51st IEEE Conference on Decision and Control
December 10-13, 2012. Maui, Hawaii, USA

A Chernoff Relaxation on the Problem of Application-Oriented Finite Sample Experiment Design

Dimitrios Katselis, Cristian R. Rojas, Håkan Hjalmarsson and Mats Bengtsson

***Abstract*— In this paper, application-oriented experiment design formulated as a chance constrained problem is investigated. The chance constraint is based on the presumption that the estimated model can be used in an application to achieve a given performance level with a prescribed probability. The aforementioned performance level is dictated by the particular application of interest. The resulting optimization problem is known to be nonconvex in most cases. To this end, convexification is attempted by employing a Chernoff relaxation. As an application, we focus on the identification of multiple input multiple output (MIMO) wireless channel models based on a general L-optimality type of performance measure.**

I. INTRODUCTION

A basic subproblem in the context of system identification is that of experiment design. Overviews of this topic over the last decade can be found in [1], [2], [3], [4]. Contributions include convexification [5], robust design [6], least-costly design [7], and closed vs open loop experiments [8].

In the context of application-oriented experiment design, the experiment is designed to optimize a performance metric associated with the particular application where the estimated model will be used, [9], [10], [11]. A conceptual framework for application-oriented experiment design was outlined in [4]. The framework hinges on introducing a function J which quantifies the degradation in performance, when a model that differs from the true system is used, in the design of the application. Suppose that the performance is deemed acceptable if $J \leq 1/\gamma$ for some parameter γ, which we will call *accuracy*. Clearly, J is dependent on a model G and the set of admissible models is denoted as $\mathcal{E}_{adm} = \{G : J \leq 1/\gamma\}$. The system identification objective is then to select a model in $\mathcal{E}_{adm}$. Therefore, the least-costly experiment is given as follows:

$$\begin{aligned} &\min_{\text{Experiment}} \text{Experimental effort} \\ &s.t. \quad \hat{G} \in \mathcal{E}_{adm} \end{aligned} \tag{1}$$

where $\hat{G}$ is the identified model. For the experimental effort, different measures commonly used are input or output power, and experimental length.

In this paper, assuming complex-valued parametric model identification, our model is represented by a parameter vector $\mathbf{h} \in \mathbb{C}^n$ and its estimator by $\widehat{\mathbf{h}} \in \mathbb{C}^n$. Therefore, we can formulate the chance constrained version of (1) as follows:

$$\begin{aligned} \min\nolimits_{\text{Experiment}} \quad & \text{Experimental effort} \\ \text{s.t.} \quad & \mathbf{P}_{\mathcal{X}}\{\widehat{\mathbf{h}} \in \mathcal{E}_{adm}\} \geq 1 - \varepsilon. \end{aligned} \tag{2}$$

Here, $\mathbf{P}_{\mathcal{X}}\{Y\}$ is the probability of the event Y over the probability space corresponding to $\mathcal{X}$. In our context, $\mathcal{X}$ corresponds to $\widehat{\mathbf{h}}$ or to $(\widehat{\mathbf{h}}, \mathbf{h})$ depending on the employed parameter estimator. To this end, the parameter estimators employed in this paper are the minimum variance unbiased (MVU) and the minimum mean squared error (MMSE) estimators [12]. For the MVU case, the parameter estimator has the distribution:

$$\widehat{\mathbf{h}} \sim \mathcal{CN}(\mathbf{h}, \mathcal{I}_{\text{F,MVU}}^{-1}), \tag{3}$$

while for the MMSE case, the posterior parameter distribution is:

$$\mathbf{h}\,|\mathbf{y}, \mathbf{u} \sim \mathcal{CN}(\widehat{\mathbf{h}}, \mathcal{I}_{\text{F,MMSE}}^{-1}). \tag{4}$$

In the last two equations, $\mathcal{CN}(\bar{\mathbf{x}}, \mathbf{Q})$ stands for the circularly symmetric complex Gaussian distribution with mean $\bar{\mathbf{x}}$ and covariance $\mathbf{Q}$, $\mathcal{I}_{\text{F,MVU}}$ [12] and $\mathcal{I}_{\text{F,MMSE}}$ are the inverse covariance matrices for the MVU and MMSE estimators, respectively and $\mathbf{u} \in \mathbb{C}^{n_T \times B}$, $\mathbf{y} \in \mathbb{C}^{n_R \times B}$ are the (complex-valued) input and output data. More details will be provided in Section III.

From the definition of $\mathcal{E}_{adm}$ and with the rest of our assumptions, it is clear that the chance constraint in the last problem can be alternatively written as

$$\mathbf{P}_{\mathcal{X}}\{J(\widetilde{\mathbf{h}}, \mathbf{h}) \geq 1/\gamma\} \leq \varepsilon, \tag{5}$$

where $\widetilde{\mathbf{h}} = \widehat{\mathbf{h}} - \mathbf{h}$ denotes the parameter estimation error. This constraint is typically nonconvex. In [4], this chance constraint is replaced by a convex linear matrix inequality (LMI) equivalent to the relation $\mathcal{E}_{id} \subseteq \mathcal{E}_{adm}$, where $\mathcal{E}_{id}$ corresponds to a confidence ellipsoid for $\hat{G}$. Alternative convex relaxations have been subsequently suggested in [13]. In this paper, following the analysis in [13], a Chernoff relaxation is employed to approximate the chance constraint (5). This approximation is the tightest among those described in [13]. However, the tightness comes at the cost of an increased computational burden due to the necessity of tuning a certain scalar parameter via line search, as we illustrate in the following. We propose here a way to annihilate this burden, thus making attractive the complexity of the aforementioned Chernoff relaxation. An alternative analysis to the one in this paper has been presented in [14]. There, a Markov

D. Katselis, C. R. Rojas, H. Hjalmarsson and M. Bengtsson are with the ACCESS Linnaeus Center, Electrical Engineering, KTH – Royal Institute of Technology,S-100 44 Stockholm, Sweden. Emails: {dimitrios.katselis—cristian.rojas—hakan.hjalmarsson—mats.bengtsson} @ee.kth.se, Post: KTH School of Electrical Engineering, Automatic Control, SE-100 44 Stockholm, Sweden.

978-1-4673-2066-5/12/$31.00 ©2012 IEEE

bound relaxation of the chance constraint is pursued, yielding semidefinite programming (SDP) relaxations of (2). The used bound in [14] is looser than the one presented in this paper.

The paper is organized as follows. In Section II, the Chernoff bound is used to convexify the chance constraint in (5) based on the analysis in [13]. Section III introduces the MIMO communication system model and other necessary definitions for our purposes. In Section IV, we adjust the earlier derived Chernoff relaxation to the case of MIMO systems. Section V summarizes the specific L-optimality performance measure of interest. Some simulation examples are given in Section VI. Finally, the paper is concluded in Section VII.

Notation: ^{T}, ^{H} and * denote the transposition, Hermitian transposition and complex conjugation operators. $\|\cdot\|$ is the Euclidean norm of a (complex) vector or a (complex) matrix. $\Re\{\cdot\}$ and $\Im\{\cdot\}$ denote the real and imaginary parts of a complex number or vector, respectively. For a matrix $\mathbf{A}$, $\mathbf{A}_{i,j}$ denotes its (i,j)th element. For a square matrix $\mathbf{A}$, $\lambda_i(\mathbf{A})$ denotes its ith eigenvalue, while $\lambda_{\max}(\mathbf{A})$ is its maximum eigenvalue. Moreover, $\mathbf{1}_{m,n}$ denotes the all ones $m \times n$ matrix. Finally, $\mathrm{vec}(\cdot)$ denotes the vectorization of a matrix, i.e., the stacking of its columns into a single vector.

II. Chernoff Relaxation

A possible convex approximation of the chance constraint (5) can be based on the Chernoff inequality [15]. To this end, we assume that J is or can be approximated by a quadratic form with respect to $\widetilde{\mathbf{h}}$:

$$J(\widetilde{\mathbf{h}}, \mathbf{h}) = \widetilde{\mathbf{h}}^H \boldsymbol{\mathcal{I}}_{adm} \widetilde{\mathbf{h}}, \tag{6}$$

where $\boldsymbol{\mathcal{I}}_{adm}$ is a Hermitian positive semidefinite matrix possibly dependent on $\mathbf{h}$.

Using the analysis in [13], the chance constraint can be relaxed as follows:

$$\begin{aligned}
&\mathbf{P}_{\mathcal{X}}\{J(\widetilde{\mathbf{h}}, \mathbf{h}) \geq 1/\gamma\} = \\
&\mathbf{P}_{\mathbf{x}}\{\mathbf{x}^H \boldsymbol{\mathcal{I}}_{\mathrm{F}}^{-1/2} \boldsymbol{\mathcal{I}}_{adm} \boldsymbol{\mathcal{I}}_{\mathrm{F}}^{-1/2} \mathbf{x} \geq 1/\gamma\} \leq \\
&E_{\widetilde{\mathbf{x}}}\left\{\exp\left[\sum_{i=1}^{n} \frac{1}{t} \lambda_i \left(\boldsymbol{\mathcal{I}}_{\mathrm{F}}^{-1/2} \boldsymbol{\mathcal{I}}_{adm} \boldsymbol{\mathcal{I}}_{\mathrm{F}}^{-1/2}\right) |\widetilde{x}_i|^2 - \frac{1}{\gamma t}\right]\right\}
\end{aligned}$$

where $\boldsymbol{\mathcal{I}}_{\mathrm{F}}$ is ether $\boldsymbol{\mathcal{I}}_{\mathrm{F,MVU}}$ or $\boldsymbol{\mathcal{I}}_{\mathrm{F,MMSE}}$ depending on the employed estimator, $\mathbf{x} = \boldsymbol{\mathcal{I}}_{\mathrm{F}}^{1/2} \widetilde{\mathbf{h}} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I})$ and $\widetilde{\mathbf{x}} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I})$ is another complex, circularly symmetric, Gaussian random vector corresponding to a suitable rotation of $\mathbf{x}$. Furthermore, $t > 0$ is an arbitrary constant that can be used to control the tightness of the Chernoff bound to $\mathbf{P}_{\mathcal{X}}\{J(\widetilde{\mathbf{h}}, \mathbf{h}) \geq 1/\gamma\}$.

Based on the circular symmetry of $\widetilde{\mathbf{x}}$, the random variable $2|\widetilde{x}_i|^2 = 2\left(\Re\{\widetilde{x}_i\}\right)^2 + 2\left(\Im\{\widetilde{x}_i\}\right)^2$ is chi-squared distributed with two degrees of freedom. Note that the variances of the real and imaginary parts are both equal to $1/2$. The independence of $\left(\Re\{\widetilde{x}_i\}\right)^2$ and $\left(\Im\{\widetilde{x}_i\}\right)^2$ alleviates us from the analytical burden of using the chi-squared distribution in the following analysis, by exploring the existence of the exponential function inside the expectation. To this end, following [13], we may use the result:

$$E\left[\exp(ty^2)\right] = \frac{1}{\sqrt{1-t}} \tag{7}$$

for $t \in (-\infty, 1)$. Here, $y = \Re\{\widetilde{x}_i\}$ or $y = \Im\{\widetilde{x}_i\}$, i.e., $y \sim \mathcal{N}(0, 1/2)$. Using the previous results, we end up with the following chance constraint bound:

$$\begin{aligned}
&\mathbf{P}_{\mathcal{X}}\{J(\widetilde{\mathbf{h}}, \mathbf{h}) \geq 1/\gamma\} \leq \\
&\exp\left(-\frac{1}{\gamma t}\right) \prod_{i=1}^{n} \frac{1}{1 - \frac{1}{t}\lambda_i\left(\boldsymbol{\mathcal{I}}_{\mathrm{F}}^{-1/2} \boldsymbol{\mathcal{I}}_{adm} \boldsymbol{\mathcal{I}}_{\mathrm{F}}^{-1/2}\right)} = \\
&\exp\left(-\frac{1}{\gamma t}\right) \frac{1}{\det\left(\mathbf{I} - \frac{1}{t}\boldsymbol{\mathcal{I}}_{\mathrm{F}}^{-1/2} \boldsymbol{\mathcal{I}}_{adm} \boldsymbol{\mathcal{I}}_{\mathrm{F}}^{-1/2}\right)},
\end{aligned} \tag{8}$$

where t is allowed to take any value in $\left(\lambda_{\max}\left(\boldsymbol{\mathcal{I}}_{\mathrm{F}}^{-1/2} \boldsymbol{\mathcal{I}}_{adm} \boldsymbol{\mathcal{I}}_{\mathrm{F}}^{-1/2}\right), \infty\right)$. Based on this bound, a sufficient condition for (5) to hold is:

$$\exp\left(-\frac{1}{\gamma t}\right) \frac{1}{\det\left(\mathbf{I} - \frac{1}{t}\boldsymbol{\mathcal{I}}_{\mathrm{F}}^{-1/2} \boldsymbol{\mathcal{I}}_{adm} \boldsymbol{\mathcal{I}}_{\mathrm{F}}^{-1/2}\right)} \leq \varepsilon.$$

Using the analysis in [13], this is equivalent to:

$$\begin{aligned}
&-t \ln \varepsilon - t \ln \det\left(\mathbf{I} - \frac{1}{t}\mathbf{M}\right) \leq \frac{1}{\gamma}, \\
&\begin{bmatrix} \mathbf{M} & \boldsymbol{\mathcal{I}}_{adm}^{1/2} \\ \boldsymbol{\mathcal{I}}_{adm}^{1/2} & \boldsymbol{\mathcal{I}}_{\mathrm{F}} \end{bmatrix} \geq 0, \begin{bmatrix} t\mathbf{I} & \boldsymbol{\mathcal{I}}_{adm}^{1/2} \\ \boldsymbol{\mathcal{I}}_{adm}^{1/2} & \boldsymbol{\mathcal{I}}_{\mathrm{F}} \end{bmatrix} \geq 0.
\end{aligned} \tag{9}$$

Here, $\mathbf{M} = \mathbf{M}^H \in \mathbb{C}^{n_T n_R \times n_T n_R}$ is an auxiliary (free) matrix. Note that the first equation in (9) is a convex constraint jointly in t and $\mathbf{M}$, although it cannot be written as a linear matrix inequality.

III. MIMO System Model

We consider a MIMO communication system with n_T antennas at the transmitter and n_R antennas at the receiver [16]. The received signal at time t is modelled as

$$\mathbf{y}(t) = \mathbf{H}\mathbf{x}(t) + \mathbf{n}(t)$$

where $\mathbf{x}(t) \in \mathbb{C}^{n_T}$ and $\mathbf{y}(t) \in \mathbb{C}^{n_R}$ are the baseband representations of the transmitted and received signals, respectively. The impact of background noise and interference from adjacent communication links is represented by the additive term $\mathbf{n}(t) \in \mathbb{C}^{n_R}$. We will further assume that $\mathbf{n}(t)$ is a (weakly) stationary signal[1]. The channel response is modeled by $\mathbf{H} \in \mathbb{C}^{n_R \times n_T}$, which is assumed constant during the transmission of one block of data. In the context of either the MVU or the MMSE estimators, two different models of the channel will be considered:

i) A deterministic model.
ii) A stochastic Rayleigh fading model, i.e. $\mathrm{vec}(\mathbf{H}) \in \mathcal{CN}(\mathbf{0}, \mathbf{R})$, where, for mathematical tractability, we will

[1]Note that the statistical characterization of $\mathbf{x}(t)$ is irrelevant in our context, thus omitted.

assume that the known covariance matrix $\mathbf{R}$ possesses the Kronecker model used, e.g., in [17], [18]:

$$\mathbf{R} = \mathbf{R}_T^T \otimes \mathbf{R}_R \tag{10}$$

where $\mathbf{R}_T \in \mathbb{C}^{n_T \times n_T}$ and $\mathbf{R}_R \in \mathbb{C}^{n_R \times n_R}$ are the spatial covariance matrices at the transmitter and receiver side, respectively. Here, $\otimes$ denotes the Kronecker product [19]. This model has been experimentally verified in [20], [21] and further motivated in [22], [23].

We consider training signals of arbitrary length B, represented by $\mathbf{P} \in \mathbb{C}^{n_T \times B}$, whose columns are the transmitted signal vectors during training. Placing the received vectors in $\mathbf{Y} = \begin{bmatrix}\mathbf{y}(1) & \dots & \mathbf{y}(B)\end{bmatrix} \in \mathbb{C}^{n_R \times B}$, we have:

$$\mathbf{Y} = \mathbf{H}\mathbf{P} + \mathbf{N},$$

where $\mathbf{N} = \begin{bmatrix}\mathbf{n}(1) & \dots & \mathbf{n}(B)\end{bmatrix} \in \mathbb{C}^{n_R \times B}$ is the combined noise and interference matrix.

Defining $\widetilde{\mathbf{P}} = \mathbf{P}^T \otimes \mathbf{I}$, we can then write

$$\mathrm{vec}(\mathbf{Y}) = \widetilde{\mathbf{P}}\,\mathrm{vec}(\mathbf{H}) + \mathrm{vec}(\mathbf{N}). \tag{11}$$

As, for example, in [17], [18], we assume that $\mathrm{vec}(\mathbf{N}) \in \mathcal{CN}(\mathbf{0}, \mathbf{S})$, where the covariance matrix $\mathbf{S}$ also possesses a Kronecker structure:

$$\mathbf{S} = \mathbf{S}_Q^T \otimes \mathbf{S}_R. \tag{12}$$

Here, $\mathbf{S}_Q \in \mathbb{C}^{B \times B}$ represents the temporal covariance matrix[2] and $\mathbf{S}_R \in \mathbb{C}^{n_R \times n_R}$ represents the received spatial covariance matrix.

In the case of the MMSE estimator, the channel and noise statistics will be assumed known to the receiver during estimation, while in the case of the MVU estimator only the noise statistics will be considered known. Statistics can often be achieved by long-term estimation and tracking [24].

As far as the different ways to estimate the MIMO channel $\mathbf{H}$ are concerned, the MVU channel estimator for the signal model (11), subject to a deterministic channel (Assumption i), is given by:

$$\mathrm{vec}(\widehat{\mathbf{H}}_{\mathrm{MVU}}) = (\widetilde{\mathbf{P}}^H \mathbf{S}^{-1} \widetilde{\mathbf{P}})^{-1} \widetilde{\mathbf{P}}^H \mathbf{S}^{-1}\,\mathrm{vec}(\mathbf{Y}). \tag{13}$$

For this estimate, the inverse covariance is

$$\boldsymbol{\mathcal{I}}_{\mathrm{F,MVU}} = \widetilde{\mathbf{P}}^H \mathbf{S}^{-1} \widetilde{\mathbf{P}}. \tag{14}$$

For the case of a stochastic channel model (Assumption ii), the first and second moments of the posterior parameter vector are

$$\begin{aligned}\mathrm{vec}(\widehat{\mathbf{H}}_{\mathrm{MMSE}}) &= (\mathbf{R}^{-1} + \widetilde{\mathbf{P}}^H \mathbf{S}^{-1} \widetilde{\mathbf{P}})^{-1} \widetilde{\mathbf{P}}^H \mathbf{S}^{-1}\,\mathrm{vec}(\mathbf{Y}) \\ \boldsymbol{\mathcal{I}}_{\mathrm{F,MMSE}}^{-1} &= (\mathbf{R}^{-1} + \widetilde{\mathbf{P}}^H \mathbf{S}^{-1} \widetilde{\mathbf{P}})^{-1}.\end{aligned} \tag{15}$$

[2] We set the subscript Q to $\mathbf{S}_Q$ to highlight its temporal nature and the fact that its size is $B \times B$, thus $\mathbf{S}_Q \neq \mathbf{R}_T$. In this paper, the matrices with subscript R are $n_R \times n_R$.

IV. Convex Formulations Based on the Chernoff Relaxation

The experimental effort of interest in our context is assumed to be the input power. Focusing on the case of the MVU estimator and using (9), the optimization problem (2) is relaxed as follows:

$$\begin{aligned}\min_{\mathbf{P},\mathbf{M},t} \quad & \mathrm{Tr}\left[\mathbf{P}\mathbf{P}^H\right] = \mathrm{Tr}\left[\mathbf{P}^*\mathbf{P}^T\right] \\ \text{s.t.} \quad & \begin{bmatrix} \mathbf{M} & \boldsymbol{\mathcal{I}}_{adm}^{1/2} \\ \boldsymbol{\mathcal{I}}_{adm}^{1/2} & \mathbf{P}^*\mathbf{S}_Q^{-T}\mathbf{P}^T \otimes \mathbf{S}_R^{-1} \end{bmatrix} \geq 0, \\ & \begin{bmatrix} t\mathbf{I} & \boldsymbol{\mathcal{I}}_{adm}^{1/2} \\ \boldsymbol{\mathcal{I}}_{adm}^{1/2} & \mathbf{P}^*\mathbf{S}_Q^{-T}\mathbf{P}^T \otimes \mathbf{S}_R^{-1} \end{bmatrix} \geq 0, \\ & -t\ln\varepsilon - t\ln\det\left(\mathbf{I} - \tfrac{1}{t}\mathbf{M}\right) \leq \tfrac{1}{\gamma}.\end{aligned}$$

Clearly, the last formulation is a relaxation of (2) based on the Chernoff bound and no further relaxations. Furthermore, we set $\mathbf{P}_Q = \mathbf{P}^*\mathbf{S}_Q^{-T/2}$. Using an additional free variable $\beta \in \mathbb{R}$, the last optimization problem takes the form

$$\begin{aligned}\min_{\beta,\mathbf{P}_Q,\mathbf{M},t} \quad & \beta \\ \text{s.t.} \quad & \mathrm{Tr}\left[\mathbf{P}_Q\mathbf{S}_Q^T\mathbf{P}_Q^H\right] \leq \beta \\ & \begin{bmatrix} \mathbf{M} & \boldsymbol{\mathcal{I}}_{adm}^{1/2} \\ \boldsymbol{\mathcal{I}}_{adm}^{1/2} & \mathbf{P}_Q\mathbf{P}_Q^H \otimes \mathbf{S}_R^{-1} \end{bmatrix} \geq 0, \\ & \begin{bmatrix} t\mathbf{I} & \boldsymbol{\mathcal{I}}_{adm}^{1/2} \\ \boldsymbol{\mathcal{I}}_{adm}^{1/2} & \mathbf{P}_Q\mathbf{P}_Q^H \otimes \mathbf{S}_R^{-1} \end{bmatrix} \geq 0, \\ & -t\ln\varepsilon - t\ln\det\left(\mathbf{I} - \tfrac{1}{t}\mathbf{M}\right) \leq \tfrac{1}{\gamma}.\end{aligned} \tag{16}$$

Since our target variable is $\mathbf{P}_Q$, it would be of interest to convexify the last formulation with respect to this decision variable. In essence, what we would like to achieve is to cope with the existence of both $\mathbf{P}_Q\mathbf{P}_Q^H$ and $\mathbf{P}_Q^H\mathbf{P}_Q$ in the last formulation. There are two alternative ways in achieving this goal, which we summarize in the following.

The first way is to introduce $\mathbf{X} = \mathrm{vec}(\mathbf{P}_Q^H)\,\mathrm{vec}^T(\mathbf{P}_Q^T) = \mathrm{vec}(\mathbf{P}_Q^H)\,\mathrm{vec}^H(\mathbf{P}_Q^H) \in \mathbb{C}^{n_T B \times n_T B}$ and use the following identity:

$$\mathrm{Tr}[\mathbf{P}_Q\mathbf{S}_Q^T\mathbf{P}_Q^H] = \mathrm{vec}^T(\mathbf{S}_Q\mathbf{P}_Q^T)\,\mathrm{vec}(\mathbf{P}_Q^H) = \mathrm{Tr}[\mathbf{X}(\mathbf{I} \otimes \mathbf{S}_Q^T)],$$

which is based on simple properties of the vectorization operator. Note that this expression is linear in $\mathbf{X}$. Furthermore, setting $\mathbf{Z} = \mathbf{P}_Q\mathbf{P}_Q^H$ we have

$$\begin{aligned}\mathbf{Z}_{i,k} = (\mathbf{P}_Q\mathbf{P}_Q^H)_{i,k} &= \sum_{m=1}^{B} (\mathbf{P}_Q)_{i,m}(\mathbf{P}_Q^H)_{m,k} \\ &= \sum_{m=1}^{B} \mathbf{X}_{m+(k-1)B,\,m+(i-1)B},\end{aligned}$$

which can be alternatively expressed as

$$\mathbf{Z} = (\mathbf{I}_{n_T} \otimes \mathbf{1}_{1\times B})\left\{(\mathbf{1}_{n_T \times n_T} \otimes \mathbf{I}_B) \odot \mathbf{X}^T\right\}(\mathbf{I}_{n_T} \otimes \mathbf{1}_{B\times 1}), \tag{17}$$

$\odot$ denoting the Hadamard or elementwise matrix product [25]. Note that this expression is also linear in $\mathbf{X}$.

Combining the previous results, we can write (16) as:

$$\begin{array}{ll}\min_{\beta,\mathbf{X},\mathbf{M},\mathbf{Z},t} & \beta \\ \text{s.t.} & \begin{bmatrix} \mathbf{M} & \boldsymbol{\mathcal{I}}_{adm}^{1/2} \\ \boldsymbol{\mathcal{I}}_{adm}^{1/2} & \mathbf{Z}\otimes\mathbf{S}_R^{-1}\end{bmatrix} \geq 0, \\ & \begin{bmatrix} t\mathbf{I} & \boldsymbol{\mathcal{I}}_{adm}^{1/2} \\ \boldsymbol{\mathcal{I}}_{adm}^{1/2} & \mathbf{Z}\otimes\mathbf{S}_R^{-1}\end{bmatrix} \geq 0, \\ & -t\ln\varepsilon - t\ln\det\left(\mathbf{I}-\frac{1}{t}\mathbf{M}\right) \leq \frac{1}{\gamma}, \\ & \mathbf{Z} = \left(\mathbf{I}_{n_T}\otimes\mathbf{1}_{1\times B}\right)\left\{\left(\mathbf{1}_{n_T\times n_T}\otimes\mathbf{I}_B\right)\odot\mathbf{X}^T\right\}\cdot \\ & \left(\mathbf{I}_{n_T}\otimes\mathbf{1}_{B\times 1}\right), \\ & \mathrm{Tr}\left[\mathbf{X}\left(\mathbf{I}\otimes\mathbf{S}_Q^T\right)\right] \leq \beta,\ \mathbf{X}\geq\mathbf{0},\ \mathrm{rank}\left[\mathbf{X}\right]=1.\end{array} \tag{18}$$

It is clear that by introducing $\mathbf{X}$, we have managed to eliminate the formulation from the existence of both products $\mathbf{P}_Q\mathbf{P}_Q^H$ and $\mathbf{P}_Q^H\mathbf{P}_Q$ and at the same time to come up with a formulation that is linear in the substituting variable $\mathbf{X}$.

Problem (18) is nonconvex. Ignoring the rank constraint, this problem is convex. We can use similar handling as in the case of rank-constrained SDP problems to obtain a solution for our target variable $\mathbf{P}_Q$. Upon obtaining the optimal $\mathbf{X}$, say $\mathbf{X}_*$, the rank-one solution is selected to be equal to $\sqrt{\lambda_1}\mathbf{q}_1$, where λ_1 is the greatest eigenvalue of $\mathbf{X}_*$ and $\mathbf{q}_1$ the corresponding eigenvector. We underline here that this is an intuitive but otherwise *ad hoc* solution, which has been observed to deliver good performance in practice, in the context of many rank-one constrained problems that lead to SDP formulations after the elimination of this rank-one constraint [26]. Nevertheless, within the context of MIMO channel estimation, it has been observed that this solution does not necessarily deliver good performance [14], mainly because of the geometry-destructive de-vectorization step that has to be performed to yield $\mathbf{P}_Q^H$ and therefore the optimal training matrix $\mathbf{P}_*$.

Trying to avoid the de-vectorization, we also come to the second way of how to cope with the existence of both $\mathbf{P}_Q\mathbf{P}_Q^H$ and $\mathbf{P}_Q^H\mathbf{P}_Q$ in (16). Using similar identities as before for $\mathrm{Tr}\left[\mathbf{P}_Q\mathbf{S}_Q^T\mathbf{P}_Q^H\right]$ and the Cauchy-Schwarz inequality, we can obtain the following bound:

$$\mathrm{Tr}\left[\mathbf{P}_Q\mathbf{S}_Q^T\mathbf{P}_Q^H\right] \leq \left\|\mathrm{vec}\left(\mathbf{S}_Q^*\right)\right\| \left\|\mathrm{vec}\left(\mathbf{Z}\right)\right\|, \tag{19}$$

where we have used the fact that $\|\mathrm{vec}(\mathbf{P}_Q^H\mathbf{P}_Q)\| = \|\mathrm{vec}(\mathbf{P}_Q\mathbf{P}_Q^H)\|$.

Using (19), a constrained version of (16) is obtained:

$$\begin{array}{ll}\min_{\beta,\mathbf{Z},\mathbf{M},t} & \beta \\ \text{s.t.} & \begin{bmatrix} \mathbf{M} & \boldsymbol{\mathcal{I}}_{adm}^{1/2} \\ \boldsymbol{\mathcal{I}}_{adm}^{1/2} & \mathbf{Z}\otimes\mathbf{S}_R^{-1}\end{bmatrix} \geq 0, \\ & \begin{bmatrix} t\mathbf{I} & \boldsymbol{\mathcal{I}}_{adm}^{1/2} \\ \boldsymbol{\mathcal{I}}_{adm}^{1/2} & \mathbf{Z}\otimes\mathbf{S}_R^{-1}\end{bmatrix} \geq 0, \\ & -t\ln\varepsilon - t\ln\det\left(\mathbf{I}-\frac{1}{t}\mathbf{M}\right) \leq \frac{1}{\gamma}, \\ & \|\mathrm{vec}(\mathbf{Z})\| \leq \beta/\|\mathrm{vec}(\mathbf{S}_Q^*)\|, \quad \mathbf{Z}=\mathbf{Z}^H\geq\mathbf{0},\end{array} \tag{20}$$

which is a convex problem.

Although (20) is a convex problem, the usual optimization packages can not handle the $\ln\det$ for the free variable t. This is due to the division $1/t$ in the case of CVX. The usual way to deal with this problem is to tune t via line search [13]. However, this translates to great computational burden. First, a good initial value for t has to be guessed and an appropriate stepsize for the update of t to be selected. These choices are problem dependent. Then, the line search has to be established on the basis of descent directions and the employed package has to solve a sequence of optimization problems for every new t, until the training matrix of minimum power is reached within a predetermined proximity level.

An idea to overcome this problem relies on lower bounding the left hand side of the $\ln\det$ constraint in such a way that the division $1/t$ vanishes. Clearly, the tightest the lower bound, the better. We propose here such a bound, although better bounds may exist. Note that $-t\ln\det\left(\mathbf{I}-\frac{1}{t}\mathbf{M}\right) = -t\sum_{i=1}^{n_T n_R}\ln\left(1-\lambda_i(\mathbf{M})/t\right)$, where $\{\lambda_i(\mathbf{M})\}_{i=1}^{n_T n_R}$ are the eigenvalues of $\mathbf{M}$. Since it is also required that $t^{-1}\mathbf{M} < \mathbf{I}$ (so that the $\ln\det$ expression is defined in $\mathbb{R}$) , we have that $t^{-1}\lambda_i(\mathbf{M}) < 1$ for all $i = 1, 2, \ldots, n_R n_T$. Therefore, using the Taylor expansion theorem, we obtain

$$\ln\left(1-\frac{1}{t}\lambda_i(\mathbf{M})\right) = -\sum_{k=1}^{\infty}\frac{1}{k}\left(\frac{1}{t}\lambda_i(\mathbf{M})\right)^k < -\frac{1}{t}\lambda_i(\mathbf{M}),$$

since all the summands are negative. Combining the above results, the following lower bound is obtained:

$$-t\ln\varepsilon + \mathrm{tr}\left(\mathbf{M}\right) \leq -t\ln\varepsilon - t\ln\det\left(\mathbf{I}-\frac{1}{t}\mathbf{M}\right) \leq \frac{1}{\gamma} \tag{21}$$

We therefore replace the $\ln\det$ constraint in (20) by $-t\ln\varepsilon + \mathrm{tr}\left(\mathbf{M}\right) \leq 1/\gamma$. This corresponds to a relaxation of the $\ln\det$ constraint.

Upon obtaining the optimal $\mathbf{Z}$, say $\mathbf{Z}_*$, the selection of the training matrix $\mathbf{P}$ has to be performed. To this end, assume that the eigenvalue decomposition (EVD) of $\mathbf{Z}_*$ is $\mathbf{U}_{Z_*}\mathbf{D}_{Z_*}\mathbf{U}_{Z_*}^H$, where $\mathbf{U}_{Z_*}\in\mathbb{C}^{n_T\times n_T}$ is its modal matrix and $\mathbf{D}_{Z_*}\in\mathbb{R}^{n_T\times n_T}$ a diagonal matrix containing its eigenvalues in decreasing order. Assume also that the EVD of $\mathbf{S}_Q^T$ is $\mathbf{U}_Q\mathbf{D}_Q\mathbf{U}_Q^H$, where $\mathbf{U}_Q\in\mathbb{C}^{B\times B}$ is its modal matrix and $\mathbf{D}_Q\in\mathbb{R}^{B\times B}$ is a diagonal matrix containing its eigenvalues in order that will be determined in the following. We denote as $\mathbf{U}_{P^*}\mathbf{D}_{P^*}\mathbf{V}_{P^*}^H$ the singular value decomposition (SVD) of $\mathbf{P}^*$. Since $\mathbf{Z} = \mathbf{P}_Q\mathbf{P}_Q^H = \mathbf{P}^*\mathbf{S}_Q^{-T}\mathbf{P}^H$, it is clear that there is an infinite number of $\mathbf{P}_Q$'s which can produce $\mathbf{Z}_*$, since if $\mathbf{P}_{Q_*}$ is such a choice then $\mathbf{P}_{Q_*}\mathbf{\Gamma}$ is also a valid choice, where $\mathbf{\Gamma}\in\mathbb{C}^{B\times B}$ is an arbitrary unitary matrix[3]. This argument shows that our main concern with respect to the selection of $\mathbf{P}$ is the formation of $\mathbf{Z}_*$. An immediate and intuitive way to make such a choice is to select $\mathbf{U}_{P^*} = \mathbf{U}_{Z_*}$ and $\mathbf{V}_{P^*} = \mathbf{U}_Q$. This implies that $\mathbf{D}_{P^*}$ must satisfy the following relationship:

$$\mathbf{D}_{Z_*} = \mathbf{D}_{P^*}\mathbf{D}_Q^{-1}\mathbf{D}_{P^*} \tag{22}$$

[3]If $\mathbf{P}_Q$ is tall, $\mathbf{Z} = \mathbf{P}_Q\mathbf{P}_Q^H$ implies a rank constraint on $\mathbf{Z}$. Here, we assume that B is large enough so that this is not an issue.

i.e., $(\mathbf{D}_{P^*}(i,i))^2 = \mathbf{D}_{Z_*}(i,i)\mathbf{D}_Q(i,i), i = 1,\ldots,n_T$. Additionally, the ordering of the eigenvalues $\mathbf{D}_Q(i,i), i = 1,\ldots,n_T$, should be such that

$$\mathrm{Tr}[\mathbf{P}^*\mathbf{P}^T] = \sum_{i=1}^{n_T}(\mathbf{D}_{P^*}(i,i))^2 \quad (23)$$

is minimized. By our assumptions and using Lemma 1 in [27], the diagonal entries of $\mathbf{D}_Q$ should be arranged in ascending order. The above choices determine the optimal $\mathbf{P}^*$, i.e., the optimal $\mathbf{P}$, say $\mathbf{P}_*$.

As for the MMSE channel estimator, the corresponding optimization problem will be as follows:

$$\begin{array}{ll}\min_{\beta,\mathbf{Z},\mathbf{M},t} & \beta \\ \text{s.t.} & \begin{bmatrix}\mathbf{M} & \mathcal{I}_{adm}^{1/2} \\ \mathcal{I}_{adm}^{1/2} & \mathbf{R}^{-1}+\mathbf{Z}\otimes\mathbf{S}_R^{-1}\end{bmatrix} \geq 0 \\ & \begin{bmatrix}t\mathbf{I} & \mathcal{I}_{adm}^{1/2} \\ \mathcal{I}_{adm}^{1/2} & \mathbf{R}^{-1}+\mathbf{Z}\otimes\mathbf{S}_R^{-1}\end{bmatrix} \geq 0, \\ & -t\ln\varepsilon + \mathrm{tr}\,(\mathbf{M}) \leq 1/\gamma, \\ & \|\,\mathrm{vec}(\mathbf{Z})\| \leq \beta/\|\,\mathrm{vec}(\mathbf{S}_Q^*)\|, \quad \mathbf{Z}=\mathbf{Z}^H \geq \mathbf{0},\end{array}$$

and the choice of the optimal $\mathbf{P}$ will be exactly the same as in the case of the MVU estimator.

Remarks:

1) The proposed choice of $\mathbf{P}$ in both cases is *one* possible solution among infinite others that achieve $\mathbf{Z}_*$. Its optimality is with respect to this aspect. However, in strict mathematical sense this choice is *ad hoc* to the same degree as the usual best rank-one approximation of $\mathbf{X}_*$ in the case of SDP-like handling. This is due to the numerical nature of the presented approach. Nevertheless, the proposed $\mathbf{P}$ is intuitive in the sense that it resembles many analytically derived optimal training matrices in the MIMO channel context, e.g., [28], [29], [30], [17], [27], [31].
2) In the case of the MMSE estimator, the information of $\mathbf{R}$ is indirectly encoded in the proposed training matrix $\mathbf{P}$ through the eigenvectors and the eigenvalues of the corresponding $\mathbf{Z}_*$.

V. Application: Optimal Training for the L-Optimality Performance Measure

Consider a performance index of the form:

$$J_W(\widetilde{\mathbf{H}},\mathbf{H}) = \mathrm{vec}^H(\widetilde{\mathbf{H}})\mathbf{W}\,\mathrm{vec}(\widetilde{\mathbf{H}}),$$

for some positive semidefinite weighting matrix $\mathbf{W}$. Taking the expected value of this performance metric with respect to either $\widetilde{\mathbf{H}}$ or both $\widetilde{\mathbf{H}}$ and $\mathbf{H}$ leads to the well-known L-optimality criterion for optimal experiment design in statistics [32]. In the context of MIMO communication systems, such a performance metric may arise, e.g., if we want to estimate the MIMO channel having some deficiencies in either the transmit and/or the receive antenna arrays. The simplest case would be that of a diagonal $\mathbf{W}$ with nonzero entries in the interval $[0,1]$. More general matrices can be considered if we assume crosscouplings between the transmit and/or receive antenna elements.

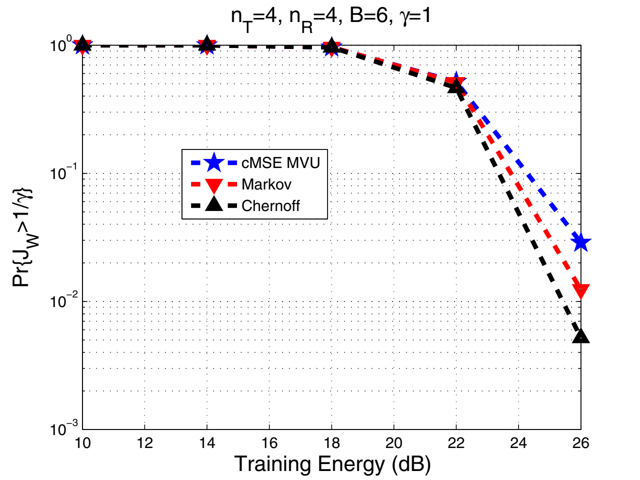

Fig. 1. $n_T = 4, n_R = 4, B = 6, \gamma = 1$: Outage probability for the L-optimality criterion with the MVU estimator. The accuracy parameter is $\gamma = 1$.

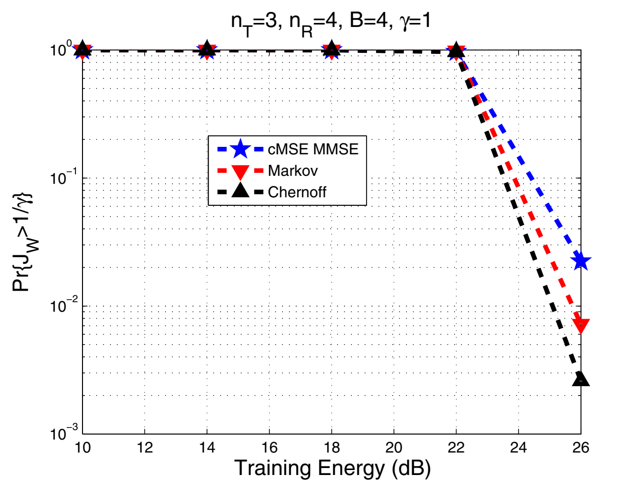

Fig. 2. $n_T = 3, n_R = 4, B = 4, \gamma = 1$: Outage probability for the L-optimality criterion with the MMSE estimator. The accuracy parameter is $\gamma = 1$.

VI. Simulations

The purpose of this section is to examine the performance of the Chernoff-relaxed chance constrained training sequence designs presented in this paper, and compare them with other existing methods. In all figures, fair comparison among the presented schemes is ensured via training energy equalization. Additionally, the matrices $\mathbf{R}_T, \mathbf{R}_R, \mathbf{S}_Q, \mathbf{S}_R$ follow the exponential model, that is, they are built according to

$$(\mathbf{R})_{i,j} = r^{j-i}, \;\; j \geq i, \quad (24)$$

where r is the (complex) normalized correlation coefficient with magnitude $\rho = |r| < 1$. We choose to examine the high correlation scenario for all the presented schemes. Therefore, in both plots $|r| = 0.9$ for all matrices $\mathbf{R}_T, \mathbf{R}_R, \mathbf{S}_Q, \mathbf{S}_R$.

Figs. 1 and 2 demonstrate the outage probability of different training designs, when the employed channel estimator is either the MVU or the MMSE, respectively. The accuracy parameter γ equals 1 in both figures. The corresponding parameter values for n_T, n_R and B are $n_T = 4, n_R = 4, B = 6$ in Fig. 1 and $n_T = 3, n_R = 4, B = 4$ in Fig. 2. The matrices $\mathbf{W}$ in both figures are randomly selected Hermitian positive semidefinite matrices, omitted here due to their large sizes. In Fig. 1, the scheme "cMSE MVU" corresponds to the optimal training for the MVU channel estimator aiming at minimizing the channel mean square error (MSE) in [33]. The corresponding scheme "cMSE MMSE" in Fig. 2 in the optimal training for MMSE channel estimation in [31]. The schemes "Markov" in both plots are the training matrices for MVU and MMSE channel estimation based on the SDP relaxations in [14][4]. Finally, the schemes "Chernoff" are the training matrices for MVU and MMSE channel estimation developed in this paper. Both figures demonstrate the superiority of the Chernoff relaxation over the other methods with respect to their corresponding outage probability, as it is intuitively expected.

VII. Conclusions

Convex relaxations for the chance constrained input design problem based on a Chernoff bound have been presented in this paper. The derived formulations were used in the context of channel estimation based on L-optimality criteria in the context of MIMO communication systems. Numerical results have verified that the Chernoff-relaxed input designs provide lower violation probabilities of the chance constraint over other existing methods, while maintaining an attractive computational complexity.

References

[1] M. Gevers, "Identification for control: From the early achievements to the revival of experiment design," *European Journal of Control*, vol. 11, pp. 1–18, 2005.

[2] H. Hjalmarsson, "From experiment design to closed-loop control," *Automatica*, vol. 41, no. 3, pp. 393–438, March 2005.

[3] L. Pronzato, "Optimal experimental design and some related control problems," *Automatica*, vol. 44(2), pp. 303–325, February 2008.

[4] H. Hjalmarsson, "System identification of complex and structured systems," *Plenary address European Control Conference / European Journal of Control*, vol. 15, no. 4, pp. 275–310, 2009.

[5] H. Jansson and H. Hjalmarsson, "Input design via LMIs admitting frequency-wise model specifications in confidence regions," *IEEE Trans. Autom. Control*, vol. 50, no. 10, pp. 1534–1549, 2005.

[6] C. R. Rojas, J. S. Welsh, G. C. Goodwin, and A. Feuer, "Robust optimal experiment design for system identification," *Automatica*, vol. 43, no. 6, pp. 993–1008, 2007.

[7] X. Bombois, G. Scorletti, M. Gevers, P. Van den Hof, and R. Hildebrand, "Least costly identification experiment for control," *Automatica*, vol. 42, no. 10, pp. 1651–1662, October 2006.

[8] J. C. Agüero and G. C. Goodwin, "Choosing between open- and closed-loop experiments in linear system identification," *IEEE Trans. Aut. Control*, vol. 52(8), pp. 1475–1480, August 2007.

[9] M. Gevers and L. Ljung, "Optimal experiment designs with respect to the intended model application," *Automatica*, vol. 22, no. 5, pp. 543–554, 1986.

[10] U. Forssell and L. Ljung, "Some results on optimal experiment design," *Automatica*, vol. 36, no. 5, pp. 749–756, May 2000.

[11] M. Barenthin, X. Bombois, H. Hjalmarsson, and G. Scorletti, "Identification for control of multivariable systems: Controller validation and experiment design via LMIs," *Automatica*, vol. 44(12), pp. 3070–3078, 2008.

[12] S. M. Kay, *Fundamentals of Statistical Signal Processing. Estimation Theory*. Englewood Cliffs, New Jersey: Prentice-Hall, 1993.

[13] C. R. Rojas, D. Katselis, H. Hjalmarsson, R. Hildebrand, and M. Bengtsson, "Chance constrained input design," in *Proceedings of CDC-ECC*, Orlando, Florida, USA, December 2011.

[14] D. Katselis, C. R. Rojas, H. Hjalmarsson, and M. Bengtsson, "Application-oriented finite sample experiment design: A semidefinite relaxation approach," in *SYSID 2012*, Brussels, Belgium, July 2012, invited.

[15] A. Papoulis, *Probability, Random Variables, and Stochastic Processes, 3rd Edition*. McGraw-Hill, 1991.

[16] A. Paulraj, R. Nabar, and D. Gore, *Introduction to Space-Time Wireless Communications*. Cambridge, United Kingdom: Cambridge University Press, 2003.

[17] Y. Liu, T. F. Wong, and W. W. Hager, "Training signal design for estimation of correlated MIMO channels with colored interference," *IEEE Trans. Signal Process.*, vol. 55, no. 4, pp. 1486–1497, Apr. 2007.

[18] M. Biguesh, S. Gazor, and M. H. Shariat, "Optimal training sequence for MIMO wireless systems in colored environments," *IEEE Trans. Signal Process.*, vol. 57, no. 8, pp. 3144–3153, Aug. 2009.

[19] J. W. Brewer, "Kronecker products and matrix calculus in system theory," *IEEE Transactions on Circuits and Systems*, vol. 25(9), pp. 772–781, 1978.

[20] J. P. Kermoal, L. Schumacher, K. I. Pedersen, P. E. Mogensen, and F. Fredriksen, "A stochastic MIMO radio channel model with experimental validation," *IEEE J. Sel. Areas Commun.*, vol. 20, no. 6, pp. 1211–1226, Aug. 2002.

[21] K. Yu, M. Bengtsson, B. Ottersten, D. McNamara, P. Karlsson, and M. Beach, "Modeling of wideband MIMO radio channels based on NLOS indoor measurements," *IEEE Trans. Veh. Technol.*, vol. 53, no. 3, pp. 655–665, May 2004.

[22] S. Gazor and H. S. Rad, "Space-time frequency characterization of MIMO wireless channels," *IEEE Trans. Wireless Commun.*, vol. 5, no. 9, pp. 2369–2376, Sep. 2006.

[23] H. S. Rad and S. Gazor, "The impact of non-isotropic scattering and directional antennas on MIMO multicarrier mobile communication channels," *IEEE Trans. Commun.*, vol. 56, no. 4, pp. 642–652, Apr. 2008.

[24] K. Werner and M. Jansson, "Estimating MIMO channel covariances from training data under the Kronecker model," *Signal Processing*, vol. 89, no. 1, pp. 1–13, Jan. 2009.

[25] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge: Cambridge University Press, 1985.

[26] I. R. Manchester, "Input design for system identification via convex relaxation," in *Proceedings of CDC*, Atlanta, Georgia, USA, December 2010.

[27] D. Katselis, E. Kofidis, and S. Theodoridis, "Training-based estimation of correlated MIMO fading channels in the presence of colored interference," *Signal Processing*, vol. 87, no. 0, pp. 2177–2187, Sep. 2007.

[28] J. H. Kotecha and A. M. Sayeed, "Transmit signal design for optimal estimation of correlated MIMO channels," *IEEE Trans. Signal Process.*, vol. 52, no. 2, pp. 546–557, Feb. 2004.

[29] T. F. Wong and B. Park, "Training sequence optimization in MIMO systems with colored interference," *IEEE Trans. Commun.*, vol. 52, no. 11, pp. 1939–1947, Nov. 2004.

[30] M. Biguesh and A. B. Gershman, "Training-based MIMO channel estimation: a study of estimator tradeoffs and optimal training signals," *IEEE Trans. Signal Process.*, vol. 54, no. 3, pp. 884–893, Mar. 2006.

[31] E. Björnson and B. Ottersten, "A framework for training-based estimation in arbitrarily correlated Rician MIMO channels with rician disturbance," *IEEE Trans. Signal Process.*, vol. 58, no. 3, Mar. 2010.

[32] J. Kiefer, "General equivalence theory for optimum designs (aproximate theory)," *Annals of Statistics*, vol. 2(5), pp. 849–879, 1974.

[33] D. Katselis, E. Kofidis, and S. Theodoridis, "On training optimization for estimation of correlated mimo chsnnels in the presence of multiuser interference," *IEEE Trans. Signal Process.*, vol. 56, no. 10, pp. 4892–4904, Oct. 2008.

[4] The reader may appreciate the fact that a Chernoff relaxation should be tighter than a Markov relaxation. This is verified by both plots in this paper.