EEC 266: Final Project Report

Mason del Rosario – *Department of Electrical and Computer Engineering*
University of California Davis
Davis, CA
mdelrosa@ucdavis.edu

For my term project, I reviewed "Maximum Mutual Information Design for MIMO Systems With Imperfect Channel Knowledge" [1]. In the paper, the authors seek to find an expression for the optimal **transmit covariance matrix**, $\mathbf{Q}$, for a **capacity lower bound** given **channel mean** and **channel correlation information** at both the transmitter and the receiver. This progress report provides summaries of the major sections of the paper and a summary of my efforts in replicating numerical results.

1. Introduction

Multiple Input Multiple Output (MIMO) systems constitute a promising technology for future wireless communications networks, as they are known to increase the achievable capacity of communications networks [2]. However, determination of the MIMO channel's capacity is dependent on the availability of accurate **Channel State Information (CSI)** at the transmitter and at the receiver.

A. Background

1) Perfect CSI: To develop an intuition for the importance of CSI, consider a canonical MIMO channel,

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n}$$

with additive Gaussian noise, $\mathbf{n}$, and complex fading channel coefficients, $\mathbf{H}$. *Perfect CSI* is equivalent to exact knowledge of $\mathbf{H}$. If the receiver can determine $\mathbf{H}$ by utilizing pilot signals, then it can send this information back to the transmitter. The transmitter may then use $\mathbf{H}$ to *precode* its transmissions, which would result in a received symbol vector

$$\mathbf{y} = \mathbf{H}^{-1}\mathbf{H}\mathbf{x} + \mathbf{n} = \mathbf{x} + \mathbf{n}$$

In this case, the determination of $\mathbf{x}$ at the receiver is trivial, as it involves filtering of Gaussian noise.

2) Imperfect CSI: Reality is often disappointing, and unfortunately, CSI is no exception. The receiver does not have a priori knowledge of the channel, and the resulting feedback to the transmitter is an estimate, $\hat{\mathbf{H}}$, which we call *Imperfect CSI*. This yields a received symbol of the form

$$\mathbf{y} = \hat{\mathbf{H}}^{-1}\mathbf{H}\mathbf{x} + \mathbf{n}.$$

Clearly, errors in estimation may have adverse effects on the recovery of $\mathbf{x}$ from $\mathbf{y}$. An illustration of the dataflow in a MIMO system can be seen in Figure 1.

This paper was written to fulfill the term paper requirement for the UC Davis graduate course, *EEC266 Information Theory and Coding*.

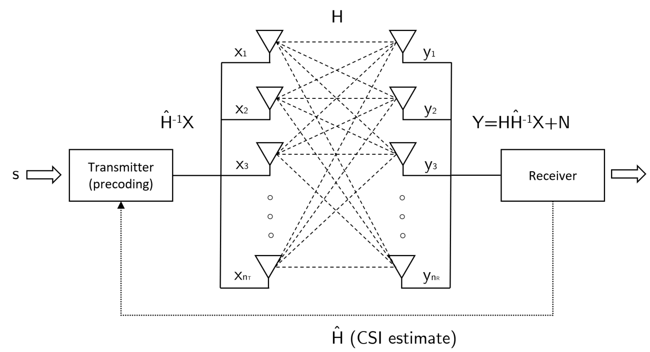

Fig. 1: Illustration of a MIMO system with imperfect feedback.

3) CSI and MIMO Capacity: As discussed in lecture, the Shannon capacity of a fading MIMO channel with additive Gaussian noise is found by way of an eigendecomposition of $\mathbf{H}\mathbf{H}^T$. Assuming an appreciably large transmit power, P, the MIMO capacity can be written as

$$\begin{aligned} C_{\text{MIMO}} &= \max_{\text{Tr}(Q_{ii})} \sum_{i=1}^{t} \frac{1}{2} \log\left(1 + Q_{ii}\lambda_{ii}\right) \\ &\approx \sum_{i=1}^{t} \frac{1}{2} \log\left(1 + \frac{P}{t}\lambda_{ii}\right) \\ &\approx \frac{t}{2}\log P \\ &\leq \min(k,\ell)\frac{1}{2}\log P \end{aligned} \tag{1.1}$$

where $\mathbf{Q}$ is the transmit covariance matrix, λ_{ii} is the i-th eigenvalue of $\mathbf{H}\mathbf{H}^T$, t is the rank of $\mathbf{H}\mathbf{H}^T$ (i.e., number of λ_{ii}), and k and ℓ are the number of transmit and receive antennas, respectively.

However, *this analysis depends on perfect CSI*, and the achievable capacity of a MIMO channel under imperfect CSI requires consideration of $\hat{\mathbf{H}}$. This brings us to the question which [1] attempts to answer: **What is the effect of Imperfect CSI on the achievable capacity of the MIMO channel?**

B. Prior Work

The authors note the following papers as prior work making the following contributions:

- Perfect CSI at transmitter and receiver [3]–[5].
- Perfect CSI at receiver (CSIR) and channel mean/covariance information at transmitter (CSIT) [6].

- Noncoherent case with no instantaneous CSIT or CSIR [7], [8]
- Perfect coherent system (perfect CSIR) [9].
- Noncoherent system (no instantaneous CSIR) [9].
- Channel estimation for CSIR with estimation errors/CSIT assumed to be identical to CSIR and fed back to transmitter with no loss [10].
- CSIR and CSIT independently estimated via channel mean and channel correlation [11], [15].

C. Main Contributions

The authors claim the following contributions:

- Formalization of globally optimum transmit covariance matrix for capacity lower bound.
- Determination of relationship between maximum mutual information design and minimum mean-square error (MSE) design with imperfect CSI.
- Derive different results than [11] w.r.t. effect of different amounts of transmit/receive correlation.

2. Preliminaries

A. Notation

The notation from [1] used in this report includes:

- Upper (lower) case bold letters represent matrices (vectors).
- $\mathbb{E}[\cdot]$ is the expectation, $\mathrm{tr}(\cdot)$ is the matrix trace.
- $|\mathbf{A}|$ is the determinant of $\mathbf{A}$, $|a|$ is the magnitude of a
- $(\cdot)^*$ and $(\cdot)^H$ represent the complex conjugate and Hermitian, respectively
- $\mathbf{I}_a$ is the $a \times a$ identity
- $\mathcal{N}_c(\cdot,\cdot)$ is the circularly symmetric complex Gaussian distribution
- $\mathbf{A} \succeq 0$ means that $\mathbf{A}$ is positive semidefinite

B. System Model

A single-user MIMO system is described by n_T transmitter antennas and n_R receiver transmitters. The channel is modeled as $\mathbf{H}$, and received signal is $\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n}$ where $\mathbf{x}$ is the $n_t \times 1$ original coded message and $\mathbf{n} \sim \mathcal{N}_c(0, \sigma_n^2 \mathbf{I}_{n_R})$ is an $n_R \times 1$ Gaussian noise vector. The input covariance matrix is written as $\mathbf{Q} = \mathbb{E}\left[\mathbf{x}\mathbf{x}^H\right]$. The channel model $\mathbf{H} = \mathbf{R}_R^{\frac{1}{2}} H_w \mathbf{R}_T^{\frac{1}{2}}$ is from [12] with normalized (i.e., 1's along their diagonals) receive and transmit correlation matrices $\mathbf{R}_R$ and $\mathbf{R}_T$ and with all entries of $\mathbf{H}_w$, $h_{w,ij} \sim \mathcal{N}(0,1)$.

The imperfect CSIR model is adapted from [11] as follows,

$$\mathbf{H} = \hat{\mathbf{H}} + \mathbf{E}, \quad \hat{\mathbf{H}} = \mathbf{R}_R^{\frac{1}{2}} \hat{\mathbf{H}}_w \mathbf{R}_T^{\frac{1}{2}}, \quad \hat{\mathbf{E}} = \mathbf{R}_R^{\frac{1}{2}} \hat{\mathbf{E}}_w \mathbf{R}_T^{\frac{1}{2}} \tag{2.1}$$

with CSIR estimate $\hat{\mathbf{H}}$ of $\mathbf{H}$ and estimation error $\mathbf{E}$. $\hat{\mathbf{H}}$ and $\mathbf{E}$ are zero-mean, uncorrelated with i.i.d. entries $\mathcal{N}_c(0, 1-\sigma_E^2)$ and $\mathcal{N}_c(0, \sigma_E^2)$, respectively (i.e., σ_E^2 is the estimation error variance). Here, the authors of [1] make two key assumptions as per [11]:

1) *Lossless feedback is assumed*,
2) *CSIR and CSIT are assumed to be identical.*

While these assumptions limit the practicality of the results, the consideration of estimated CSIR/CSIT is still more realistic than the perfect CSIR/CSIT used in the prior work. With these assumptions in hand, the full CSI, $\hat{\mathbf{H}}$, $\mathbf{R}_R$, $\mathbf{R}_T$, $\mathbf{R}_T$, σ_E^2, and σ_n^2, are known at both the transmitter and the receiver. The full channel output can be rewritten as $\mathbf{y} = \hat{\mathbf{H}}\mathbf{x} + \mathbf{E}\mathbf{x} + \mathbf{n}$, and the total noise can be written as $\mathbf{n}_{\text{total}} = \mathbf{E}\mathbf{x} + \mathbf{n}$ with covariance written as

$$\mathbf{R}_{n_{\text{total}}} = \sigma_E^2 \mathrm{tr}\left(\mathbf{R}_T \mathbf{Q}\right) \mathbf{R}_R + \sigma_n^2 \mathbf{I}_{n_R} \tag{2.2}$$

Note that $R_{n_{\text{total}}}$ is not strictly Gaussian, meaning an explicit capacity expression is difficult to derive. Instead, lower and upper bounds on the capacity have been proposed as per [10], [11],

$$\underline{I}_{\text{low}} \leq I\left(\mathbf{x}; \mathbf{y}|\hat{\mathbf{H}}\right) \leq \underline{I}_{\text{up}} \tag{2.3}$$

with

$$\underline{I}_{\text{low}} = \log_2 |\mathbf{I}_{n_R} + \hat{\mathbf{H}}\mathbf{Q}\hat{\mathbf{H}}^H \mathbf{R}_{n_{\text{total}}}^{-1}| \tag{2.4}$$

$$\underline{I}_{\text{up}} = \underline{I}_{\text{low}} + \log_2 |\mathbf{R}_{n_{\text{total}}}| - \mathbb{E}\left[\log_2 |\sigma_E^2 (\mathbf{x}^H \mathbf{R}_T \mathbf{x}^H \mathbf{R}_R) + \sigma_n^2 \mathbf{I}_{n_R}|\right] \tag{2.5}$$

where the expectation in the upper bound is taken w.r.t. $p(\mathbf{x})$. $\underline{I}_{\text{low}}$ and $\underline{I}_{\text{up}}$ denote the lower and upper bounds on mutual information, respectively.

3. Problem Formulation

The capacity lower-bound is adopted as the design criterion with the following optimization problem, [10], [15]

$$I_{\text{low}} = \max_{\mathbf{Q} \succeq 0, \mathrm{tr}(\mathbf{Q}) \leq P_T} \log_2 |\mathbf{I}_{n_R} + \hat{\mathbf{H}}\mathbf{Q}\hat{\mathbf{H}}^H \mathbf{R}_{n_{\text{total}}}^{-1}| \tag{3.1}$$

where the lower-bound on the ergodic capacity is the following expectation taken w.r.t. the channel distribution,

$$C_{\text{low}} = \mathbb{E}\left[I_{\text{low}}\right]. \tag{3.2}$$

4. Methodology

To find the optimal $\mathbf{Q}$, the authors of [1] form an equivalent problem to (3.1) by considering a *precoder-decoder pair*, $(\mathbf{F}, \mathbf{G})$, where $\mathbf{x} = \mathbf{F}\mathbf{s}$ for zero-mean, unit variance i.i.d. data vector $\mathbf{s}$ with dimension $r_g \times 1$. Observe that $\mathbf{Q} = \mathbb{E}\left[\mathbf{x}\mathbf{x}^H\right] = \mathbf{F}\mathbf{F}^H$. The channel output can be rexpressed as

$$\mathbf{y} = \hat{\mathbf{H}}\mathbf{F}\mathbf{s} + \mathbf{E}\mathbf{F}\mathbf{s} + \mathbf{n} \tag{4.1}$$

The received vector after the decoder is given as $\mathbf{r} = \mathbf{G}\mathbf{y}$. Define the Mean-Square Error (MSE) matrix w.r.t. $(\mathbf{F}, \mathbf{G})$

$$\begin{aligned}\mathrm{MSE}(\mathbf{F}, \mathbf{G}) &\overset{\text{def}}{=} \mathbb{E}\left[(\mathbf{r}-\mathbf{s})(\mathbf{r}-\mathbf{s})^H\right] \\ &= \mathbf{G}\hat{\mathbf{H}}\mathbf{F}\mathbf{F}^H\hat{\mathbf{H}}^H\mathbf{G}^H - \mathbf{G}\hat{\mathbf{H}}\mathbf{F} - \mathbf{F}^H\hat{\mathbf{H}}^H\mathbf{G}^H + \mathbf{I}_{r_g} \\ &\quad + \mathbf{G}\mathbf{R}_{n_{\text{total}}}\mathbf{G}^H \end{aligned} \tag{4.2}$$

where the expectation is taken w.r.t. the distributions of $\mathbf{s}, \mathbf{n}, \mathbf{E}_w$. The problem in (3.1) is equivalent to

$$\min_{\mathbf{F}, \mathbf{G}, \mathrm{tr}(\mathbf{F}\mathbf{F}^H) \leq P_T} \ln |\mathrm{MSE}(\mathbf{F}, \mathbf{G})| \tag{4.3}$$

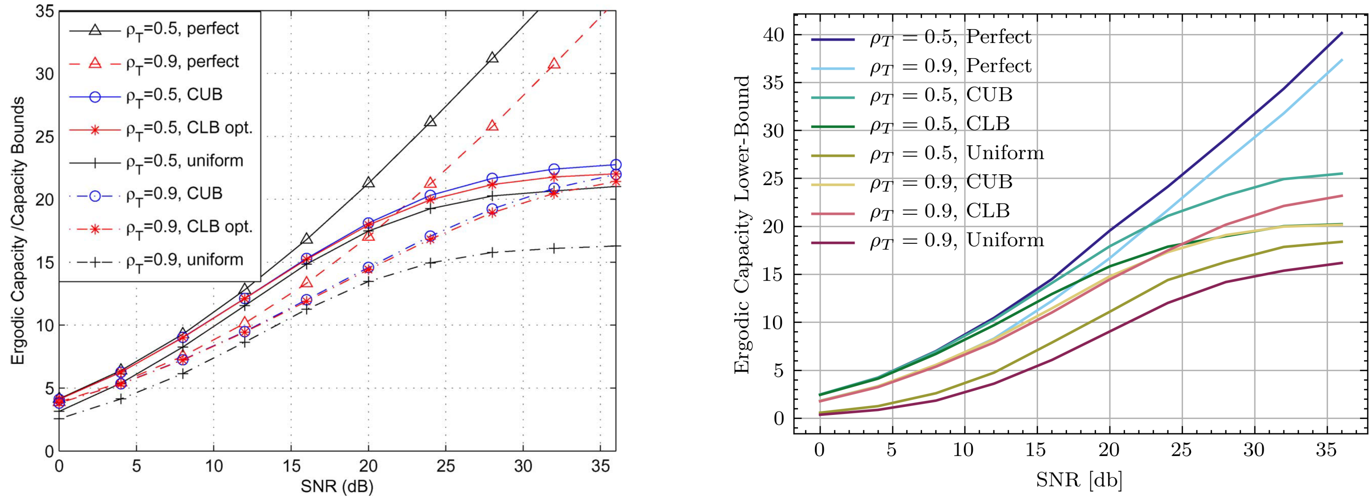

Fig. 2: **Left**: Justification for the capacity lower bound (CLB) as a design criterion for the transmit covariance matrix. All curves assume no receive correlation (i.e., $\rho_R = 0.0$, $\mathbf{R}_R = \mathbf{I}_{n_R}$). **Right**: Replication using the iterative algorithm.

As a constrained optimization problem, the Lagrangian with inequality constraint can be written as

$$\mathcal{L}_c(\mathbf{F}, \mathbf{G}, \mu) = \ln |\text{MSE}(\mathbf{F}, \mathbf{G})| + \mu_g \left[\text{tr}\left(\mathbf{F}\mathbf{F}^H\right) - P_T\right]. \quad (4.4)$$

After some simplification of the Lagrangian with KKT conditions, the relevant equations for the precoder/decoder are

$$\mathbf{G} = \mathbf{F}^H \hat{\mathbf{H}}^H \left[\hat{\mathbf{H}}\mathbf{F}\mathbf{F}^H\hat{\mathbf{H}}^H + \mathbf{R}_{n_{\text{total}}}\right]^{-1} \quad (4.5)$$

$$\mu_g = \frac{\sigma_n^2}{P_T}\text{tr}(\mathbf{G}\mathbf{R}_{n_{\text{total}}}\hat{\mathbf{H}}\mathbf{F}) \quad (4.6)$$

$$\alpha_g = \text{tr}\left\{\mathbf{G}\mathbf{R}_R\mathbf{R}_{n_{\text{total}}}\hat{\mathbf{H}}\mathbf{F}\right\} \quad (4.7)$$

$$\mathbf{F} = \left[\mu_g \mathbf{I} + \alpha_g \sigma_E^2 \mathbf{R}_T\right]^{-1} \hat{\mathbf{H}}^H \mathbf{G}^H \quad (4.8)$$

The authors show that a global optimum, $(\mathbf{F}_{g_{\text{opt}}}, \mathbf{G}_{g_{\text{opt}}})$ exists for (4.3) and that $\mathbf{Q}_{g_{\text{opt}}}$ is related by $\mathbf{Q}_{g_{\text{opt}}} = \mathbf{F}_{g_{\text{opt}}}\mathbf{F}_{g_{\text{opt}}}^H$.

In general, when each of the correlation matrices, $\mathbf{R}_R$ and $\mathbf{R}_T$, are not strictly uncorrelated (i.e., $\mathbf{R}_T \neq \mathbf{I}_{n_T}$ and $\mathbf{R}_R \neq \mathbf{I}_{n_R}$), the authors write the expression for $(\mathbf{F}_{g_{\text{opt}}}, \mathbf{G}_{g_{\text{opt}}})$ can be found numerically with an iterative algorithm based on [13]. Table I outlines this algorithm.

1.	Initialize $\mathbf{F} = \mathbf{F}_0$ as diagonal matrix with $\text{tr}(\mathbf{F}_0\mathbf{F}_0^H) = P_T$
2.	Update $\mathbf{R}_{n_{\text{total}}}$ and $\mathbf{G}$ using (2.2) and (4.5), respectively.
3.	Update μ_g and α_g using (4.6) and (4.7), respectively.
4.	Update $\mathbf{F}$ using (4.8); scale $\mathbf{F}$ s.t. $\text{tr}(\mathbf{F}\mathbf{F}^H) = P_T$.
5.	If $\|\|\mathbf{F}_i - \mathbf{F}_{i-1}\|\|_F \leq \epsilon$, return $\mathbf{F}_i$. Otherwise, return to Step 2.

TABLE I: Iterative algorithm for Lagrangian optimization of $\mathbf{F}, \mathbf{G}$.

In the special case where the received signals are uncorrelated (i.e., $\mathbf{R}_R = \mathbf{I}_{n_R}$ and $\mathbf{R}_T \neq \mathbf{I}_{n_T}$), $\mathbf{Q}_{\text{opt}}$ may be found by way of a closed-form expression.

5. Numerical Results

The authors conduct three experiments summarized in the following three figures:

1) Fig. 2 provides a justification for adopting the capacity lower bound (CLB) as a design criterion for $\mathbf{Q}$.
2) Fig. 3 shows the effect of increasing correlation on the channel capacity.
3) Fig. 4 compares equal amounts of correlation at the transmitter and the receiver on the channel capacity.

For the purposes of this project, I have replicated Fig. 2, 3, and 4.

To define the spatial correlation of the antennas, the authors use the exponential correlation models given in [14], [15] as

$$(\mathbf{R}_T)_{ij} = \rho_T^{|i-j|} \text{ for } i, j \in \{1, \ldots, n_T\} \quad (5.1)$$

$$(\mathbf{R}_R)_{ij} = \rho_R^{|i-j|} \text{ for } i, j \in \{1, \ldots, n_R\} \quad (5.2)$$

where $n_T = n_R = 4$ (i.e., 4 antennas are used at both the transmitter and the receiver). The SNR is defined as P_T/σ_n^2.

A. *Comparison of Different Capacity Bounds*

Fig. 2 shows the CLB in comparison to the capacity upper bound (CUB). The authors show that in the case of no receive correlation, these bounds are tight, meaning adoption of the lower bound provides close-to-optimal capacity. Furthermore, two more cases are plotted, including

1) **Perfect**: Assumes knowledge of **Perfect CSI** (σ_E^2) and utilizes the water-filling algorithm along with knowledge of the channel's eigenmodes.
2) **Uniform**: Allocates equal power to each antenna.

The **Perfect** case reflects the Shannonn capacity bound (1.1) and serves as an upper bound for an ideal situation while the **Uniform** case serves as the analogous lower bound.

B. *Effect of Correlation on Capacity*

Figures 3 and 4 show the effect of increasing correlation at either the transmitter or the receiver. In Fig. 3, the authors illustrate that increasing correlation at both ends of the channel

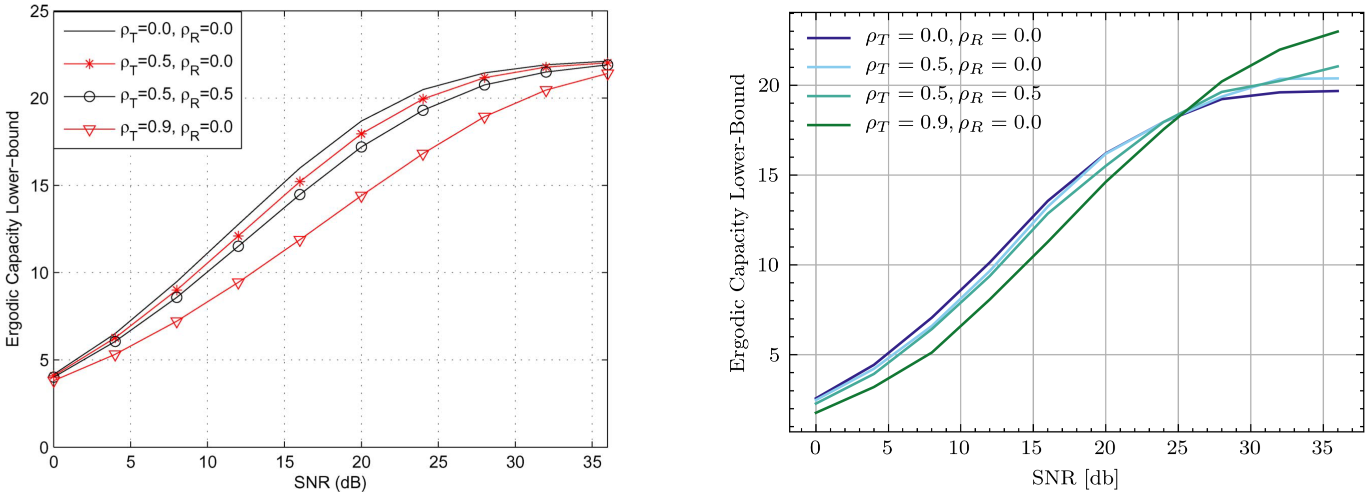

Fig. 3: **Left**: Figure 3 from [1] showing the capacity lower-bound for $\sigma_E^2 = 0.01$ using the iterative algorithm (Table I. Increasing correlation degrades the capacity of the channel. **Right**: Replication using the iterative algorithm.

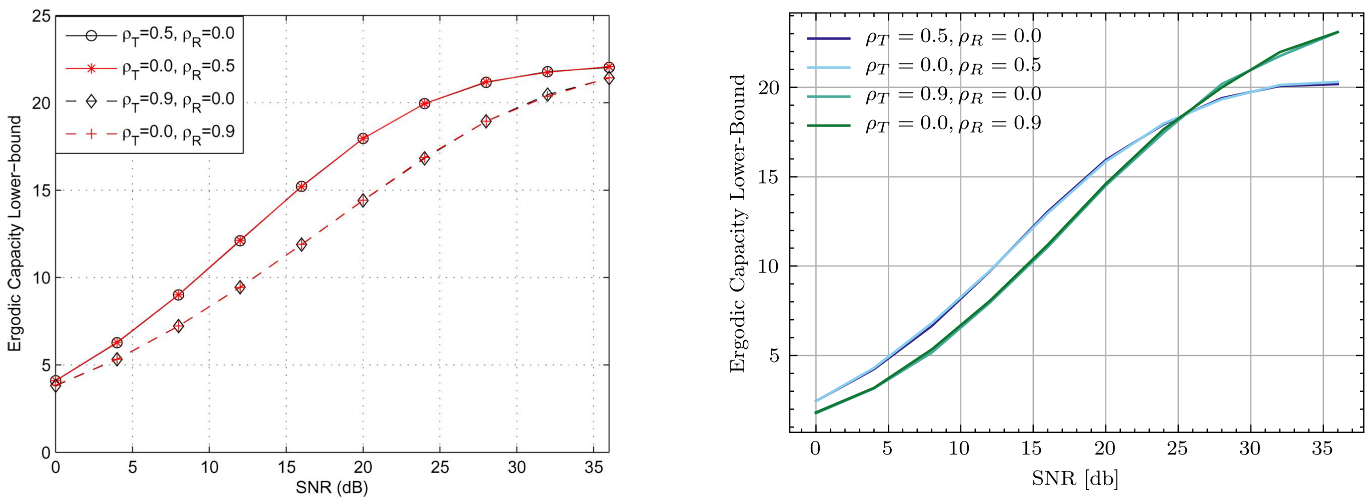

Fig. 4: **Left**: Figure 4 from [1] showing the capacity lower-bound for $\sigma_E^2 = 0.01$ using the iterative algorithm (Table I. Equivalent amounts of correlation at the transmitter/receiver result in equivalent channel capacity. **Right**: Replication using the iterative algorithm.

decreases the achievable capacity. In Fig. 4, the authors show that equal amounts of correlation result in equal channel capacity.

C. Replication

1) Performance: The replicated results do not perfectly reflect the paper's results, as the cases with high correlation ($\rho_T = 0.9$ or $\rho_R = 0.9$) reach higher capacity values than the cases with low correlation ($\rho_T \leq 0.5$ or $\rho_R \leq 0.5$) at high SNRs (see RHS of Fig. 3 and 4). The crossover point where high correlation capacity exceeds low correlation capacity appears to be 25dB.

The key piece of this replication is the algorithm outlined in Table I, and to generate results identical to [1], parameters for the algorithm may need to be tweaked. Some potential causes of this discrepancy are

- **The convergence criterion,** ϵ. The paper cites $\epsilon = 0.01$ as a *possible value* but does not make explicit what value is reflected in their figures. For this project, $\epsilon \in \{0.1, 0.01, 0.001\}$ were adopted, but none of them resulted in a significant change in behavior.
- **Number of samples**. The capacity it taken as the expectated value of the MI lower and upper bounds, $\underline{I}_{\text{low}}$ and $\underline{I}_{\text{up}}$. The number of samples used for the replication is 500, which was chosen to minimize the runtime of the experiments, but it is possible that a larger number of samples needs to be taken to accurately approximate the expectation.

2) File Location: The codebase for generating the system model described in Section 2.2 and the algorithm described in Table I is in the following Github repository: https://github.com/mdelrosa/eec266_project. This code is in a Jupyter notebook, and for ease of portability, this can be viewed and run in an internet browser using the following Google Colab: https://colab.research.google.com/github/mdelrosa/eec266_project/blob/master/eec266_project.ipynb.

3) Runtime: On my local device (Dell XPS 13 9350 Signature edition, Windows 10) and on Google Colab, all experiments took **14 minutes** to complete.

4) Formatting Notes: If run on a device without the proper `matplotlib` styles (see https://github.com/garrettj403/SciencePlots), Fig. 2 through 4 will render with default `matplotlib` settings. Thus, the content of the Figures will be identical while the formatting will be different.

References

[1] M. Ding and S.D. Blostein, "Maximum Mutual Information Design for MIMO Systems With Imperfect Channel Knowledge," in IEEE Transactions on Information Theory, vol. 56, no. 10, pp. 4793-4801, Oct. 2010.

[2] A. Goldsmith, S. A. Jafar, N. Jindal and S. Vishwanath, "Capacity limits of MIMO channels," in IEEE Journal on Selected Areas in Communications, vol. 21, no. 5, pp. 684-702, June 2003.

[3] E. Telatar, "Capacity of multi-antenna Gaussian channels," Eur. Trans. Telecommun. (ETT), vol. 10, no. 6, pp. 585–596, Nov./Dec. 1999.

[4] A. Scaglione, P. Stoica, S. Barbarossa, G. B. Giannakis, and H. Sampath, "Optimal designs for space-time linear precoders and decoders," IEEE Trans. Signal Process., vol. 50, no. 5, pp. 1051–1064, May 2002.

[5] H. Sampath, P. Stoica, and A. Paulraj, "Generalized linear precoder and decoder design for MIMO channels using the weighted MMSE criterion," IEEE Trans. Commun., vol. 49, pp. 2198–2206, Dec. 2001.

[6] S. A. Jafar and A. Goldsmith, "Transmitter optimization and optimality of beamforming for multiple antenna systems," IEEE Trans. Wireless Commun., vol. 3, no. 4, pp. 1165–1175, Jul. 2004.

[7] T. L. Marzetta and B. M. Hochwald, "Capacity of a mobile multiple antenna communication link in Rayleigh flat fading," IEEE Trans. Inf. Theory, vol. 45, no. 1, pp. 139–157, Jan. 1999.

[8] S. A. Jafar and A. Goldsmith, "Multiple-antenna capacity in correlated Rayleigh fading with channel covariance information," IEEE Trans. Wireless Commun., vol. 4, no. 3, pp. 990–997, May 2005.

[9] A. Goldsmith, S. A. Jafar, N. Jindal, and S. Vishwanath, "Capacity limits of MIMO channels," IEEE J. Sel. Areas. Commun., vol. 21, no. 5, pp. 684–702, Jun. 2003.

[10] T. Yoo and A. Goldsmith, "Capacity and power allocation for fading MIMO channels with channel estimation error," IEEE Trans. Inf. Theory, vol. 52, no. 5, pp. 2203–2214, May 2006.

[11] L. Musavian, M. R. Nakhai, M. Dohler, and A. H. Aghvami, "Effect of channel uncertainty on the mutual information of MIMO fading channels," IEEE Trans. Veh. Technol., vol. 56, pp. 2798–2806, Sep. 2007.

[12] D. Shiu, G. J. Foschini, M. J. Gans, and J. M. Kahn, "Fading correlation and its effect on the capacity of multielement antenna systems," IEEE Trans. Commun., vol. 48, pp. 502–513, Mar. 2000.

[13] S. Serbetli and A. Yener, "Transceiver optimization for multiuser MIMO systems," IEEE Trans. Signal Process., vol. 52, no. 1, pp. 214–226, Jan. 2004.

[14] V. A. Aalo, "Performance of maximal-ratio diversity systems in a correlated Nakagami-fading environment," IEEE Trans. Commun., vol. 43, pp. 2360–2369, Aug. 1995.

[15] T. Yoo, E. Yoon, and A. Goldsmith, "MIMO capacity with channel uncertainty: Does feedback help?," in Proc. IEEE Globecom, Dallas, TX, Dec. 2004, pp. 96–100.