Optimizing the Input Covariance for MIMO Channels

Alex J. Grant
Institute for Telecommunications Research,
University of South Australia, Australia
alex.grant@unisa.edu.au

Leif W. Hanlen
Wireless Signal Processing Program
National ICT Australia
leif.hanlen@nicta.com.au

***Abstract*— A multivariate Gaussian input density achieves capacity for multiple-input multiple-output flat fading channels with additive white Gaussian noise and perfect receiver channel state information. Capacity computation for these channels reduces to the problem of finding the best input covariance matrix, and is in general a convex semi-definite program. This paper presents Kuhn-Tucker optimality conditions for this problem, and iterative algorithms for determination of the optimal covariance matrix. For channels with a separable covariance structure, the optimal eigenvectors are known, and only the eigenvalues need to be determined. In the completely general case, both the eigenvectors and the eigenvalues must be found by numerical solutions. Large systems asymptotics are also considered.**

I. Introduction

Consider transmission is over a t-input, r-output additive white Gaussian noise channel of the form

$$y[k] = \sqrt{\gamma} H[k] x[k] + z[k] \tag{1}$$

where $y[k] \in \mathbb{C}^{r\times 1}$ is a complex column vector of matched filter outputs at symbol time $k = 1, 2, \ldots, N$ and $H[k] \in \mathbb{C}^{r\times t}$ is the corresponding matrix of complex channel coefficients. The element at row i and column j of $H[k]$ is the complex channel coefficient from transmit element j to receive element i. The channel matrices $H[k]$ will be selected independently at each symbol according to a matrix density p_H. It is assumed that $H[i]$ is known at the receiver, but that only the statistics p_H are known at the transmitter.

The vector $x[k] \in \mathbb{C}^{t\times 1}$ is the vector of complex baseband input signals, and $z[k] \in \mathbb{C}^{r\times 1}$ is a complex, circularly symmetric Gaussian vector with $\mathrm{E}\left[n[k]n[k]^\dagger\right] = I_r$. The superscript $(\cdot)^\dagger$ means Hermitian adjoint and I_r is the $r \times r$ identity matrix. The notation $(A)_{ij}$ denotes element ij of A. Let $n = \max(t, r)$ and $m = \min(t, r)$.

A power constraint $E[\|x[k]\|_2^2] \leq 1$ is enforced and the signal-to-noise ratio is γ. The input covariance is

$$Q = \mathrm{E}\left[x[k]x[k]^\dagger\right] \tag{2}$$

with $\mathrm{tr}(Q) \leq 1$. The assumptions of additive Gaussian noise and receiver knowledge of $H[k]$ mean that the optimal input density is Gaussian and determination of capacity C corresponds to optimizing Q. Defining

$$\Psi(Q) = I(x; y|H) = \mathrm{E}\left[\log\det\left(I + HQH^\dagger\right)\right], \tag{3}$$

we have the following convex semi-definite problem.

Problem 1:

$$C = \max_Q \Psi(Q) \text{ subject to}$$
$$\mathrm{tr}(Q) \leq 1$$
$$Q > 0$$
$$Q = Q^\dagger$$

The resulting optimal Q can depend on the channel statistics, but not on the channel realizations. For $H[k]$ with i.i.d. Gaussian entries, Telatar [1] showed that the optimizing $Q = I_t/t$. Telatar also gave an expression for computation of $\Psi(I)$, and several other expressions have subsequently been found [2–4].

Let $\mathcal{N}_{t,r}(M, \Sigma)$ be a multivariate Gaussian density with $r \times t$ mean matrix M and $rt \times rt$ covariance matrix $\Sigma = \mathrm{E}\left[hh^\dagger\right]$ where h is formed by stacking the columns of the matrix H into a single column vector. The Kronecker correlation model has $H \sim \mathcal{N}_{t,r}(M, R \otimes T)$ corresponding to separable transmit T and receive correlation R. This distribution may be generated via $H = M + R^{1/2} G T^{1/2}$ where $G \sim \mathcal{N}_{t,r}(0, I)$, and M corresponds to a "line-of-sight" channel component. In [5] it was shown that for $H \sim \mathcal{N}_{t,r}(0, I \otimes T)$ it is optimal to transmit independently on the eigenvectors of T. Closed form solutions have been obtained for $\Psi(I/t)$ where $H \sim \mathcal{N}_{t,r}(0, I \otimes T)$ [4, 6] and for $H \sim \mathcal{N}_{t,r}(0, R \otimes T)$, [7, 8]. Majorization results have also been obtained showing that stronger modes should be allocated higher powers [5]. Section II describes the Kuhn-Tucker optimality condition for these "diagonalizable" channels and a resulting iterative algorithm for computation of the optimal eigenvalues of Q. This approach is extended to arbitrary channel distributions in Section III. In this case both the eigenvectors and the eigenvalues must be determined.

Large systems ($r, t \to \infty$ with $r/t \to$ a constant) limits of $\Psi(I/t)$ have been obtained in [9], for $H \sim \mathcal{N}_{t,r}(0, R \otimes T)$. Asymptotic results for arbitrary Q were considered in [10], where Ψ was found to be asymptotically normal. Large-systems results have also been obtained in [11], concentrating on the case where the eigenvectors of the optimal Q can be

National ICT Australia is funded through the Australian Government's *Backing Australia's Ability initiative*, in part through the Australian Research Council.

identified by inspection. In Section IV, we show how to find the optimal power allocation for large systems with single-sided transmitter correlation.

All proofs can be found in [4, 12] and are omitted.

II. Diagonalizable Covariance

In certain cases Problem 1 can be simplified, and convenient optimality conditions can be obtained from the Kuhn-Tucker conditions. The simplest case, $S \sim \mathcal{N}_{r,t}(0, I \otimes I)$ was solved in [1]. Other special cases have been solved in [5, 13]. Independent work finding similar results to those described below has appeared in [11]. Suppose it can be determined that the optimal Q has the form

$$Q = U\hat{Q}U^\dagger \tag{4}$$

$$\hat{Q} = \operatorname{diag}(q_1, q_2, \ldots, q_t) \tag{5}$$

for some fixed U. Then the problem reduces to finding the best allocation of power to each column of U.

One important example is $H[k] \sim \mathcal{N}_{m,m}(0, R \otimes T)$, the Kronecker model with no line-of-sight. Then U diagonalizes T and optimal transmission is independent on each eigenvector of T. In such cases, $Q > 0 \implies \hat{Q} > 0$ allows the application of the Kuhn-Tucker conditions [14, p. 87]

$$\psi_i = \frac{\partial \Psi(Q)}{\partial q_i} \begin{cases} = \mu & q_i > 0 \\ \leq \mu & q_i = 0 \end{cases} \tag{6}$$

where μ is a constant independent of q_i. Differentiation leads to the following theorem, proved in [4].

Theorem 1: Consider the ergodic channel (1) with p_H such that the optimal input covariance is known to be of the form (4)-(5) for unitary U. A necessary and sufficient condition for the optimality of the diagonal $\hat{Q}$ in (5) is

$$\psi_i = \mathrm{E}_S\left[\left(\left(I + S\hat{Q}\right)^{-1} S\right)_{kk}\right] \begin{cases} = \mu & q_i > 0 \\ \leq \mu & q_i = 0 \end{cases} \tag{7}$$

for $k = 1, 2, \ldots, t$ and some constant μ. The expectation is with respect to the random matrix $S = \gamma U^\dagger H^\dagger H U$.

When $Q > 0$, the condition (7) may be re-written

$$\hat{Q} = \nu\, \mathrm{E}_S\left[\left(\hat{Q}^{-1} + S\right)^{-1} S\right], \tag{8}$$

which suggests the following iterative procedure for numerically finding the optimal $\hat{Q}$. Starting from an initial diagonal $\hat{Q}^{(0)} > 0$, compute

$$q_k^{(i+1)} = \nu^{(i)} \left[\mathrm{E}_S\left[\left((\hat{Q}^{(i)})^{-1} + S\right)^{-1} S\right]\right]_{kk}, \tag{9}$$

selecting $\nu^{(i)}$ at each step to keep $\operatorname{tr}\left(\hat{Q}^{(i)}\right) = \gamma$. In the general case, there is no known closed form solution for $\mathrm{E}_S\left[\left(\hat{Q}^{-1} + S\right)^{-1} S\right]$, and typically it may be accurately estimated using monte-carlo integration. Note that the numerical procedure may be applied to each entry $q_k = Q_{kk}$ separately for a given $\hat{Q}^{(i)}$. Numerically, each fixed point iteration is performed once and the t non-zero diagonal entries of $\hat{Q}$ are updated.

In the case $H[k] \sim \mathcal{N}_{t,r}(0, I \otimes T)$ however, a closed form expression is possible, using the expression for $\Psi(Q)$ developed in [4]. Let $a_i = \tau_i q_i$ and $t = r$ (for clarity only, other values can easily be accommodated). Then

$$\Psi(Q) = \frac{1}{t|V|} \sum_{k=1}^{t} |A^{(k)}| \tag{10}$$

where the $t \times t$ matrix V has elements $V_{ij} = a_j^{1-i}$ and the matrix $A^{(k)}$ has elements

$$A_{ij}^{(k)} = \begin{cases} \dfrac{a_j^{i-t}}{\Gamma(k)} \displaystyle\int_0^\infty \log(1 + a_j \lambda) \lambda^{k-1} e^{-\lambda}\, d\lambda & i = k \\ a_j^{i-t} & i \neq k \end{cases}$$

The partial derivative with respect to q_i is given by

$$\psi_i = \frac{\tau_i}{t|V|} \left(|V| \sum_k |A_i^{(k)}| - |V_i| \sum_k |A^{(k)}| \right)$$

where the matrices V_i, $A_i^{(k)}$ are formed from V and $A^{(k)}$ respectively via differentiation of column i with respect to $a_i = \tau_i q_i$ (other columns are unchanged). Extension to $\mathcal{N}_{t,r}(0, I \otimes T)$ may be possible using results from [7, 8].

It is interesting to compare the conditions (7), with the case of perfect transmitter side information. Suppose now $H[i] = H$ is time-invariant and known perfectly at the transmitter with $HH^\dagger = USU^\dagger$ being the eigenvalue decomposition of $HH^\dagger$. The Kuhn-Tucker condition for optimality of the input covariance $Q = U^\dagger \hat{Q} U^\dagger$ can be written in the following form,

$$\left(\left(I + S\hat{Q}\right)^{-1} S\right)_{kk} \begin{cases} = \mu & q_i > 0 \\ \leq \mu & q_i = 0 \end{cases} \tag{11}$$

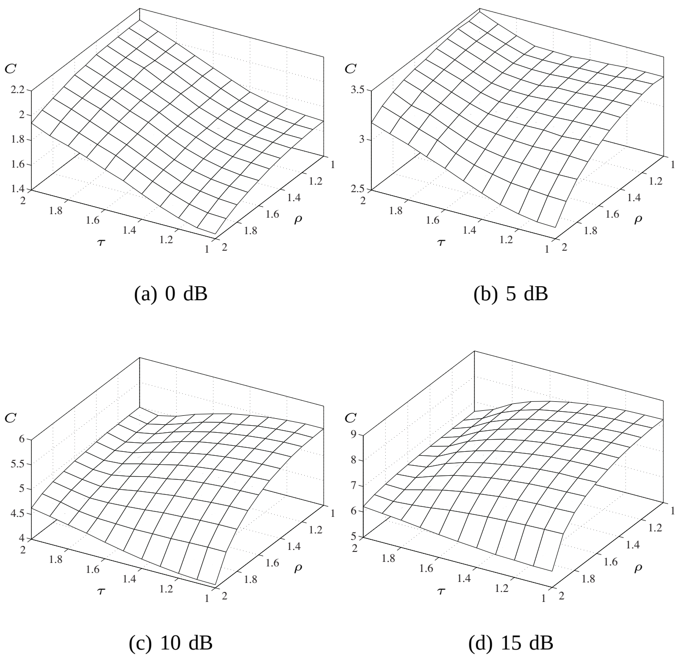
with $\hat{Q}$ satisfying (5). Solution of these equations is straightforward and leads easily to the well-known water filling result.

Comparing (7) with (11) it can be seen that the *only* difference is the presence of the expectation in (7). This is no real surprise, and is due to the interchangeability of differentiation and expectation. Thus Theorem 1 is a direct generalization of the classical water-filling result for parallel channels.

For the deterministic case, it is clear that increasing γ can only increase the power allocated to any particular eigenvector (water-level raises). The same thing happens in the ergodic case, as demonstrated by the following theorem.

Theorem 2: Let $\hat{Q} = \operatorname{diag}(q_1, \ldots, q_t)$ be the eigenvalues of the optimal covariance matrix for a channel with signal-to-noise ratio γ, satisfying the conditions of Theorem 1. Then $\partial q_k / \partial \gamma \geq 0$, $k = 1, 2, \ldots, t$.

Example 1: Let $H \sim \mathcal{N}_{2,2}(0, R \otimes T)$, where $T = \operatorname{diag}(\tau, 2 - \tau)$ and $R = \operatorname{diag}(\rho, 2 - \rho)$. Figure 1 shows the resulting capacity as a function of $1 \leq \tau \leq 2$ and $1 \leq \rho \leq 2$ (capacity is invariant to row permutations of either T or R, so these results cover the entire range of possible correlations).



(a) 0 dB (b) 5 dB

(c) 10 dB (d) 15 dB

Fig. 1. Capacity of $t = r = 2$ zero mean Rayleigh with separable correlation.

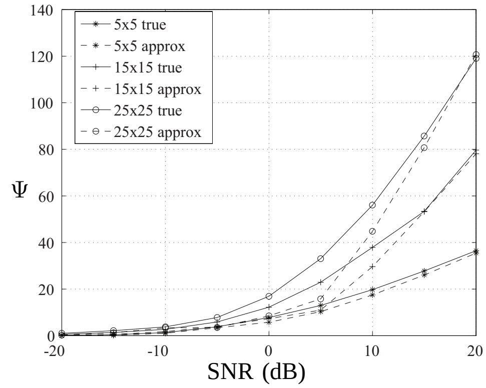


Fig. 2. Accuracy of the central Wishart approximation.

The effect of correlation on capacity can be either positive or negative, subject to the trace restriction $\mathrm{tr}(R) = \mathrm{tr}(T) = 2$. For fixed trace covariance matrices, at low SNR, high correlation increases capacity, since the channel is assisting with beam-forming.

Theorem 1 takes care of zero-mean Rayleigh fading channels with separable correlation structure. Suppose now $H \sim \mathcal{N}_{r,t}(M, I \otimes T)$. Then $S = HQH^\dagger$ may be approximated by a *central* Wishart matrix [15, p. 125] $S \sim W_t(0, \Sigma)$, where $\Sigma = T^{1/2}QT^{1/2} + \frac{1}{t}M^\dagger M$. This motivates heuristic application of Theorem 1 to Ricean channels $H[k] \sim \mathcal{N}_{r,t}(M, I \otimes T)$. The relation between correlation and line-of-sight (non-zero mean) has been heuristically established in MIMO channel measurement literature [16–18].

Example 2: An example of the accuracy of this approximation is investigated in Figure 2 for a $t = r$ channel with rank-one mean $M = \mathrm{diag}\{t, 0, \ldots, 0\}$ and non-diagonal transmit covariance $T = \tau\mathbf{1}_t + (\tau - 1)I_t$, where $\mathbf{1}_t$ is a $t \times t$ all-one matrix. The plot compares the capacity – solid lines (optimal input covariance, with *true* probability law) with the mutual information – dashed lines ($I(Q_a)$ where Q_a is optimal according to the central Wishart approximation *approximation*) for various SNR and numbers of transmit and receive elements. Note that the approximated covariance matrix is a linear combination of the transmit-end covariance T and the mean, and thus approximated input covariance is a dominated by beam-forming on M at low SNR, and T at higher SNR.

III. The General Case

We now wish to solve Problem 1, without the requirement of diagonal input covariance. In particular we do not wish to restrict ourselves to the zero-mean Kronecker Gaussian model. Specific examples of interest include the non-central channel, where the channel covariance and mean are not jointly diagonalizable, and several other models in the literature [19, 20].

To accommodate the positive definite constraint on Q, we apply the Cholesky factorization, so the constraint becomes implicit in the final solution. By adopting this approach we force the optimization to only consider the minimum number of independent variables required for solution, $t(t+1)/2$ rather than t^2.

Any non-negative matrix A may be written as $A = \Gamma^\dagger\Gamma$ for upper triangular matrix Γ, with real diagonal elements $d_{ii} \geq 0$. Similarly, for a given upper triangular matrix Γ, the product $\Gamma^\dagger\Gamma$ is positive definite. Problem 1 becomes

Problem 2:

$$\max_{\Gamma} \Psi(\Gamma^\dagger\Gamma) \text{ subject to}$$
$$\sum_{i\leq j} d_{ij}^2 = 1$$
$$d_{ii} \geq 0, \quad \forall i$$

Problem 2 admits a quadratic optimization approach, using Lagrange multipliers [21]. The optimization in Problem 2 occurs on the (upper triangular) matrix Γ which has *exactly* $t(t+1)/2$ independent (complex) variables. In order to solve Problem 2, consider the modified cost function $J(\nu, \mu, \phi)$ where $\nu = \vec{\Gamma}$, μ and ϕ are vectors of Lagrange multipliers corresponding to equality and inequality constraints. In general terms, given a convex $\cap$ function $f(\nu)$ of a vector ν, constrained by:

$$\sum_{i<j} \nu_{ij}^2 = 1 \quad \nu_{ii} \geq 0$$

the sufficient condition for optimality is

$$\frac{\partial f(\nu)}{\partial \nu_{ij}} \begin{cases} = 2\mu\nu_{ij}, & i \neq j, \mu > 0 \\ = 2\mu\nu_{ii}, & \nu_{ii} > 0, \mu > 0 \\ < 0 & \nu_{ii} = 0 \end{cases}$$

Which is the basis of the following result.

Theorem 3: Given the ergodic channel (1) with $H \sim p_H$, the capacity achieving input is Gaussian with covariance $Q = \Gamma^\dagger\Gamma$ where the elements d_{ij} of the upper triangular matrix Γ

satisfy

$$\mathrm{E}\Big[\mathrm{tr}\left[(I+S\Gamma^\dagger\Gamma)^{-1}SE^{(ij)}\right]\Big] \begin{cases} = 2\mu d_{ij} & i\neq j, \mu>0 \\ = 2\mu d_{ii} & d_{ii}>0, \mu>0 \\ <0 & d_{ii}=0 \end{cases} \tag{12, 13}$$

where the expectation is with respect to $S = H^\dagger H$, the constant μ is chosen to satisfy the power constraint and

$$E^{(ij)} = \frac{\partial \Gamma^\dagger \Gamma}{\partial d_{ij}}$$
$$\left(E^{(ij)}\right)_{mn} = d_{in}\delta_{mj} + d_{im}\delta_{nj}.$$

with $\delta_{ij} = 1$ when $i = j$ and zero otherwise.

The form of (12) suggests the following projected gradient algorithm [12], which is guaranteed to converge to the optimal covariance matrix.

Algorithm 1:

1) Update

$$\Gamma^{(k+1)} \to \Gamma^{(k)}\left(M + M^\dagger\right), \text{ where} \tag{14}$$
$$M = \mathrm{E}\left[(I + S\,\Gamma^\dagger\Gamma)^{-1}S\right] \tag{15}$$

2) Scale

$$\left[\Gamma^{(k+1)}\right]_{ij} \to \begin{cases} \frac{1}{\mu}\left[\Gamma^{(k+1)}\right]_{ij} & i \le j \\ 0 & \text{otherwise} \end{cases} \tag{16}$$

with μ constant for all i, j so that $\mathrm{tr}\left(\Gamma^\dagger\Gamma\right) = 1$.

3) Repeat

This algorithm may be initiated with any (upper triangular) Γ satisfying $\mathrm{tr}(\Gamma^\dagger\Gamma) = 1$. The expectation (14) is typically intractable and may be evaluated using monte-carlo integration.

The stability of the algorithm is directly affected by the stability of the expectation. In particular, at high-SNR, the off-diagonal entries of Γ will approach zero (since $Q = \alpha I$ is optimal). In this case, the elements of Γ may fluctuate as small movements over the Haar manifold (small changes in eigenvectors) result in large changes in the entries of Γ.

IV. Asymptotic Analysis

Large systems analysis is a powerful technique for analysis of linear matrix channels. The large systems assumption is $r, t \to \infty$ with $t/r \to \beta$, a constant. This asymptotic regime allows convenient application of the theory of large random matrices. Typically convergence to asymptotic values is quite fast, e.g. $r, t > 10$. In this section we consider the optimization of α in the large systems limit.

Suppose in the large systems limit, the empirical eigenvalue distribution of T/t converges to a given density $t(\lambda)$. Without loss of generality, we can assume that $T = \mathrm{diag}(\tau_1, \ldots, \tau_t)$ is diagonal, and that in the limit, the τ_i are chosen i.i.d. according to $t(\lambda)$. The corresponding input covariance matrix Q will be diagonal. The matrix Q can be specified by a transmit power allocation function $q(\lambda)$, which specifies how much power is transmitted on an eigenvector whose eigenvalue is λ.

Under these conditions, the objective function is

$$\frac{\Psi}{t} \to \eta_f(\gamma) = \int \log\left(1 + \gamma\lambda\right) f(\lambda)\, d\lambda$$

where $f(\lambda)$ is the limiting eigenvalue distribution of $GTQG^\dagger/t$, where $G \sim \mathcal{N}_{t,r}\left(0, I\right)$ (this eigenvalue density exists under the assumptions that we have made). Let $\xi = h(\lambda) = \lambda q(\lambda)$. The eigenvalue density of TQ is therefore $g(\xi) = t\left(h^{-1}(\xi)\right) dh^{-1}(\xi)/d\xi$.

Information theorists have recently labeled the function η_f as the *Shannon transform* of $f(x)$, although this integral has a long history in connection with the determination of the eigenvalues of random matrices. Its use in information theory is typically in the other direction, namely given an eigenvalue density $f(x)$, determine η_f as the corresponding mutual information.

The following parametric relation shows how to compute η_f in terms of the eigenvalue density g (which is easy to find), rather than f, which is hard to find.

$$\eta_f(\gamma) = \frac{1}{\beta}\eta_g(s) - \ln\frac{s}{\gamma} + \frac{s}{\gamma} - 1 \tag{17}$$
$$\gamma(s) = s\left(1 - \frac{s}{\beta}\eta_g'(s)\right)^{-1}. \tag{18}$$

This avoids working with Stieltjes transforms. It is easily verified that (17)-(18) is the same as

$$\eta_f(\gamma) = \frac{1}{\beta}\zeta_q(s) - \ln\frac{s}{\gamma} + \frac{s}{\gamma} - 1 \tag{19}$$
$$\gamma(s) = s\left(1 - \frac{s}{\beta}\zeta_q'(s)\right)^{-1}, \tag{20}$$
$$\zeta_q(s) = \int \log\left(1 + s\lambda q(\lambda)\right) t(\lambda)\, d\lambda. \tag{21}$$

The large-systems optimization problem is therefore

$$\max_{s,q} \frac{1}{\beta}\zeta_q(s) - \ln\frac{s}{\gamma} + \frac{s}{\gamma} - 1 \tag{22}$$
$$\text{subject to } q(\lambda) \ge 0 \tag{23}$$
$$s \ge 0 \tag{24}$$
$$\gamma - s \ge 0 \tag{25}$$
$$\int q(\lambda)t(\lambda)\, d\lambda - 1 = 0 \tag{26}$$
$$s\left(1/\gamma + \zeta_q'(s)/\beta\right) - 1 = 0 \tag{27}$$

Note that (20) is enforced in the constraint (27), and that (26) enforces the transmit power constraint $\mathrm{tr}(Q) \le 1$. It can be verified that the objective function is convex and that the constraints define a convex set, and standard numerical optimization procedures can be applied.

V. Conclusion

We have investigated the capacity achieving input covariance for MIMO channels in the case where the transmitter has statistical CSI. We have presented a method for calculating the optimal input covariance for arbitrary Gaussian vector channels. We have provided an iterative algorithm which converges to the optimal input covariance, by considering the covariance in terms of a Cholesky factorization. We have demonstrated the algorithm on several difficult channels, where the appropriate "diagonal" Q input cannot be readily found by inspection. Although the diagonalizing decomposition $Q = U\hat{Q}U^{\dagger}$ always exists, we have shown that the matrix U may be non-trivially related to the pdf of the channel.

For special cases, the optimal input covariance can be a-priori diagonalized by inspection – such as for zero-mean Kronecker correlated Rayleigh channels. In such cases we gave a simpler fixed point equation that characterizes the optimal transmit covariance. This particular characterization reveals a close link between the optimality condition for deterministic channels (water filling) and that for ergodic channels.

REFERENCES

[1] I. Emre Telatar, "Capacity of multi-antenna Gaussian channels," *Euro. Trans. Telecomm.*, vol. 10, no. 6, pp. 585–595, Nov. 1999.

[2] Alex J. Grant, "Rayleigh fading multiple-antenna channels," *EURASIP J. App. Sig. Proc., Special Issue on Space-Time Coding, Part I*, , no. 3, pp. 316 – 329, Mar. 2002.

[3] E. Biglieri and G. Taricco, "Transmission and reception with multiple antennas: Theoretical foundations," *Foundations and Trends in Communications and Information Theory*, vol. 1, no. 2, pp. 183–332, 2004.

[4] Leif W. Hanlen and Alex J. Grant, "Capacity analysis of correlated MIMO channels," Submitted to IEEE Trans. Inform. Theory., Sept. 2004.

[5] Syed Ali Jafar, Sriram Vishwanath, and Andrea Goldsmith, "Channel capacity and beamforming for multiple transmit and receive antennas with covariance feedback," in *Proc. IEEE Intl. Conf. Commun., ICC*, June 11–14 2001, vol. 7, pp. 2266–2271.

[6] Guiseppa Alfano, Antonia M. Tulino, Angel Lozano, and Sergio Verdú, "Capacity of MIMO channels with one-sided correlation," in *Proceedings IEEE Intl. Symp. Spread Spectrum Techniques and Applications ISSSTA*, 2004, pp. 515–519.

[7] M. Kiessling and J. Speidel, "Exact ergodic capacity of MIMO channels in correlated Rayleigh fading environments," *Int. Zurich Seminar*, Feb. 2004.

[8] Steven H. Simon and Aris L. Moustakas, "Eigenvalue density of correlated complex random Wishart matrices," *Phys E. Review*, vol. 69, pp. 065101–1 – 065101–4, June 11 2004.

[9] C.-N. Chuah, D. N. C. Tse, J. M. Kahn, and R. A. Valenzuela, "Capacity scaling in MIMO systems under correlated fading," *IEEE Trans. Inform. Theory*, vol. 48, no. 3, pp. 637–650, Mar. 2002.

[10] C. Martin and B. Ottersten, "Asymptotic eigenvalue distributions and capacity for MIMO channels under correlated fading," *IEEE Trans. Wireless Commun.*, vol. 3, no. 4, pp. 13500–1359, jul 2004.

[11] Antonia M. Tulino, Angel Lozano, and Sergio Verdu, "Impact of antenna correlation on the capacity of multiantenna channels," *IEEE Trans. Inform. Theory*, vol. 51, no. 7, pp. 2491–2509, July 2005.

[12] Leif W. Hanlen and Alex J. Grant, "Optimal transmit covariance for ergodic MIMO channels," submitted to IEEE Trans. Inform. Theory, Oct. 2005.

[13] Syed Ali Jafar and Andrea Goldsmith, "Multi-antenna capacity in correlated Rayleigh fading with channel covariance information," in *Proc. IEEE Intl. Symp. Inform. Theory, ISIT*, Yokohama, Japan, June 29–July 4 2003, p. 470.

[14] Robert G. Gallager, *Information Theory and Reliable Communication*, John Wiley & Sons, New York, USA, 1968.

[15] A. K. Gupta and D. K. Nagar, *Matrix Variate Distributions*, vol. 104 of *Monographs and Surveys in Pure and Applied Mathematics*, Chapman & Hall/CRC, Boca Raton, 2000.

[16] P. Kyritsi, D. C. Cox, R. A. Valenzuela, and P.W. Wolniansky, "Correlation analysis based on MIMO channel measurements in an indoor environment," *IEEE J. Select. Areas Commun.*, vol. 21, no. 5, pp. 713–720, June 2003.

[17] K. Yu, M. Bengtsson, B. Ottersten, D. McNamara, P. Karlsson, and M. Beach, "Second order statistics of NLOS indoor MIMO channels based on 5.2 GHz measurements," *IEEE Trans. Signal Processing*, vol. 49, no. 5, pp. 1002–1012, May 2001.

[18] Kai Yu and B. Ottersten, "Models for MIMO propagation channels: A review," *Wireless Communications and Mobile Computing*, vol. 2, pp. 653–666, Nov. 2002.

[19] Huseyin Ozcelik, Nicolai Czink, and Ernst Bonek, "What makes a good MIMO channel model?," in *Proc. IEEE Vehic. Techn. Conf., VTC*, 2005.

[20] H. Ozcelik, M. Herdin, W. Weichselberger, J., and E. Bonek, "Deficiencies of 'Kronecker' MIMO radio channel model," *IEE Elect. Lett.*, vol. 39, no. 16, pp. 1209–1210, Aug. 7 2003.

[21] Stephen Boyd and Lieven Vandenberghe, *Convex Optimization*, Cambridge University Press, Cambridge, United Kingdom, 2004.