Transmitter Optimization for Correlated MISO Fading Channels with Generic Mean and Covariance Feedback

Jiangyuan Li and Q. T. Zhang, Senior Member, IEEE

Abstract—Transmitter optimization for correlated multiple-input single-output (MISO) channels with average channel-information feedback is of theoretical and practical importance. It was tackled in the past by mainly focusing on cases with either mean or covariance feedback alone. The difficulty with generic MISO optimization arises from the fact that the treatment of optimal beam directions and power allocation are no longer separable, thereby making the problem nearly impossible to be tackled on the basis of its current formulation. We therefore take a different philosophy by representing the derivative of the system performance as a single integral and linking it to the channel characteristic function. This new formulation allows us to obtain the sufficient and necessary conditions, simply in the form of two vector equations which can be efficiently solved by a Newton-Raphson type algorithm developed in this paper. Numerical results are presented for illustration.

Index Terms—Channel capacity, transmitter optimization, general mean/covariance feedback, MISO system optimization, multiple-antenna systems.

I. INTRODUCTION

Wireless transmission employing multiple antennas has captured the attention of many researchers due to its huge potential channel capacity [1]. One challenging issue for the system designer is to optimize the transmitted signal, so that the system data rate can be maximized. Such an optimization problem is well addressed for the case with perfect channel state information (CSI) at the transmitter [2],[3]. In many practical applications, however, the transmitter knows only some average information about the channel such as its distribution or partial statistics. For a Rayleigh or Rician fading channel, its distribution is completely specified by its mean value and covariance matrix. Exploiting the knowledge of channel mean and covariance matrix to enhance the system performance is therefore of practical and theoretical importance [4].

Transmitter optimization along this line starts from multiple-input single-output (MISO) antenna systems, aiming to find the optimal input covariance matrix for a given constraint on transmit power. For tractability, research is usually focused on two extreme cases. In the first case with mean feedback, it is assumed that the channel covariance matrix, up to a factor, is equal to the identity matrix. In the second case with covariance feedback, it is assumed that the channel mean is zero. With these channel constraints, the MISO optimization is tackled, for the first time, by Visotsky and Madhow [5], revealing that the optimal transmit strategy essentially consists of a set of beamformers defined either by the eigenvectors of the channel covariance matrix or by the channel mean vector. Subsequently, Jafar and Goldsmith obtain, under the same channel conditions, the necessary and sufficient conditions for the optimal beamforming to achieve the Shannon capacity [6]. Power allocation on different transmit beams, on the other hand, is not so easy to handle. Jorswieck and Boche therefore develop a numerical optimization algorithm for determining the eigenvalues of the optimal input covariance matrix for the case with covariance feedback [7]. Such an algorithm is further refined in [8], [9]. Jafar and Goldsmith extend the methodology of [5], again with the same channel constraints, to tackle the issue of transmitter optimization for multiple-input and multiple-output (MIMO) systems [10], [11], [12].

The transmitter optimization for MISO systems with hybrid channel mean and covariance feedback remains unsolved in the literature [13], [14]. This general case cannot be treated in the same way as the two extreme cases mentioned above. The difficulty lies in that the optimization for the eigenvalues and for the eigenvectors of the input covariance matrix is no longer separable. Their optimization must be done simultaneously, thereby lending a direct formulation and analysis nearly impossible. We bypass this difficulty by devising a different formulation for the channel capacity, such that the multi-fold expectation can be easily converted to a single integral, making it possible to obtain simple expressions for the sufficient and necessary conditions for the global solution.

The remainder of the paper is organized as follows. The problem is formulated in Section II, followed by the derivation of the sufficient and necessary conditions in Section III and an efficient algorithm for optimal solution in IV. Section V presents numerical results and Section VI concludes the paper with a summary.

II. SYSTEM MODEL AND FORMULATION

We use superscripts *, T, and † to signify conjugation, transposition, and conjugate transposition, respectively. Use $\mathbb{E}_a[\cdot]$ to denote the expectation taking over random vector $a$, and use $\text{Tr}\{A\}$ to denote the trace of matrix $A$. Notation $z \sim \mathcal{CN}(\mu, R)$ means that vector $z$ follows the complex Gaussian distribution with mean $\mu$ and covariance matrix $R$. Consider a MISO system with $M$ transmit antennas and one receive antenna, so that the received baseband signal can be written as

$$y = h^\dagger x + n$$  (1)
where \( x \sim \mathcal{CN}(0, R_x) \) is an \( M \times 1 \) input complex vector, and \( n \sim \mathcal{CN}(0, \sigma_n^2) \) is circularly symmetric complex Gaussian noise. The signal \( x \) is transmitted over a Rayleigh fading channel defined by the \( M \times 1 \) vector \( h \sim \mathcal{CN}(\mu_h, R_h) \). We can represent \( R_x \) in terms of the average signal energy \( E_s \) and normalized channel covariance matrix \( Q \), so that \( R_x = E_s Q \) and \( \text{Tr}(Q) = 1 \). From (1) and using the definition, it is easy to obtain the ergodic channel capacity:

\[
C = E_h \left\{ \log (1 + \rho h^H Q_h) \right\}
\]  

(2)

where \( \rho = E_s / \sigma_n^2 \).

The transmitter optimization problem is to find the optimal positive semi-definite matrix \( Q \), for which \( C \) is maximized. So far, only three special cases can be handled in the literature; they correspond to \( h \sim \mathcal{CN}(0, \alpha I_M) \), \( h \sim \mathcal{CN}(\mu_h, \alpha I_M) \) and \( h \sim \mathcal{CN}(0, R_h) \), respectively. Telatar [1] tackles the first case showing that the optimal strategy is to evenly distribute the total power over any set of orthogonal independent transmit directions. For the second case, it is shown that the optimal transmission direction is the same as the channel mean vector \( \mu_h \) [5], whereas the optimal eigenvalues are determined by numerical search. For the third case, the optimal covariance matrix has the same eigenvector structure as \( R_h \) [5] while its eigenvalues are determined, again, by resorting to numerical optimization [7]. These results are consistent with our intuition, since the statistical feedback information in the three cases defines, respectively, no direction, one direction and \( M \) directions.

The situation with the unsolved general case with \( h \sim \mathcal{CN}(\mu_h, R_h) \) is quite different. We are now given \((M + 1)\) \textit{a priori} vectors defined by mean vector \( \mu \) and the eigenvectors of \( R_h \) and yet, we need to determine only up to \( M \) optimal beam directions. Therefore, the optimal directions must be a function of \( \mu, R_h \), and SNR in a complex unknown manner. Determining their relationship and the corresponding power allocation is the focus of this paper.

By denoting \( \gamma = \rho^{-1} \), the optimization problem is equivalent to

\[
\max \quad J_1(Q) = E_h \left\{ \log (\gamma + h^H Q_h) \right\},
\]  

(3)

s.t. \( \text{Tr}(Q) = 1, Q \succeq 0 \).

Note that \( J_1(Q) \) is strictly convex with respect to \( Q \). Thus, the optimization problem (3) has a unique solution \( Q^* \). To obtain the optimal solution, we need to differentiate \( J_1(Q) \), leading to

\[
\Theta = \frac{\partial J_1}{\partial Q} = E_h \left\{ \frac{h h^H}{\gamma + h^H Q_h} \right\}.
\]  

(4)

The presence of \( h \) in the denominator prevents the calculation of the expectation. The central idea to avoid this difficulty is to invoke the identity of \( \frac{1}{a} = \int_0^\infty e^{-ax} dt, a > 0 \) to convert the denominator into an exponent whereby we can relate the derivative to the well-known characteristic function (CHF) of \( h^H Q_h \) [15]. It turns out that

\[
\Theta = \int_0^\infty e^{-z^H} E_h \left\{ \frac{h h^H e^{-z^H}}{\gamma + h^H Q_h} \right\} dz
\]  

(5)

where

\[
\mu = R_h^{-1} \mu_h, \quad K = (R_h^{-1} + z Q)^{-1}.
\]  

(6)

The above expression consists of only a single integral thereby laying a convenient basis for our subsequent optimization.

### III. SUFFICIENT AND NECESSARY CONDITIONS FOR OPTIMAL SOLUTION

To proceed, let \( \Lambda_q = \text{diag}\{\lambda_k^q \geq 0\} \) denote the diagonal matrix formed by the eigenvalues of \( Q \), in descending order, and let \( U_q = [u_1^q, \ldots, u_M^q] \) denote the corresponding unitary eigenvectors such that \( Q = U_q \Lambda_q U_q^H \). The unity transmit-power constraint implies \( \text{Tr}(Q) = \text{Tr}(\Lambda_q) = 1 \). By the same token, we eigen decompose \( R_h \) as \( R_h = U_h \Lambda_h U_h^H \), with the unitary matrix \( U_h = [u_1^h, \ldots, u_M^h] \) and diagonal matrix of eigenvalues \( \Lambda_h = \text{diag}\{\lambda_1^h, \ldots, \lambda_M^h\} \).

Using Lagrange multipliers enables the construction of a single objective function:

\[
L(Q, \zeta_1, \Psi) = J_1(Q) - \zeta_1 (\text{Tr}(Q) - 1) + \text{Tr}(\Psi Q)
\]  

(7)

where \( \zeta_1 \) is the Lagrange multiplier associated with the equality constraint \( \text{Tr}(Q) = 1 \), and \( \Psi \succeq 0 \) is the Lagrange multiplier, in matrix form, associated with the inequality constraint \( Q \succeq 0 \). Denote the eigenvalues of \( \Psi \) and \( \Theta \) by \( \psi_1, \ldots, \psi_M \) and by \( \theta_1, \ldots, \theta_M \), respectively. Denote \( \Lambda_\psi = \text{diag}\{\psi_1, \ldots, \psi_M\} \), and \( \Lambda_\theta = \text{diag}\{\theta_1, \ldots, \theta_M\} \).

We take the derivative of \( L \) with respect to \( Q \) by using (5) and the rule for matrix differentiation [16]. Then set the resulting derivatives to zero and appropriately characterize the constraints, resulting in a set of KKT necessary conditions [17]:

\[
\Theta - \zeta_1 I + \Psi = 0
\]  

(8)

\[
\text{Tr}(Q) - 1 = 0
\]  

(9)

\[
\text{Tr}(\Psi Q) = 0
\]  

(10)

\[
Q \succeq 0
\]  

(11)

\[
\Psi \succeq 0
\]  

(12)

We need to remove the multipliers to make the KKT conditions useful in practice. The three conditions (10), (11) and (12) mean that all eigenvalues of the positive semidefinite product \( \Psi Q \) are zero. Thus, \( \Psi \) and \( Q \) must have the same eigenvectors, and their eigenvalue patterns are complementary.
in the sense that if \( \lambda_i(Q) > 0 \), then \( \lambda_i(\Psi) = 0 \), and vice versa [18]. This enables us to assert that

\[
\Psi = U_q A_\psi U_q^\dagger, \quad A_q A_\psi = 0
\]  
(13)

and hence \( \Psi Q = O \). We next right-multiply both sides of (8) with \( Q \) and employ the result of \( \Psi Q = O \) to obtain \( \Theta Q = \zeta_1 Q \) which, when taking the trace noting that \( \text{Tr}(Q) = 1 \), allows us to write \( \zeta_1 \) explicitly as

\[
\zeta_1 = \text{Tr}(\Theta Q).
\]  
(14)

Inserting \( \zeta_1 \) into \( \Theta Q = \zeta_1 Q \) produces

\[
\text{Tr}(\Theta Q) Q = \Theta Q
\]  
(15)

which is a basic equation for our optimization.

Equation (15) reveals an important relation between Hermitian matrices \( Q \) and \( \Theta \). To be explicit, we take Hermitian transforms of both sides of (15) yielding

\[
Q \Theta = \Theta Q
\]  
(16)

which asserts that \( \Theta \) and \( Q \) commute. According to a lemma in matrix theory [19], the sufficient and necessary condition for two matrices to be commutable is the existence of a unitary matrix that can simultaneously diagonalize both matrices. Since \( U_q \) diagonalizes \( Q \), it does the same for \( \Theta \), namely,

\[
U_q^\dagger \Theta U_q = \Lambda_\theta.
\]  
(17)

We use \( U_q \) to diagonalize (8) and invoke (13) to simplify yielding

\[
A_q A_\theta = \zeta_1 A_q
\]  
(18)

which, when written explicitly, produces

\[
\lambda_j^q \theta_j = \zeta_1 \lambda_j^q, j = 1, \cdots, M.
\]  
(19)

Eq. (19) implies that if \( \lambda_j^q > 0 \), then \( \theta_j = \zeta_1 \). On the other hand, we can directly conclude from (8) that \( \zeta_1 - \theta_j = \psi_j > 0, j = 1, \cdots, M \). Hence, we can assert \( \max_{1 \leq k \leq M} \theta_k = \zeta_1 \) which, when combined with (14), leads to another fundamental equation that

\[
\max_{1 \leq k \leq M} \theta_k = \text{Tr}(\Theta Q).
\]  
(20)

So far, we have simplified the KKT conditions to obtain two fundamental equations, (15) and (20), by using some linear transforms. However, there is no guarantee that the results so obtained are equivalent to the original KKT conditions. In the Appendix, we show that (15) and (20) are indeed necessary and sufficient for \( Q \) to be optimal.

Though simple, (15) is not always convenient to use. To further simplify, we note that the optimal transmitter is likely not to transmit signals along the eigenvectors of the channel covariance matrix corresponding to small eigenvalues in order to avoid cross-beam interference at the receiver. This implies that \( Q \) is not necessarily of full rank. Thus, we may assume \( \text{rank}(Q) = m \), so that

\[
Q = \sum_{k=1}^m \lambda_k^q u_k^q u_k^q\dagger.
\]  
(21)

Inserting (21) into (15) and using the fact that \( \text{Tr}(\Theta Q) = \zeta_1 \) allow us to write

\[
\Theta u_k^q = \zeta_1 u_k^q, \quad i = 1, \cdots, m
\]  
(22)

which implies that \( \zeta_1 \) is the \( m \)-fold eigenvalue of \( \Theta \) with corresponding eigenvectors \( u_k^q, i = 1, \cdots, m \). We may summarize the foregoing analysis as follows.

**Proposition 3.1:** There exist a unique \( m \) and a set of \( \{u_k^q, \lambda_k^q > 0, k = 1, \cdots, m\} \), such that the optimal solution \( Q = \sum_{k=1}^m \lambda_k^q u_k^q u_k^q\dagger \) meets the requirements specified in (20) and (22).

Proposition 3.1 reveals that: 1) All the non-zero eigenvalues of \( Q \) correspond to the same eigenvalue of \( \Theta \); i.e., if \( \lambda_j^q > 0 \) and \( \lambda_j^q > 0 \), then \( \theta_i = \theta_j = \zeta_1 \). 2) All the eigenvalues of \( \Theta \) corresponding to the zero eigenvalues of \( Q \) are less than or equal to \( \zeta_1 \). This proposition also provides a means to identify the optimum number of beams to be used in the transmitter.

**IV. EFFICIENT ALGORITHM FOR OPTIMAL SOLUTION**

**A. Algorithm Description**

To reduce the number of parameters to be optimized, let us define \( v_k = (\lambda_k^q)^{1/2} u_k \) whereby \( Q = \sum_{k=1}^m v_k v_k\dagger \). We need to find a set of \( v \)'s to meet (15) such that \( \text{Tr}(\Theta Q) v_k = \Theta v_k, k = 1, \cdots, m \). An iterative algorithm to achieve this goal is described below.

a) Start from \( m = 1 \).

b) Find \( v_k, k = 1, \cdots, m \) satisfying

\[
\text{Tr}(\Theta Q) v_k = \Theta v_k, k = 1, \cdots, m.
\]  
(23)

c) Update \( Q \) using \( Q = \sum_{k=1}^m v_k v_k\dagger \) and update \( \Theta \) using (5).

d) Check whether (20) holds. If yes, go to step f); otherwise, set \( m = m + 1 \) and go to b).

f) Stop

To initialize, we suggest using the first \( m \) eigenvectors of \((R_k + \mu_k M_k \dagger) \) as the initial values of \( v_1, \cdots, v_m \). We use Gauss-Laguerre numerical integration method to compute the related integrals which is efficient for single variable integrals. The algorithm would be incomplete without showing how to determine \( v_k \) in step b). To proceed, denote \( x = [v_1^T, \cdots, v_m^T]^T \). Rewrite (23) as

\[
F(x, x^c) = [F_1, \cdots, F_m\dagger]^T = 0
\]  
(24)

whose \((l - 1)M + k\)th entry is defined as

\[
F_{(l-1)M+k} = \text{Tr}(\Theta Q) v_k - (\Theta v_i)_k,
\]  
\( i = 1, \cdots, m; k = 1, \cdots, M \).  
(25)

We can solve these equations by developing a Newton-Raphson type iterative procedure as described next.

**B. Multidimensional Complex Iterative Algorithm**

The Newton algorithm was originally developed in the real domain on the basis of real Taylor series. Following the same principle, we can obtain its counterpart for complex variables. For a vector function \( F(x, x^c) \) of complex vector argument \( x = u + jv \), define its derivatives as \( dF(x_0, x_{0c})/dx = A_1, \)
\( dF(x_0, \lambda_0)/dx^* = A_2 \) where \( x - x_0 = \Delta x = \Delta u + j\Delta v \).

We expand \( F \) as a Taylor's series in the neighborhood \( x_0 \) (cf. [20]), denote \( D_1 = \text{Re}\{A_1 + A_2\} \), \( D_2 = -\text{Im}\{A_1 - A_2\} \), \( D_3 = -\text{Re}\{F(x_0, \lambda_0)\} \), \( D_4 = \text{Im}\{A_1 + A_2\} \), \( D_5 = \text{Re}\{A_1 - A_2\} \), and \( D_6 = -\text{Im}\{F(x_0, \lambda_0)\} \), set the linear part of the expansion to zero, and solve the resulting equations for \( \Delta u \) and \( \Delta v \). It ends up with an update equation for the \((k+1)\)th iteration:

\[
x^{k+1} = x^k + \Delta x^k
\]

(26)

where \( \Delta x^k = \Delta u^k + j\Delta v^k \) can be evaluated from the \( k \)th iterative results of

\[
\Delta v = (D_1D_3^{-1}D_2 - D_5)^{-1}(D_4D_3^{-1}D_3 - D_6)
\]

(27)

\[
\Delta u = D_1^{-1}(D_3 - D_2\Delta v).
\]

(28)

For a proof of the convergence of this complex Newton-type algorithm, reader is referred to [21] for details.

To use the above algorithm for our optimization problem, we need to work out the derivatives required in (25). The results are summarized below.

\[
\frac{\partial \text{Tr}(\Theta Q)}{\partial v_{ij}} = \text{Tr} \left\{ \frac{\partial \Theta}{\partial v_{ij}} \sum_{t=1}^{m} v_{t}v_{t}^\dagger + \Theta J_{t}J_{t}^\dagger \right\}
\]

(29)

\[
\frac{\partial \text{Tr}(\Theta Q)}{\partial v_{ij}^*} = \text{Tr} \left\{ \frac{\partial \Theta}{\partial v_{ij}^*} \sum_{t=1}^{m} v_{t}v_{t}^\dagger + \Theta J_{t}J_{t}^\dagger \right\}
\]

(30)

\[
\frac{\partial (\Theta v_{ij})_{k}}{\partial v_{ij}} = \left( \frac{\partial \Theta}{\partial v_{ij}} v_{t} + \Theta J_{t}J_{t}^\dagger \right)_{k}
\]

(31)

\[
\frac{\partial (\Theta v_{ij})_{k}}{\partial v_{ij}^*} = \left( \frac{\partial \Theta}{\partial v_{ij}^*} v_{t} \right)_{k}
\]

(32)

where

\[
\frac{\partial \Theta}{\partial v_{ij}} = -\int_{0}^{\infty} dz \text{det}(KR_{h}^{-1})e^{-z\gamma}R_{n}^{-1}\mu_{i}\mu_{j}\mu_{i}\mu_{j}K_{\mu}
\]

\[
\cdot \text{Tr} \left[ KJ_{i}v_{t}^\dagger + AJ_{t}v_{t}^\dagger \right] A + KJ_{i}v_{t}^\dagger K + KJ_{i}v_{t}^\dagger A + AJ_{t}v_{t}^\dagger K
\]

\[
+ KJ_{i}v_{t}^\dagger A + AV_{i}J_{i}^\dagger K \right) dz,
\]

(33)

\[
\frac{\partial \Theta}{\partial v_{ij}^*} = -\int_{0}^{\infty} dz \text{det}(KR_{h}^{-1})e^{-z\gamma}R_{n}^{-1}\mu_{i}\mu_{j}\mu_{i}\mu_{j}K_{\mu}
\]

\[
\cdot \text{Tr} \left[ KJ_{i}v_{t}^\dagger + AV_{i}J_{i}^\dagger \right] A + KJ_{i}v_{t}^\dagger K
\]

\[
+ KJ_{i}v_{t}^\dagger A + AV_{i}J_{i}^\dagger K \right) dz,
\]

(34)

and \( A = K\mu_{i}\mu_{j}K \). Here, \( J_{ij} \) denotes a matrix with all zero elements except for unity in its \((i,j)\)th entry, whereas \( J_{j} \) signifies an all-zero vector except for unity in its \( j \)th element. With these derivatives, we can differentiate both sides of (25) to obtain

\[
\frac{\partial F_{(l-1)M+k}}{\partial v_{ij}} = \frac{\partial \text{Tr}(\Theta Q)}{\partial v_{ij}} v_{l,k} + \text{Tr}(\Theta Q) J_{l}J_{l}^\dagger J_{j}J_{j}^\dagger
\]

\[
\frac{\partial F_{(l-1)M+k}}{\partial v_{ij}} = \frac{\partial \text{Tr}(\Theta Q)}{\partial v_{ij}^*} v_{l,k}^* - \frac{\partial (\Theta v_{ij})_{k}}{\partial v_{ij}}
\]

(35)

(36)

for the update equation (26).

V. NUMERICAL RESULTS

Let us examine the correctness of the theory and algorithms developed above by considering a MISO system with \( M = 4 \) transmit antennas to operate on a Rician fading channel with mean vector \( \mu_{h} = [0.1, 0.3, 0.4, 0.2 - j0.1, 0.5 + j0.25]^T \) and Hermitian covariance matrix \( R_{h} \). For brevity, we only show the lower triangle elements of \( R_{h} \), given by \([0.615; 0.263 + j0.160, 0.571; 0.358 - j0.321, 0.126 - j0.159, 0.926; 0.054 + j0.158, 0.354 + j0.248, 0.216 - j0.274, 1.013] \).

Our algorithm consists of two iteration loops, one for \( m \) while another for \( k \). For illustration, we set \( SNR = 5 \text{ dB} \). The iteration starts from \( m = 1 \), and continues until the norm of \( F \) is smaller than a preset error, say, \( \varepsilon = 10^{-6} \). The iteration procedure is shown in Table I. For \( m = 1 \), the Newton algorithm leads to \( v_{1} = [0.3441 + j0.0341, 0.4698 - j0.1384, 0.3801 - j0.1305, 0.6921 + j0.0021]^T \) which allows us to determine \( Q \) and \( \Theta \) using \( Q = v_{1}v_{1}^\dagger \) and (5), respectively. It ends up with \( \text{Tr}(\Theta Q) = 0.7417 \) whereas the maximum eigenvalue of \( \Theta \) equal to 1.0725, indicating that the requirement of (20) is not met when \( m = 1 \). The algorithm then proceeds to \( m = m + 1 = 2 \), for which \( \text{Tr}(\Theta Q) = 0.7724 \) and is identical to the maximum eigenvalue of \( \Theta \). The condition (20) is satisfied and the iteration stops.

We repeat the same procedure for different values of \( SNR \). The result is shown in Fig. 1 for the optimal power allocation along different beams as a function of \( SNR \). It is observed that the power tends to more evenly distribute over the two beams as \( SNR \) increases. The optimal \( Q \) obtained can be used to determine the system capacity. The MISO ergodic capacity vs. \( SNR \) is depicted in Fig. 2 where the system capacity of equal power allocation is also included for comparison. Clearly, the optimized system considerably outperforms its counterpart with equal power allocation, in terms of channel capacity.

### Table I

<table>
<thead>
<tr>
<th>SNR (dB)</th>
<th>Q ( \lambda_1 )</th>
<th>Q ( \lambda_2 )</th>
<th>Q ( \lambda_3 )</th>
<th>Q ( \lambda_4 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.0725, 0.7417</td>
<td>0.2545, 0.1212</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0.7724, 0.7724</td>
<td>0.2221, 0.1065</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The optimal Q obtained can be used to determine the system capacity. The MISO ergodic capacity vs. SNR is depicted in Fig. 2 where the system capacity of equal power allocation is also included for comparison. Clearly, the optimized system considerably outperforms its counterpart with equal power allocation, in terms of channel capacity.
VI. CONCLUSIONS

In this paper, we tackled the issue of transmitter optimization problem for MISO systems with hybrid mean feedback and covariance feedback. By representing the objective function as a single integral, we are able to link the derivative of the ergodic channel capacity to the channel characteristic function (CHF) whereby the sufficient and necessary conditions for optimal transmitter can be derived in some cases. It is revealed that there exists an optimal number of transmit beams, and the optimal beam directions and power allocations are completely specified by two vector equations. A Newton-Raphson type algorithm has been developed for the efficient search of the optimal solution. The theory is illustrated and examined by numerical simulations, and its correctness and effectiveness are demonstrated.

APPENDIX

PROOF OF PROPOSITION 3.1

Proof: The nontrivial positive semidefinite $Q^\circ$ satisfies (15) and (20). Denote $\zeta^2 = \text{Tr}(\Theta^0Q^\circ) > 0$. It follows from (16) that $\Theta^0Q^\circ = Q^\circ\Theta^0$ and further

\[ \Theta^0 = \zeta^2 I_M - U^0_q \text{diag}(\Theta^0) U^0_q. \]  

(37)

According to (4) and (5), we obtain

\[ \frac{\partial J_1(Q^\circ)}{\partial Q} = \Theta^0T. \]  

(38)

Note that $Q$ is Hermitian implying that $q_{ji} = q_{ij}^*$ and $q_{ij}$'s are real variables. We denote $q_{ij} = q_{ij}^R + i q_{ij}^I$, $i > j$. $J_1$ is a real function with respect to $y = [q_{11}^R, q_{21}^R, q_{22}^R, \ldots, q_{MM-1,MM}^R]^T$. Represent $J_1(y)$ as a Taylor’s series to obtain

\[ J_1(y) = J_1(y^0) + (y - y^0)^T \frac{\partial J_1(y^0)}{\partial y} + \frac{1}{2} (y - y^0)^T H (\eta(y - y^0)) (y - y^0), 0 < \eta < 1 \]  

(39)

where $H$ is the Hessian matrix.

Note that $J_1$ is a \( \cap \)-convex function of $y$. Thus, matrix $H$ is negative semidefinite, and the third term on the right hand side of (39) is less than or equal to zero. Next, we consider the second term, which can be expressed as

\[ (y - y^0)^T \frac{\partial J_1(y^0)}{\partial y} = \text{Tr} \left\{ (Q - Q^\circ)^T \frac{\partial J_1(Q^\circ)}{\partial Q} \right\}. \]  

(40)

Note that $\text{diag}(\Theta^0) \Lambda_q = O$, $Q \geq 0$, and $\text{Tr}(Q) = 1$. The use of these results alongside (37) and (38) yields

\[ \text{Tr} \left\{ (Q - Q^\circ)^T \frac{\partial J_1(Q^\circ)}{\partial Q} \right\} \]

\[ = - \text{Tr} \left\{ \text{diag}(\Theta^0)U_q^0 Q U_q^0 + \text{diag}(\Theta^0)A_q^0 \right\} \]

\[ = - \text{Tr} \left\{ \text{diag}(\Theta^0) \text{diag}(U_q^0 Q U_q^0) \right\} \]

\[ \leq 0 \]  

(41)

which implies $J_1(y) \leq J_1(y^0)$ and thus, confirming the optimality of $Q^\circ$. 

ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewer for his suggestion that leads to a simplified derivation of the KKT equations in this paper.

REFERENCES