

Optimal input design for system identification using spectral decomposition

Shravan Mohan, Mithun IM and Bharath Bhikkaji

Department of Electrical Engineering,
Indian Institute of Technology Madras, Chennai, India

Abstract

The aim of this paper is to design a band-limited optimal input with power constraints for identifying a linear multi-input multi-output system. It is assumed that the nominal system parameters are specified. The key idea is to use the spectral decomposition theorem and write the power spectrum as $\phi_u(j\omega) = \frac{1}{2}H(j\omega)H^*(j\omega)$. The matrix $H(j\omega)$ is expressed in terms of a truncated basis for $\mathcal{L}^2([- \omega_{\text{cut-off}}, \omega_{\text{cut-off}}])$. With this parameterization, the elements of the Fisher Information Matrix and the power constraints turn out to be homogeneous quadratics in the basis coefficients. The optimality criterion used are the well-known \mathcal{D} -optimality, \mathcal{A} -optimality, \mathcal{T} -optimality and \mathcal{E} -optimality. The resulting optimization problem is non-convex in general. A lower bound on the optimum is obtained through a bi-linear formulation of the problem, while an upper bound is obtained through a convex relaxation. These bounds can be computed efficiently as the associated problems are convex. The lower bound is used as a sub-optimal solution, the sub-optimality of which is determined by the difference in the bounds. Interestingly, the bounds match in many instances and thus, the global optimum is achieved. A discussion on the non-convexity of the optimization problem is also presented. Simulations are provided for corroboration.

Keywords: System Identification; Optimal Input Design; Spectral Decomposition; Fisher Information Matrix; Convex Optimization

1 Introduction

In this paper, the problem of designing a band-limited optimal input for identifying a continuous time linear multi-input multi-output (MIMO) system, subject to input and output power constraints, is considered. The linear system considered is given by

$$y(t) = G(s; \theta)u(t) + e(t), \quad (1)$$

where

$$G_{ij}(s; \theta) = \frac{\sum_{k=0}^{m_{ij}} b_{ij}^{(k)} s^k}{\sum_{k=0}^{n_{ij}} a_{ij}^{(k)} s^k}, \quad (2)$$

$$\theta^{ij} = \left\{ b_{ij}^{(0)}, \dots, b_{ij}^{(m_{ij})}, a_{ij}^{(0)}, \dots, a_{ij}^{(n_{ij})} \right\} \quad \& \quad \theta = \bigcup_{\{i,j\}} \theta^{ij}.$$

For all $t \in \mathcal{R}^+$, the input $u(t) \in \mathcal{R}^{r \times 1}$, the output $y(t) \in \mathcal{R}^{p \times 1}$ and the additive white noise $e(t) \sim \mathcal{N}(0, \Sigma)$ ($\Sigma \in \mathcal{S}_p^+$, the space of positive definite $p \times p$ matrices). It is assumed that the linear system under consideration is stable and the order of elements of the matrix $G(s; \theta)$ are specified. The aim of system identification, in general, is to estimate the set of parameters θ , the nominal values of which are specified, by exciting the system with inputs and recording the outputs. The schematic in Figure 1 shows the general system identification framework.

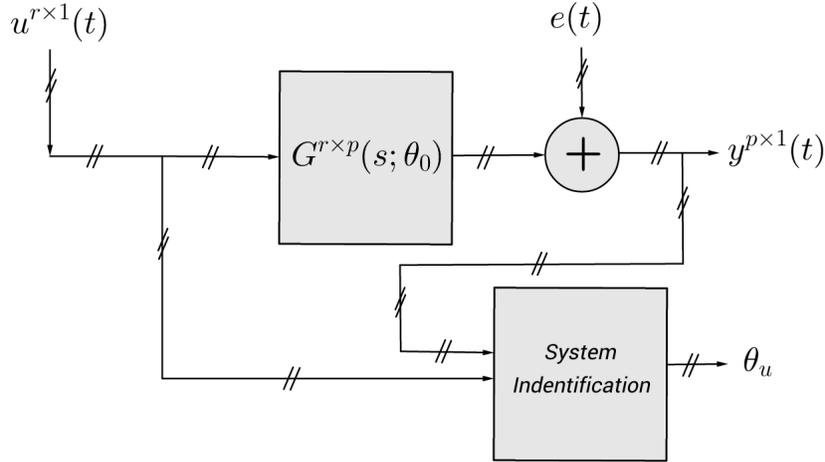


Figure 1: This figure shows a general optimal input design framework. $G^{r \times p}(s; \theta_0)$ represents the linear MIMO system with nominal parameters θ_0 . The input vector is given by $u^{r \times 1}(t)$, $e(t)$ is a normal random vector, *i.e.* $e(t) \sim \mathcal{N}^{p \times 1}(0, \Sigma)$, $y^{p \times 1}$ is the observed output. The estimation of the system parameters is done using subspace methods, which is represented by the box titled *system identification*. The estimated system parameters is given by the set θ_u . The double dashes in the figure represent multi-dimensional signals.

The quality of an unbiased estimator is judged by a suitable functional defined on the covariance matrix of the estimated system. It turns out that the covariance matrix of an unbiased estimator of a linear system is a matrix multiple of the inverse of the Fisher Information Matrix (FIM). Suppose that the set of parameters θ is indexed and its cardinality is N , then the FIM for (1) is defined as (see Kumar et al. [2015]):

$$[\mathcal{M}_\theta]_{l,m} = \int_{-\infty}^{\infty} \text{Tr} \left(\phi_e^{-1}(j\omega) \frac{\partial G(j\omega)}{\partial \theta_l} \phi_u(j\omega) \frac{\partial G(j\omega)^*}{\partial \theta_m} \right) d\omega, \quad 1 \leq l, m \leq N, \quad (3)$$

where, $\phi_u(j\omega) = [\phi_{ij}(j\omega)]_{\{1 \leq i, j \leq r\}}$ and $\phi_e(j\omega) = \Sigma$.

For simplicity of analysis, it is assumed that Σ is the identity matrix. This assumption simplifies the expression of FIM to:

$$[\mathcal{M}_\theta]_{l,m} = \int_{-\infty}^{\infty} \text{Tr} \left(\frac{\partial G(j\omega)}{\partial \theta_l} \phi_u(j\omega) \frac{\partial G(j\omega)^*}{\partial \theta_m} \right) d\omega, \quad 1 \leq l, m \leq N. \quad (4)$$

Since the FIM depends on the input power spectrum, it is obvious that the covariance matrix of the estimated system would depend on the input chosen for exciting the system. Therefore, the goal is to determine a power spectrum $\phi_u(j\omega)$ which is optimal with respect one of the following commonly used functionals defined as :

- \mathcal{D} -Optimality which maximizes the determinant of \mathcal{M}_θ (Sanchez et al. [2012], Kumar et al. [2015]),
- \mathcal{A} -Optimality which minimizes the trace of \mathcal{M}_θ^{-1} (Aoki and Staley [1970], Nahi and DE Wallis [1969]),
- \mathcal{T} -Optimality which maximizes the trace of \mathcal{M}_θ (Srivastava and Anderson [1974]), and

<p>Given $G(s; \theta)$ and \mathcal{M}_θ defined as in (15), find ϕ^* such that</p> $\phi^* = \arg \max_{\phi_u} \log(\det(\mathcal{M}_\theta)) \text{ or } -\text{Tr}(\mathcal{M}_\theta^{-1}) \text{ or } \sigma_{\min}(\mathcal{M}_\theta) \text{ or } \text{Tr}(\mathcal{M}_\theta)$ <p>subject to</p> $\int_{-1}^1 \text{Tr}(\phi_u(j\omega)) d\omega \leq K_u$ $\int_{-1}^1 \text{Tr}(G(j\omega)\phi_u(j\omega)G(j\omega)^*) d\omega \leq K_y.$ <p style="text-align: right;">(6)</p> <p>Generate a signal $u(t) \in \mathcal{R}^{r \times 1}$ with a spectral density equal to ϕ_u^*.</p>

Figure 2: The optimal input design problem.

- \mathcal{E} -Optimality which maximizes the minimum eigenvalue of \mathcal{M}_θ (Cheng [1980], Constantine [1981]).

It is also natural to consider power constraints on the input signals and the output signals. The input power constraint may arise due to limitations of the excitation system. If the linear system has peak responses at some frequencies, the output power constraint would limit possible damages during excitation. The expressions for the input and the output power constraints are given by:

$$\int_{-\infty}^{\infty} \text{Tr}(\phi_u(j\omega)) d\omega \leq K_u \quad \text{and} \quad \int_{-\infty}^{\infty} \text{Tr}(G(j\omega)\phi_u(j\omega)G(j\omega)^*) d\omega \leq K_y, \quad (5)$$

where K_u and K_y are specified. It is assumed that the elements of the transfer function matrix are strictly proper. This assumption holds good in many practical systems and allows for the integral limits to be limited to a band of frequencies around zero. With the aforementioned definitions and constraints, the mathematical problem at hand is outlined in Figure 2.

Optimal input design for linear system identification has been studied for a long time and is still an active area of research (see Mehra [1974], Annergren et al. [2017]). Although not comprehensive, this paragraph would touch upon some of the seminal pieces of work in this area and discuss in brief, the methodologies used therein. In each of Barenthin et al. [2008], Kumar et al. [2015], the power spectrum is parameterized as

$$\phi_u(\omega) = \sum_{k=-M}^M c_k e^{-j\omega k}. \quad (7)$$

The optimal input design problem is cast as a convex optimization problem in terms of the covariance matrices $\{c_k\}_{k=0}^M$. In particular, the Kalman-Yakubovich-Popov lemma and the Hamburger's moment condition aid the formulation. The former ensures the positive definiteness of the power spectrum while the latter guarantees the existence of a valid solution. The idea of optimal input design has many applications. Ljung et al. [2011] is an excellent article which discusses the applications of system identification in communication systems, sensor networks and machine learning algorithms. In addition, automotive systems and chemical engineering processes also find the need for system identification and optimal input design (Alberer et al. [2011], Kumar et al. [2015]).

The way this paper differs from the earlier references is that the spectral density matrix is parameterized in terms of $H(j\omega)$, such that $\phi_u(j\omega) = \frac{1}{2}H(j\omega)H^*(j\omega)$. The matrix $H(j\omega)$ is expressed as a linear combination of a suitable truncated basis of $\mathcal{L}^2(-\omega_{\text{cut-off}}, \omega_{\text{cut-off}})$ and the coefficients become the optimization variables. Observe that this idea readily enforces positivity of $\phi_u(j\omega)$, which otherwise requires the Fourier basis and the Kalman-Yakubovich-Popov lemma. This also implies that the proposed method is independent of the

choice of the basis. Indeed, the resulting optimization problem in this paper is non-convex in general. The optimization problem has an interesting structure which has been studied in the literature and has wide applications in signal processing and control systems. Moreover, solving this class of optimization problems still remains open.

The paper is organized as follows. The second section sets up the optimization problem, presents methods to compute lower and upper bounds, and discusses aspects related to the convexity of the problem. The third section presents simulation results, *i.e.*, the input and the output waveforms, and the estimated parameters of the system. The concluding remarks are presented in the last section followed by references. The appendices to the paper present a couple of theorems and discuss the methodology of generating inputs from the optimal power spectrum.

2 The proposed method

This section discusses the formulation of the optimization problem, the non-convex nature of it and a methodology for obtaining a sub-optimal solution. As mentioned earlier, the input power spectrum is limited to a pre-defined band given by $D = [-\omega_c, \omega_c]$. Now, each element of $\phi_u(j\omega)$ must belong to the space $C^\infty(D)$. Spectral decomposition theorem states that one can find $H(j\omega)$ belonging to $C^1(D)$ such that $\phi_u(j\omega) = H(j\omega)H^*(j\omega)$. With the knowledge that $C^1(D)$ is dense in $\mathcal{L}^2(D)$, one can as well approximate elements $H(j\omega)$ through a linear combination of a truncated basis of $\mathcal{L}^2(D)$. Therefore, each element of $H(j\omega)$ is assumed to the space of square integrable functions over D denoted by $\mathcal{L}^2(D)$ (see Nair [2001]).

2.1 Setting up the optimization problem

As mentioned in the beginning of this section, it is assumed that each element of $H(j\omega)$ belongs to $\mathcal{L}^2(D)$. Thus, $H(j\omega)$ can be approximated with as a linear combination of a truncated basis for $\mathcal{L}^2(D)$. Suppose the truncated basis elements are given by the set $\{w_1, w_2, \dots, w_m\}$. Then $H(j\omega)$ is expressed as:

$$\begin{aligned} H_m(j\omega) &= W\mathcal{H} \\ \text{where, } W &= \begin{bmatrix} I_{r \times r} & w_1 I_{r \times r} & \dots & w_m I_{r \times r} \end{bmatrix} \\ \text{and } \mathcal{H} &= \begin{bmatrix} H_0 & H_1 & \dots & H_m \end{bmatrix}^\top, \quad \forall i \ H_i \in \mathcal{R}^{r \times r}. \end{aligned} \quad (8)$$

The matrix \mathcal{H} comprises of the coefficients which need to be determined by solving an optimization problem. As a consequence, the input power spectral density can be expressed as:

$$\phi_u^m(j\omega) = \frac{1}{2} W(j\omega) \mathcal{H} \mathcal{H}^\top W(j\omega)^*. \quad (9)$$

Using (9), the constraint on the total input power is given by:

$$\int_{-\infty}^{\infty} \text{Tr} \left(\frac{1}{2} W(j\omega) \mathcal{H} \mathcal{H}^\top W(j\omega)^* \right) d\omega \leq K_u. \quad (10)$$

Interchanging the integral operator and the trace operator, one obtains

$$\begin{aligned} \int_{-\infty}^{\infty} \text{Tr} \left(\frac{1}{2} W(j\omega) \mathcal{H} \mathcal{H}^\top W(j\omega)^* \right) d\omega &= \text{Tr} \left(\mathcal{H}^\top \left(\int_{-1}^1 \frac{1}{2} W(j\omega)^* W(j\omega) d\omega \right) \mathcal{H} \right) \\ &= \text{Tr} \left(\mathcal{H}^\top \mathcal{W} \mathcal{H} \right) \\ &= h^\top P_I h, \end{aligned} \quad (11)$$

$$\text{where, } h = \text{vec}(\mathcal{H}) \quad \text{and} \quad P_I = \begin{bmatrix} \mathcal{W} & & \\ & \ddots & \\ & & \mathcal{W} \end{bmatrix}.$$

Since \mathcal{W} is a positive definite matrix by (24), P_I is also positive definite. In short, the total input power constraint given by (10) is equivalent to

$$h^\top P_I h \leq K_u. \quad (12)$$

Note that $h \in \mathcal{R}^{mr^2}$. Similarly the total output power constraint given by

$$\int_{-\infty}^{\infty} \text{Tr} (G(j\omega)W(j\omega)\mathcal{H}\mathcal{H}^\top W(j\omega)^*G(j\omega)^*) d\omega \leq K_y - p. \quad (13)$$

can be rewritten as

$$P_O = \begin{bmatrix} \mathcal{W}_G & & \\ & \ddots & \\ & & \mathcal{W}_G \end{bmatrix} \quad \text{and} \quad \mathcal{W}_G = \frac{1}{2} \int_{-1}^1 W(j\omega)^*G(j\omega)^*G(j\omega)W(j\omega)d\omega. \quad (14)$$

$h^\top P_O h \leq K_y$, where

Note that, as in (24), it can be shown that \mathcal{W}_G is positive definite and so is P_O . Finally, the FIM \mathcal{M}_θ can be rewritten as

$$[\mathcal{M}_\theta]_{l,m} = \int_{-\infty}^{\infty} \text{Tr} \left(\frac{\partial G(j\omega)}{\partial \theta_l} \phi_u(j\omega) \frac{\partial G(j\omega)}{\partial \theta_m} \right) d\omega = \text{Tr} \left(\left(\mathcal{H}^\top \tilde{G}_l \right) \left(\mathcal{H}^\top \tilde{G}_m \right)^\top \right), \quad (15)$$

where $\tilde{G}_l = \int_{-\infty}^{\infty} \frac{\partial G(j\omega)}{\partial \theta_l} W$. In other words, every element of the FIM is a quadratic in the coefficients h . Now assuming $\Phi(\cdot) = \log\det(\cdot)$ or $\sigma_{\min}(\cdot)$ or $-\text{trace-inv}(\cdot)$ or $\text{trace}(\cdot)$, the optimization problem (6) is equivalent to

$$\begin{aligned} & \max_{h \in \mathcal{R}^{mr^2}} \Phi(\mathcal{M}_\theta(h)) \\ & \text{subject to} \\ & \quad h^\top P_I h \leq K_u \\ & \quad h^\top P_O h \leq K_y. \end{aligned} \quad (16)$$

2.2 Finding a suboptimal solution

To the best of the authors' knowledge, the optimization problem (16) does not have an efficient algorithm. The problem, due to its quadratic structure, renders itself to bi-linear forms, as well as semi-definite convex relaxations. The method proposed here depends on these derived forms and has two steps in it: (i) find a lower bound to (16) and (ii) find an upper bound to (16). The first step, which obtains a lower bound to the optimization problem through a bi-linear form, is iterative and in each iteration, a convex problem is solved. The procedure is asymptotically convergent and produces a feasible solution. The second step find an upper bound to the optimal value of (16) through a convex relaxation. Hence, it can be computed efficiently with semi-definite programming. The solution corresponding to the lower bound is a used as a sub-optimal solution to (16). The sub-optimality of the solution is determined by the difference between the upper and lower bounds.

Lower bound: To obtain a lower bound, the quadratic form in each of the elements of \mathcal{M}_θ is converted into a bi-linear form. The input and output power constraints, which are convex, are retained as they are. In other words,

$$x^\top M_{i,j} x \text{ is converted to } x^\top M_{i,j} y \text{ for all } i, j. \quad (17)$$

The bi-linear optimization problem is described, for example, in Konno [1976]. Note that once one of the variables (x or y) is fixed to some vector, (16) becomes a convex optimization problem in the other variable.

```

y = y0

while(count ≤ 100 || tol ≤ 1e-6) do

    max_{x ∈ ℝ^N} Φ(M(x, y)), where M_{ij}(x, y) = x^T M_{i,j} y and M_{i,j} ∈ ℝ^{N × N}
    & Φ(.) = logdet(.)|σ_min(.)|-trace-inv(.)|trace(.)

    subject to

        x^T P_I x ≤ 1
        x^T P_O x ≤ 1

    tol = ||x - y||_2
    y = x

end

```

(18)

Figure 3: Lower bound algorithm: y_0 is a random initial vector, the maximum iteration count is set to 100 and the tolerance bound for convergence is set to 1e-6. $\mathcal{M}(x, y)$ represents the bi-linear form of the Fisher Information Matrix, P_I represents the total input power matrix and P_O , the total output power matrix. The functional Φ can be chosen either of the following functions: $\log\det(\cdot)$ or $\sigma_{\min}(\cdot)$ or $-\text{trace-inv}(\cdot)$ or $\text{trace}(\cdot)$.

Starting from an arbitrary feasible point $y = y_0$, the convex optimization problem in x is solved to obtain x^* . For the next iteration, y is fixed to the vector the x^* . The resulting vector of each iteration acts as a lower bound to (16) since it is a feasible solution. Note that with each iteration, the value of the cost function can only increase and hence, the lower bound becomes tighter. The process asymptotically converge to a value but however, a discussion on the rate of convergence is out of the scope of this paper. The algorithm, also termed as the ‘hill-climbing’ procedure, is outlined in Figure 3.

Upper bound: The upper bound is determined by a standard semi-definite convex relaxation of (16). It can be shown that any homogeneous quadratic $x'Ax$ can also be written as (see Boyd and Vandenberghe [2004])

$$x'Ax = \text{Tr}(AX), \quad \text{where } X = xx^\top. \quad (19)$$

This transformation implies that (16) is equivalent to

$$\begin{aligned}
& \max_{X \in \mathbb{R}^{N \times N}} \Phi(\mathcal{M}_\theta(X)), \quad \text{where } [\mathcal{M}_\theta]_{ij}(X) = \text{Tr}(A^{ij}X) \\
& \text{subject to} \\
& \quad \text{trace}(P_I X) \leq 1 \\
& \quad \text{trace}(P_O X) \leq 1 \\
& \quad X \succeq 0 \quad \text{and} \quad \text{Rank}(X) = 1
\end{aligned} \quad (20)$$

Dropping the Rank-1 in (20) results in a convex semi-definite optimization problem shown in Figure 4. Being a relaxation of the original optimization problem, the solution to (20) gives an upper bound to (16). Moreover, the bound can be obtained efficiently as the associated problem is convex. However, unlike the lower bound algorithm, the solution matrix X^* need not be have a rank of 1, in which case deriving a feasible solution for (16) from it may not be possible.

Remark: It might be necessary to have more power constraints in the optimization problem. In particular, it might be necessary to constrain input and output powers at the individual ports, in addition to the total power constraints. Moreover, the power constraints may also be non-convex, particularly when the constraint is to deliver a minimum amount of input or output power. In all the cases mentioned here, the algorithm for the upper bound does not change, except for the increase in the number of constraints. The

$\begin{aligned} & \max_{X \in \mathcal{R}^{N \times N}} \Phi(\mathcal{M}(X)), \text{ where } \mathcal{M}_{ij}(X) = \text{trace}(M_{i,j}X), \quad M_{i,j} \in \mathcal{R}^{N \times N} \\ & \quad \quad \quad \& \Phi(\cdot) = \log\det(\cdot) \sigma_{\min}(\cdot) \text{trace-inv}(\cdot) \text{trace}(\cdot) \\ \text{subject to} & \\ & \quad \text{Tr}(P_I X) \leq 1 \\ & \quad \text{Tr}(P_O X) \leq 1 \\ & \quad X \succeq 0 \end{aligned}$	(21)
--	------

Figure 4: Upper bound algorithm: $\mathcal{M}(X)$ represents the relaxed form of the Fisher Information Matrix, P_I represents the total input power matrix and P_O , the total output power matrix. The functional Φ can be chosen either of the following functions: $\log\det(\cdot)$ or $\sigma_{\min}(\cdot)$ or $\text{-trace-inv}(\cdot)$ or $\text{trace}(\cdot)$.

lower bound algorithm would need a few changes. Firstly, the constraints are also changed to their corresponding bi-linear forms. Secondly, the step “ $y = x$ ” in Figure 3 needs to be replaced by “ $y = \frac{x_k + x_{k-1}}{2}$ ”. In other words, y is replaced by the average of the resulting vectors of the last two iterates.

2.3 A discussion on quadratic maps

This subsection discusses the convexity properties of the optimization problem given in (16). As mentioned earlier, every element of the FIM $\mathcal{M}_\theta(h)$ is a homogeneous quadratic in h . For clarity, the optimization problem is written in an expanded form as:

$$\begin{aligned} & \max_{h \in \mathcal{R}^{mr^2}} \Phi \left(\begin{bmatrix} h^\top M_{1,1}h & \dots & h^\top M_{1,N}h \\ \vdots & & \vdots \\ h^\top M_{N,1}h & \dots & h^\top M_{N,N}h \end{bmatrix} \right) \\ \text{subject to} & \\ & \quad h^\top P_I h \leq K_u \\ & \quad h^\top P_O h \leq K_y, \end{aligned} \tag{22}$$

where $M_{i,j}$'s are symmetric matrix of dimension $mr \times mr$. Moreover, by construction, the symmetric matrices are such that the FIM in the argument of Φ is positive semidefinite for all $h \in \mathcal{R}^{mr}$. The convexity of this optimization problem depends on the convexity of the image of \mathcal{R}^{mr} under the quadratic map $Q : \mathcal{R}^{mr^2} \rightarrow \mathcal{R}^{\frac{N(N+1)}{2}+2}$ given by:

$$Q(x) = [x^\top M_{1,1}x \dots x^\top M_{i,j}x \dots x^\top M_{N,N}x \quad x^\top P_I x \quad x^\top P_O x]^\top; \quad i \leq j. \tag{23}$$

It is assumed here that the total number of coefficients given by mr is much larger than the number of parameters to be estimated given by N . Quadratic maps and the convexity of its image have been studied earlier in literature and some of the important results are pointed out here.

Ramana and Goldman [1994] were one of the first few to study, in detail, the convexity of images under quadratic transformations. They give necessary and sufficient condition for characterizing such maps. However, the condition are not checkable for general cases. Polyak [1998] proved that the image of the unit disk under any “definite” quadratic map $Q : \mathcal{R}^n \rightarrow \mathcal{R}^3$ is convex for $n \geq 3$. Sheriff [2013] presented results for “stable convexity” and discussed its applications in analyzing coupled systems. Dymarsky [2014] showed the convexity of the image of an open ball in \mathcal{R}^n under quadratic maps. Since the dimension of the range space in this paper is much larger than 3, these results cannot be applied here. Beck [2009] presented some startling generalizations of earlier results on the subject. One of the results can be applied to obtain a convex relaxation of the optimization problem in this paper, but only at the cost of a large number of redundant variables. Hiriart-Urruty and Torki [2002] provided an excellent overview on the convexity aspects of the images of quadratic maps in their paper. These references point out that quadratic maps with convex images are rather exceptions than the rule.

2.4 Input generation and system identification

The methodology for generating inputs real inputs with the optimal power spectrum obtained from the optimization procedure is outlined in Appendix B. The methodology essentially discretizes the power spectrum, concatenates it with a random phase and performs inverse Fourier transform to obtain the inputs. The system is excited with the the designed optimal inputs and the outputs are recorded. The system identification is done using the Instrument Variables method, which is an unbiased estimator of the system parameters. For details on the system identification algorithm, the reader is directed to Söderström and Stoica [2002].

3 Simulations

In most practical applications, a system is fabricated with a desired/nominal set of parameters. But fabrication is seldom perfect and therefore, the actual parameters vary to a small extent around the desired locations. For optimal input design, the nominal system parameters are used for generating the optimal inputs, while the system with actual parameters generates the outputs when actuated with the optimal inputs. For simulations, a 2-input and 2-output system was chosen. The nominal transfer function matrix and the actual transfer function matrix are given by:

$$\underbrace{\begin{bmatrix} \frac{3.2}{s^2 + 1.6s + 1.5} & \frac{3.1}{s^2 + 1.9s + 1} \\ \frac{5.2}{s^2 + 0.5s + 0.95} & \frac{1.5}{s^2 + 0.9s + 0.3} \end{bmatrix}}_G, \quad \underbrace{\begin{bmatrix} \frac{3.125}{s^2 + 1.5s + 1.563} & \frac{3.188}{s^2 + 2s + 1.063} \\ \frac{5.313}{s^2 + 0.5s + 1.063} & \frac{1.563}{s^2 + s + 0.312} \end{bmatrix}}_{G_{nom}}.$$

The total input power and the total output power were constrained to a maximum of 1000 units. Chebyshev's polynomials up to the 13th order were considered for parameterizing $H(j\omega)$. It was observed that lower bounds and the upper bounds matched in the \mathcal{D} , \mathcal{A} and \mathcal{E} optimality criterion, while they were off by less than 0.1% in case of \mathcal{T} optimality. It has been observed that the lower and upper bounds match in several instances. However, there is no mathematical proof for the observation yet. The optimal power spectrums obtained using the proposed method have been depicted in Figure 5. The inputs generated according to the optimal spectrums corresponding to the four different optimality criterion are shown in Figures 6, 7, 8. In addition, the corresponding outputs, corrupted with uncorrelated additive Gaussian white noise vectors, are also shown in Figures 6, 7, 8. The standard deviation of the measurement noise is assumed to be equal to one-tenth of the minimum of the root-mean-squared values of the outputs. The system identification was done using Instrument Variables method and the results corresponding to the different optimality criteria are given below.

$$\underbrace{\begin{bmatrix} \frac{3.21}{s^2 + 1.603s + 1.497} & \frac{3.105}{s^2 + 1.897s + 0.998} \\ \frac{5.073}{s^2 + 0.492s + 0.951} & \frac{1.387}{s^2 + 0.856s + 0.284} \end{bmatrix}}_{G_{\mathcal{D}}}, \quad \underbrace{\begin{bmatrix} \frac{3.555}{s^2 + 1.639s + 1.501} & \frac{2.742}{s^2 + 1.863s + 0.961} \\ \frac{5.252}{s^2 + 0.502s + 0.952} & \frac{1.458}{s^2 + 0.884s + 0.295} \end{bmatrix}}_{G_{\mathcal{A}}},$$

$$\underbrace{\begin{bmatrix} \frac{3.197}{s^2 + 1.606s + 1.506} & \frac{3.082}{s^2 + 1.920s + 0.999} \\ \frac{5.172}{s^2 + 0.498s + 0.951} & \frac{1.425}{s^2 + 0.873s + 0.292} \end{bmatrix}}_{G_{\mathcal{T}}}, \quad \underbrace{\begin{bmatrix} \frac{2.787}{s^2 + 1.502s + 1.488} & \frac{3.144}{s^2 + 1.905s + 1.006} \\ \frac{5.308}{s^2 + 0.499s + 0.952} & \frac{1.473}{s^2 + 0.887s + 0.296} \end{bmatrix}}_{G_{\mathcal{E}}}.$$

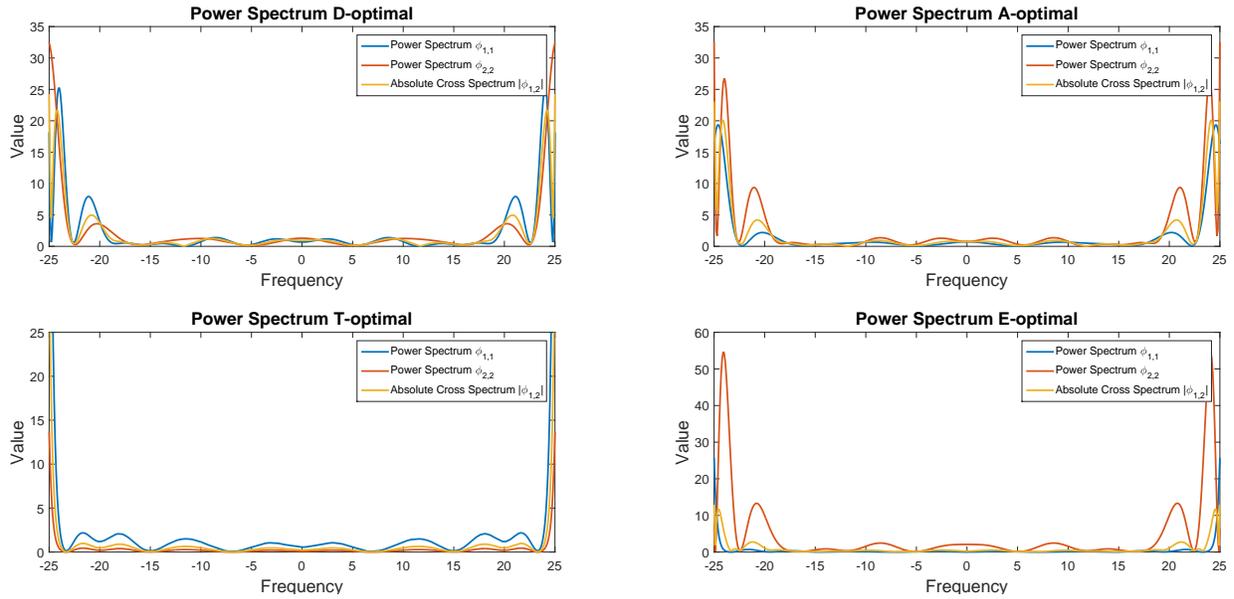


Figure 5: This figure shows the optimal power spectrum obtained using the proposed method for the four different optimality criterion. The calculations were done by parameterizing $H(j\omega)$ with Chebyshev's polynomials upto the 13th order. The absolute value of the cross-spectrum is also plotted.

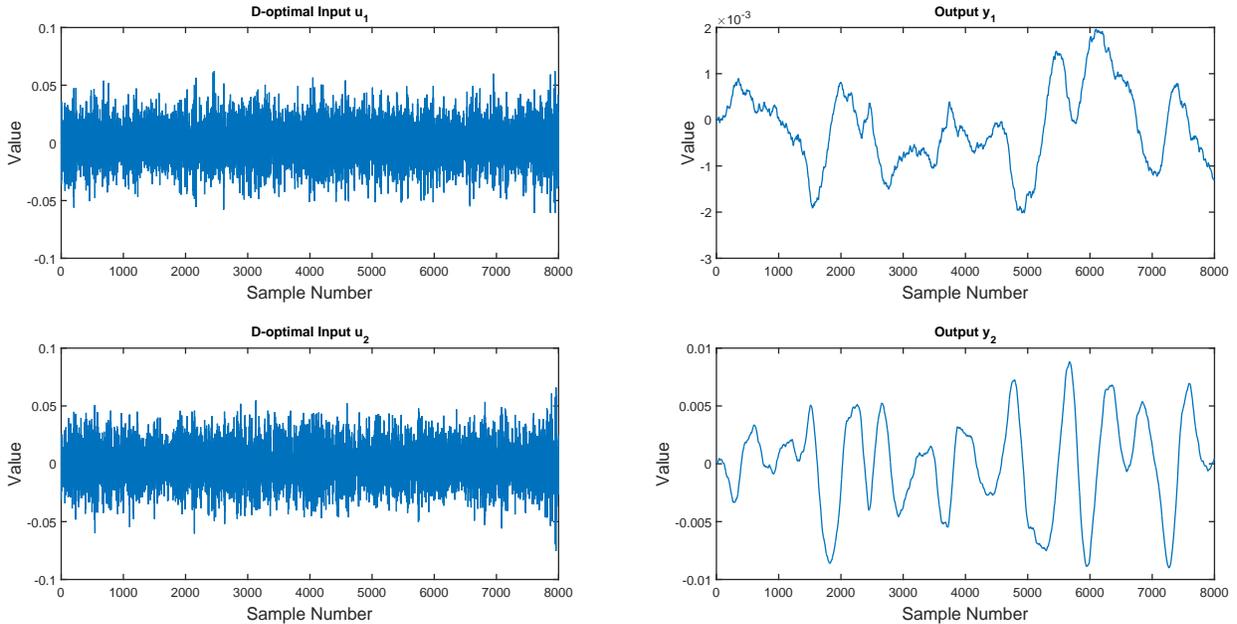


Figure 6: The inputs corresponding to the D -optimal spectrum are shown on the left. The outputs obtained by exciting the system with the inputs are shown to the right.

4 Conclusion

In this paper, a distinct paradigm to design a band-limited optimal inputs for identifying a linear continuous time MIMO system was presented. Unlike conventional schemes where the power spectral density is

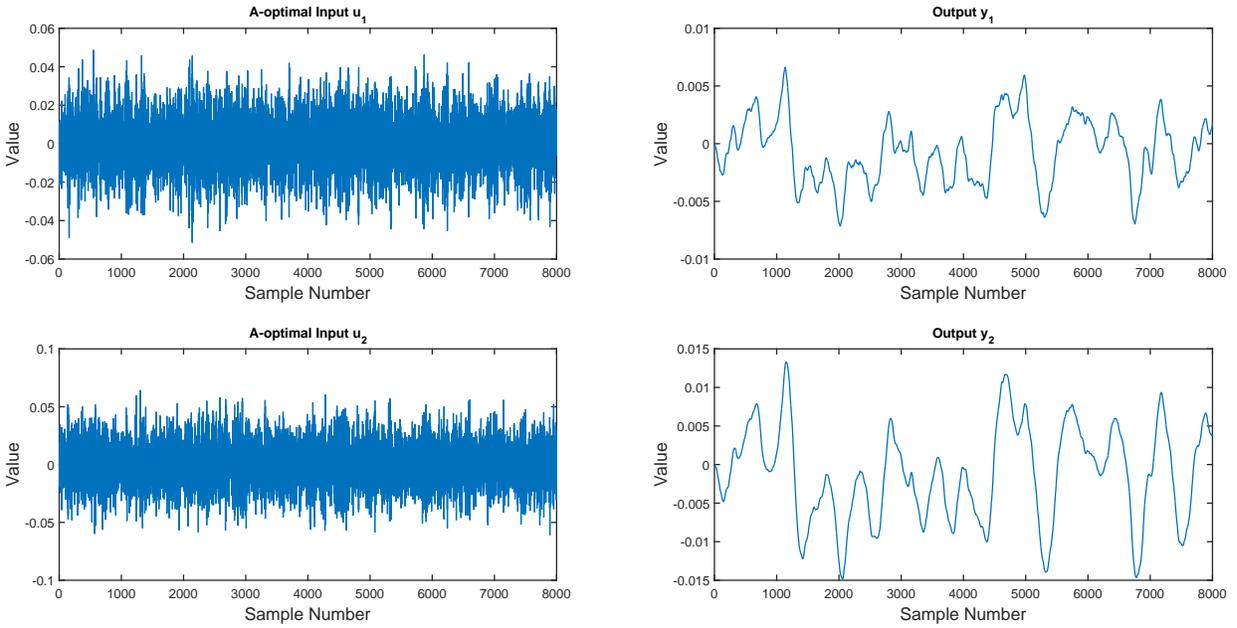


Figure 7: The inputs corresponding to the \mathcal{A} -optimal spectrum are shown on the left. The outputs obtained by exciting the system with the inputs are shown to the right.

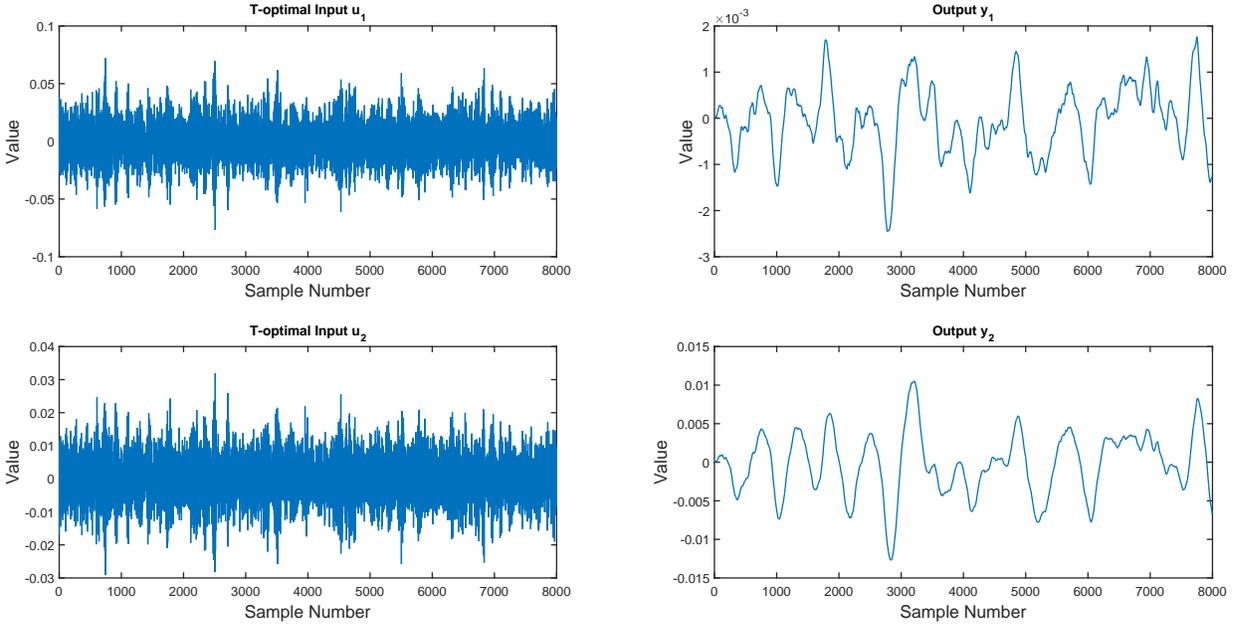


Figure 8: The inputs corresponding to the \mathcal{T} -optimal spectrum are shown on the left. The outputs obtained by exciting the system with the inputs are shown to the right.

parametrized with covariance matrices, this paper used spectral decomposition. The input spectra $\phi_u(j\omega)$ was factored as $\frac{1}{2}H(j\omega)H^*(j\omega)$ and $H(j\omega)$ was approximated by a truncated basis for $\mathcal{L}^2([- \omega_c, \omega_c])$, where ω_c is the cut-off frequency. The coefficients of the finite linear combination thus became the optimization

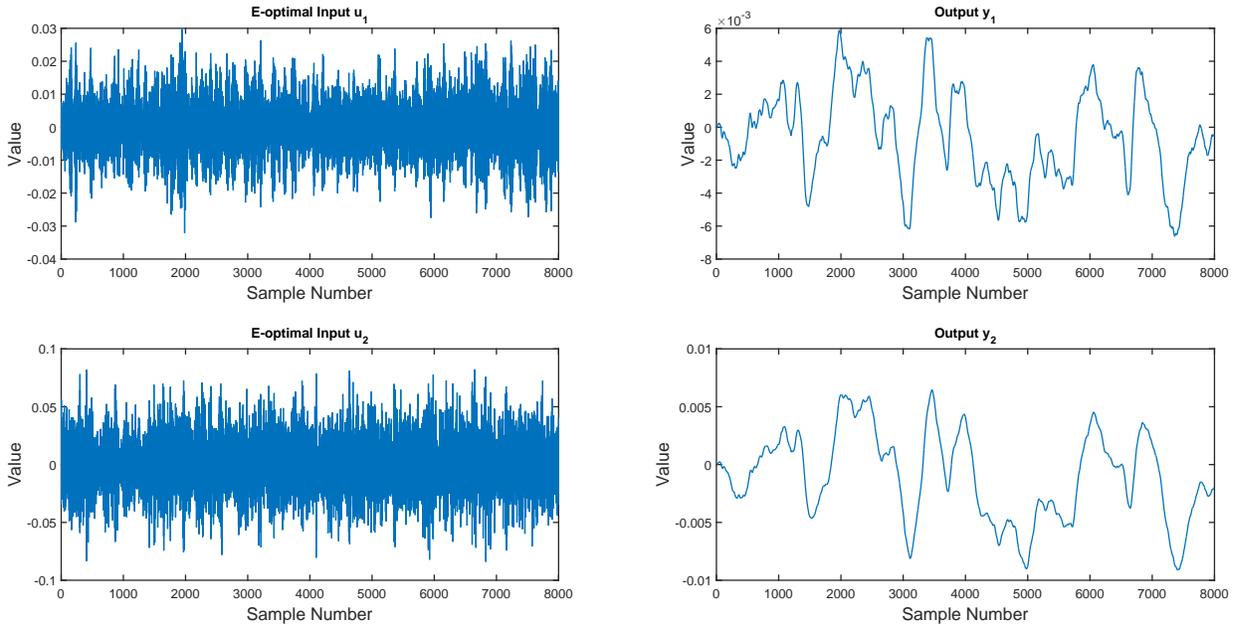


Figure 9: The inputs corresponding to the \mathcal{E} -optimal spectrum are shown on the left. The outputs obtained by exciting the system with the inputs are shown to the right.

parameters. The parameterization ensured the positivity of the power spectral density matrix, irrespective of the basis chosen for $\mathcal{L}^2(\mathcal{D})$. The optimization problem of determining the optimal coefficients for \mathcal{D} , \mathcal{A} , \mathcal{T} or \mathcal{E} optimality criteria turned out to be a non-standard optimization problem. A discussion on the convexity aspect of the problem was also presented. A lower bound for the optimal value was obtained by first converting the optimization problem into a bi-linear form and then solving it using an iterative procedure. In each of the iterations, a convex problem was solved which made the algorithm computationally efficient. An upper bound for the optimal value was obtained by first converting the optimization problem into a rank constrained semi-definite problem. The Rank-1 constraint was then dropped and the resulting convex semi-definite program was solved using standard routines. If the upper bound turned out to be a Rank-1 solution, the vector defining its column space was chosen as the optimal solution; else the solution corresponding to the lower bound was chosen. The sub-optimality of the final solution was determined by the difference in the bounds. Interestingly, the bounds matched in many instances of simulations. With the optimal spectral density, the corresponding inputs were generated. The linear MIMO system was stimulated with these inputs and the outputs, corrupted with measurement noise, were recorded. The Instrument Variables method was used to estimate the system parameters. The estimated system was observed to be in close agreement with the original system.

As for future directions of work, several interesting questions can be posed. The authors would like to point to a few: (i) The choice of basis for $\mathcal{L}^2(\mathcal{D})$ and its effect on the obtained optimality, and (ii) the ratio of the lower bound to the upper bound, in general.

References

- D. Alberer, H. Hjalmarsson, and L. Del Re. *Identification for automotive systems*, volume 418. Springer Science & Business Media, 2011.
- M. Annergren, C. A. Larsson, H. Hjalmarsson, X. Bombois, and B. Wahlberg. Application-oriented input

-
- design in system identification: Optimal input design for control [applications of control]. *IEEE Control Systems*, 37(2):31–56, 2017.
- M. Aoki and R. Staley. On input signal synthesis in parameter identification. *Automatica*, 6(3):431–440, 1970.
- M. Barentin, X. Bombois, H. Hjalmarsson, and G. Scorletti. Identification for control of multivariable systems: Controller validation and experiment design via lmis. *Automatica*, 44(12):3070–3078, 2008.
- A. Beck. Convexity properties associated with nonconvex quadratic matrix functions and applications to quadratic programming. *Journal of Optimization Theory and Applications*, 142(1):1–29, 2009.
- S. Boyd and L. Vandenberghe. *Convex optimization*. Cambridge university press, 2004.
- C.-S. Cheng. On the e-optimality of some block designs. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 199–204, 1980.
- G. M. Constantine. Some e-optimal block designs. *The Annals of Statistics*, pages 886–892, 1981.
- A. Dymarsky. On the convexity of image of a multidimensional quadratic map. *arXiv preprint arXiv:1410.2254*, 2014.
- J.-B. Hiriart-Urruty and M. Torchi. Permanently going back and forth between the “quadratic world” and the “convexity world” in optimization. *Applied Mathematics & Optimization*, 45(2):169–184, 2002.
- H. Konno. Maximization of a convex quadratic function under linear constraints. *Mathematical programming*, 11(1):117–127, 1976.
- A. Kumar, M. Nabil, and S. Narasimhan. Economical input design for identification of multivariate systems. *IFAC-PapersOnLine*, 48(28):1313–1318, 2015.
- L. Ljung, H. Hjalmarsson, and H. Ohlsson. Four encounters with system identification. *European Journal of Control*, 17(5-6):449–471, 2011.
- R. K. Mehra. Optimal input signals for parameter estimation in dynamic systems—survey and new results. *Automatic Control, IEEE Transactions on*, 19(6):753–768, 1974.
- N. Nahi and J. DE Wallis. Optimal inputs for parameter estimation in dynamic systems with white observation noise. 1969.
- M. T. Nair. *Functional analysis: A first course*. PHI Learning Pvt. Ltd., 2001.
- B. Polyak. Convexity of quadratic transformations and its use in control and optimization. *Journal of Optimization Theory and Applications*, 99(3):553–583, 1998.
- M. Ramana and A. Goldman. *Quadratic maps with convex images*. Rutgers University. Rutgers Center for Operations Research [RUTCOR], 1994.
- B. Sanchez, C. R. Rojas, G. Vandersteen, R. Bragos, and J. Schoukens. On the calculation of the d-optimal multisine excitation power spectrum for broadband impedance spectroscopy measurements. *Measurement Science and Technology*, 23(8):085702, 2012.
- J. L. Sheriff. *The convexity of quadratic maps and the controllability of coupled systems*. PhD thesis, 2013.
- T. Söderström and P. Stoica. Instrumental variable methods for system identification. *Circuits, Systems and Signal Processing*, 21(1):1–9, 2002.
- J. Srivastava and D. Anderson. A comparison of the determinant, maximum root, and trace optimality criteria. *Communications in Statistics-Theory and Methods*, 3(10):933–940, 1974.
-

A Theorems and proofs

Theorem. Suppose $\mathcal{W} = \frac{1}{2} \int_{-1}^1 W(j\omega)^* W(j\omega) d\omega$. Then \mathcal{W} is an $r(m+1) \times r(m+1)$ positive definite matrix.

Proof. It is clear that \mathcal{W} is positive semidefinite. Suppose for some $x \neq 0$

$$\begin{aligned} 0 = x^* \mathcal{W} x &= \int_0^1 x^* W(j\omega)^* W(j\omega) x d\omega, \\ &\Rightarrow x^* W(j\omega)^* W(j\omega) x = 0, \forall \omega \\ &\Rightarrow W(j\omega) x = 0, \forall \omega \\ &\Rightarrow x = 0, \text{ a contradiction.} \end{aligned}$$

□

Theorem. Suppose Φ is a matrix function defined on $[-\pi, \pi]$ such that (i) $\Phi(\theta) \in \mathbb{C}^{q \times q}$ is a Hermitian nonnegative definite matrix for all $\theta \in [-\pi, \pi]$, (ii) $\Phi(\theta) = \Phi^\top(-\theta)$ and (iii) Φ is integrable and admits a Fourier series. Then there exists a matrix function Ψ defined on $[-\pi, \pi]$ such that $\Phi(\theta) = \Psi(\theta)\Psi^*(\theta)$. Moreover, Ψ admits a Fourier series.

Proof. See Wilson [1972].

□

B Design of inputs from optimal power spectrum

An algorithm for generating a $u(t)$ from $H(j\omega)$, which is obtained from the proposed method, is discussed in this section. For brevity, the case for generating a 2×1 input is illustrated. The extension for generating a $r \times 1$ input is presented in Figure 10. Conventionally, the input is generated as a sample path of a stochastic process with the desired power spectral density. The same principle is adopted in this paper as well. To elaborate further, let

$$H(j\omega) = \begin{bmatrix} H_{11}(j\omega) & H_{12}(j\omega) \\ H_{21}(j\omega) & H_{22}(j\omega) \end{bmatrix} \quad \text{and} \quad \phi(j\omega) = \frac{1}{2} H(j\omega) H^*(j\omega) = \begin{bmatrix} \phi_{11}(j\omega) & \phi_{12}(j\omega) \\ \phi_{21}(j\omega) & \phi_{22}(j\omega) \end{bmatrix}. \quad (24)$$

Naturally,

$$\begin{aligned} \phi_{11}(j\omega) &= \frac{1}{2} \left(|H_{11}(j\omega)|^2 + |H_{12}(j\omega)|^2 \right), \quad \phi_{22}(j\omega) = \frac{1}{2} \left(|H_{21}(j\omega)|^2 + |H_{22}(j\omega)|^2 \right) \\ \phi_{12}(j\omega) &= \frac{1}{2} \left(H_{11}(j\omega) H_{21}^*(j\omega) + H_{12}(j\omega) H_{22}^*(j\omega) \right) = \phi_{21}(j\omega)^*. \end{aligned} \quad (25)$$

The goal is then to determine a real-valued random vector $u(t) = [u_1(t), u_2(t)]^\top$ such that

$$\mathcal{R}_u(\tau) = \mathbb{E} \{ u(t) u^*(t - \tau) \} = \frac{1}{2\pi} \int_{-\omega_c}^{\omega_c} \phi_u(j\omega) e^{j\omega\tau} d\omega, \quad (26)$$

where $\mathbb{E}\{\cdot\}$ denotes the expected value. To that end, consider the one sided frequency functions *i.e.* defined for $\omega \geq 0$,

$$U_1(j\omega) = \sqrt{\pi} (H_{11}(j\omega) e^{j\Phi_1^\omega} + H_{12}(j\omega) e^{j\Phi_2^\omega}) \quad \text{and} \quad U_2(j\omega) = \sqrt{\pi} (H_{21}(j\omega) e^{j\Phi_1^\omega} + H_{22}(j\omega) e^{j\Phi_2^\omega}), \quad (27)$$

where the families of random variables $\{\Phi_1^\omega, \omega \geq 0\}$ and $\{\Phi_2^\omega, \omega \geq 0\}$ are such that

$$\begin{aligned} \mathbb{E}\{e^{j\Phi_i^\omega}\} &= 0, \quad \forall \omega \geq 0, \quad i = 1, 2 \quad \text{and} \quad \mathbb{E}\{e^{j\Phi_1^{\omega_1}} e^{-j\Phi_2^{\omega_2}}\} = 0, \quad \forall \omega_1, \omega_2 \geq 0 \\ \mathbb{E}\{e^{j\Phi_i^{\omega_1}} e^{-j\Phi_i^{\omega_2}}\} &= \delta(\omega_1 - \omega_2), \quad \forall \omega_1, \omega_2 \geq 0, \quad i = 1, 2. \end{aligned} \quad (28)$$

For the input signals have to be real-valued, extend U_1 and U_2 to the negative imaginary axis as:

$$U_1(-j\omega) = U_1^*(j\omega) \quad \text{and} \quad U_2(-j\omega) = U_2^*(j\omega). \quad (29)$$

Define the inputs as:

$$u_1(t) = \frac{1}{2\pi} \int_{-\omega_c}^{\omega_c} U_1(j\omega) e^{j\omega t} d\omega \quad \text{and} \quad u_2(t) = \frac{1}{2\pi} \int_{-\omega_c}^{\omega_c} U_2(j\omega) e^{j\omega t} d\omega. \quad (30)$$

The claim is that the power spectral density of $[u_1(t), u_2(t)]$ is equal to $\phi_u(j\omega)$. For proving the claim, note that by using Fubini's Theorem Nair [2001],

$$\begin{aligned} \mathcal{R}_{11}(\tau) &= \mathbb{E}\{u_1(t)u_1^*(t-\tau)\} \\ &= \mathbb{E}\left\{\frac{1}{2\pi} \int_{-\omega_c}^{\omega_c} \frac{1}{2\pi} \int_{-\omega_c}^{\omega_c} U_1(j\omega_1)U_1^*(j\omega_2) e^{j\omega_1 t} e^{-j\omega_2(t-\tau)} d\omega_1 d\omega_2\right\} \\ &= \frac{1}{2\pi} \int_{-\omega_c}^{\omega_c} \frac{1}{2\pi} \int_{-\omega_c}^{\omega_c} \mathbb{E}\{U_1(j\omega_1)U_1^*(j\omega_2)\} e^{j\omega_1 t} e^{-j\omega_2(t-\tau)} d\omega_1 d\omega_2 \end{aligned} \quad (31)$$

From (27) and (28), it can be deduced that

$$\mathbb{E}\{U_1(j\omega_1)U_1^*(j\omega_2)\} = \pi \left(|H_{11}(j\omega_1)|^2 + |H_{12}(j\omega_2)|^2 \right) \delta(\omega_1 - \omega_2). \quad (32)$$

Therefore,

$$\mathcal{R}_{11}(\tau) = \frac{1}{2\pi} \int_{-\omega_c}^{\omega_c} \frac{1}{2} \left(|H_{11}(j\omega)|^2 + |H_{12}(j\omega)|^2 \right) e^{j\omega\tau} d\omega. \quad (33)$$

On similar lines, it follows that

$$\begin{aligned} \mathcal{R}_{22}(\tau) &= \frac{1}{2\pi} \int_{-\omega_c}^{\omega_c} \frac{1}{2} \left(|H_{21}(j\omega)|^2 + |H_{22}(j\omega)|^2 \right) e^{j\omega\tau} d\omega \quad \text{and} \\ \mathcal{R}_{12}(\tau) &= \frac{1}{2\pi} \int_{-\omega_c}^{\omega_c} \frac{1}{2} \left(H_{11}(j\omega)H_{21}^*(j\omega) + H_{12}(j\omega)H_{22}^*(j\omega) \right) e^{j\omega\tau} d\omega. \end{aligned} \quad (34)$$

It is now easy to note from the relation (26) applied to (34) that the power spectral density of $[u_1(t), u_2(t)]^\top$ is equal to $\phi(j\omega) = H(j\omega)H^*(j\omega)$. Now, a closed form expression for the integrals in (30) is difficult to obtain and therefore it needs to be evaluated numerically by discretization of the frequency band. To this end, choose Φ_1^ω and Φ_2^ω to be independent Gaussian white noise vectors. Since $U_1(j\omega)$ is zero outside the interval $[-\omega_c, \omega_c]$,

$$u_1(t) = \frac{1}{2\pi} \int_{-\omega_c}^{\omega_c} U_1(j\omega) e^{j\omega t} d\omega. \quad (35)$$

Generate independent white random vectors Ψ_1, Ψ_2, \dots , and Ψ_r

Define $U_i \left(j \frac{k}{N} \right) = \sqrt{\pi} \sum_{m=1}^r H_m \left(j \frac{k}{N} \right) e^{j\Psi_m(k)}$, $\forall k = 1, \dots, N$, $\forall i = 1, \dots, r$.

Compute $\tilde{u}_i(t) = \frac{1}{2\pi N} \sum_{k=-N}^N U_i \left(j \frac{k}{N} \right) e^{j\frac{k}{N}t}$, $\forall k = 1, \dots, r$, $\forall i = 1, \dots, r$.

(37)

Figure 10: Input Design Methodology.

With a positive integer N such that $\frac{1}{N} \ll 1$, a Riemann sum approximation of (35) would be

$$\tilde{u}_1(t) = \frac{1}{2\pi N} \sum_{k=-N}^N U_1 \left(j \frac{k}{N} \right) e^{j\frac{k}{N}t}. \quad (36)$$

In a similar way, $\tilde{u}_2(t)$ can also be defined. The signal $\tilde{u}(t) = [\tilde{u}_1(t) \ \tilde{u}_2(t)]$ is then used for simulations and computations. For a general r -input system, the procedure has been shown in Figure 10.