AN ORDER FITTING RULE FOR OPTIMAL SUBSPACE AVERAGING

I. Santamaria[1], L. L. Scharf[2], C. Peterson[2], M. Kirby[2], J. Francos[3],

[1]Dept. of Communications Engineering, University of Cantabria, Spain
[2]Dept. of Mathematics, Colorado State University, Fort Collins, CO, USA
[3] Dept. of Electrical and Computer Engineering, Ben-Gurion University, Israel

ABSTRACT

The problem of estimating a low-dimensional subspace from a collection of experimentally measured subspaces arises in many applications of statistical signal processing. In this paper we address this problem, and give a solution for the average subspace that minimizes an extrinsic mean-squared error, defined by the squared Frobenius norm between projection matrices. The solution automatically returns the dimension of the optimal average subspace, which is the novel result of the paper. The proposed order fitting rule is based on thresholding the eigenvalues of the average projection matrix, and thus it is free of penalty terms or other tuning parameters commonly used by other rank estimation techniques. Several numerical examples demonstrate the usefulness and applicability of the proposed criterion, showing how the dimension of the average subspace captures the variability of the measured subspaces.

Index Terms— Subspace signal processing, subspace averaging, order-fitting, extrinsic mean, Grassmann manifold, flag manifold.

1. INTRODUCTION

Given a subspace model for a signal, there are many ways to detect the signal [1], or estimate where it lies in the subspace [2], but when only experimental data is available, the question of extracting an average or prototype subspace model arises. Input data in many signal processing applications admit a subspace representation. Examples include detection and recognition of one-dimensional and multi-dimensional geometrically warped signals, where an invariant representation of the signal is shown to have the form of a subspace [3], or pattern recognition applications in which features obtained after a dimensionality reduction stage such as principal component analysis (PCA) are commonly used [4]. Also, in multi-antenna wireless communications systems, subspaces play a central role in problems such as spectrum sensing for cognitive radio [5], non-coherent communications [6], [7], and interference alignment [8], [9].

In this paper we address the following question, which is central to all these applications: "given a sequence of experimentally derived subspaces, how are these subspaces to be averaged and the dimension of the averaged subspace determined, from this collection of experimental subspaces?" Our key result is an order fitting rule that minimizes the mean-squared error between the projection matrix associated with the average subspace (an equivalent representation of a subspace in its Grassmanian manifold) and the projections associated with each of the experimental subspaces.

This mean-squared error may be considered an extrinsic or chordal mean-squared error associated with a chordal mean, to distinguish it from other intrinsic mean-squared errors such as the mean-squared error associated with the Riemannian center of mass [10], [11] (sometimes also named as Karcher mean [12]). This order fitting rule is determined only by our definition of mean-squared error. Consequently, it relies on no statistical model for the experimentally-derived subspaces, and it uses no penalizing term to control order.

1.1. Notation

In this paper we use $\langle \mathbf{A} \rangle$ to denote a subspace of $\mathcal{C}^n$ (a point in the complex Grassmann manifold), whereas $\mathbf{A}$ is used to denote a matrix whose columns form a unitary basis for that subspace. The superscripts $(\cdot)^T$ and $(\cdot)^H$ denote transpose and Hermitian, respectively. The trace and Frobenius norm of a matrix $\mathbf{B}$ will be denoted, respectively, as $\mathrm{tr}(\mathbf{B})$ and $\|\mathbf{B}\|_F$.

2. PROBLEM FORMULATION

Let us consider a collection of subspaces $\{\langle \mathbf{V}_m \rangle\}_{m=1}^M$ of $\mathcal{C}^n$, each with respective dimension $\dim(\langle \mathbf{V}_m \rangle) = q_m < n$. Each subspace $\langle \mathbf{V}_m \rangle$ is a point on the Grassmann manifold $\mathbb{G}(q_m, n)$, and the collection of subspaces lives on a disjoint union of Grasmannians. Let $D = \dim\left(\bigcup_{m=1}^M \langle \mathbf{V}_m \rangle\right) \leq n$ be the dimension of the union. In a pattern recognition problem, for instance, the measured or observation subspaces, $\{\langle \mathbf{V}_m \rangle\}_{m=1}^M$, might be noisy versions of a common object under different pose and illumination conditions, or other forms of elastic deformations [3]. As another example, in a noncoherent multiple-input multiple-output (MIMO) communications system the observed subspaces can be perturbed versions of a Grassmaniann constellation point received from M different transmitters [7], [6].

Let $\mathbf{V}_m \in \mathcal{C}^{n \times q_m}$ be a matrix whose columns form a unitary basis for $\langle \mathbf{V}_m \rangle$. Then $\mathbf{V}_m^H \mathbf{V}_m = \mathbf{I}_{q_m}$, and $\mathbf{P}_m = \mathbf{V}_m \mathbf{V}_m^H$ is the idempotent orthogonal projection onto $\langle \mathbf{V}_m \rangle$. Notice that $\mathbf{P}_m$ is a unique representation of $\langle \mathbf{V}_m \rangle$, whereas $\mathbf{V}_m$ is not unique, because if $\mathbf{G}$ is an arbitrary unitary $q_m \times q_m$ matrix, then $\mathbf{V}_m \mathbf{G}$ will be another representation of $\langle \mathbf{V}_m \rangle$ with orthonormal columns.

The problem we consider in this paper is to determine the dimension s of the subspace $\langle \mathbf{V}_s \rangle$ that "best approximates" the collection of subspaces $\{\langle \mathbf{V}_m \rangle\}_{m=1}^M$. More specifically, we aim to solve

The work of I. Santamaria was supported by the Spanish Government through grants PRX14/0028 (Estancias de Movilidad de Profesores, Ministerio de Educación) and by project RACHEL (TEC2013-47141-C4-3-R) funded by the Ministerio de Economa y Competitividad (MINECO). The work of L. L. Scharf was supported by the National Science Foundation (NSF) under grant CCF-1018472.

$$\left(s^*, \mathbf{V}_s^*\right) = \underset{\substack{s\in\{0,1,\ldots,D\}\\ \langle\mathbf{V}_s\rangle\in\mathbb{G}(s,n)}}{\operatorname{argmin}} \quad \frac{1}{M}\sum_{m=1}^{M} d\left(\langle\mathbf{V}_s\rangle, \langle\mathbf{V}_m\rangle\right)^2, \qquad (1)$$

where $d\left(\langle\mathbf{V}_s\rangle, \langle\mathbf{V}_m\rangle\right)$ is the extrinsic distance metric between the subspaces $\langle\mathbf{V}_s\rangle$ and $\langle\mathbf{V}_m\rangle$, which is given by the Frobenius norm of the difference between the respective projection matrices [13], [11]. That is

$$d\left(\langle\mathbf{V}_s\rangle, \langle\mathbf{V}_m\rangle\right)^2 = \left\|\mathbf{P}_s - \mathbf{P}_m\right\|_F^2, \qquad (2)$$

where $\mathbf{P}_s = \mathbf{V}_s\mathbf{V}_s^H$. There is motivation for this definition. Let $\{\mathbf{e}_i\}_{i=1}^n$ denote the standard basis for the ambient space $\mathcal{C}^n$. Then, the error in resolving $\mathbf{e}_i$ onto the subspace $\langle\mathbf{V}_s\rangle$ as opposed to the subspace $\langle\mathbf{V}_m\rangle$ is $(\mathbf{P}_s - \mathbf{P}_m)\mathbf{e}_i$, and the squared error computed over the basis $\{\mathbf{e}_i\}_{i=1}^n$ is

$$\sum_{i=1}^{n}\mathbf{e}_i^T(\mathbf{P}_s - \mathbf{P}_m)^H(\mathbf{P}_s - \mathbf{P}_m)\mathbf{e}_i = \operatorname{tr}\left[(\mathbf{P}_s - \mathbf{P}_m)^H(\mathbf{P}_s - \mathbf{P}_m)\right] = \|\mathbf{P}_s - \mathbf{P}_m\|_F^2. \qquad (3)$$

An argument in favor of this distance measure is that projections operate in the ambient space, and it is where they operate that we wish to measure error.

The extrinsic distance metric, as defined in (2), can also be computed in terms of the cosines of the principal angles between the two subspaces [14]; specifically, it is easy to show that

$$\left\|\mathbf{P}_s - \mathbf{P}_m\right\|_F^2 = q_m + s - 2\sum_{r=1}^{\min(q_m,s)}\cos(\theta_r)^2. \qquad (4)$$

Notice finally that $s = 0$ in (1) can be viewed as a null or noise-only hypothesis, meaning that the collection of measured subspaces does not contain a central or signal subspace.

3. SUBSPACE AVERAGING AND ORDER FITTING RULE

Using the extrinsic distance between subspaces, the proposed order estimation criterion becomes

$$\left(s^*, \mathbf{P}_s^*\right) = \underset{\substack{s\in\{0,1,\ldots,D\}\\ \mathbf{P}\in\mathcal{P}_s}}{\operatorname{argmin}} \quad \frac{1}{M}\sum_{m=1}^{M}\left\|\mathbf{P} - \mathbf{P}_m\right\|_F^2, \qquad (5)$$

where $\mathcal{P}_s$ denotes the set of all projection matrices of rank s.

Let us first consider the following inner optimization problem: given $s \in \{1,\ldots,D\}$, find the subspace $\langle\mathbf{V}_s\rangle$ (unequivocally determined by its projection matrix) that minimizes the following cost function

$$E(s) = \min_{\mathbf{P}\in\mathcal{P}_s} \quad \frac{1}{M}\sum_{m=1}^{M}\left\|\mathbf{P} - \mathbf{P}_m\right\|_F^2, \qquad (6)$$

It is now easy to show that (6) can be expanded as follows:

$$E(s) = \min_{\mathbf{P}\in\mathcal{P}_s} \quad \operatorname{tr}\left[(\mathbf{P} - \overline{\mathbf{P}})^H(\mathbf{P} - \overline{\mathbf{P}})^H + \overline{\mathbf{P}} - \overline{\mathbf{P}}^2\right]. \qquad (7)$$

Here the symmetric matrix

$$\overline{\mathbf{P}} = \frac{1}{M}\sum_{m=1}^{M}\mathbf{P}_m, \qquad (8)$$

is the average of the projection matrices[1]. In words, the inner problem amounts to finding a central or average subspace of dimension s given the set of input (observed) subspaces.

Now, discarding constant terms and writing the projection matrix as $\mathbf{P} = \mathbf{U}_s\mathbf{U}_s^H$, where $\mathbf{U}_s$ is a unitary $n \times s$ matrix, problem (7) can be rewritten as

$$\min_{\mathbf{U}_s\in\mathbb{S}(s,n)} \quad \operatorname{tr}\left[\left(\mathbf{U}_s\mathbf{U}_s^H - \overline{\mathbf{P}}\right)^H\left(\mathbf{U}_s\mathbf{U}_s^H - \overline{\mathbf{P}}\right)\right], \qquad (9)$$

where $\mathbb{S}(s,n)$ denotes the complex Stiefel manifold of orthonormal s-frames in $\mathcal{C}^n$. Writing the compact eigendecomposition of the average projection matrix in (8) as $\overline{\mathbf{P}} = \mathbf{F}\mathbf{K}\mathbf{F}^H$, where $\mathbf{K} = \operatorname{diag}(k_1,\ldots,k_D)$ with $1 \geq k_1 \geq k_2 \geq \ldots \geq k_D$[2], our optimization problem can alternatively be written as

$$\max_{\mathbf{U}_s\in\mathbb{S}(s,n)} \quad \operatorname{tr}\left[\mathbf{U}_s^H\mathbf{F}\mathbf{K}\mathbf{F}^H\mathbf{U}_s\right]. \qquad (10)$$

It is known that the solution of (10) is given by any unitary matrix whose column space is the same as the subspace spanned by the s principal eigenvectors of $\mathbf{F}$ [15], [16], i.e.,

$$\mathbf{U}_s^* = (\mathbf{f}_1, \mathbf{f}_2, \ldots, \mathbf{f}_s) = \mathbf{F}_s. \qquad (11)$$

The nested sequence of optimal subspaces, $\langle\mathbf{U}_1^*\rangle \subset \langle\mathbf{U}_2^*\rangle \subset \cdots \subset \langle\mathbf{U}_D^*\rangle$, is a flag of means as defined in [17].

Plugging the solution for the optimal subspace of dimension s in (6), the minimum mean squared error (MSE) $E(s)$ takes the value

$$E(s) = \sum_{i=1}^{s}(1 - k_i) + \sum_{i=s+1}^{D} k_i. \qquad (12)$$

Interestingly, the MSE decomposition in (12) admits a bias-variance tradeoff interpretation in which the first term is the variance due to the selected dimensions of $\overline{\mathbf{P}}$, whereas the second term is a squared-bias cost associated with the discarded dimensions.

The proposed subspace order fitting rule then becomes

$$s^* = \underset{s\in\{0,1,\ldots,D\}}{\operatorname{argmin}} \quad \sum_{i=1}^{s}(1 - k_i) + \sum_{s+1}^{D} k_i. \qquad (13)$$

A simple analysis reveals that $E(s+1) \leq E(s)$ if $k_{s+1} \geq \frac{1}{2}$, and therefore the fitting rule amounts to selecting the largest s such that $k_s > \frac{1}{2}$. Recently, we found out that a similar rule was developed for the problem of designing optimum time-frequency (TF) subspaces with a specified TF pass region [18].

According to the proposed criterion, the eigenvectors of the average projection matrix whose eigenvalues are above the threshold $\eta = 1/2$ determine the *signal subspace*. Unlike other rank estimation approaches in the literature, this optimality criterion is free of tuning parameters and does not rely on any statistical model for the generated data.

Why should the selected order not be the full order D? The answer is that the average $\overline{\mathbf{P}}$ is not a projection, and to force its approximation with a projection $\mathbf{P}$ is to risk the fitting of subdominant modes of $\overline{\mathbf{P}}$, with small eigenvalues, with modes of $\mathbf{P}$ which are

[1] Notice, however, that the average of projection matrices is not a projection matrix, and therefore $\overline{\mathbf{P}}$ is not idempotent.

[2] In practice, the measured subspaces will contain noise and therefore the eigenvalues will be all distinct with probability one. Also, notice that since $\overline{\mathbf{P}}$ is the mean of M projection matrices, its eigenvalues are less than or equal to 1, with equality iff all the projection matrices are identical.

forced to have unit eigenvalues. The order fitting rule automatically protects against this.

Remark 1: The term $v = \mathrm{tr}\left(\overline{\mathbf{P}} - \overline{\mathbf{P}}^2\right)$ in the cost function (7), measures how far the mean projection matrix $\overline{\mathbf{P}}$ is from a true, idempotent, projection matrix. In terms of the eigenvalues, k_i, this term can be written as

$$v = \sum_{i=1}^{n} k_i(1 - k_i), \tag{14}$$

which is a measure of variability in the eigenvalues of $\overline{\mathbf{P}}$ or, in other words, a measure of how spread out are the measured subspaces. In fact, if all subspaces were identical, then $\overline{\mathbf{P}}$ would be idempotent and (14) would be zero. On the other hand, when the subspaces are far apart from each other, v increases. This aspect will be further analyzed by means of numerical simulations in the next section.

4. SIMULATION RESULTS

In this section we evaluate the performance of the proposed order fitting rule by means of some numerical examples. In the first example, we generate M perturbed versions of a central subspace $\langle \mathbf{V}_c \rangle \in \mathbb{G}(k, n)$, as follows: we first generate

$$\mathbf{G}_m = \left[\mathbf{V}_c \mid \mathbf{0}_{n\times(n-k)}\right] + \sigma \mathbf{Z}_m, \qquad m = 1, \ldots, M \tag{15}$$

where $\mathbf{V}_c \in \mathcal{C}^{n\times k}$ is a matrix whose columns form an orthonormal basis for a central subspace $\langle \mathbf{V}_c \rangle$, $\mathbf{0}_{n\times(n-k)}$ is an $n \times (n-k)$ zero matrix, and $\mathbf{Z}_m \in \mathcal{C}^{n\times n}$ is a matrix whose entries are independent and identically distributed complex Gaussian random variables with zero mean and variance $1/n$. The value of σ determines the signal-to-noise-ratio, which is defined for this example as $\mathrm{SNR} = 10\log_{10}\left(\frac{k}{n\sigma^2}\right)$. An orthogonal basis for the m-th subspace, $\mathbf{V}_m$, is then constructed from the first k orthonormal vectors of the QR decomposition of $\mathbf{G}_m$. Notice also that for this example all subspaces in the collection have exactly the same dimension.

Fig. 1 shows the estimated order as a function of the SNR for different values of (k, n) and a total number of $M = 200$ subspaces. The curves represent averaged results of 500 independent simulations. As we can see, there is phase-transition behavior between $s^* = 0$ (no central subspace) and the right order $s^* = k$. This phase-transition behavior might be related to some concentration-of-measure phenomenon in the eigenvalues of the average projection matrix, and deserves further theoretical study[3].

Fig. 2 represents the estimated probability density function (pdf) of the random variable v/n with v defined as in (14), which gives us a measure of how clustered the subspaces are. For this example our data set is a collection of $M = 20$ subspaces with $(k, n) = (3, 30)$ and different SNRs. We observe that for low SNR values the subspaces are more spread out and, consequently, v/n takes higher values. Interestingly, this random variable is also sharply concentrated around its mean value.

In the second example, we evaluate the performance of the proposed order fitting rule when the input subspaces have different dimensions. More precisely, each measured subspace in this example is generated by extracting the first r columns of the DFT matrix, where r is a discrete uniform random variable taking values in the set $r \in \{1, ..., R\}$. Fig. 3 shows the estimated pdf of the optimal order s^* when the dimension of the ambient space is $n = 100$, $R = 30$, and the number of input subspaces varies from $M = 10$ to $M = 200$.

[3]Similar phase-transition phenomena have been reported for the classical problem of estimating the rank of a sample covariance matrix in the asymptotic regime [19], [20].

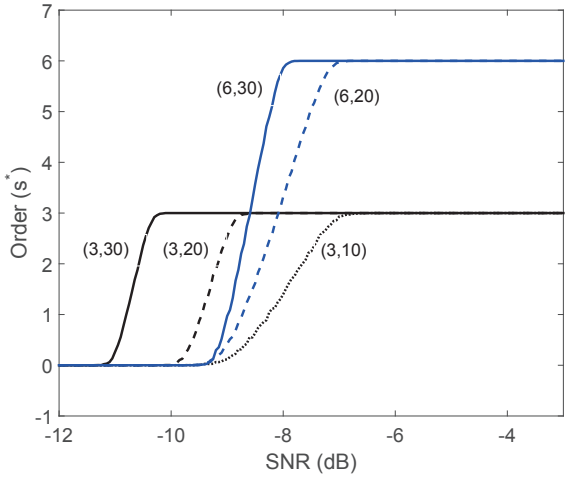

Fig. 1. Estimated order as a function of the SNR for different values of (k, n). In all examples the number of measured subspaces is $M = 200$.

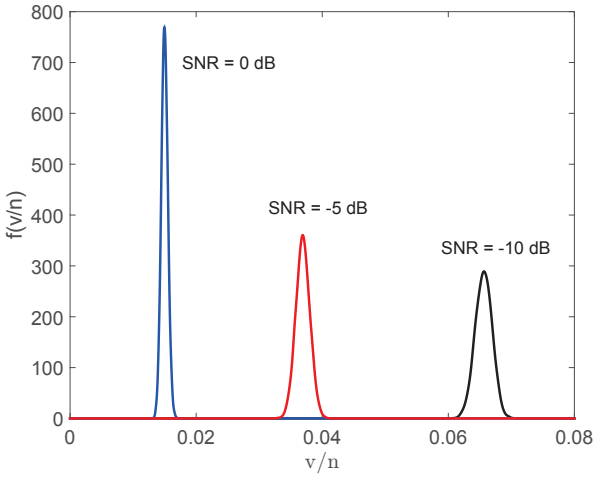

Fig. 2. Probability density function for the random variable $v/n = \frac{1}{n}\sum_{i=1}^{n} k_i(1 - k_i)$ when $(k, n) = (3, 30)$ for different values of the SNR and $M = 20$ subspaces.

Although the estimated order is obviously a discrete variable, for representation purposes we have estimated its pdf using the Parzen windowing method and hence a continuous density results. As expected, the estimated pdf shrinks around the mean value $s^* = 15$ as the number of input subspaces increases.

5. CONCLUSIONS

In this paper we have solved a subspace averaging problem that minimizes an extrinsic mean-squared error between the experimental subspace projections and the projection onto an extrinsic mean of subspaces. The minimization automatically returns an order-fitting

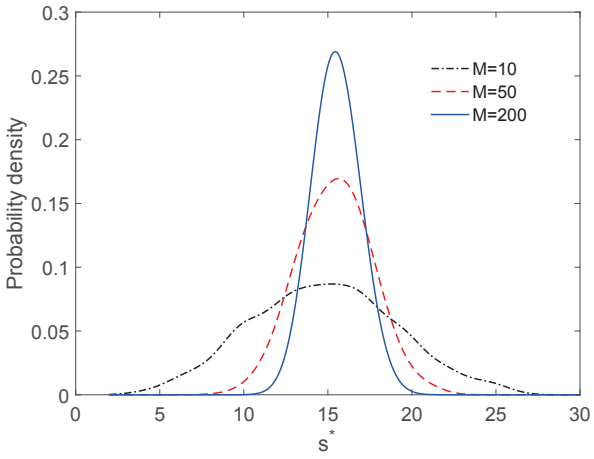

Fig. 3. Estimated pdf of s^* with subspaces of dimensions uniformly distributed on the interval $r \in \{1, ..., 30\}$, for $M = 10$, $M = 50$ and $M = 200$ subspaces. The ambient dimension is $n = 100$.

rule for the dimension of the extrinsic subspace mean. The rule is simple: when the eigenvalues of the average of experimental projections are smaller than $1/2$, the extrinsic mean of subspaces discards eigenvectors of the average, and uses only those eigenvectors for which eigenvalues of the average are greater than $1/2$ to construct its projection. In fact, the extrinsic mean of subspaces is constructed by quantizing the eigenvalues of the average projections at 0 or 1, around the slicing level $1/2$, and building its projection from these quantized eigenvalues. The order fitting rule follows directly from the definition of extrinsic mean-squared error, with no dependence on a statistical model for the data or a penalty term that would control for order.

As future work, we will consider extensions of the proposed rule to subspace clustering problems [21], where the goal is to simultaneously find the average subspaces, and determine their orders, that best fit each cluster of experimental subspaces. Also, we will consider applications in the context of multi-antenna communications and subspace tracking algorithms.

6. REFERENCES

[1] L. L. Scharf, B. Friedlander, "Matched subspace detectors," *IEEE Trans. Signal Process.*, vol. 42, no. 8, pp. 2146-2157, Aug. 1994.

[2] Y. C. Eldar, T. G. Dvorkind, "A minimum squared-error framework for generalized sampling," *IEEE Trans. Signal Process.*, vol. 54, no. 6, pp. 2155-2167, Jun. 2006.

[3] R. R. Hagege, J. M. Francos, "Universal manifold embedding for geometric deformations estimation," *IEEE Int. Symp. on Information Theory*, San Petersburg, Russia, pp. 1589-1593, July, 2011.

[4] C. J. C. Burges, "Dimension reduction: a guided tour," *Foundations and Trends in Machine Learning*, vol. 2, no. 4, pp. 275-365, 2009.

[5] D. Ramirez, G. Vazquez-Vilar, R. Lopez-Valcarce, J. Via, I. Santamaria, "Detection of rank-P signals in cognitive radio networks with uncalibrated multiple antennas," *IEEE Trans. Signal Process.*, vol. 59, no. 8, pp. 3764-3774, Aug. 2011.

[6] R. H. Gohary, T. N. Davidson, "Noncoherent MIMO communication: Grassmannian constellations and efficient detection," *IEEE Trans. Inf. Theory*, vol. 55, no. 3, pp. 1176-1205, March 2009.

[7] L. Zheng, D. N. C. Tse, "Communication on the Grassmann manifold: A geometric approach to the noncoherent multiple-antenna channel," *IEEE Trans. Inf. Theory*, vol. 48, no. 2, pp. 359-383, Feb. 2002.

[8] K. S. Gomadam, V. R. Cadambe, S. A. Jafar, "A distributed numerical approach to interference alignment and applications to wireless interference networks," *IEEE Trans. on Inf. Theory,* vol. 57, no. 6, pp. 3309-3322, Jun. 2011.

[9] I. Santamaria, O. Gonzalez, R. W. Heath, S. W. Peters, "Maximum sum-rate interference alignment algorithms for MIMO channels," *IEEE Global Communications Conference*, GLOBECOM, Miami, Fl, Dec. 2010.

[10] H. Karcher, "Riemannian center of mass and mollifier smoothing," *Comm. Pure Appl. Math.* vol. 5, no. 30, pp. 509-541, 1977.

[11] T. Marrinan, J. R. Beveridge, B. Draper, M. Kirby, C. Peterson, "Finding the subspace mean or median to fit your need," in *Proc. of Computer Vision and Pattern Recognition* (CVPR), 2014.

[12] H. Karcher, "Riemannian center of mass and the so called Karcher mean," Available in Arxiv:1407.2087v1, 2014.

[13] A. Srivastava, E. Klassen, "Monte Carlo extrinsic estimators of manifold-valued parameters," *IEEE Trans. Signal Process.*, vol. 50, no. 2, pp. 299-308, Aug. 2002.

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

[15] J. H. Manton, "Optimization algorithms exploiting unitary constraints," *IEEE Trans. Signal Process.*, vol. 50, no. 3, pp. 635-650, March 2002.

[16] S. Shahbazpanahi, A. B. Gershman, J. H. Manton, "Closed-form blind MIMO channel estimation for orthogonal spacetime block codes," *IEEE Trans. Signal Process.*, vol. 53, no. 12, pp. 4506-4517, Dec. 2005.

[17] B. Draper, M. Kirby, J. Marks, T. Marrinan, C. Peterson, "A flag representation for finite collections of subspaces of mixed dimensions," *Linear Algebra and its Applications*, vol. 451, pp. 15-32, 2014.

[18] F. Hlawatsch, W. Kozek, "Time-frequency projection filters and time-frequency signal expansions," *IEEE Trans. Signal Process.*, vol. 42, no. 12, pp. 3321-3334, Dec. 1994.

[19] P. O. Perry, P. J. Wolfe, "Minimax rank estimation for subspace tracking," *IEEE J. Sel. Topics Signal Process.*, vol. 4, no. 3, pp. 504-513, June 2010.

[20] R. R. Nadakuditi, A. Edelman, "Sample eigenvalue based detection of high-dimensional signals in white noise using relatively few samples," *IEEE Trans. Signal Process.*, vol. 56, no. 7, pp. 2625-2638, July 2008.

[21] R. Vidal, "Subspace clustering," *IEEE Signal Process. Magazine*, vol.28 , no. 2, pp. 52-68, July 2011.