MULTI-STREAM ITERATIVE SVD FOR MASSIVE MIMO COMMUNICATION SYSTEMS UNDER TIME VARYING CHANNELS

Junting Chen and Vincent K. N. Lau

Department of Electronic and Computer Engineering
The Hong Kong University of Science and Technology, Hong Kong

ABSTRACT

Singular value decomposition (SVD) plays an important role in signal processing for multi-input multi-output (MIMO) communication systems. Under massive MIMO scenarios, as the channel matrix is very large, implementing SVD at every frame is highly inefficient. Existing literature on iterative SVD algorithms are mostly heuristic based, and the associated tracking performance under time-varying channels is not clear. The difficulties of deriving and analyzing SVD algorithms are due to the non-convexity of the associated optimization problem and the time-varying nature of the MIMO channel. In this paper, we formulate the problem on Grassmann manifolds and derive a multi-stream iterative SVD algorithm using optimization techniques. To enhance the tracking performance under time-varying channels, we propose a compensation algorithm to offset the motion of the time-varying target eigenspace. We analyze the convergence behavior of the proposed algorithm, where we show that under some mild conditions, the proposed iterative SVD algorithm with compensations has zero tracking error, despite the underlying problem being non-convex and the channel being time-varying. The complexity of the algorithm is only $\mathcal{O}(n^2 p)$ for estimating p singular vectors, compared with $\mathcal{O}(n^3)$ for the SVD of a $n \times n$ channel matrix.

***Index Terms*—** SVD, Iterative algorithm, Convergence analysis, Grassmann manifold, Optimization

1. INTRODUCTION

1.1. Motivations

Singular value decomposition (SVD) is a key technology in multi-input multi-output (MIMO) systems. Applying SVD to a MIMO channel matrix, parallel eigen-subchannels can be created for multiple data stream transmission. There are many applications of SVD in MIMO systems. For example, a MIMO-SVD joint processing on the transmitter and receiver in a multi-user MIMO (MU-MIMO) system was studied in [1]. The author in [2] proposed a SVD-based beamforming strategy under noisy channel state information (CSI), where SVD is used to create parallel eigen-subchannels from the MIMO channel.

The computational complexity of SVD for a $n \times n$ matrix is $\mathcal{O}(n^3)$ [3]. While this is acceptable in current MIMO systems with small channel matrices [1,2,4,5], the complexity may be too high for massive MIMO systems [6–8], where the number of antennas n is typically very large. Since the MIMO channel is temporal correlated, it is not efficient to compute a new SVD for the large channel matrix once every transmission frame. Instead, it is desirable to utilize the channel temporal correlation and derive recursive algorithms to reduce the complexity. One systematic approach is to reverse engineer an optimization objective so that the associated optimizer is the desired eigenspace of the channel matrix. However, this is very challenging, because the underlying optimization problem is *non-convex* and there might be multiple *non-isolated* local optima that affect the convergence of the algorithm. The more challenging problem is, as the channel matrix is time-varying, the optimal *target* eigenvectors are time-varying as well, and therefore, the algorithm needs to track the solution to a series of time-varying optimization problems.

In this paper, we propose an iterative SVD algorithm to track a small number of active eigen-subchannels in a time-varying massive MIMO system. To deal with the non-isolated local optima problem, we adopt an optimization based approach over *Grassmann manifold* [9, 10] and derive the iterative SVD algorithm by extending the gradient over a manifold [9, 11]. To enhance the tracking of the time-varying eigen-subchannels, we propose a compensation algorithm by compensating the motion of the target eigen-subchannel. We show in our analysis that, despite the underlying problem being non-convex, the proposed algorithms converge to the global optimal solution almost surely when the channel matrix is static. In time-varying channels, the tracking error of the continuous-time dynamics for the proposed compensation algorithm convergences to zero almost surely under some mild conditions. We also demonstrate in the simulation that the proposed algorithms have superior tracking capability over the baseline schemes and only have a complexity of $\mathcal{O}(n_t^2)$, where n_t is the number of transmitter antennas.

1.2. Relation to Prior Work

There are a number of iterative SVD algorithms developed in the past, most of which can be classified into four families: methods based on QR factorization or Jacobi rotation, power methods, Krylov subspace methods and gradient based methods. Specifically, however, the incremental Jacobi methods [12] are not attractive in large systems due to their high computational complexity. The power methods, such as power iteration [13], subspace iteration [14] and Rayleigh-Ritz method [15], have a fast convergence speed for a static matrix, but they perform poorly when the matrix is time-varying. The Krylov subspace method, such as Lanczos method [16] and fast subspace decomposition [17], either have poor tracking performance under time-varying channels or require a high complexity to update the Krylov subspace at each time slot for the time-varying matrix. On the other hand, the existing gradient type algorithms, e.g., PAST-like algorithms [18], gradient descent with deflations [19] and gradient flows on manifolds [9,20], still did not fully exploit the time variation property of the channel to enhance their eigen-subspace tracking performance. In general, although the traditional gradient-like

This work was supported by RGC 614913.

algorithms are suitable for the tracking in noisy time-varying channels, their convergence speed is slow. Moreover, in these existing works, there is no analytical convergence analysis for the tracking under time-varying channels.

2. SYSTEM MODEL

Consider a point-to-point massive MIMO wireless communication system, where the transmitter has n_t antennas and the receiver has n_r antennas. Both n_t and n_r are assumed to be large, i.e., at the order of 100. The received signal at the receiver is modeled as $\mathbf{y} = H\mathbf{s} + \mathbf{z}$, where $H \in \mathbb{C}^{n_r \times n_t}$ is the MIMO channel, $\mathbf{s} \in \mathbb{C}^{n_t}$ is the transmit signal, $\mathbf{y} \in \mathbb{C}^{n_r}$ is the received signal and $\mathbf{z} \sim \mathcal{CN}(\mathbf{0}, I_{n_r})$ is the additive complex Gaussian noise.

We assume that there is rich local scattering at the receiver. On the other hand, since the transmitter is usually installed on the roof of a building or on a tower, when it equips with a large number of antennas, there is usually not enough local scattering around the transmitter. As a result, the angular spread (AS) of the transmit signal is narrow. This phenomenon is captured by the widely used one-ring local scattering model [21, 22] for massive MIMO systems. We adopt the one-ring model in [21, 22], and the MIMO channel matrix H is rank deficient, i.e., rank$(H) \ll \min(n_r, n_t)$. Moreover, the time-varying channel $H(t)$ is assumed to be ergodic stationary.

Since there are only a few active eigen-modes for the channel H, we only deliver p data streams ($p <$ rank$(H) \ll \min(n_r, n_t)$) from the transmitter to the receiver. Suppose we know the SVD of the channel H and write $H = U\Sigma V^\dagger$, where Σ is a $n_r \times n_t$ matrix with only non-zero main diagonal elements sorted in descending order as the singular values of H. Then the transmitter can apply the first p columns of V as the precoder $\overline{V} \in \mathbb{C}^{n_t \times p}$, and the receiver uses the first p columns of U as the decorrelator. The MIMO channel is then transformed to p parallel eigen-subchannels. We assume the receiver has perfect CSI knowledge, and at each time slot, it computes the precoder $\overline{V} \in \mathbb{C}^{n_t \times p}$ and feeds back to the transmitter.

The above precoding strategy is widely used in both the industry and the literature, but it requires $\mathcal{O}(n_t^3)$ arithmetic operations as SVD is involved. There are some complexity-reduced recursive SVD algorithms in the literature [12–20], but since the performance of a massive MIMO system is very sensitive to the SVD convergence errors, these algorithms may not have satisfactory tracking performance under time-varying massive MIMO channels.

3. MULTI-STREAM ITERATIVE SVD ALGORITHM DESIGN FOR TIME-VARYING CHANNELS

We note that finding the first p right-singular vectors of H is equivalent to finding the eigenvectors of $A = H^\dagger H$ corresponding to the p largest eigenvalues. As a result, the desired precoder $\overline{V}$ corresponds to the solution to the following problem

$$\max_{X \in \mathbb{C}^{n_t \times p}} \overline{f}(X) = \text{tr}[X^\dagger A(t) X], \qquad \text{subject to } X^\dagger X = I_p \quad (1)$$

However, when $p > 1$, the above problem becomes very hard to solve, because it is non-convex due to the unitary constraint. Moreover, the optimum X_* is not unique and non-isolated, since the rotated solution $X_* M$ gives another optimum under any unitary matrix $M \in \mathbb{C}^{p \times p}$.

3.1. Problem Transformation on Grassmann Manifold

To resolve this problem, notice that all the semi-orthogonal matrices that span the same subspace yield the same value in the objective function $\overline{f}(X)$. Therefore, we should quotient out such equivalence and treat all the matrices X that span the same column space as the same element to optimize. Such observation leads to the notion of the Grassmann manifold. Specifically, a *Grassmann manifold* Grass(p, n_t) is the set of all p-dimensional subspaces: Grass$(p, n_t) = \{$span$(X) : X \in \mathbb{C}^{n_t \times p}, $rank$(X) = p\}$, where span$(X)$ denotes the column space spanned by X.

Let f : Grass$(p, n_t) \mapsto \mathbb{R}$: span$(X) \mapsto$ tr$((X^\dagger X)^{-1} X^\dagger A X)$. Then precoder optimization problem (1) can be reformulated into the eigenspace optimization on the Grassmann manifold

$$\max_{\text{span}(X) \in \text{Grass}(p, n_t)} f(X) = \text{tr}[(X^\dagger X)^{-1} X^\dagger A(t) X] \quad (2)$$

where the optimization variable is any p-dimensional subspace in $\mathbb{C}^{n_t \times p}$.

Note that, if A has distinct p-th and $(p+1)$-th eigenvalues $\lambda_p \neq \lambda_{p+1}$, the above problem (2) has a unique global optimal solution span(X_*) on the Grassmann manifold [23]. For notation convenience, we consider the matrix X to represent the subspace span$(X) \in$ Grass(p, n_t), where, without loss of generality, X is considered to be the matrix that satisfies $X^\dagger X = I_p$ and $X^\dagger A X = \Phi$, diagonal[1].

3.2. Eigenspace Tracking Algorithm with Compensations

An intuitive way to solve the problem (2) is to generalize the gradient algorithm to the Grassmann manifold. Using the calculus on Grassmann manifold [10, 11], the gradient of (2) is given by

$$\nabla f(X) := AX - X(X^\dagger X)^{-1} X^\dagger A X. \quad (3)$$

As a result, the *gradient algorithm* on Grassmann manifold is given by

$$X_{n+1} = X_n + \gamma_n \nabla f(X_n). \quad (4)$$

In general, when the parameter A is static, gradient-based algorithms can converge to a local optimal point under a good choice of step size sequence γ_n. However, under the time-varying parameter $A(t)$, a constant step size $\gamma_n \equiv \gamma$ (that violates the convergence-guaranteed step size rule) may be used for the tracking. In addition, since the optimal solution $X_*(t)$ is now also time-varying, there is always a convergence gap between X_n and $X_*(t_n)$.

Intuitively, to reduce the tracking error gap, one would like to estimate the motion of the moving target $X_*(t_n)$ and *compensate* to it:

$$X_{n+1} = X_n + \gamma \nabla f(X_n; t_n) + \widehat{\triangle X_{*,n}} \quad (5)$$

where $\widehat{\triangle X_{*,n}}$ is a *compensation* for the moving target $X_*(t_{n+1}) - X_*(t_n)$.

However, it is challenging to obtain $X_*(t_{n+1}) - X_*(t_n)$, because it requires the knowledge of the optimal solution $X_*(t_n)$. To tackle this problem, consider the *optimality condition* [24] $g(X; A) \triangleq \nabla f(X) = \mathbf{0}$ at the optimal point span$(X_*) \in$ Grass(p, n_t). In this implicit function, taking the differentiation on $g(\cdot)$ with respect to (w.r.t.) t, we obtain

$$h(dX_*; X_*, A) + g(X_*, dA) = \mathbf{0} \quad (6)$$

[1] Under such specification, there are still multiple X to represent the same subspace.

where $g(X_*, dA) = g(X_*, A(t+dt)) - g(X_*, A(t))$ is the partial differential of $g(X_*; A)$ on A, and $h(dX_*; X_* A)$ is the partial differential of $g(X; A)$ on span(X_*) along the direction dX_*, i.e., the Hessian of the objective function $f(X)$ [9,20]. Note that, under the assumption $\lambda_p \neq \lambda_{p+1}$, the unique global optimum span(X_*) is non-degenerate and the function $g(X; A) = \mathbf{0}$ has a unique solution in the neighborhood of span(X_*). Moreover, the Hessian $h(\xi; X_* A)$ is linear in ξ. Then by the *implicit function theorem* [10], there exists a unique solution $dX_* = \xi$ to the equation (6).

Suppose that the iterate X_n is a good approximation of X_*. We can thus obtain an estimation of the compensation $\widehat{\triangle X_{*,n}} := \hat{\xi}$ by solving the following equation over $\hat{\xi}$

$$h(\hat{\xi}; X_n, A(t_n)) + g(X_n, dA(t_n)) = \mathbf{0}. \tag{7}$$

3.3. Low Complexity Implementation

Using the property of Grassmann manifold, the compensation equation (7) can be solved efficiently. From the results in [20, 23], the Hessian term in (7) is given by

$$h(\hat{\xi}; X, A) = (I_{n_t} - XX^\dagger)\left[A\hat{\xi} - \hat{\xi}(X^\dagger X)^{-1} X^\dagger A X\right]. \tag{8}$$

As X represents a subspace in the Grassmann manifold, we can simply consider X to be semi-orthonormal and $X^\dagger A X = \Phi$, a diagonal matrix. This is because, we can always take $X^{'} = XM$ to span the same column subspace as X, where M is obtained from $X^\dagger A X = M\Phi M^\dagger$ and Φ is a diagonal matrix with diagonal elements $\beta_1, \ldots, \beta_p$. As a result, the equation (7) can be decomposed into p parallel linear matrix equations,

$$(I_{n_t} - XX^\dagger)\left[A - \beta_i I_{n_t}\right]\hat{\xi}^i + g(X, dA)^i = \mathbf{0} \tag{9}$$

where $\hat{\xi}^i$ and $g(X, dA)^i$ are the i-th ($1 \leq i \leq p$) columns of $\hat{\xi}$ and $g(X, dA)$, respectively.

The linear equations (9) can be solved efficiently by *conjugate gradient* (CG) algorithm, which has a fast convergence speed and low complexity. In fact, as $g(X, dA)$ is small and X_n is close to $X^*(t_n)$, computing only one step of the CG algorithm would be enough to obtain a first order approximation for $X_*(t_{n+1}) - X_*(t_n)$.

Furthermore, to make the algorithm practical, it is desirable to choose a good representative X_n for the subspace span(X_n) from each iteration. One strategy is to apply the Gram-Schmidt procedure $q(\cdot)$ on each iteration to guarantee $X_n^\dagger X_n = I_p$. On the other hand, we want X_n to diagonalize A, which can be easily realized by taking $X_{n+1}^{'} := X_{n+1}M$, where M is a unitary matrix given by the diagonalization of the $p \times p$ matrix $X_{n+1}^\dagger A X_{n+1} = M\Phi M^\dagger$.

4. TRACKING PERFORMANCE ANALYSIS FOR MULTI-STREAM ITERATIVE SVD

In this section, we analyze the performance of the iterative SVD algorithms derived in the previous section. Existing literature only focuses on the convergence under static channel. However, we also need to understand the tracking performance under time-varying massive MIMO channels. We introduce the convergence result in the following.

4.1. Convergence under Static Channel H

When the channel H is static, the compensation term is zero and the proposed compensation algorithm (5) degenerates to the gradient algorithm (4) on the Grassmann manifold. We can derive the following convergence result.

Theorem 1 (Global convergence under static CSI)**.** *Consider that the initial solution X_0 is random and the step size rule in (4) satisfies $\sum_n \gamma_n = \infty$ and $\sum_n \gamma_n^2 < \infty$. Then the iterate X_n in algorithm (4) converges to X_* almost surely.*

This result has been well established in the literature and the associated analysis can be found in [9]. The result shows that, despite the original eigen precoding problem (1) being non-convex, the algorithm on the Grassmann manifold still converges to the global optimal solution. However, we are most interested in the case when $H(t)$ is time-varying.

4.2. Convergence under Time-varying Channel $H(t)$

We first study the tracking error of the conventional gradient algorithm (4) by dropping the compensation term in (5).

Denote $\triangle\lambda_p(\iota) = \lambda_p(\iota) - \lambda_{p+1}(\iota)$ as the eigenvalue gap of $A(\iota)$ for the p-th and $(p+1)$-th eigenvalues, $\nu(\iota) \triangleq \|dA(\iota)\|_2/d\iota$ as the channel variation speed, $\overline{\nu} = \sup_{\iota<T}\nu(\iota)$, and τ as the frame duration. We have the following result.

Theorem 2 (Tracking error for the gradient algorithm)**.** *Suppose the eigenvalue gap satisfies $\triangle\lambda_p(\iota) \geq \lambda_0 > 0$, for $\iota \in [0, T)$ w.p.1. Then, for $0 < \epsilon < \lambda_0$, there exists $\delta(\epsilon) > 0$, such that if the channel variation speed $\overline{\nu} < \frac{\gamma}{p\tau}\lambda_0(\lambda_0 - \epsilon)\delta$ and the initial tracking error $\|X_0 - X_*(0)\|_F < \delta$, then the average mean square tracking error is bounded by*

$$\frac{1}{T/\tau}\sum_{n=1}^{T/\tau}\mathbb{E}\|X_n - X_*(t_n)\|_F^2 \leq \frac{\overline{\nu}^2\tau^2p^2}{\gamma^2\alpha}\phi + \mathcal{O}(\gamma + \frac{1}{T})$$

where $\alpha = \sup_{\lambda_0 \leq c < \infty}\{(\lambda_0 - \epsilon)F_{\triangle|\lambda_0}(c) + (c-\epsilon)\left(1 - F_{\triangle|\lambda_0}(c)\right)\}$, $\phi = \mathbb{E}\left[(\triangle\lambda_p - \epsilon)^{-1}(\triangle\lambda_p)^{-2}\right]$, and $F_{\triangle|\lambda_0}(x) \triangleq \Pr(\triangle\lambda_p \leq x|\triangle\lambda_p > \lambda_0)$ is the distribution function of the eigenvalue gap $\triangle\lambda_p$.

Proof. Please refer to [25] for the proof. □

The above result shows that the channel variation speed $\overline{\nu}$ and the channel singular value gap distribution (captured by α) have a significant impact on the tracking error of the gradient algorithm. Small $\overline{\nu}$ and large α are both desired for maintaining small tracking errors. With this observation, one can reduce the tracking error by increasing α. One strategy is to use more antennas, so that there is a higher chance to have a large singular value gap $\triangle\lambda_p > \lambda_0$ for the particular p-th and $(p+1)$-th singular values [26], which results in a larger α.

We now consider the tracking performance of the compensation algorithm (5). From (5), consider the equivalent *continuous-time algorithm dynamics* $X^c(t)$ as the solution to

$$dX^c = g(X^c; A(t))dt + \widehat{dX_*}, \quad X^c(0) = X_0 \tag{10}$$

where $\widehat{dX_*}(t)$ satisfies the dynamic equation $h(\widehat{dX_*}; X^c, A) + g(X^c, dA) = \mathbf{0}$ for all $t > 0$. We have the following convergence result.

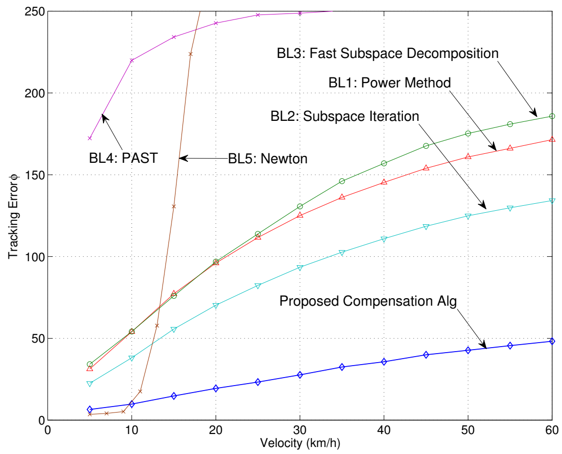

Fig. 1. Aggregate angle difference versus the terminal mobility.

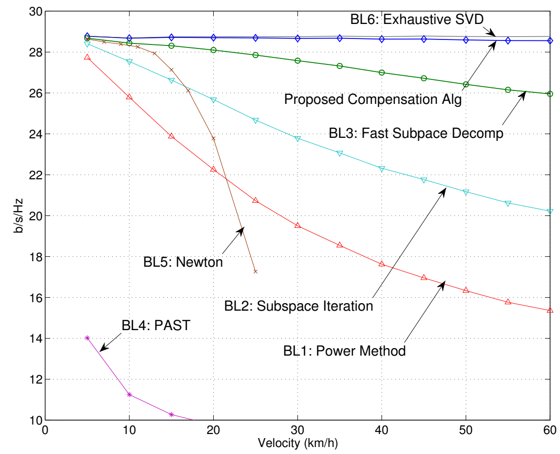

Fig. 2. Achievable data rate versus the terminal mobility under 10 dB SNR.

Theorem 3 (Convergence of the compensation algorithm)**.** *Suppose the largest singular value* σ_1 *of* $H(t)$ *is bounded w.p.1. Then there exists a* $\epsilon > 0$*, such that for* $\|X^c(0) - X_*(0)\|_F < \epsilon$*, we have* $\|X^c(t) - X_*(t)\|_F \to 0$*, as* $t \to \infty$ *almost surely.*

Proof. Please refer to [25] for the proof. □

The compensation algorithm in (5) can be viewed as a discretization from the continuous-time algorithm flow $X^c(t)$ in (10). The above result provides a close insight on the superior tracking capability of the proposed multi-stream iterative SVD algorithm with compensations. It implies that, despite the eigen precoding optimization problem (1) being non-convex and the massive MIMO channel being time-varying, by choosing a proper initial state, it is possible to achieve a sufficiently small tracking error from the proposed compensation algorithm in (5)[2].

5. NUMERICAL RESULTS

We consider a point-to-point MIMO wireless communication channel with $n_t = 40$ transmit antennas, $n_r = 8$ receive antennas and $p = 4$ data streams to deliver. In the simulation, we follow a similar model in LTE standard [28] to specify the spatial correlation of the MIMO channel. The temporal correlation is modeled by the widely used autoregressive (AR) model [29] given by $H_n = \theta H_{n-1} + \sqrt{1-\theta^2} W$, with $\theta = J_0(2\pi f_d \tau)$, where $J_0(\centerdot)$ is the zero-th order Bessel function, f_d is the maximum Doppler frequency, $\tau = 1$ ms is the frame duration, and W is a zero mean complex Gaussian random matrix with covariance specified by the channel spatial correlation. Water-filling power allocation is applied. The step size of the proposed tracking algorithm is $\gamma = 0.5$. The simulation is run for over $T = 100$ s.

The performance of the proposed compensation algorithm is compared with the following baselines (BL). **BL1**: Power iteration with deflations [19, 30], **BL2**: Subspace iteration with the Rayleigh-Ritz method [15], **BL3**: Fast subspace decomposition [17] (implemented in an iterative way with similar per frame (i.e., time slot) complexity), **BL4**: PAST algorithm [18], **BL5**: Newton algorithm on the Grassmann manifold [20], and **BL6**: Exhaustive SVD, where a complete SVD is computed for $H(t)$ at every time slot.

Fig. 1 shows the tracking error versus the terminal mobility. The tracking error ϕ is defined as $\phi \triangleq \mathbb{E}\|X_n - X_*\|_F$. The proposed compensation algorithm significantly outperforms BL1 - BL5 from medium to high mobility ($>$ 10 km/h). Fig. 2 shows the aggregate achievable data rate versus the terminal mobility under 10 dB SNR. The proposed compensation algorithm outperforms BL1 - BL5 and its performance is very close to BL6 employing exhaustive SVD at every time slot. We can also observe that, although the Newton algorithm (BL5) has a good performance under slow mobility, it is not reliable for high mobility.

To study the computational complexity, we count the number of arithmetic operations (addition, multiplication, etc.) per frame. Omitting the small order terms, the proposed compensation algorithm requires around $14n_t^2 p$ arithmetic operations, compared to $4n_t^2 n_r + 8n_t n_r^2 + 9n_r^3$ (for $n_t \geq n_r$) of BL6 (exhaustive SVD) [3]. In particular, for $n_t \approx n_r \approx n$, our algorithms have a complexity order $\mathcal{O}(n^2 p)$, which is substantially lower than $\mathcal{O}(n^3)$ for the conventional exhaustive SVD algorithm.

6. CONCLUSIONS

In this paper, we have derived a multi-stream iterative SVD algorithm with compensations to track a few principal eigen-subchannels of a massive MIMO system. The algorithm is derived by solving an optimization problem formulated in the Grassmann manifold. To fully utilize the channel temporal correlation, we introduce a compensation term to offset the motion of the target time-varying eigenspace. We show in our analysis that the proposed compensation algorithm converges to the global optimal solution under static channels, and its tracking error is negligible in time-varying channels. We have also demonstrated with numerical results that the proposed compensation algorithm has superior performance advantage over various baselines.

[2] There would be a $\mathcal{O}(\gamma\tau)$ discretization error due to the use of constant step size γ in each frame with duration τ [25, 27].

7. REFERENCES

[1] L.-U. Choi and R. D. Murch, "A transmit preprocessing technique for multiuser MIMO systems using a decomposition approach," *IEEE Transactions on Wireless Communications*, vol. 3, no. 1, pp. 20–24, 2004.

[2] G. Lebrun, J. Gao, and M. Faulkner, "MIMO transmission over a time-varying channel using SVD," *IEEE Transactions on Wireless Communications*, vol. 4, no. 2, pp. 757–764, 2005.

[3] R. A. Horn and C. R. Johnson, *Matrix analysis*. Cambridge university press, 2012.

[4] C.-B. Chae, T. Tang, R. W. Heath, and S. Cho, "MIMO relaying with linear processing for multiuser transmission in fixed relay networks," *IEEE Transactions on Signal Processing*, vol. 56, no. 2, pp. 727–738, 2008.

[5] R. S. Prabhu and B. Daneshrad, "Energy-efficient power loading for a MIMO-SVD system and its performance in flat fading," in *IEEE Global Telecommunications Conference (GLOBECOM 2010)*, 2010, pp. 1–5.

[6] K. Zu, R. C. de Lamare, and M. Haardt, "Low-complexity lattice reduction-aided channel inversion methods for large multi-user MIMO systems," in *The 46th Asilomar Conference on Signals, Systems and Computers (ASILOMAR)*, 2012, pp. 463–467.

[7] J. Tong, P. J. Schreier, and S. R. Weller, "Design and analysis of large MIMO systems with Krylov subspace receivers," *IEEE Transactions on Signal Processing*, vol. 60, no. 5, pp. 2482–2493, 2012.

[8] G. Lee, J. Park, Y. Sung, and J. Seo, "A new approach to beamformer design for massive MIMO systems based on k-regularity," in *IEEE Globecom Workshops*, 2012, pp. 686–690.

[9] P.-A. Absil, R. Mahony, and R. Sepulchre, *Optimization algorithms on matrix manifolds*. Princeton University Press, 2009.

[10] L. W. Tu, *An introduction to manifolds*. Springer, 2011.

[11] C. Udriste, *Convex functions and optimization methods on Riemannian manifolds*. Springer, 1994, vol. 297.

[12] P. Comon and G. H. Golub, "Tracking a few extreme singular values and vectors in signal processing," *Proceedings of the IEEE*, vol. 78, no. 8, pp. 1327–1343, 1990.

[13] J. B. Roseborough and H. Murase, "Partial eigenvalue decomposition for large image sets using run-length encoding," *Pattern Recognition*, vol. 28, no. 3, pp. 421–430, 1995.

[14] K.-J. Bathe, "The subspace iteration method – revisited," *Computers & Structures*, 2012.

[15] G. Stewart, "Simultaneous iteration for computing invariant subspaces of non-hermitian matrices," *Numerische Mathematik*, vol. 25, no. 2, pp. 123–136, 1976.

[16] C. Lanczos, *An iteration method for the solution of the eigenvalue problem of linear differential and integral operators*. United States Governm. Press Office, 1950.

[17] G. Xu and T. Kailath, "Fast subspace decomposition," *IEEE Transactions on Signal Processing*, vol. 42, no. 3, pp. 539–551, 1994.

[18] B. Yang, "Projection approximation subspace tracking," *IEEE Transactions on Signal Processing*, vol. 43, no. 1, pp. 95–107, 1995.

[19] A. S. Poon, D. N. TSe, and R. W. Brodersen, "An adaptive multiantenna transceiver for slowly flat fading channels," *IEEE Transactions on Communications*, vol. 51, no. 11, pp. 1820–1827, 2003.

[20] E. Lundström and L. Eldén, "Adaptive eigenvalue computations using Newton's method on the Grassmann manifold," *SIAM journal on matrix analysis and applications*, vol. 23, no. 3, pp. 819–839, 2002.

[21] A. Abdi and M. Kaveh, "A space-time correlation model for multielement antenna systems in mobile fading channels," *IEEE Journal on Selected Areas in Communications*, vol. 20, no. 3, pp. 550–560, 2002.

[22] M. Zhang, P. J. Smith, and M. Shafi, "An extended one-ring MIMO channel model," *IEEE Transactions on Wireless Communications*, vol. 6, no. 8, pp. 2759–2764, 2007.

[23] P. Absil, "Continuous-time flows on quotient spaces for principal component analysis," in *Proceedings of the 16th International Symposium on Mathematical Theory of Networks and Systems (MTNS2004)*, 2004.

[24] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.

[25] J. Chen and V. K. N. Lau, "Multi-stream iterative singular value decomposition for massive MIMO systems under time varying channels," *submitted to IEEE Transactions on Signal Processing*, 2013.

[26] C. Martin and B. Ottersten, "Asymptotic eigenvalue distributions and capacity for MIMO channels under correlated fading," *IEEE Transactions on Wireless Communications*, vol. 3, no. 4, pp. 1350–1359, 2004.

[27] J. Chen and V. K. N. Lau, "Convergence analysis of mixed timescales cross-layer stochastic optimization," *submitted to IEEE Transactions on Information Theory*, April 2013. [Online]. Available: http://arxiv.org/abs/1305.0153

[28] *TS 36.101 Evolved Universal Terrestrial Radio Access (E-UTRA); User Equipment (UE) radio transmission and reception, Release 10*, March 2012.

[29] K. Baddour and N. Beaulieu, "Autoregressive models for fading channel simulation," in *Global Telecommunications Conference, 2001. IEEE GLOBECOM '01.*, vol. 2, 2001, pp. 1187 –1192 vol.2.

[30] D. Tufts and C. Melissinos, "Simple, effective computation of principal eigenvectors and their eigenvalues and application to high-resolution estimation of frequencies," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 34, no. 5, pp. 1046–1053, 1986.