On Rate-Optimal MIMO Signalling with Mean and Covariance Feedback

Ramy H. Gohary, *Member, IEEE*, and Timothy N. Davidson, *Member, IEEE*

***Abstract*—We consider a single-user multiple-input multiple-output (MIMO) communication system in which the transmitter has access to both the channel covariance and the channel mean. For this scenario, we provide an explicit second-order approximation of the ergodic capacity of the channel, and we use this approximation to show that when the channel has a non-zero mean, the basis of the optimal input covariance matrix depends on the input signal power. (This basis is independent of the signal power in the zero-mean case.) The second-order approximation also provides insight into the way in which the low-signal-to-noise-ratio (SNR) optimal input covariance matrix is related to the optimal input covariance matrix at arbitrary SNRs. Furthermore, we show that the design of the input covariance matrix that optimizes the second-order approximation can be cast as a convex optimization problem for which the Karush-Kuhn-Tucker (KKT) conditions completely characterize the optimal solution. Using these conditions, we provide an efficient algorithm for obtaining second-order optimal input covariance matrices. The resulting covariances confirm our theoretical observation that, in general, the low-SNR optimal signal basis does not coincide with the optimal basis at higher SNRs. Finally, we show how our second-order design algorithm can be used to efficiently obtain input covariance matrices that provide ergodic rates that approach the ergodic capacity of the system.**

***Index Terms*—MIMO communication systems, ergodic capacity, statistical channel state information, correlated channel with non-zero mean, Kronecker channel model.**

I. Introduction

THE availability of channel state information (CSI) at the transmitter of a multiple-input multiple-output (MIMO) communication system can have a significant impact on the maximum ergodic rate that can be reliably communicated over the channel [1]. While it is desirable that the transmitter has access to perfect (instantaneous) CSI, this may not be realistic in many practical scenarios. Instead, it may be more practical to assume that the transmitter is given access to statistical channel information, while the receiver has access to instantaneous CSI.

Manuscript received December 18, 2007; revised March 13, 2008 and April 25, 2008; accepted May 5, 2008. The associate editor coordinating the review of this paper and approving it for publication was R. Mallik.

This work was supported in part by a Premier's Research Excellence Award from the Government of Ontario. The work of the second author is also supported by the Canada Research Chair program.

R. H. Gohary was with the Department of Electrical and Computer Engineering, McMaster University, Hamilton, Ontario, L8S 4K1, Canada. He is now with the Communications Research Centre, Industry Canada, Ottawa, Ontario, K2H 8S2, Canada (e-mail: ramy.gohary@crc.ca).

T. N. Davidson is with the Department of Electrical and Computer Engineering, McMaster University, Hamilton, Ontario, L8S 4K1, Canada (e-mail: davidson@mcmaster.ca).

Digital Object Identifier 10.1109/TWC.2009.071417

The impact that statistical CSI at the transmitter has on the maximum achievable ergodic rate depends on the signal-to-noise ratio (SNR) at which the system operates. In particular, while transmitter CSI typically plays a significant role in low-SNR regimes in which the transmit power is only sufficient to excite a subset of the eigen modes of the channel [2], [3], transmitter CSI may be less significant at high SNRs [4], [5].

The exploitation of statistical CSI at the transmitter is typically accomplished via the design of the input covariance matrix. With an ergodic rate objective in mind, the design of optimal input covariance matrices for a variety of scenarios have been considered. In particular, multiple-input single-output (MISO) systems were considered in [6], [7] and a comprehensive characterization of rate-optimal transmission strategies for these systems was provided in [8]. For MIMO systems, the work in [6], [7] was extended in [9]–[11] to zero-mean channels with Kronecker-structured covariance (cf. [1]), and to non-zero mean channels in [12]. A prominent theme in those studies was to identify scenarios under which beamforming is the optimal transmission strategy; that is, scenarios in which the rate-optimal input covariance matrix is rank one. The structure of rate-optimal designs for scenarios in which the channel is zero-mean but the covariance has a more general structure than the Kronecker model was considered in [13]. While obtaining closed-form expressions for ergodic rate-optimal input covariance matrices seems to be quite difficult in general [14], significant insight into the structure of those matrices can be drawn by studying asymptotically low-SNR and high-SNR regimes, [2], [3], [15] and [16], [17], respectively. In addition to ergodic rate objectives, the design of the input covariance matrices can be tailored to meet other pragmatic objectives; e.g., outage capacity [18] and various error rate perspectives [19], [20].

Our goal in this paper is to study the optimization of the input covariance with an ergodic rate objective for communication scenarios in which the transmitter has access to both the covariance and the mean of the channel. We draw on our earlier results in [15] in which we provided explicit formulae for the rate-optimal input covariance matrix at low SNR. That design was based on a first-order approximation of the ergodic capacity. In [15] we considered arbitrarily correlated non-zero mean channels, but in order to maintain mathematical tractability, in this paper, we restrict our attention to non-zero mean channels with Kronecker-structured covariance. In particular, we provide a second-order approximation of the ergodic capacity of these channels, and we study optimal input covariance design for this approximation. By comparing the second-order designs with their first-order counterparts, we will show that unlike zero-mean channels (cf. [11]), the basis (i.e., the eigenvectors) of the rate-optimal input covariance matrix depends on the input signal power. Although our results were originally devised to draw insight into low-SNR signalling regimes, we will show how those results are intimately related to the second-order approximation of the ergodic capacity at arbitrary SNRs. In order to use this observation to devise a rate-efficient signalling strategy at any SNR, we analyze the second-order approximation and expose its underlying convexity in the transmit covariance for every choice of the centre point in the Taylor expansion. This convexity, and the structure of the constraint set, ensure that satisfying the Karush-Kuhn-Tucker (KKT) conditions is both necessary and sufficient for obtaining a transmit covariance that maximizes the second-order approximation [21], [22]. By solving these conditions, the design problem is reduced to finding two scalars that can be efficiently obtained using a bisection search.[1] In addition to the effectiveness of the obtained solutions, those solutions confirm our earlier observation that, in general, the optimal signalling basis is power-dependent. Finally, we provide numerical results that show that by adjusting the centre point in the Taylor expansion, the covariance matrices obtained through the convex optimization of the second-order approximation yield ergodic rates that approach the ergodic capacity.

Notation: Throughout the paper we will adopt the convention of using upper case bold symbols to denote random matrices and lower case bold symbols to denote random scalars. Deterministic matrices, vectors and scalars will be denoted by regular weight symbols.

II. Channel Model and Previous Results

We consider a MIMO communication system with M transmit and N receive antennas. The channel, $\mathbf{H}$, is assumed to be Ricean narrowband block-fading with a non-zero mean, $\bar{H}$, and a Kronecker-structured covariance (cf. [1]) with an $M\times M$ transmit covariance matrix T and an $N\times N$ receive covariance matrix R, where both T and R are positive semidefinite matrices. In this case the channel model can be written as

$$\mathbf{H} = R^{1/2}\mathbf{H}_w T^{1/2} + \bar{H}, \tag{1}$$

where $\mathbf{H}_w$ is a standard zero-mean circularly-symmetric complex Gaussian matrix with unit-variance independent identically distributed (i.i.d.) entries. Throughout this paper we will normalise the matrices T and R so that $\mathrm{Tr}(R) = \mathrm{Tr}(T) = 1$. Although the Kronecker model is not the most general model available, it will be adopted in this paper model because its mathematical tractability generates considerable insight into the design of the transmit covariance.

Using the results in [23] pertaining to MIMO systems with perfect (instantaneous) CSI at the receiver, the ergodic capacity for additive circularly-symmetric zero-mean Gaussian noise with identity covariance can be expressed as

$$C = \max_{Q\succeq 0,\mathrm{Tr}(Q)=P} \mathrm{E}_{\mathbf{H}_w}\{\log\det(I_N + \mathbf{H}Q\mathbf{H}^\dagger)\}, \tag{2}$$

[1]Note that even for zero-mean channels, in which the optimal signalling bases are power-independent, finding the optimal signalling strategy using existing methods involves a rather tedious multidimensional stochastic optimization over the eigenvalues of the input covariance matrix; e.g., [11].

where Q is the input covariance matrix, P is the power budget, $\mathbf{H}$ is related to $\mathbf{H}_w$ via (1), and '$\succeq$' is the standard Lowener ordering whereby $A \succeq B$ indicates that $A - B$ is positive semidefinite [24]. Solving (2) directly for a general case of the model in (1) appears to be a difficult problem. Instead, some progress can be made by using the Taylor expansion. In particular, if the power budget P is sufficiently small so that the maximum eigenvalue of $\mathbf{H}Q\mathbf{H}^\dagger$, $\lambda_{\max}(\mathbf{H}Q\mathbf{H}^\dagger) < 1$ with high probability, then (e.g., [15]) one can express the the right hand side of (2) as

$$C = \max_{Q\succeq 0,\mathrm{Tr}(Q)=P} \mathrm{E}_{\mathbf{H}_w}\big\{\mathrm{Tr}(\mathbf{H}Q\mathbf{H}^\dagger) - \tfrac{1}{2}\mathrm{Tr}\big((\mathbf{H}Q\mathbf{H}^\dagger)^2\big) + \tfrac{1}{3}\mathrm{Tr}\big((\mathbf{H}Q\mathbf{H}^\dagger)^3\big) + \cdots\big\}. \tag{3}$$

In [15] we studied the first-order term in (3) under a correlation model that is more general than the one in (1). When this model is restricted to the Kronecker model in (1), it was shown that [15] the optimal signalling strategy is to perform beamforming along the principal eigenvector of

$$X = \mathrm{Tr}(R)T + \bar{H}^\dagger\bar{H} = T + \bar{H}^\dagger\bar{H}. \tag{4}$$

As pointed out in [15], for the zero-mean case, the receiver correlation matrix R plays no role in determining the beam direction, which agrees with the results in [11], whereas in the case of a non-zero mean, the trace of R affects the angle by which the beam is 'steered' away from the direction of the mean. While the results in [15] were derived for the low-power regime, we showed therein that there are instances in which signalling along the eigenvectors of the matrix X in (4) can provide significant gains over standard signalling strategies at quite large SNRs. These observations prompted us to investigate whether the low-SNR optimality of signalling along the eigen basis of (4) extends to higher SNRs (as it does in the zero-mean case). In Section IV we will use the second-order approximation of (3) to provide a negative answer to that question. In particular, we will show that, unlike the zero-mean case, the optimal signalling basis is power dependent in the general case.

III. Second-Order Approximation of the Low-SNR Ergodic Capacity

In order to answer the question posed in the previous section regarding the optimality of signalling along the eigen basis of (4) at any SNR, we study the second-order term in (3),

$$\tau = \mathrm{E}_{\mathbf{H}_w}\big\{\mathrm{Tr}\big(\mathbf{H}Q\mathbf{H}^\dagger\big)^2\big\}. \tag{5}$$

Substituting (1) into (5), we have

$$\tau = \mathrm{E}_{\mathbf{H}_w}\big\{\mathrm{Tr}\big(\mathbf{A}^2 + 2\mathbf{A}(\mathbf{B}+\mathbf{B}^\dagger) + 2\mathbf{A}C + 2(\mathbf{B}+\mathbf{B}^\dagger)C + (\mathbf{B}+\mathbf{B}^\dagger)(\mathbf{B}+\mathbf{B}^\dagger) + C^2\big)\big\}, \tag{6}$$

where

$$\mathbf{A} = R^{1/2}\mathbf{H}_w T^{1/2} Q T^{1/2}\mathbf{H}_w^\dagger R^{1/2},$$
$$\mathbf{B} = R^{1/2}\mathbf{H}_w T^{1/2} Q \bar{H}^\dagger, \quad C = \bar{H}Q\bar{H}^\dagger. \tag{7}$$

To simplify the expression in (6), we observe that, using the definitions in (7), $\mathrm{E}_{\mathbf{H}_w}\{\mathrm{Tr}(\mathbf{B}C)\} = 0$, and $\mathrm{E}_{\mathbf{H}_w}\{\mathrm{Tr}(\mathbf{B}^\dagger C)\} = 0$. Hence, one can write (6) as

$$\tau = \mathrm{E}_{\mathbf{H}_w}\big\{\mathrm{Tr}\big(\mathbf{A}^2 + 2\mathbf{A}(\mathbf{B}+\mathbf{B}^\dagger) + 2\mathbf{A}C + \mathbf{B}\mathbf{B} + \mathbf{B}^\dagger\mathbf{B}^\dagger + 2\mathbf{B}^\dagger\mathbf{B}\big)\big\} + \mathrm{Tr}(C^2). \quad (8)$$

This expression can be further simplified by showing (as we do in Appendix A) that

$$\mathrm{E}_{\mathbf{H}_w}\{\mathrm{Tr}(\mathbf{A}\mathbf{B})\} = \mathrm{E}_{\mathbf{H}_w}\{\mathrm{Tr}(\mathbf{A}\mathbf{B}^\dagger)\} = 0, \quad (9)$$
$$\mathrm{E}_{\mathbf{H}_w}\{\mathrm{Tr}(\mathbf{B}\mathbf{B})\} = \mathrm{E}_{\mathbf{H}_w}\{\mathrm{Tr}(\mathbf{B}^\dagger\mathbf{B}^\dagger)\} = 0, \quad (10)$$

and hence that[2] $\tau = \mathrm{E}_{\mathbf{H}_w}\big\{\mathrm{Tr}\big(\mathbf{A}^2 + 2\mathbf{A}C + 2\mathbf{B}^\dagger\mathbf{B}\big)\big\} + \mathrm{Tr}(C^2)$. In Appendix B we show that

$$\mathrm{E}_{\mathbf{H}_w}\{\mathrm{Tr}(\mathbf{A}C)\} = \mathrm{Tr}(TQ)\,\mathrm{Tr}(R\bar{H}Q\bar{H}^\dagger), \quad (11)$$
$$\mathrm{E}_{\mathbf{H}_w}\{\mathrm{Tr}(\mathbf{B}^\dagger\mathbf{B})\} = \mathrm{Tr}(TQ\bar{H}^\dagger\bar{H}Q)\,\mathrm{Tr}(R), \quad (12)$$

and in Appendix C we show that

$$\mathrm{E}_{\mathbf{H}_w}\{\mathrm{Tr}(\mathbf{A}^2)\} = \mathrm{Tr}(R^2)\big(\mathrm{Tr}(TQ)\big)^2 + \mathrm{Tr}(TQTQ)\big(\mathrm{Tr}(R)\big)^2. \quad (13)$$

Substituting from (11), (12) and (13) into (8), and regrouping terms, we have that

$$\tau = \mathrm{Tr}(R^2)\big(\mathrm{Tr}(TQ)\big)^2 + 2\,\mathrm{Tr}(TQ)\,\mathrm{Tr}(R\bar{H}Q\bar{H}^\dagger) + \mathrm{Tr}\Big(\big((\mathrm{Tr}(R)T + \bar{H}^\dagger\bar{H})Q\big)^2\Big). \quad (14)$$

In the sequel, we will show how this expression can be used to draw insight into the optimal signalling basis at any SNR. This expression will also be instrumental in developing an efficient algorithm for generating second-order optimal input covariance matrices that perform better than other commonly-used signalling techniques.

IV. First-order versus Second-order optimal basis

The goal in this section is to show that, unlike zero mean channels, when the channel is non-zero mean the basis of the rate-optimal input covariance depends on the signal power. In order to do that, it is sufficient to find an instance of T, R and $\bar{H}$ for which the basis of the first-order rate-optimal input covariance matrix does not coincide with that of the second-order rate-optimal matrix.

For the first-order rate-optimal transmission, the basis of the input covariance matrix is given by the eigenvectors of $X = T + \bar{H}^\dagger\bar{H}$ (cf. (4)), whereas for up-to second-order optimality, the optimum input covariance matrix must maximize (cf. (14))

$$F = \mathrm{Tr}(XQ) - \tfrac{1}{2}\big(\mathrm{Tr}(R^2)\big(\mathrm{Tr}(TQ)\big)^2 + 2\,\mathrm{Tr}(TQ)\,\mathrm{Tr}(R\bar{H}Q\bar{H}^\dagger) + \mathrm{Tr}(XQXQ)\big). \quad (15)$$

Now, consider the case in which $T = \frac{1}{M}I_M$. Substituting for T into (15), and using the fact that $\mathrm{Tr}(Q) = P$, we obtain

$$F = \mathrm{Tr}\big((\tfrac{1}{M}I_M + \bar{H}^\dagger\bar{H})Q\big) - \tfrac{1}{2}P^2\,\mathrm{Tr}(R^2) - P\,\mathrm{Tr}(R\bar{H}Q\bar{H}^\dagger) - \tfrac{1}{2}\,\mathrm{Tr}\Big(\big((\tfrac{1}{M}I_M + \bar{H}^\dagger\bar{H})Q\big)^2\Big).$$

[2] In fact, due to the circular symmetry of the complex entries of H_w, one can show that, in general, if H_w and $H_w^\dagger$ do not appear an equal number of times in an expression, then the expected value of that expression is equal to zero.

When $T = \frac{1}{M}I_M$, the basis of the first-order optimal input covariance matrix coincides with the eigen-basis of $\bar{H}^\dagger\bar{H}$ irrespective of the value of R. Now, if we consider the case in which $\mathrm{Tr}(R) \gg \lambda_{\max}(\bar{H}^\dagger\bar{H})$, it can be shown, using [24, Example 7.4.13], that the basis of the second-order optimal solution approaches the eigen-basis of $\bar{H}^\dagger R\bar{H}$. Hence, we conclude that for channels with non-zero mean the first-order optimal basis does not necessarily coincide with the optimal basis for higher-order approximations. In Section VII we provide a numerical example that illustrates the difference between first and second order rate-optimal bases.

V. Power-series expansion of capacity at any SNR

In this section we demonstrate the relevance of low-SNR optimal signalling to optimal signalling at any SNR. In particular, we show that, apart from a constant (covariance-independent) term, the capacity is dominated by the low-order terms in a power series that is intimately related to the low-SNR expansion in (3) of the ergodic capacity. Towards that end, we denote the eigen decomposition of $\mathbf{H}Q\mathbf{H}^\dagger$ as $\mathbf{\Psi}\mathbf{\Theta}\mathbf{\Psi}^\dagger$ and write

$$\begin{aligned} C_Q &= \mathrm{E}_{\mathbf{H}_w}\big\{\log\det(I_N + \mathbf{H}Q\mathbf{H}^\dagger)\big\} \\ &= \mathrm{E}_{\mathbf{H}_w}\big\{\mathrm{Tr}\log(I_N + \mathbf{H}Q\mathbf{H}^\dagger)\big\} \\ &= \mathrm{E}_{\mathbf{\Theta}}\big\{\mathrm{Tr}\log(I_N + \mathbf{\Theta})\big\}, \end{aligned} \quad (16)$$

where in the last equality in (16) we have denoted the diagonal matrix of the ordered non-negative (random) eigenvalues of $\mathbf{H}Q\mathbf{H}^\dagger$ by $\mathbf{\Theta}$, and the expectation is taken over $\{\boldsymbol{\theta}_i\}$, where $\boldsymbol{\theta}_i$ denotes the i-th eigenvalue of $\mathbf{H}Q\mathbf{H}^\dagger$. For the case in which the channel is zero-mean and has a Kronecker-structured correlation, a closed-form expression for this expectation is available [25]. However, that expression is not readily extensible to the nonzero-mean case, and, instead, we seek insight into the expectation by writing it as

$$\mathrm{E}_{\mathbf{\Theta}}\big\{\mathrm{Tr}\log(I_N + \mathbf{\Theta})\big\} = \int_{\infty>\theta_1\geq\cdots\geq\theta_N\geq 0} p_{\mathbf{\Theta}}(\Theta)\,\mathrm{Tr}\log(I_N + \Theta)d\Theta, \quad (17)$$

where $p_{\mathbf{\Theta}}(\Theta)$ is the joint probability density function of $\{\boldsymbol{\theta}_i\}$. Partitioning the integral, we write

$$\mathrm{E}_{\mathbf{\Theta}}\big\{\mathrm{Tr}\log(I_N + \mathbf{\Theta})\big\} = \int_{M_\epsilon\geq\theta_1\geq\cdots\geq\theta_N\geq 0} p_{\mathbf{\Theta}}(\Theta)\,\mathrm{Tr}\log(I_N + \Theta)d\Theta + \int_{\infty>\theta_1\geq\cdots\geq\theta_N> M_\epsilon} p_{\mathbf{\Theta}}(\Theta)\,\mathrm{Tr}\log(I_N + \Theta)d\Theta, \quad (18)$$

where the scalar $M_\epsilon < \infty$ is chosen to be sufficiently large so as to ensure that

$$\int_{\infty>\theta_1\geq\cdots\geq\theta_N> M_\epsilon} p_{\mathbf{\Theta}}(\Theta)\,\mathrm{Tr}\log(I_N + \Theta)d\Theta \leq \epsilon, \quad (19)$$

for some sufficiently small $\epsilon > 0$. The existence of such a scalar, M_ϵ, is guaranteed for all statistical distributions of $\mathbf{\Theta}$ for which the improper integral in (17) converges [26]; that is, for all distributions of $\mathbf{\Theta}$ for which the ergodic capacity is finite.

Assuming ϵ to be sufficiently small and choosing M_ϵ to satisfy (19), we can write

$$\begin{aligned} C_Q &\approx \mathrm{E}_{\boldsymbol{\Theta}}\big\{\operatorname{Tr}\log\big(\boldsymbol{\Theta} - M_\epsilon I_N + (M_\epsilon+1)I_N\big)\big\} \\ &= N\log(M_\epsilon+1) + \mathrm{E}_{\boldsymbol{\Theta}}\big\{\operatorname{Tr}\log\big(\tfrac{1}{M_\epsilon+1}(\boldsymbol{\Theta} - M_\epsilon I_N) + I_N\big)\big\}, \end{aligned} \tag{20}$$

where in (20) we assume that $\{\boldsymbol{\theta}_i\}$ are restricted to the domain of the first integral in (18). Since in this domain $\left|\frac{\boldsymbol{\theta}_1 - M_\epsilon}{M_\epsilon+1}\right| < 1$, we can use the Taylor expansion about $M_\epsilon I_N$ to write (20) as

$$\begin{aligned} C_Q \approx N\log(M_\epsilon+1) + \mathrm{E}_{\boldsymbol{\Theta}}\big\{\operatorname{Tr}\big(&\tfrac{1}{(M_\epsilon+1)}(\boldsymbol{\Theta} - M_\epsilon I_N) \\ &- \tfrac{1}{2(M_\epsilon+1)^2}(\boldsymbol{\Theta} - M_\epsilon I_N)^2 \\ &+ \tfrac{1}{3(M_\epsilon+1)^3}(\boldsymbol{\Theta} - M_\epsilon I_N)^3 + \cdots\big)\big\}. \end{aligned} \tag{21}$$

Observe that, similar to the low-SNR case (which corresponds to $M_\epsilon = 0$), the expansion in (21) is dominated by the low-order terms. In fact, the first-order optimal input covariance for low SNRs is also first-order optimal at any SNR and is independent of M_ϵ. However, M_ϵ plays a key role in determining the range of input powers for which the optimal signalling strategy is dominated by its second-order optimal signalling strategy. In particular, if M_ϵ is not sufficiently large, the Taylor expansion in (21) may not converge and one can no longer claim that the series is dominated by the first few terms. On the other hand, if M_ϵ is set to be too large, the series in (21) will eventually converge but the rate of convergence will be too slow for the first few terms to capture the dominant components of the series. This observation suggests that one ought to carefully choose the value of M_ϵ around which the series is expanded for each given value of the input power budget, P. In Section VII, we will show that by properly selecting M_ϵ, we obtain second-order optimal input covariance matrices that provide higher achievable rates than other commonly-used signalling strategies.

VI. Second-order Optimal Input Covariance via Convex Optimization

In this section we exploit the Taylor expansion in (21) to optimize the input signal covariance at an arbitrary SNR. We will derive an efficient algorithm for obtaining the second-order optimal input covariance for a given value of M_ϵ; i.e., the matrix that maximizes the sum of the first- and second-order terms in (21). This matrix achieves a certain ergodic rate and hence one can think of the ergodic rate as being parameterized by M_ϵ. The largest such ergodic rate can be determined by using a one-dimensional stochastic optimization over M_ϵ. In contrast, most related methods (e.g., [11]) require multi-dimensional stochastic optimization over the eigenvalues of the input covariance matrix.

A. Convex Formulation of the Second-order Optimal Design Problem

Our goal now is to show that for any M_ϵ, the input covariance matrix that maximizes the second-order approximation of the ergodic achievable rate can be obtained via the solution of a convex optimization problem. We will then employ some duality arguments to develop an efficient method for computing this matrix.

Using the expansion in (21) and the expectations computed in Section III, our second-order design problem can be formulated as follows,

$$\max_{Q:\, Q\succeq 0, \operatorname{Tr}(Q)=P} \quad F(Q),$$

where

$$\begin{aligned} F(Q) = \tfrac{2M_\epsilon+1}{(M_\epsilon+1)^2}\operatorname{Tr}(XQ) - \tfrac{1}{2(M_\epsilon+1)^2}\big(&\operatorname{Tr}(R^2)\big(\operatorname{Tr}(TQ)\big)^2 \\ &+ 2\operatorname{Tr}(TQ)\operatorname{Tr}(R\bar{H}Q\bar{H}^\dagger) \\ &+ \operatorname{Tr}(XQXQ)\big), \end{aligned} \tag{22}$$

where X is defined in (4) and M_ϵ is appropriately chosen so as to ensure that (19) is satisfied for the SNR of interest; see Section V. By the virtue of being the intersection of an M-dimensional simplex and a positive semidefinite cone [21], the feasible set of (22) is convex. Furthermore, it is shown in Appendix D that the objective in (22) is a convex function of Q.

B. KKT Conditions and Second-order Optimal Input Covariance around $M_\epsilon I$

It is easy to check that for any $P > 0$, the problem in (22) is strictly feasible. Hence, using Slater's condition [21], [22], the convexity of (22) implies that strong duality holds and the KKT conditions are sufficient and necessary to attain both the primal and the dual optimal solutions. The Lagrangian of (22) can written as

$$\begin{aligned} L(Q,Z,\nu) = \beta\big(&\operatorname{Tr}(R^2)\big(\operatorname{Tr}(TQ)\big)^2 + \\ &2\operatorname{Tr}(TQ)\operatorname{Tr}(R\bar{H}Q\bar{H}^\dagger) + \operatorname{Tr}(XQXQ)\big) \\ &- \alpha\operatorname{Tr}(XQ) - \operatorname{Tr}(ZQ) + \nu(\operatorname{Tr}(Q) - P), \end{aligned} \tag{23}$$

where $\alpha = \frac{2M_\epsilon+1}{(M_\epsilon+1)^2}$, $\beta = \frac{1}{2(M_\epsilon+1)^2}$, and $Z \succeq 0$ and ν are the Lagrange dual variables. Using this Lagrangian, we can write the KKT conditions [21], [22] for the optimization problem in (22):

$$\begin{aligned} \nabla_Q L = 2\beta\big(&\operatorname{Tr}(R^2)\operatorname{Tr}(TQ) + \operatorname{Tr}(\bar{H}^\dagger R\bar{H}Q)\big)T \\ &+ 2\beta\operatorname{Tr}(TQ)\bar{H}^\dagger R\bar{H} + 2\beta XQX \\ &- (Z + \alpha X - \nu I) = 0, \end{aligned} \tag{24a}$$

$$\operatorname{Tr}(Q) = P, \tag{24b}$$

$$Q \succeq 0, \tag{24c}$$

$$Z \succeq 0, \tag{24d}$$

$$\operatorname{Tr}(ZQ) = 0. \tag{24e}$$

Assuming that the matrix X in (4) is non-singular,[3] we obtain from (24a) that

$$\begin{aligned} Q = X^{-1}\big(\tfrac{1}{2\beta}(Z + \alpha X - \nu I) - \big(&\operatorname{Tr}(R^2)\operatorname{Tr}(TQ) \\ &+ \operatorname{Tr}(\bar{H}^\dagger R\bar{H}Q)\big)T - \operatorname{Tr}(TQ)\bar{H}^\dagger R\bar{H}\big)X^{-1}. \end{aligned} \tag{25}$$

In order to simplify the notation, we introduce the following transformations:

[3] If X is singular, a similar expression involving the Moore-Penrose pseudo-inverse applies, but for ease of exposition we will restrict our attention to cases in which X is non-singular.

$$\tilde{Z} = X^{-1/2} Z X^{-1/2}, \qquad \tilde{T} = X^{-1/2} T X^{-1/2}, \\ \tilde{S} = X^{-1/2} \bar{H}^{\dagger} R \bar{H} X^{-1/2}. \tag{26}$$

Using this notation, we can write (25) as

$$Q = \tfrac{1}{2\beta} X^{-1/2} \big(\tilde{W} - \gamma_{Z_2} \tilde{T} - \gamma_{Z_1} \tilde{S}\big) X^{-1/2}, \tag{27}$$

where $\tilde{W} = \tilde{Z} + \alpha I - \nu X^{-1}$,

$$\gamma_{Z_1} = 2\beta \operatorname{Tr}(TQ) = \tfrac{1}{\Delta}\Big(\big(1 + \operatorname{Tr}(\tilde{S}\tilde{T})\big) \operatorname{Tr}(\tilde{W}\tilde{T}) \\ - \operatorname{Tr}(\tilde{T}^2) \operatorname{Tr}(\tilde{W}\tilde{S})\Big), \tag{28}$$

$$\begin{aligned} \gamma_{Z_2} &= 2\beta \operatorname{Tr}(R^2) \operatorname{Tr}(TQ) + 2\beta \operatorname{Tr}(\bar{H}^{\dagger} R \bar{H} Q) \\ &= \operatorname{Tr}(R^2) \gamma_{Z_1} \\ &+ \tfrac{1}{\Delta}\Big(\big(1 + \operatorname{Tr}(R^2)\operatorname{Tr}(\tilde{T}^2) + \operatorname{Tr}(\tilde{T}\tilde{S})\big) \operatorname{Tr}(\tilde{W}\tilde{S}) \\ &- \big(\operatorname{Tr}(R^2)\operatorname{Tr}(\tilde{S}\tilde{T}) + \operatorname{Tr}(\tilde{S}^2)\big) \operatorname{Tr}(\tilde{W}\tilde{T})\Big), \end{aligned} \tag{29}$$

where the constant scalar Δ is given by $\Delta = \big(1 + \operatorname{Tr}(\tilde{T}\tilde{S})\big)^2 + \operatorname{Tr}(\tilde{T}^2)\big(\operatorname{Tr}(R^2) - \operatorname{Tr}(\tilde{S}^2)\big)$.

Using these simplifications, we now proceed with the analysis of the KKT conditions in (24). The condition in (24c) and the structure of Q in (27) imply that

$$(\tilde{W} - \gamma_{Z_2}\tilde{T} - \gamma_{Z_1}\tilde{S}) \succeq 0. \tag{30}$$

Furthermore, from (24e) and (27) we have

$$\begin{aligned} \operatorname{Tr}(QZ) &= \operatorname{Tr}\big(\tilde{Z}(\tilde{W} - \gamma_{Z_2}\tilde{T} - \gamma_{Z_1}\tilde{S})\big) \\ &= \operatorname{Tr}\big(\tilde{Z}(\tilde{Z} + \alpha I - \nu X^{-1} - \gamma_{Z_2}\tilde{T} - \gamma_{Z_1}\tilde{S})\big) \\ &= 0. \end{aligned} \tag{31}$$

Let $U_Z \Lambda_Z U_Z^{\dagger}$ be the eigen decomposition of $\tilde{Z}$, and let $U \Lambda U^{\dagger}$ be the eigen decomposition of $(\alpha I - \nu X^{-1} - \gamma_{Z_2}\tilde{T} - \gamma_{Z_1}\tilde{S})$. Furthermore, denote the non-negative entries of Λ by Λ^+ and the negative ones by Λ^-. Now condition (31) can be re-cast as

$$\operatorname{Tr}(U \Lambda U^{\dagger} U_Z \Lambda_Z U_Z^{\dagger}) + \operatorname{Tr}(\Lambda_Z^2) = 0. \tag{32}$$

If we set[4]

$$U_Z = U, \tag{33}$$

condition (32) will yield

$$\operatorname{Tr}\left(\left(\begin{bmatrix} \Lambda^+ & 0 \\ 0 & \Lambda^- \end{bmatrix} + \begin{bmatrix} \Lambda_{Z_1} & 0 \\ 0 & \Lambda_{Z_2} \end{bmatrix}\right) \begin{bmatrix} \Lambda_{Z_1} & 0 \\ 0 & \Lambda_{Z_2} \end{bmatrix}\right) = 0, \tag{34}$$

where Λ_{Z_1}, Λ_{Z_2} are of the same dimension as Λ^+ and Λ^-, respectively.

Now, from condition (30), we know that $\Lambda_{Z_2} + \Lambda^- \succeq 0$. Furthermore, from condition (24d), it is clear that for any positive diagonal entry in $\begin{bmatrix} \Lambda^+ & 0 \\ 0 & \Lambda^- \end{bmatrix} + \begin{bmatrix} \Lambda_{Z_1} & 0 \\ 0 & \Lambda_{Z_2} \end{bmatrix}$, the corresponding entry in Λ_Z must be equal to zero. Hence, the only solution that satisfies (34) is the one in which

$$\Lambda_{Z_1} = 0, \quad \text{and} \quad \Lambda_{Z_2} = -\Lambda^-. \tag{35}$$

Using (33) and (35), if γ_{Z_1}, γ_{Z_2} and ν were known, one would have been able to obtain $\tilde{Z}$ directly, and hence also Z and Q. However, analytical computation of these values appears to be intractable, and hence we propose to perform a bisection search over these scalars. In this search, the values of ν are changed in an outer loop. For every value of ν, an inner loop performs a bisection search on γ_{Z_1} and γ_{Z_2}. For each test pair $(\gamma_{Z_1}, \gamma_{Z_2})$, a test matrix Z is computed as per (33) and (35). The inner loop stops the search whenever a test pair is attained that yields a matrix $\tilde{Z}$ that satisfies (28) and (29). On the other hand, the bisection search of the outer loop stops if a value of ν is attained for which the condition in (24b) is satisfied. Observe that for the bisection search over the pair $(\gamma_{Z_1}, \gamma_{Z_2})$ we have from (28) that $\gamma_{Z_1} \in [0, 2\beta P \operatorname{Tr}(T)]$ and from (29) that $\gamma_{Z_2} \in [\operatorname{Tr}(R^2)\gamma_{Z_1}, \operatorname{Tr}(R^2)\gamma_{Z_1} + 2\beta P \operatorname{Tr}(\bar{H}^{\dagger} R \bar{H})]$. The flow chart in Figure 1 shows the steps of the proposed solution method.

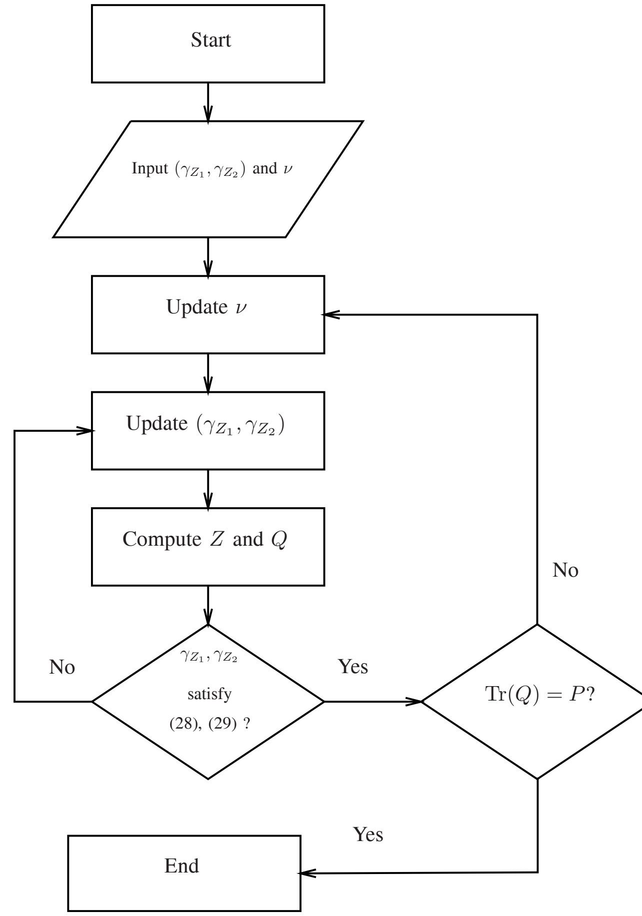

Fig. 1. A flow chart of the solution steps in the proposed algorithm.

We now revisit the issue that was addressed in Section IV regarding the fact that the basis of the second-order optimal input-covariance does not coincide, in general, with that of the first-order optimal one. We recall that for the second-order optimal solution, it is sufficient for U_Z, the eigenvectors of $\tilde{Z}$ to be the same as U, the eigenvectors of $(\alpha I - \nu X^{-1} - \gamma_{Z_2}\tilde{T} - \gamma_{Z_1}\tilde{S})$. Now, substituting this result into (27), we obtain

$$Q = \tfrac{1}{2\beta} X^{-1/2} U \big(\Lambda_Z - \Lambda_a\big) U^{\dagger} X^{-1/2}, \tag{36}$$

where we have used Λ_a to denote the eigenvalues of $(\alpha I - \nu X^{-1} - \gamma_{Z_2}\tilde{T} - \gamma_{Z_1}\tilde{S})$. Now, consider the following lemma.

[4] Observe that setting U_Z to be equal to U is without loss of optimality, because the optimality of the solution will be guaranteed by finding values for Q, Z and ν that satisfy the KKT conditions. Once such values are obtained they serve as a certificate of optimality [21].

Lemma 1: Let G, K, and L be square Hermitian matrices and let their eigen decompositions be given by $U_G\Lambda_G U_G^\dagger$, $U_K\Lambda_K U_K^\dagger$, and $U_L\Lambda_L U_L^\dagger$, respectively. Let K be non-singular, L have distinct eigenvalues, and G be given by $G = HLH^\dagger$. Then, $U_G = U_K$ if and only if $U_K = U_L\Pi$, where Π is a permutation matrix that is determined by the order of the diagonal entries of Λ_G, Λ_K and Λ_L.

Proof: See Appendix E ■

We now apply the result of Lemma 1 to the right hand side of (36). Doing so, it is seen that if the diagonal entries of $(\Lambda_a - \Lambda_Z)$ are distinct, then the eigenvectors of Q will be the same as those of X (up to permutation) if and only if $U = U_X$; that is, if and only if the eigenvectors of X are the same as those of $\left(\nu X^{-1} - \gamma_{Z_2}\tilde{T} - \gamma_{Z_1}\tilde{S}\right)$. However, this condition holds if and only if the matrices X and $\gamma_{Z_2}\tilde{T}+\gamma_{Z_1}\tilde{S}$ commute. From (4), it is seen that one particular situation in which this condition is satisfied is when $\bar{H} = 0$; i.e., when the channel is zero mean.

It is worth pointing out at this point that although the second-order optimal signalling directions do not conform to the first-order ones for any M_ϵ, in practice these directions can be quite close. The reason for that can be seen by applying the result in [24, Example 7.4.13] to the objective in (22). That result implies that the eigen basis of Q that maximizes $\mathrm{Tr}(XQ)$ (in the first order approximation), and at the same time minimizes $\mathrm{Tr}(XQXQ)$ (part of the second order approximation), coincides with the eigen basis of X. Hence, the deviation of the second-order optimal directions from the first-order ones is due to the term $\mathrm{Tr}(R^2)\big(\mathrm{Tr}(TQ)\big)^2 + 2\,\mathrm{Tr}(TQ)\,\mathrm{Tr}(\bar{H}^\dagger R\bar{H})$.

VII. Numerical Examples

In the two numerical examples considered in this section the channel model is as per (1). In the first example, the number of transmit and receive antennas is four, and the matrices T, R and $\bar{H}$ were chosen at random such that $\mathrm{Tr}(T) = \mathrm{Tr}(R) = 1$ and $\|\bar{H}\|_F = 3.5$, and were held fixed while channel realizations were generated according to (1). Figure 2 shows the ergodic rates that can be achieved by signalling along the first-order directions and those that can be achieved by second-order signalling using the algorithm outlined in Section VI-B. Recall that optimized signalling along the first-order directions involves multidimensional stochastic optimization over the eigenvalues of the covariance matrix [15], whereas for our second-order design, stochastic optimization is only required for the optimization of the scalar M_ϵ. As described in Section VI-B, for any M_ϵ the corresponding second-order optimal covariance matrix can be generated efficiently using the technique described in Section VI-B. From Figure 2 it can be seen that the proposed second-order design can provide a higher ergodic rate than optimal signalling along the first-order directions. This confirms the observation made in Section IV that the optimal signalling basis is power dependent and that this basis does not necessarily conform to the first-order one.

Although the example in Figure 2 shows an appreciable difference between the ergodic rates that can be achieved via (optimal) signalling along the first-order basis and those that can be achieved via our proposed second-order design, in many cases this difference is quite small. However, as pointed out earlier, one of the key features of the proposed design is that it enables us to obtain input covariance matrices that achieve ergodic rates that exceed those that can be obtained via significantly more computationally demanding approaches.

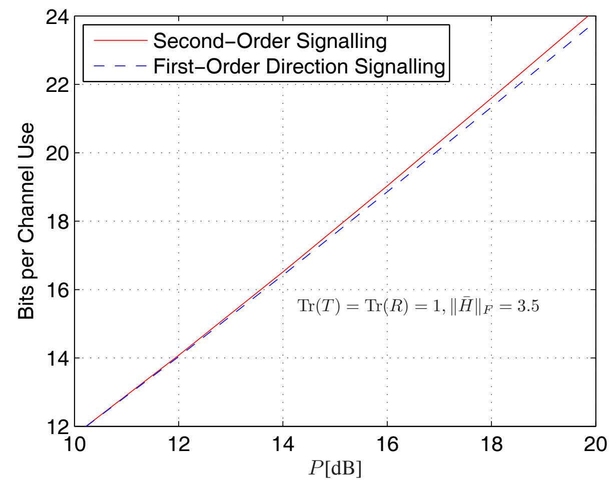

Fig. 2. Comparison between second-order optimal signalling and optimal signalling along the first-order directions.

In order to compare the achievable ergodic rate of the proposed signalling strategy with that of other strategies, we consider a second example in which the number of transmit and receive antennas is six, and the matrices T, R and $\bar{H}$ are randomly generated with $\mathrm{Tr}(T) = \mathrm{Tr}(R) = 1$ but $\|\bar{H}\|_F = 0.77$. The matrices T, R and $\bar{H}$ were held fixed while channel realizations were generated according to (1). For this setup Figure 3 compares the ergodic rates that can be achieved by the following signalling strategies:

1) Uniform power loading [23]: In this technique the transmitter ignores the available CSI and transmits isotropically; i.e., it transmits in all directions with equal power.
2) Mean-optimal signalling [23]: The transmitter ignores the covariance information, treats the mean, $\bar{H}$, as if it were the true channel, and performs signalling along the eigen basis of the channel mean and 'water-fills' over its eigenvalues.
3) Covariance-optimal signalling [11]: The transmitter ignores the mean information and transmits along the eigenvectors of the transmit covariance matrix, T. The power allocation to those eigenvectors that maximizes the ergodic rate in (2) is determined via stochastic optimization over the eigenvalues of the input covariance matrix.
4) First-order low-SNR signalling [15]: The transmitter performs low-SNR optimal signalling; a technique which amounts to beamforming along the principal eigenvector of X in (4).
5) First-order directions signalling [15]: The transmitter selects the low-SNR signalling directions, and then performs stochastic optimization to find the power loads (eigenvalues) that maximize the ergodic rate; cf. (2).
6) Second-order optimal signalling (proposed herein): The transmitter selects an appropriate value of M_ϵ and solves the convex optimization problem in (22) using the tech-

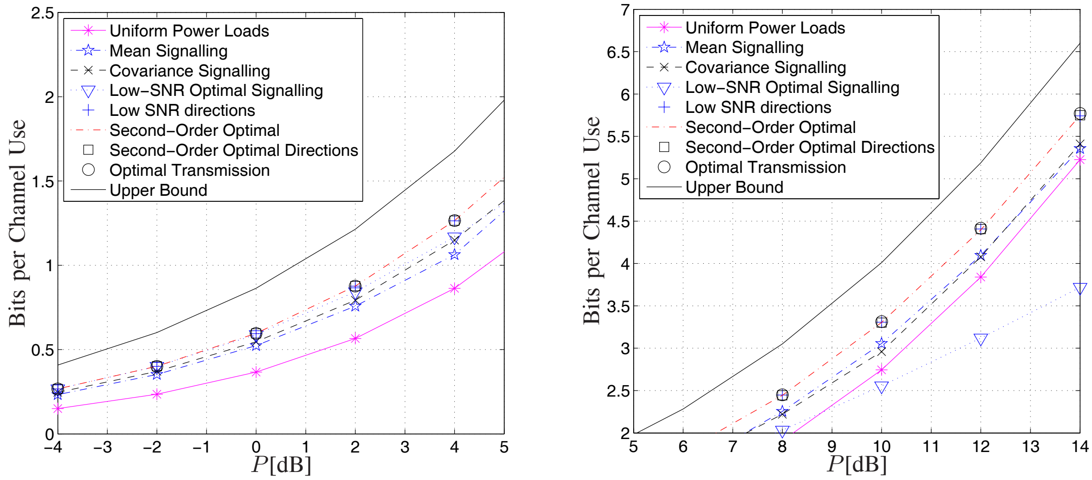

Fig. 3. Comparison between the ergodic rates of Strategies 1–8. An upper bound that corresponds to perfect (instantaneous) CSI at the transmitter is also shown.

nique proposed in Section VI-B. Recall that finding an appropriate value of M_ϵ is much less involved than stochastically optimizing M eigenvalues.

7) Second-order signalling directions: In a fashion similar to Strategy 5, the transmitter signals along the eigen basis of the input covariance matrix obtained in Strategy 6, but the eigenvalues are stochastically optimized so as to maximize the ergodic capacity; cf. (2).
8) Capacity-achieving signalling: In this technique, the optimal-input covariance is generated using the fixed-point iteration algorithm provided in [27].[5] The expectation in each iteration of this algorithm is evaluated numerically using Monte-Carlo integration.

In order to set a benchmark for our comparisons, Figure 3 includes the upper bound on the ergodic capacity that corresponds to perfect (instantaneous) transmitter CSI is available.

In Figure 3, it can be seen that for the instance considered in this example, the uniform power loading (Strategy 1) performs rather poorly in comparison with the performance of mean signalling (Strategy 2) and covariance signalling (Strategy 3). It can also be seen from this figure that although low-SNR optimal beamforming (Strategy 4) incurs a considerable rate loss at moderate-to-high input powers, signalling along the low-SNR optimal directions (Strategy 5) provides higher ergodic rates than Strategies 1–4. In addition, it can be seen that the performance of the proposed second-order design (Strategy 6) coincides with signalling along the second-order directions (Strategy 7). Strategies 6 and 7 perform slightly better than Strategy 5, but this is not clear from the scale in which the figure is plotted. Using the capacity-achieving signalling technique in Strategy 8, it can be seen from this figure that in essence, Strategies 5–7 yield covariance matrices that perform very closely to the ergodic capacity of the MIMO system. However, we emphasize that, in comparison with Strategies 5, 7 and 8, Strategy 6 appears to be a more computationally efficient method for obtaining input covariance matrices that perform close to the ergodic capacity.

[5]We would like to thank an anonymous reviewer for bringing this paper to our attention.

VIII. Conclusions

We considered rate-efficient signalling over MIMO systems in which both the channel mean and the Kronecker-structured covariance are known at the transmitter. By obtaining an explicit expression for the second-order approximation of the ergodic capacity, we were able to show that, in contrast to the case of zero-mean channels, the optimal signalling directions for a channel with a non-zero mean are power dependent. After having exposed the inherent convexity of the second-order approximation, we then used Lagrange duality theory to devise an efficient technique for obtaining the second-order optimal input covariance matrix. By adjusting the centre point in the Taylor expansion that underlies the second-order approximation, this technique was used to obtain second-order optimized input covariance matrices at any given SNR. Finally, we showed, numerically, that the second-order optimized input covariance is not only easier to design, but that it also provides higher ergodic rates than other commonly-used signalling strategies.

Appendix A
Proof of (9) and (10)

In order to prove (9), we denote the eigen decomposition of $T^{1/2}QT^{1/2}$ by $U\Lambda U^\dagger$, and that of R by $V\Sigma V^\dagger$. We also note that for any (white) complex Gaussian matrix with zero mean and i.i.d. entries, and for any deterministic matrices U and V, we have

$$\mathbf{H}_w \overset{d}{=} V^\dagger \mathbf{H}_w U, \tag{37}$$

where $\overset{d}{=}$ denotes equality in distribution. Using these observations, we have

$$\begin{aligned}
&\mathrm{E}_{\mathbf{H}_w}\{\mathrm{Tr}(\mathbf{AB})\}\\
&= \mathrm{E}_{\mathbf{H}_w}\big\{\mathrm{Tr}\big(R^{1/2}\mathbf{H}_w T^{1/2}QT^{1/2}\mathbf{H}_w^\dagger R\mathbf{H}_w T^{1/2}Q\bar{H}^\dagger\big)\big\}\\
&= \mathrm{E}_{\mathbf{H}_w}\big\{\mathrm{Tr}\big(V\Sigma^{1/2}\mathbf{H}_w\Lambda\mathbf{H}_w^\dagger\Sigma\mathbf{H}_w T^{-1/2}\bar{H}^\dagger\big)\big\}\\
&= \mathrm{Tr}\big(V\Sigma^{1/2}\,\mathrm{E}_{\mathbf{H}_w}\{\mathbf{H}_w\Lambda\mathbf{H}_w^\dagger\Sigma\mathbf{H}_w\}T^{-1/2}\bar{H}^\dagger\big).
\end{aligned} \tag{38}$$

In order to compute the expectation in (38), we use $[\mathbf{H}_w]_i$ to denote the i-th column of $\mathbf{H}_w$ and λ_i to denote the i-th diagonal entry of Λ. We now write

$$\begin{aligned}
&\mathrm{E}_{\mathbf{H}_w}\{\mathbf{H}_w\Lambda\mathbf{H}_w^\dagger\Sigma\mathbf{H}_w\}\\
&=\sum_{i=1}^{N}\lambda_i\,\mathrm{E}_{\mathbf{H}_w}\Big\{[\mathbf{H}_w]_i[\mathbf{H}_w]_i^\dagger\Sigma\Big[[\mathbf{H}_w]_1,\cdots,\\
&\qquad [\mathbf{H}_w]_i\ ,\cdots,[\mathbf{H}_w]_N\Big]\Big\}\\
&=\sum_{i=1}^{N}\lambda_i\,\mathrm{E}_{\mathbf{H}_w}\Big\{[\mathbf{H}_w]_i\Big[[\mathbf{H}_w]_i^\dagger\Sigma[\mathbf{H}_w]_1,\ \cdots,\\
&\qquad [\mathbf{H}_w]_i^\dagger\Sigma[\mathbf{H}_w]_i,\ \cdots,[\mathbf{H}_w]_i^\dagger\Sigma[\mathbf{H}_w]_N\Big]\Big\}.
\end{aligned}$$

Observe that, for $i\neq j$

$$\begin{aligned}
&\mathrm{E}_{\mathbf{H}_w}\{[\mathbf{H}_w]_i[\mathbf{H}_w]_i^\dagger\Sigma[\mathbf{H}_w]_j\}\\
&=\mathrm{E}_{[\mathbf{H}_w]_i}\Big\{\mathrm{E}_{[\mathbf{H}_w]_j}\{[\mathbf{H}_w]_i[\mathbf{H}_w]_i^\dagger\Sigma[\mathbf{H}_w]_j\ \Big|\ [\mathbf{H}_w]_i\}\Big\}=0.
\end{aligned}\tag{39}$$

Furthermore,

$$\begin{aligned}
&\mathrm{E}_{\mathbf{H}_w}\{[\mathbf{H}_w]_i[\mathbf{H}_w]_i^\dagger\Sigma[\mathbf{H}_w]_i\} &(40)\\
&=\mathrm{E}_{[\mathbf{H}_w]_i}\left\{\sum_{r=1}^{N}\sigma_r|[\mathbf{H}_w]_{i,r}|^2\begin{bmatrix}[\mathbf{H}_w]_{i,1}\\ \vdots\\ [\mathbf{H}_w]_{i,N}\end{bmatrix}\right\}\\
&=\sum_{r=1}^{N}\sigma_r\,\mathrm{E}_{[\mathbf{H}_w]_i}\Big\{\Big[|[\mathbf{H}_w]_{i,r}|^2[\mathbf{H}_w]_{i,1},\ \cdots,\\
&|[\mathbf{H}_w]_{i,r}|^2[\mathbf{H}_w]_{i,r},\cdots,|[\mathbf{H}_w]_{i,r}|^2[\mathbf{H}_w]_{i,N}\Big]^T\Big\}, &(41)
\end{aligned}$$

where in (41), we have used $[\mathbf{H}_w]_{i,r}$ to denote the r-th entry of $[\mathbf{H}_w]_i$. For $r\neq p$,

$$\begin{aligned}
&\mathrm{E}_{[\mathbf{H}_w]_i}\{|[\mathbf{H}_w]_{i,r}|^2[\mathbf{H}_w]_{i,p}\}=\\
&\mathrm{E}_{[\mathbf{H}_w]_{i,r}}\Big\{\mathrm{E}_{[\mathbf{H}_w]_{i,p}}\{|[\mathbf{H}_w]_{i,r}|^2[\mathbf{H}_w]_{i,p}\ \Big|\ [\mathbf{H}_w]_{i,r}\}\Big\}=0.
\end{aligned}\tag{42}$$

Also, using the fact that the real and imaginary components of $[\mathbf{H}_w]_{i,r}$ are Gaussian and independent with zero mean, it is easy to show that

$$\mathrm{E}_{[\mathbf{H}_w]_i}\{|[\mathbf{H}_w]_{i,r}|^2[\mathbf{H}_w]_{i,r}\}=0.\tag{43}$$

This completes that proof of the first statement of (9); that is, the proof that $\mathrm{E}_{\mathbf{H}_w}\{\mathrm{Tr}(\mathbf{AB})\}=0$. By interchanging the roles of $\mathbf{H}_w$ and $\mathbf{H}_w^\dagger$, it immediately follows that $\mathrm{E}_{\mathbf{H}_w}\{\mathrm{Tr}(\mathbf{AB}^\dagger)\}=0$.

We now prove the statements of (10). Let $U_1\Sigma_1V_1^\dagger$ be the singular value decomposition of $T^{1/2}Q\bar{H}^\dagger R^{1/2}$. Using (37), we have

$$\begin{aligned}
&\mathrm{E}_{\mathbf{H}_w}\{\mathrm{Tr}(\mathbf{BB})\}\\
&=\mathrm{E}_{\mathbf{H}_w}\{\mathrm{Tr}(R^{1/2}\mathbf{H}_wU_1\Sigma_1V_1^\dagger\mathbf{H}_wT^{1/2}Q\bar{H}^\dagger)\}\\
&=\mathrm{E}_{\mathbf{H}_w}\{\mathrm{Tr}(R^{1/2}V_1V_1^\dagger\mathbf{H}_wU_1\Sigma_1V_1^\dagger\mathbf{H}_wU_1U_1^\dagger T^{1/2}Q\bar{H}^\dagger)\}\\
&=\mathrm{E}_{\mathbf{H}_w}\{\mathrm{Tr}(R^{1/2}V_1\mathbf{H}_w\Sigma_1\mathbf{H}_wU_1^\dagger T^{1/2}Q\bar{H}^\dagger)\} &(44)\\
&=\mathrm{Tr}(R^{1/2}V_1\,\mathrm{E}_{\mathbf{H}_w}\{\mathbf{H}_w\Sigma_1\mathbf{H}_w\}U_1^\dagger T^{1/2}Q\bar{H}^\dagger). &(45)
\end{aligned}$$

In order to compute the expectation in (45), we write

$$\mathrm{E}_{\mathbf{H}_w}\{\mathbf{H}_w\Sigma_1\mathbf{H}_w\}=\sum_{i=1}^{N}\sigma_i\,\mathrm{E}_{\mathbf{H}_w}\{[\mathbf{H}_w]_i[\mathbf{H}_w^\dagger]_i^\dagger\}.\tag{46}$$

Now, the (p,q)-th entry of $\mathrm{E}_{\mathbf{H}_w}\{[\mathbf{H}_w]_i[\mathbf{H}_w^\dagger]_i^\dagger\}$ can be written as

$$\big[\mathrm{E}_{\mathbf{H}_w}\{[\mathbf{H}_w]_i[\mathbf{H}_w^\dagger]_i^\dagger\}\big]_{pq}=\mathrm{E}_{\mathbf{H}_w}\{[\mathbf{H}_w]_{i,p}[\mathbf{H}_w]_{q,i}\}=0.\tag{47}$$

In (47), we use the fact that for $p\neq q\neq i$, $p=q\neq i$ $p\neq q=i$ and $p=i\neq q$, $[\mathbf{H}_w]_{i,p}$ and $[\mathbf{H}_w]_{q,i}$ are independent and zero mean. For $p=q=i$, we use the fact that real and imaginary parts of $[\mathbf{H}_w]_{p,p}$ are zero mean i.i.d. random variables. This completes the proof of (10).

Appendix B
Proof of (11) and (12)

In order to prove (11), we denote the eigen decomposition of $T^{1/2}QT^{1/2}$ by $U\Lambda U^\dagger$ and write

$$\begin{aligned}
&\mathrm{E}_{\mathbf{H}_w}\{\mathrm{Tr}(\mathbf{A}C)\}\\
&=\mathrm{E}_{\mathbf{H}_w}\{\mathrm{Tr}(R^{1/2}\mathbf{H}_wT^{1/2}QT^{1/2}\mathbf{H}_w^\dagger R^{1/2}\bar{H}Q\bar{H}^\dagger)\}\\
&=\mathrm{Tr}(R^{1/2}\,\mathrm{E}_{\mathbf{H}_w}\{\mathbf{H}_w\Lambda\mathbf{H}_w^\dagger\}R^{1/2}\bar{H}Q\bar{H}^\dagger) &(48)\\
&=\mathrm{Tr}\Big(R^{1/2}\,\mathrm{E}_{\mathbf{H}_w}\Big\{\sum_{i=1}^{N}\lambda_i[\mathbf{H}_w]_i[\mathbf{H}_w]_i^\dagger\Big\}R^{1/2}\bar{H}Q\bar{H}^\dagger\Big) &(49)\\
&=\mathrm{Tr}(TQ)\,\mathrm{Tr}(R\bar{H}Q\bar{H}^\dagger), &(50)
\end{aligned}$$

where in (48) we have used (37), in (49) we have denoted the i-th diagonal entry of Λ by λ_i, and in (50) we have used $\mathrm{E}_{\mathbf{H}_w}\Big\{[\mathbf{H}_w]_i[\mathbf{H}_w]_i^\dagger\Big\}=I_N$. Using a similar technique, we can prove (12). In particular, we denote the eigen decomposition of $T^{1/2}Q\bar{H}^\dagger\bar{H}QT^{1/2}$ by $WSW^\dagger$ and write

$$\begin{aligned}
\mathrm{E}_{\mathbf{H}_w}\{\mathrm{Tr}(\mathbf{B}^\dagger\mathbf{B})\}&=\mathrm{Tr}(S)\,\mathrm{Tr}(R)\\
&=\mathrm{Tr}(T^{1/2}Q\bar{H}^\dagger\bar{H}QT^{1/2})\,\mathrm{Tr}(R).
\end{aligned}\tag{51}$$

Appendix C
Proof of (13)

In order to derive (13), we again denote the eigen decomposition of $T^{1/2}QT^{1/2}$ by $U\Lambda U^\dagger$, and that of R by $V\Sigma V^\dagger$ and use (37) to write

$$\begin{aligned}
&\mathrm{E}_{\mathbf{H}_w}\{\mathrm{Tr}(\mathbf{A}^2)\}\\
&=\mathrm{E}_{\mathbf{H}_w}\{\mathrm{Tr}(V\Sigma^{1/2}\mathbf{H}_w\Lambda\mathbf{H}_w^\dagger\Sigma\mathbf{H}_w\Lambda\mathbf{H}_w^\dagger\Sigma^{1/2}V^\dagger)\} &(52)\\
&=\mathrm{E}_{\mathbf{H}_w}\Big\{\mathrm{Tr}\Big(\Sigma^{1/2}\Big(\sum_{i=1}^{N}\lambda_i[\mathbf{H}_w]_i[\mathbf{H}_w]_i^\dagger\Big)\\
&\qquad\times\Sigma\Big(\sum_{j=1}^{N}\lambda_j[\mathbf{H}_w]_j[\mathbf{H}_w]_j^\dagger\Big)\Sigma^{1/2}\Big)\Big\} &(53)\\
&=\sum_{i,j=1}^{N}\lambda_i\lambda_j\,\mathrm{E}_{\mathbf{H}_w}\{\mathrm{Tr}(\Sigma^{1/2}[\mathbf{H}_w]_i[\mathbf{H}_w]_i^\dagger\Sigma[\mathbf{H}_w]_j[\mathbf{H}_w]_j^\dagger\Sigma^{1/2})\} &(54)\\
&=\sum_{i=1}^{N}\lambda_i^2\,\mathrm{E}_{\mathbf{H}_w}\{\mathrm{Tr}(\Sigma^{1/2}[\mathbf{H}_w]_i[\mathbf{H}_w]_i^\dagger\Sigma[\mathbf{H}_w]_i[\mathbf{H}_w]_i^\dagger\Sigma^{1/2})\}\\
&\quad+\sum_{i,j=1,i\neq j}^{N}\lambda_i\lambda_j\,\mathrm{E}_{\mathbf{H}_w}\{\mathrm{Tr}(\Sigma^{1/2}[\mathbf{H}_w]_i[\mathbf{H}_w]_i^\dagger\Sigma[\mathbf{H}_w]_j[\mathbf{H}_w]_j^\dagger\Sigma^{1/2})\} &(55)\\
&=\sum_{i=1}^{N}\lambda_i^2\,\mathrm{E}_{\mathbf{H}_w}\{([\mathbf{H}_w]_i^\dagger\Sigma[\mathbf{H}_w]_i)^2\}+\sum_{i,j=1,i\neq j}^{N}\lambda_i\lambda_j\,\mathrm{Tr}(\Sigma^2), &(56)
\end{aligned}$$

where we have used the fact that for $i \neq j$, $[\mathbf{H}_w]_i$ and $[\mathbf{H}_w]_j$ are independent and identically distributed and $\mathrm{E}_{\mathbf{H}_w}\Big\{[\mathbf{H}_w]_i[\mathbf{H}_w]_i^\dagger\Big\} = I_N$. Now,

$$\begin{aligned} &E_{\mathbf{H}_w}\{([\mathbf{H}_w]_i^\dagger\Sigma[\mathbf{H}_w]_i)^2\} \\ &= E_{\mathbf{H}_w}\Big\{\Big(\sum_{r=1}^N \sigma_r\big|[\mathbf{H}_w]_{i,r}\big|^2\Big)^2\Big\} \qquad (57)\\ &= \sum_{r=1}^N \sigma_r^2\,\mathrm{E}_{\mathbf{H}_w}\big\{\big|[\mathbf{H}_w]_{i,r}\big|^4\big\} \\ &\quad + \sum_{\substack{r,s=1\\ r\neq s}}^N \sigma_r\sigma_s\,\mathrm{E}_{\mathbf{H}_w}\big\{\big|[\mathbf{H}_w]_{i,r}\big|^2\big|[\mathbf{H}_w]_{i,s}\big|^2\big\} \qquad (58)\\ &= 2\sum_{r=1}^N\sigma_r^2 + \sum_{\substack{r,s=1\\ r\neq s}}^N \sigma_r\sigma_s = \mathrm{Tr}(R^2) + \big(\mathrm{Tr}(R)\big)^2, \qquad (59)\end{aligned}$$

where we have denoted the r-th diagonal entry of Σ by σ_r. In the first equality of (59), we have used the fact that $\big|[\mathbf{H}_w]_{i,r}\big|^2$ is a Chi-square distributed random variable with two degrees of freedom, $\chi_2^2(\sigma)$, with $\sigma = \frac{1}{2}$. Substituting from (59) into (56), we obtain (13).

Appendix D
Convexity of $-F$ in (22)

For notational convenience we will use s to denote $\mathrm{vec}(\bar{H}^\dagger R\bar{H})$, and t to denote $\mathrm{vec}(T)$. We will also use the definition of X in (4), and the fact that for any two Hermitian symmetric positive semidefinite matrices, A and B,

$$\mathrm{Tr}(AB) = \mathrm{Tr}(A^\dagger B) = \big(\mathrm{vec}(A)\big)^\dagger\mathrm{vec}(B) = k, \qquad (60)$$

where k is a non-negative real scalar. Using this observation the function F can be written as

$$F = \big(\mathrm{vec}(X)\big)^\dagger q - \beta q^\dagger\big(\mathrm{Tr}(R^2)tt^\dagger + 2ts^\dagger + X^*\otimes X\big)q, \quad (61)$$

where $q = \mathrm{vec}\,(Q)$, and $(\cdot)^*$ denotes complex conjugation. In deriving (61) we have used the fact [28] that for any three matrices A, B, and C, $\mathrm{vec}(ABC) = (C^T\otimes A)\,\mathrm{vec}(B)$. Observe that $2q^\dagger ts^\dagger q = q^\dagger(ts^\dagger + st^\dagger)q$. Hence, by completing squares, one can write (61) as

$$\begin{aligned} F = \alpha\big(\mathrm{vec}(X)\big)^\dagger q - \beta q^\dagger\Big(\mathrm{Tr}(R^2)\big(t + \tfrac{1}{2\,\mathrm{Tr}(R^2)}s\big) \\ \times\big(t + \tfrac{1}{2\,\mathrm{Tr}(R^2)}s\big)^\dagger - \frac{1}{4\,\mathrm{Tr}(R^2)}ss^\dagger + X^*\otimes X\Big)q. \quad (62)\end{aligned}$$

In order to show that the optimization problem in (22) is convex, we will show that the Hessian of the function $-F$ is positive semidefinite. Since F is quadratic, we have

$$\begin{aligned} \nabla_q^2(-F) &= 2\beta\Big(\mathrm{Tr}(R^2)\big(t + \tfrac{1}{2\,\mathrm{Tr}(R^2)}s\big)\big(t + \tfrac{1}{2\,\mathrm{Tr}(R^2)}s\big)^\dagger \\ &\quad - \tfrac{1}{4\,\mathrm{Tr}(R^2)}ss^\dagger + X^*\otimes X\Big) \\ &= 2\beta\Big(\mathrm{Tr}(R^2)\big(t + \tfrac{1}{2\,\mathrm{Tr}(R^2)}s\big)\big(t + \tfrac{1}{2\,\mathrm{Tr}(R^2)}s\big)^\dagger \\ &\quad - \tfrac{1}{4\,\mathrm{Tr}(R^2)}(\bar{H}^T\otimes\bar{H}^\dagger)\,\mathrm{vec}(R)\big(\mathrm{vec}(R)\big)^\dagger(\bar{H}^*\otimes\bar{H}) \\ &\quad + \big(\mathrm{Tr}(R)\big)^2(T^*\otimes T) + \mathrm{Tr}(R)(T^*\otimes\bar{H}^\dagger\bar{H} \\ &\quad + \bar{H}^T\bar{H}^*\otimes T) + \bar{H}^T\bar{H}^*\otimes\bar{H}^\dagger\bar{H}\Big) \qquad (63)\\ &= 2\beta\Big(\mathrm{Tr}(R^2)\big(t + \tfrac{1}{2\,\mathrm{Tr}(R^2)}s\big)\big(t + \tfrac{1}{2\,\mathrm{Tr}(R^2)}s\big)^\dagger \\ &\quad + \big(\mathrm{Tr}(R)\big)^2(T^*\otimes T) + \mathrm{Tr}(R)(T^*\otimes\bar{H}^\dagger\bar{H} \\ &\quad + \bar{H}^T\bar{H}^*\otimes T) + (\bar{H}^T\otimes\bar{H}^\dagger)\times \\ &\quad \big((I_{N^2} - \tfrac{1}{4\,\mathrm{Tr}(R^2)}\,\mathrm{vec}(R)\big(\mathrm{vec}(R)\big)^\dagger\big)(\bar{H}^*\otimes\bar{H})\Big), \qquad (64)\end{aligned}$$

where in (63) we have used the fact that $s = \mathrm{vec}(\bar{H}^\dagger R\bar{H}) = (\bar{H}^T\otimes\bar{H}^\dagger)\,\mathrm{vec}(R)$, and in (64), we have used the mixed product rule [28]. In order to show that $\nabla_q^2(-F)$ is indeed positive semidefinite, it is sufficient to show that the minimum eigenvalue of $\big((I_{N^2} - \frac{1}{4\,\mathrm{Tr}(R^2)}\,\mathrm{vec}(R)\big(\mathrm{vec}(R)\big)^\dagger\big)$ in (64) is greater than or equal to zero. Since $\mathrm{vec}(R)\big(\mathrm{vec}(R)\big)^\dagger$ is a dyadic matrix, we have that

$$\begin{aligned}\lambda_{\min}\big((I_{N^2} - \tfrac{1}{4\,\mathrm{Tr}(R^2)}\,\mathrm{vec}(R)\big(\mathrm{vec}(R)\big)^\dagger\big) = \\ 1 - \tfrac{1}{4\,\mathrm{Tr}(R^2)}\big(\mathrm{vec}(R)\big)^\dagger\,\mathrm{vec}(R) = 0.75 > 0, \quad (65)\end{aligned}$$

where for a matrix A, we use $\lambda_{\min}(A)$ to denote its minimum eigenvalue. Hence, $\nabla_q^2(-F)$ is positive definite, which establishes the (strict) convexity of $-F$.

Appendix E
Proof of Lemma 1

The proofs of the direct part and the converse of this lemma are quite similar. We will hence focus on proving the converse. Assume that $U_G = U_K$. Therefore, we can write $U_K\Lambda_G U_K^\dagger = U_K\Lambda_K U_K^\dagger U_L\Lambda_L U_L^\dagger U_K\Lambda_K U_K^\dagger$, and since K is non-singular, we have $U_K^\dagger U_L\Lambda_L U_L^\dagger U_K = \Lambda_G\Lambda_K^{-2}$. Since the right hand side of that expression is diagonal, and L has distinct eigenvalues, it follows from the uniqueness property of the eigen decomposition [24] that $U_K^\dagger U_L = \Pi$, where Π is a permutation matrix that is detemined by the order of diagonal entries of Λ_L and $\Lambda_G\Lambda_K^{-2}$. This completes the proof of the lemma.

References

[1] A. Goldsmith, S. A. Jafar, N. Jindal, and S. Vishwanath, "Capacity limits of MIMO channels," *IEEE J. Select. Areas Commun.*, vol. 21, pp. 684–701, June 2003.

[2] S. Verdú, "Spectral efficiency in the wideband regime," *IEEE Trans. Inform. Theory*, vol. 48, pp. 1319–1343, June 2002.

[3] A. Lozano, A. M. Tulino, and S. Verdú, "Multiple-antenna capacity in the low-power regime," *IEEE Trans. Inform. Theory*, vol. 49, pp. 2527–2544, Oct. 2003.

[4] A. M. Tulino, A. Lozano, and S. Verdú, "Impact of antenna correlation on the capacity multiantenna channels," *IEEE Trans. Inform. Theory*, vol. 51, pp. 2491–2509, July 2005.

[5] G. J. Foschini and M. J. Gans, "On limits of wireless communication in a fading environment when using multiple antennas," *Wireless Personal Commun.*, vol. 6, pp. 311–335, Mar. 1998.

[6] E. Visotsky and U. Madhow, "Space-time transmit precoding with imperfect feedback," *IEEE Trans. Inform. Theory*, vol. 47, pp. 2632–2639, Sept. 2001.

[7] A. L. Moustakas and S. H. Simon, "Optimizing multiple-input single-output MISO communication systems with general Gaussian channels: nontrivial covariance and nonzero mean," *IEEE Trans. Inform. Theory*, vol. 49, pp. 2770–2780, Oct. 2003.

[8] E. A. Jorswieck and H. Boche, "Optimal transmission strategies and impact of correlation in multi-antenna systems with different types of channel state information," *IEEE Trans. Signal Processing*, vol. 52, pp. 3440–3453, Dec. 2004.

[9] S. A. Jafar, S. Vishwanath, and A. Goldsmith, "Channel capacity and beamforming for multiple transmit and receive antennas with covariance feedback," in *Proc. IEEE Int. Conf. Commun.*, Helsinki, pp. 2266–2270, June 2001.

[10] S. H. Simon and A. L. Moustakas, "Optimizing MIMO antenna systems with channel covariance feedback," *IEEE J. Select. Areas Commun.*, vol. 21, pp. 406–417, Apr. 2003.

[11] E. A. Jorswieck and H. Boche, "Channel capacity and capacity-range of beamforming in MIMO wireless systems under correlated fading with covariance feedback," *IEEE Trans. Wireless Commun.*, vol. 3, pp. 1543–1553, Sept. 2004.

[12] S. H. Simon and A. L. Moustakas, "Optimality of beamforming in multiple transmitter multiple receiver communication systems with partial channel knowledge," in *Proc. DIMACS Wkshp Signal Processing Wireless Commun.*, Rutgers Univ., Oct. 2002.

[13] I. Bjelaković and H. Boche, "Structure of optimal input covariance matrices for MIMO systems with covariance feedback under general correlated fading," in *Proc. IEEE Int. Symp. Inform. Theory*, Seattle, pp. 1041–1045, July 2006.

[14] M. Vu and A. Paulraj, "On the capacity of MIMO wireless channels with dynamic CSIT," *IEEE J. Select. Areas Commun.*, vol. 25, pp. 1269–1283, Sept. 2007.

[15] R. H. Gohary, W. Mesbah, and T. N. Davidson, "Rate-optimal MIMO transmission with mean and covariance feedback at low SNR," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, Las Vegas, pp. 3101–3104, Mar. 2008.

[16] A. Lozano, A. Tulino, and S. Verdú, "High-SNR power offset in multi-antenna communication," in *Proc. IEEE Int. Symp. Inform. Theory*, Chicago, p. 288, July 2004.

[17] A. Lozano, A. Tulino, and S. Verdú, "Multiantenna capacity: Myths and realities," in *Space-Time Wireless Systems: From Array Processing to MIMO Communications* (H. Boelcskei, D. Gesbert, C. Papadias, and A. J. van der Veen, eds.), ch. 8, Cambridge, UK: Cambridge University Press, 2006.

[18] A. L. Moustakas and S. H. Simon, "On the outage capacity of correlated multiple-path MIMO channels," *IEEE Trans. Inform. Theory*, vol. 53, pp. 3887–3903, Nov. 2007.

[19] M. Vu and A. Paulraj, "Optimal linear precoders for MIMO wireless correlated channels with nonzero mean in space-time coded systems," *IEEE Trans. Signal Processing*, vol. 54, pp. 2318–2332, June 2006.

[20] A. Hjørungnes and D. Gesbert, "Precoding of orthogonal space-time block codes in arbitrarily correlated MIMO channels: Iterative and closed-form solutions," *IEEE Trans. Wireless Commun.*, vol. 6, pp. 1072–1082, Mar. 2007.

[21] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, UK: Cambridge University Press, 2004.

[22] D. P. Bertsekas, A. Nedic, and A. E. Ozdaglar, *Convex Analysis and Optimization*. Nashua, NH: Athena Scientific, 2003.

[23] I. E. Telatar, "Capacity of multiantenna Gaussian channels," *Eur. Trans. Telecom.*, vol. 10, pp. 585–595, Nov. 1999.

[24] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge, UK: Cambridge University Press, 1999.

[25] M. Kießling, "Unifying analysis of ergodic MIMO capacity in correlated Rayleigh fading environment," *Europ. Trans. Telecommun.*, vol. 16, pp. 17–35, 2005.

[26] N. Piskunov, *Differential and Integral Calculus*. Moscow: MIR Press, 1969.

[27] L. W. Hanlen and A. Grant, "Optimal transmit covariance for ergodic MIMO channels," 2005. Available at http://arxiv.org/abs/cs/0510060.

[28] A. Graham, *Kronecker Products and Matrix Calculus: with Applications*. New York: Elis Horwood Ltd., 1981.

Ramy H. Gohary (M'00) received the B.Sc. (Hons.) degree in Electronics and Communications Engineering from Assiut University, Egypt in May 1996. In January 2000, he received the M.Sc. degree in Communications Engineering from Cairo University, Egypt. He received the Ph.D. degree in Electrical Engineering from McMaster University, Ontario, Canada in April 2006. In 2007 he received the NSERC visiting fellowship award and he is currently a visiting fellow with the Terrestrial Wireless Systems Branch, Communications Research Centre, Industry Canada. His current interests include analysis and design of MIMO wireless communication systems, applications of optimization techniques in signal processing, information theoretic aspects of multiuser communication systems, and applications of iterative detection and decoding techniques in multiple antenna and multiuser systems.

Tim Davidson (M'96) received the B.Eng. (Hons. I) degree in Electronic Engineering from the University of Western Australia (UWA), Perth, in 1991 and the D.Phil. degree in Engineering Science from the University of Oxford, U.K., in 1995.

He is currently an Associate Professor in the Department of Electrical and Computer Engineering at McMaster University, Hamilton, Ontario, Canada, where he holds the (Tier II) Canada Research Chair in Communication Systems, and is currently serving as Acting Director of the School of Computational Engineering and Science. His research interests lie in the general areas of communications, signal processing and control. He has held research positions at the Communications Research Laboratory at McMaster University, the Adaptive Signal Processing Laboratory at UWA, and the Australian Telecommunications Research Institute at Curtin University of Technology, Perth, Western Australia.

Dr. Davidson was awarded the 1991 J. A. Wood Memorial Prize (for "the most outstanding [UWA] graduand" in the pure and applied sciences) and the 1991 Rhodes Scholarship for Western Australia. He is currently serving as an Associate Editor of the IEEE TRANSACTIONS ON SIGNAL PROCESSING and as an Editor of the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS. He has also served as an Associate Editor of the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS II, and as a Guest Co-editor of issues of the IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS and the IEEE JOURNAL ON SELECTED TOPICS IN SIGNAL PROCESSING.