V. Concluding Comments

We have derived the basic statistical properties for the estimated rotary coefficient. These depend on the true value of the rotary coefficient, and the conjugate coherence γ_*^2, a nuisance parameter. Fortunately when the latter is estimated and debiased constructed confidence intervals maintain appropriate coverage probabilities, so such confidence intervals have practical utility as illustrated by the Labrador Sea current data analysis.

Acknowledgment

The authors would like to thank J. Lilly for making the Labrador Sea data available to them. Helpful comments and observations by the reviewers were much appreciated.

References

[1] H. H. Andersen, M. Højbjerre, D. Sørensen, and P. S. Eriksen, *Linear and Graphical Models*. New York: Springer-Verlag, 1995.
[2] C. E. Barton, "Analysis of palaeomagnetic time series—Techniques and applications," *Surveys Geophys.*, vol. 5, pp. 335–368, 1983.
[3] V. A. Benignus, "Estimation of the coherence spectrum and its confidence interval using the fast Fourier transform," *IEEE Trans. Audio Electroacoust.*, vol. 17, no. 2, pp. 145–150, 1969.
[4] D. R. Brillinger, *Time Series: Data Analysis and Theory*, Expanded ed. New York: McGraw-Hill, 1981.
[5] E. M. Carter, C. G. Khatri, and M. S. Srivastava, "Nonnull distribution of likelihood ratio criterion for reality of covariance matrix," *J. Multivariate Anal.*, vol. 6, pp. 176–184, 1976.
[6] W. J. Emery and R. E. Thomson, *Data Analysis Methods in Physical Oceanography*. New York: Pergamon, 1998.
[7] J. Gonella, "A rotary component method for analysing meteorological and oceanographic vector time series," *Deep-Sea Res.*, vol. 19, pp. 833–846, 1972.
[8] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products*, Corrected and Enlarged ed. New York: Academic, 1980.
[9] Y. Hayashi, "Space-time spectral analysis of rotary vector series," *J. Atmos. Sci.*, vol. 36, pp. 757–766, 1979.
[10] J. M. Lilly, P. B. Rhines, M. Visbeck, R. Davis, J. R. Lazier, F. Schott, and D. Farmer, "Observing deep convection in the Labrador Sea during winter 1994/95," *J. Phys. Oceanogr.*, vol. 29, pp. 2065–2098, 1999.
[11] J. M. Lilly and P. B. Rhines, "Coherent eddies in the Labrador Sea observed from a mooring," *J. Phys. Oceanogr.*, vol. 32, pp. 585–598, 2002.
[12] T. Maitani, "Statistics of wind direction fluctuations in the surface layer over plant canopies," *Boundary-Layer Meteorol.*, vol. 26, pp. 15–24, 1983.
[13] T. Medkour and A. T. Walden, "Attenuation estimation from correlated sequences," *IEEE Trans. Signal Process.*, vol. 55, no. 1, pp. 378–383, Jan. 2007.
[14] T. Medkour and A. T. Walden, "A variance equality test for two correlated complex Gaussian variables with application to spectral power comparison," *IEEE Trans. Signal Process.*, vol. 55, no. 3, pp. 881–888, Mar. 2007.
[15] T. Medkour and A. T. Walden, "Statistical properties of the estimated degree of polarization," *IEEE Trans. Signal Process.*, vol. 56, no. 1, pp. 408–414, Jan. 2008.
[16] K. S. Miller, *Hypothesis Testing With Complex Distributions*. New York: Krieger, 1980.
[17] C. N. K. Mooers, "A technique for the cross spectrum analysis of pairs of complex-valued time series, with emphasis on properties of polarized components and rotational invariants," *Deep-Sea Res.*, vol. 20, pp. 1129–1141, 1973.
[18] I. Olkin and J. W. Pratt, "Unbiased estimation of certain correlation coefficients," *Ann. Math. Stat.*, vol. 21, pp. 201–211, 1958.
[19] M. Orlić, B. Penzar, and I. Penzar, "Adriatic sea and land breezes: Clockwise versus anticlockwise rotation," *J. Appl. Meteorol.*, vol. 27, pp. 675–679, 1988.
[20] D. B. Percival and A. T. Walden, *Spectral Analysis for Physical Applications*. Cambridge, U.K.: Cambridge Univ. Press, 1993.
[21] M. S. S. Sarma and L. V. G. Rao, "Spectra of currents and temperature off Godavari (East Coast of India)," *Mahasagar*, vol. 22, pp. 29–36, 1989.
[22] B. Picinbono and P. Bondon, "Second-order statistics of complex signals," *IEEE Trans. Signal Process.*, vol. 45, no. 2, pp. 411–420, 1997.
[23] P. J. Schreier, "Polarization ellipse analysis of nonstationary random signals," *IEEE Trans. Signal Process.*, vol. 56, no. 9, pp. 4330–4339, Sep. 2008.
[24] P. J. Schreier and L. L. Scharf, "Second-order analysis of improper complex random vectors and processes," *IEEE Trans. Signal Process.*, vol. 51, no. 3, pp. 714–725, Mar. 2003.
[25] P. J. Schreier and L. L. Scharf, *Statistical Signal Processing of Complex-Valued Data*. Cambridge, U.K.: Cambridge Univ. Press, 2010.
[26] H. van Haren and C. Millot, "Rectilinear and circular inertial motions in the Western Mediterranean Sea," *Deep-Sea Res. Part I*, vol. 51, pp. 1441–1455, 2004.
[27] A. T. Walden, E. J. McCoy, and D. B. Percival, "The variance of multitaper spectrum estimates for real Gaussian processes," *IEEE Trans. Signal Process.*, vol. 42, no. 2, pp. 479–482, 1994.
[28] A. T. Walden, E. J. McCoy, and D. B. Percival, "The effective bandwidth of a multitaper spectral estimator," *Biometrika*, vol. 82, pp. 201–214, 1995.
[29] L. Zhiliang, H. Dunxin, T. Xiaohui, and W. Enbo, "Rotary spectrum analysis of tidal current in the southern Yellow Sea," *Chinese J. Oceanol. Limnol.*, vol. 25, pp. 286–291, 2007.

An Explicit Expression for the Newton Direction on the Complex Grassmann Manifold

Ramy H. Gohary and Timothy N. Davidson

***Abstract*—Several important design problems in signal processing for communications can be cast as optimization problems in which the objective is a function of the subspaces spanned by tall complex matrix variables with orthonormal columns. Such problems can be viewed as optimization problems on the complex Grassmann manifold, and an effective means for performing this optimization is to use a Grassmannian version of Newton's method. To facilitate the implementation of that method, we provide an explicit expression for the Grassmannian Newton direction for an arbitrary twice differentiable function. We also use an example in which the pairwise chordal Frobenius norm between subspaces is to be optimized to outline a systematic procedure for obtaining the Hessian matrix.**

***Index Terms*—Levi–Civita connection, Newton's method, optimization on manifolds, orthogonality constraints, principal angles, Wirtinger derivatives.**

I. Introduction

A number of important engineering design problems can be cast in the form of optimizing a twice differentiable real scalar objective $F(Y)$ over the set of $m \times n$ matrices with $m \geq 2n$ that satisfy $Y^\dagger Y = I_n$, where I_n is the $n \times n$ identity matrix and $Y^\dagger$ denotes the transpose or Hermitian transpose of Y, as appropriate. When $F(Y) = F(YQ)$ for all $n \times n$ matrices satisfying $QQ^\dagger = I_n$, the function $F(Y)$ is said to be homogeneous. That is, it depends on the subspace spanned by the columns of Y, but not on the particular basis that spans this subspace. The set of n-dimensional subspaces in $\mathbb{R}^m$ or $\mathbb{C}^m$ is typically

Manuscript received May 21, 2010; revised September 15, 2010; accepted November 10, 2010. Date of publication November 22, 2010; date of current version February 09, 2011. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Soontorn Oraintara.

R. H. Gohary is with the Department of Systems and Computer Engineering, Carleton University, Ottawa, ON K1S 5B6, Canada (e-mail: gohary@sce.carleton.ca).

T. N. Davidson is with the Department of Electrical and Computer Engineering, McMaster University, Hamilton, ON L8S 4K1, Canada.

Digital Object Identifier 10.1109/TSP.2010.2094615

referred to as the Grassmann manifold, and is denoted by $\mathbb{G}_n(\mathbb{R}^m)$ or $\mathbb{G}_n(\mathbb{C}^m)$, [1]. A homogeneous function $F(Y)$ can be regarded as a mapping from the Grassmannian manifold to $\mathbb{R}$.

Design problems that can be cast as optimization over the complex Grassmann manifold include that of designing constellations for noncoherent multiple antenna wireless communication systems [2]–[6], and that of designing quantization codebooks for multiple antenna systems with limited feedback [7]–[9]. A common theme in these applications is that the design problem can be cast as one in which the variable is a block diagonal matrix and the objective is a function of the angles between the subspaces spanned by its blocks. When that function is chosen to be the projection Frobenius norm [2], [10], optimal packings can be obtained for specific dimensions by exploiting the projection structure of the embedding space [11]. However, it was argued in [6] that a more appropriate metric for designing constellations is the chordal Frobenius norm (cf. [1]), which follows from embedding the Grassmann manifold in Euclidean space [1].

In conventional optimization problems defined over Euclidean space, Newton's method exhibits faster local convergence than other standard optimization techniques, including the conjugate gradient method [12]. In [1], these standard techniques were adapted to problems defined over the Grassmannian manifold, and it was shown that the superior local convergence of Newton's method is retained. (A somewhat different approach to adapting these techniques to the Grassmann manifold was taken in [13].) In Euclidean space, the Newton direction is obtained by solving a single set of linear equations that involves the Hessian matrix and the gradient. On the Grassmann manifold, Newton's method also involves the evaluation of the Hessian and the gradient, but obtaining the Newton direction requires the solution of a sequence of linear matrix equations with a special structure [1]. Despite the numerical illustration of the desirable properties of the Grassmannian Newton method in [1], the structure of these equations has been an impediment to the application of the method to other engineering design problems. Another impediment arises when the design problem is defined over the complex Grassmann manifold rather than the real one. Although the approach in [1] is conceptually extensible from the real case to the complex case, the expressions for the complex case cannot be obtained by simple analogy with the real case, and must be derived separately.

The goal of this paper is to facilitate the application of the methodology developed in [1] to design problems on the complex Grassmann manifold. We do so in three ways. First, we use the notion of Wirtinger derivatives [14] to provide an explicit extension of the framework in [1] to the complex case. Second, we derive a closed-form expression for the Newton direction for an arbitrary twice-differentiable function $F(Y)$ in terms of the gradient F_Y and the second derivatives F_{YY} and F_{YY^*}. Third, we outline a procedure for obtaining the gradient and the Hessian. In particular, we consider a case in which the optimization variable is a block diagonal matrix, $\bar{Y}$, and the objective $F(\bar{Y})$ involves the chordal Frobenius norm between the subspaces spanned by the diagonal blocks of $\bar{Y}$ as well as the subspaces spanned by their orthogonal complements.

Notation: We will use $(\,\cdot\,)^T$, $(\,\cdot\,)^\dagger$ and $(\,\cdot\,)^*$ to denote the transpose, Hermitian transpose and conjugate, respectively, and $\mathrm{Tr}(\,\cdot\,)$ to denote the trace. The operator $\mathrm{vec}(\,\cdot\,)$ will be used to stack the columns of the matrix argument in one vector, and $\Re\{\cdot\}$ and $\Im\{\cdot\}$ will be used to extract the real and imaginary components, respectively. The inner product on $\mathbb{C}^{m\times n}$ is $\langle A,B\rangle = \Re\{\mathrm{Tr}(A^\dagger B)\}$, and this is also the canonical metric on $\mathbb{G}_n(\mathbb{C}^m)$; cf. [1].

II. The Newton Direction on $\mathbb{G}_n(\mathbb{C}^m)$

In [1], Newton's method was adapted to functions $F(Y)$ on the real Grassmann manifold, $\mathbb{G}_n(\mathbb{R}^m)$. In this section, we will use the Wirtinger convention for differentiation with respect to a complex variable [14], to provide an explicit extension of that approach to $\mathbb{G}_n(\mathbb{C}^m)$. (The Wirtinger convention is briefly reviewed in Appendix A.) Given the current iterate, Y, Newton's method on $\mathbb{G}_n(\mathbb{C}^m)$ involves the construction of the Newton direction, Δ_{N}, which lies in the tangent space of $\mathbb{G}_n(\mathbb{C}^m)$ at Y, $\mathbf{T}_Y\mathbb{G}_n(\mathbb{C}^m) = \{\Delta \in \mathbb{C}^{m\times n} \mid \Delta^\dagger Y = 0\}$, [1]. The Newton update of Y is then obtained by taking a step of length t along the geodesic specified by Δ_{N}. For an arbitrary tangent direction Δ with compact singular-value decomposition (SVD) $\Delta = U\Sigma V^\dagger$, a step of length t from Y along the corresponding geodesic can be written as $Y^+(t) = YV\cos(\Sigma t)V^\dagger + U\sin(\Sigma t)V^\dagger$ [1].

Following the approach in [1], the characterization of the Newton direction begins with the notion of the Grassmannian gradient at Y, which we will denote by G. To define G, we first define the real scalar $\mathrm{Grad}\ F(\Delta) = \frac{d}{dt}F(Y^+(t))|_{t=0}$. The Grassmannian gradient at Y is the direction $G \in \mathbf{T}_Y\mathbb{G}_n(\mathbb{C}^m)$ such that $\langle G,X\rangle = \mathrm{Grad}\ F(X)$ for all $X \in \mathbf{T}_Y\mathbb{G}_n(\mathbb{C}^m)$. By explicit evaluation of these terms, it can be shown that the Grassmannian gradient can be written as $G = \mathcal{P}_Y F_Y$, where $F_Y \in \mathbb{C}^{m\times n}$ is the matrix with elements $[F_Y]_{ij} = \frac{\partial F}{\partial Y ij}$, and $\mathcal{P}_Y = (I_m - YY^\dagger)$ is the projector to the tangent space at Y.

Now, let us define the real scalar $\mathrm{Hess}\ F(\Delta,\Delta) = \frac{d^2}{dt^2}F(Y^+(t))|_{t=0}$. Applying the chain rule, we have

$$\begin{aligned}\mathrm{Hess}\ F(\Delta,\Delta) &= \sum_{ij}\frac{d}{dt}\left(\frac{\partial F}{\partial\Re\{Y_{ij}\}}\right)\bigg|_{t=0}\Re\{\Delta_{ij}\} + \Re\{[F_Y]_{ij}\}\frac{d^2}{dt^2}\Re\{Y_{ij}\}\bigg|_{t=0}\\ &\quad + \frac{d}{dt}\left(\frac{\partial F}{\partial\Im\{Y_{ij}\}}\right)\bigg|_{t=0}\Im\{\Delta_{ij}\} - \Im\{[F_Y]_{ij}\}\frac{d^2}{dt^2}\Im\{Y_{ij}\}\bigg|_{t=0}.\end{aligned} \tag{1}$$

By evaluating the derivatives it can be shown that $\mathrm{Hess}\ F(\Delta,\Delta)$ is quadratic in Δ, and hence can be polarized (cf. [1, p. 312]) to obtain $\mathrm{Hess}\ F(\Delta,X)$. By arranging the components of $\mathrm{Hess}\ F(\Delta,X)$ in a convenient matrix-based form, we can write $\mathrm{Hess}\ F(\Delta,X) = F_{YY}(\Delta,X) - \Re\{\mathrm{Tr}(F_Y^\dagger Y\Delta^\dagger X)\}$, where

$$F_{YY}(\Delta,X) = \Re\{\mathrm{Tr}(X^\dagger K^*(\Delta))\}, \tag{2}$$

and $[K(\Delta)]_{ij} = \frac{\partial}{\partial Y_{ij}}\Re\{\mathrm{Tr}(F_Y^T\Delta)\}$. The Newton direction at Y is then the direction $\Delta_{\mathrm{N}} \in \mathbf{T}_Y\mathbb{G}_n(\mathbb{C}^m)$ such that $\mathrm{Hess}\ F(\Delta_{\mathrm{N}},X) = -\langle G,X\rangle$ for all $X \in \mathbf{T}_Y\mathbb{G}_n(\mathbb{C}^m)$. Following [1], we now simplify that condition. For a given Δ, let us define $F_{YY}(\Delta)$ to be the unique element of $\mathbf{T}_Y\mathbb{G}_n(\mathbb{C}^m)$ such that

$$F_{YY}(\Delta,X) = \Re\{\mathrm{Tr}(X^\dagger F_{YY}(\Delta))\}\quad \text{for all } X \in \mathbf{T}_Y\mathbb{G}_n(\mathbb{C}^m). \tag{3}$$

Then, the Newton direction can be characterized as the solution of

$$F_{YY}(\Delta_{\mathrm{N}}) - \Delta_{\mathrm{N}}(Y^\dagger F_Y) = -G. \tag{4}$$

In the following section, we derive a closed-form expression for Δ_{N}.

III. An Explicit Expression for the Newton Direction

In [1], the Newton direction, Δ_{N}, for a specific objective on $\mathbb{G}_n(\mathbb{R}^m)$ is obtained by inspection. In this section, we derive a closed-form expression for Δ_{N} for an arbitrary twice differentiable objective on $\mathbb{G}_n(\mathbb{C}^m)$. To do so, we first obtain a more convenient expression for the term $K(\Delta) \in \mathbb{C}^{m\times n}$ in (2). We then use that expression to provide an explicit solution to (3). Using that solution, we obtain a closed-form expression for the set of linear equations that specifies Δ_{N} by explicitly solving (4).

To obtain a more convenient expression for $K(\Delta)$, we define the matrices of second derivatives, F_{YY} and F_{YY^*} so that the $k\ell$th entries of the ijth blocks are $[F_{YY}]_{ij,k\ell} = \frac{\partial^2 F}{\partial Y_{ij}\partial Y_{k\ell}}$ and $[F_{YY^*}]_{ij,k\ell} = \frac{\partial^2 F}{\partial Y_{ij}\partial Y^*_{k\ell}}$, respectively. By defining the matrices $A = \frac{1}{2}(F_{YY} + F_{YY^*})$ and $B = \frac{1}{2}(F_{YY} - F_{YY^*})$, using $[[\cdot]]_{ij}$ to denote the ijth block, and employing the Wirtinger convention, we can write

$$\begin{aligned}[K(\Delta)]_{ij} &= \mathrm{Tr}\left(\left(\frac{\partial^2 F}{\partial\Re\{Y_{ij}\}\partial\Re\{Y\}} - \jmath\frac{\partial^2 F}{\partial\Im\{Y_{ij}\}\partial\Re\{Y\}}\right)^T \Re\{\Delta\}\right.\\ &\quad\left. + \left(\frac{\partial^2 F}{\partial\Re\{Y_{ij}\}\partial\Im\{Y\}} - \jmath\frac{\partial^2 F}{\partial\Im\{Y_{ij}\}\partial\Im\{Y\}}\right)^T \Im\{\Delta\}\right),\\ &= \Re\left\{\mathrm{Tr}\left([[A]]_{ij}^T\Delta\right)\right\} + \jmath\Im\left\{\mathrm{Tr}\left([[B]]_{ij}^T\Delta\right)\right\}. \end{aligned} \quad (5)$$

Since each $m \times n$ block of the matrices A and B is multiplied by Δ, $[K(\Delta)]_{ij}$ can be expressed as the trace of the ijth block of $\Re\{A^T(I_m \otimes \Delta)\} + \jmath\Im\{B^T(I_m \otimes \Delta)\}$. To use this observation to write $K(\Delta)$ in a simplified form, we define the matrix $E^{p,q}_{r,s}$ to be the all zero $p \times q$ matrix with rsth entry replaced by unity. Let P_j and Q_j be the matrices defined as, $P_j = \sum_{r=1}^n E^{n,n^2}_{r,n(r-1)+j}$, and $Q_j = \sum_{r=1}^m E^{mn,m}_{n(r-1)+j,r}$, $\quad j = 1,\ldots,n$. Using this notation, it can be shown that

$$K(\Delta) = \sum_{j=1}^n P_j(\Re\{A^T(I_m \otimes \Delta)\} + \jmath\Im\{B^T(I_m\otimes\Delta)\})Q_j. \quad (6)$$

Our second step is to substitute the expression in (6) into (2) and derive an explicit expression for $F_{YY}(\Delta)$ by solving (3). Since X in (3) lies in $\mathbf{T}_Y\mathbb{G}_n(\mathbb{C}^m)$, it can be written as $X = \mathcal{P}_Y\tilde{X} = \mathcal{P}_Y^2\tilde{X}$, where $\tilde{X}$ is an arbitrary element of $\mathbb{C}^{m\times n}$ and the second equality follows from the idempotency of $\mathcal{P}_Y$. Using this fact in (2) and (3),

$$\Re\left\{\mathrm{Tr}\left(\tilde{X}^\dagger\mathcal{P}_Y^2 K^*(\Delta)\right)\right\} = \Re\left\{\mathrm{Tr}(\tilde{X}^\dagger\mathcal{P}_Y F_{YY}(\Delta))\right\}. \quad (7)$$

Since (7) must hold for all $\tilde{X} \in \mathbb{C}^{m\times n}$, we have that $F_{YY}(\Delta) = \mathcal{P}_Y K^*(\Delta)$, and hence that

$$F_{YY}(\Delta) = \mathcal{P}_Y\sum_{j=1}^n P_j(\Re\{A^T(I_m\otimes\Delta)\} - \jmath\Im\{B^T(I_m\otimes\Delta)\})Q_j. \quad (8)$$

That is, $F_{YY}(\Delta)$ is the projection of $K^*(\Delta)$ onto the tangent space of the Grassmann manifold at Y.

Having obtained (8), the Newton direction can be determined by solving (4); i.e., solving

$$\mathcal{P}_Y\sum_{j=1}^n P_j(\Re\{A^T(I_m\otimes\Delta_{\mathrm{N}})\} - \jmath\Im\{B^T(I_m\otimes\Delta_{\mathrm{N}})\})Q_j - \Delta_{\mathrm{N}}(F_Y^\dagger Y) = -G. \quad (9)$$

Notice that since the first term and G lie in $\mathbf{T}_Y\mathbb{G}_n(\mathbb{C}^m)$, the solution to (9) automatically lies in $\mathbf{T}_Y\mathbb{G}_n(\mathbb{C}^m)$. In order to make the linear nature of (9) explicit, we will apply the $\mathrm{vec}(\,\cdot\,)$ operator to the real and imaginary parts of both sides. When we do so, we obtain expressions that involve $\mathrm{vec}(I_m\otimes\Re\{\Delta_{\mathrm{N}}\})$ and $\mathrm{vec}(I_m\otimes\Im\{\Delta_{\mathrm{N}}\})$. To simplify those expressions, we make the following definition.

Definition 1 (The S-Operator): For any matrix $A\in\mathbb{C}^{J\times MNL^2}$ and integers $J,L,M,N\in\mathbb{N}$, define the S-operator to be $S_{LMN}(A) = \sum_{\ell=1}^L [A_{(\ell-1)LN+\ell} \quad A_{(\ell-1)LN+L+\ell} \quad \cdots \quad A_{(\ell-1)LN+(N-1)L+\ell}]$, where A_i is the ith $J\times M$ block of A. □

Using this definition, we have the following result, which is easy to verify.

Lemma 1: Let $Z\in\mathbb{C}^{M\times N}$ and $A\in\mathbb{C}^{J\times MNL^2}$. Then, $A\,\mathrm{vec}(I_L\otimes Z) = S_{LMN}(A)\mathrm{vec}(Z)$. □

Now, applying the $\mathrm{vec}(\,\cdot\,)$ operator to the real and imaginary parts of both sides of (9), and making the simplifying definitions $C_j = \mathcal{P}_Y P_j$ for $j=1,\ldots,n$, $D = F_Y^\dagger Y$,

$$R_1 = \sum_{j=1}^n S_{mnm}\left(Q_j^T\otimes(\Re\{C_j\}\Re\{A\} + \Im\{C_j\}\Im\{B\})\right) - (\Re\{D^T\}\otimes I_m), \quad (10)$$

$$R_2 = \sum_{j=1}^n S_{mnm}\left(Q_j^T\otimes(\Im\{C_j\}\Re\{B\} - \Re\{C_j\}\Im\{A\})\right) + (\Im\{D^T\}\otimes I_m) \quad (11)$$

$$R_3 = \sum_{j=1}^n S_{mnm}\left(Q_j^T\otimes(\Im\{C_j\}\Re\{A\} - \Re\{C_j\}\Im\{B\})\right) - (\Im\{D^T\}\otimes I_m), \quad (12)$$

$$R_4 = \sum_{j=1}^n S_{mnm}\left(Q_j^T\otimes(\Im\{C_j\}\Im\{A\} + \Re\{C_j\}\Re\{B\})\right) + (\Re\{D^T\}\otimes I_m), \quad (13)$$

we have

$$\begin{bmatrix} R_1 & R_2\\ R_3 & -R_4\end{bmatrix}\begin{bmatrix}\mathrm{vec}(\Re\{\Delta_{\mathrm{N}}\})\\ \mathrm{vec}(\Im\{\Delta_{\mathrm{N}}\})\end{bmatrix} = -\begin{bmatrix}\mathrm{vec}(\Re\{G\})\\ \mathrm{vec}(\Im\{G\})\end{bmatrix}. \quad (14)$$

We record this result in the following proposition:

Proposition 1: For an arbitrary twice-differentiable function on the complex Grassmann manifold, $F:\mathbb{G}_n(\mathbb{C}^m)\to\mathbb{R}$, the real and imaginary components of the Newton direction are given by the solution to the set of linear equations in (14). □

To determine the Newton direction Δ_{N} in (14), matrices of the form $S_{mnm}(\sum_{j=1}^n Q_j^T\otimes T_j) = \sum_{j=1}^n S_{mnm}(Q_j^T\otimes T_j)$ have to be computed; cf. (10)–(13). However, direct computation of such matrices can be difficult to implement because $(Q_j^T\otimes T_j)$ is of size $nm\times m^3n$. When the values of m and n are large, the memory required to store the n^2m^4 complex entries can be larger than that typically available on a regular computer. This difficulty can be mitigated by exploiting the Kronecker structure of the argument of the S-operator. To show this, consider the matrices $X\in\mathbb{C}^{N\times L^2}$, $Y\in\mathbb{C}^{M\times MN}$ and $A = X\otimes Y$. Now, partition the matrix $S = S_{LMN}(A)$ into N blocks of M columns each; i.e., $S = [S_1 \ \cdots \ S_N]$, where $S_i = \sum_{\ell=1}^L A_{(\ell-1)LN+\ell+(i-1)L}$, and $S_i, A_i\in\mathbb{C}^{MN\times M}$. Let X_i denote the ith column of X and $Y_{j:k}$ denote the block of Y that consists of the jth column to the kth column. Then [see (15), shown at the bottom of the page], where $I(\ell,i) = (\ell-1)LN+\ell+(i-1)L$, $J_1(\ell,i) = \left\lceil\frac{(I(\ell,i)-1)M+1}{MN}\right\rceil$, $J_2(\ell,i) = ((I(\ell,i)-1)M) \bmod MN+1$, $K(\ell,i) = (MI(\ell,i)-1)\bmod MN+$

$$A_{I(\ell,i)} = \begin{cases}\left[X_{J_1(\ell,i)}\otimes Y_{J_2(\ell,i):MN} \quad X_{\lceil\frac{I(\ell,i)}{N}\rceil}\otimes Y_{1:K(\ell,i)}\right] & \text{if } K(\ell,i) < J_2(\ell,i),\\ X_{J_1(\ell,i)}\otimes Y_{J_2(\ell,i):K(\ell,i)} & \text{otherwise,}\end{cases} \quad (15)$$

1, and $\lceil\cdot\rceil$ is the ceiling operator. Using (15) in (10)–(13) it is seen that the number of complex entries to be stored is reduced from n^2m^4 to n^2m^2.

The expression for the Newton direction on $\mathbb{G}_n(\mathbb{C}^m)$ in Proposition 1 was obtained by analyzing the characterization in (3) and (4). The corresponding expression for the real case can be obtained by applying analogous techniques to the characterization of the Newton direction on $\mathbb{G}_n(\mathbb{R}^m)$ in [1, (2.72)].

IV. The Pairwise Distance Between Multiple Points on the Grassmann Manifold

Now that we have obtained the expression in (14), what remains to be done in order to implement Newton's method on the Grassmann manifold is to obtain the gradient, F_Y, and the second derivatives, F_{YY} and F_{YY^*}, of $F(Y)$. In this section we outline the process of obtaining F_Y and F_{YY} in a special case in which the objective is a function of the pairwise distance between multiple points on the Grassmann manifold, $\mathbb{G}_n(\mathbb{C}^m)$. ($F_{YY^*}$ can be obtained in an analogous way to F_{YY}.) Let these points be represented by $Y_1, Y_2, \ldots, Y_L$. To enable joint optimization over these points, we construct the block diagonal matrix $\bar{Y} = Y_1 \oplus Y_2 \cdots \oplus Y_L \in \mathbb{G}_n(\mathbb{C}^m) \times \cdots \times \mathbb{G}_n(\mathbb{C}^m)$, where $\oplus$ denotes the direct sum operation [15]. We consider the optimization of the pairwise distances between the subspace spanned by the ith block and that spanned by the jth block. Using the block diagonal representation, the optimization of $\{Y_\ell\}_{\ell=1}^L$ can be regarded as the optimization of $\bar{Y}$ on $\mathbb{G}_{nL}(\mathbb{C}^{mL})$.

Denoting the SVD of $Y_i^\dagger Y_j$ by $U_{ij}\Sigma_{ij}V_{ij}^\dagger$, it can be shown that the pairwise distance between the subspaces spanned by the columns of Y_i and Y_j is a function of Σ_{ij} [1]. (In fact, the elements of Σ_{ij} are the cosines of the principal angles between the subspaces.) To obtain Σ_{ij} from $\bar{Y}$, we define $I_m^{(i)} \in \mathbb{C}^{m\times mL}$ to be the matrix that contains L square blocks of size m with all blocks are zero except the ith block which is replaced by the identity matrix, I_m. Then, $Y_j = I_m^{(j)}\bar{Y}(I_n^{(j)})^\dagger$, and hence

$$\Sigma_{ij} = U_{ij}^\dagger I_n^{(i)}\bar{Y}^\dagger \left(I_m^{(i)}\right)^\dagger I_m^{(j)}\bar{Y}\left(I_n^{(j)}\right)^\dagger V_{ij}. \tag{16}$$

The squared chordal Frobenius norm between the ith and jth blocks of $\bar{Y}$ is $2M - 2\mathrm{Tr}(\Sigma_{ij})$. Hence, maximizing this distance metric is equivalent to minimizing $\mathrm{Tr}(\Sigma_{ij})$. We now consider functions $F(\bar{Y})$ that are smooth functions of $\mathrm{Tr}(\Sigma_{ij})$. For the case in which $m = 2n$, we will consider the situation in which $F(\bar{Y})$ is also a function of the chordal Frobenius norms between the ith diagonal block and the orthogonal complement of the jth diagonal block; i.e., $F(\bar{Y})$ is also a function of the terms $\mathrm{Tr}((I_n - \Sigma_{ij}^2)^{1/2})$. More specifically, when $m = 2n$, we let $H_{ij}(\bar{Y}) = \mathrm{Tr}(\Sigma_{ij})$ and $\tilde{H}_{ij}(\bar{Y}) = \mathrm{Tr}((I_n - \Sigma_{ij})^{1/2})$, and we will express $F(\bar{Y})$ as the composite function $\Gamma(\{H_{ij}\}, \{\tilde{H}_{ij}\})$ for some twice differentiable function $\Gamma : \mathbb{R}^{L(L-1)} \to \mathbb{R}$. The remainder of the paper focuses on this case.

To compute $F_{\bar{Y}}$ and $F_{\bar{Y}\bar{Y}}$ for a function $F(\bar{Y})$ of the considered form we need Wirtinger derivatives for certain complex matrix functions of a complex matrix, and compatible versions of the product and chain rules. In Appendix B, we use the product rule to obtain the required Wirtinger derivatives, which are summarized in Table I. In Section IV-B we provide an explicit statement of the chain rule for differentiation of composite complex matrix functions with respect to a complex matrix.

A. Computation of First Order Derivatives

Using the conventional chain rule, the gradient of $F(\bar{Y}) = \Gamma(\{H_{ij}\}, \{\tilde{H}_{ij}\})$ can be expressed as $\sum_{ij} \frac{\partial\Gamma}{\partial H_{ij}}\frac{\partial H_{ij}}{\partial \bar{Y}} + \frac{\partial\Gamma}{\partial \tilde{H}_{ij}}\frac{\partial \tilde{H}_{ij}}{\partial \bar{Y}}$. Since Γ, $\{H_{ij}\}$ and $\{\tilde{H}_{ij}\}$ are real scalars, $\frac{\partial\Gamma}{\partial H_{ij}}$ and $\frac{\partial\Gamma}{\partial \tilde{H}_{ij}}$ can be obtained in a straightforward manner. To compute the derivatives $\frac{\partial}{\partial\bar{Y}}\mathrm{Tr}(\Sigma_{ij})$ and $\frac{\partial}{\partial\bar{Y}}\mathrm{Tr}((I_n - \Sigma_{ij}^2)^{1/2})$, we use the fact that the variation of Σ_{ij} is due to the variation of the subspaces spanned by the columns of Y_i and Y_j and is independent of the bases that span these subspaces. To see this, let $\{Q_\ell\}_{\ell=1}^L$ be a set of arbitrary $n\times n$ unitary matrices. Then, the subspace spanned by $\breve{Y}_\ell = Y_\ell Q_\ell$ is the same as that spanned by Y_ℓ. Furthermore, the SVD of $\breve{Y}_i^\dagger \breve{Y}_j$ can be expressed as $\breve{U}_{ij}\Sigma_{ij}\breve{V}_{ij}^\dagger$, where $\breve{U}_{ij} = Q_i^\dagger U_{ij}$ and $\breve{V}_{ij} = Q_j^\dagger V_{ij}$. Hence, it is seen that varying the particular bases that span the subspaces of Y_i and Y_j does not change Σ_{ij}. From this observation, it can be readily seen that assuming that the unitary matrices U_{ij} and V_{ij} are fixed does not affect the computation of the derivatives $\frac{\partial}{\partial\bar{Y}}\mathrm{Tr}(\Sigma_{ij})$ and $\frac{\partial}{\partial\bar{Y}}\mathrm{Tr}((I_n - \Sigma_{ij}^2)^{1/2})$. That enables us to write

$$\frac{\partial}{\partial\bar{Y}}\mathrm{Tr}(\Sigma_{ij}) = \left(I_m^{(i)}\right)^\dagger I_m^{(j)}\bar{Y}\left(I_n^{(j)}\right)^\dagger V_{ij}U_{ij}^\dagger I_n^{(i)} + \left(I_m^{(j)}\right)^\dagger I_m^{(i)}\bar{Y}\left(I_n^{(i)}\right)^\dagger U_{ij}V_{ij}^\dagger I_n^{(j)}. \tag{17}$$

To derive an expression for $\frac{\partial}{\partial\bar{Y}}\mathrm{Tr}((I - \Sigma_{ij}^2)^{1/2})$, we note that $\frac{\partial}{\partial\bar{Y}}\mathrm{Tr}((I_n - \Sigma_{ij}^2)^{1/2}) = -\sum_{\ell=1}^M \frac{\sigma_\ell}{\sqrt{1-\sigma_\ell^2}}\frac{\partial\sigma_\ell}{\partial\bar{Y}}$, where σ_ℓ is the ℓth entry of Σ_{ij} and can be written as

$$\sigma_\ell = e_\ell^\dagger U_{ij}^\dagger I_n^{(i)}\bar{Y}^\dagger A_m^{ij}\bar{Y}\left(I_n^{(j)}\right)^\dagger V_{ij}e_\ell = \mathrm{Tr}\left(\bar{Y}^\dagger A_m^{ij}\bar{Y}\left(I_n^{(j)}\right)^\dagger V_{ij}e_\ell e_\ell^\dagger U_{ij}^\dagger I_n^{(i)}\right)$$

where e_ℓ is the $n\times 1$ all zero vector with unity in the ℓth position. Using this notation, we have

$$\frac{\partial\sigma_\ell}{\partial\bar{Y}} = A_m^{ij}\bar{Y}\left(I_n^{(j)}\right)^\dagger V_{ij}e_\ell e_\ell^\dagger U_{ij}^\dagger I_n^{(i)} + A_m^{ji}\bar{Y}\left(I_n^{(i)}\right)^\dagger U_{ij}e_\ell e_\ell^\dagger V_{ij}^\dagger I_n^{(j)}. \tag{18}$$

Using (18), denoting $U_{ij}\Sigma_{ij}(I - \Sigma_{ij}^2)^{-1/2}V_{ij}^\dagger$ by Ω_{ij}, and simplifying, we have

$$\frac{\partial}{\partial\bar{Y}}\mathrm{Tr}\left(\left(I_n - \Sigma_{ij}^2\right)^{1/2}\right) = -A_m^{ij}\bar{Y}\left(I_n^{(j)}\right)^\dagger\Omega_{ij}^\dagger I_n^{(i)} - A_m^{ji}\bar{Y}\left(I_n^{(i)}\right)^\dagger\Omega_{ij}I_n^{(j)}. \tag{19}$$

Although the expressions in (17) and (19) are convenient forms of the first derivatives, in order to compute the second derivatives we need to obtain an expression for these gradients in terms of $\bar{Y}$. To do so, we use (16) to express $V_{ij}U_{ij}^\dagger$ in terms of $\bar{Y}$, namely

$$V_{ij}U_{ij}^\dagger = \left(I_n^{(i)}\bar{Y}^\dagger A_m^{ij}\bar{Y}\left(I_n^{(j)}\right)^\dagger\right)^{-1} \times \left(I_n^{(i)}\bar{Y}^\dagger A_m^{ij}\bar{Y}A_n^{jj}\bar{Y}^\dagger A_m^{ji}\bar{Y}\left(I_n^{(i)}\right)^\dagger\right)^{1/2} \tag{20}$$

TABLE I
WIRTINGER DERIVATIVES WITH RESPECT TO COMPLEX MATRICES

Let $Z\in\mathbb{C}^{M\times N}$, $\Phi\in\mathbb{C}^{K\times L}$, $W\in\mathbb{C}^{I\times J}$, $X\in\mathbb{C}^{P_1\times P_2}$, $Y\in\mathbb{C}^{P_2\times P_3}$.
Let $U = \sum_{rs} E_{rs}^{M,N}\otimes(E_{rs}^{M,N})^T$ and $\bar{U} = \sum_{rs} E_{rs}^{M,N}\otimes E_{rs}^{M,N}$.

Product rule [17]	$\frac{\partial(XY)}{\partial Z} = \frac{\partial X}{\partial Z}(I_N\otimes Y) + (I_M\otimes X)\frac{\partial Y}{\partial Z}$
Chain rule [17]	$\frac{\partial W(\Phi(Z))}{\partial Z} = \left(\frac{\partial(\mathrm{vec}(\Phi))^T}{\partial Z}\otimes I_I\right)\left(I_N\otimes\frac{\partial W}{\partial\,\mathrm{vec}(\Phi)}\right)$
Z with indep. entries	$\frac{\partial Z}{\partial Z} = 2\bar{U}$ and $\frac{\partial Z^\dagger}{\partial Z} = 0$
Hermitian Z	$\frac{\partial Z}{\partial Z} = 2\bar{U} - \sum_{r=1}^M E_{rr}^{M,M}\otimes E_{rr}^{M,M}$
Hermitian Z	$\frac{\partial Z^*}{\partial Z} = 2U - \sum_{r=1}^M E_{rr}^{M,M}\otimes E_{rr}^{M,M}$
General Z	$\frac{\partial Z^{-1}}{\partial Z} = -(I_M\otimes Z^{-1})\frac{\partial Z}{\partial Z}(I_M\otimes Z^{-1})$
General Z	$\frac{\partial(Z^{1/2})}{\partial Z} = T(Z)\,\mathrm{vec}\left(\frac{\partial Z}{\partial Z}\right)$, see (36)
General Z	$\frac{\partial Z^{-1/2}}{\partial Z} = -(I_M\otimes Z^{-1/2})\frac{\partial Z^{1/2}}{\partial Z}(I_M\otimes Z^{-1/2})$
Z with indep. entries	$\frac{\partial(Z^{1/2})^\dagger}{\partial Z} = 0$ and $\frac{\partial(Z^{-1/2})^\dagger}{\partial Z} = 0$

where $A_p^{ij} = (I_p^{(i)})^\dagger I_p^{(j)} = (A_p^{ij})^T$. Denoting the matrix $I_n^{(i)}\bar{Y}^\dagger A_m^{ij}\bar{Y}A_n^{jj}\bar{Y}^\dagger A_m^{ji}\bar{Y}(I_n^{(i)})^\dagger$ by $\tilde{\Omega}_{ij}$ and the matrix $\bar{Y}^\dagger A_m^{ij}\bar{Y}$ by $\bar{\Psi}_{ij}$, and substituting from (20) into (17), we obtain,

$$\frac{\partial}{\partial \bar{Y}}\mathrm{Tr}(\Sigma_{ij}) = A_m^{ij}\bar{Y}\left(I_n^{(j)}\right)^\dagger\left(I_n^{(i)}\bar{\Psi}_{ij}\left(I_n^{(j)}\right)^\dagger\right)^{-1}\tilde{\Omega}_{ij}^{1/2}I_n^{(i)} + A_m^{ji}\bar{Y}\left(I_n^{(i)}\right)^\dagger\tilde{\Omega}_{ij}^{1/2}\left(I_n^{(j)}\bar{\Psi}_{ij}^\dagger\left(I_n^{(i)}\right)^\dagger\right)^{-1}I_n^{(j)}. \quad (21)$$

For the terms related to the orthogonal complement, we write

$$V_{ij}\Sigma_{ij}\left(I-\Sigma_{ij}^2\right)^{-1/2}U_{ij}^\dagger = I_n^{(j)}\bar{Y}^\dagger A_m^{ji}\bar{Y}\left(I_n^{(i)}\right)^\dagger \times\left(I_n - I_n^{(i)}\bar{Y}^\dagger A_m^{ij}\bar{Y}A_n^{jj}\bar{Y}^\dagger A_m^{ji}\bar{Y}\left(I_n^{(i)}\right)^\dagger\right)^{-1/2}. \quad (22)$$

Substituting from (22) into (19) we have

$$\frac{\partial}{\partial \bar{Y}}\mathrm{Tr}\left(\left(I_n-\Sigma_{ij}^2\right)^{1/2}\right) = -A_m^{ij}\bar{Y}A_n^{jj}\bar{\Psi}_{ij}^\dagger\left(I_n-\tilde{\Omega}_{ij}\right)^{-1/2} \times I_n^{(i)} - A_m^{ji}\bar{Y}\left(I_n^{(i)}\right)^\dagger\left(I_n-\tilde{\Omega}_{ij}\right)^{-1/2}I_n^{(i)}\bar{\Psi}_{ij}A_n^{ji}. \quad (23)$$

B. Computation of Second Order Derivatives

Our goal in this section is to compute the Hessian, $F_{\bar{Y}\bar{Y}}$; that is, the derivatives of the components of $F_{\bar{Y}}$. To simplify the notation, we will actually compute the derivatives with respect to a general matrix Y that is not necessarily block diagonal. (It can be shown that if the Newton method is initialized with a block diagonal matrix, subsequent iterates retain the block diagonal structure.) To further simplify the notation, we observe that the structure of $\frac{\partial\Gamma}{\partial \bar{Y}}$, and the expressions in (21) and (23) means that the functions that need to be differentiated take the forms $\theta(Y)W_k(Y), k=1,\ldots,4$, where $\theta(Y)$ denotes either $\frac{\partial\Gamma}{\partial H_{ij}}$ or $\frac{\partial\Gamma}{\partial \bar{H}_{ij}}$, and the matrices $\{W_k(Y)\}_{k=1}^4$ take the following forms:

$$W_1(Y) = AYB(CY^\dagger AYB)^{-1}(CY^\dagger AYDY^\dagger A^\dagger YC^\dagger)^{1/2}C \quad (24a)$$

$$W_2(Y) = A^\dagger YC^\dagger(CY^\dagger AYDZ^\dagger A^\dagger YC^\dagger)^{1/2}(B^\dagger Y^\dagger A^\dagger YC^\dagger)^{-1}B^\dagger \quad (24b)$$

$$W_3(Y) = AYDY^\dagger A^\dagger YC^\dagger(I_M - CY^\dagger AYDY^\dagger A^\dagger YC^\dagger)^{-1/2}C \quad (24c)$$

$$W_4(Y) = A^\dagger YC^\dagger(I_M - CY^\dagger AYDY^\dagger A^\dagger YC^\dagger)^{-1/2}CY^\dagger AYE \quad (24d)$$

where $A, B, C, D,$ and E are constant matrices. To calculate the derivative with respect to Y of a function of the form $\{\theta(Y)W_k(Y)\}_{k=1}^4$, we note that for a scalar function $\theta(Y)$ and a matrix function $W(Y)$ we have [16]

$$\frac{\partial\left(\theta(Y)W(Y)\right)}{\partial Y} = \theta(Y)\frac{\partial W(Y)}{\partial Y} + \frac{\partial\theta(Y)}{\partial Y}\otimes W(Y). \quad (25)$$

The derivative $\frac{\partial\theta(Y)}{\partial Y}$ can be computed by direct analogy with the differentiation techniques for real matrices (e.g., [16] and [17]), but computing $\frac{\partial W(Y)}{\partial Y}$ requires Wirtinger derivatives of a complex-valued matrix function with respect to a complex matrix and compatible versions of the product and chain rules. In general, the analogy with the real case does not extend to matrix functions, and the required derivatives are obtained in Appendix B. For convenience, we have summarized the key results in Table I.

Expressions for $\frac{\partial W_k(Y)}{\partial Y}$ can be constructed by applying a combination of the product and chain rules and the expressions in Table I. Applying the product rule on the matrix products in (24), we have

$$\frac{\partial AZB}{\partial Z} = 2(I\otimes A)\bar{U}(I\otimes B),$$
$$\frac{\partial Z^\dagger A^\dagger ZC^\dagger}{\partial Z} = 2(I\otimes Z^\dagger A^\dagger)\bar{U}(I\otimes C^\dagger) \quad (26)$$
$$\frac{\partial (CZ^\dagger AZB)}{\partial Z} = 2(I\otimes CZ^\dagger A)\bar{U}(I\otimes B),$$
$$\frac{\partial (CZ^\dagger AZB)^*}{\partial Z} = 2(I\otimes C^*)U(I\otimes A^*Z^*B^*) \quad (27)$$

$$\frac{\partial\Psi(Z)}{\partial Z} = \frac{\partial CZ^\dagger AZD}{\partial Z}(I\otimes Z^\dagger A^\dagger ZC^\dagger) + (I\otimes CZ^\dagger AZD)\frac{\partial Z^\dagger A^\dagger ZC^\dagger}{\partial Z} \quad (28)$$

$$\frac{\partial\Psi^*(Z)}{\partial Z} = \frac{\partial C^*Z^TA^*Z^*D^*}{\partial Z}(I\otimes Z^TA^TZ^*C^T) + (I\otimes C^*Z^TA^*Z^*D^*)\frac{\partial Z^TA^TZ^*C^T}{\partial Z} \quad (29)$$

where $\frac{\partial Z^TA^TZ^*C^T}{\partial Z} = 2U(I\otimes A^TZ^*C^T)$, and $\Psi(Z) = CZ^\dagger AZDZ^\dagger A^\dagger ZC^\dagger$.

To complete the expressions for $\frac{\partial W_k(Y)}{\partial Y}$ we need to apply the chain rule for the cases in which $\Phi(Z) = Z^{1/2}, Z^{-1/2}$ and Z^{-1}. However, applying the expression in Table I, for the case of complex matrices is quite cumbersome, and hence in Appendix C, we derive the following explicit expression:

$$\frac{\partial W(\Phi(Z))}{\partial Z} = \Upsilon_1 + \jmath\Upsilon_2 \quad (30)$$

where

$$\begin{aligned}\Upsilon_1 = &\left(\frac{\partial(\mathrm{vec}(\Re\{\Phi\}))^T}{\partial\Re\{Z\}}\otimes I_I\right)\left(I_N\otimes\frac{\partial\Re\{W\}}{\partial\mathrm{vec}(\Re\{\Phi\})}\right)\\ &+\left(\frac{\partial(\mathrm{vec}(\Im\{\Phi\}))^T}{\partial\Re\{Z\}}\otimes I_I\right)\left(I_N\otimes\frac{\partial\Re\{W\}}{\partial\mathrm{vec}(\Im\{\Phi\})}\right)\\ &+\left(\frac{\partial(\mathrm{vec}(\Re\{\Phi\}))^T}{\partial\Im\{Z\}}\otimes I_I\right)\left(I_N\otimes\frac{\partial\Im\{W\}}{\partial\mathrm{vec}(\Re\{\Phi\})}\right)\\ &+\left(\frac{\partial(\mathrm{vec}(\Im\{\Phi\}))^T}{\partial\Im\{Z\}}\otimes I_I\right)\left(I_N\otimes\frac{\partial\Im\{W\}}{\partial\mathrm{vec}(\Im\{\Phi\})}\right),\\ \Upsilon_2 = &\left(\frac{\partial(\mathrm{vec}(\Re\{\Phi\}))^T}{\partial\Re\{Z\}}\otimes I_I\right)\left(I_N\otimes\frac{\partial\Im\{W\}}{\partial\mathrm{vec}(\Re\{\Phi\})}\right)\\ &+\left(\frac{\partial(\mathrm{vec}(\Im\{\Phi\}))^T}{\partial\Re\{Z\}}\otimes I_I\right)\left(I_N\otimes\frac{\partial\Im\{W\}}{\partial\mathrm{vec}(\Im\{\Phi\})}\right)\\ &-\left(\frac{\partial(\mathrm{vec}(\Re\{\Phi\}))^T}{\partial\Im\{Z\}}\otimes I_I\right)\left(I_N\otimes\frac{\partial\Re\{W\}}{\partial\mathrm{vec}(\Re\{\Phi\})}\right)\\ &-\left(\frac{\partial(\mathrm{vec}(\Im\{\Phi\}))^T}{\partial\Im\{Z\}}\otimes I_I\right)\left(I_N\otimes\frac{\partial\Re\{W\}}{\partial\mathrm{vec}(\Im\{\Phi\})}\right)\end{aligned}$$

where, for complex matrices, X_1 and X_2, the real and imaginary components of the derivatives can be computed directly by applying the Wirtinger convention (cf. (35)) to entries of $\frac{\partial X_1}{\partial X_2}$. Doing so, we have

$$\frac{\partial\Re\{X_1\}}{\partial\Re\{X_2\}} = \frac{1}{2}\Re\left\{\frac{\partial X_1}{\partial X_2}+\frac{\partial X_1^*}{\partial X_2}\right\},\quad \frac{\partial\Re\{X_1\}}{\partial\Im\{X_2\}} = -\frac{1}{2}\Im\left\{\frac{\partial X_1}{\partial X_2}+\frac{\partial X_1^*}{\partial X_2}\right\}, \quad (31a)$$

$$\frac{\partial\Im\{X_1\}}{\partial\Re\{X_2\}} = \frac{1}{2}\Re\left\{\frac{\partial X_1}{\partial X_2}-\frac{\partial X_1^*}{\partial X_2}\right\},\quad \frac{\partial\Im\{X_1\}}{\partial\Im\{X_2\}} = -\frac{1}{2}\Im\left\{\frac{\partial X_1}{\partial X_2}-\frac{\partial X_1^*}{\partial X_2}\right\}. \quad (31b)$$

The computation of the expression in (30) requires the determination of $\frac{\partial(\mathrm{vec}(\Phi))^T}{\partial Z}$ and $\frac{\partial W}{\partial\mathrm{vec}(\Phi)}$ from $\frac{\partial\Phi}{\partial Z}$ and $\frac{\partial W}{\partial\Phi}$. To do so, it can be shown that the $M\times NKL$ matrix $\frac{\partial(\mathrm{vec}(\Phi))^T}{\partial Z}$ can be expressed as

$$\left[\frac{\partial(\mathrm{vec}(\Phi))^T}{\partial Z}\right]_{r,(s-1)KL+1:sKL} = \left(\left[\mathrm{vec}\left(\frac{\partial\Phi}{\partial Z}\right)\right]_{(r-1)K+1:rK,(s-1)L+1:sL}\right)^T, \quad (32)$$

for $r=1,\ldots,M, s=1,\ldots,N$, and that the $IKL\times J$ matrix $\frac{\partial W}{\partial\mathrm{vec}(\Phi)} = \mathrm{bvec}(\frac{\partial W}{\partial\Phi}, J)$, where $\mathrm{bvec}(\cdot,k)$ is used to denote the operator that stacks blocks of k columns in one tall matrix.

C. Steps for Constructing the Hessian

We are now ready to compute the Hessian required for obtaining the Newton direction in (14). For brevity, we will not write the Hessian F_{YY} explicitly, but the method to obtain it is outlined below. (Recall that F_{YY^*} can be obtained in an analogous manner.)

- Using the expression for $\frac{\partial \Gamma}{\partial Y}$, and the expressions in (21) and (23), identify the terms that need to be differentiated. These terms take the form $\theta(Y)W_k(Y)$.
- For a given scalar function, $\theta(Y)$, use (25) to express $\frac{\partial \theta(Y)W_k(Y)}{\partial Y}$ in terms of $\frac{\partial W_k(Y)}{\partial Y}$.
- Use the multiplication rule in Table I and the expressions in (26)–(28) to express the derivatives of $\{W_k(Y)\}_{k=1}^4$ in terms of the derivatives of the terms with composite functions of the form of matrix square root, inverse and inverse square root. For each case, denote the composite function by W.
- For each composite function let the intermediate function be denoted by Φ. Use the expressions in Table I to compute $\frac{\partial W}{\partial \Phi}$ and use the expressions that follow (32) to obtain $\frac{\partial W}{\partial \mathrm{vec}(\Phi)}$.
- Use (31) with (27), (28), (27), and (29) to compute $\frac{\partial \Re\{\Phi\}}{\partial \Re\{Y\}}$, $\frac{\partial \Re\{\Phi\}}{\partial \Im\{Y\}}$, $\frac{\partial \Im\{\Phi\}}{\partial \Re\{Y\}}$ and $\frac{\partial \Im\{\Phi\}}{\partial \Im\{Y\}}$, and use (32) to obtain $\frac{\partial \mathrm{vec}(\Re\{\Phi\})}{\partial \Re\{Y\}}$, $\frac{\partial \mathrm{vec}(\Re\{\Phi\})}{\partial \Im\{Y\}}$, $\frac{\partial \mathrm{vec}(\Im\{\Phi\})}{\partial \Re\{Y\}}$ and $\frac{\partial \mathrm{vec}(\Im\{\Phi\})}{\partial \Im\{Y\}}$.
- Use (30) with the multiplication rule in Table I to obtain F_{YY}.

V. Conclusion

Optimization on the Grassmann manifold is an important tool for approaching several important design problems in signal processing for communications, including those that involve sphere packing for non-coherent MIMO communications and codebook designs for limited feedback systems. In this paper, we provided a structured procedure for obtaining the Newton direction required for optimizing arbitrary twice differentiable objectives on the complex Grassmann manifold. To illustrate the procedure, we considered a class of objectives that involves the pairwise chordal Frobenius norm between multiple points on the Grassmann manifold and we provided an expression for the gradient and a structured method for obtaining the second derivatives.

Appendix

A. The Wirtinger Convention for Differentiation of Complex Scalars

Let $z = z_r + \jmath z_i$ be a complex function with differentiable real and imaginary parts. The Wirtinger derivative of z with respect to the complex variable $y = y_r + \jmath y_i$ is given by (e.g., [14] and [18])

$$\begin{aligned}\frac{dz}{dy} &= \frac{\partial z}{\partial y_r} + \frac{\partial z}{\partial \jmath y_i} = \frac{\partial z}{\partial y_r} - \jmath \frac{\partial z}{\partial y_i}\\ &= \frac{\partial z_r}{\partial y_r} + \frac{\partial z_i}{\partial y_i} + \jmath\left(\frac{\partial z_i}{\partial y_r} - \frac{\partial z_r}{\partial y_i}\right). \quad (33)\end{aligned}$$

Using this definition, we have $\frac{dz^*}{dy} = \frac{\partial z_r}{\partial y_r} - \frac{\partial z_i}{\partial y_i} - \jmath\left(\frac{\partial z_i}{\partial y_r} + \frac{\partial z_r}{\partial y_i}\right)$. Observe that if z is an analytic function of y, then $\frac{\partial z_r}{\partial y_r} = \frac{\partial z_i}{\partial y_i}$, and $\frac{\partial z_i}{\partial y_r} = -\frac{\partial z_r}{\partial y_i}$, which leads to $\frac{dz}{dy} = 2(\frac{\partial z_r}{\partial y_r} + \jmath \frac{\partial z_i}{\partial y_r}) = 2(\frac{\partial z_i}{\partial y_i} - \jmath \frac{\partial z_r}{\partial y_i})$, and to

$$\frac{dz^*}{dy} = 0. \quad (34)$$

For any smooth complex function, not necessarily analytic, we have

$$\begin{aligned}\frac{1}{2}\left(\frac{dz}{dy} + \frac{dz^*}{dy}\right) &= \frac{\partial \Re(z)}{\partial y} = \frac{\partial z_r}{\partial y_r} - \jmath \frac{\partial z_r}{\partial y_i}, \quad \text{and}\\ \frac{1}{2\jmath}\left(\frac{dz}{dy} - \frac{dz^*}{dy}\right) &= \frac{\partial \Im(z)}{\partial y} = \frac{\partial z_i}{\partial y_r} - \jmath \frac{\partial z_i}{\partial y_i}. \quad (35)\end{aligned}$$

B. Wirtinger Derivatives of Complex Matrices

The results in this section are obtained by invoking the Wirtinger convention outlined in Appendix A. (See [19] for another approach to differentiation of complex-valued matrices.) Using the notation defined in Table I, the entries of that table can be obtained in the following ways.

- If the entries of $Z \in \mathbb{C}^{M\times N}$ are independent, then using Appendix A, we have that $\frac{\partial Z}{\partial Z} = 2\bar{U}$.
- If $Z \in \mathbb{C}^{M\times M}$ is Hermitian, $\frac{\partial Z}{\partial Z} = 2\bar{U} - \bar{E}$, and $\frac{\partial Z^*}{\partial Z} = 2U - \bar{E}$, where $\bar{E} = \sum_{r=1}^{M} E_{rr}^{M,M} \otimes E_{rr}^{M,M}$.
- If the entries of $Z \in \mathbb{C}^{M\times N}$ are independent, then it follows from (34) that $\frac{\partial (Z^\dagger)}{\partial Z} = 0$, and $\frac{\partial (Z^*)}{\partial Z} = 0$.
- For a nonsingular matrix $Z \in \mathbb{C}^{M\times M}$, applying the product rule in Table I with $X = Z^{-1}$ and $Y = Z$ yields $\frac{\partial Z^{-1}}{\partial Z} = -2(I_M \otimes Z^{-1})\frac{\partial Z}{\partial Z}(I_M \otimes Z^{-1})$. If Z has independent entries, $\frac{\partial (Z^{-1})^\dagger}{\partial Z} = 0$.
- For a matrix $Z \in \mathbb{C}^{M\times M}$ for which $Z^{1/2}$ exists, using $X = Y = Z^{1/2}$ in the product rule yields

$$\begin{aligned}\frac{\partial (Z^{1/2})}{\partial Z} = \mathrm{mat}_{M^2}&\left(I_M \otimes Z^{T/2} \otimes I_{M^2} + I_{M^3} \otimes Z^{1/2}\right)^{-1}\\ &\times \mathrm{vec}\left(\frac{\partial Z}{\partial Z}\right) \quad (36)\end{aligned}$$

  where the operator $\mathrm{mat}_k(\cdot)$ reverses the action of the $\mathrm{vec}(\cdot)$ operator; it forms a k-column matrix from a vector of ℓk entries for some positive integer ℓ. If Z has independent entries, $\frac{\partial (Z^{1/2})^\dagger}{\partial Z} = 0$.
- For a nonsingular matrix $Z \in \mathbb{C}^{M\times M}$ for which $Z^{1/2}$ exists, using $X = Y^{-1} = Z^{1/2}$ in the product rule yields $\frac{\partial Z^{-1/2}}{\partial Z} = -(I_M \otimes Z^{-1/2})\frac{\partial Z^{1/2}}{\partial Z}(I_M \otimes Z^{-1/2})$. If Z has independent entries $\frac{\partial (Z^{-1/2})^\dagger}{\partial Z} = 0$.

C. An Explicit Expression for the Chain Rule for Complex Matrices

We now use the Wirtinger convention to derive the expression in (30) for the chain rule. The alternative approach to complex matrix differentiation in [19] yields a different expression for the chain rule.

The expression for the chain rule in Table I can be written as

$$\begin{aligned}\frac{\partial W(\Phi(Z))}{\partial Z} &= \sum_{r,s} E_{rs}^{M,N} \otimes \sum_{i,j} E_{ij}^{M,N} \frac{\partial [W]_{ij}}{\partial [Z]_{rs}}\\ &= \sum_{r,s} E_{rs}^{M,N} \otimes \sum_{i,j} E_{ij}^{M,N} \sum_{\alpha,\beta} \frac{\partial [W]_{ij}}{\partial [\Phi]_{\alpha\beta}} \frac{\partial [\Phi]_{\alpha\beta}}{\partial [Z]_{rs}}. \quad (37)\end{aligned}$$

Hence, $\frac{\partial [W]_{ij}}{\partial [Z]_{rs}} = \sum_{\alpha,\beta} \frac{\partial [W]_{ij}}{\partial [\Phi]_{\alpha\beta}} \frac{\partial [\Phi]_{\alpha\beta}}{\partial [Z]_{rs}}$. When both W and Z are complex matrices, the expression in (33) is used to compute the scalars $\frac{\partial [W]_{ij}}{\partial [Z]_{rs}} = \frac{\partial \Re\{[W]_{ij}\}}{\partial \Re\{[Z]_{rs}\}} + \frac{\partial \Im\{[W]_{ij}\}}{\partial \Im\{[Z]_{rs}\}} + \jmath\left(\frac{\partial \Im\{[W]_{ij}\}}{\partial \Re\{[Z]_{rs}\}} - \frac{\partial \Re\{[W]_{ij}\}}{\partial \Im\{[Z]_{rs}\}}\right)$, where $\frac{\partial \Re\{[W]_{ij}\}}{\partial \Re\{[Z]_{rs}\}} = \sum_{\alpha,\beta} \frac{\partial \Re\{[W]_{ij}\}}{\partial \Re\{[\Phi]_{\alpha\beta}\}} \frac{\partial \Re\{[\Phi]_{\alpha\beta}\}}{\partial \Re\{[Z]_{rs}\}} + \frac{\partial \Re\{[W]_{ij}\}}{\partial \Im\{[\Phi]_{\alpha\beta}\}} \frac{\partial \Im\{[\Phi]_{\alpha\beta}\}}{\partial \Re\{[Z]_{rs}\}}$, and expressions for $\frac{\partial \Im\{[W]_{ij}\}}{\partial \Im\{[Z]_{rs}\}}$, $\frac{\partial \Re\{[W]_{ij}\}}{\partial \Im\{[Z]_{rs}\}}$, and $\frac{\partial \Im\{[W]_{ij}\}}{\partial \Re\{[Z]_{rs}\}}$ are obtained in an analogous way. The expression in (30) is obtained by using these expressions in (37) and arranging the elements in a matrix form.

Acknowledgment

R. H. Gohary would like to thank Dr. H. Yanikomeroglu of Carleton University, Ottawa, ON, Canada, for his support.

References

[1] A. Edelman, T. Arias, and S. Smith, "The geometry of algorithms with orthogonality constraints," *SIAM J. Matrix Anal. Appl.*, vol. 20, no. 2, pp. 303–353, 1998.

[2] B. M. Hochwald and T. L. Marzetta, "Unitary space-time modulation multiple-antenna communications in Rayleigh flat fading," *IEEE Trans. Inf. Theory*, vol. 46, pp. 543–564, Mar. 2000.
[3] L. Zheng and D. N. C. Tse, "Communication on the Grassmann manifold: A geometric approach to the noncoherent multiple-antenna channel," *IEEE Trans. Inf. Theory*, vol. 48, pp. 359–383, Feb. 2002.
[4] O. Henkel, "Geometrical relations between space time block code designs and complexity reduction," *IEEE Trans. Inf. Theory*, vol. 52, pp. 5324–5335, Dec. 2006.
[5] I. Kammoun, A. M. Cipriano, and J.-C. Belfiore, "Non-coherent codes over the Grassmannian," *IEEE Trans. Wireless Commun.*, vol. 6, pp. 3657–3667, Oct. 2007.
[6] R. H. Gohary and T. N. Davidson, "Non-coherent MIMO communication: Grassmannian constellations and efficient detection," *IEEE Trans. Inf. Theory*, vol. 55, pp. 1176–1205, Mar. 2009.
[7] D. J. Love and R. W. Heath, Jr., "Limited feedback unitary precoding for orthogonal space-time block codes," *IEEE Trans. Signal Process.*, vol. 53, no. 1, pp. 64–73, Jan. 2005.
[8] M. B. Shenouda and T. N. Davidson, "A design framework for limited feedback MIMO systems with zero-forcing DFE," *IEEE J. Sel. Areas Commun.*, vol. 26, pp. 1578–1587, Oct. 2008.
[9] D. Love, R. Heath, V. Lau, D. Gesbert, B. Rao, and M. Andrews, "An overview of limited feedback in wireless communication systems," *IEEE J. Sel. Areas Commun.*, vol. 26, pp. 1341–1365, Oct. 2008.
[10] D. Agrawal, T. Richardson, and R. Urbanke, "Multiple-antenna signal constellations for fading channels," *IEEE Trans. Inf. Theory*, vol. 47, pp. 2618–2626, Sep. 2001.
[11] J. H. Conway, R. H. Hardin, and N. J. A. Sloane, "Packing lines, planes, etc.: Packings in Grassmannian spaces," *Exper. Math.*, vol. 5, no. 2, pp. 139–159, 1996.
[12] J. Nocedal and S. J. Wright, *Numerical Optimization*, 2nd ed. New York: Springer, 2006.
[13] J. H. Manton, "Optimization algorithms exploiting unitary constraints," *IEEE Trans. Signal Process.*, vol. 50, no. 3, pp. 635–650, Mar. 2002.
[14] K. Kreutz-Delgado, "The complex gradient operator and the $\mathbb{CR}$-calculus," Dept. Electr. Comput. Eng., Univ. of California, San Diego, CA, Tech. Rep., Jun. 2009 [Online]. Available: http://arxiv.org/abs/0906.4835
[15] R. A. Horn and C. R. Johnson, *Topics in Matrix Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 1994.
[16] J. W. Brewer, "Kronecker products and matrix calculus in system theory," *IEEE Trans. Circuits Sys.*, vol. CAS-25, pp. 772–781, Sep. 1978.
[17] A. Graham, *Kronecker Products and Matrix Calculus: With Applications*. New York: Elis Horwood Ltd., 1981.
[18] S. Haykin, *Adaptive Filter Theory*, 4th ed. Upper Saddle River, NJ: Prentice-Hall, 2002.
[19] A. Hjørungnes and D. Gesbert, "Complex-valued matrix differentiation: Techniques and key results," *IEEE Trans. Signal Process.*, vol. 55, no. 6, pp. 2740–2746, Jun. 2007.

Turning Tangent Empirical Mode Decomposition: A Framework for Mono- and Multivariate Signals

Julien Fleureau, Jean-Claude Nunes, Amar Kachenoura, Laurent Albera, and Lotfi Senhadji

***Abstract*—A novel empirical mode decomposition (EMD) algorithm, called 2T-EMD, for both mono- and multivariate signals is proposed in this correspondence. It differs from the other approaches by its computational lightness and its algorithmic simplicity. The method is essentially based on a redefinition of the signal mean envelope, computed thanks to new characteristic points, which offers the possibility to decompose multivariate signals without any projection. The scope of application of the novel algorithm is specified, and a comparison of the 2T-EMD technique with classical methods is performed on various simulated mono- and multivariate signals. The monovariate behaviour of the proposed method on noisy signals is then validated by decomposing a fractional Gaussian noise and an application to real life EEG data is finally presented.**

***Index Terms*—Analysis of nonlinear and nonstationary signals, EEG denoising, extrema and barycenters of oscillation, filter bank structure, Hurst exponent estimation, intrinsic mode functions, mono- and multivariate empirical mode decomposition, time varying representation.**

I. Introduction

Empirical mode decomposition (EMD) was originally introduced in the late 1990's to study water surface wave evolution [1]. The EMD can be considered as an emerging technique in signal processing with a very important topic of research and development in various fields such as biomedical signal analysis [2], Hurst exponent estimation [3], speech processing [4], texture analysis [5], etc. It decomposes adaptively a given signal, s, into a sum of N AM-FM components, d_n [referred to as the intrinsic mode functions (IMFs)], plus a residue a_N. An IMF is defined [1] as a locally centered function where the number of extrema and the number of zero-crossings must differ at most by one. More precisely, for a given signal $s = a_0$, the EMD sequentially computes the N IMFs d_n, and N corresponding trends a_n, such that $a_{n-1} = a_n + d_n$. The EMD key issue is then the extraction of the N IMFs d_n. In practice, such a signal is obtained by stopping a so-called *sifting process*, using a Cauchy-like criterion [6]. If k denotes the number of iterations in the sifting process, the so-called sifting process can be summarized as follows:

1) Initialization with $d_{n,0} = a_{n-1}$.
2) Computation of the mean envelope $\mathcal{M}(d_{n,k})$.
3) Extraction of the detail $d_{n,k+1} = d_{n,k} - \mathcal{M}(d_{n,k})$.
4) Incrementation of k and return to step 2 if $d_{n,k+1}$ is not designated as an IMF, else stop of the procedure.

Manuscript received February 17, 2010; revised July 16, 2010, October 29, 2010; accepted November 16, 2010. Date of publication December 06, 2010; date of current version February 09, 2011. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Ut-Va Koc. This work was supported by the National Research Agengy (ANR) of France by Grant mv-EMD BLAN07-0314-02.

The authors are with the INSERM, U642, Rennes F-35000, France. They are also with the Université de Rennes 1, LTSI, Rennes F-35000, France (e-mail: julien.fleureau@univ-rennes1.fr; jean-claude.nunes@univ-rennes1.fr; amar.kachenoura@univ-rennes1.fr; laurent.albera@univ-rennes1.fr; lotfi.senhadji@univ-rennes1.fr).

Color versions of one or more of the figures in this correspondence are available online at http://ieeexplore.ieee.org.
Digital Object Identifier 10.1109/TSP.2010.2097254