Indefinite Hermitian Quadratic Forms in Gaussian Random Variables: Distribution, Scaling, and Application to the Broadcast Channel

Tareq Al-Naffouri

Electrical Engineering Department

King Fahd University of Petroleum and Minerals

Dhahran, Saudi Arabia

www.kfupm.edu.sa/faculty/ee/naffouri

Fulbright Research Visitor

Electrical Engineering Department

USC

Part of is this work was done jointly with Masoud Sharif, and Babak Hassibi

Quadratic Forms in Gaussian Variables

- Gaussian variables play a very important role in statistics, signal processing, and communications
- Quadratic forms in Gaussian random variables are of particular importance
- Let A be a Hermitian matrix of size M and consider the random quadratic form

$$Y = \|H\|_A^2 \triangleq H^* A H \tag{1}$$

where H is a white cicularly symmetric Gaussian random variable, i.e. H i.e. $H \sim \mathcal{CN}(0, R)$.

- Without loss of generality, we can assume H white as we can absorb the correlation into the weight matrix.

Evaluating the CDF the Normal Way is Difficult

- Consider the random Hermitian quadratic form

  $$Y = \|H\|_A^2$$

  The CDF of Y is defined by

  $$\begin{aligned} F_Y(y) &= P\{Y \le y\} && (2) \\ &= \int_{\mathcal{A}} p(H) dH && (3) \end{aligned}$$

  where $\mathcal{A}$ is area in M multidimensional H plane defined by the inequality

  $$\|H\|_A^2 \le y \quad (4)$$

- Such an integral would in general be very difficult to evaluate.

An Alternative Way to Evaluating the CDF

- An alternative way to do so is to express the inequality that appears in (4) as

$$y - \|H\|_A^2 \geq 0$$

So, the CDF takes the form

$$F_Y(y) = \frac{1}{2\pi^M} \int e^{-H^* H} u(y - \|H\|_A^2) dH$$

The constraint appears in the integrand and not in the integration limits

- Difficult to deal with the unit step. So replace it with its Fourier transform

$$u(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{e^{x(j\omega+\beta)}}{j\omega + \beta} d\omega$$

which is valid for any $\beta > 0$

An Alternative Way to Evaluating the CDF

- An alternative way to do so is to express the inequality that appears in (4) as

$$y - \|H\|_A^2 \geq 0$$

So, the CDF takes the form

$$F_Y(y) = \frac{1}{2\pi^M} \int e^{-H^* H} u(y - \|H\|_A^2) dH$$

The constraint appears in the integrand and not in the integration limits

- Difficult to deal with the unit step. So replace it with its Fourier transform

$$u(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{e^{x(j\omega+\beta)}}{j\omega + \beta} d\omega$$

which is valid for any $\beta > 0$

The CDF as a 1-D Integral

- CDF can be written as $M+1$ integral

$$F_Y(y) = \frac{1}{2\pi^{M+1}} \int d\omega \frac{e^{y(j\omega+\beta)}}{j\omega+\beta} \int dH e^{-H^*(I+A(j\omega+\beta))H} dH \qquad (5)$$

- By examining (5), we note that inner integral looks like a Gaussian integral. Intuition suggests that this integral can be written as

$$\frac{1}{\pi^M} \int e^{-H^*(I+A(j\omega+\beta))H} dH = \frac{1}{\det(I+A(j\omega+\beta))} \qquad (6)$$

- We are left with

$$F_Y(y) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{e^{y(j\omega+\beta)}}{j\omega+\beta} \prod_{i=1}^{M} \frac{1}{1+\lambda_i(x)(j\omega+\beta)} d\omega \qquad (7)$$

The CDF in Closed Form

- Use partial fraction expansion and contour integration to write CDF in closed form

$$F_Y(y) = u(y) + \sum_{l=1}^{M} \frac{\lambda_i^{M+1}}{\prod_{l \neq i}(\lambda_i - \lambda_l)} \frac{1}{|\lambda_i|} e^{-\frac{y}{\lambda_l}} u(\frac{y}{\lambda_l})$$

- Note that this result applies irrespective of the correlation of H and irrespective of the weight matrix A

Dealing with the Nonzero Mean Problem

- In the nonzero mean problem, we have the quadratic form

$$Y = \|H - a\|_A^2$$

- We can evaluate the CDF up to a 1-D integral

$$Pr\left\{Y \le y\right\} \quad = \quad \frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{e^{y(j\omega+\beta)}}{j\omega+\beta} e^{-c} \frac{1}{det(I+(j\omega+\beta)\Lambda)} d\omega \quad (8)$$

where

$$c \quad = \quad \bar{m}^*(I + \frac{1}{j\omega+\beta}\Lambda^{-1})^{-1}\bar{m}$$

- We could not find the distribution in closed form

Dealing with Real Gaussian Variables

- When the Gaussian variables are real, the quadratic form can be expressed as
$$Y = \|H_r\|^2_{A_r} = H_r^T A_r H_r$$
where A_r is now symmetric.

- Remember the difference between complex and real Gaussian variables is that the determinant of covariance matrix appears under the square root

- Because of that, we have
$$Pr\{Y \leq y\} = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{e^{y(j\omega+\beta)}}{j\omega + \beta} \frac{1}{\sqrt{\det(I + A_r(j\omega + \beta))}} \tag{9}$$

Further Results

- We can use this approach to find the distribution of a ratio of Gaussian norms

$$\Pr\left\{\frac{\epsilon_1 + \|H\|_{B_1}^2}{\epsilon_2 + \|H\|_{B_2}^2} \leq x\right\} = \Pr\left\{\|H\|_{B_1 - xB_2} \leq \epsilon_2 x - \epsilon_1\right\}$$

- We can use this approach to find the *joint* distribution of two or more weighted norms

$$F_{X_a, X_b}(x_a, x_b) = \Pr\left\{\|H\|_A^2 \leq x_a, \|H\|_B^2 \leq x_b\right\}$$

- All results can be extended to isotropic distributions: we can find the distribution of $\|\phi\|_A^2$ in closed form

An Application to Broadcast Channels

- Multiple antennas add tremendous value to point to point systems
- Research shifted recently to the role of multiple antennas in multiuser systems
  - (Uplink) Multiple Access (MAC)
  - (Downlink) Broadcast (BC)
- Broadcast scenarios (point to multi-point) are especially important because downlink scheduling is the major bottleneck for broadband wireless networks

Introduction to Broadcast Channels

- Multiple antennas add tremendous value to point to point systems
- Research shifted recently to the role of multiple antennas in multiuser systems
  - (Uplink) Multiple Access (MAC)
  - (Downlink) Broadcast (BC)
- Broadcast scenarios (point to multi-point) are especially important because downlink scheduling is the major bottleneck for broadband wireless networks

Three Main Questions in a Broadcast Scenario (1)

Q1) Quantify the maximum sum rate possible to all users

A1) Sum-rate is achieved using dirty paper coding (DPC) (Caire and Shamai '02, Viswanath and Tse '02, Vishwanath et al. '02, Yu and Cioffi '02)

- **(-)** DPC is computationally complex at both Tx and Rx
- **(-)** Requires a great deal of Feedback (CSI for all users at Tx)

Three Main Questions in a Broadcast Scenario (2)

The second question is motivated by the drawbacks of DPC

Q2) Devise computationally efficient algorithms for capturing capacity

A2) Utilize multi-user diversity to achieve performance close to capacity

- **(+)** Opportunist multiple random beamforming coincides asymptotically with DPC (Sharif and Hassibi '06)

$$R = M \log \log n + M \log \frac{P}{M} + o(1)$$

- **(+)** Requires simply SINR feedback to Tx

Three Main Questions in a Broadcast Scenario (3)

Q3) With this promising performance, how does opportunist beam-forming perform under various non-idealities

A3) **(i)** Time correlation (Kountouris and Gesbert '05)

(ii) Frequency correlation (Fakhereddin, Sharif, and Hassibi'06)

(iii) Channel estimation error (Vikali, Sharif, and Hassibi '06)

(iv) Spatial correlation (D. Park and S Y. Park '05)

Main Problem to be Addressed:

- For a Gaussian broadcast channel, we would like to quantify the hit that transmit correlation causes to scaling laws of the sum-rate capacity. We consider DPC and various beamforming schemes.

System Model

- Base station with M antennas broadcasting to n single-antenna users
- Received signal at each antenna

$$Y_i = \sqrt{P} H_i S + W_i, \qquad i = 1, \ldots, n$$

  with $E[S^* S] = 1$ and Gaussian noise $W_i \sim CN(0, I)$

- Channel H_i of i-th user is $1 \times M$ vector
  - Distributed as $CN(0, R)$; R is nonsingular with $\mathrm{tr}(R) = M$
  - Known perfectly at receiver
  - Follows a bock fading model (with coherence interval T)
  - H_i is independent from one user to another

Scaling of DPC under Correlation

- Sum-rate capacity of DPC

$$R_{DPC} = E\left\{ \max_{\{P_1,\ldots,P_n,\sum P_i = P\}} \log\det\left(I + \sum_{i=1}^{n} H_i^* P_i H_i \right) \right\}$$

- For large n we can show that RHS is both an upper and lower bound

$$\boxed{R_{DPC} = M \log\log n + M \log\left(\tfrac{P}{M}\right) + M \log \sqrt[M]{\det R}}$$

Since $\mathrm{tr}(R) = M$, the geometric mean satisfies

$$\sqrt[M]{\det(R)} \leq \frac{\mathrm{tr}(R)}{M} \leq 1$$

- Compare with rate for spatially uncorrelated channel

$$\boxed{R_{DPC} = M \log\log n + M \log\left(\tfrac{P}{M}\right)}$$

What is Random Beam Forming?

- Choose M random orthonormal vectors ϕ_m, $m = 1, \ldots, M$ (according to an isotropic distribution)
- Construct the signal

$$S(t) = \sum_{m=1}^{M} \phi_m s_m(t), \qquad t = 1, \ldots, T$$

where T is less than the coherence interval of the channel.

- After T channel uses we independently choose another isotropic set of orthonormal vectors $\{\phi_m\}$, and so on. So we are transmitting M random beams.
- This is a generalization of the scheme "Opportunistic Beamforming" (Viswanath et al. '02) in which only one random beam is transmitted and proportional fairness is guaranteed.

Exploit Multi-User Diversity

- Each receiver $i = 1, \ldots, n$ computes the following M SINRs

$$\mathrm{SINR}_{i,m} = \frac{|H_i \phi_m|^2}{1/\rho + \sum_{n \neq m} |H_i \phi_n|^2}, \qquad m = 1, \ldots, M$$

  and feeds back the best SINR

- Rather than randomly assigning the beams, the transmitter assigns signal s_m to the user with the best SINR for that signal. Therefore

$$C = E \sum_{m=1}^{M} \log \left(1 + \max_{i=1,\ldots,n} \mathrm{SINR}_{i,m} \right)$$

- Due to the symmetry of all the random variables involved:

$$C = M E \log \left(1 + \max_{i=1,\ldots,n} \mathrm{SINR}_{i,1} \right)$$

Other Beamforming Schemes

- Random Beam forming (RBF) $S(t) = \sum_{m=1}^{M} \phi_m s_m(t)$
- RBF with Channel whitening

$$S(t) = \sum_{m=1}^{M} \sqrt{\alpha} R^{-1/2} \phi_m s_m(t)$$

- RBF with general precoding

$$S(t) = \sum_{m=1}^{M} \sqrt{\alpha} A \phi_m s_m(t)$$

- Deterministic beamforming

$$S(t) = \sum_{m=1}^{M} \phi_m s_m(t), \quad \phi_m\text{'s are fixed}$$

How to Determine Scaling of BF Schemes

1. Sum rate

$$\begin{aligned} R_{\mathrm{BF}} &= E \sum_{m=1}^{M} \log\left(1 + \max_{i=1,\ldots,n} \mathrm{SINR}_{i,m}\right) \\ &= ME\left(1 + \max_{i=1,\ldots,n} \mathrm{SINR}_{i,m}\right) \end{aligned}$$

2. To calculate expectation, condition on beams

$$R_{\mathrm{BF}|\Phi} = ME_{H_i|\Phi}\left(1 + \max_{i=1,\ldots,n} \mathrm{SINR}_{i,m}\right)$$

- $\mathrm{SINR}_{i,m}|\Phi$ is iid over i
- Find the distribution of $\mathrm{SINR}_{i,m}|\Phi$
- Employ extreme value theory to find $\max_{i=1,\ldots,n} \mathrm{SINR}_{i,m}$

3. Average $R_{\mathrm{BF}|\Phi}$ over Φ

Statistics of $\mathrm{SINR}_{i,m}$ (White Channel)

- $\mathrm{SINR}_{i,m}$ is defined by

$$\mathrm{SINR}_{i,m} = \frac{|H_i \phi_m|^2}{1/\rho + \sum_{n \neq m} |H_i \phi_n|^2}, \qquad m = 1, \ldots, M$$

- Easy to find distribution of $\mathrm{SINR}_{i,m}|\Phi$ when H_i is white

$$\begin{aligned} f(x) &= \frac{e^{-\frac{x}{\rho}}}{(1+x)^M} \left(\frac{1}{\rho}(1+x) + M - 1 \right) \\ F(x) &= 1 - \frac{e^{-\frac{x}{\rho}}}{(1+x)^M} \end{aligned}$$

- Finding these statistics in the correlated case is challenging

Statistics of $\mathrm{SINR}_{i,m}$ given Φ (Correlated Case)

- We can show that the CDF of SINR in the correlated case

$$\boxed{F(x) = 1 - \frac{1}{2\pi^M \det(R)} \lambda_M \prod_{i=1}^{M-1} \frac{\lambda_i \lambda_M}{x(\lambda_i - \lambda_M)} e^{-\frac{1}{\rho}\frac{x}{\lambda_M}}}$$

  where $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_M$ are the eigenvalues of the matrix

$$A = (1+x)\Lambda^{1/2}\phi_m\phi_m^*\Lambda^{1/2} - x\Lambda \qquad \rho = \frac{P}{M}$$

  Note that eigenvalues are a function of x.

- pdf is given by

$$f(x) = \frac{1}{2\pi^M \det(R)} e^{-\frac{1}{\rho}\frac{x}{\lambda_M}} \prod_{i=1}^{M-1} \frac{\lambda_i \lambda_M}{x(\lambda_i - \lambda_M)} \times \left\{ \frac{1}{\rho} \frac{\|q_M\|_C^2}{\lambda_M} - \|q_M\|_B^2 - \sum_{i=1}^{M} \frac{1}{\lambda_i} \frac{\lambda_M^2 \|q_i\|_C^2 - \lambda_i^2 \|q_M\|_C^2}{x(\lambda_i - \lambda_M)} \right\}$$

  where $B = \Lambda^{1/2}(\phi_m\phi_m^* - I)\Lambda^{1/2} \qquad C = \Lambda^{1/2}\phi_m\phi_m^*\Lambda^{1/2}$

Scaling of Maximum SINR

- Can now show

$$\lim_{x\to\infty} \frac{1-F(x)}{f(x)} = \frac{P}{M} \frac{1}{\|\phi_m\|^2_{\Lambda^{-1}}}$$

- Using extreme value theory, we can show that for large n

$$\max_{i=1,\ldots,n} \mathrm{SINR}_{i,m} = \frac{P}{M} \frac{1}{\|\phi_m\|^2_{\Lambda^{-1}}} \log n$$

- Conditional sum-rate capacity scales as

$$R_{\mathrm{BF}|\Phi} = M \log\log n + M \log \frac{P}{M} + M \log\left(\frac{1}{\|\phi_m\|^2_{\Lambda^{-1}}}\right)$$

- Sum-rate capacity of random beam-forming

$$\boxed{R_{\mathrm{RBF}} = M \log\log n + M \log \tfrac{P}{M} + M E_\Phi \log\left(\tfrac{1}{\|\phi_m\|^2_{\Lambda^{-1}}}\right)}$$

Averaging over the Random Beams

- Need to obtain CDF of $\frac{1}{\|\phi_m\|^2_{\Lambda-1}}$ which is challenging.
- The CDF of $y = \frac{1}{\|\phi\|^2_{\Lambda-1}}$ is given by

$$\boxed{G(x) = Pr(\frac{1}{\|\phi\|^2_{\Lambda-1}} < x) = 1 - \sum_i \eta_i \left(\frac{1}{x} - \frac{1}{\lambda_i(\Lambda)}\right)^{M-1} u\left(1 - \frac{x}{\lambda_i(\Lambda)}\right)}$$

where $\eta_i = \frac{1}{\prod_{j \neq i}(\frac{1}{\lambda_j(\Lambda)} - \frac{1}{\lambda_i(\Lambda)})}$

- Use CDF to show that

$$\begin{aligned} R_{RBF} &= M \log\log n + M \log \frac{P}{M} + \\ & \log \lambda_1(\Lambda) + \sum_{i=1}^{M} \eta_i \log \left(\frac{\lambda_i}{\lambda_1}\right) \sum_{k=1}^{M-1} \frac{1}{k+2} (\frac{-1}{\lambda_i})^{M-1-k} \left. \frac{1}{y^{k+2}} \right|_{\lambda_1}^{\lambda_i} \end{aligned}$$

Sum Rate of Deterministic Beam Forming

- Sum-rate of deterministic beam forming

$$R_{BF-D} = M \log\log n + M \log \frac{P}{M} + \sum_{i=1}^{M} \log \left(\frac{1}{\phi_i^* U^* \Lambda^{-1} U \phi_i} \right)$$

$U^* \Lambda^{-1} U$ is the eigenvalue decomposition of R^{-1}

- Special case: $U\phi_i$'s are the columns of identity matrix

$$\boxed{R_{BF-D} = M \log\log n + M \log \tfrac{P}{M} + M \log \sqrt[M]{\det R}}$$

Since $\mathrm{tr}(R) = M$, the geometric mean satisfies

$$\sqrt[M]{\det(R)} \leq \frac{\mathrm{tr}(R)}{M} \leq 1$$

Sum rate of RBF with Channel Whitening

- For random beam forming with channel whitening,

$$S(t) = \sum_{m=1}^{M} \sqrt{\alpha} R^{-1/2} \phi_m s_m(t)$$

- Set $\alpha = \frac{\mathrm{tr}(R^{-1})}{M}$ to guarantee $E[S^* S] \leq 1$
- Scaling becomes the same as for white channel case with reduced signal power

$$\boxed{R_{BF-W} = M \log\log n + M \log \frac{P}{M} + M \log \frac{M}{\mathrm{tr}(R^{-1})}}$$

Simulations

- Consider a base station with $M = 2$ and $M = 3$ antennas
- The corresponding correlation matrix is

$$R = \begin{bmatrix} 1 & \alpha \\ \alpha & 1 \end{bmatrix}$$

$$R = \begin{bmatrix} 1 & \alpha & \alpha^2 \\ \alpha & 1 & \alpha \\ \alpha^2 & \alpha & 1 \end{bmatrix}$$

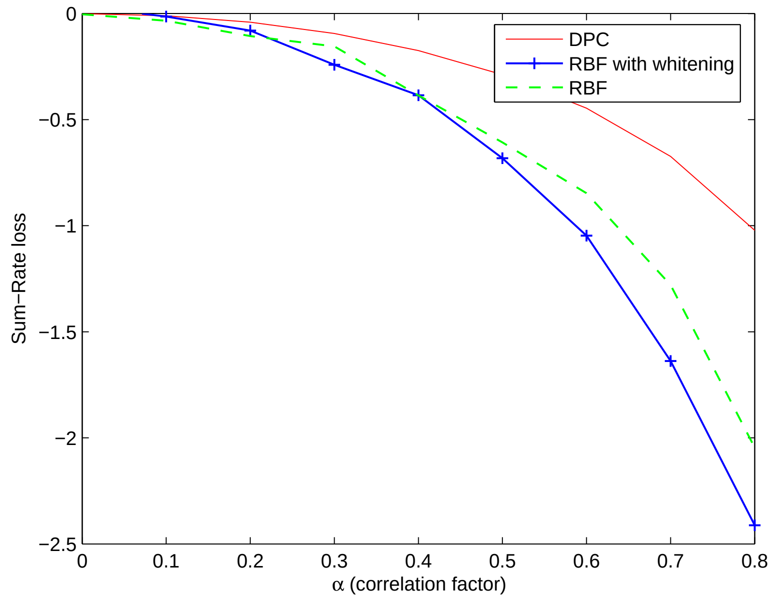

Figure 1: Sum-rate loss versus the correlation factor α for a system with $M = 2$ and $n = 100$.

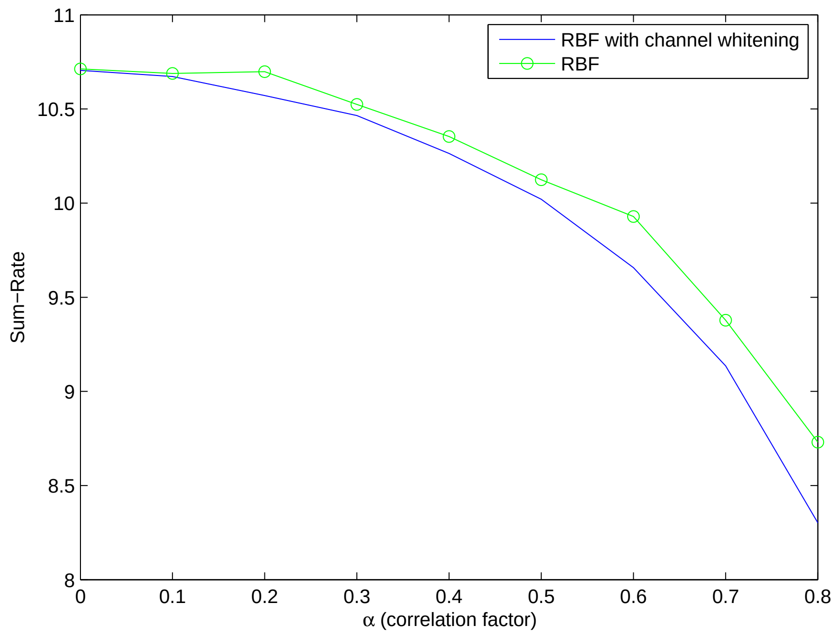

Figure 2: Sum-rate versus the correlation factor α for a system with $M = 2$, $P = 10$, and $n = 100$.

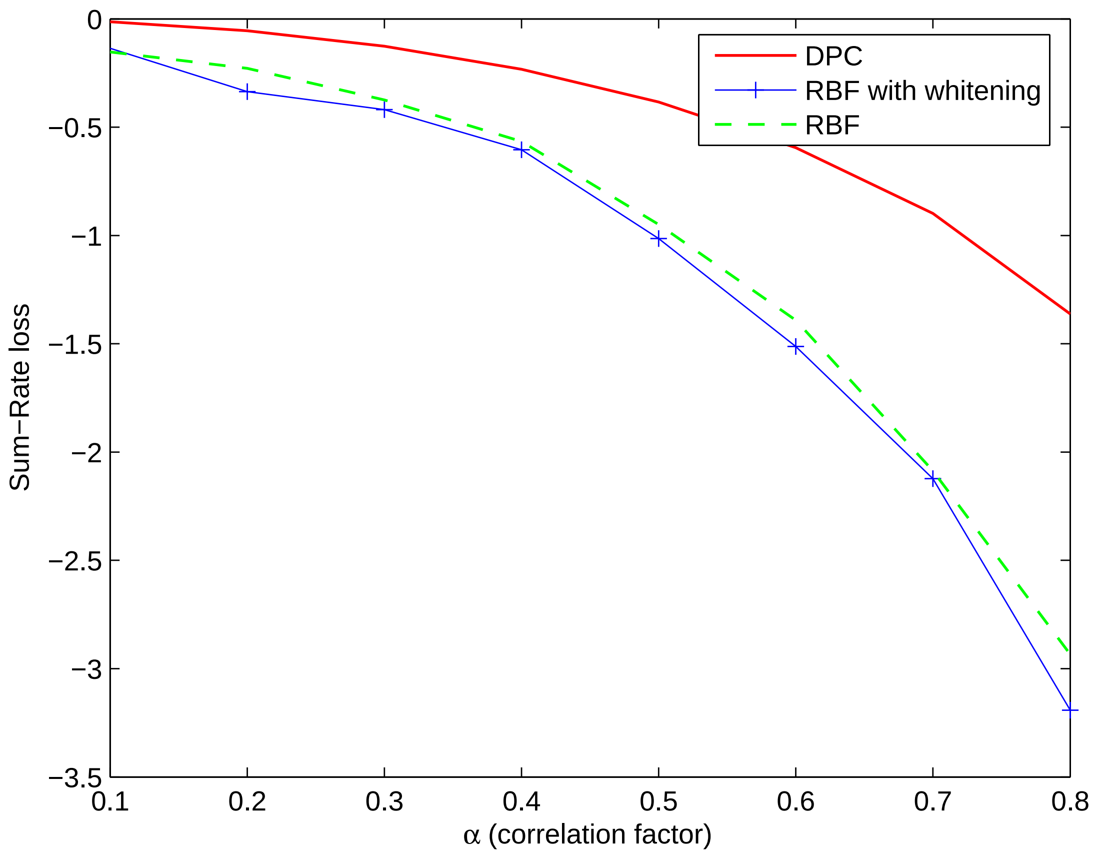

Figure 3: Sum-rate loss versus the correlation factor α for a system with $M = 3$ and $n = 100$.

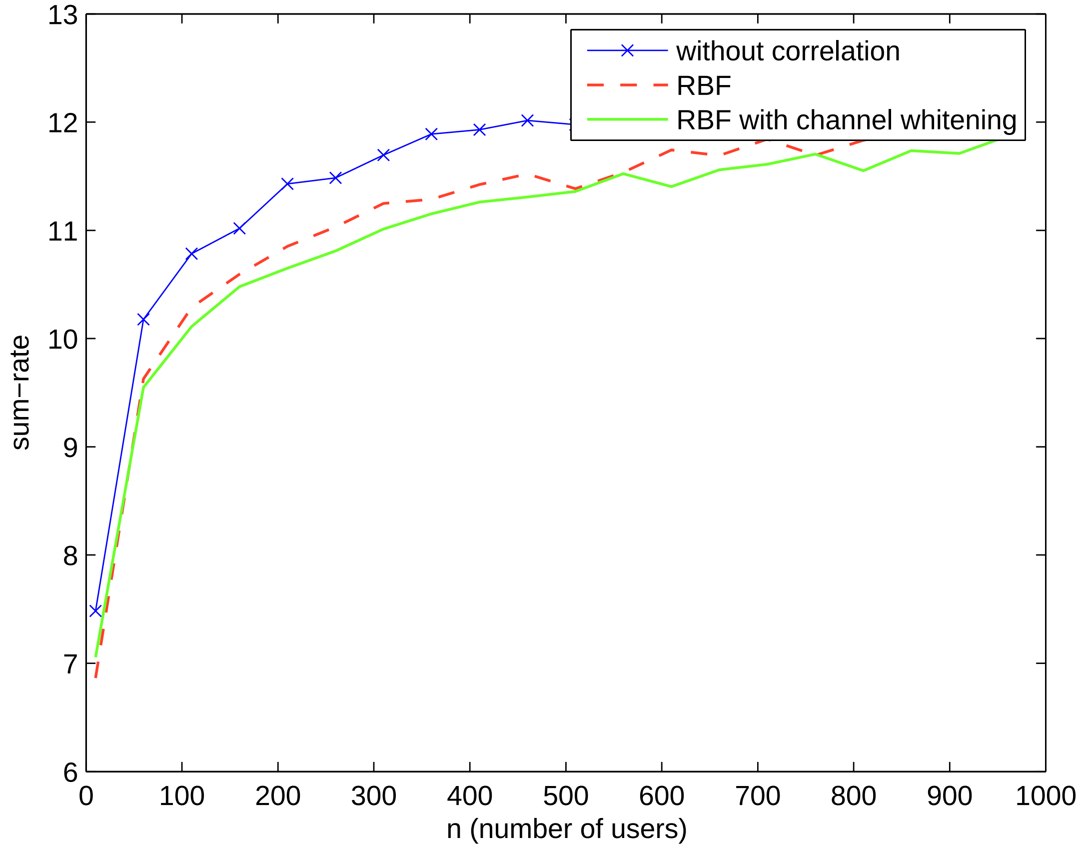

Figure 4: Sum-rate versus the number of users in a system with $M = 2$ and $\alpha = 0.5$

Can We Do Better?

- Apply a general precoding matrix

$$\alpha A S(t) = \alpha A \sum_{m=1}^{M} \phi_m(t) s_m(t), \qquad t = 1, \ldots, T$$

- The factor α ensures that we have a fixed power constraint

$$\alpha \leq \sqrt{\frac{M}{\operatorname{tr}(A^* A)}}$$

- This produces the effective channel

$$\tilde{H}_i = \alpha H_i A$$

with correlation $\alpha^2 \tilde{R} = \alpha^2 A^* R A$.

What is the Sum-Rate with a General Precoding?

- Sum-rate is given by

$$R_{\mathrm{PC}} = M \log \log n + M \log \frac{P}{M} - h_{\mathrm{PC}} \tag{10}$$

where h_{PC} is the hit incurred by using a general precoding matrix A

$$\boxed{h_{\mathrm{PC}} = M \log \frac{\operatorname{tr}(A^* A)}{M} + M E \log \|\phi_m\|^2_{\tilde{R}^{-1}}} \tag{11}$$

- Finding the optimum A that minimizes hit is difficult. But we can show that optimum precoding matrix A_{opt} can be written as

$$A_{opt} = Q_{Aopt} D_{Aopt}$$

where Q_{Aopt} is an orthornormal matrix and D_{opt} is diag with positive entries.

Special choices of A

- Difficult to optimize Q_{opt} and D_{opt} jointly.
- Set $Q_{opt} = Q_R$ as this will diagonalize R and optimize over D_{opt}.
- Zero forcing

  $$A_{\text{ZF}} = Q_R \Lambda_R^{-\frac{1}{2}}$$

  and resulting hit

  $$h_{ZF} = M \log \frac{tr(R^{-1})}{M}$$

Special Choices of A

- MMSE precoding

$$A = Q_R(\Lambda + \beta I)^{-\frac{1}{2}}$$

with β obtained as a solution to a fixed-pt problem

$$\frac{Tr(\Lambda + \beta^* I)^{-2}}{Tr(\Lambda + \beta^* I)^{-1}} = E\left(\frac{1}{\beta^* + \frac{1}{\|\Phi_m\|^2_{\Lambda^{-1}}}}\right)$$

- More generally, we can set $Q_{opt} = Q_R$ and find the optimum D_{opt}. Need to solve a set of M implicit equations

$$\frac{1}{d_i} E\left[\frac{\frac{1}{d_i \lambda_i}|\phi(i)|^2}{\|\phi\|^2_{D^{-1}\Lambda^{-1}}}\right] = \frac{1}{tr(D)}$$

Minimize an Upper Bound Instead

- Difficulty in minimizing h_{PC} due to the *phi* term

$$h_{\mathrm{PC}} = M \log \frac{\mathrm{tr}(A^* A)}{M} + M E \log \|\phi_m\|^2_{\tilde{R}^{-1}}$$

- Minimize an upper bound

$$h \leq M \log \mathrm{tr}(A^* A) + M \log \mathrm{tr}((A^* R A)^{-1})$$

- Can show that optimum A in this case is

$$\boxed{A = Q_R \Lambda_R^{-1/4}}$$

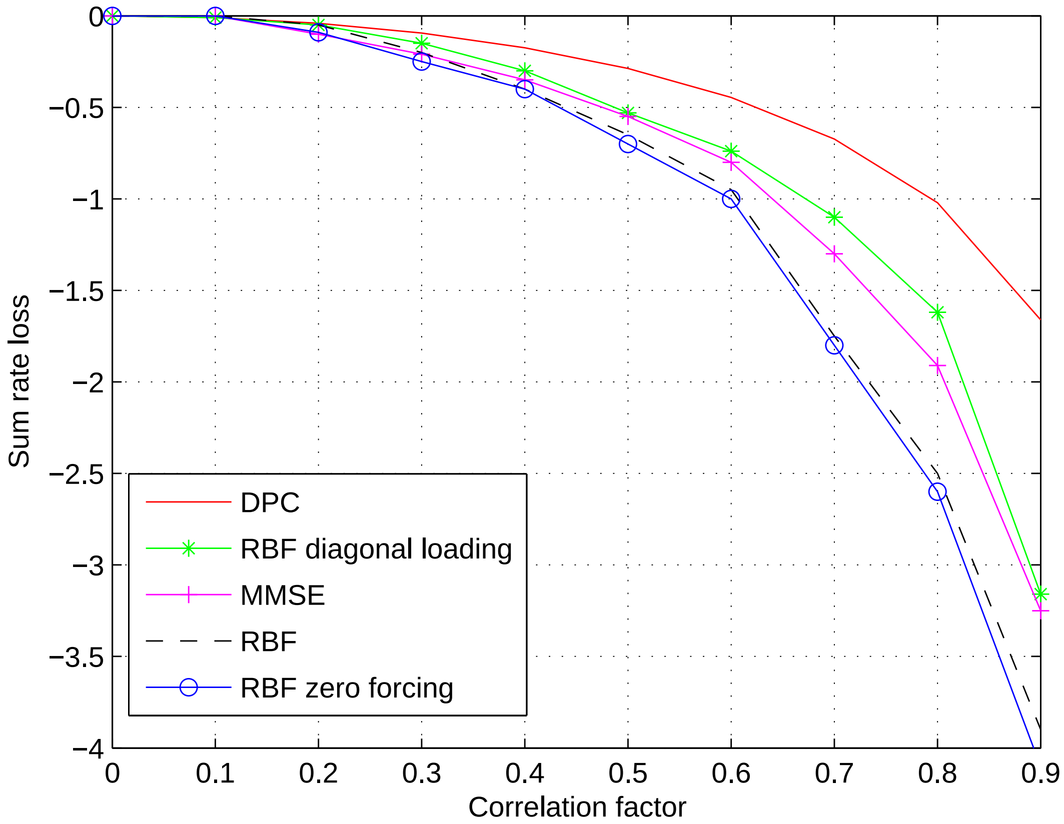

Figure 5: Sum-rate loss versus correlation factor α for a system with M = 3, P=10 and n = 200.

Conclusion

- Presented a new approach for calculating the (joint) distribution of indefinite quadratic forms in (Gaussian) random variables.
- Used these results to study the effect of spatial correlation on various multiuser schemes for MIMO broadcast channels.
- Considered DPC and random, deterministic, and channel whitening schemes.
- All these techniques exhibit the same scaling for iid channels

$$\boxed{R_{\text{sum-rate}} = M \log \log n + M \log \frac{P}{M}}$$

- In the presence of correlation between transmit antennas, scaling is

$$\boxed{R_{\text{sum-rate}} = M \log \log n + M \log \frac{P}{M} + M \log c}$$

Other Research Interests

- Compressive sensing for impulsive noise in OFDM (Giuseppe Caire)
- Group Broadcast Channels (Amir Dana and Babak Hassibi)
- Adaptive filtering analysis and design (Babak Hassibi and Vitor Nascimento; previously with Ali Sayed)
- Receiver design for (MIMO) OFDM in (block) time-variant channels (Naofal Al-Dhahir + Students) (previously with A. Paulraj)
- Blind channel estimation (Students)