Fig. 2. Pareto boundary for a sample channel realization with $N = 2$ two transmit antennas at high SNR 30 dB.

VI. CONCLUDING REMARKS

The motivation for this correspondence has been the recent, huge interest in IFCs as a model for spectrum resource conflicts (see, e.g., [5], [7], [9], [10], and [12], and the references therein). Our main contribution has been a characterization of the MISO IFC for arbitrary SNR, and specifically a parametrization of the Pareto boundary of the rate region. Our hope is that the results will be useful for future research on resource allocation and spectrum sharing for situations that are well modeled via the MISO IFC.

REFERENCES

[1] R. Ahlswede, "The capacity region of a channel with two senders and two receivers," *Ann. Prob.*, vol. 2, pp. 805–814, 1974.

[2] A. B. Carleial, "Interference channels," *IEEE Trans. Inf. Theory*, vol. 24, no. 1, pp. 60–70, Jan. 1978.

[3] T. Han and K. Kobayashi, "A new achievable rate region for the interference channel," *IEEE Trans. Inf. Theory*, vol. 27, no. 1, pp. 49–60, Jan. 1981.

[4] M. H. M. Costa, "On the Gaussian interference channel," *IEEE Trans. Inf. Theory*, vol. 31, no. 5, pp. 607–615, Sep. 1985.

[5] X. Shang, B. Chen, and M. J. Gans, "On the achievable sum rate for MIMO interference channels," *IEEE Trans. Inf. Theory*, vol. 52, no. 9, pp. 4313–4320, Sep. 2006.

[6] S. A. Jafar and M. Fakhereddin, "Degrees of freedom for the MIMO interference channel," *IEEE Trans. Inf. Theory*, vol. 53, no. 7, pp. 2637–2642, Jul. 2007.

[7] M. Charafeddine, A. Sezgin, and A. Paulraj, "Rate region frontiers for n-user interference channel with interference as noise," in *Proc. 45th Allerton Conf. Communication, Control, and Computing*, Monticello, IL, Sep. 26–28, 2007.

[8] X. Shang and B. Chen, "Achievable rate region for downlink beamforming in the presence of interference," in *Proc. IEEE Asilomar Conf. Signgals, Systems, and Computers*, Pacific Grove, CA, Nov. 4–7, 2007, pp. 1684–1688.

[9] E. A. Jorswieck and E. G. Larsson, "The MISO interference channel from a game-theoretic perspective: A combination of selfishness and altruism achieves Pareto optimality," in *Proc. Int. Conf. Acoustics, Speech, Signal Processing (ICASSP)*, 2008.

[10] S. Vishwanath and S. A. Jafar, "On the capacity of vector Gaussian interference channels," in *Proc. IEEE Information Theory Workshop*, San Antonio, TX, Oct. 24–29, 2004, pp. 365–369.

[11] E. G. Larsson and P. Stoica, *Space-Time Block Coding for Wireless Communications*. Cambridge, U.K.: Cambridge Uni. Press, 2003.

[12] E. G. Larsson and E. A. Jorswieck, "Competition versus collaboration on the MISO interference channel," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 7, pp. 1059–1069, Sep. 2008.

[13] E. A. Jorswieck and E. G. Larsson, "Linear precoding in multiple antenna broadcast channels: Efficient computation of the achievable rate region," in *Proc. IEEE/ITG Int. Workshop on Smart Antennas*, Darmstadt, Grtmsny, Feb. 26–27, 2008, pp. 21–28.

[14] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge, U.K.: Cambridge Univ. Press, 2005.

Multimode Precoding for MIMO Systems: Performance Bounds and Limited Feedback Codebook Design

Xiaofei Song and Heung-No Lee, *Member, IEEE*

Abstract—This correspondence investigates the problem of designing the precoding codebook for limited feedback multiple-input multiple-output (MIMO) systems. We first analyze the asymptotic capacity loss of a suboptimal multimode precoding scheme as compared to optimal waterfilling and show that the suboptimal scheme is sufficient when negligible capacity loss is allowed. This knowledge is then applied to the design of the limited feedback codebook. In the design, the generalized Lloyd algorithm is employed, where the computation of the centroid is formulated as an optimization problem and solved optimally. Numerical results show that the proposed codebook design outperforms the comparable algorithms reported in the literature.

Index Terms—Given's rotation, limited feedback codebook design, Lloyd algorithm, waterfilling.

I. INTRODUCTION

A well-known result of information theory establishes that feedback does not improve the capacity of a discrete memoryless channel [1]. Nonetheless, for the cases where the channel is selective in either time, frequency, or space, feedback of the channel state to the transmitter can bring substantial benefits to the forward communications system in terms of either capacity, performance, or complexity. The theoretical study of capacity and coding with channel state information at the transmitter (CSIT) can be traced back as early as to Shannon [2]. More recently, information-theoretic capacity on channels with both perfect [3]–[5] and imperfect [6] CSIT and practical coding schemes using CSIT [7], [8] have been studied.

With the advent of multiple-input and multiple-output (MIMO) antenna systems, investigation on the potential benefits of CSIT for MIMO systems has been intensified and design of a practical scheme to achieve the potential benefits as closely as possible becomes very important. The channel estimation done at the receiver needs to be sent back to the transmitter to provide the potential CSIT benefit. Thus, the study of MIMO system with limited feedback is of practical interests. In the past, various options in MIMO transmit beamforming

with limited rate feedback have been considered in [9]–[13]. In the beamforming setting, however, one notes that the capacity loss is usually large, as compared to the optimal water-filling (WF) solution [14]. To remedy this shortcoming, in [15]–[17], the problem is approached from the perspective of designing a codebook to achieve the WF gain. The optimal WF solution is a mixture of optimal antenna phase rotation and power adaptation, which changes subject to a particular realization of the channel. The proposed codebook design methodology in [15] includes both the phase rotation and power allocation matrix. Whereas in the multimode precoding scheme [16], [17], the codebook contains only the first m columns of the phase rotation matrix (first m eigenmodes) and the total transmit power is equally allocated to the m eigenmodes.

In this correspondence, we also address the codebook design problem in the framework of multimode precoding. We first bound the asymptotic capacity loss of multimode precoding, and show the sufficiency of it when negligible capacity loss is allowed. We then use the generalized Lloyd algorithm [18] to design the multimode precoding codebook. One novelty of our design methodology lies in the computation of the centroid, which is formulated as an optimization problem by the Given's angle parametrization of unitary matrix (orthonormal column matrix). The proposed algorithm, although it is based on a suboptimal multimode precoding scheme rather than on the optimal WF as in [15], outperforms the algorithm in [15] as the numerical results indicate. We claim the gain in performance is largely owing to the optimal computation of the centroid. Our proposed design also brings performance superior, we show, to that of the same multimode precoding scheme in [16]. The gain over [16] comes mainly from the fact that the cardinality distribution of the designed codebook employing our proposed algorithm is much closer to optimal (if not exactly optimal) as compared to the prederived one in [16].

The remainder of the correspondence is organized as follows. In Section II, the MIMO system with multimode precoding is introduced. In Section III, the asymptotic capacity loss of multimode precoding as compared to WF is developed. The codebook design algorithm is presented in Section IV, and the numerical results are shown in Section V. Finally, conclusions are drawn in Section VI.

II. SYSTEM OVERVIEW

We consider a MIMO system with N_t inputs and N_r outputs. The $[N_r \times 1]$ output signal vector is modeled as $\mathbf{r} = \mathbf{Hs} + \mathbf{n}$, where $\mathbf{H}$ is a $[N_r \times N_t]$ channel matrix with circularly symmetric complex Gaussian entries of zero mean and unit variance; $\mathbf{s}$ is a $[N_t \times 1]$ transmitted signal with total power constraint P_t, such as $\mathbb{E}[\mathbf{s}^*\mathbf{s}] \leq P_t$; $\mathbf{n}$ is the $[N_r \times 1]$ additive white Gaussian noise present at the receiver, with $\mathbb{E}[\mathbf{nn}^*] = \sigma^2 \mathbf{I}_{N_r \times N_r}$. Without loss of generality, we assume $\sigma^2 = 1$ throughout the correspondence.

By the singular value decomposition, the channel matrix $\mathbf{H}$ can be written as $\mathbf{H} = \mathbf{U}\mathbf{\Lambda}\mathbf{V}^*$, where both $\mathbf{U}$ and $\mathbf{V}$ are the unitary matrices and $\mathbf{\Lambda}$ is a diagonal matrix with the singular values of $\mathbf{H}$ on its diagonal. When the channel is known to both the transmitter and the receiver, the capacity achieving power allocation solution is the well-known WF, and the instantaneous WF capacity [14] is

$$C_w(\mathbf{H}) = \sum_{i=1}^{m^*} \log(\mu \lambda_i). \tag{1}$$

$\lambda_i^{1/2}$ is the ith largest singular value of $\mathbf{H}$ and $\log$ is the logarithm to base 2. For convenience, we define the subchannel loss as $g_i := 1/\lambda_i$ and the set of *good* sub-channels as $\mathcal{E} = \{g_i : \mu - g_i > 0, i = $

$1, 2, \ldots, N_t\}$. Then m^* denotes the cardinality of the *good* subchannels, i.e., $m^* = |\mathcal{E}|$. The water level μ relates to P_t by satisfying the total power constraint, i.e., $P_t = \sum_{i=1}^{m^*}(\mu - g_i)$ or

$$S_0 := \frac{P_t}{m^*} = \mu - \frac{1}{m^*}\sum_{i=1}^{m^*} g_i = \mu - \bar{g} \tag{2}$$

where $\bar{g} := 1/m^* \sum_{i=1}^{m^*} g_i$ is the mean.

While perfect channel knowledge at the receiver can be assumed, perfect CSIT may not be available, in particular for frequency division duplex (FDD) systems where no "reciprocity" exists between the forward and reverse channel. In such a case, the CSI estimated at the receiver needs to be fed back to the transmitter through a feedback channel. While the optimal WF requires knowing both $\mathbf{V}$ and $\mathbf{\Lambda}$ at the transmitter, under the constraint on the feedback channel capacity, we consider a suboptimal transmission scheme which is termed multimode precoding in [16]. With limited feedback, only the quantized first m columns of the eigenmode matrix $\mathbf{V}$ are used at the transmitter under multimode precoding, and an equal amount of power P_t/m is allocated to each eigenmode. The $N_t \times m$ unitary matrix $\hat{\mathbf{V}}_m$ serves as the precoding matrix at the transmitter. This then yields a modified input–output relationship

$$\mathbf{r} = \mathbf{H}\hat{V}_m \mathbf{s}_m + \mathbf{n} \tag{3}$$

where $\mathbf{s}_m$ is $m \times 1$ input with $\mathbb{E}[\mathbf{s}_m \mathbf{s}_m^*] = (P_t/m)\mathbf{I}_{m \times m}$. The instaneous channel capacity then is

$$C_p(\mathbf{H}, \hat{\mathbf{V}}_m) = \log \det \left(\mathbf{I}_{N_r} + \frac{P_t}{m}\mathbf{H}\hat{\mathbf{V}}_m \hat{\mathbf{V}}_m^* \mathbf{H}^* \right). \tag{4}$$

III. ASYMPTOTIC CAPACITY LOSS OF MULTIMODE PRECODING

As the multimode precoding scheme is based on a suboptimal rather than optimal WF, it is worthwhile to investigate its loss in capacity as compared to the WF based optimal method. However, the capacity of the channel for either scheme can not be easily obtained with a given feedback rate, thus we resort to the asymptotic capacity loss of the multimode precoding scheme when the feedback rate is assumed infinite.

The optimal multimode precoding scheme will use the exact first m' columns of $\mathbf{V}$ and $\mathbf{V}_{m'}$ as the precoeding matrix, where $m' = \arg\max_{1 \leq m \leq N_t} C_p(\mathbf{H}, V_m)$. The instaneous capacity then is

$$C_p(\mathbf{H}, \mathbf{V}_{m'}) = \sum_{i=1}^{m'} \log \left(\frac{1 + P_t \lambda_i}{m'} \right). \tag{5}$$

The asymptotic instaneous capacity loss of multimode precoding, using (5) and (1) can be then be upperbounded as

$$\begin{aligned}
\Delta_C &:= C_w(\mathbf{H}) - C_p(\mathbf{H}, \mathbf{V}_{m'}) \leq C_w(\mathbf{H}) - C_p(\mathbf{H}, \mathbf{V}_{m^*}) \\
&= -\sum_{i=1}^{m^*} \log \frac{\frac{1 + P_t \lambda_i}{m^*}}{\mu \lambda_i} \\
&= -\sum_{i=1}^{m^*} \log \left(1 + \frac{g_i - \bar{g}}{\mu} \right)
\end{aligned} \tag{6}$$

where the inequality is by the fact $C_p(\mathbf{H}, \mathbf{V}_{m'}) \geq C_p(\mathbf{H}, \mathbf{V}_{m^*})$ and the last equality follows from (2).

The capacity loss (6) depends on the optimal power level μ. Thus, it can only be evaluated numerically as no analytical solution to μ exists. To offer an insight into the capacity loss, we expand the summand in (6) by the Taylor series and obtain the following two upper bounds.

Theorem 1: The asymptotic capacity loss of multimode precoding scheme is upper bounded by

$$\Delta_C \leq \frac{1}{\ln 2} m^* \frac{\sigma^2(g)}{\mu^2} \tag{7}$$

where $\sigma^2(g)$ denotes the sample-variance of the subchannel losses (inverse sub-channel gains) in the set of good subchannels $\mathcal{E}$. Meanwhile, by using the absolute value of the first-order term, we have

$$\Delta_C \leq \frac{1}{\ln 2} \sum_{i=1}^{m^*} \frac{|g_i - \bar{g}|}{\mu}. \tag{8}$$

Proof: See Appendix A.

These bounds, though easy to obtain, provide useful tools to make inference on the loss of a suboptimal solution. It is noteworthy that the asymptotic multimode precoding scheme subsumes the beamforming scheme in low signal-to-noise ratio (SNR) regime and the equal power allocation scheme to all transmit antennas in high SNR regime (when $N_t \leq N_r$). Therefore, making use of these upper bounds, we can explain why beamforming and the equal power allocation are optimal in the low and the high SNR regime, respectively. In the high SNR regime, i.e., $P_t \gg 1$ and $\mu \gg 1$, the upper bound (7) becomes zero, and thus the equal power allocation is optimal; in the low SNR regime, i.e., $P_t \ll 1$, $m^* = 1$ and $\sigma^2(g) = 0$, the upper bound also goes to zero, thus beamforming is optimal in this case.

Based on the ensemble of randomly generated channel matrix $\mathbf{H}$, the average asymptotic capacity loss of multimode precoding is plotted in Fig. 1. The capacity loss is also compared with the bounds derived. Selected system parameters are $N_t = N_r = 16$. Yu–Cioffi bound [19] is also provided for the purpose of comparison. From the simulation results, we can see that the capacity loss of multimode precoding is indeed small as reflected in this figure. This suggests the suboptimal scheme based multimode precoding can be used in practice at the cost of negligible capacity loss.

IV. FEEDBACK CODEBOOK DESIGN

Motivated by the small asymptotic capacity loss of multimode precoding, we now consider the problem of designing the multimode precoding codebook $\mathcal{C}$ under the cardinality constraint $|\mathcal{C}| = N_{\text{tot}}$. Assume the feedback channel is error free, this precoding codebook $\mathcal{C}$ at the transmitter is also the feedback codebook used at the receiver. Under the multimode precoding scheme, each codeword $\hat{\mathbf{V}}^k$ ($1 \leq k \leq N_{\text{tot}}$) is a $N_t \times m$ unitary matrix, and the codebook $\mathcal{C}$ is a finite subset of unitary matrices, i.e., $\mathcal{C} \subset \mathcal{U} = \bigcup_{m=1}^{N_t} \mathcal{U}_m$ where $\mathcal{U}_m$ is the set of complex $N_t \times m$ unitary matrix. With a slight change of notation, we drop the dimensionality signifying subscript m of the kth codeword $\hat{\mathbf{V}}^k$.

The problem of codebook design is essentially a vector quantization problem and hence the conventional generalized Lloyd algorithm is applied here to find a codebook that optimizes an overall distortion measure. According to the generalized Lloyd algorithm, a set of channel matrix $\mathbf{H}$ will be generated randomly according to a given channel statistics as the training sequence. We also randomly generate N_{tot} number of orthonormal column matrices, $\{\hat{\mathbf{V}}^k\}_{k=1}^{N_{\text{tot}}}$, as the initial codebook $\mathcal{C}$.

Depending upon the interests of system performance, various criteria such as probability of error [16], [20], capacity [15], [16] and the error exponent [21] have been utilized to design the codebook. In this section, our design goal is to find the codebook $\mathcal{C}$ that maximizes the forward channel capacity. It is equivalently the codebook that minimizes capacity loss compared to the optimal WF. The capacity loss is thus employed as our distortion measure, i.e.,

$$d(\mathbf{H}, \hat{\mathbf{V}}^k) = C_w(\mathbf{H}) - C_p(\mathbf{H}, \hat{\mathbf{V}}^k) \tag{9}$$

where $C_p(\mathbf{H}, \hat{\mathbf{V}}^k)$ is defined in (4).

Given a codebook $\mathcal{C}$, we first find the optimal partition of the training sequence according to the distortion measure, i.e.,

$$\hat{\mathbf{V}}^{k^*} = \arg \min_{\hat{\mathbf{V}}^k \in \mathcal{C}} d(\mathbf{H}, \hat{\mathbf{V}}^k)$$
$$= \arg \max_{\hat{\mathbf{V}}^k \in \mathcal{C}} C_p(\mathbf{H}, \hat{\mathbf{V}}^k) := T(\mathbf{H}). \tag{10}$$

Define the kth cluster as $\mathcal{R}^k = \left\{ \mathbf{H} \ : \ T(\mathbf{H}) = \hat{\mathbf{V}}^k \right\}$. Then, the kth partial distortion is a conditional expectation $D(\hat{\mathbf{V}}^k) = \mathbb{E}\left[d(\mathbf{H}, \hat{\mathbf{V}}^k) | \mathbf{H} \in \mathcal{R}^k \right]$. To update the codebook, we then recompute the optimal centroid (new codeword) within each cluster $\mathcal{R}^k$ such that the distortion in the cluster is minimized, i.e.,

$$\hat{\mathbf{V}}^k = \arg \min_{\hat{\mathbf{V}}^k \in \mathcal{U}} D(\hat{\mathbf{V}}^k). \tag{11}$$

Once the new codebook is generated, we can repeat the previous process until the overall distortion $D = \sum_{k=1}^{N_{\text{tot}}} D(\hat{\mathbf{V}}^k) p_k$ has changed little since the last iteration, where p_k is the probability that a specific channel matrix $\mathbf{H}$ falls into the cluster $\mathcal{R}^k$.

The signal design procedure is summarized in the following two-step algorithm.

S1) Specify P_t the total available transmit power from a given SNR value. Randomly generate the ensemble of channel matrices $\mathbf{H}$ according to the given channel distribution which serve as the training set. Randomly generate an initial codebook $\mathcal{C}$ with N_{tot} codewords.

S2) Repeat the following sub-steps until the change of distortion D becomes negligible.
   a) Given a codebook $\mathcal{C}$, redistribute each channel matrix $\mathbf{H}$ into one of the clusters in $\mathcal{C}$ by selecting the one whose centroid is closer to $\mathbf{H}$, i.e., $\mathbf{H} \in \mathcal{R}^{k^*} \iff d(\mathbf{H}, \hat{\mathbf{V}}^{k^*}) \leq d(\mathbf{H}, \hat{\mathbf{V}}^k)$ for $k^* \neq k$.
   b) Recompute the centroid for each cluster $\mathcal{R}^k$ created, i.e., $\hat{\mathbf{V}}^k = \arg \min_{\hat{\mathbf{V}}^k \in \mathcal{U}} D(\hat{\mathbf{V}}^k)$ to obtain a new codebook $\mathcal{C}$. If an empty cluster is generated in (a), randomly generate another replacement centroid.
   c) Compute the overall distortion D for the new generated codebook $\mathcal{C}$.

We next discuss a few issues with the design algorithm.

A. Computation of the Centroid

One difficult part of the Lloyd algorithm is the computation of the centroid (11). In conventional scalar or vector quantization problems with the Euclidian distance measure [22], an explicit expression of computing the centroid can be obtained. However, this is mathematically intractable in our problem as the distortion measure (9) employed is non-linear. In [15], the authors employed a similar Lloyd algorithm with a difference such that their design combines optimal antenna phase rotation and power adaptation, as compared to phase rotation only in the multimode precoding scheme. They used a heuristic approximation to compute the centroid rather than an exact derivation. Their approximation enjoyed a closed form expression, but it inevitably caused a performance degradation.

In this correspondence, we setup an optimization problem to compute the centroid (11), where the optimization is taken over the space

$\mathcal{U} = \bigcup_{m=1}^{N_t} \mathcal{U}_m$. The optimization is solved in two steps. First by confining the optimization space to be a subspace $\mathcal{U}_m$ for each $1 \leq m < N_t$, we can obtain $N_t - 1$ unitary matrices, with each one lying in a different subspace $\mathcal{U}_m$. For $m = N_t$, from the fact that $C_p(\mathbf{H}, \hat{\mathbf{V}}^k) = C_p(\mathbf{H}, \mathbf{I}_{N_t})$ we set $\hat{\mathbf{V}}^k = \mathbf{I}_{N_t}$. We then choose from those N_t optimized unitary matrices the one that gives the minimum $D(\hat{\mathbf{V}}^k)$ as the kth centroid.

To facilitate the solution of the optimization problem for each m such that $1 \leq m < N_t$, we parameterize the unitary matrix using the Given's angle rotation. For parametrization, we define a $N_t \times N_t$ complex rotation matrix $\mathbf{U}^{pq}(\phi_{p,q}, \sigma_{p,q})$ with $p < q$ and $\phi_{p,q}, \sigma_{p,q} \in [-\pi, \pi)$ as

$$
\mathbf{U}^{pq}_{jk}(\phi_{p,q}, \sigma_{p,q})
= \begin{cases}
1 & \text{if } j = k \text{ and } j \neq p, q \\
\cos(\phi_{p,q}) & \text{if } j = k \text{ and } j = p, q \\
-\sin(\phi_{p,q})\exp(-i\sigma_{p,q}) & \text{if } j = p \text{ and } k = q \\
\sin(\phi_{p,q})\exp(i\sigma_{p,q}) & \text{if } j = q \text{ and } k = p \\
0 & \text{otherwise.}
\end{cases}
$$

Then, any $N_t \times m$ $(m < N_t)$ unitary matrix $\mathbf{U}$ can be written as the product of the Given's rotation matrices and a diagonal matrix, i.e.,

$$
\mathbf{U} = \mathbf{U}_{\mathbf{\Lambda}} \prod_{p=1}^{N_t-1} \prod_{q=N_t}^{p+1} \mathbf{U}^{p,q}(\phi_{p,q}, \sigma_{p,q}) \cdot \begin{bmatrix} \mathbf{I}_m \\ \mathbf{0}_{(N_t-m) \times m} \end{bmatrix} \quad (12)
$$

where $\mathbf{U}_{\mathbf{\Lambda}} = \mathrm{diag}\left(\exp(i\delta_1), \ldots, \exp(i\delta_{N_t})\right)$.

Let Θ be the collection of N_t^2 parameters (σ, ϕ and δ). By this parametrization, the computation of centroid problem (11) becomes an unconstraint optimization over the parameter set Θ. To obtain the optimal solution, we randomly generate the initial values of the parameters and then update them along the direction of the gradient. The gradient with respect to the parameter set Θ is derived in Appendix B.

B. SNR Adaptive Codebook Design

In theory, there is an optimal codebook for each SNR point and the designed codebook shall vary with the SNR value. However, codebook for each SNR is impossible as it causes too much overhead in design as well as in adaptive feedback practice. As the operating SNR of a system may drift over time, we take the approach to design a codebook which works well for a range of SNR values. We call this SNR adaptive codebook design in this correspondence. In practice, we can partition the dynamic range of SNR of a certain system into smaller regions and apply the SNR adaptive codebook design for each region. As long as the operating SNR of the system does not change very fast, we can designate a codebook to use with little overhead incurred.

The goal of designing the codebook that works for a range of SNR can be achieved with slight modifications of the proposed algorithm. First, Step 1 of the algorithm needs to be modified. We assume the instaneous operating SNR is a random variable taking values in a range according to a given distribution. Then, in Step 1 of the algorithm, instead of specifying a single power P_t for all randomly generated channel matrices, we randomly generate P_t, and use, according to the given distribution for each channel matrix $\mathbf{H}$. Second, to strike a balance among different SNR values, the normalized capacity difference, i.e., $d'(\mathbf{H}, \hat{\mathbf{V}}^k) = 1 - C_p(\mathbf{H}, \hat{\mathbf{V}}^k)/C_w(\mathbf{H})$, is employed as the distortion measure.

V. NUMERICAL RESULTS

In Fig. 1, the normalized (with respect to the optimal WF capacity) average channel capacity with feedback is shown for 4×4 MIMO system. Our codebook is designed according to the proposed algorithm
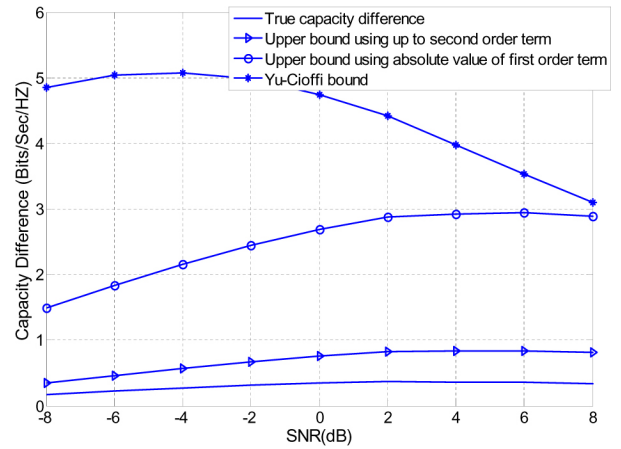

Fig. 1. Bounds on asymptotic capacity loss of multimode precoding.

TABLE I
CODEWORD ALLOCATION IN DIFFERENT MODES

SNR(dB)	mode 1	mode 2	mode 3	mode 4
-10	6	2	0	0
-5	1	7	0	0
0	0	2	6	0
5	0	0	8	0
10	0	0	7	1

given in Section IV. In the design, 1000 channel matrices $\mathbf{H}$ were randomly generated as the training sequence. The scheme by Lau, Liu, and Chen (covariance feedback scheme) in [15] and the multimode precoding scheme by Love and Heath in [16] have also been implemented and their performances are depicted for comparison purpose. In the multimode precoding by Love and Heath, the applied codebook distribution is the one derived from the capacity allocation criterion in [16] for fair comparison. For example, for 3 bit feedback case, the number of codewords lies in $\mathcal{U}_1, \mathcal{U}_2, \mathcal{U}_3$, and $\mathcal{U}_4$ is respectively 4, 3, 0, and 1. The codebook by Love and Heath is generated so as to minimize the Fubini-study distance of each mode as in [16]. The codebook in covariance feedback scheme is designed according to [15].

From the simulation results, we see our designed codebook outperforms the other two schemes throughout all SNR regions. The gain over the codebook by Love and Heath comes from the fact that the codebook distribution (shown in Table I) according to the proposed algorithm is closer to optimal as compared to the derived one in [16]. Meanwhile, our designed codebook based on a suboptimal multimode precoding scheme surprisingly outperforms the optimal WF based covariance feedback scheme. We claim that the gain comes from the more exact computation of the centroid.

In Fig. 3, we depict the normalized forward channel capacity of our codebook design in the case when the operating SNR of the system is set to vary over a range. We also compare our multimode precoding scheme with the one by Love and Heath for fair comparison. In our SNR adaptive codebook design, the SNR of the system is assumed to be uniformly distributed from -3 to 3 dB. Different from the previous setting, the training sequence used to design the codebook has 10 000 channel matrices. Seen from Fig. 3, our adaptive codebook design again outperforms the multimode precoding scheme by Love and Heath for
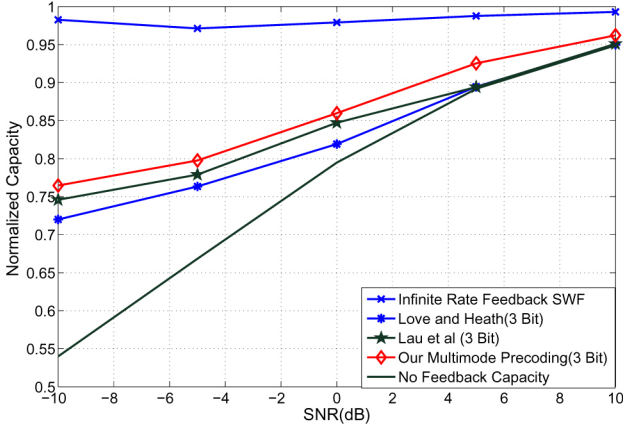
Fig. 2. Capacity comparison of our multimode precoding with covariance feedback [15] and the multimode precoding by Love and Heath [16] for 4×4 MIMO system.
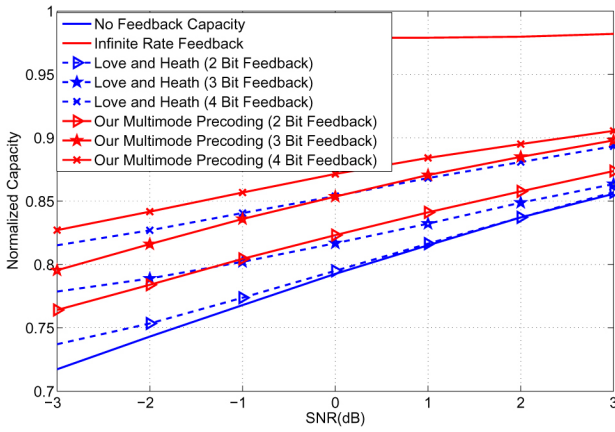

Fig. 3. Capacity comparison of our SNR adaptive multimode precoding codebook with the multimode precoding by Love and Heath [16] for 4×4 MIMO system.

the same feedback rate, although the difference is smaller as compared to that in Fig. 2. In some SNR region, even the performance of a lower rate feedback adaptive codebook surpasses that of a higher rate multimode scheme. For example, our 2-bit feedback outperforms 3-bit feedback multimode precoding scheme by Love and Heath at SNR from -1 to 3 dB.

VI. CONCLUSION

We have considered the problem of designing the feedback codebook under the feedback channel rate constraint under multimode precoding scheme. A number of bounds on the asymptotic capacity loss of multimode precoding have been obtained in this correspondence. These bounds explain why multimode precoding is close to optimal WF in throughput. We have then utilized the generalized Lloyd algorithm to design the optimal limited feedback codebook with respect to a distortion measure—the loss in forward channel capacity as compared to the optimal WF. In each iteration of the Lloyd algorithm, the codebook is constructed by a gradient search method applied on Given's angle parameterization of the unitary matrix (orthonormal column matrix). The proposed algorithm can also be adaptive to the SNR value with slight modifications. Numerical results showed that the proposed algorithm outperforms comparable algorithms reported in the literature.

APPENDIX A
PROOF OF THEOREM 1

For all $g_i \in \mathcal{E}$ we have $g_i < \mu$, i.e., for all $i \leq m^*$, $|g_i - \bar{g}/\mu| < 1$. Applying the Taylor series expansion to (6), we have $-\ln(1 + g_i - \bar{g}/\mu) = -\sum_{i=1}^{\infty} (g_i - \bar{g})^i / i \cdot \mu^i$. In this series, the absolute value of ith-order term is greater than the summation from the $(i+1)$th-order term up. By this property, both the second-order and first-order upperbound in Theorem 1 can be obtained trivially.

APPENDIX B
COMPUTATION OF THE GRADIENT

As the partial derivative is a linear operator, we can exchange the order of expectation and partial derivative to compute the gradient of the partial distortion $D(\hat{\mathbf{V}}^k)$ with respect to Θ, i.e.,

$$\nabla_\Theta D(\hat{\mathbf{V}}^k) = \nabla_\Theta \left\{ \mathbb{E}_{\mathbf{H} \in \mathcal{R}^k} \left[C_w - C_p \left(\mathbf{H}, \mathbf{U}_m(\Theta) \right) \right] \right\}$$
$$= -\mathbb{E}_{\mathbf{H} \in \mathcal{R}^k} \left\{ \nabla_\Theta C_p(\mathbf{H}, \mathbf{U}_m(\Theta)) \right\}.$$

If $\Theta = \delta_k$ with $1 \leq k \leq N_t$, to compute $\nabla_\Theta C_p(\mathbf{H}, \mathbf{U}_m(\Theta))$, we rewrite the Given's parametrization (12) as $\mathbf{U}_m = \mathbf{U}_\Lambda \mathbf{U}_{+1}$, where

$$\mathbf{U}_{+1} = \prod_{p=1}^{N_t-1} \prod_{q=N_t}^{p+1} \mathbf{U}^{p,q}(\phi_{p,q}, \sigma_{p,q}) \begin{bmatrix} \mathbf{I}_m \\ \mathbf{0}_{(N_t-m) \times m} \end{bmatrix}.$$

As $C_p\left(\mathbf{H}, \mathbf{U}_m(\Theta)\right)$ depends on δ_k only through $\mathbf{U}_\Lambda$, using the results (Theorem 2 and Lemma 3) in [23], we have $\nabla_\Theta C_p(\mathbf{H}, \mathbf{U}_m(\Theta)) = 2\mathrm{Re}\mathrm{Tr}\left\{ \mathbf{H}^*\mathbf{H}\mathbf{U}_m\mathbf{E}\mathbf{U}_{+1}^* \times \nabla_\Theta \mathbf{U}_\Lambda^* \right\}$ where $\mathbf{E} = (m/N_t\mathbf{I}_m + \mathbf{U}_m^*\mathbf{H}^*\mathbf{H}\mathbf{U}_m)^{-1}$ is the MMSE matrix defined in [23] and

$$\nabla_\Theta \mathbf{U}_\Lambda^* |_{\Theta = \delta_k} = -i \exp(-i\delta_k) \mathbf{e}_k \mathbf{e}_k^*,$$

with $\mathbf{e}_k$ being a unit norm column vector of length N_t with the kth element being 1.

To compute the gradient $\nabla_\Theta C_p(\mathbf{H}, \mathbf{U}_m(\Theta))$ when $\Theta = \phi_{p,q}$ or $\Theta = \sigma_{p,q}$, we rewrite $\mathbf{U}_m$ as

$$\mathbf{U}_m = \mathbf{U}_\Lambda \prod_{p=1}^{N_t-1} \prod_{q=N_t}^{p+1} \mathbf{U}^{p,q}(\phi_{p,q}, \sigma_{p,q}) \begin{bmatrix} \mathbf{I}_m \\ \mathbf{0}_{(N_t-m) \times m} \end{bmatrix}$$
$$= \underbrace{\mathbf{U}_\Lambda \mathbf{U}^{N_t-1, N_t}(\phi_{N_t-1, N_t}, \sigma_{N_t-1, N_t}) \cdots}_{\mathbf{U}_{-1}}$$
$$\underbrace{\mathbf{U}^{p,q}(\phi_{p,q}, \sigma_{p,q})}_{\mathbf{U}_0}$$
$$\underbrace{\cdots \mathbf{U}^{1, N_t}(\phi_{1, N_t}, \sigma_{1, N_t}) \begin{bmatrix} \mathbf{I}_m \\ \mathbf{0}_{(N_t-m) \times m} \end{bmatrix}}_{\mathbf{U}_{+1}}.$$

The input–output relationship with this precoding matrix $\mathbf{U}_m$ is then $\mathbf{r} = \mathbf{H}\mathbf{U}_{-1}\mathbf{U}_0\mathbf{U}_{+1}\mathbf{s}_m + \mathbf{n}$. Use the results in [23] again, we obtain $\nabla_\Theta C_p(\mathbf{H}, \mathbf{U}_m(\Theta)) = 2\mathrm{Re}\mathrm{Tr}\left\{ \mathbf{U}_{-1}^*\mathbf{H}^*\mathbf{H}\mathbf{U}_m\mathbf{E}\mathbf{U}_{+1}^* \times \nabla_\Theta \mathbf{U}_0^* \right\}$, where

$$\nabla_\Theta \mathbf{U}_0^* |_{\Theta = \phi_{p,q}}$$
$$= \begin{cases} -\sin(\phi_{p,q}), & \text{if } j = k \text{ and } j = p, q \\ \cos(\phi_{p,q}) \exp(-i\sigma_{p,q}), & \text{if } j = p \text{ and } k = q \\ -\cos(\phi_{p,q}) \exp(i\sigma_{p,q}), & \text{if } j = q \text{ and } k = p \\ 0, & \text{otherwise} \end{cases}$$

and

$$\nabla_\Theta \mathbf{U}_0^* |_{\Theta=\sigma_{p,q}}$$
$$= \begin{cases} -i\sin(\phi_{p,q})\exp(-i\sigma_{p,q}), & \text{if } j = p \text{ and } k = q \\ -i\sin(\phi_{p,q})\exp(i\sigma_{p,q}), & \text{if } j = q \text{ and } k = p \\ 0, & \text{otherwise.} \end{cases}$$

REFERENCES

[1] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.

[2] C. E. Shannon, "Channels with side information at the transmitter," *IBM J. Res. Develop.*, vol. 2, pp. 289–293, Oct. 1958.

[3] A. J. Goldsmith and P. Varaiya, "Capacity, mutual information, and coding for finite-state markov channels," *IEEE Trans. Inf. Theory*, vol. 42, no. 3, pp. 868–886, May 1996.

[4] A. J. Goldsmith and P. Varaiya, "Capacity of fading channels with channel side information," *IEEE Trans. Inf. Theory*, vol. 43, pp. 1986–1992, Nov. 1997.

[5] G. Caire and S. S. Shamai, "On the capacity of some channels with channel state information," *IEEE Trans. Inf. Theory*, vol. 45, no. 6, pp. 2007–2019, Sep. 1999.

[6] H. Viswanathan, "Capacity of markov channels with receiver CSI and delayed feedback," *IEEE Trans. Inf. Theory*, vol. 45, no. 2, pp. 761–771, Mar. 1999.

[7] A. J. Goldsmith and S.-G. Chua, "Variable-rate variable-power MQAM for fading channels," *IEEE Trans. Commun.*, vol. 46, pp. 595–602, May 1998.

[8] D. L. Goeckel, "Adaptive coding for time-varying channels using outdated fading estimates," *IEEE Trans. Commun.*, vol. 47, pp. 844–855, Jun. 1999.

[9] A. Narula, M. J. Lopez, M. D. Trott, and G. W. Wornell, "Efficient use of side information in multiple-antenna data transmission over fading channels," *IEEE J. Sel. Areas Commun.*, vol. 16, no. 8, pp. 1423–1436, Oct. 1998.

[10] K. Mukkavilli, A. Sabharwal, E. Erkip, and B. Aazhang, "On beamforming with finite rate feedback in multiple-antenna systems," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2562–2579, Oct. 2003.

[11] D. J. Love, R. W. Heath, and T. Strohmer, "Grassmannian beamforming for multiple-input multiple-output wireless systems," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2735–2747, Oct. 2003.

[12] P. Xia and G. Giannakis, "Design and analysis of transmit-beamforming based on limited-rate feedback," *IEEE Trans. Signal Process*, vol. 54, no. 5, pp. 1853–1863, May 2006.

[13] J. C. Roh and B. D. Rao, "Transmit beamforming in multiple-antenna systems with finite rate feedback: A vq-based approach," *IEEE Trans. Inf. Theory*, vol. 52, no. 3, pp. 1101–1112, Mar. 2006.

[14] I. E. Telatar, "Capacity of multi-antenna Gaussian channels," *Eur. Trans. Telecommun.*, vol. 10, no. 6, pp. 585–595, Nov./Dec. 1999.

[15] V. K. N. Lau, Y. Liu, and T.-A. Chen, "On the design of MIMO block-fading channels with feedback-link capacity constraint," *IEEE Trans. Commun.*, vol. 52, pp. 62–70, Jan. 2004.

[16] D. J. Love and J. R. W. Heath, "Multimode precoding for MIMO wireless systems," *IEEE Trans. Signal Process*, vol. 53, no. 10, pp. 3674–3687, Oct. 2005.

[17] J. Roh and B. Rao, "Design and analysis of MIMO spatial multiplexing systems with quantized feedback," *IEEE Trans. Signal Process.*, vol. 54, no. 8, pp. 2874–2886, Aug. 2006.

[18] S. Lloyd, "Least squares quantization in pcm," *IEEE Trans. Inf. Theory*, vol. 28, pp. 129–137, Mar. 1982.

[19] W. Yu and J. Cioffi, "Constant-power waterfilling: Performance bound and low-complexity implementation," *IEEE Trans. Commun.*, vol. 54, no. 1, pp. 23–28, Jan. 2006.

[20] S. Zhou and B. Li, "BER criterion and codebook construction for finite-rate precoded spatial multiplexing with linear receivers," *IEEE Trans. Signal Process*, vol. 54, no. 5, pp. 1653–1665, May 2006.

[21] V. K. N. Lau, Y. Liu, and T.-A. Chen, "Capacity of memoryless channels and block-fading channels with designable cardinality-constrained channel state feedback," *IEEE Trans. Inf. Theory*, vol. 50, no. 9, pp. 2038–2049, Sep. 2004.

[22] A. Gersho and R. Gray, *Vector Quantization and Signal Compression*. Norwell, MA: Kluwer Academic, 1992.

[23] D. Palomar and S. Verdu, "Gradient of mutual information in linear vector Gaussian channels," *IEEE Trans. Inf. Theory*, vol. 52, no. 1, pp. 141–154, Jan. 2006.

An Efficient Water-Filling Solution for Linear Coherent Joint Estimation

Wenbin Guo, *Member, IEEE*, Jin-Jun Xiao, *Member, IEEE*, and Shuguang Cui, *Member, IEEE*

Abstract—We consider a sensor network with distributed sensors observing multiple sources and transmitting local observations over a Gaussian multiple access channel to a fusion center where the signals are coherently combined. We develop a joint estimation method based on the linear analog forwarding scheme at each local node, which can be cast as a nonlinear optimization problem. By minimizing the gap to the performance benchmark, we obtain a closed-form solution that follows the water-filling strategy. Numerical results are given to demonstrate the average estimation performance in a fading communication environment.

Index Terms—Linear coherent estimation, water-filling, wireless sensor networks.

I. INTRODUCTION

Consider a sensor network where sensors observe multiple sources and transmit the local observations to a fusion center. The fusion center aggregates the received data to generate a final estimate. We assume that the sensors encode their observations in a distributed way, since in many applications cooperative encoding is infeasible. For example, in a battleground scenario, there may be multiple sensors tracking the movement of the enemy tank and sending their observations to a commander. The cooperative communications between sensors would cost local energy and increase the system complexity, which may not be desirable. Therefore, distributed (noncooperative locally) schemes that achieve acceptable performance are of more practical importance.

In wireless sensor networks, the limited energy and communication bandwidth are two main constraints. Specifically, due to the difficulty of changing sensor batteries, low power consumption is important to guarantee a long lifetime for a sensor network. Communication bandwidth becomes a precious resource when a large number of sensors need to communicate with the fusion center under a strict delay requirement. There are two ways to model the communication bandwidth constraint. One is to directly limit the transmitted bits of each sensor node, which usually applies to a digital communication scheme, for which some of the recent work in the literature can be found in [1]–[7]. The second one is to limit the number of real messages sent from each sensor, which applies to analog forwarding schemes [8]. The analog forwarding schemes with linear processing at both local sensors and the