FEEDBACK OF DIFFERENTIAL PRECODER FOR GEOMETRICAL MEAN DECOMPOSITION SYSTEMS

Hung-Chun Chen and Yuan-Pei Lin
Dept. Electrical Engr., National Chiao Tung Univ., Hsinchu, Taiwan

ABSTRACT
For a time-correlated channel, we consider the differential feedback of the geometrical mean decomposition (GMD) precoder, which is known to be optimal for a number of criteria. When the channel varies slowly, we can expect the optimal GMD precoders of consecutive channel uses to be close. We consider the feedback of the so-called differential precoder and show that it lies in a neighborhood of the identity matrix using matrix perturbation theory. Furthermore the radius of the neighborhood is proportional to a time-correlation parameter. Such a characterization is crucial for efficient quantization of the differential precoder. Simulations are given to demonstrate that, with a small feedback rate, the performance of the proposed differential GMD comes close to the case when perfect channel state information is available to the transmitter.

Index Terms— MIMO system, precoder, differential feedback, time-correlated channel, geometrical mean decomposition.

1. INTRODUCTION
Multiple-input multiple-output (MIMO) systems with limited feedback have been widely studied in recent years [1]-[6]. It has been demonstrated that limited feedback of channel state information improves the performance significantly. Various feedback schemes have been proposed. Feedback of precoder has been extensively investigated, e.g., [2]-[4]. When there is decision feedback at the receiver, the GMD precoder is shown to be optimal for minimizing bit error rate in [4] and codebook design is addressed therein. Feedback of bit loading is considered in [5][6]. The channel considered in these works is an independent fading channel, not time correlated.

In practical systems, there is substantial temporal correlation between consecutive channel uses. Exploitation of the channel correlation leads to more efficient feedback of channel information [7]-[11]. Channel diagonalizing precoders are parameterized in [7] and the differential changes of the parameters are fed back to the transmitter. With the assumption that the channel Gram matrices of consecutive time instants are around a geodesic curve, differential feedback of the channel Gram matrix is presented in [8]. The temporally correlated channel is modeled as a first-order Gauss-Markov process in [9]-[11] to further exploit the statistics of the channel. Differential feedback of precoder based on rotation matrices is proposed in [9]. In [10], the difference of consecutive channel matrices is fed back to transmitter and differential feedback of bit loading is considered in [11] using predictive quantization.

In this paper, we consider differential quantization of the GMD precoder for a time-correlated channel when the receiver has decision feedback. Differential feedback of precoder has also been addressed earlier [7]-[9], but the precoder considered therein are not optimized for error rate. When the terminal speed is low and the channel varies slowly, we can expect the optimal GMD precoders of consecutive channel uses to be very close. We define the so-called differential precoder between two consecutive optimal precoders and exploit properties of GMD to characterize the differential precoder. Using matrix perturbation theory, we show that the differential precoder lies in a small neighborhood around the identity matrix for a slowly varying Gauss-Markov channel. The radius of the neighborhood is quantified and is shown to be proportional to a time-correlation parameter. The radius can be used in the design of codebooks for efficient quantization of the differential precoder. Simulations are given to demonstrate that, with a small feedback rate, the performance of the proposed differential GMD is only 0.5 dB away from the case when the transmitter has full channel information for a moderate moving speed. The sections are organized as follows. In Sec. 2, we introduce the system model for the time-varying MIMO system. Statistical characterization of the differential precoder is presented in Sec. 3. Simulation examples are shown in Sec. 4 and a conclusion is given in Sec. 5.

2. SYSTEM MODEL
Consider a MIMO communication system with $M_t$ transmit antennas and $M_r$ receive antennas. At time $n$, the channel is modeled by an $M_r \times M_t$ matrix $H_n$, whose entries are independent and identically distributed circularly symmetric complex Gaussian random variables with zero mean and unit variance. A useful time correlated channel model is the first-order Gauss-Markov process [12]

$$H_{n+1} = \sqrt{1-\epsilon^2}H_n + \epsilon W_{n+1},$$

(1)

where $W_{n+1}$ is independent of $H_n$ and its entries have the same statistics as those of $H_n$. Using Jake’s model [13], $\epsilon = \sqrt{1-(\sigma_0(2\pi f_d T_s))^2}$, where $\sigma_0(\cdot)$ is the zeroth order Bessel function, $f_d$ is the maximum Doppler frequency and $T_s$ is the time interval between consecutive channel uses. The $M_r \times 1$ channel noise vector $q_n$ is additive white Gaussian with zero mean and variance $N_0$. The precoder $F_n$ is an $M_t \times M$ matrix, where $M$ is the number of substreams with $M \leq \min(M_r,M_t)$. The input vector $s_n$ is assumed to be uncorrelated, and zero mean with $E[|s_n s_n^\dagger|] = P_t/M M_s$, where $P_t$ is the total transmission power.

Fig. 1. A MIMO communication system.

Let the eigenvalue decomposition of $H_n^H H_n$ be $V_n A_n V_n^H$, where $V_n$ is an $M_r \times M_t$ unitary matrix and the diagonal matrix $A_n$ contains the eigenvalues in nonincreasing order, i.e., $\lambda_{n,0} \geq \lambda_{n,1} \geq \cdots \geq \lambda_{n,M_t-1}$. Let the $M \times M$ leading principal
matrix of \( A_n \) be \( [A_n]_M \). Compute the geometric mean decomposition of \( [A_n]_M^{1/2} \) [14],
\[
[A_n]_M^{1/2} = Q_n R_n P_n^T,
\]
where \( Q_n \) and \( P_n \) are \( M \times M \) unitary matrices and \( R_n \) is upper triangular with \( [R_n]_{n,k} = (\prod_{m=1}^{M} \lambda_m)^{1/2M} \). When the zero-forcing decision feedback equalizer is used, the optimal precoder that minimizes the mean square error and average error rate have been shown to be [4][5]
\[
F_n = V_{n,0} P_n,
\]
where \( V_{n,0} \) contains the first \( M \) columns of \( V_n \). The optimal precoder consists of two parts, \( V_{n,0} \) and \( P_n \). The first part \( V_{n,0} \) is formed by the singular vectors of \( H_n \), diagonalizes the channel while the second part \( P_n \) has the effect of equalizing the subchannel SNRs [4] and will be termed the SNR equalizer.

3. Feedback of Differential GMD Precoder
When the channel is time correlated, we can expect the optimal precoder of consecutive channels uses to be correlated as well. Without loss of generality, the precoder at time \( n + 1 \) can be expressed as
\[
F_{n+1} = F_{n,E_n} E_n,
\]
where \( E_n, \) an \( M \times M \) semi-unitary matrix satisfying \( E_n^* E_n = I_M \), will be called the differential precoder. The notation \( A^T \) denotes the transpose conjugate of a matrix \( A \). The \( M \times M \) matrix \( F_{n,E_n} \) is given by \( F_{n,E_n} = [F_n, U_n] \), where \( U_n \) is an \( M \times (M - M) \) matrix chosen in a deterministic manner from \( F_n \), such that \( F_{n,E_n} \) is unitary. For example, we can choose \( F_{n,E_n} = [F_n, V_{n,1}] \), where \( V_{n,1} \) is the matrix that contains the last \( M - M \) columns of \( V_n \). With the feedback of \( E_n \), the transmitter can compute the precoder at \( n + 1 \) from the current precoder. From (4), we can write \( E_n \) as
\[
E_n = F_{n,E_n} F_{n+1}.
\]
Consider the special case \( H_{n+1} = H_n \). The differential precoder matrix \( E_n, \) in (5) becomes \( [I_M]_M^T \). When the channel varies slowly, i.e., \( H_{n+1} \approx H_n \), it can be expected that \( E_n \) is in the neighborhood of \( [I_M]_M^T \). As we will see such a neighborhood can be characterized. To simply the notations, the time index is omitted in the following discussion and \( A_{n+1}, A_n \) are denoted by \( A \) and \( A^T \), respectively.

Define the average distance between \( E \) and \( [I_M]_M^T \) as
\[
D_e = \sqrt{E[\|E - [I_M]_M^T]\|^2_F},
\]
where \( \| \cdot \|_F \) denotes the Frobenius norm of a matrix \( A \). The optimal precoder at time \( n \) depends on \( V \) and \( P \). Thus \( D_e \) depends on \( V \) and \( P \) as well \( V \) and \( P \). From [4], we know that \( V \) and \( P \) are statistically independent. It is reasonable that \( V \) and \( P \) (likewise \( V \) and \( P \)) are also statistically independent. In the following, we derive a bound of \( D_e \) that will give us some insight on how \( D_e \) is related to the variation of the channel.

Lemma 1. Assume \( V \) and \( P \) are respectively statistically independent of \( V \) and \( V \). The average distance \( D_e \) satisfies
\[
D_e \leq \sqrt{E[D_{e,0} + D_{e,1} + D_p]} + 2\sqrt{E[D_{e,0}]E[D_p]}\]
\[
D_{e,0} = \|\tilde{V}_0^T V_0 - [I_M]_M^T\|_F^2
\]
\[
D_{e,1} = \|\tilde{V}_1^T V_0\|_F^2
\]
and
\[
D_p = \|\tilde{P}^T P - [I_M]_M^T\|_F^2.
\]
See Appendix A for a proof. Observe from (8) that \( \sqrt{D_{e,0} + D_{e,1}} \)
represents the distance from \( \tilde{V}_0^T V_0 \) to \( [I_M]_M^T \). On the other hand, \( \sqrt{D_p} \) is the distance between \( P \) and the perturbed matrix \( \tilde{P} \). Thus \( E[D_p] \) corresponds to the perturbation of SNR equalizer and \( E[D_{e,0} + D_{e,1}] \) corresponds to the perturbation of eigenspace. In the following two subsections, we quantify the perturbations of eigenspace and SNR equalizer for the Gauss-Markov channel in (1).

3.1. Perturbation of Eigenspace
Consider the time-correlated channel model given in (1), we have
\[
\tilde{H}^T \tilde{H} = H^T H + \epsilon \sqrt{1 - \epsilon^2} \Delta_0 + \epsilon^2 \Delta_1,
\]
where \( \Delta_0 = H^T W + \tilde{W}^T H \) and \( \Delta_1 = W^T W - H^T H \). When the time-correlated channel is changing slowly, i.e., \( \epsilon \) is small, \( \tilde{H}^T \tilde{H} \) can be viewed as a perturbation of \( H^T H \). There are many results in the literature on the perturbation of matrices [15][16]. In these studies, \( D_{e,1} = \| \tilde{V}_1^T V_0 \|_F^2 \) is regarded as eigenspace variation and various bounds have been derived. However there is no discussion on \( D_{e,0} + D_{e,1} \), to the best of our knowledge. In the following, we use a technique similar to that in [15] to derive a bound for \( D_{e,0} + D_{e,1} \). Then \( D_{e,0} + D_{e,1} \) can be bounded in a similar manner.

The column vectors of \( V_0 \) correspond to the eigenvectors of \( H^T H \) and they are not uniquely determined. However we can always choose \( V_0 \) such that \( \tilde{V}_0^T V_0 \) is a positive real number for all \( j \). In this case, we have the following second order approximation of the perturbation of eigenspace when \( \epsilon \) is small.

Lemma 2. Assume \( \tilde{\lambda}_i \neq \lambda_j \) for \( 0 \leq i \leq M - 1 \) and \( 0 \leq j \leq M - 1 \). A second order approximation of \( D_{e,0} + D_{e,1} \) is given by
\[
D_{e,0} + D_{e,1} \leq \epsilon^2 \max_{i,j \in S_1} (\Delta_0^T V_0)^2 / (\tilde{\lambda}_i - \lambda_j)^2.
\]
See Appendix B for a proof. Notice that an upper bound for the term \( \sum_{i=0}^{M-1} \sum_{j=0}^{M-1} (\tilde{V}_1^T V_0)^2 / (\tilde{\lambda}_i - \lambda_j)^2 \), where \( S_1 = \{ j \neq i, i,j \in N \ | \ 0 \leq i \leq M - 1, 0 \leq j \leq M - 1 \} \). The above expression depends the distance between \( \tilde{\lambda}_i \) and \( \lambda_j \). Using perturbation theory for matrix eigenvalues [17], the distance satisfies \( |\tilde{\lambda}_i - \lambda_j| \leq \epsilon \|G\|_2 \), where \( \|G\|_2 \) denotes the two norm of a matrix \( G \). Using this result and ignoring the third or higher order terms of \( \epsilon \), we obtain an upper bound that depends on the singular values of the current channel but not those of the previous channel, given by
\[
D_{e,0} + D_{e,1} \leq \epsilon^2 \max_{i,j \in S_0} (\Delta_0^T V_0)^2 / (\tilde{\lambda}_i - \lambda_j)^2.
\]
Similarly, we can show \( D_{e,0} \leq \epsilon^2 \max_{i,j \in S_0} (\Delta_0^T V_0)^2 / (\tilde{\lambda}_i - \lambda_j)^2 \), where \( S_0 = \{ j \neq i, i,j \in N \ | \ 0 \leq i \leq M - 1, 0 \leq j \leq M - 1 \} \). These results lead to the following result on the average perturbation of eigenspace.

Theorem 1. Consider the Gauss-Markov channel in (1). For a small \( \epsilon \), we have \( E[D_{e,0} + D_{e,1}] \leq \epsilon^2 \rho_{e,1} \) and \( E[D_{e,0}] \leq \epsilon^2 \rho_{e,0} \), where
\[
\rho_{e,k} = \|H\| \max_{i,j \in S_k} (\tilde{\lambda}_i - \lambda_j)^2.
\]
and \( t_3 = M \sum_{\ell=0}^{M-1} \lambda_\ell \).

Proof. (12) can be obtained by taking expectation of the bound in (11). A sketch of proof is given below. The computation requires averaging over the random matrices \( \mathbf{H} \) and \( \mathbf{W} \). To do this, we observe that \( \mathbf{W} \) is independent of the channel \( \mathbf{H} \), so the expectation \( E[D_{d1}, a, D_{d1}] \) can be obtained using \( \mathbb{E}[E_{\mathbf{W}}(E_{\mathbf{W}}[D_{d1}, a, D_{d1}]+\mathbf{H})] \).

Also observe that the elements of \( \mathbf{W} \) are i.i.d. Gaussian with zero mean and unit variance, so \( E_{\mathbf{W}}[\mathbf{W}^\dagger \mathbf{W} \mathbf{K} \mathbf{W}] = \text{tr}(K)I_M \), and \( E_{\mathbf{W}}[\mathbf{W} \mathbf{G} \mathbf{W}^\dagger] = 0 \) for a deterministic \( M \times M \) matrix \( \mathbf{K} \) and \( M \times M \) matrix \( \mathbf{G} \). The theorem can be proved using these two observations.

The bound in Theorem 1 shows that the perturbation of eigenspace is proportional to \( \epsilon^2 \). The constants \( \rho_{\epsilon, \ell} \) in (12) depend on \( \{\lambda_\ell\} \), the eigenvalues of \( \mathbf{H}^\dagger \mathbf{H} \). As the elements of \( \mathbf{H} \) are i.i.d complex Gaussian random variables with zero mean and unit variance, the matrix \( \mathbf{H}^\dagger \mathbf{H} \) has a Wishart distribution. Thus the joint probability density function for the ordered eigenvalues of a Wishart matrix in [18] can be used to compute the expectation \( \rho_{\epsilon, \ell} \) in (12).

### 3.2. Perturbation of SNR equalizer

We first review a closed form solution of \( \rho \) that helps to establish a connection between \( D_p \) and the perturbation of eigenvalues. To ease the derivation of the perturbation of SNR equalizer, we consider \( M = 3 \). The case \( M = 2 \) follows directly. Derivation for a general \( M \) can be found in [19].

It is known that the SNR equalizer \( \rho \) can be expressed as a product of permutation matrices and Givens rotations [14]. Define \( \alpha_k = \sqrt{(\lambda_k - \lambda)/(\lambda_k - d_0)} \), \( \beta_k = \sqrt{(\lambda - d_k)/(\lambda_k - d_k)} \) for \( k = 0, 1 \), where \( \lambda = (\lambda_0 \lambda_1 \lambda_2)^{1/3} \), \( d_0 = \lambda_2 \) and \( d_1 = \lambda_0 \lambda_2 / \lambda \). Then \( \mathbf{P} = \mathbf{A}_0 \mathbf{A}_1 \) [14], where

\[
\mathbf{A}_0 = \begin{bmatrix} \alpha_0 & -\beta_0 & 0 \\ 0 & 0 & 1 \\ \beta_0 & 0 & \alpha_0 \end{bmatrix}, \quad \mathbf{A}_1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & -\beta_1 \\ 0 & \beta_1 & \alpha_1 \end{bmatrix}. \tag{13}
\]

Similarly, \( \mathbf{P} \) can be given in terms of \( \tilde{\alpha}_k \) and \( \tilde{\beta}_k \). We also define \( \theta_k = \sin^{-1} \alpha_k \), \( \tilde{\theta}_k = \sin^{-1} \tilde{\alpha}_k \). Then \( \delta_k = \tilde{\theta}_k - \theta_k \) represents the perturbation in rotation angles. We show in Appendix C that \( D_p \) can be approximated nicely as

\[
D_p \approx 2(\delta_0^2 + \delta_1^2). \tag{14}
\]

Although the statistical properties of \( \delta_0^2 \) are not readily available, they can be bounded from above using eigenvalues \( \lambda_i \) that are statistically more tractable, as we will see next. Using \( \delta_k \approx \sin \delta_k \) and \( \sin \delta_k = \sin(\theta_k - \tilde{\theta}_k) \), we have \( \delta_k \approx \alpha_k \beta_k (\alpha_k \alpha_k^{-1} - \tilde{\beta}_k \tilde{\beta}_k^{-1}) \) for a small \( \epsilon \). Notice that \( \alpha_k \alpha_k^{-1} \) and \( \tilde{\beta}_k \tilde{\beta}_k^{-1} \) are both close to one. For \( x \approx 1 \), Taylor approximation yields \( \sqrt{x} \approx (x + 1)/2 \). This means

\[
\alpha_k \alpha_k^{-1} \approx \frac{\alpha_k^2}{2} + 1/2 \quad \text{and} \quad \tilde{\beta}_k \tilde{\beta}_k^{-1} \approx \frac{\tilde{\beta}_k^2}{2} + 1/2.
\]

Thus we have

\[
\delta_k \approx (1 + \alpha_k \alpha_k^{-1}) (\alpha_k - \tilde{\alpha}_k )/2(\alpha_k). \tag{15}
\]

Substituting the definitions of \( \alpha_k \) and \( \tilde{\alpha}_k \) to the approximation and ignoring the second and higher order terms of \( \epsilon \), we obtain

\[
\delta_k \approx \nu_k \epsilon \mathbf{c}_1^T \mathbf{y},
\]

where \( \mathbf{y} = [\tilde{\lambda}_0 - \lambda_0 \tilde{\lambda}_2 - \lambda_2 \lambda_1 - \lambda_1] \).

\[
\mathbf{c}_0 = \begin{bmatrix} \frac{\tilde{\lambda}_0 - \lambda_0 \tilde{\lambda}_2 - \lambda_2 \lambda_1 - \lambda_1}{\lambda_0} \\ \frac{\tilde{\lambda}_0 - \lambda_0 \tilde{\lambda}_2 - \lambda_2 \lambda_1 - \lambda_1}{\lambda_2} \\ \frac{\tilde{\lambda}_0 - \lambda_0 \tilde{\lambda}_2 - \lambda_2 \lambda_1 - \lambda_1}{\lambda_1} \end{bmatrix}, \quad \mathbf{c}_1 = \begin{bmatrix} \frac{\tilde{\lambda}_0 - \lambda_0 \tilde{\lambda}_2 - \lambda_2 \lambda_1 - \lambda_1}{\lambda_0} \\ \frac{\tilde{\lambda}_0 - \lambda_0 \tilde{\lambda}_2 - \lambda_2 \lambda_1 - \lambda_1}{\lambda_2} \\ \frac{\tilde{\lambda}_0 - \lambda_0 \tilde{\lambda}_2 - \lambda_2 \lambda_1 - \lambda_1}{\lambda_1} \end{bmatrix}.
\]
[9] and the geodesic curves in [8] are labeled as ‘differential rotation’ and ‘geodesic’, respectively. For ‘differential channel’ [10], the feedback information is the difference of consecutive channels. We also compare with the GMD precoder [4] that does not take the correlation into consideration. The results are shown in Fig. 3 for ϵ = 0.06. We can see that the proposed feedback scheme has a good BER performance. This is because GMD with full channel information achieves the minimum BER and the performance of our proposed method is very close to the unquantized GMD.

5. CONCLUSION
In this paper, we consider differential quantization of GMD precoder for a time-correlated channel. Modeling the time-correlated channel as a first-order Gauss-Markov process, we show the differential precoder is in the neighborhood of the identity matrix for a small ϵ. Moreover we derive an upper bound of the radius of the neighborhood and show it is proportional to ϵ. The result is very useful towards the codebook design for the differential precoder, as demonstrated by simulations. With a small feedback rate, the performance of the proposed differential GMD comes close to that of GMD with perfect channel state information at the transmitter.

6. APPENDIX
6.1. Appendix A: Proof of Lemma 1

From (3) and (5), E can be expressed as $E = \{V_oP - V_0\}^\dagger \hat{V}_o\hat{P}$, so we have $\|E - [I_M 0]\|_F^2 = \|P[V_oV_oP - I_M]\|_F^2 + \|V_0\hat{V}_o\hat{P}\|_F^2$. Using $P[V_oV_oP - I_M] = P[V_oV_oP - I_M]P + \hat{P}[\hat{V}_o\hat{V}_o - I_M]P$ and $\|A + B\|_F \leq \|A\|_F + \|B\|_F$, we obtain $\|P[V_oV_oP - I_M]\|_F^2 \leq \|P[V_oV_oP - I_M]P\|_F^2 + 2\|P[V_oV_oP - I_M]P\|_F\|P\|_F + \|\hat{P}\|_F - I_M\|_F^2$. We know that the Frobenius norm is an unitary invariance norm and $\|V_0\hat{V}_o\|_F = \|V_0\|_F$. Thus we have the upper bound for $D_x$, $\sqrt{E[D_{x,0} + D_{x,1} + D_P] + 2E[\sqrt{D_{x,0}}\sqrt{D_P}]}$.

When $P$ and $\hat{P}$ are respectively statistically independent of $\hat{V}$ and $V$, we have $E[\sqrt{D_{x,0}D_P}] = E[\sqrt{D_{x,1}}E[\sqrt{D_P}]]$. Using the Jensen’s inequality $E[\sqrt{X}] \leq \sqrt{E[X]}$, the result follows.

6.2. Appendix B: Proof of Lemma 2

Pre- and post-multiplying (10) by $\hat{V}^\dagger$ and $V$, respectively and considering the $j$th element of the left and right hand side, we obtain $[\hat{V}^\dagger V]_{j|i}(\hat{\lambda}_i - \lambda_j) = \epsilon\sqrt{1 - \epsilon^2}\hat{V}^\dagger\Delta_0 V_{j|i} + \epsilon^2[\hat{V}^\dagger \Delta_0 V_{j|i}]/(\hat{\lambda}_i - \lambda_j)$. When $\hat{\lambda}_i \neq \lambda_j$ for $0 \leq i \leq M_t - 1$ and $0 \leq j \leq M - 1$, the $j$th element of $\hat{V}^\dagger V$ can be written as

$$[\hat{V}^\dagger V]_{j|i} = (\epsilon\sqrt{1 - \epsilon^2}\hat{V}^\dagger\Delta_0 V_{j|i} + \epsilon^2[\hat{V}^\dagger \Delta_0 V_{j|i}]/(\hat{\lambda}_i - \lambda_j)).$$

When $[\hat{V}^\dagger V_{0}]_{j|i}$ is a positive real number for all $j$, $D_{x,0} + D_{x,1}$ in (8) becomes

$$D_{x,0} + D_{x,1} = \sum_{j=0}^{M_t-1} (1 - \|V_{0j}\|_F^2)^2 + \sum_{i=0,j\neq i}^{M_t-1} \sum_{j=0,j\neq i}^{M_t-1} \|V_0V_{0j}\|_F^2 \|V_{0i}\|_F^2.$$ (20)

Using (19) and $\|\hat{V}_0\hat{V}_{0j}\|_F^2 = 1 - \sum_{i=0}^{M_t-1} \|V_0V_{ij}\|^2_2$, it can be shown that $(1 - \|V_0V_{0j}\|^2_2)$ is in the order of $\epsilon^4$. The proof can be found in [19]. Thus (20) can be approximated as $D_{x,0} + D_{x,1} \approx \sum_{i=0}^{M_t-1} \sum_{j=0,j\neq i}^{M_t-1} \|V_0V_{0j}\|^2_2$. Combining the approximation and (19), we obtain the result.

6.3. Appendix C: Proof of (14)

Using the fact that $P$ and $\hat{P}$ are real and unitary and $\|A\|^2_F = tr(A^\dagger A)$, we can rewrite $D_P$ in (9) as $D_P = 2\sum_{i=0}^{M_t-1} (1 - \|p_i\|_F^2)$. With (13), it can be verified that $p_i^\dagger p_0 = \alpha_0\hat{\alpha}_0 + \beta_0\hat{\beta}_0$, $p_i^\dagger p_i = \alpha_0\hat{\alpha}_0 + \beta_0\hat{\beta}_0 + \beta_0\hat{\beta}_0 + \beta_0\hat{\beta}_0 + \beta_0\hat{\beta}_0 + \beta_0\hat{\beta}_0$. On the other hand, $\cos \delta_k = \cos (\hat{\theta}_k - \theta_k)$, i.e., $\cos \delta_k = \alpha_0\hat{\alpha}_0 + \beta_0\hat{\beta}_0$. When $\delta_k$ is small, we obtain $\cos \delta_k \approx 1 - \delta_k^2/2$. Using these results, we arrive at (14).
7. REFERENCES


