A_j, A_j) - f(A_i, A_j) - f(A_j, A_i)] \quad (\text{B.49})$$

Substituting (B.45) and (B.49) into (B.44) and noting that the $f(A_i, A_i)$ and $f(A_j, A_j)$ terms cancel yields

$$[F_{\alpha\alpha}]_{ij} = \frac{1}{2}(f(A_i, A_j) + f(A_j, A_i)) . \quad (\text{B.50})$$

Using (B.46) this evaluates to

$$= K \left(\text{tr}(-\Lambda A_i \Lambda_s^{-1} A_j - \sigma^2 A_i A_j \Lambda_s^{-1}) + \text{tr}(-\Lambda A_j \Lambda_s^{-1} A_i - \sigma^2 A_j A_i \Lambda_s^{-1}) \right), \quad (\text{B.51})$$

$$= K \left(\text{tr}(-\Lambda A_i \Lambda_s^{-1} A_j - \Lambda A_j \Lambda_s^{-1} A_i) + \text{tr}(-\sigma^2 A_i A_j \Lambda_s^{-1} - \sigma^2 A_j A_i \Lambda_s^{-1}) \right). \quad (\text{B.52})$$

As shown below, each of the two terms in the sum vanish for $i \neq j$ so that

$$[F_{\alpha\alpha}]_{ij} = 0 \text{ for } i \neq j. \quad (\text{B.53})$$

FIM submatrix $F_{\alpha\alpha}$ with respect to the Standard Basis for A_Y

To evaluate the terms in the sum, consider the specific form of the basis vectors A_k . Let $A_{(ij)}$ denote matrix with 1 at the i, j location and -1 at the j, i location and zeros elsewhere. Make the identification $A_1 = A_{(12)}$, $A_2 = A_{(13)}$, etc as shown below.

Example: Recall that $N_a = p(p-1)$ (213)) so that for $p = 3$ we have $N_a = 6$.

The standard basis matrices are A_k ; $k = 1 \dots 6$ are

$$A_1 = \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \quad A_3 = \begin{bmatrix} 0 & 0 & -1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix} \quad A_5 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{bmatrix} \quad (\text{B.54})$$

$$A_2 = i \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \quad A_4 = i \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix} \quad A_6 = i \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}. \quad (\text{B.55})$$

For convenience in the calculations that follow record the ΛA products

$$\Lambda A_1 = \begin{bmatrix} 0 & \lambda_1 & 0 \\ -\lambda_2 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad \Lambda A_2 = i \cdot \begin{bmatrix} 0 & \lambda_1 & 0 \\ \lambda_2 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}. \quad (\text{B.56})$$

Compute Off-Diagonal

The $\Lambda A_1 \Lambda_s^{-1} A_2$ and $\Lambda A_2 \Lambda_s^{-1} A_1$ are respectively

$$= i \cdot \begin{bmatrix} 0 & \lambda_1 & 0 \\ \lambda_2 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & \lambda_{s;1}^{-1} & 0 \\ -\lambda_{s;2}^{-1} & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} = i \cdot \begin{bmatrix} -\lambda_1 \lambda_{s;2}^{-1} & 0 & 0 \\ 0 & \lambda_2 \lambda_{s;1}^{-1} & 0 \\ 0 & 0 & 0 \end{bmatrix} \quad (\text{B.57})$$

and

$$= i \cdot \begin{bmatrix} 0 & \lambda_1 & 0 \\ -\lambda_2 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & \lambda_{s;1}^{-1} & 0 \\ \lambda_{s;2}^{-1} & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} = i \cdot \begin{bmatrix} \lambda_1 \lambda_{s;2}^{-1} & 0 & 0 \\ 0 & -\lambda_2 \lambda_{s;1}^{-1} & 0 \\ 0 & 0 & 0 \end{bmatrix}. \quad (\text{B.58})$$

We see that $\Lambda A_1 \Lambda_s^{-1} A_2 = -\Lambda A_2 \Lambda_s^{-1} A_1$ so that their sum which appears in (B.52) vanishes. In addition

$$A_1 A_2 = i \cdot \begin{bmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad A_2 A_1 = i \cdot \begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \quad (\text{B.59})$$

so that the second terms in the sum (B.52) also vanishes.

Compute Diagonal

The second term, $\Lambda A_k \Lambda_s^{-1} A_k$, in the sum (B.46) for both $k = 1$ and $k = 2$ is

$$\begin{bmatrix} 0 & \lambda_1 & 0 \\ -\lambda_2 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & \lambda_{s;1}^{-1} & 0 \\ -\lambda_{s;2}^{-1} & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} -\lambda_1 \lambda_{s;2}^{-1} & 0 & 0 \\ 0 & -\lambda_2 \lambda_{s;1}^{-1} & 0 \\ 0 & 0 & 0 \end{bmatrix} \quad (\text{B.60})$$

so that

$$-tr(\Lambda A_k \Lambda_s^{-1} A_k) = \frac{\lambda_1}{\lambda_{s;2}} + \frac{\lambda_2}{\lambda_{s;1}}; \quad k = 1, 2. \quad (\text{B.61})$$

In general for $k \mapsto ij$

$$\Lambda \cdot A_k = \begin{bmatrix} \lambda_i \text{ at } ij \text{ location} \\ -\lambda_j \text{ at } ji \text{ location} \\ \text{zeros elsewhere} \end{bmatrix}. \quad (\text{B.62})$$

Then following (B.60) the product $\Lambda A_k \Lambda_s^{-1} A_k$ is

$$-tr(\Lambda A_k \Lambda_s^{-1} A_k) = \frac{\lambda_i}{\lambda_{s;j}} + \frac{\lambda_j}{\lambda_{s;i}}. \quad (\text{B.63})$$

The first and third terms in the sum (B.46) involve the products $A_k A_k$. These are

$$A_1 A_1 = A_{(12)} \cdot A_{(12)} = \begin{bmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad (\text{B.64})$$

$$A_2 A_2 = i \cdot S_{(12)} \cdot i \cdot S_{(12)} = -1 \cdot \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}. \quad (\text{B.65})$$

In the general case, $S_{(ij)}$ is a symmetric matrix given by

$$S_{(ij)} = \begin{bmatrix} 1 \text{ at the } ij \text{ and } ji \text{ off diagonals} \\ \text{zeros elsewhere} \end{bmatrix} \quad (\text{B.66})$$

and $A_{(ij)}$ is an anti-symmetric matrix given by

$$A_{(ij)} = \begin{bmatrix} -1 \text{ at the } ij \text{ th location and } 1 \text{ at } ji \text{ th off diagonals} \\ \text{zeros elsewhere} \end{bmatrix}. \quad (\text{B.67})$$

In general for $k \mapsto ij$

$$A_k \cdot A_k = \begin{bmatrix} -1 \text{ at } ii \text{ and } jj \text{ diagonals} \\ \text{zeros elsewhere} \end{bmatrix} \quad (\text{B.68})$$

It follows, then, that for all k

$$\text{tr}(A_k A_k) = -2 \quad (\text{B.69})$$

which gives the first term in the sum. To compute the third term use (B.68)

$$-\text{tr}(\sigma^2 A_k A_k \Lambda_s^{-1}) = \frac{\sigma^2}{\lambda s; i} + \frac{\sigma^2}{\lambda s; j}. \quad (\text{B.70})$$

Substituting (B.63), (B.69) and (B.70) into (B.43) and making the identification $k = (ij)$ yields

$$K^{-1}[F_{\alpha\alpha}]_{kk} = \gamma_{ij} \equiv -2 + \frac{\lambda_i}{\lambda s; j} + \frac{\lambda_j}{\lambda s; i} + \frac{\sigma^2}{\lambda s; i} + \frac{\sigma^2}{\lambda s; j}. \quad (\text{B.71})$$

For convenience, define γ_{ij} by the right hand side of the above. Rearranging terms yields

$$\gamma_{ij} = \frac{\lambda_i \lambda s; j - 2\lambda s; i \lambda s; j + \lambda_j \lambda s; i + \sigma^2(\lambda s; j + \lambda s; i)}{\lambda s; i \lambda s; j}. \quad (\text{B.72})$$

Using $\lambda s; i = \lambda_i + \sigma^2$ we have

$$\gamma_{ij} = \frac{\lambda_i(\lambda s; i + \sigma^2) - 2(\lambda_i + \sigma^2)(\lambda_j + \sigma^2) + \lambda_j(\lambda s; j + \sigma^2) + 2\sigma^4}{(\lambda_i + \sigma^2)(\lambda_j + \sigma^2)} \quad (\text{B.73})$$

$$\gamma_{ij} = \frac{\lambda_i(\lambda_i + 2\sigma^2) - 2(\lambda_i + \sigma^2)(\lambda_j + \sigma^2) + \lambda_j(\lambda_j + 2\sigma^2) + 2\sigma^4}{(\lambda_i + \sigma^2)(\lambda_j + \sigma^2)} \quad (\text{B.74})$$

For

$$\lambda_i \gg \sigma^2 \quad (\text{B.75})$$

$$\gamma_{ij} = \frac{(\lambda_i - \lambda_j)^2}{\lambda_i \lambda_j} \quad (\text{B.76})$$

$F_{\alpha\alpha}$ is the $N_a \times N_a$ diagonal matrix ($N_a = p(p-1) = 6$)

$$F_{\alpha\alpha} = K \begin{bmatrix} \gamma_{(12)} I_2 & 0 & 0 \\ 0 & \gamma_{(13)} I_2 & 0 \\ 0 & 0 & \gamma_{(23)} I_2 \end{bmatrix} \quad (\text{B.77})$$

This can alternatively be expressed as

$$F_{\alpha\alpha} = K (G \otimes I_2) \quad (\text{B.78})$$

where $G = \text{diag}(\gamma)$, $\gamma = (\gamma_{(12)}, \gamma_{(13)}, \gamma_{(23)})$.

Compute $F_{\beta\beta}$

For $\Delta = Y_{\perp} B$, the coefficient matrix B arbitrary,

$$\delta R_1(0, B) = Y_{\perp} B \Lambda Y^H + Y \Lambda B^H Y_{\perp}^H. \quad (\text{B.79})$$

Right multiple by $R^{-1} = Y \Lambda_s^{-1} Y^H + \sigma^{-2} Y_{\perp} Y_{\perp}^H$ to obtain

$$\delta R_1(0, B) R^{-1} = Y_{\perp} B \Lambda \Lambda_s^{-1} Y^H + \sigma^{-2} Y \Lambda B^H Y_{\perp}^H. \quad (\text{B.80})$$

Squaring the above yields

$$(\delta R_1(0, B) R^{-1})^2 = \sigma^{-2} Y_{\perp} B \Lambda \Lambda_s^{-1} \Lambda B^H Y_{\perp}^H + \sigma^{-2} Y \Lambda B^H B \Lambda \Lambda_s^{-1} Y^H \quad (\text{B.81})$$

and

$$\text{tr}(\delta R_1(0, B) R^{-1})^2 = \sigma^{-2} \text{tr}(\Lambda_{OPT} B^H B + \Lambda_{OPT} B^H B), \quad (\text{B.82})$$

$$= 2\sigma^{-2} \text{tr}(\Lambda_{OPT} B^H B) \quad (\text{B.83})$$

where

$$\Lambda_{OPT} \equiv \Lambda \Lambda_s^{-1} \Lambda = \Lambda_s^{-1} \Lambda^2. \quad (\text{B.84})$$

Replacing the B by B_k and substituting (B.83) into (411) yields the diagonal elements

$$[F_{\beta\beta}]_{kk} = \frac{2K}{\sigma^2} \text{Re} \left(\text{tr}(\Lambda_{OPT} B_k^H B_k) \right); \quad k = 1 \cdots N_b. \quad (\text{B.85})$$

The off diagonal terms are computed using a polarization approach

$$[F_{\beta\beta}]_{ij} = \frac{1}{4} [\mathbf{F}[\mathbf{b}_i + \mathbf{b}_j, \mathbf{b}_i + \mathbf{b}_j] - \mathbf{F}[\mathbf{b}_i - \mathbf{b}_j, \mathbf{b}_i - \mathbf{b}_j]]. \quad (\text{B.86})$$

Replacing B_k and $B_i + B_j$ in (413) we have

$$\mathbf{F}[\mathbf{b}_i + \mathbf{b}_j, \mathbf{b}_i + \mathbf{b}_j] = \frac{2K}{\sigma^2} \text{Re} \left(\text{tr}(\Lambda_{OPT} (B_i + B_j)^H (B_i + B_j)) \right) \quad (\text{B.87})$$

and expanding

$$= \frac{2K}{\sigma^2} \text{tr}(\Lambda_{OPT} (B_i^H B_i + B_i^H B_j + B_j^H B_i + B_j^H B_j)) . \quad (\text{B.88})$$

$\mathbf{F}[\mathbf{b}_i - \mathbf{b}_j, \mathbf{b}_i - \mathbf{b}_j]$ is the same except the cross terms have a negative sign. Substituting into (B.86) yields

$$[F_{\beta\beta}]_{ij} = \frac{1}{4} \left[\frac{2K}{\sigma^2} \text{tr}(\Lambda_{OPT} \cdot 2(B_i^H B_j + B_j^H B_i)) \right] . \quad (\text{B.89})$$

Noting that

$$2(B_i^H B_j + B_j^H B_i) = 4\text{Re}(B_i^H B_j) \quad (\text{B.90})$$

results in

$$= \frac{1}{4} \left[\frac{2K}{\sigma^2} \text{tr}(\Lambda_{OPT} \cdot 4\text{Re}(B_i^H B_j)) \right] , \quad (\text{B.91})$$

$$= \frac{1}{4} \left[4 \frac{2K}{\sigma^2} \text{tr}(\Lambda_{OPT} \cdot \text{Re}(B_i^H B_j)) \right] , \quad (\text{B.92})$$

$$= \frac{2K}{\sigma^2} [\text{tr}(\Lambda_{OPT} \cdot \text{Re}(B_i^H B_j))] . \quad (\text{B.93})$$

Taken together we have (compare to [1, equation 143])

$$[F_{\beta\beta}]_{ij} = \frac{2K}{\sigma^2} \text{Re} \left(\text{tr}(\Lambda_{OPT} B_i^H B_j) \right); \quad i, j = 1 \cdots N_b . \quad (\text{B.94})$$

Using the identity

$$\text{tr}(\Lambda_{OPT} B^H B) = b^H (\Lambda_{OPT} \otimes I_{N-p}) b \quad (\text{B.95})$$

(see B.1.10)

$$[F_{\beta\beta}]_{ij} = \frac{2K}{\sigma^2} \text{Re}(b_j^H (\Lambda_{OPT} \otimes I_{N-p}) b_i); \quad i, j = 1 \cdots N_b \quad (\text{B.96})$$

where $b_k = \text{vec}(B_k)$. The $N_b \times N_b$ matrix form is

$$[F_{\beta\beta}] = \frac{2K}{\sigma^2} \text{Re} \left(E_b^H (\Lambda_{OPT} \otimes I_{N-p}) E_b \right) \quad (\text{B.97})$$

where E_b is the $\bar{N}_b \times N_b$ matrix

$$E_b = \begin{bmatrix} b_i & \cdots & b_{N_b} \end{bmatrix} . \quad (\text{B.98})$$

Remark:

If $E_b = \begin{bmatrix} I_{\bar{N}_b} & i \cdot I_{\bar{N}_b} \end{bmatrix}$ $N_b = 2p(N - p)$, $\bar{N}_b = p(N - p)$

$$[F_{\beta\beta}] = \frac{2K}{\sigma^2} \text{Re} \left(\begin{bmatrix} I_{\bar{N}_b} \\ -i \cdot I_{\bar{N}_b} \end{bmatrix} (\Lambda_{OPT} \otimes I_{N-p}) \begin{bmatrix} I_{\bar{N}_b} & i \cdot I_{\bar{N}_b} \end{bmatrix} \right) \quad (\text{B.99})$$

$$[F_{\beta\beta}] = \frac{2K}{\sigma^2} \text{Re} \left(\begin{bmatrix} (\Lambda_{OPT} \otimes I_{N-p}) & i \cdot (\Lambda_{OPT} \otimes I_{N-p}) \\ -i \cdot (\Lambda_{OPT} \otimes I_{N-p}) & (\Lambda_{OPT} \otimes I_{N-p}) \end{bmatrix} \right) \quad (\text{B.100})$$

$$[F_{\beta\beta}] = \frac{2K}{\sigma^2} \begin{bmatrix} (\Lambda_{OPT} \otimes I_{N-p}) & 0 \\ 0 & (\Lambda_{OPT} \otimes I_{N-p}) \end{bmatrix} \quad (\text{B.101})$$

Compute Cross Term $F_{\alpha\beta}$

To compute the cross terms recall (411)

$$\mathbf{F}[\Delta, \Delta] = K \cdot \text{tr}(\delta R_1(A, B) R^{-1})^2 \quad (\text{B.102})$$

$$[F_{\alpha\beta}]_{ij} \equiv E[\nabla^2 L[\mathbf{a}_i, \mathbf{b}_j]] \quad (\text{B.103})$$

The polarization method is used to compute the cross terms of the bilinear form. Polarization for the cross terms

$$\mathbf{F}[\mathbf{a}_i, \mathbf{b}_j] = \frac{1}{4} (\mathbf{F}[\mathbf{a}_i + \mathbf{b}_j, \mathbf{a}_i + \mathbf{b}_j] - \mathbf{F}[\mathbf{a}_i - \mathbf{b}_j, \mathbf{a}_i - \mathbf{b}_j]) \quad (\text{B.104})$$

Let $\mathbf{a}_i = Y A_i$ and $\mathbf{b}_j = Y_\perp B_j$ and for convenience drop the subscripts on A and B in the computation below. The two terms in the sum on the right hand side of (B.104) are

$$\mathbf{F}[\mathbf{a} + \mathbf{b}, \mathbf{a} + \mathbf{b}] = \text{tr}(\delta R_1(A, B) R^{-1})^2 \quad (\text{B.105})$$

$$\mathbf{F}[\mathbf{a} - \mathbf{b}, \mathbf{a} - \mathbf{b}] = \text{tr}(\delta R_1(A, -B) R^{-1})^2 . \quad (\text{B.106})$$

Using (B.16), for $\delta R_1(A, B)$ the factor $\delta R_1(A, B)(R^{-1})$ evaluates to

$$\delta R_1(A, B)(R^{-1}) = \delta R_1(A, B)(Y\Lambda_s^{-1}Y^H + \sigma^{-2}Y_\perp Y_\perp^H) \quad (\text{B.107})$$

$$= (Y A \Lambda \Lambda_s^{-1} Y^H) + Y \Lambda A^H \Lambda_s^{-1} Y^H + Y_\perp B \Lambda \Lambda_s^{-1} Y^H + \sigma^{-2} Y \Lambda B^H Y_\perp^H. \quad (\text{B.108})$$

Squaring and taking the trace yields

$$\text{tr}((\delta R_1(A, B))R^{-1})^2 = \text{tr}(X) + \text{tr}(Y(C_1 + C_2)Y_\perp^H + Y_\perp(C_3 + C_4)Y^H) \quad (\text{B.109})$$

where

$$C_1 = \sigma^{-2} A \Lambda \Lambda_s^{-1} \Lambda B^H, \quad (\text{B.110})$$

$$C_2 = \sigma^{-2} \Lambda A^H \Lambda_s^{-1} \Lambda B^H, \quad (\text{B.111})$$

$$C_3 = B \Lambda \Lambda_s^{-1} A \Lambda \Lambda_s^{-1}, \quad (\text{B.112})$$

and

$$C_4 = B \Lambda \Lambda_s^{-1} \Lambda A^H \Lambda_s^{-1}. \quad (\text{B.113})$$

The term X denotes the factors involving A only and B only quadratic (squared) terms.

Since the factor $Y^H Y_\perp = 0$ in the second and third terms (B.109), these terms vanish, leaving only $\text{tr}(X)$. Thus (B.109) reduces to

$$\text{tr}((\delta R_1(A, B))R^{-1})^2 = \text{tr}(X). \quad (\text{B.114})$$

Similarly

$$\text{tr}((\delta R_1(A, -B))R^{-1})^2 = \text{tr}(X), \quad (\text{B.115})$$

and noting (B.105) and (B.106), conclude that

$$[F_{\alpha\beta}]_{ij} = \text{tr}(X) - \text{tr}(X) = 0. \quad (\text{B.116})$$

This hold true for arbitrary choice of A and B so conclude that cross term matrix is all zeros

$$[F_{\alpha\beta}] = 0. \quad (\text{B.117})$$

B.1.4 Proof of Corollary 13 (Eigenvector CRB)

Let $\Delta = \exp_{Y_0}^{-1} \hat{Y}$. The perturbation to the k th column of Y_0 is

$$\hat{s}_k - s_k \mapsto \Delta \cdot i_k; \quad m = 1 \cdots p \quad (\text{B.118})$$

where i_m is the standard unit vector consisting of 1 in the m -th location and zeros elsewhere. This tangent vector Δ can be represented with respect to arbitrarily chosen orthonormal basis set as

$$\Delta = \sum_{k=1}^{N_a} \mathbf{a}_k \alpha^k + \sum_{k=1}^{N_b} \mathbf{b}_k \beta^k. \quad (\text{B.119})$$

In this expression the real N_a -tuple α , and the real N_b -tuple β , are the stochastic components. Then

$$\Delta \cdot i_m = \left(\sum_{k=1}^{N_a} \mathbf{a}_k \alpha^k + \sum_{k=1}^{N_b} \mathbf{b}_k \beta^k \right) i_m; \quad m = 1 \cdots p \quad (\text{B.120})$$

$$\Delta \cdot i_m = \sum_{k=1}^{N_a} \alpha^k \mathbf{a}_k i_m + \sum_{k=1}^{N_b} \beta^k \mathbf{b}_k i_m \quad (\text{B.121})$$

$$\Delta \cdot i_m = P_A \Delta \cdot i_m + P_B \Delta \cdot i_m \quad (\text{B.122})$$

We compute the terms, $P_A \Delta \cdot i_m$ and $P_B \Delta \cdot i_m$, separately. Let the basis vectors have the numerical form $\mathbf{a}_k = Y A_k$ and $\mathbf{b}_k = Y_\perp B_k$.

Compute $P_A \Delta \cdot i_m$

Let the basis vectors for A_{Y_0} be $\mathbf{a}_k = Y A_k$. Then

$$P_A \Delta \cdot i_m = \sum_{k=1}^{N_a} \alpha^k \mathbf{a}_k i_m = \sum_{k=1}^{N_a} \alpha^k Y A_k i_m. \quad (\text{B.123})$$

For $p = 3$, $N_a = 6$ and the standard basis matrices are A_k given by (B.54) and (B.55).

For $m = 1$ the 6 terms in the sum (B.123) are

$$\alpha^1 Y A_1 i_1 = \alpha^1 Y \begin{bmatrix} 0 \\ -1 \\ 0 \end{bmatrix} = Y \begin{bmatrix} 0 \\ -\alpha^1 \\ 0 \end{bmatrix} \quad (\text{B.124})$$

$$\alpha^3 Y A_3 i_1 = \alpha^3 Y \begin{bmatrix} 0 \\ 0 \\ -1 \end{bmatrix} = Y \begin{bmatrix} 0 \\ 0 \\ -\alpha^3 \end{bmatrix} \quad (\text{B.125})$$

$$\alpha^5 Y A_5 i_1 = \alpha^5 Y \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} = 0 \quad (\text{B.126})$$

$$\alpha^2 Y A_2 i_1 = \alpha^2 Y \begin{bmatrix} 0 \\ i \\ 0 \end{bmatrix} = Y \begin{bmatrix} 0 \\ i\alpha^2 \\ 0 \end{bmatrix} \quad (\text{B.127})$$

$$\alpha^4 Y A_4 i_1 = \alpha^4 Y \begin{bmatrix} 0 \\ 0 \\ i \end{bmatrix} = Y \begin{bmatrix} 0 \\ 0 \\ i\alpha^4 \end{bmatrix} \quad (\text{B.128})$$

$$\alpha^6 Y A_6 i_1 = \alpha^6 Y \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} = 0 \quad (\text{B.129})$$

Then (B.123) is the sum of (B.124) to (B.129)

$$P_A \Delta \cdot i_1 = \sum_{m=1}^{N_a} \alpha^m Y A_m i_1 = Y \begin{bmatrix} 0 \\ -\alpha^1 + i\alpha^2 \\ -\alpha^3 + i\alpha^4 \end{bmatrix}. \quad (\text{B.130})$$

This expression and the corresponding ones for $P_A \Delta \cdot i_2, \dots, P_A \Delta \cdot i_p$ can be written in the matrix form as

$$P_A \Delta \cdot i_k = \sum_{m=1}^{N_a} \alpha^m Y A_m i_k = Y T_k \alpha; \quad k = 1, \dots, p \quad (\text{B.131})$$

For the $p = 3$ the set of T_k matrices are

$$T_1 = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ -1 & i & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & i & 0 & 0 \end{bmatrix}, \quad (\text{B.132})$$

$$T_2 = \begin{bmatrix} -1 & i & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & i \end{bmatrix}, \quad (\text{B.133})$$

$$T_3 = \begin{bmatrix} 0 & 0 & 1 & i & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & i \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}. \quad (\text{B.134})$$

Note that substituting T_1 given by (B.132) into (B.131) yields (B.130), confirming the T_1 matrix.

Compute $P_B \Delta \cdot i_m$

Let the basis vectors for B_{Y_0} be $\mathbf{b}_k = Y_{0\perp} B_k$. Then

$$P_B \Delta \cdot i_m = \sum_{k=1}^{N_b} \mathbf{b}_k \beta^k i_m = \sum_{k=1}^{N_b} Y_{\perp} B_k \beta^k i_m = Y_{\perp} M_m \beta . \quad (\text{B.135})$$

p=2

We first develop the formulas for the case $p = 2, N = 5$ and then generalize the result. In this case $N - p = 3$ and $N_b = 2p(N - p) = 12$. For $m = 1$ the $P_B \Delta \cdot i_m$ term is computed as follows. Using the standard basis, the B_k matrices are such that

$$Y_{\perp} B_k i_1 = 0; \quad k = 4, 5, 6, 10, 11, 12 . \quad (\text{B.136})$$

The non-zeros terms corresponding to the $k = 1, 2, 3$ and $k = 7, 8, 9$ basis vectors are

$$Y_{\perp} B_1 i_1 = Y_{\perp} \begin{bmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = Y_{\perp} \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} = g_1 \quad (\text{B.137})$$

$$Y_{\perp} B_2 i_1 = Y_{\perp} \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = Y_{\perp} \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} = g_2 \quad (\text{B.138})$$

$$Y_{\perp} B_3 i_1 = Y_{\perp} \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = Y_{\perp} \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = g_3 = g_{N-p} \quad (\text{B.139})$$

and for $k = 7, 8, 9$ (with $B_7 = i \cdot B_1$, $B_8 = i \cdot B_2$ and $B_9 = i \cdot B_3$)

$$Y_{\perp} B_7 i_1 = i g_1 , \quad (\text{B.140})$$

$$Y_{\perp} B_8 i_1 = i g_2 , \quad (\text{B.141})$$

$$Y_{\perp} B_9 i_1 = i g_3 . \quad (\text{B.142})$$

Using the above, the sum (B.135) for $m = 1$ can be expressed as

$$P_B \Delta \cdot i_1 = Y_{\perp} M_1 \beta \quad (\text{B.143})$$

where M_1 is $((N - p) \times N_b)$ given by

$$M_1 = \begin{bmatrix} X_1 & i \cdot X_1 \end{bmatrix} \quad (\text{B.144})$$

and X_1 is $((N - p) \times p(N - p))$. For the example calculation at hand $p = 2$ and the X_1 matrix is

$$X_1 = \begin{bmatrix} I_{N-p} & 0 \end{bmatrix}. \quad (\text{B.145})$$

Similarly for $m = 2$ the non-zero terms in the sum corresponding to the $k = 4, 5, 6$ and $k = 10, 11, 12$ basis vectors are

$$Y_{\perp} B_4 i_2 = Y_{\perp} \begin{bmatrix} 0 & 1 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} = Y_{\perp} \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} = g_1 \quad (\text{B.146})$$

$$Y_{\perp} B_5 i_2 = Y_{\perp} \begin{bmatrix} 0 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} = Y_{\perp} \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} = g_2 \quad (\text{B.147})$$

$$Y_{\perp} B_{N_b} i_2 = Y_{\perp} \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} = Y_{\perp} \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = g_3 = g_{N-p} \quad (\text{B.148})$$

and for $k = 10, 11, 12$

$$Y_{\perp} B_{n+9} i_2 = i \cdot g_n; \quad n = 1, 2, 3 \quad (\text{B.149})$$

Using the above the sum (B.135) for $m = 2$ can be expressed as

$$P_B \Delta \cdot i_2 = Y_{\perp} M_2 \beta \quad (\text{B.150})$$

where $M_2 = \begin{bmatrix} X_2 & i \cdot X_2 \end{bmatrix}$ and $X_2 = \begin{bmatrix} 0 & I_{N-p} \end{bmatrix}$.

p arbitrary

In the general case, for arbitrary p , the X_k matrices are $(N - p) \times p(N - p)$ and are given by

$$X_k = i_k^t \otimes I_{N-p}; \quad k = 1 \cdots p. \quad (\text{B.151})$$

It follows that the M_k matrices are $(N - p) \times N_b$, $N_b = 2p(N - p)$ and are given by

$$M_k = \begin{bmatrix} X_k & i X_k \end{bmatrix}; \quad k = 1 \cdots p. \quad (\text{B.152})$$

The factor $Y_{\perp} M_k$ is then $N \times N_b$.

(B.153)

Using (B.151) we note that for $p = 3$ the X_2 matrix, for example, is $X_2 = \begin{bmatrix} 0 & I_{N-p} & 0 \end{bmatrix}$.

Compute $P_A \Delta \cdot i_m + P_B \Delta \cdot i_m$

Substituting (B.131) and (B.135) into (B.121) yields

$$\Delta \cdot i_k = Y T_k \alpha + Y_{\perp} M_k \beta \quad (\text{B.154})$$

or

$$\Delta \cdot i_k = \begin{bmatrix} Y T_k & Y_{\perp} M_k \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix}. \quad (\text{B.155})$$

The covariance matrix of $\Delta \cdot i_k$, the k -th column of Δ , is

$$\text{cov}(\Delta \cdot i_k) = \begin{bmatrix} Y T_k & Y_{\perp} M_k \end{bmatrix} \text{cov} \left(\begin{bmatrix} \alpha \\ \beta \end{bmatrix} \right) \begin{bmatrix} Y T_k & Y_{\perp} M_k \end{bmatrix}^H, \quad (\text{B.156})$$

$$\text{cov}(\Delta \cdot i_k) = \begin{bmatrix} Y T_k & Y_{\perp} M_k \end{bmatrix} \begin{bmatrix} C_{\alpha\alpha} & C_{\beta\alpha} \\ C_{\alpha\beta} & C_{\beta\beta} \end{bmatrix} \begin{bmatrix} Y T_k & Y_{\perp} M_k \end{bmatrix}^H. \quad (\text{B.157})$$

To establish the bound use the fact that $C \succeq F^{-1}$ (see theorem 10) so that

$$\text{cov}(\Delta \cdot i_k) \succeq \begin{bmatrix} Y T_k & Y_{\perp} M_k \end{bmatrix} \begin{bmatrix} F_{\alpha\alpha} & F_{\alpha\beta} \\ F_{\beta\alpha} & F_{\beta\beta} \end{bmatrix}^{-1} \begin{bmatrix} Y T_k & Y_{\perp} M_k \end{bmatrix}^H \quad (\text{B.158})$$

Since the off diagonal sub-matrices ($F_{\alpha\beta}$ and $F_{\beta\alpha}$) of the matrix F vanish, this becomes

$$\text{cov}(\Delta \cdot i_k) \succeq \begin{bmatrix} Y T_k & Y_{\perp} M_k \end{bmatrix} \begin{bmatrix} F_{\alpha\alpha}^{-1} & 0 \\ 0 & F_{\beta\beta}^{-1} \end{bmatrix} \begin{bmatrix} Y T_k & Y_{\perp} M_k \end{bmatrix}^H, \quad (\text{B.159})$$

or equivalently

$$\text{cov}(\Delta \cdot i_k) \succeq Y T_k F_{\alpha\alpha}^{-1} T_k^H Y^H + Y_{\perp} M_k F_{\beta\beta}^{-1} M_k^H Y_{\perp}^H. \quad (\text{B.160})$$

For convenience, this is expressed as

$$\text{cov}(\Delta \cdot i_k) \succeq Y G^{(k)} Y^H + Y_{\perp} H^{(k)} Y_{\perp}^H \quad (\text{B.161})$$

where the $p \times p$ matrix $G^{(k)}$ is

$$G^{(k)} \equiv T_k F_{\alpha\alpha}^{-1} T_k^H \quad (\text{B.162})$$

and where the $(N-p) \times (N-p)$ matrix $H^{(k)}$ is

$$H^{(k)} = M_k F_{\beta\beta}^{-1} M_k^H. \quad (\text{B.163})$$

Consider the first term on the right hand side of (B.161). For $k = 1$ using (B.132) and (439) we have yields

$$G^{(k)} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & \gamma_{12}^{-1} & 0 \\ 0 & 0 & \gamma_{13}^{-1} \end{bmatrix}. \quad (\text{B.164})$$

Using the notation $Y = \begin{bmatrix} s_1 & s_2 & s_3 \end{bmatrix}$ we have

$$Y G^{(1)} Y^H = \gamma_{12}^{-1} s_2 s_2^H + \gamma_{13}^{-1} s_3 s_3^H. \quad (\text{B.165})$$

Using (439) for γ_{ij} , that is

$$\gamma_{ij} = \frac{(\lambda_i - \lambda_j)^2}{\lambda_i \lambda_j} \quad (\text{B.166})$$

yields

$$Y G^{(1)} Y^H = \frac{\lambda_1 \lambda_2}{(\lambda_1 - \lambda_2)^2} s_2 s_2^H + \frac{\lambda_1 \lambda_3}{(\lambda_1 - \lambda_3)^2} s_3 s_3^H \quad (\text{B.167})$$

$$Y G^{(1)} Y^H = \lambda_1 \left(\frac{\lambda_2}{(\lambda_1 - \lambda_2)^2} s_2 s_2^H + \frac{\lambda_3}{(\lambda_1 - \lambda_3)^2} s_3 s_3^H \right) \quad (\text{B.168})$$

or

$$Y G^{(1)} Y^H = \lambda_1 \sum_{k=1, k \neq 1}^p \frac{\lambda_k}{(\lambda_1 - \lambda_k)^2} s_k s_k^H. \quad (\text{B.169})$$

This generalizes to

$$Y G^{(i)} Y^H = \lambda_i \sum_{k=1, k \neq i}^p \frac{\lambda_k}{(\lambda_i - \lambda_k)^2} s_k s_k^H \quad (\text{B.170})$$

$$\Lambda_{kk} = \frac{\lambda_k - \tilde{\sigma}^2}{\lambda_k}. \quad (\text{B.171})$$

Next, to evaluate the second term on the right hand side of (B.161), recall that

$$F_{\beta\beta}^{-1} = \frac{\sigma^2}{2K} \begin{bmatrix} (\Lambda_{OPT}^{-1} \otimes I_{N-p}) & 0 \\ 0 & (\Lambda_{OPT}^{-1} \otimes I_{N-p}) \end{bmatrix} \quad (\text{B.172})$$

Using this the factor $M_k F_{\beta\beta}^{-1} M_k^H$ (B.163) is

$$M_k F_{\beta\beta}^{-1} M_k^H = \frac{\tilde{\sigma}_0^2}{K(\lambda_{OPT})_k} I_{N-p} . \quad (\text{B.173})$$

(Note that the 2 has cancelled.) As a result we have

$$Y_{\perp} M_k F_{\beta\beta}^{-1} M_k^H Y_{\perp}^H = \frac{\tilde{\sigma}_0^2}{K(\lambda_{OPT})_k} Y_{\perp} Y_{\perp}^H . \quad (\text{B.174})$$

Using the notation $Y_{\perp} = [g_1 \cdots g_{N-p}]$ this can be expressed as

$$Y_{\perp} M_k F_{\beta\beta}^{-1} M_k^H Y_{\perp}^H = \frac{\tilde{\sigma}_0^2}{K(\lambda_{OPT})_k} \sum_{m=1}^{N-p} g_m g_m^H; \quad k = 1 \cdots p . \quad (\text{B.175})$$

Using $(\lambda_{OPT})_k = \frac{\lambda_k^2}{\lambda_k - \tilde{\sigma}^2} = \frac{\lambda_{s;k} - \tilde{\sigma}^2}{\lambda_{s;k}}$ the above becomes

$$= \frac{\lambda_k \tilde{\sigma}_0^2}{K(\lambda_k - \tilde{\sigma}^2)} \sum_{m=1}^{N-p} g_m g_m^H; \quad k = 1 \cdots p . \quad (\text{B.176})$$

Lastly we have

$$E[(\Delta \cdot i_i)(\Delta \cdot i_j)^H] = 0; \quad i \neq j \quad (\text{B.177})$$

which follows since

$$T_i F_{\alpha\alpha}^{-1} T_j^H = 0, \quad i \neq j \quad (\text{B.178})$$

and

$$M_i F_{\beta\beta}^{-1} M_j^H = 0, \quad i \neq j, . \quad (\text{B.179})$$

B.1.5 Proof of Lemma 15

The columns of $E_{\mathcal{S}}$ are given by $E_{\mathcal{S}}(:, k) = \text{vec}(\acute{S}_k)$ where $\acute{S}_k$ is (518)

$$\acute{S}_k = Y_{\perp}^H D J_{kk} \cdot T . \quad (\text{B.180})$$

Using this we have

$$\text{Re}(E_{\mathcal{S}}^H E_{\mathcal{S}})_{ij} = \text{Re}(\text{tr}(\acute{S}_i^H, \acute{S}_j)) = \text{Re}(\text{tr}(T^H J_{ii} D P_{\perp} D J_{jj} T)) \quad (\text{B.181})$$

where $P_{\perp} \equiv Y_{\perp} Y_{\perp}^H$. Using the selection matrix identity (B.260) the factor on the right hand side of the above can be expressed as

$$\text{tr}(T^H J_{ii} D P_{\perp} D J_{jj} T) = \text{tr}(T^H (D^H P_{\perp} D)_{ij} \cdot J_{ij} \cdot T) , \quad (\text{B.182})$$

$$= \text{tr}((D^H P_{\perp} D)_{ij} \cdot J_{ij} \cdot T T^H) . \quad (\text{B.183})$$

Factoring out the scalar term yields

$$= (D^H P_{\perp} D)_{ij} \text{tr}(J_{ij} \cdot T T^H) . \quad (\text{B.184})$$

Applying the selection matrix identity (B.256) we have

$$(E_s^H E_s)_{ij} = (D^H P_{\perp} D)_{ij} \cdot (T T^H)_{ij} . \quad (\text{B.185})$$

This result can be restated in matrix form using the Hadamard product as

$$E_s^H E_s = D^H P_{\perp} D \odot (T T^H) . \quad (\text{B.186})$$

Since, by definition, $Y = HT$ we have $H = Y T^{-1}$ so that

$$H^H H = T^{-H} Y^H Y T^{-1} = T^{-H} T^{-1} \quad (\text{B.187})$$

and

$$T T^H = (H^H H)^{-1} . \quad (\text{B.188})$$

Making this substitution for $T T^H$ in (B.186) yields the desired result

$$\text{Re}(E_s^H E_s) = \text{Re}((D^H P_{\perp} D) \odot (H^H H)^{-1}) . \quad (\text{B.189})$$

B.1.6 Proof of 18 (M-Dim Orthogonalizing Matrix Lemma)

The columns of E_s are given by $E_s(:, l_1) \equiv \text{vec}(\acute{S}_{l_1})$ where $\acute{S}_k$ is

$$\acute{S}_{l_1} = Y_{\perp}^H D J_{m_1 k_1 k_1} T; \quad l_1 = 1 \cdot q . \quad (\text{B.190})$$

The indices on the right and left-hand sides are related by

$$l_1 = p(m_1 - 1) + k_1, \quad m_1 = 1 \cdots M, k_1 = 1 \cdots p. \quad (\text{B.191})$$

Using this expression the product

$$[E_{\dot{s}}^H E_{\dot{s}}]_{l_1 l_2} = [E_{\dot{s}}^H(:, l_1) E_{\dot{s}}(:, l_2)] \quad (\text{B.192})$$

is given by

$$[E_{\dot{s}}^H E_{\dot{s}}]_{l_1 l_2} = \text{tr}((J_{m_1 k_1 k_1})^T D^H P_{\perp} D \cdot J_{m_2 k_2 k_2} (T T^H)) . \quad (\text{B.193})$$

Using the selection matrix identity (B.263) we have

$$(J_{m_1 k_1 k_1})^T D^H P_{\perp} D \cdot J_{m_2 k_2 k_2} = [D^H P_{\perp} D]_{l_1 l_2} \cdot J_{k_1 k_2} . \quad (\text{B.194})$$

Substituting this into (B.193) yields

$$= [D^H P_{\perp} D]_{l_1 l_2} \cdot \text{tr}(J_{k_1 k_2} T T^H) . \quad (\text{B.195})$$

From the selection matrix identity (B.256), the factor above simplifies to

$$\text{tr}(J_{k_1 k_2} T T^H) = [T T^H]_{k_1 k_2} \quad (\text{B.196})$$

and back substituting yields

$$[E_{\dot{s}}^H E_{\dot{s}}]_{l_1 l_2} = [D^H P_{\perp} D]_{l_1 l_2} \cdot [T T^H]_{k_1 k_2} . \quad (\text{B.197})$$

This can be expressed in full matrix form using Hadamard product notation

$$E_{\dot{s}}^H E_{\dot{s}} = (D^H P_{\perp} D) \odot (T T^H \otimes 1_M \otimes 1_M^t) . \quad (\text{B.198})$$

Using (B.188) to substitute for $T T^H$ above yields

$$E_{\dot{s}}^H E_{\dot{s}} = (D^H P_{\perp} D) \odot ((H^H H)^{-1} \otimes 1_M \otimes 1_M^t) . \quad (\text{B.199})$$

$$l = k(m - 1) + k \quad m = \text{floor}(l/K) + 1 \quad k = \text{mod}(l, M)$$

$$l = P(m - 1) + k$$

$$i_m \otimes J_{kk} T = (e_m \otimes (e_k \otimes e_k^t)) T = J_{mkk} T . \quad (\text{B.200})$$

$$l_1 = P(m_1 - 1) + k_1 \text{ and } l_2 = P(m_2 - 1) + k_2$$

B.1.7 Proof of Lemma 16 (Equivalence of Forms)

Let Y be the eigen vector matrix corresponding to the signal covariance $R_s = HPH^H = Y\Lambda Y^H$, (346). Then $Y = HT$ for some invertible T so that

$$HPH^H = Y\Lambda Y^H = H(T\Lambda T^H)H^H. \quad (\text{B.201})$$

Identifying parts, we see that source correlation matrix P is equal to $P = T\Lambda T^H$ so that

$$T^{-1}P = \Lambda T^H. \quad (\text{B.202})$$

The signal-plus-noise model covariance is

$$R = HPH^H + \sigma^2 I = Y\Lambda_s Y^H + \sigma^2 Y_\perp Y_\perp^H \quad (\text{B.203})$$

and its inverse is

$$R^{-1} = Y\Lambda_s^{-1}Y^H + \sigma^{-2}Y_\perp Y_\perp^H. \quad (\text{B.204})$$

Substituting this into $P_s = (PH^H R^{-1} HP)$ yields

$$P_s = PH^H (Y\Lambda_s^{-1}Y^H + \sigma^{-2}Y_\perp Y_\perp^H) HP \quad (\text{B.205})$$

$$P_s = PH^H (Y\Lambda_s^{-1}Y^H) HP. \quad (\text{B.206})$$

Noting that $Y^H H = Y^H (YT^{-1}) = T^{-1}$ we have

$$P_s = PT^{-H}\Lambda_s^{-1}T^{-1}P \quad (\text{B.207})$$

substituting (B.202) into the above expression yields the desired result (527)

$$P_s = T\Lambda\Lambda_s^{-1}\Lambda T^H = T\Lambda^2\Lambda_s^{-1} = T\Lambda_{OPT}T^H. \quad (\text{B.208})$$

B.1.8 Proof of Theorem 17

The general form of the $F_{\dot{s}\dot{s}}$ matrix is (471)

$$F_{\dot{s}\dot{s}} = \frac{2K}{\sigma^2} \text{Re}(E_{\dot{s}}^H (\Lambda_{OPT} \otimes I_{N-p}) E_{\dot{s}}) . \quad (\text{B.209})$$

This can be expressed in component form as

$$[F_{\dot{s}\dot{s}}]_{ij} = \frac{2K}{\sigma^2} [\text{Re}(\dot{E}_s^H (\Lambda_{OPT} \otimes I_{N-p}) \dot{E}_s)]_{ij} . \quad (\text{B.210})$$

The factor on the right hand side is equal to

$$(E_s^H (\Lambda_{OPT} \otimes I_{N-p}) E_s)_{ij} = (E_s(:, i)^H (\Lambda_{OPT} \otimes I_{N-p}) E_s(:, j)) . \quad (\text{B.211})$$

Recall that the matrix E_s is given by (518)

$$E_s(:, i) = \text{vec}(\dot{S}_i) = \text{vec}(Y_\perp^H D J_{kk} T) = (T^t \otimes Y_\perp^H D) j_{kk} \quad (\text{B.212})$$

where j_{kk} is the vectorized selection matrix, that is $j_{kk} = \text{vec}(J_{kk})$. Substituting (B.212) into (B.211) yields

$$= j_{ii}^T (T^t \otimes Y_\perp^H D)^H (\Lambda_{OPT} \otimes I_{N-p}) (T^t \otimes Y_\perp^H D) j_{jj} \quad (\text{B.213})$$

$$= j_{ii}^T (T^* \otimes D^H Y_\perp) (\Lambda_{OPT} \otimes I_{N-p}) (T^t \otimes Y_\perp^H D) j_{jj} \quad (\text{B.214})$$

Using the Kronecker product identity (B.250) the previous expression becomes

$$j_{ii}^T \left((T^* \Lambda_{OPT} T^t) \otimes (D^H Y_\perp I_{N-p} Y_\perp^H D) \right) j_{jj} \quad (\text{B.215})$$

and defining $P_\perp \equiv Y_\perp Y_\perp^H$ we have

$$j_{ii}^t \left((T^* \Lambda_{OPT} T^t) \otimes D^H P_\perp D \right) j_{jj} . \quad (\text{B.216})$$

Noting that the factor $((T^* \Lambda_{OPT} T^t) \otimes D^H P_\perp D) j_{jj}$ is of the form $(C^t \otimes A)b$, and using the identity $(C^t \otimes A)b = \text{vec}(ABC)$, we can express the above as

$$((T^* \Lambda_{OPT} T^t) \otimes D^H P_\perp D) j_{jj} = \text{vec}(D^H P_\perp D J_{jj} (T^* \Lambda_{OPT} T^t)) . \quad (\text{B.217})$$

Substituting this form into (B.216) gives

$$= j_{ii}^t \text{vec}(D^H P_\perp D J_{jj} (T \Lambda_{OPT} T^H)) \quad (\text{B.218})$$

where we have used

$$(T^* \Lambda_{OPT} T^t)^t = T \Lambda_{OPT} T^H . \quad (\text{B.219})$$

Using $a^H b = \text{tr}(AB)$, $a = \text{vec}(A)$, $b = \text{vec}(B)$, (B.218) (with $a = j_{ii}$ and $b = \text{vec}(D^H P_\perp D J_{jj} (T \Lambda_{OPT} T^H))$) (B.218) becomes

$$= \text{tr}(J_{ii} D^H P_\perp D J_{jj} (T \Lambda_{OPT} T^H)) . \quad (\text{B.220})$$

Note that

$$J_{ii} D^H P_\perp D J_{jj} = (D^H P_\perp D)_{ij} J_{ij} \quad (\text{B.221})$$

where we have used the identity $J_{ii} R J_{jj} \equiv [R]_{ij} \cdot J_{ij}$ (see (B.260)). Substituting the above into (B.220) and factoring out the scalar term yields

$$= (D^H P_\perp D)_{ij} \text{tr}(J_{ij} T \Lambda_{OPT} T^H) . \quad (\text{B.222})$$

Using the matrix identity (B.256) (note the transposition)

$$(E_s^H (\Lambda_{OPT} \otimes I_{N-p}) E_s)_{ij} = (D^H P_\perp D)_{ij} (T \Lambda_{OPT} T^H)_{ji} \quad (\text{B.223})$$

$$(E_s^H (\Lambda_{OPT} \otimes I_{N-p}) E_s)_{ij} = (D^H P_\perp D)_{ij} (T \Lambda_{OPT} T^H)_{ij} . \quad (\text{B.224})$$

Substituting the above into (B.210) and using the Hadamard product notation this is expressed in matrix form as

$$F_{ss} = \frac{2K}{\sigma^2} \text{Re} \left((D P D) \odot (T \Lambda_{OPT} T^H) \right) . \quad (\text{B.225})$$

Using the the equivalence of forms lemma (527), $P_s = (T \Lambda_{OPT} T^H)$, yields desired result

$$F_{ss} = \frac{2K}{\sigma^2} \text{Re} \left(D^H P_\perp D \odot P_s \right) . \quad (\text{B.226})$$

B.1.9 Proof of Theorem 19

F_{ss} for the modal model $M \geq 1$ case is

$$F_{ss} = \frac{2K}{\sigma^2} \text{Re} \left((D^H P_\perp D) \odot P_s^{(M)} \right) \quad (\text{B.227})$$

$$P_s^{(M)} \equiv P_s \otimes (1_M \otimes I_M) \quad (\text{B.228})$$

where $P_s = (PH^H R^{-1}HP)$ (see (526)) .

The general form of the $F_{\dot{s}\dot{s}}$ matrix is (471)

$$F_{\dot{s}\dot{s}} = \frac{2K}{\sigma^2} \text{Re}(E_{\dot{s}}^H (\Lambda_{OPT} \otimes I_{N-p}) E_{\dot{s}}) . \quad (\text{B.229})$$

The factor $E_{\dot{s}}^H (\Lambda_{OPT} \otimes I_{N-p}) E_{\dot{s}}$ is given in component form as

$$[E_{\dot{s}}^H (\Lambda_{OPT} \otimes I_{N-p}) E_{\dot{s}}]_{l_1 l_2} = E_{\dot{s}}^H(:, l_1) (\Lambda_{OPT} \otimes I_{N-p}) E_{\dot{s}}(:, l_2) . \quad (\text{B.230})$$

For the natural coordinate basis, the matrix $E_{\dot{s}}$ in the general case is given by (547)

$$E_{\dot{s}}(:, l) = (T^t \otimes Y_{\perp}^H D) j_{mjj}; \quad l = p(m-1) + j . \quad (\text{B.231})$$

Substituting the above into (B.230) with the two indexes $l_1 = p(m_1 - 1) + i_1$, and $l_2 = p(m_2 - 1) + i_2$ yields

$$j_{m_1 i_1 i_1}^t (T^t \otimes Y_{\perp}^H D)^H (\Lambda_{OPT} \otimes I_{N-p}) (T^t \otimes Y_{\perp}^H D) j_{m_2 j_2 j_2} \quad (\text{B.232})$$

$$j_{m_1 i_1 i_1}^t (T^* \otimes D^H Y_{\perp}) (\Lambda_{OPT} \otimes I_{N-p}) (T^t \otimes Y_{\perp}^H D) j_{m_2 j_2 j_2} . \quad (\text{B.233})$$

Using the Kronecker product identity (B.250) this is

$$j_{m_1 i_1 i_1}^t (T^* \Lambda_{OPT} T^t \otimes D^H P_{\perp} D) j_{m_2 j_2 j_2} . \quad (\text{B.234})$$

Using $P_s = (T \Lambda_{OPT} T^H)$ in the above yields

$$j_{m_1 i_1 i_1}^t (P_s^t \otimes D^H P_{\perp} D) j_{m_2 j_2 j_2} . \quad (\text{B.235})$$

Noting that $(P_s^t \otimes D^H P_{\perp} D) j_{kii}$ is in the form $(C^T \otimes A)b$ and we can express it as $\text{vec}(ABC)$. Substituting this form into (B.235) gives

$$j_{m_1 i_1 i_1}^t \text{vec}(D^H P_{\perp} D J_{m_2 j_2 j_2} P_s) . \quad (\text{B.236})$$

Next using the fact that $a^H b = \text{tr}(AB)$, $a = \text{vec}(A)$, $b = \text{vec}(B)$ we have

$$= \text{tr}[J_{m_1 i_1 i_1}^t D^H P_\perp D J_{m_2 j_2 j_2} P_s] . \quad (\text{B.237})$$

Using (B.263) the factor appearing in the above can be expressed as

$$J_{m_1 i_1 i_1}^t D^H P_\perp D J_{m_2 i_2 i_2} = [D^H P_\perp D]_{l_1 l_2} J_{i_1 i_2} \quad (\text{B.238})$$

and substituting this into the above and factoring out the scalar term yields

$$= [D^H P_\perp D]_{l_1 l_2} \text{tr}[J_{i_1 i_2} P_s] . \quad (\text{B.239})$$

Using (B.256) (note the transposition)

$$X_{l_1 l_2} = [D^H P_\perp D]_{l_1 l_2} [P_s]_{i_2 i_1} . \quad (\text{B.240})$$

To see the indexing, write out the first two rows of this matrix for the $p = 3$, $M = 2$ case

$$X_{1,1} = [D^H P_\perp D]_{11} [P_s]_{11} \quad (\text{B.241})$$

$$X_{1,2} = [D^H P_\perp D]_{12} [P_s]_{12} \quad (\text{B.242})$$

$$X_{1,3} = [D^H P_\perp D]_{13} [P_s]_{13} \quad (\text{B.243})$$

$$X_{1,4} = [D^H P_\perp D]_{14} [P_s]_{11} \quad (\text{B.244})$$

$$X_{1,5} = [D^H P_\perp D]_{15} [P_s]_{12} \quad (\text{B.245})$$

$$X_{1,6} = [D^H P_\perp D]_{16} [P_s]_{13} . \quad (\text{B.246})$$

This has the form

$$(P_s \otimes (1_M \otimes 1_M)) = \begin{bmatrix} P_s & P_s \\ P_s & P_s \end{bmatrix} . \quad (\text{B.247})$$

For the $M = 3$ case

$$(P_s \otimes (1_M \otimes 1_M)) = \begin{bmatrix} P_s & P_s & P_s \\ P_s & P_s & P_s \\ P_s & P_s & P_s \end{bmatrix} . \quad (\text{B.248})$$

The general case is given by

$$= ((D^H P_\perp D) \odot (P_s \otimes (1_M \otimes 1_M^t))) . \quad (\text{B.249})$$

B.1.10 Useful Matrix Identities

$$(A \otimes B)(C \otimes D) = (AC \otimes BD) \quad (\text{B.250})$$

$$(A \otimes B)^H = (A^H \otimes B^H) \quad (\text{B.251})$$

Notation $a = \text{vec}(A)$, $b = \text{vec}(B)$

$$\text{tr}(AB) = \text{tr}(BA) \quad (\text{B.252})$$

$$\text{tr} B_i^H B_j = b_i^H b_j \quad (\text{B.253})$$

Selection Matrix Identities In this appendix let $\mathbf{i}_k$ denote the $N \times 1$ equal to one at the k th location and zero elsewhere. Let The $p \times p$ selection matrix J_{ij} is define as

$$J_{(ij)} \equiv \mathbf{e}_i \otimes \mathbf{e}_i^T. \quad (\text{B.254})$$

This matrix has 1 in the ij th location and zeros elsewhere. The vectorized selection matrix $j_{(ij)} \equiv \text{vec}(J_{(ij)})$ is $p^2 \times 1$ and equal to one at the $(i-1)p + j$ location and zeros elsewhere.

Matrix Identities and Examples

$$AJ_{kk} = A(:, k) \otimes \mathbf{e}_k^H \quad (\text{B.255})$$

$$\text{tr}(J_{ij}R) = [R]_{ji} \quad (\text{B.256})$$

$$\text{tr}(cJ_{ij}R) = c[R]_{ji} \quad (\text{B.257})$$

$$\text{tr}(RJ_{ij}) = [R]_{ij} \quad (\text{B.258})$$

$$J_{kk}A = J_{kk} \begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix} = \begin{bmatrix} 0 \\ c_k \\ 0 \end{bmatrix} \quad (\text{B.259})$$

$$J_{ii}RJ_{jj} \equiv [R]_{ij} \cdot J_{ij} \quad (\text{B.260})$$

I.5 ij the element of a matrix

$$[D^H R D]_{ij} = d_i^H R d_j \quad (\text{B.261})$$

d_j is the j th column of D

The selection matrix J_{mij} is an $Mp \times p$ matrix defined as

$$J_{mij} \equiv \mathbf{i}_m \otimes (\mathbf{e}_i \otimes \mathbf{e}_j^t), \quad m = 1 \cdots M. \quad (\text{B.262})$$

Let R be any $Mp \times Mp$ matrix so that the product is the $p \times p$ matrix given by

$$(J_{m_1 k_1 k_1})^T R \cdot J_{m_2 k_2 k_2} = R_{l_1 l_2} \cdot J_{k_1 k_2} \quad (\text{B.263})$$

where $l_1 = p(m_1 - 1) + k_1$ and $l_2 = p(m_2 - 1) + k_2$ In this expression $R_{l_1 l_2}$ is a scalar and $J_{k_1 k_2}$ is the $p \times p$ selection matrix

$$J_{k_1 k_2} = \mathbf{e}_{k_1} \otimes \mathbf{e}_{k_2}^t \quad (\text{B.264})$$

Matrix Identity $\text{tr}(D^2 A^H B) = a^H (D^2 \otimes I_M) b$

Lemma 27. *Let A and B be arbitrary $M \times p$ complex matrices and D an arbitrary $p \times p$ diagonal matrix. Then*

$$\text{tr}(D^2 A^H B) = a^H (D^2 \otimes I_M) b \quad (\text{B.265})$$

where

$$a \equiv \text{vec}(A); b \equiv \text{vec}(B) \quad (\text{B.266})$$

Proof:

$$\text{tr}(D^2 A^H B) = \text{tr}(A^H \cdot B D^2) \quad (\text{B.267})$$

Using general fact that $\text{tr}(X^H Y) = \text{vec}(X)^H \text{vec}(Y)$ we have

$$= \text{vec}(A)^H \text{vec}(B D^2) = a^H \text{vec}(I_M B D^2) \quad (\text{B.268})$$

Applying the standard matrix identity $\text{vec}(ABC) = (C^t \otimes A)$ to the second factor yields the result (B.265).

$$= a^H (D^2 \otimes I_M) b \quad (\text{B.269})$$

Corrollary 28. Let the complex vector above be represented as

$$\alpha \equiv \begin{bmatrix} \alpha_R \\ \alpha_I \end{bmatrix} \equiv \begin{bmatrix} \text{Re}(a) \\ \text{Im}(a) \end{bmatrix}; \quad \beta \equiv \begin{bmatrix} \beta_R \\ \beta_I \end{bmatrix} \equiv \begin{bmatrix} \text{Re}(b) \\ \text{Im}(b) \end{bmatrix}. \quad (\text{B.270})$$

Then from the previous lemma (B.265)

$$\text{Re}[\text{tr}(D^2 A^H B)] = \alpha^T \begin{bmatrix} (D^2 \otimes I_M) & 0 \\ 0 & (D^2 \otimes I_M) \end{bmatrix} \beta \quad (\text{B.271})$$

or equivalently

$$\text{Re}[\text{tr}(D^2 A^H B)] = \alpha^T (I_2 \otimes (D^2 \otimes I_M)) \beta. \quad (\text{B.272})$$

Proof: From the definition of α and β above $a = \begin{bmatrix} I_{Mp} & iI_{Mp} \end{bmatrix} \alpha$ and $b = \begin{bmatrix} I_{Mp} & iI_{Mp} \end{bmatrix} \beta$. Substituting these into (B.265) yields

$$= \alpha^t \begin{bmatrix} I_{Mp} \\ -iI_{Mp} \end{bmatrix} (D^2 \otimes I_M) \begin{bmatrix} I_{Mp} & iI_{Mp} \end{bmatrix} \beta. \quad (\text{B.273})$$

Note that $(D^2 \otimes I_M)$ is $Mp \times Mp$

$$= \alpha^t \begin{bmatrix} I_{Mp} \\ -iI_{Mp} \end{bmatrix} \begin{bmatrix} (D^2 \otimes I_M) & i(D^2 \otimes I_M) \end{bmatrix} \beta \quad (\text{B.274})$$

$$= \alpha^t \begin{bmatrix} (D^2 \otimes I_M) & i(D^2 \otimes I_M) \\ -i(D^2 \otimes I_M) & (D^2 \otimes I_M) \end{bmatrix} \begin{bmatrix} \beta_R \\ \beta_I \end{bmatrix} \quad (\text{B.275})$$

$$= \begin{bmatrix} \alpha_R^t & \alpha_I^t \end{bmatrix} \begin{bmatrix} (D^2 \otimes I_M)\beta_R + i(D^2 \otimes I_M)\beta_I \\ -i(D^2 \otimes I_M)\beta_R + (D^2 \otimes I_M)\beta_I \end{bmatrix} \quad (\text{B.276})$$

$$= \alpha_R^t (D^2 \otimes I_M)\beta_R + i\alpha_R^t (D^2 \otimes I_M)\beta_I - i\alpha_I^t (D^2 \otimes I_M)\beta_R + \alpha_I^t (D^2 \otimes I_M)\beta_I. \quad (\text{B.277})$$

Taking the real part yields

$$\text{Re}[\text{tr}(D^2 A^H B)] = \alpha_R^t (D^2 \otimes I_M)\beta_R + \alpha_I^t (D^2 \otimes I_M)\beta_I. \quad (\text{B.278})$$

Expressed in the matrix form this is

$$\text{Re}[tr(D^2 A^H B)] = \begin{bmatrix} \alpha_R^t & \alpha_I^t \end{bmatrix} \begin{bmatrix} (D^2 \otimes I_M) & 0 \\ 0 & (D^2 \otimes I_M) \end{bmatrix} \begin{bmatrix} \beta_I \\ \beta_R \end{bmatrix} \quad (\text{B.279})$$

Using (B.270) yields the desired result (B.271).

List of References

- [1] S. Smith, "Covariance, subspace, and intrinsic cramér-rao bounds," *IEEE Trans. Signal Process.*, vol. 53, no. 5, p. 1610 1630, 2005.

APPENDIX C

appendix C

C.1 Proof of Lemma 20

From lemma B.1.1 the log likelihood function along the geodesic $Y(t)$ emanating from $Y(0) = Y$ in the direction $\Delta = Y A + Y_{\perp} B$ is

$$L(Y(t)) = L(Y(0)) + \delta L(t) \quad (\text{C.1})$$

where

$$\delta L(t) = K \cdot \left(t \cdot \text{tr}((\delta R_1(A, B)) R^{-1} (I - \tilde{R} R^{-1}) + O(t^2) \right) . \quad (\text{C.2})$$

The directional derivative in the $\Delta = Y A + Y_{\perp} B$ direction therefore is

$$\nabla L[\Delta] \equiv \frac{dL(Y(t))}{dt} = \frac{d\delta L(t)}{dt} . \quad (\text{C.3})$$

Substituting for $\delta L(t)$ and differentiating yields the desired result

$$\nabla L[\Delta] \equiv \frac{dL(Y(t))}{dt} = K \cdot \text{tr}((\delta R_1(A, B)) R^{-1} (I - \tilde{R} R^{-1}) \quad (\text{C.4})$$

or equivalently

$$\nabla L[\Delta] = K \cdot \text{tr}(\delta R_1(A, B) R^{-1}) - \text{tr}(\delta R_1(A, B) \tilde{R} R^{-1}) . \quad (\text{C.5})$$

For the specific coordinate basis vector $\Delta = \mathbf{b}_k = Y_{\perp} B_k$ (C.5) is

$$\nabla L[\mathbf{b}_k] = \text{tr}(\delta R_1(0, B_k) R^{-1}) - \text{tr}(\delta R_1(0, B_k) \tilde{R} R^{-1}) \quad (\text{C.6})$$

To compute the directional derivative in the coordinate directions recall (410)

$$\delta R_1(A, B) = Y A \Lambda Y^H + Y \Lambda A^H Y^H + Y_{\perp} B \Lambda Y^H + Y \Lambda B^H Y_{\perp}^H . \quad (\text{C.7})$$

For notional simplicity, in the remainder of the calculation we drop the subscript on the matrix B_k . The factor $\delta R_1(0, B) R^{-1}$ is therefore (see (B.80))

$$\delta R_1(0, B) R^{-1} = Y_{\perp} B \Lambda \Lambda_s^{-1} Y^H + \sigma^{-2} Y \Lambda B^H Y_{\perp}^H \quad (\text{C.8})$$

where we have used $R^{-1} = Y\Lambda_s^{-1}Y^H + \sigma^{-2}Y_{\perp}Y_{\perp}^H$. Rearranging terms within the trace operator results in

$$\text{tr}(\delta R_1(0, B)R^{-1}) = 0 \quad (\text{C.9})$$

which follows since $Y^HY_{\perp} = 0$.

The second term in (C.6) is a function of the data through the $\tilde{R}$, the SCM. Re-ordering (C.6) within the trace operator, and substituting for $\delta R_1(0, B)$ this term is

$$\text{tr}(R^{-1}\delta R_1(0, B)\tilde{R}) = \text{tr}(R^{-1}(Y_{\perp}B\Lambda_s^{-1}Y^H + \sigma^{-2}Y\Lambda B^HY_{\perp}^H)\tilde{R}) \quad (\text{C.10})$$

$$= \text{tr}(\sigma^{-2}(Y_{\perp}B\Lambda\Lambda_s^{-1}Y^H + Y\Lambda_s^{-1}\Lambda B^HY_{\perp}^H)\tilde{R}) \quad (\text{C.11})$$

$$= \text{tr}(\sigma^{-2}(Y_{\perp}B\Lambda\Lambda_s^{-1}Y^H\tilde{R} + Y\Lambda_s^{-1}\Lambda B^HY_{\perp}^H\tilde{R})) \quad (\text{C.12})$$

$$= \text{tr}(\sigma^{-2}(B\Lambda\Lambda_s^{-1}Y^H\tilde{R}Y_{\perp} + \Lambda_s^{-1}\Lambda B^HY_{\perp}^H\tilde{R}Y)) . \quad (\text{C.13})$$

Defining $C \equiv B\Lambda\Lambda_s^{-1}$ the above is

$$= \sigma^{-2}\text{tr}((CY^H\tilde{R}Y_{\perp} + C^HY_{\perp}^H\tilde{R}Y)) \quad (\text{C.14})$$

$$= \sigma^{-2}2\text{Re}\left(\text{tr}(CY^H\tilde{R}Y_{\perp})\right) . \quad (\text{C.15})$$

Replacing C by its definition above and returning the index to B yields the desired result (590)

$$\nabla L[\mathbf{b}_k] = \sigma^{-2}2\text{Re}\left(\text{tr}(\Lambda\Lambda_s^{-1}Y^H\tilde{R}Y_{\perp}B_k)\right) . \quad (\text{C.16})$$

Substituting $\Delta = \mathbf{a}_k = YA_k$ where A_k is anti-Hermitian into (C.5), yields

$$\nabla L[\mathbf{a}_k] = \text{tr}(\delta R_1(A_k, 0)R^{-1}) - \text{tr}(\delta R_1(A_k, 0)\tilde{R}R^{-1}) . \quad (\text{C.17})$$

For notional simplicity, in the remainder of the calculation we drop the subscript on the matrix A_k . Using (C.7) the factor $\delta R_1(A, 0)R^{-1}$ evaluates to

$$\delta R_1(A, 0)R^{-1} = (Y\Lambda Y^H + Y\Lambda A^HY^H)R^{-1} \quad (\text{C.18})$$

$$= Y A \Lambda \Lambda_s^{-1} Y^H + Y \Lambda A^H \Lambda_s^{-1} Y^H . \quad (\text{C.19})$$

Using this expression the first term in (C.17) becomes

$$= \text{tr}(Y A \Lambda \Lambda_s^{-1} Y^H + Y \Lambda A^H \Lambda_s^{-1} Y^H) \quad (\text{C.20})$$

$$= \text{tr}(A \Lambda \Lambda_s^{-1} Y^H Y + \Lambda A^H \Lambda_s^{-1} Y^H Y) \quad (\text{C.21})$$

$$= \text{tr}(A \Lambda \Lambda_s^{-1} + \Lambda A^H \Lambda_s^{-1}) \quad (\text{C.22})$$

$$= \text{tr}(\Lambda \Lambda_s^{-1} (A + A^H)) \quad (\text{C.23})$$

$$= 0 . \quad (\text{C.24})$$

This follows since A is anti-Hermitian (i.e., $A = -A^H$).

The second term in (C.6) is a function of the data $\tilde{R}$ (SCM). Reordering within the trace operator, and substituting for $\delta R_1(A, 0)$ this term is

$$\text{tr}(R^{-1} \delta R_1(A, 0) \tilde{R}) = \text{tr}(R^{-1} (Y A \Lambda \Lambda_s^{-1} Y^H + Y \Lambda A^H \Lambda_s^{-1} Y^H) \tilde{R}) . \quad (\text{C.25})$$

Carrying out the multiplication with the expanded form of R^{-1} yields the desired result (591)

$$\nabla L[\mathbf{a}_k] = \text{tr}(Y (\Lambda_s^{-1} A \Lambda \Lambda_s^{-1} + \Lambda_s^{-1} \Lambda A^H \Lambda_s^{-1}) Y^H \tilde{R}) \quad (\text{C.26})$$

C.2 Proof of Lemma 21

Proof: The derivative in the $\mathbf{b}_k$ direction is (590)

$$\nabla L[\mathbf{b}_k] = \sigma^{-2} 2 \text{tr}(\text{Re}(\Lambda \Lambda_s^{-1} \cdot Y^H \tilde{R} Y_{\perp} \cdot B_k)) \quad (\text{C.27})$$

or rearranging terms

$$\nabla L[\mathbf{b}_k] = \sigma^{-2} 2 \text{Re} \left(\text{tr} \left((B_k \Lambda \Lambda_s^{-1}) Y^H \tilde{R} Y_{\perp} \right) \right) . \quad (\text{C.28})$$

For arbitrary B_k , this will vanish if the $N \times p$ matrix Y is chosen such that

$$Y_{\perp}^H \tilde{R} Y = 0 . \quad (\text{C.29})$$

Writing $\tilde{R}$ in eigen decomposition partitioned form the above becomes

$$Y_{\perp}^H \begin{bmatrix} Y_1 & Y_2 \end{bmatrix} \begin{bmatrix} \Lambda_1 & 0 \\ 0 & \Lambda_2 \end{bmatrix} \begin{bmatrix} Y_1^H \\ Y_2^H \end{bmatrix} Y = 0. \quad (\text{C.30})$$

In this expression, Y_1 is an $N \times p$ matrix whose columns are p arbitrarily selected eigenvectors of $\tilde{R}$ and Y_2 is the $N \times (N - p)$ matrix of the remaining eigenvectors. The above equation is satisfied if Y is chosen as

$$Y = Y_1 Q \quad (\text{C.31})$$

for Q an arbitrary $p \times p$ unitary matrix. To verify this substitute (C.31) into (C.30)

$$= Y_{1\perp}^H \begin{bmatrix} Y_1 & Y_2 \end{bmatrix} \begin{bmatrix} \Lambda_1 & 0 \\ 0 & \Lambda_2 \end{bmatrix} \begin{bmatrix} Y_1^H \\ Y_2^H \end{bmatrix} Y_1 Q \quad (\text{C.32})$$

$$= \begin{bmatrix} 0 & Y_{1\perp}^H Y_2 \end{bmatrix} \begin{bmatrix} \Lambda_1 & 0 \\ 0 & \Lambda_2 \end{bmatrix} \begin{bmatrix} Q \\ 0 \end{bmatrix} = 0_{p \times p}. \quad (\text{C.33})$$

This establishes that any set of p -eigenvectors Y_1 (not necessarily the p -dominant eigenvectors) and all their rotations, (C.31), satisfies the first condition $\nabla L[\mathbf{b}_k] = 0$. Substituting this solution, that is $Y = Y_1 Q$, into the second equation $\nabla L[\mathbf{a}_k]$ yields

$$\nabla L[\mathbf{a}_k] = \text{tr}(Y(\Lambda_s^{-1} A \Lambda_s^{-1} + \Lambda_s^{-1} \Lambda A^H \Lambda_s^{-1}) Q^H \tilde{\Lambda}_s Q Y^H) \quad (\text{C.34})$$

$$= \text{tr}((\tilde{\Lambda}_s^{-1} A \Lambda \tilde{\Lambda}_s^{-1} + \tilde{\Lambda}_s^{-1} \Lambda A^H \tilde{\Lambda}_s^{-1}) Q^H \tilde{\Lambda}_s Q). \quad (\text{C.35})$$

In general, this term is non-zero for arbitrary choice of valid A and Q . If Q is restricted to be unitary diagonal then the factor $Q \tilde{\Lambda}_s Q^H$ appearing above is

$$Q \tilde{\Lambda}_s Q^H = \tilde{\Lambda}_s Q Q^H = \tilde{\Lambda}_s \quad (\text{C.36})$$

which follows since diagonal matrices commute. Using this result we have

$$\nabla L[\mathbf{a}_k] = \text{tr}((\Lambda_s^{-1} A \Lambda_s^{-1} + \Lambda_s^{-1} \Lambda A^H \Lambda_s^{-1}) \tilde{\Lambda}_s). \quad (\text{C.37})$$

Evaluating the parameter Λ_s at its ML value

$$\Lambda_s = \tilde{\Lambda}_s \quad (\text{C.38})$$

(C.37) reduces to

$$= \text{tr}(\Lambda \Lambda_s^{-1} A + \Lambda_s^{-1} \Lambda A^H) \quad (\text{C.39})$$

$$= \text{tr}(\Lambda \Lambda_s^{-1} (A + A^H)) . \quad (\text{C.40})$$

This vanishes for arbitrary Λ and Λ_s , since by hypothesis the $p \times p$ matrix A is anti-Hermitian so that $A + A^H = 0$. The choice of $Y \in \mathcal{B}$ that simultaneously satisfies both sets of equations is

$$\hat{Y} = \tilde{Y} \Lambda_Q . \quad (\text{C.41})$$

We next show that the maximum value of likelihood function is achieved by choosing the p eigenvectors of Y_1 to correspond to the p -largest magnitude eigenvalues. The value of the likelihood function for this choice is

$$L(\tilde{R}, R) = -K(\text{tr}(R^{-1} \tilde{R}) + \log \det R) \quad (\text{C.42})$$

$$R^{-1} \tilde{R} = Q_1^{-1} Q \quad (\text{C.43})$$

$$R^{-1} = Y \Lambda_s^{-1} Y^H + \sigma^{-2} (I - Y Y^H) . \quad (\text{C.44})$$

To maximize, we choose the set of p vectors corresponding to the largest p eigenvalues.

C.3 Proof of lemma 23

Consider the damped exponential signal waveform $h(n) = z^n$ where $z = \rho \exp(i\theta)$. Then

$$d_\rho(n) \equiv \frac{\partial h}{\partial \rho} = n z^{n-1} \frac{\partial z}{\partial \rho} = n z^n \rho^{-1} \equiv \rho^{-1} n h(n) \quad (\text{C.45})$$

$$d_\theta(n) \equiv \frac{\partial h(n)}{\partial \theta} = n z^{n-1} \frac{\partial z}{\partial \theta} = i n z^n \equiv i n h(n) \quad (\text{C.46})$$

where we have used

$$\frac{\partial z}{\partial \rho} = \exp(i\theta) = z\rho^{-1} \quad \frac{\partial z}{\partial \theta} = i\rho \exp(i\theta) = iz \quad (\text{C.47})$$

Comparing (C.46) and (C.45) see that

$$d_\theta = i\rho d_\rho . \quad (\text{C.48})$$

The vectors d_θ and d_ρ are independent in the real vector space formulation since $d_\theta \neq c \cdot d_\rho$ for any real constant c . The $N \times q$, matrix derivatives D is defined as

$$D = \begin{bmatrix} D_\theta & D_\rho \end{bmatrix} \quad (\text{C.49})$$

where $q = 2Mp$ and

$$D_\theta = \begin{bmatrix} d_{\theta_1}, \dots, d_{\theta_{Mp}} \end{bmatrix} . \quad (\text{C.50})$$

The natural coordinate basis is (518) $\mathbf{s}_k = Y_\perp \check{S}_k$ where

$$\check{S}_{\rho;k} = Y_\perp^H D_\rho J_{kk} T; \quad k = 1 \dots \bar{q} \quad (\text{C.51})$$

$$\check{S}_{\theta;k} = Y_\perp^H D_\theta J_{kk} T \quad k = 1 \dots \bar{q} \quad (\text{C.52})$$

where $\bar{q} = Mp$. The basis matrix E_s has the partitioned form

$$E_s = \begin{bmatrix} E_{s;\theta} & E_{s;\rho} \end{bmatrix} \quad (\text{C.53})$$

where

$$E_{s;\theta} \equiv \begin{bmatrix} \text{vec}(\check{S}_{\theta;1}) & \dots & \text{vec}(\check{S}_{\theta;p}) \end{bmatrix} \quad (\text{C.54})$$

and

$$E_{s;\rho} \equiv \begin{bmatrix} \text{vec}(\check{S}_{\rho;1}) & \dots & \text{vec}(\check{S}_{\rho;p}) \end{bmatrix} . \quad (\text{C.55})$$

Since J_{kk} selects the k th column, using (C.48), we see that in vectorized form

$$\text{vec}(\check{S}_{\theta;k}) = i \cdot \rho_k \text{vec}(\check{S}_{\rho;k}) . \quad (\text{C.56})$$

The columns of the first set are scaled version of the second set, and therefore spanned by $\bar{E}_s$. Therefore, if $\bar{E}_s$ is the orthogonal basis for $E_{s;\theta}$, that is

$$\bar{E}_s \equiv \text{orth}(E_{s;\theta}) \quad (\text{C.57})$$

then $i \cdot \bar{E}_s$ is a basis for the set $E_{s;\rho}$. The real vector space S_Y therefore has the orthonormal basis given by

$$E_s = \begin{bmatrix} \bar{E}_s & i \cdot \bar{E}_s \end{bmatrix} \quad (\text{C.58})$$

or expressed alternatively as

$$E_s = \bar{E}_s E_0 \quad (\text{C.59})$$

where $E_0 \equiv \begin{bmatrix} I & iI \end{bmatrix}$.

C.4 Proof of theorem 24

Substituting (C.58) into the first derivative defined by (660) yields

$$L_s = \text{Re} \left(E_s^H \tilde{\Lambda}_{bb} b_n \right) = \text{Re} \left(\begin{bmatrix} \bar{E}_s^H \\ -i \bar{E}_s^H \end{bmatrix} \tilde{\Lambda}_{bb} b_n \right) \quad (\text{C.60})$$

$$L_s = \begin{bmatrix} \text{Re} \left(\bar{E}_s^H \tilde{\Lambda}_{bb} b_n \right) \\ \text{Im} \left(\bar{E}_s^H \tilde{\Lambda}_{bb} b_n \right) \end{bmatrix} = \begin{bmatrix} \text{Re} \left(\bar{L}_s \right) \\ \text{Im} \left(\bar{L}_s \right) \end{bmatrix} \quad (\text{C.61})$$

where we have defined

$$\bar{L}_s \equiv \bar{E}_s^H \tilde{\Lambda}_{bb} b_n. \quad (\text{C.62})$$

Substituting (C.58) into the second derivative defined by (659) yields

$$L_{ss} = \text{Re} \left(E_s^H \tilde{\Lambda}_{bb} E_s \right) = \text{Re} \left(\begin{bmatrix} \bar{E}_s^H \tilde{\Lambda}_{bb} \bar{E}_s & i \bar{E}_s^H \tilde{\Lambda}_{bb} \bar{E}_s \\ -i \bar{E}_s^H \tilde{\Lambda}_{bb} \bar{E}_s & \bar{E}_s^H \tilde{\Lambda}_{bb} \bar{E}_s \end{bmatrix} \right) \quad (\text{C.63})$$

Defining ¹

$$\bar{L}_{ss} \equiv \bar{E}_s^H \tilde{\Lambda}_{bb} \bar{E}_s \quad (\text{C.64})$$

¹ Note that $\bar{L}_{ss}$ is real since $\tilde{\Lambda}_{bb}$ is a real diagonal matrix. $\text{Re}(i \cdot X) = -\text{Im}(X)$

we have

$$L_{ss} = \begin{bmatrix} \text{Re}(\bar{L}_{ss}) & -\text{Im}(\bar{L}_{ss}) \\ \text{Im}(\bar{L}_{ss}) & \text{Re}(\bar{L}_{ss}) \end{bmatrix}. \quad (\text{C.65})$$

² The inverse is given by

$$L_{ss}^{-1} = \frac{1}{2} \begin{bmatrix} \text{Re}(\bar{L}_{ss}^{-1}) & \text{Im}(\bar{L}_{ss}^{-1}) \\ -\text{Im}(\bar{L}_{ss}^{-1}) & \text{Re}(\bar{L}_{ss}^{-1}) \end{bmatrix}. \quad (\text{C.66})$$

This expression for L_{ss}^{-1} is verified by direct multiplication and noting that

$$\text{Re}(\bar{L}_{ss})\text{Re}(\bar{L}_{ss}^{-1}) + \text{Im}(\bar{L}_{ss})\text{Im}(\bar{L}_{ss}^{-1}) = 2 \cdot I \quad (\text{C.67})$$

$$-\text{Re}(\bar{L}_{ss})\text{Im}(\bar{L}_{ss}^{-1}) + \text{Im}(\bar{L}_{ss})\text{Re}(\bar{L}_{ss}^{-1}) = 0 \quad (\text{C.68})$$

which follows since multiplication of symmetric matrices commutes ($\bar{L}_{ss}$ is Hermitian).

Partitioning L_s as

$$L_s = \begin{bmatrix} L_{s;R} \\ L_{s;I} \end{bmatrix} \quad (\text{C.69})$$

and using (C.66) the product $L_{ss}^{-1}L_s$ is expressed as

$$s = L_{ss}^{-1}L_s = \frac{1}{2} \begin{bmatrix} \text{Re}(\bar{L}_{ss}^{-1})L_{s;R} - \text{Im}(\bar{L}_{ss}^{-1})L_{s;I} \\ \text{Im}(\bar{L}_{ss}^{-1})^t L_{s;R} + \text{Re}(\bar{L}_{ss}^{-1})L_{s;I} \end{bmatrix}. \quad (\text{C.70})$$

The real and imaginary components of the complex form

$$\bar{s} = \bar{L}_{ss}^{-1}\bar{L}_s \quad (\text{C.71})$$

are given by ³

$$\text{Re}(\bar{s}) = \text{Re}(\bar{L}_{ss}^{-1})\text{Re}(\bar{L}_s) - \text{Im}(\bar{L}_{ss}^{-1})\text{Im}(\bar{L}_s) \quad (\text{C.72})$$

$$\text{Im}(\bar{s}) = \text{Im}(\bar{L}_{ss}^{-1})\text{Re}(\bar{L}_s) + \text{Re}(\bar{L}_{ss}^{-1})\text{Im}(\bar{L}_s) \quad (\text{C.73})$$

Comparing with (C.70) conclude that

$$s = \frac{1}{2} \begin{bmatrix} \text{Re}(\bar{s}) \\ \text{Im}(\bar{s}) \end{bmatrix} \quad (\text{C.74})$$

²(C.63) is symmetric since the term in the bracket is Hermitian since $(i\bar{E}_s^H L_{bb}\bar{E}_s)^H = -ii\bar{E}_s^H L_{bb}\bar{E}_s$

³ $\bar{L}_{ss}^{-1} = \text{Re}(\bar{L}_{ss}^{-1}) + i \cdot \text{Im}(\bar{L}_{ss}^{-1})$

where

$$\bar{L}_{ss}^{-1} \equiv \text{Re}(\bar{L}_{ss}^{-1}) + i \cdot \text{Im}(\bar{L}_{ss}^{-1}) \quad (\text{C.75})$$

and

$$\bar{L}_s = L_{s;R} + i \cdot L_{s;I} . \quad (\text{C.76})$$

C.5 Proof of Lemma 25

The expanded form of (C.79) is

$$\Phi_k(Y) \equiv ((J_k Y)^H J_k Y)^{-1} (J_k Y)^H J_0 Y \quad (\text{C.77})$$

$$\equiv (Y^H J_k^t J_k Y)^{-1} Y^H J_k^t J_0 Y . \quad (\text{C.78})$$

Substituting YQ for Y yields

$$\Phi_k(YQ) \equiv Q^H (Y^H J_k^t J_k Y)^{-1} Y^H J_k^t J_0 Y Q = Q^H \Phi_k(Y) Q . \quad (\text{C.79})$$

As a result, (687) for YQ is

$$X_k(YQ) \equiv (J_k YQ \cdot Q^H \Phi_k(YQ) - J_0 YQ) = (J_k Y \Phi_k(Y) - J_0 Y) Q \quad (\text{C.80})$$

$$X_k(YQ) = X_k(Y) Q . \quad (\text{C.81})$$

Therefore if $X_k(Y) = 0$ then $X_k(YQ) = 0$.

C.6 MLE using Reference Solution and Geometric Newton's Method

The discussion in section 4.2 made use of a reference point Y_0 on the submanifold $\bar{\mathcal{B}}$. Given an S-N Basis at this point we established the submanifold membership condition (641) as (655) . When the constraint submanifold is defined by a parametric model the reference point Y_0 may be specified by a reference parameter θ_0 via the singular vector matrix of $H = H(\theta_0)$. The basis vectors that define the tangent space $T_{Y_0} \bar{\mathcal{B}}$ are derived from the differentiable signal model $H(\theta)$. The reference parameter θ_0 is typically provided by a computationally efficient suboptimal estimator (e.g., ESPRIT)

or is the result of a previous iteration as in a tracking application. Using this reference point approach, the estimator in section 4.2 may be transformed into a geometric form of Newton's method, with the q -tuple θ parameter interpreted as a coordinate system on $\bar{\mathcal{B}}$.

The formulation of the S-N basis approach leading to the solution form (655) required a basis for the space $S_{\tilde{Y}}$. Notionally this space was defined by parallel translating the basis vectors for S_{Y_0} , where $Y_0 \in \bar{\mathcal{B}}$, to the unconstrained MLE $\tilde{Y}$ (in general $\tilde{Y} \notin \bar{\mathcal{B}}$). Using this, the constrained MLE $\hat{Y}$ was expressed in terms of exponential map at $\tilde{Y}$ (664) and the tangent vector $\hat{\Delta}$ (665).

Given a reference solution, it is preferable to develop the ML estimate using an exponential map based at Y_0 rather than $\tilde{Y}$. That is, we seek the estimate in the form $\hat{Y} = \exp_{Y_0} \hat{\Delta}$ where

$$\hat{\Delta} = \sum_{k=1}^p \acute{s}_k \widehat{\delta\theta}^k, \quad (\text{C.82})$$

and where $\acute{s}_k$ is natural coordinate basis at Y_0 . The components $\widehat{\delta\theta}$ represent the change in θ required to move from Y_0 to $\hat{Y}$, that is

$$\hat{\theta} = \theta_0 + \widehat{\delta\theta} \quad (\text{C.83})$$

where θ_0 and $\hat{\theta}$ correspond to Y_0 and $\hat{Y}$, respectively. In the following sections we show that the ML tangent vector estimate is given by

$$\widehat{\delta\theta} = L_{\acute{s}\acute{s}}^{-1} L_{\acute{s}}|_{Y_0}. \quad (\text{C.84})$$

The main point of this section is to justify the notation, $L_s|_{Y_{MD}} \equiv 2Re(E_s^H \tilde{\Lambda}_{bb} b_n)$ (see eqn. (660)), used earlier.

Constrained ML Criterion and Solution in S-N Coordinates

Let the partitioned basis $\mathbf{s}, \mathbf{n}$ for $B_Y = S_Y \oplus N_Y$ have the numerical form $\mathbf{s}_k = Y_{\perp} S_k$ and $\mathbf{n}_k = Y_{\perp} N_k$. These basis elements are related to the basis $\mathbf{b}_k$ for B_Y by a

linear transformation

$$\mathbf{s}_k = \sum_{m=1}^q \mathbf{b}_k [T_s]_k^m; \quad k = 1 \cdots q \quad (\text{C.85})$$

$$\mathbf{n}_k = \sum_{m=1}^{N_n} \mathbf{b}_k [T_n]_k^m; \quad k = 1 \cdots N_b - q. \quad (\text{C.86})$$

This can be expressed in equivalent matrix form as (see (A.232))

$$\begin{bmatrix} E_s & E_n \end{bmatrix} = E_b T = E_b \begin{bmatrix} T_s & T_n \end{bmatrix} \quad (\text{C.87})$$

$$E_s = E_b T_s, \quad E_n = E_b T_n \quad (\text{C.88})$$

where E_s, E_n and E_b are defined in terms of the matrices S_k, N_k and B_k by (653),(654) and (654), respectively.

Given an arbitrary $\Delta \in B_Y$, represented as $\Delta = \sum_{k=1}^{N_b} \mathbf{b}_k \beta^k$, the components β transform to the $\begin{bmatrix} s & n \end{bmatrix}^t$ components in the S-N basis as (see (A.236))

$$\begin{bmatrix} s \\ n \end{bmatrix} = T^{-1} \beta = \begin{bmatrix} X_s^t \\ X_n^t \end{bmatrix} \beta \quad (\text{C.89})$$

(see eqn. (A.239)).

The coefficient matrix $L_{\beta\beta}$ transforms as a 2nd order covariant tensor and the partitioned form with respect to the S-N basis is defined by this transformation is (see (A.254))

$$\begin{bmatrix} L_{ss} & L_{sn} \\ L_{ns} & L_{nn} \end{bmatrix} \equiv \begin{bmatrix} T_s^t \\ T_n^t \end{bmatrix} L_{\beta\beta} \begin{bmatrix} T_s & T_n \end{bmatrix}. \quad (\text{C.90})$$

Combining these two results (C.89) and (C.90), the quadratic objective function in terms of the S-N basis is

$$\begin{bmatrix} s^t & n^t \end{bmatrix} \begin{bmatrix} L_{ss} & L_{sn} \\ L_{ns} & L_{nn} \end{bmatrix} \begin{bmatrix} s \\ n \end{bmatrix} = \beta^t L_{\beta\beta} \beta. \quad (\text{C.91})$$

Substituting the constraint-satisfying pair $(s, \hat{n})$ (647) into the quadratic objective function (637), and using (C.91), yields an unconstrained minimization problem in the free

p -tuple s given by

$$\hat{s} = \min_s \left(\begin{bmatrix} s^t & \hat{n}^t \end{bmatrix} \begin{bmatrix} L_{ss} & L_{sn} \\ L_{ns} & L_{nn} \end{bmatrix} \begin{bmatrix} s \\ \hat{n} \end{bmatrix} \right). \quad (\text{C.92})$$

This minimization problem has the well known solution

$$\hat{s} = L_{ss}^{-1} L_{sn} \hat{n}. \quad (\text{C.93})$$

For $L_{\beta\beta} = \frac{2K}{\tilde{\sigma}^2} \text{Re}(E_b^H \tilde{\Lambda}_{bb} E_b)$ (see (607)) the matrices L_{ss} and L_{sn} are

$$L_{ss} = T_s^t L_{\beta\beta} T_s = \frac{2K}{\tilde{\sigma}^2} \text{Re}(E_s^H \tilde{\Lambda}_{bb} E_s), \quad (\text{C.94})$$

$$L_{sn} = T_s^t L_{\beta\beta} T_n = \frac{2K}{\tilde{\sigma}^2} \text{Re}(E_s^H \tilde{\Lambda}_{bb} E_n). \quad (\text{C.95})$$

The product $L_{sn} \hat{n}$ in (C.93) is

$$L_{sn} \hat{n} = \frac{2K}{\tilde{\sigma}^2} \text{Re}(E_s^H \tilde{\Lambda}_{bb} E_n) \hat{n} = \frac{2K}{\tilde{\sigma}^2} \text{Re}(E_s^H \tilde{\Lambda}_{bb} b_n), \quad (\text{C.96})$$

where we have used $b_n \equiv E_n \hat{n}$.

We observe that (C.96) is equal to (660), that is $L_s|_{Y_{MD}} = L_{sn} \hat{n}$. (this relation is explained more fully in the next section). As a result we have

$$\hat{s} = L_{ss}^{-1} L_{sn} \hat{n} = L_{ss}^{-1} L_s|_{Y_{MD}}. \quad (\text{C.97})$$

Evaluating Derivatives and $L_{sn} \hat{n}$ and $\hat{c}$

The first derivatives of L in the neighborhood of the unconstrained MLE $\tilde{Y}$ may be expressed using a Taylor expansion. By definition all the first derivatives of L vanish at $\tilde{Y}$. Since the ML tangent vector $\hat{\Delta}$ at $\tilde{Y}$ is such that $P_A \hat{\Delta} = 0$ (i.e., $\alpha = 0$), we need only consider the expansion along tangent vectors in $B_{\tilde{Y}}$ (see eqn. (585)).

Let Y be an arbitrary point with normal coordinates (s, n) relative to the $\mathbf{s}, \mathbf{n}$ basis for $B_{\tilde{Y}}$. Then

$$\begin{bmatrix} L_s \\ L_n \end{bmatrix}_Y = \begin{bmatrix} 0 \\ 0 \end{bmatrix} + \begin{bmatrix} L_{ss} & L_{sn} \\ L_{ns} & L_{nn} \end{bmatrix}_{\tilde{Y}} \begin{bmatrix} s \\ n \end{bmatrix} + h.o.t.. \quad (\text{C.98})$$

In particular, the L_s component is

$$L_s|_Y = L_{ss}s + L_{sn}n . \quad (\text{C.99})$$

To simplify notation in this section define $c \equiv \frac{2K}{\bar{\sigma}^2}$. Substituting (A.254) and (A.256) for L_{ss} and L_{sn} , respectively, into the above yields

$$L_s|_Y = c \cdot \text{Re}(E_s^H \tilde{\Lambda}_{bb} E_s) s + c \cdot \text{Re}(E_s^H L_{bb} E_n) n \quad (\text{C.100})$$

$$= c \cdot \text{Re}(E_s^H \tilde{\Lambda}_{bb} (E_s s + E_n n)) . \quad (\text{C.101})$$

Substituting $b \equiv E_s s + E_n n$ (see eqn. (652)) into the above, the first derivative at Y in the neighborhood of the unconstrained MLE $\tilde{Y}$ is

$$L_s|_Y = c \cdot \text{Re}(E_s^H \tilde{\Lambda}_{bb} b) . \quad (\text{C.102})$$

Using this general expression we can compute the first derivatives at particular points of interest. The first derivative L_s at the minimum distance point $Y_{MD} \in \bar{\mathcal{B}}$ (see (650)) with coordinates $(0, 0, \hat{n})$, and at reference point $Y_0 \in \bar{\mathcal{B}}$, with coordinates $(0, s_0, \hat{n})$, are given respectively by

$$L_s|_{Y_{MD}} = L_{sn}\hat{n} = c \cdot \text{Re}(E_s^H \tilde{\Lambda}_{bb} b_n) \quad (\text{C.103})$$

and

$$L_s|_{Y_0} = L_{ss}s_0 + L_{sn}\hat{n} = c \cdot \text{Re}(E_s^H \tilde{\Lambda}_{bb} \tilde{b}) \quad (\text{C.104})$$

where

$$\tilde{b} \equiv b_n + E_s s_0 \quad (\text{C.105})$$

and where we have used the definition $b_n \equiv E_n \hat{n}$ (see eqn. (656)).

We can use this result to express the estimate (C.93) in a more familiar form. Substituting (C.103) in place of $L_{sn}\hat{n}$ in (C.93) yields

$$\hat{s}_{\tilde{Y}} = L_{ss}^{-1} L_s|_{Y_{MD}} . \quad (\text{C.106})$$

The estimate is thus the product of the inverse Hessian matrix, and the first derivative vector, which is recognized as the standard Newton's method form.

The desired ML estimate $\hat{Y}$ then is

$$\hat{Y} = \exp_{\tilde{Y}} \hat{\Delta}_{\tilde{Y}} \quad (\text{C.107})$$

where

$$\hat{\Delta}_{\tilde{Y}} = \sum_{k=1}^q \mathbf{s}_k \hat{s}^k + \sum_{k=1}^q \mathbf{n}_k \hat{n}^k . \quad (\text{C.108})$$

Tangent vector based at the reference point Y_0

An alternate and more practically useful form is found noting that, from (C.103) and (C.104),

$$L_s|_{Y_{MD}} = L_{ss}s_0 - L_s|_{Y_0} \quad (\text{C.109})$$

where the Hessian components (L_{ss} , etc.) are evaluated at $\tilde{Y}$. This formulation assumes that the Hessian is approximately constant near $\tilde{Y}$. Substituting (C.109) into (C.106) yields

$$\hat{s}_{\tilde{Y}} = L_{ss}^{-1}(L_{ss}s_0 - L_s|_{Y_0}) \quad (\text{C.110})$$

$$\hat{s}_{\tilde{Y}} = s_0 - \hat{s}_{Y_0} \quad (\text{C.111})$$

where we have used the notation

$$\hat{s}_{Y_0} = L_{ss}^{-1}L_s|_{Y_0} \quad (\text{C.112})$$

(see figure C.1). Denote the tangent vector from $\tilde{Y}$ to Y_0 as $\tilde{\Delta}_{\tilde{Y}}$, that is $\tilde{\Delta}_{\tilde{Y}} = \exp_{\tilde{Y}}^{-1}Y_0$. In the S-N basis this tangent vector has components $(s_0, \hat{n})$. If $\Delta_{\tilde{Y}}$ is arbitrary with components (s, n) then $(\Delta_{\tilde{Y}} - \tilde{\Delta}_{\tilde{Y}})$ has components $(s - s_0, n - \hat{n})$. In particular if $\Delta_{\tilde{Y}}$ has components $(\hat{s} - s_0, \hat{n})$ then the difference vector $(\Delta_{\tilde{Y}} - \tilde{\Delta}_{\tilde{Y}})$ has components $(\hat{s}, 0)$ (see figure C.1).

Define the parallel translated difference vector as Δ_{Y_0} , that is

$$\Delta_{Y_0} \equiv \tau(\Delta_{\tilde{Y}} - \tilde{\Delta}_{\tilde{Y}}) \approx \Delta_{\tilde{Y}} - \tilde{\Delta}_{\tilde{Y}} . \quad (\text{C.113})$$

Figure C.1. Newton's Method

The translated components are $(s - s_0, n - \hat{n})_{Y_0}$. Since the basis vectors for $\tilde{Y}$ were formally defined by paralleling translating the S_{Y_0} basis vectors defined at Y_0 to $\tilde{Y}$, the components of Δ_{Y_0} are $(s - s_0, n - \hat{n})_{Y_0}$. If $n = \hat{n}$ then $(\hat{s}, 0)$ or $(\hat{s}_{Y_0}, 0)$.

The ML estimate can be expressed relative to the reference Y_0 by parallel translating the ML tangent vector from $\tilde{Y}$ to Y_0 :

$$\hat{\Delta}_{Y_0} = \tau(\hat{\Delta}_{\tilde{Y}} - \tilde{\Delta}) \approx \hat{\Delta}_{\tilde{Y}} - \tilde{\Delta}, \quad (\text{C.114})$$

so that

$$\hat{\Delta}_{Y_0} = \tau\left(\sum_{k=1}^q \mathbf{s}_k (\hat{s} - s_0)^k\right) = \sum_{k=1}^q \tau(\mathbf{s}_k) (\hat{s} - s_0)^k. \quad (\text{C.115})$$

The component with respect to the basis at Y_0 is therefore $\hat{s}_{Y_0} \equiv \hat{s}_{\tilde{Y}} - s_0$. Substituting

(C.110) for $\hat{s}_{\hat{Y}}$ and noting (C.112) yields

$$\hat{s}_{Y_0} = L_{ss}^{-1} L_s|_{Y_0} . \quad (\text{C.116})$$

The corresponding ML tangent vector then is

$$\hat{\Delta}_{Y_0} = \sum_{k=1}^p \mathbf{s}_k \hat{s}_{Y_0}^k \quad (\text{C.117})$$

and the resultant estimate of $Y \in \bar{\mathcal{B}}$ is given by $\hat{Y} = \exp_{Y_0} \hat{\Delta}_{Y_0}$.

Parameter Estimation

Given the signal model $H(\theta)$, a natural coordinate basis $\acute{s}_k$ can be computed for S_{Y_0} in a straightforward way. As shown in section 3.9.2, a natural coordinate basis $\acute{s}_k$ for the tangent space S_{Y_0} can be expressed in terms of the derivative matrix D of the signal matrix H evaluated at $\hat{\theta}_0$. This, in turn, defines the matrix the $E_{\acute{s}}$ that appears in $L_{\acute{s}\acute{s}}$ and $L_{\acute{s}}$ (see eqn. (520)).

The natural coordinate basis $\acute{s}_k$ is related to the orthonormal basis $\mathbf{s}_k$ used in the earlier development by a real matrix $\mathbf{J}$. The corresponding basis matrices are then related as

$$E_{\acute{s}} = E_s \mathbf{J}^{-1} . \quad (\text{C.118})$$

The components of the ML tangent $\hat{\Delta}$ with respect to these two basis sets are related as

$$\widehat{\delta\theta} = \mathbf{J} \hat{s} \quad (\text{C.119})$$

(see eqn. (509)). The Hessian is found either directly or via the tensorial transformation

$$L_{\acute{s}\acute{s}} = \mathbf{J}^{-t} L_{ss} \mathbf{J}^{-1} = \frac{2K}{\tilde{\sigma}^2} \text{Re}(E_{\acute{s}}^H (\tilde{\Lambda}_{OPT} \otimes I_{N-p}) E_{\acute{s}}) . \quad (\text{C.120})$$

Likewise the first derivative is

$$L_{\acute{s}}|_{Y_0} = \mathbf{J}^{-t} L_s = \frac{2K}{\tilde{\sigma}^2} \text{Re}(E_{\acute{s}}^H (\tilde{\Lambda}_{OPT} \otimes I_{N-p}) \tilde{b}) . \quad (\text{C.121})$$

In this natural coordinate basis, the MLE tangent vector (C.122) expands as

$$\Delta_{Y_0} = \sum_{k=1}^p \hat{s}_k \widehat{\delta\theta}^k \quad (\text{C.122})$$

where

$$\widehat{\delta\theta} = L_{\hat{s}\hat{s}}^{-1} L_{\hat{s}}|_{Y_0} . \quad (\text{C.123})$$

The coefficient estimate $\widehat{\delta\theta}$ represents a correction to the reference value θ_0 and the improved estimate is

$$\hat{\theta} = \theta_0 + \widehat{\delta\theta} . \quad (\text{C.124})$$

Establishing a Reference point Y_0

The above discussion supposed a reference point $Y_0 \in \bar{\mathcal{B}}$ rotationally aligned with $\tilde{Y}$; that is $\Delta = \exp_{\tilde{Y}}^{-1} Y_0$ was such that $P_A \Delta = 0$. This rotationally-aligned reference is established as follows. Given a prior parameter estimate $\hat{\theta}_0$, the semi-unitary matrix $\check{Y}_0$ is computed as the left singular value matrix of $H(\hat{\theta}_0)$.

$$\check{Y}_0 = \text{svd}(H(\hat{\theta}_0)) . \quad (\text{C.125})$$

The point Y_0 is the rotationally aligned version of $\check{Y}_0$ and is given by

$$Y_0 = \check{Y}_0 Q \quad (\text{C.126})$$

where the $p \times p$ unitary matrix Q is given by

$$Q = U_1 V_1^H \quad (\text{C.127})$$

where the unitary matrices U_1 and V_1 are defined by

$$U_1 D_1 V_1^H = \text{svd}(\check{Y}_0^H \tilde{Y}) . \quad (\text{C.128})$$

By this construction,

$$\text{svd}(Y_0^H \tilde{Y}) = \text{svd}(Q^H \check{Y}_0^H \tilde{Y}) = Q^H U_1 D_1 V_1^H . \quad (\text{C.129})$$

The rotation matrix between Y_0 and $\tilde{Y}$ then is

$$Q_0 \equiv Q^H U_1 V_1^H = I \quad (\text{C.130})$$

as desired. As a consequence of the fact that $Q_0 = I$, we have $P_A \Delta = 0$ and, as a result, $\Delta \in B_{Y_0}$; there is no component in A_{Y_0} (see eqn. (252)).

If this reference point Y_0 is close to $\tilde{Y}$ (i.e. $\text{distSS}(\tilde{Y}, \hat{Y})$ small) then $D_1 \approx I$ and the matrix Q (C.127), (C.128) is well approximated by

$$Q \approx \check{Y}_0^H \tilde{Y} \quad (\text{C.131})$$

so that

$$Y_0 = \check{Y}_0 Q \approx P(\theta_0) \tilde{Y} \quad (\text{C.132})$$

where $P(\theta_0) \equiv \check{Y}_0 \check{Y}_0^H$. The transformation matrix T , defined by $Y_0 = H(\theta)T$ or $T = H^\# Y_0$, is then approximated by

$$T = H^\# \check{Y}_0 Q \approx H^\# \check{Y}_0 \check{Y}_0^H \tilde{Y} = H^\# P(\theta_0) \tilde{Y} = H^\# \tilde{Y} \quad (\text{C.133})$$

where we used the fact that $P(\theta_0)H^\# = H^\#$.

C.6.1 Estimation for M-Dimensional Modal Signal Model

When the signal matrix H is specified by a modal form (section 4.3), the above results take a simple form. As shown in section B.1.3, a natural coordinate basis $\acute{s}_k$ for the tangent space S_{Y_0} can be expressed in terms of the derivative matrix D of $H(\theta)$ at θ_0 . This basis defines the matrix the $E_{\acute{s}}$ that appears in definitions of $L_{\acute{s}\acute{s}}$ (C.120) and $L_{\acute{s}}$ (C.121). The $E_{\acute{s}}$ matrix is defined by (546) for the general multi-dimensional case (the 1-dimensional case is given by (520)).

Lemma 29 (Modal form of Hessian $L_{\acute{s}\acute{s}}$ in Natural Coordinates). *The subspace block of the Hessian of the log likelihood function with respect to the natural coordinate basis is*

$$L_{\acute{s}\acute{s}} = \frac{2K}{\sigma^2} \text{Re}((D^H P_\perp D) \odot \hat{P}_s^{(M)}) \quad (\text{C.134})$$

where

$$\hat{P}_s^{(M)} \equiv \tilde{P}_c \otimes (1_M \otimes 1_M^t) \quad (\text{C.135})$$

where $\tilde{P}_c = T\tilde{\Lambda}_{OPT}T^H$ and $T = H^\#Y_0$ (C.133).

Proof:

The following lemma records the form of $L_{\dot{s}}|_{Y_0}$ appearing in (C.123) for the general modal case.

Lemma 30. *If Y_0 is the rotationally aligned left singular vector matrix of $H(\theta)$, as defined by (C.126), then for $\tilde{\Delta} \in B_{\tilde{Y}}$ given by $\tilde{\Delta} = \exp_{\tilde{Y}}^{-1}Y_0$ the following holds*

$$\tilde{\Delta}\tilde{\Lambda}_{OPT}T^H = -P_\perp(\theta)\bar{R}_sH^{+H} \quad (\text{C.136})$$

where $T = H^\#Y_0$ and $\bar{R}_s \equiv \tilde{Y}\tilde{\Lambda}_{OPT}\tilde{Y}^H$

Proof:

Using the first-order approximation of $\tilde{\Delta}$ as $\tilde{\Delta} = \exp_{\tilde{Y}}^{-1}Y_0 \approx Y_0 - \tilde{Y}$ and using $Y_0 \approx P(\theta_0)\tilde{Y}$ (C.132) yields

$$\tilde{\Delta} = P(\theta_0)\tilde{Y} - \tilde{Y} = -P_\perp(\theta_0)\tilde{Y} . \quad (\text{C.137})$$

Using (C.133) and (C.137) we have

$$\tilde{\Delta}\tilde{\Lambda}_{OPT}T^H = -P_\perp(\theta)\tilde{Y}\tilde{\Lambda}_{OPT}\tilde{Y}^HH^{+H} . \quad (\text{C.138})$$

Using the definition $\bar{R}_s \equiv \tilde{Y}\tilde{\Lambda}_{OPT}\tilde{Y}^H$ gives the desired result.

Note that at high SNR $\tilde{R}_s \approx \bar{R}_s$ since $\tilde{\Lambda}_{OPT} \approx \tilde{\Lambda}$.

Lemma 31 (Modal form of first derivative $L_{\dot{s}}$). *The derivative of L in the arbitrary direction $\tilde{\Delta} \in B_{\tilde{Y}}$ or $\tilde{\Delta} \in B_{Y_0}$ is*

$$L_{\dot{s}}|_{Y_0} \equiv \nabla L \left[\tilde{\Delta} \right] = \frac{2K}{\sigma^2} G^t \text{vec}(\text{Re}(D^H \tilde{\Delta}\tilde{\Lambda}_{OPT}T^H)) \quad (\text{C.139})$$

where $G = \begin{bmatrix} G_1, & \dots, & G_M \end{bmatrix}$ is a selection matrix with submatrices G_m defined by (549). This is well approximated by

$$L_{\dot{s}}|_{Y_0} \approx -\frac{2K}{\sigma^2} G^t \text{vec}(\text{Re}(D^H P_{\perp}(\theta) \tilde{R}_s H^{+H})) \quad (\text{C.140})$$

where

$$\tilde{R}_s \equiv \tilde{Y} \tilde{\Lambda}_s \tilde{Y}^H. \quad (\text{C.141})$$

Proof: Represent the vector $\tilde{\Delta} \in B_{\tilde{Y}}$ as $\tilde{\Delta} = Y_{0\perp} \tilde{B}$. Substituting the modal expression for the $E_{\dot{s}}$ matrix, given by (551), into the general expression for $L_{\dot{s}}$, given by (C.121), yields

$$L_{\dot{s}}|_{Y_0} = c \cdot E_{\dot{s}}^H (\tilde{\Lambda}_{OPT} \otimes I_{N-p}) \tilde{b} = c \cdot G^t (T^t \otimes \tilde{Y}_{\perp}^H D)^H (\tilde{\Lambda}_{OPT} \otimes I_{N-p}) \tilde{b} \quad (\text{C.142})$$

where, to simplify notation, $c \equiv \frac{2K}{\sigma^2}$. Repeated application of the identity $\text{vec}(C^t \otimes A) \text{vec}(b) = \text{vec}(ABC)$ yields

$$= c \cdot G^t (T^* \otimes D^H \tilde{Y}_{\perp}) (\tilde{\Lambda}_{OPT} \otimes I_{N-p}) \tilde{b} \quad (\text{C.143})$$

$$= c \cdot G^t (T^* \otimes D^H \tilde{Y}_{\perp}) \text{vec}(\tilde{B} \tilde{\Lambda}_{OPT}) \quad (\text{C.144})$$

$$= c \cdot G^t \text{vec}(D^H \tilde{Y}_{\perp} \tilde{B} \tilde{\Lambda}_{OPT} T^H). \quad (\text{C.145})$$

Making the identification $\tilde{\Delta} = \tilde{Y}_{\perp} \tilde{B}$ yields the first desired result (C.139) Using (C.136) into the above yields the second desired result (C.141).

Direct computation of $L_{\dot{s}}$

Previously we computed the derivative $L_{\dot{s}}$ at Y_0 using an indirect Taylor series approach based on the unconstrained MLE $\tilde{Y}$. In this section we compute the derivative directly in the standard way and show that the two approaches yield the same result. The directional derivative of interest is

$$[L_{\dot{s}}]_k \equiv \frac{\partial L}{\partial \theta^k} \equiv \nabla L[\dot{s}_k]. \quad (\text{C.146})$$

In lemma 20 we showed the first derivative of the likelihood function in the direction of the arbitrary tangent vector $\Delta = Y_{0\perp} B_k$, $\Delta \in B_{Y_0}$ is (590)

$$\nabla L[\Delta] = c \cdot \text{tr}(\text{Re}(\Lambda \Lambda_s^{-1} Y_0^H \tilde{R} Y_{0\perp} B_k)) . \quad (\text{C.147})$$

Let $\Delta_{(k)}$ denote the tangent to the coordinate curve $\theta_{(k)}(t)$ to $t = 0$. This tangent in general has components in both A_Y and S_Y . Denote the projection of $\Delta_{(k)}$ onto $S_Y < B_Y$ as the vector $\acute{s}_k$, that is $\acute{s}_k = P_B \Delta_{(k)}$ and $\acute{s}_k \in S_Y < B_Y$.

The vector $\acute{s}_l$ has the general form $\acute{s}_l = Y_{0\perp} \acute{S}_l$, and for the modal case

$$\acute{S}_l = Y_{\perp}^H D J_{mkk} \cdot T \quad (\text{C.148})$$

where $l = f(m, k)$. Substituting this expression for $\acute{S}_l$ in place of B in (C.147) yields

$$[L_{\acute{s}}]_l = c \cdot \text{tr}(\text{Re}(\Lambda \Lambda_s^{-1} Y_0^H \tilde{R} Y_{\perp} Y_{\perp}^H D J_{mkk} \cdot T)) \quad (\text{C.149})$$

or equivalently

$$[L_{\acute{s}}]_l = c \cdot \text{tr}(\text{Re}(\Lambda \Lambda_s^{-1} Y_0^H \tilde{R} P_{\perp} D J_{mkk} \cdot H^{\#} Y_0)) \quad (\text{C.150})$$

where we have used $P_{\perp} = Y_{\perp} Y_{\perp}^H$ and $T = H^{\#} Y_0$. Rearranging terms within the trace operation yields

$$= c \cdot \text{tr}(H^{\#} Y_0 \Lambda \Lambda_s^{-1} Y_0^H \tilde{R} P_{\perp} D J_{kk}) . \quad (\text{C.151})$$

Using the fact that $\Lambda \Lambda_s^{-1} = (\Lambda_s - \sigma^2 I) \Lambda_s^{-1} = (I - \sigma^2 \Lambda_s^{-1})$ yields

$$= c \cdot \text{tr}(H^{\#} Y_0 Y_0^H \tilde{R} P_{\perp} D J_{kk}) - \text{tr}(H^{\#} Y_0 \sigma^2 \Lambda_s^{-1} Y_0^H \tilde{R} P_{\perp} D J_{kk}) . \quad (\text{C.152})$$

At high SNR $\sigma^2 \Lambda_s^{-1}$ appearing in the second term above is very small yielding the approximation

$$= c \cdot \text{tr}(H^{\#} P \tilde{R} P_{\perp} D J_{mkk}) \quad (\text{C.153})$$

where we have used the fact that $H^{\#} P(\theta) = H^{\#}$.

Using $\text{tr}(A^H B) = b^t \text{vec}(A)$ and identifying $A^H = H^\# P \tilde{R} P_\perp D$ and $B = J_{kk}$, $b = \text{vec}(B) = j_{kk}$ yields

$$[L_{\dot{s}}]_l = -c \cdot j_{mkk}^t \text{vec}(D^H P_\perp \tilde{R} P(\theta) H^{+H}(\theta)); k = 1 \cdots Mp \quad (\text{C.154})$$

where we have used $A = D^H P_\perp \tilde{R} H^\# H$ and $P(\theta) H^{+H}(\theta) = H^{+H}(\theta)$.

In $q \times 1$ vector form we have

$$L_{\dot{s}} = -c \cdot G^t \text{vec}(\text{Re}(D^H P_\perp(\theta) \tilde{R} H^{+H}(\theta))) \quad (\text{C.155})$$

where G is a selection matrix (549). This result agrees with (C.140). The first derivative

$L_\theta = \begin{bmatrix} \frac{\partial L}{\partial \theta^1} & \cdots & \frac{\partial L}{\partial \theta^q} \end{bmatrix}$ is the $q \times 1$ vector, $q = rMp$.

C.6.2 Geometric Newton's Method Refinement Algorithm Summary

Let the eigen-decomposition of the SCM be $\tilde{R} = \tilde{Y} \tilde{\Lambda} \tilde{Y}^H + \tilde{Y}_\perp \tilde{\Lambda}_\perp \tilde{Y}_\perp^H$. Assume that an initial parameter estimate θ_0 is given. The refinement $\widehat{\delta\theta}$ is computed as follows.

1. Compute $D, \check{Y}_0$ at θ_0

$$\check{Y}_0 = \text{svd}(H(\theta_0)) \quad (\text{C.156})$$

$$D = \begin{bmatrix} d(\theta_0^1) & \cdots & d(\theta_0^p) \end{bmatrix} \quad (\text{C.157})$$

2. Compute $\hat{\sigma}^2$ and $\tilde{\Lambda}_{OPT}$

$$\hat{\sigma}^2 = \frac{1}{N-p} \text{tr}(\tilde{\Lambda}_\perp) \quad (\text{C.158})$$

$$\tilde{\Lambda}_{OPT} = \tilde{\Lambda}^2 (\tilde{\Lambda} - \hat{\sigma}^2 I)^{-1} \quad (\text{C.159})$$

3. Compute the rotationally aligned reference point Y_0 and the transformation matrix T

$$U_1 D_1 V_1^H = \text{svd}(\check{Y}_0^H \tilde{Y}) \quad (\text{C.160})$$

then

$$Q = U_1 V_1^H \quad (\text{C.161})$$

and

$$Y_0 \equiv \check{Y}_0 Q \quad (\text{C.162})$$

$$T = H^\#(\theta_0) Y_0 \quad (\text{C.163})$$

3. Compute Tangent Vector $\tilde{\Delta} = \exp_{\tilde{Y}}^{-1} Y_0$

$\tilde{\Delta}$ is the tangent vector based at $\tilde{Y}$ with tail at the rotated reference solution Y_0 and is given by $\tilde{\Delta} = \exp_{\tilde{Y}}^{-1} Y_0$. Compute the first-order approximation of this vector as

$$\tilde{\Delta} \approx Y_0 - \tilde{Y} . \quad (\text{C.164})$$

4. Compute the Weighted Data vector

$$J_{mkk} = i_m \otimes (e_k \otimes e_k), \quad m = 1 \dots, M \quad k = 1 \dots, p \quad (\text{C.165})$$

$$G = \begin{bmatrix} \text{vec}(J_{1kk}) & \dots & \text{vec}(J_{Mpp}) \end{bmatrix} \quad (\text{C.166})$$

$$w = G^t \text{vec}(D^H \tilde{\Delta} \tilde{\Lambda}_{OPT} T^H) \quad (\text{C.167})$$

$$w = G^t \text{vec}(D^H \tilde{\Delta} T^{-1} T \tilde{\Lambda}_{OPT} T^H) \quad (\text{C.168})$$

$$w = G^t \text{vec}(D^H \tilde{\Delta} T^{-1} \tilde{P}_c) \quad (\text{C.169})$$

$$w = G^t \text{vec}(D^H \tilde{\Delta} T^{-1} \tilde{P}_c) \quad (\text{C.170})$$

5. Compute correction and Update parameter Estimate

$$P_s^{(M)} \equiv (T \tilde{\Lambda}_{OPT} T^H) \otimes (1_M \otimes 1_M^t) \quad (\text{C.171})$$

$$\widehat{\delta\theta} = \frac{\hat{\sigma}^2}{K} \text{Re}((D^H P_\perp D) \odot \hat{P}_s^{(M)})^{-1} \cdot w \quad (\text{C.172})$$

$$\hat{\theta} = \theta_0 + \widehat{\delta\theta} \quad (\text{C.173})$$

Remark: The ML subspace estimate can be computed from the estimate $\widehat{\delta\theta}$ using the transformation **J** as follows:

$$\hat{\Delta} = \sum_{k=1}^q \hat{s}_k \widehat{\delta\theta}^k \quad (\text{C.174})$$

$$\hat{Y} = \exp_{Y_0} \hat{\Delta} \quad (\text{C.175})$$

where $\tilde{\Delta} = \exp_{\tilde{Y}}^{-1} Y_0$ is the tangent vector from $\tilde{Y}$ to Y_0 , $D = D(\theta_0)$ is the derivative matrix, $T_s \equiv H^\#(\theta_0) Y_0$ and

6. Update Subspace Estimate(Optional)

$$\hat{\Delta} = Y_{0\perp} \text{mat}(E_s \hat{\delta\theta}) \quad (\text{C.176})$$

$$Y = \exp_{Y_0} \hat{\Delta} \quad (\text{C.177})$$

Example: 1-Dimensional Array, Harmonic Signals

For the specific example case of $p = 2$, two source signals, the $p^2 \times p$ matrix G is

$$G = \begin{bmatrix} \text{vec}(J_{11}) & \text{vec}(J_{22}) \end{bmatrix} \quad (\text{C.178})$$

$$G = \begin{bmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 1 \end{bmatrix} \quad (\text{C.179})$$

$$L_s = c \cdot (G^t) \text{vec}(D^H \tilde{\Delta} \tilde{\Lambda}_{OPT} T^H) \quad (\text{C.180})$$

The two components of L_s are

$$[L_s]_1 = c \cdot [\text{vec}(D^H \tilde{\Delta} \tilde{\Lambda}_{OPT} T^H)]_1 \quad (\text{C.181})$$

$$[L_s]_2 = c \cdot [\text{vec}(D^H \tilde{\Delta} \tilde{\Lambda}_{OPT} T^H)]_4 \quad (\text{C.182})$$

The corresponding parameter refinement is computed using (C.84) as

$$\hat{\delta\theta}(1) = [L_{ss}^{-1}]_{11} [L_s]_1 + [L_{ss}^{-1}]_{12} [L_s]_2, \quad (\text{C.183})$$

$$\hat{\delta\theta}(2) = [L_{ss}^{-1}]_{21} [L_s]_1 + [L_{ss}^{-1}]_{22} [L_s]_2. \quad (\text{C.184})$$

For the 1-dimensional case $M = 1$, $p = 2$, the Hessian is the 2×2 given by

$$L_{ss} = c \cdot \text{Re}((D^H P_\perp D) \odot P_s) \quad (\text{C.185})$$

and $P_s = T \tilde{\Lambda}_{OPT} T^H$.

Example: 2-Dimensional Array, Harmonic Signals

For the 2-dimensional array Harmonic signal case there are 2 parameters associated with each mode so that $M = 2$. For the specific example case of $p = 2$, two source signals, the $Mp^2 \times p^2$ matrix G is 8×4 and is given by (562). Using this in (C.167) gives the four components of w as

$$[L_s]_1 = [vec(D^H \tilde{\Delta} \tilde{\Lambda}_{OPT} T^H)]_1, \quad (C.186)$$

$$[L_s]_2 = [vec(D^H \tilde{\Delta} \tilde{\Lambda}_{OPT} T^H)]_6, \quad (C.187)$$

$$[L_s]_3 = [vec(D^H \tilde{\Delta} \tilde{\Lambda}_{OPT} T^H)]_3, \quad (C.188)$$

$$[L_s]_4 = [vec(D^H \tilde{\Delta} \tilde{\Lambda}_{OPT} T^H)]_8. \quad (C.189)$$

For this $M = 2$ case the factor $P_s^{(M)}$ (C.135) is

$$P_s^{(2)} \equiv (T \tilde{\Lambda}_{OPT} T^H) \otimes (1_2 \otimes 1_2^t) \equiv \begin{bmatrix} (T \tilde{\Lambda}_{OPT} T^H) & (T \tilde{\Lambda}_{OPT} T^H) \\ (T \tilde{\Lambda}_{OPT} T^H) & (T \tilde{\Lambda}_{OPT} T^H) \end{bmatrix}. \quad (C.190)$$

BIBLIOGRAPHY

- “Differential geometry of surfaces.” [Online]. Available: https://en.wikipedia.org/wiki/Weingarten_map
- Boumal, A., “On intrinsic cramer-rao bounds for riemannian submanifolds and quotient manifolds,” *IEEE Transactions on Signal Processin*, vol. 61, no. 7, April 2013.
- C.F.Gauss, *General Investigations of Curved Surfaces*. New York: Raven Press, 1965.
- Edelman, A., Arias, T. A., and Smith, S., “The geometry of algorithms with orthogonality constraints,” *SIAM Matrix Anal. Appl.*, vol. 20, no. 2, p. 303353, 1998.
- Haardt, M. and Galdo, G. D., “Higher-order svd based subspace estimation to improve the parameter estimation accuracy in multi-dimensional harmonic retrieval problems,” *IEEE Trans. Signal Processing*, vol. 56, p. 3198 3213, July 2008.
- Haardt, M., Thoma, R. S., and Richter, A., *Multi-dimensional highresolution parameter estimation with applications to channel sounding*. New York, NY: Marcel Dekker, June 2004, ch. 5, p. 255 338.
- Jeffries, D. J. and Farrier, D. R., “Asymptotic results for eigenvector methods,” *Proc. Inst. Elec. Eng.*, vol. 132, no. 7, pp. 589 –594, June 1985.
- Krim, H. and Viberg, M., “Two decades of array signal processing research,” *IEEE Signal Processing Magazine*, pp. 67–94, July 1996.
- Maurice, R., “Improved structured least squares for the application of unitary esprit to cross arrays,” *IEEE Signal Processing Letters*, vol. 13, no. 6, pp. 349–352, 2006.
- Rao, B. D. and Hari, K. V. S., “Performance analysis of esprit and tam in determining the direction of arrival of plane waves in noise,” *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 37, pp. 1990 –1995, December 1989.
- Roemer, F. and Haardt, M., “Analytical performance assessment of 1-d structured least squares,” in *ICASSP Proceedings*, vol. 56, July 2011, p. 3198 3213.
- Roemer, F. and Haardt, M., “Performance analysis of esprit-type algorithms for strictly non-circular sources using structured least squares,” in *Computational Advances in Multi-Sensor Adaptive Processing*, vol. 56, 2013.
- Scharf, L. L., *Statistical Signal Processing: Detection, Estimation, and Time Series Analysis*. Reading, MA: Addison-Wesley, 1991.
- Scharf, L. L. and McWhorter, T., “Geometry of the cramer-rao bound,” *Signal Process*, vol. 31, p. 301 311, 1993.

- Smith, S., "Covariance, subspace, and intrinsic cramér-rao bounds," *IEEE Trans. Signal Process.*, vol. 53, no. 5, p. 1610 1630, 2005.
- Stoica, P., Larson, E., and Gershman, A., "The stochastic crb for array processing: A textbook derivation," *IEEE Transactions on Signal Processing Letters*, vol. 8, no. 5, pp. 148–150, May 2001.
- Stoica, P. and Nehorai, A., "Music, maximum likelihood, and cramér-rao bound," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, no. 5, p. 720 741, May 1989.
- Vaccaro, R. and Ding, Y., "A new state space approach for direction finding," *IEEE Transactions on Signal Processing*, pp. 3234–3237, November 1994.
- Xu, Z., "Perturbation analysis for subspace decomposition with applications in subspace-based algorithms," *IEEE Transactions on Signal Processing*, vol. 50, no. 11, p. 2820 2830, Nov 2002.