

# Interference Channel Sum Rate Optimization on the Grassmann Manifold

Mohsen Rezaee and Maxime Guillaud

Institute of Telecommunications

Vienna University of Technology

Email: {mohsen.rezaee,guillaud}@nt.tuwien.ac.at

**Abstract**—The problem of truncated unitary precoder design for the Multiple-Input Multiple-Output Interference Channel (MIMO-IC) is investigated. A simple gradient ascent method is utilized over the Grassmann manifold to maximize the sum rate achieved over the MIMO-IC. Since such optimization methods are not guaranteed to converge to the global solution, we also introduce an implementation of simulated annealing for the problem at hand and compare the performance of the two methods. The results show that the gradient on the Grassmann manifold provides gains over existing methods such as max-SINR, in particular when asymmetric channels are considered and the achieved performance is comparable to that achieved by more complex methods not restricted to unitary constraints. Furthermore it performs close to the simulated annealing method at low SNR regime.

## I. INTRODUCTION

High spectral efficiency can be achieved by transmitting independent data streams simultaneously through the use of multiple-input multiple-output (MIMO) techniques that provide spatial multiplexing. Here we consider the MIMO Interference channel (IC). The capacity characterization of the IC is still an open problem even for two users and single antenna cases [1].

Several attempts have been made to find the optimum transmit precoders for the MIMO-IC. There are different approaches to the problem depending on the amount of channel knowledge available at the transmitters. The problem is modeled as a non-cooperative game in [2] and the iterative water filling algorithm is used at the transmitters to design the transmit covariance matrices. Authors in [3] use the Nash bargaining method and show that further gains can be achieved by enabling the transmitters to act as players in a cooperative game. The problem is also investigated in [4] using the framework of Bayesian games. However due to the nature of the problem which involves interference generated by transmitters, altruistic schemes like interference alignment [5] seem to perform better than greedy methods when interference is the dominant unwanted signal. The algorithms of [6], [7] use interference alignment as the main objective, which is asymptotically optimal for the interference channel but may have suboptimal sum rate at finite signal to noise ratio (SNR). Sum rate maximization is the final objective which several methods have been trying to achieve implicitly. For example leakage minimization or max-SINR algorithms in [6] are based on surrogate objective functions that make the problem more

tractable, however the performance of the presented algorithms is benchmarked in terms of sum rate. These methods may not be optimal under general conditions hence many approaches try to directly optimize the sum rate. Authors in [8] use the weighted sum mean square error objective function with a total power constraint. In [9] the authors try to find the best truncated unitary precoders that maximize sum rate while maintaining the interference alignment conditions. Maximizing the sum rate under per-user power constraint using the gradient method is discussed in [10] without the constraint of unitary precoders.

In this paper we present two iterative methods to find truncated unitary precoders that maximize the sum rate of the IC. Orthogonality is generally desired in MIMO precoding designs to aid with the practical feedback of channel state. Comparison presented here to methods that consider general precoders shows that the performance loss is negligible by only optimizing truncated unitary precoders. Such a constraint allows us to use the Grassmann manifold as the search space for the gradient ascent method. Furthermore, using the properties of the Grassmann manifold, we propose another iterative algorithm based on simulated annealing which aims to find the global solution of the problem. The results show that the gradient algorithm performs close to the simulated annealing method at low SNR regime.

The paper is organized as follows. In Section II, the system model is described. The methods based on gradient ascent and simulated annealing are given in Section III. Simulation results are presented in Section IV and some final conclusions are drawn in Section V.

The following notations are used in this paper. Normal letters represent scalar quantities, boldface lowercase letters indicate vectors and boldface uppercase letters designate matrices. The trace, conjugate, transpose, Hermitian transpose of a matrix or vector are represented by  $\text{tr}(\cdot)$ ,  $(\cdot)^*$ ,  $(\cdot)^T$ ,  $(\cdot)^H$  respectively.  $\text{Bdiag}(\cdot)$  represents a block diagonal matrix with the argument blocks on its diagonal.  $\mathbb{E}(\cdot)$  denotes the expectation operator.  $d_{\mathbf{X}}A$  denotes the differential of  $A$  resulting from a change in  $\mathbf{X}$ . The identity matrix is represented by  $\mathbf{I}$ .

## II. SYSTEM MODEL

Consider the interference channel in which  $K$  transmitters are willing to communicate with their corresponding receivers simultaneously. Each transmitter is equipped with  $M$  antennas

while each receiver has  $N$  antennas and the number of data symbols to be transmitted simultaneously at each transmitter is equal to  $d$ . The channels are assumed to be constant during the transmission.

The received vector at receiver  $i$  can be written as

$$\mathbf{y}_i = \mathbf{H}_{ii} \sqrt{\lambda_i} \mathbf{V}_i \mathbf{x}_i + \sum_{j=1, j \neq i}^K \mathbf{H}_{ij} \sqrt{\lambda_j} \mathbf{V}_j \mathbf{x}_j + \mathbf{n}_i, \quad (1)$$

in which  $\mathbf{H}_{ij} \in \mathbb{C}^{N \times M}$  is the channel matrix between transmitter  $j$  and receiver  $i$ ,  $\mathbf{V}_j \in \mathbb{C}^{M \times d}$  and  $\mathbf{x}_j \in \mathbb{C}^{d \times 1}$  are the precoding matrix and the data vector of transmitter  $j$  respectively.  $\mathbf{n}_i \in \mathbb{C}^{N \times 1}$  is the additive noise at receiver  $i$  which is modeled as a Gaussian complex circularly symmetric vector with zero mean and covariance matrix equal to  $\sigma^2 \mathbf{I}$ . Assuming  $\mathbb{E}(\mathbf{x}_i \mathbf{x}_i^H) = \mathbf{I}_d$ ,  $i = 1, \dots, K$ , the transmit power for user  $i$  equals  $\lambda_i \text{tr}(\mathbf{V}_i^H \mathbf{V}_i)$ .

Assuming Gaussian signaling, the sum of mutual information between each pair of transmitter and receivers is written as

$$R_{\text{sum}}(\mathbf{V}_1, \dots, \mathbf{V}_K) = \sum_{i=1}^K \log_2 |\mathbf{Q}_i| - \log_2 |\hat{\mathbf{Q}}_i|, \quad (2)$$

where

$$\mathbf{Q}_i = \sum_{j=1}^K \lambda_j \mathbf{H}_{ij} \mathbf{V}_j \mathbf{V}_j^H \mathbf{H}_{ij}^H + \sigma_n^2 \mathbf{I}, \quad (3)$$

$$\hat{\mathbf{Q}}_i = \sum_{j=1, j \neq i}^K \lambda_j \mathbf{H}_{ij} \mathbf{V}_j \mathbf{V}_j^H \mathbf{H}_{ij}^H + \sigma_n^2 \mathbf{I}. \quad (4)$$

Therefore we consider the following optimization problem,

$$\begin{aligned} & \max_{\mathbf{V}_1, \dots, \mathbf{V}_K} R_{\text{sum}}(\mathbf{V}_1, \dots, \mathbf{V}_K) \\ & \text{s.t.} \quad \mathbf{V}_j^H \mathbf{V}_j = \mathbf{I}_d \quad \forall j = 1, \dots, K. \end{aligned} \quad (5)$$

### III. ITERATIVE OPTIMIZATION

We present two iterative methods exploiting the structure of the Grassmann manifold to find precoding matrices that locally maximize the sum mutual information. We are interested in cases where  $d < M$ . It is easy to check that unitary transformation of the precoding matrices will not change the sum rate, i.e.,  $R_{\text{sum}}(\mathbf{V}_1, \dots, \mathbf{V}_K) = R_{\text{sum}}(\mathbf{V}_1 \mathbf{C}_1, \dots, \mathbf{V}_K \mathbf{C}_K)$  where each  $\mathbf{C}_i$  is a  $d \times d$  unitary matrix. This property allows us to reduce the search space to the Grassmann manifold.

#### A. Gradient on the Grassmann Manifold

In this section we try to formulate the problem in an appropriate way in order to apply the gradient ascent method on the Grassmann manifold. First we define the following matrices

$$\mathbf{V} = \text{Bdiag}(\mathbf{V}_1, \mathbf{V}_2, \dots, \mathbf{V}_K), \quad (6)$$

$$\mathbf{D} = \text{Bdiag} \left( \frac{\partial}{\partial \mathbf{V}_1^*} R_{\text{sum}}, \dots, \frac{\partial}{\partial \mathbf{V}_K^*} R_{\text{sum}} \right), \quad (7)$$

in which  $\frac{\partial}{\partial \mathbf{V}_k^*} R_{\text{sum}}$  is the partial derivative of the sum rate w.r.t  $\mathbf{V}_k^*$  (evaluated in the Appendix). The gradient w.r.t the matrix  $\mathbf{V}$  is

$$\nabla_{\mathbf{V}} R_{\text{sum}} = 2 \frac{\partial}{\partial \mathbf{V}^*} R_{\text{sum}} = 2\mathbf{D}, \quad (8)$$

where the second equality is proved in the Appendix. The resulting gradient should be projected on the tangent space of the Grassmann manifold in order to find the gradient direction on the manifold [11],

$$\nabla_{\mathbf{V}}^G R_{\text{sum}} = (\mathbf{I} - \mathbf{V}\mathbf{V}^H) \nabla_{\mathbf{V}} R_{\text{sum}}. \quad (9)$$

$\nabla_{\mathbf{V}}^G R_{\text{sum}}$  is tangent to the manifold and moving along this direction preserves the orthonormality constraint. In order to move along the geodesic corresponding to the resulting tangent direction we have to evaluate the compact singular value decomposition (SVD) of the tangent direction in  $\mathbf{V}$ ,

$$\nabla_{\mathbf{V}}^G R_{\text{sum}} = \mathbf{U}\mathbf{\Sigma}\mathbf{W}^H. \quad (10)$$

At each iteration, a new point is generated by moving along the corresponding geodesic which is parameterized as

$$\mathbf{V}(t) = \mathbf{V}\mathbf{W} \cos(\mathbf{\Sigma}t) \mathbf{W}^H + \mathbf{U} \sin(\mathbf{\Sigma}t) \mathbf{W}^H, \quad t \in \mathbb{R} \quad (11)$$

in which  $\sin(\cdot)$  and  $\cos(\cdot)$  are applied component-wise on the diagonal of their argument matrices. Note that  $\mathbf{V}(t)$  is always block diagonal because we always move on this manifold along the gradient and due to the block diagonal structure of  $\mathbf{V}$  and  $\nabla_{\mathbf{V}} R_{\text{sum}}$ , the resulting gradient matrix on the Grassmann manifold,  $\nabla_{\mathbf{V}}^G R_{\text{sum}}$  also is block diagonal and the SVD operation also does not change this property. Since  $\mathbf{V}$ ,  $\mathbf{W}$  and  $\mathbf{U}$  are all block diagonal, at every iteration we get a new block diagonal matrix  $\mathbf{V}$ . Therefore it is clear that  $\mathbf{V}(t)$  belongs to a Cartesian product of  $K$  Grassmann manifolds.

The step size  $t$  should be set in a proper way to ensure improvement in the objective function at each step. In order to achieve this we use Armijo's rule [12] to have an increase at least as large as a fraction of the increase predicted by the first order Taylor expansion of the sum rate: The step size is chosen as  $t = \beta^{m_0}$  for some fixed scalars  $0 < \alpha, \beta < 1$  such that  $m_0$  is the smallest non-negative integer  $m$  that satisfies

$$R_{\text{sum}}^+ - R_{\text{sum}} \geq \alpha \beta^{m_0} \text{tr} \left( \left( \frac{\partial}{\partial \mathbf{V}^*} R_{\text{sum}} \right)^T (\mathbf{V}^+ - \mathbf{V})^* \right), \quad (12)$$

in which  $\mathbf{V}^+ = \mathbf{V}(\beta^{m_0})$  and  $R_{\text{sum}}$  and  $R_{\text{sum}}^+$  are the sum rate calculated at the points  $\mathbf{V}$  and  $\mathbf{V}^+$  respectively. The RHS is a fraction of the differential of the sum rate when we move from  $\mathbf{V}$  to  $\mathbf{V}^+$  and it is proportional to the first order term in the Taylor expansion of  $R_{\text{sum}}^+$  around  $\mathbf{V}$ .

Since the SVD operates on each block separately, we will have the same procedure for each block, therefore (12) yields

$$\mathbf{V}_k(t) = \mathbf{V}_k \mathbf{W}_k (\cos \mathbf{\Sigma}_k t) \mathbf{W}_k^H + \mathbf{U}_k (\sin \mathbf{\Sigma}_k t) \mathbf{W}_k^H, \quad \forall k \quad (13)$$

in which  $\mathbf{U}_k \Sigma_k \mathbf{W}_k^H$  represents the compact SVD of the  $k$ th block of  $\nabla_{\mathbf{V}}^G R_{\text{sum}}$  and is equal to  $(\mathbf{I} - \mathbf{V}_k \mathbf{V}_k^H) \nabla_{\mathbf{V}_k} R_{\text{sum}}$ . Equation (12) simplifies to

$$R_{\text{sum}}^+ - R_{\text{sum}} \geq \alpha \beta^m \sum_{k=1}^K \text{tr} \left( \left( \frac{\partial}{\partial \mathbf{V}_k^*} R_{\text{sum}} \right)^T (\mathbf{V}_k^+ - \mathbf{V}_k)^* \right), \quad (14)$$

with  $\mathbf{V}_k^+ = \mathbf{V}_k(\beta^m)$ . It is clear that every user who has access to global channel information can optimize its precoding vector using this gradient approach. Note that all users share a common step size  $t = \beta^{m_0}$  because they are optimizing the same objective function. Algorithm 1 summarizes the method.

---

**Algorithm 1** Gradient on the Grassmann manifold

---

- \* Initialize precoders with random orthonormal matrices  $\mathbf{V}_k$  for  $k = 1, \dots, K$
- Repeat**
  - \* Calculate the gradient  $\nabla_{\mathbf{V}} R_{\text{sum}}$  at point  $\mathbf{V}$  as in (8)
  - \* Compute the SVD of  $(\mathbf{I} - \mathbf{V}\mathbf{V}^H) \nabla_{\mathbf{V}} R_{\text{sum}}$
  - \* Choose the step size according to (12) for given  $\alpha, \beta$
  - \* Move on the geodesics according to (11)
  - \*  $\mathbf{V} \leftarrow \mathbf{V}^+ = \mathbf{V}(\beta^m)$

**until** convergence

---

Note that Algorithm 1 is guaranteed to converge since  $R_{\text{sum}}$  is non-decreasing (and can be upper-bounded).

*B. Simulated Annealing Using the Grassmann Manifold*

Simulated annealing is a well-known method to deal with non-convex optimization problems. In several problems this method is shown to converge to a global optimum solution with high probability [13]. The computational burden of this method is much higher than other deterministic optimization methods that find a locally optimum solution such as the gradient method presented above. This computational complexity restricts the use of this method for precoder design in which the space under exploration can be very large specially when there are multiple users. Here we present an application of simulated annealing restricted to the Grassmann manifold.

Unlike the usual descent methods which decrease the objective function at each iteration, the simulated annealing method tries to find the global optimum by making choices that do not necessarily improve the objective function at each step. The algorithm starts at some arbitrary initial point. At each iteration, the objective function is evaluated at several points in the vicinity of the current point. If a new point has better objective function then it will be accepted, otherwise it will be accepted with some probability  $p$ ,

$$p = \exp \left( \frac{f_{\text{new}} - f_{\text{old}}}{T} \right), \quad (15)$$

in which  $f_{\text{old}}, f_{\text{new}}$  denote the objective function (in our case, the sum rate) computed at the current point and new point respectively and  $T$  is a parameter that decreases over the iterations, referred to as "temperature" by analogy to the physical

cooling process that inspired the method. After a number of trials, the best point is recorded and the temperature is reduced ( $\eta$  denotes the temperature reduction factor) and another set of trials begins with the new temperature. The iteration continues until the change in objective function remains negligible for a number of successive temperature reductions.

In order to adapt this model to our problem we generate random directions inside the tangent space of the current point (i.e., matrices with random entries that are projected into the tangent space of the current point) for each user and then the search in the vicinity is made by moving along the geodesics with some random step size. We use different step sizes for different users denoted by  $\mathbf{t} = [t_1, t_2, \dots, t_K]$ . To generate new points, elements of the step size vector are produced randomly inside the interval of  $[-a, a]$  in which  $a$  is initially set to 1 and is adjusted frequently to maintain the probability of acceptance around  $p = 0.5$  at a given temperature [13]. The detailed procedure is presented in Algorithm 2.

---

**Algorithm 2** Simulated annealing over the Grassmann manifold

---

- \* Initialize precoders with random orthonormal matrices  $\mathbf{V}_k$  for  $k = 1, \dots, K$
- \* Initialize the temperature  $T \leftarrow T_0$
- \* Let  $N_s$  denote the number of trials for a given step size
- \* Let  $N_r$  denote the number of step size adjustments
- Repeat**
  - \*  $n_r \leftarrow 0$
  - While**  $n_r < N_r$ 
    - \*  $n_s \leftarrow 0$
    - While**  $n_s < N_s$ 
      - \* Generate random directions in the tangent space of all  $\mathbf{V}_k$  and then calculate the SVD.
      - \* Generate a random step size vector  $\mathbf{t} = [t_1, t_2, \dots, t_K]$ .
      - \* Calculate the new candidate points  $\mathbf{V}_k(t_k)$  according to (13)
      - \* Compute  $f_{\text{new}} = R_{\text{sum}}(\mathbf{V}_1(t_1), \dots, \mathbf{V}_K(t_K))$ .
      - \* Accept the new point with probability  $\max\{p, 1\}$  according to (15).
      - \*  $n_s \leftarrow n_s + 1$ .
    - EndWhile**
    - \* Adjust the step size vector variance.
    - \*  $n_r \leftarrow n_r + 1$ .
  - EndWhile**
  - \*  $\mathbf{V} \leftarrow \mathbf{V}^{\text{opt}}$  (the best point found at this temperature).
  - \*  $T \leftarrow \eta \times T$ .

**until** convergence

---

IV. SIMULATION RESULTS

In this section the performance of the proposed schemes is compared to existing methods. In the gradient method we have used  $\alpha = 0.1, \beta = 0.5$  in order to have a good convergence behavior. In the simulated annealing method the initial temperature is chosen as the absolute value of the average of the objective function computed at a number of randomly

selected points in the design space [14] and  $\eta$  is set to 0.9. The performance metric is the ergodic sum rate which is evaluated by averaging (2) over many channel realizations using the optimized precoders. The channel between each transmitter and each receiver is a matrix whose elements are modeled as independent Gaussian random variables with zero mean and unit variance. The performance of the proposed schemes is compared to the leakage minimization (denoted by IA) and max-SINR algorithms proposed in [6] (using equal-power beams) and also to the gradient method presented in [10] which is not restricted to orthonormal precoders. The max-SINR algorithm also gives matrices which are not necessarily orthogonal. Fig. 1 shows the performance comparison for the 3-user IC with 2 antennas per node and 1 data stream for each user. It is clear from Fig. 1 that the simulated annealing method matches the slope of IA and the sum rate offset at higher SNRs better than the gradient algorithm. However it shows that the gradient algorithm is not far from the optimum (represented here by simulated annealing) at least in the low SNR regime. Fig. 2 shows the sum rate for the 3-user IC with 4 antennas per node and 2 data streams for each user. Here each precoding matrix has two columns which allows for power allocation or correlation between the columns. Our gradient scheme performs close to other methods which employ more general precoders. In this case the simulated annealing method does not provide significant gain over the other methods due to the increase in the dimensions of the search space. The sum rate is plotted in Fig. 3 for the case of unequal transmit powers across the users for 3 users and 4 antennas per node. Here  $P_1, P_2, P_3$  are considered to be -10, 0, 10 dB respectively and the performance is plotted versus the noise power. Sum rate gradient based methods outperform other methods based on surrogate objective functions. Another interesting case is considered in Fig. 4 in which the receivers have different noise variances of -10, 0, 10 dB and the sum rate is plotted versus the transmit power which is the same for all users. This scenario arises when some users undergo different amount of uncoordinated interference such as cell-edge users in a cellular environment. It is clear that the gradient provides considerable gains for such non-symmetric cases w.r.t. algorithms based on other objectives than sum rate (leakage, SINR, ...).

## V. CONCLUSION

Two iterative methods were proposed to find orthonormal precoders that maximize the sum throughput over the MIMO-IC. The methods exploit the properties of the Grassmann manifold. The performance of the gradient method was presented for different scenarios. The results show that the performance loss due to the constraint of unitary precoders is negligible. Also we observe that the solution obtained by the gradient algorithm performs close to that obtained by simulated annealing specially at low SNR regime.

## APPENDIX

Starting from the following fact

$$d_{\mathbf{X}}(\ln |\mathbf{X}|) = \text{tr}(\mathbf{X}^{-1} d_{\mathbf{X}}(\mathbf{X})) \quad (16)$$

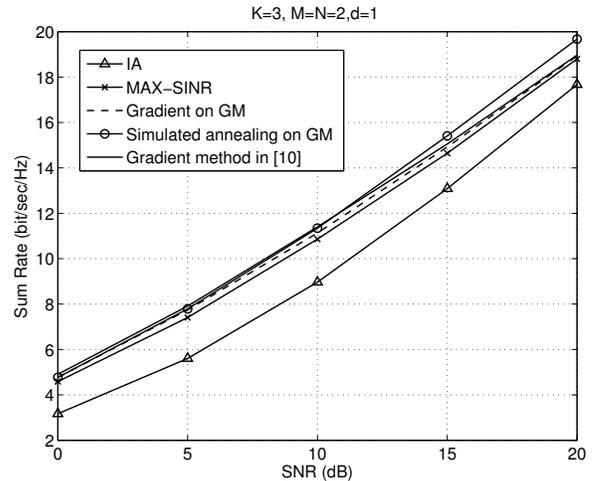


Fig. 1. Sum rate comparison for  $M = N = 2$

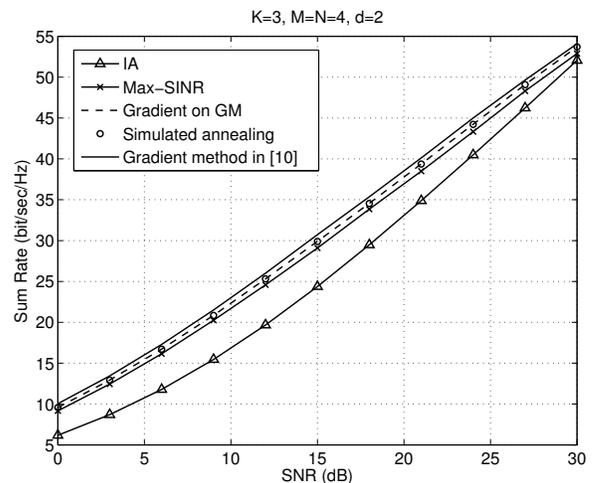


Fig. 2. Sum rate comparison for  $M = N = 4$

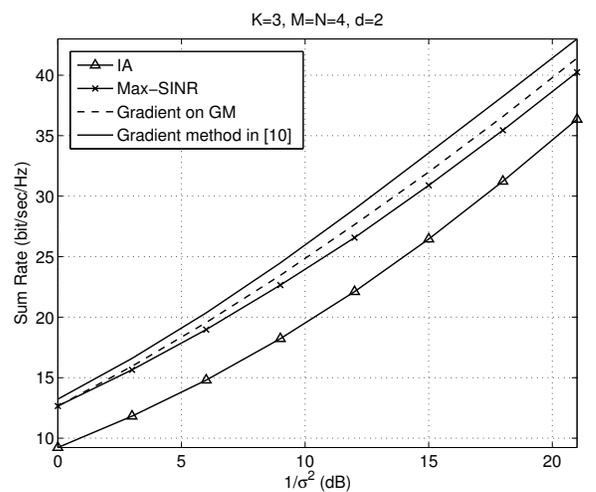


Fig. 3. Sum rate comparison for  $M = N = 4$  (unequal powers)

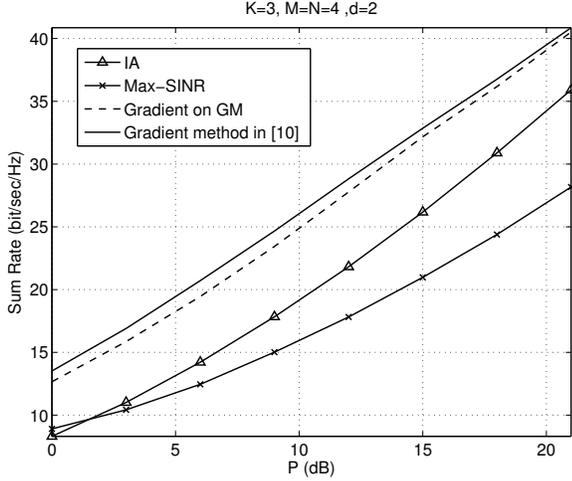


Fig. 4. Sum rate comparison for  $M = N = 4$  (different noise powers)

we write the differential of the sum rate w.r.t  $\mathbf{V}_k$

$$\begin{aligned}
d_{\mathbf{V}_k} R_{sum} &= \sum_{i=1}^K \text{tr} \left( \mathbf{Q}_i^{-1} d_{\mathbf{V}_k} (\mathbf{Q}_i) \right) \\
&\quad - \sum_{i=1, i \neq k}^K \text{tr} \left( \hat{\mathbf{Q}}_i^{-1} d_{\mathbf{V}_k} (\hat{\mathbf{Q}}_i) \right) \\
&= \lambda_k \sum_{i=1}^K \text{tr} \left( \mathbf{Q}_i^{-1} \mathbf{H}_{ik} \mathbf{V}_k d_{\mathbf{V}_k} (\mathbf{V}_k^H) \mathbf{H}_{ik}^H \right) \\
&\quad - \lambda_k \sum_{i=1, i \neq k}^K \text{tr} \left( \hat{\mathbf{Q}}_i^{-1} \mathbf{H}_{ik} \mathbf{V}_k d_{\mathbf{V}_k} (\mathbf{V}_k^H) \mathbf{H}_{ik}^H \right) \\
&= \lambda_k \sum_{i=1}^K \text{tr} \left( \mathbf{H}_{ik}^H \mathbf{Q}_i^{-1} \mathbf{H}_{ik} \mathbf{V}_k d_{\mathbf{V}_k} (\mathbf{V}_k^H) \right) \\
&\quad - \lambda_k \sum_{i=1, i \neq k}^K \text{tr} \left( \mathbf{H}_{ik}^H \hat{\mathbf{Q}}_i^{-1} \mathbf{H}_{ik} \mathbf{V}_k d_{\mathbf{V}_k} (\mathbf{V}_k^H) \right) \\
&= \text{tr} \left( \lambda_k \left[ \sum_{i=1}^K \left( \mathbf{H}_{ik}^H \mathbf{Q}_i^{-1} \mathbf{H}_{ik} \mathbf{V}_k \right)^T \right. \right. \\
&\quad \left. \left. - \sum_{i=1, i \neq k}^K \left( \mathbf{H}_{ik}^H \hat{\mathbf{Q}}_i^{-1} \mathbf{H}_{ik} \mathbf{V}_k \right)^T \right] d_{\mathbf{V}_k} (\mathbf{V}_k^*) \right). \tag{17}
\end{aligned}$$

Comparing the result to the following equality

$$d_{\mathbf{V}_k} R_{sum} = \text{tr} \left( \left( \frac{\partial}{\partial \mathbf{V}_k^*} R_{sum} \right)^T d_{\mathbf{V}_k} \mathbf{V}_k^* \right), \tag{18}$$

we have

$$\begin{aligned}
&\frac{\partial}{\partial \mathbf{V}_k^*} R_{sum} \\
&= \lambda_k \left( \sum_{i=1}^K \mathbf{H}_{ik}^H \mathbf{Q}_i^{-1} \mathbf{H}_{ik} \mathbf{V}_k - \sum_{i=1, i \neq k}^K \mathbf{H}_{ik}^H \hat{\mathbf{Q}}_i^{-1} \mathbf{H}_{ik} \mathbf{V}_k \right). \tag{19}
\end{aligned}$$

The differential of the sum rate can be written as sum of the differentials w.r.t each  $\mathbf{V}_k$

$$\begin{aligned}
dR_{sum} &= \sum_{k=1}^K d_{\mathbf{V}_k} R_{sum} \\
&= \sum_{k=1}^K \text{tr} \left( \left( \frac{\partial}{\partial \mathbf{V}_k^*} R_{sum} \right)^T d_{\mathbf{V}_k} \mathbf{V}_k^* \right) \tag{20} \\
&= \text{tr} (\mathbf{D}^T d_{\mathbf{V}} \mathbf{V}^*),
\end{aligned}$$

in which  $d_{\mathbf{V}_k} R_{sum}$  is the differential of sum rate due to variation of  $\mathbf{V}_k$  and we have used the fact that  $\text{tr} [\text{Bdiag}(\mathbf{X}_1, \dots, \mathbf{X}_K)] = \sum_{k=1}^K \text{tr}(\mathbf{X}_k)$ . It can be concluded that  $\mathbf{D} = \frac{\partial}{\partial \mathbf{V}^*} R_{sum}$  which results in (8).

#### ACKNOWLEDGMENT

This work was supported by the FP7 project HIATUS (grant 265578) of the European Commission and by the Austrian Science Fund (FWF) through grant NFN SISE (S106).

#### REFERENCES

- [1] R. H. Etkin, D. N. C. Tse, and H. Wang, "Gaussian interference channel capacity to within one bit," *IEEE Trans. Inf. Theory*, vol. 54, pp. 5534–5562, Dec. 2008.
- [2] G. Scutari, D. P. Palomar, and S. Barbarossa, "The MIMO iterative waterfilling algorithm," *IEEE Trans. on Signal Processing*, vol. 57, no. 5, pp. 1917–1935, May. 2009.
- [3] Z. Chen, S. A. Vorobyov, C.-X. Wang, and J. Thompson, "Nash bargaining over MIMO interference systems," in *Proc. IEEE Int. Conf. Commun.*, Jun. 2009.
- [4] Z. K. M. Ho and D. Gesbert, "Balancing egoism and altruism on interference channel: The MIMO case," in *Proc. IEEE Int. Conf. Commun.*, May. 2010.
- [5] V. Cadambe and S. Jafar, "Interference alignment and the degrees of freedom of the K user interference channel," *IEEE Trans. Inf. Theory*, vol. 54, no. 8, pp. 3425–3441, Aug. 2008.
- [6] K. Gomadam, V. R. Cadambe, and S. A. Jafar, "Approaching the capacity of wireless networks through distributed interference alignment," in *Proc. of IEEE GLOBECOM*, Dec. 2008.
- [7] S. W. Peters and R. W. Heath, "Interference alignment via alternating minimization," in *Proc. of IEEE ICASSP*, Apr. 2009, pp. 2445 – 2448.
- [8] S. H. Park, H. Park, Y. D. Kim, and I. Lee, "Regularized interference alignment based on weighted sum-MSE criterion for MIMO interference channels," in *Proc. IEEE Int. Conf. Commun.*, May. 2010.
- [9] I. Santamaria, O. Gonzalez, R. W. Heath, Jr., and S. W. Peters, "Maximum sum-rate interference alignment algorithms for MIMO channels," in *Proc. of IEEE Global Telecommunications Conference*, Dec. 2010.
- [10] H. Sung, S. Park, K. Lee, and I. Lee, "Linear precoder designs for K-user interference channels," *IEEE Trans. on Wireless Communication*, vol. 9, no. 1, pp. 291–301, Jan. 2010.
- [11] A. Edelman, T. A. Arias, and S. T. Smith, "The geometry of algorithms with orthogonality constraints," *SIAM J. Matrix Anal. Applicat.*, vol. 20, no. 2, pp. 303–353, 1998.
- [12] D. P. Bertsekas, *Nonlinear Programming*. Belmont, MA: Athena Scientific, 1995, sec. 2.3.
- [13] A. Corona, M. Marchesi, C. Martini, and S. Ridella, "Minimizing multimodal functions of continuous variables with the simulated annealing algorithm," *ACM Trans. Mathematical Software*, vol. 13, no. 3, pp. 262–280, 1987.
- [14] S. S. Raoh, *Engineering Optimization*. New York: Wiley, 1996.