Notes on Input Design: From Multi-Sine Design to Data-Driven Procedures

László Gerencsér, *Member, IEEE*, György Michaletzky, József Bokor, *Fellow, IEEE*, and Péter Polcz

***Abstract*—We show that a class of optimal input design problems have only discrete spectral measures as solutions. If we fix any finite set of possible frequencies then a randomized version of the resulting convex problem has a unique (sparse) solution with probability 1. We also propose a data-driven approach to optimal input design via virtual off-line estimators that coincide with the optimized PE estimator modulo a negligible error, both for open loop and closed loop systems.**

***Index Terms*—Closed loop systems, data-driven modeling, performance evaluation, system identification.**

I. Introduction

INPUT design for linear stochastic control systems is of fundamental importance in many industrial control systems, see [1] for an excellent survey. Remarkably, input design has become central also in machine learning, see [2]. Considering a family of linear stochastic control systems with input u, parameterized by θ, a central issue of input design is to minimize the expected loss in a performance index $J(\theta)$ due to uncertainty, incurred by replacing the true parameter by its estimate $\hat{\theta}_N$, subject to constraints on the input. A preliminary problem is to optimize the information matrix M, depending linearly on the spectral measure of u. It is readily seen that this is a convex problem with a unique solution, see (16) and Section II. Thus, the input design problem can be reduced to a generalized moment problem of finding a spectral measure $d\Phi^u(\cdot)$ such that it generates M^u via (13) below. This approach has been pursued in [3].

Manuscript received 8 March 2024; revised 11 May 2024; accepted 31 May 2024. Date of publication 18 June 2024; date of current version 17 July 2024. This work was supported in part by the Ministry of Innovation and Technology NRDI Office (National Research, Development and Innovation Office) within the Autonomous Systems National Laboratory Program, and in part by NRDI OTKA under Grant PD-145902. Recommended by Senior Editor T. Oomen.
(Corresponding author: László Gerencsér.)

László Gerencsér is with the Systems and Control Lab, HUN-REN Institute for Computer Science and Control, 1518 Budapest, Hungary (e-mail: gerencser.laszlo@sztaki.hun-ren.hu).

György Michaletzky is with the Faculty of Science, Eötvös Loránd University, 1117 Budapest, Hungary (e-mail: gyorgy.michaletzky@ttk.elte.hu).

József Bokor is with the Systems and Control Lab, HUN-REN Institute for Computer Science and Control, 1518 Budapest, Hungary (e-mail: bokor.jozsef@sztaki.hun-ren.hu).

Péter Polcz is with the Faculty of Information Technology and Bionics, Pázmány Péter Catholic University, 1088 Budapest, Hungary (e-mail: polcz.peter@itk.ppke.hu).
Digital Object Identifier 10.1109/LCSYS.2024.3416072

For practical reasons significant attention has been paid to finding sub-optimal solutions by restricting the search space to a convex set of feasible spectral measures with a density, depending linearly on a parameter η belonging to a compact, convex set in a Euclidean space. A benchmark example is the set of spectral measures of FIR (Finite Impulse Response) processes, with the feasible parameters being the coefficients of the half spectra, constrained by LMI-s (Linear Matrix Inequalities), as first introduced in [4].

In the case of closed loop systems, allowing a LTI (Linear Time Invariant) feedback K, a preliminary convex optimization problem over the pairs of spectral density and cross spectra, (Φ_u, Φ_{ue}), as variables can be formulated, see [5]. Elaborating this idea K itself can be treated as an additional design parameter. A smart way of describing the pairs (Φ_u, Φ_{ue}) is obtained by using the Youla-Kučera parameterization of K, leading to a linear parameterization, see [6] for an advanced exposition. A remarkable feature of closed loop identification is that feedback reduces the asymptotic covariance matrix of the estimate of the signal transfer function, see [7].

In this letter we revisit the problem of multi-sine input design, recently attracting renewed interest, see [8]. Our starting point is that the optimization problem can be redefined as a convex optimization problem over the space of spectral measures $d\Phi^u(.)$ with discrete spectrum subject to energy constraints. In turn, the latter can be approximated with arbitrary prescribed accuracy by a convex design problem over the space of spectral measures $d\Phi^u(.)$ with support on a fixed set of frequencies. We will establish the remarkable fact that the relaxed version of this problem, obtained by adding the energy constraint multiplied by a freely chosen Lagrange-multiplier, has a unique solution w.p.1 (with probability 1) when the weights are chosen randomly according to a probability law that has a density. This sub-optimal spectral measure can be readily realized by a multi-sine.

An additional issue to be considered is what was called the "Achilles' heel of optimal input design" in [1], namely the fact that the optimization problem depends on the true system. This paradox has actually been resolved in [9] for ARMAX systems exited by inputs generated by a FIR filter, with a full-scale technical analysis relying on advanced results on recursive estimation given in [10]. In Section IV we reformulate and extend the basic idea of the above paper. The key tool is the definition of a virtual off-line estimator, having a very accurate characterization. Although we can not compute it in practice, its on-line approximation, obtained along the lines of [11], [12], is computable leading to an adaptive input design method. This data-driven approach is presented first for open loop system. However, it readily generalizes to closed loop systems with a θ-dependent feedback loop, obtained for some optimal control problem, yielding the optimal spectral density for the external excitation. As far as we can see our approach is a digression from the main-stream literature, in which the transfer function of the feedback loop is a design variable for the input design problem itself.

II. Technical Preliminaries

To be specific, we consider a discrete time single input single output linear stochastic control system with input u, external noise e and output y, defined in the range $-\infty < n < +\infty$

$$y = H^u\left(\theta^*, q^{-1}\right)u + H^e\left(\theta^*, q^{-1}\right)e. \tag{1}$$

Here H^u and H^e are rational functions of the backward shift operator q^{-1} of fixed degrees, depending on a parameter θ. We assume $\theta \in D \in \mathbb{R}^p$, where D is an open domain. The true parameter will be denoted by θ^*. The associated transfer functions are obtained when replacing q^{-1} by $e^{-i\omega}$.

Condition 1: The transfer function $H^u(\theta)$ is stable, and $H^e(\theta)$ is stable and inverse stable for all $\theta \in D$. Moreover, they are three-times continuously differentiable functions of θ. In addition, we assume that the input is delayed 1 unit.

The smoothness condition imposed on $H^u(\theta)$ and $H^e(\theta)$ is interpreted via a state-space realization, see also [13].

Condition 2: The input process $u = (u_n)$ and the noise process $e = (e_n)$, with $-\infty < n < +\infty$, are jointly w.s.st. (wide sense stationary) stochastic processes. Moreover, (e_n) is a martingale difference process with respect to (w.r.t.) an increasing family of σ-algebras $(\mathcal{F}_n)$ such that

$$\mathbb{E}\left[e_n|\mathcal{F}_{n-1}\right] = 0 \quad \text{and} \quad \mathbb{E}\left[e_n^2|\mathcal{F}_{n-1}\right] = \sigma^2 \quad \text{a.s.} \tag{2}$$

Finally we assume that u is orthogonal to e, written as $u \perp e$.

For the basic concepts of the theory of w.s.st. processes and system identification see [14], [15], and [16]. It follows that e is the innovation process of $y - H^u(\theta^*, q^{-1})u$.

The spectral distribution measure of u is denoted by $\mathrm{d}\Psi^u(\omega)$, with $-\pi \leq \omega \leq \pi$. Since it is symmetric, its restriction to $[0, \pi]$, with $\mathrm{d}\Psi^u(\{0\})$ and $\mathrm{d}\Psi^u(\{\pi\})$ halved, is denoted by $\mathrm{d}\Phi^u(\omega)$.

A *multi-sine input* u is defined as

$$u_n := \sum_{k=1}^{t} \sigma_k 2\cos(\varphi_k + \omega_k n) \tag{3}$$

where $0 \leq \omega_k \leq \pi$ are different, and the random phases φ_k are independent and uniformly distributed in $[-\pi, \pi]$. Thus $\mathrm{d}\Phi^u(\cdot)$ is discrete, assigning the energy σ_k^2 to each frequency.

To fix notations for the description of the off-line PE (prediction error) estimator of θ^* define for any $\theta \in D$ the assumed innovation process $\varepsilon(\theta)$ as the w.s.st. process

$$\varepsilon(\theta) = H^e(\theta)^{-1}\left(y - H^u(\theta)u\right), \tag{4}$$

defined for $-\infty < n < \infty$. Assuming the observations are collected for $1 \leq n \leq N$ the (idealized) off-line PE method of θ^* is then obtained by minimizing the cost function

$$V_N(\theta) := \frac{1}{2}\sum_{n=1}^{N} \varepsilon_n^2(\theta). \tag{5}$$

The solution will be denoted by $\hat{\theta}_N$. A precise definition, taking into account the possibility of no solution or multiple solutions, can be obtained along the lines of [13]. In practice the w.s.st. process $\varepsilon(\theta)$ is approximated by a process defined via (4) with $u_n = y_n = 0$ for $n \leq 0$. The asymptotic cost function associated with the PE method is then

$$W(\theta) := \frac{1}{2}\mathbb{E}\varepsilon_n^2(\theta). \tag{6}$$

The Hessian of $W(\theta)$ for $\theta = \theta^*$ is given by

$$M = M\left(\theta^*\right) := \frac{\partial^2}{\partial\theta^2} W(\theta)_{\theta=\theta^*} = \mathbb{E}\varepsilon_{\theta n}\left(\theta^*\right)\varepsilon_{\theta n}^\top\left(\theta^*\right), \tag{7}$$

where $\varepsilon_{\theta n}(\theta)$ denotes the gradient of $\varepsilon_n(\theta)$ w.r.t. θ, considered to be a column vector. The system or θ^* is *locally identifiable* if M is positive definite, written as $M \in \mathbb{R}_+^{p\times p}$.

Condition 3: The system (1) is locally identifiable if the input u is a w.s.st. orthogonal process, independent of e. Equivalently, there exists a multi-sine input (3), independent of e, such that M is non-singular (see Proposition 1).

Precise conditions for the non-singularity of M are given in [17]. The existence of an asymptotic covariance matrix of $\hat{\theta}_N$ is established under a variety of conditions in the literature, see [16]. To ease reference we state two yet unpublished results that can be obtained by straightforward extensions of [13, Th. 2.1]. For the concept of L-mixing, an extremely useful extension of what was defined as exponentially stable processes in [18], see [19].

Theorem 1: Assume Conditions 1, 2 and 3, and let $M(\theta^*)$ be non-singular. In addition, let (u, e) be L-mixing w.r.t. a family of pairs of σ-algebras $(\mathcal{F}_n, \mathcal{F}_n^+)$. Then

$$\hat{\theta}_n - \theta^* = -M\left(\theta^*\right)^{-1}\sum_{n=1}^{N}\varepsilon_{\theta n}\left(\theta^*\right)e_n + r_N \tag{8}$$

where $r_N = O_M(N^{-1})$, indicating that the L_p-norms of r_N decay with rate $O(N^{-1})$ for all $p \geq 1$. It follows that

$$\Sigma_{\theta\theta} := \lim_{n\to\infty} N^{1/2}\left(\hat{\theta}_n - \theta^*\right)\left(\hat{\theta}_n - \theta^*\right)^T = \sigma^2 M\left(\theta^*\right)^{-1}.$$

An extension of this theorem for multi-sine inputs, which are far from being mixing in any sense, can be easily obtained, noting that if u is a multi-sine, independent of e, then the products of processes of the form $D(q^{-1})u$ and $D(q^{-1})e$ satisfy a law of large numbers with controlled rate.

The Hessian M, modulo a constant multiplier, is the information matrix. To capture the effect of u onto $M(\theta^*)$ consider the gradient process $(\varepsilon_{\theta n}(\theta^*))$. Equation (4) gives

$$\varepsilon_\theta\left(\theta^*\right) = -H^e\left(\theta^*\right)^{-1}\left(H_\theta^u\left(\theta^*\right)u + H_\theta^e\left(\theta^*\right)e\right), \quad \text{or} \tag{9}$$

$$\varepsilon_\theta\left(\theta^*\right) = D^u\left(\theta^*\right)u + D^e\left(\theta^*\right)e, \tag{10}$$

with $D^u(\theta^*)$ and $D^e(\theta^*)$ given by

$$D^u\left(\theta^*, e^{-i\omega}\right) = -H^e\left(\theta^*, e^{-i\omega}\right)^{-1} H_\theta^u\left(\theta^*, e^{-i\omega}\right), \tag{11}$$

$$D^e\big(\theta^*, e^{-i\omega}\big) = -H^e\big(\theta^*, e^{-i\omega}\big)^{-1} H^e_\theta\big(\theta^*, e^{-i\omega}\big). \quad (12)$$

Taking into account that $u \perp e$, we get $M = M^u + M^e$ where M^u and M^e are given via the expressions:

$$M^u = 2\int_0^\pi \Re\Big(D^u\big(e^{-i\omega}\big)D^{u\top}\big(e^{i\omega}\big)\Big)\, \mathrm{d}\Phi^u(\omega), \quad (13)$$

$$M^e = \int_0^\pi D^e\big(e^{-i\omega}\big)D^{e\top}\big(e^{i\omega}\big)\, \mathrm{d}\omega \cdot \sigma^2, \quad (14)$$

where the r.h.s. of (13) is a Riemann-Stieltjes integral. Here the transfer functions $D^u(e^{-i\omega}) := D^u(\theta^*, e^{-i\omega})$ and $D^e(e^{-i\omega}) := D^e(\theta^*, e^{-i\omega})$ are explicitly known.

The set of feasible matrices M^u is defined via (13) with $\mathrm{d}\Phi^u(\omega)$ being arbitrary subject to (s.t.) constraint as follows. Let $w(\omega), \omega \in [0, \pi]$ be a bounded, continuous function, for which $w(\omega) \geq w_0 > 0$ for all $\omega \in [0, \pi]$ holds, and impose

$$\int_0^\pi w(\omega)\, \mathrm{d}\Phi^u(\omega) \leq K. \quad (15)$$

The set of spectral distribution measures $\mathrm{d}\Phi^u(\cdot)$, satisfying the above the weighted energy constraint, is a compact convex set in the weak topology. Since the matrices M^u are obtained from $\mathrm{d}\Phi^u(\cdot)$ via a continuous linear operator (13), they also constitute a compact, convex set of symmetric, non-negative definite matrices $\mathcal{M}^u \subset \mathbb{R}^{p\times p}$.

Assume that the performance index $J(\theta)$ is sufficiently smooth in θ. Let its Hessian at $\theta = \theta^*$ be denoted by P. Obviously $P^\top = P \geq 0$. Assume that P is positive definite. The objective of the input design then is to minimize the asymptotic value of the normalized performance degradation $\lim_N N\mathbb{E}(J(\hat\theta_N) - J(\theta^*))$. The *primary input design* problem is then to optimize the information matrix M via

$$\min_{M^u \in \mathcal{M}^u} \mathrm{tr}\Big(\big(M^u + M^e\big)^{-1} P\Big)\Big), \quad (16)$$

with $\mathrm{tr}(M^{-1}P)$ defined as $+\infty$, if $M \geq 0$ is singular.

It is easily seen that $M \to M^{-1}$ is strictly convex on $\mathbb{R}^{p\times p}_+$ w.r.t. the usual ordering of symmetric matrices. It follows that $\mathrm{tr}(M^{-1}P)$ is a strictly convex function of $M \in \mathbb{R}^{p\times p}_+$. It is readily seen that the optimization problem (16) has a unique solution in $\mathcal{M}^u$, to be denoted by M^{u*}. The question remains how to construct a spectral distribution measure $\mathrm{d}\Phi^{u*}(\cdot)$ or input u^* that would generate M^{u*}.

III. Multi-Sine Input Design

The matrices $\Re(D^u(e^{-i\omega})D^{u\top}(e^{i\omega}))$, defining M^u via (13), are elements of the vector-space of real, symmetric matrices of dimension $s := p(p+1)/2$. The following result was essentially stated for the case of $w(\cdot) \equiv 1$, in [20, Ch. 3.2, Th. 1]:

Proposition 1: Let $M^u \in \mathcal{M}^u$ be defined in terms of $\mathrm{d}\Phi^u(\cdot)$, satisfying (15) with equality, via (13). Then there exist at most $s+1$ frequencies $0 \leq \omega_k \leq \pi$ and energy levels $\alpha_k \geq 0$, $k = 1, \ldots, s+1$, such that

$$M^u = 2\sum_{k=1}^{s+1} \alpha_k \Re\Big(D^u\big(e^{-i\omega_k}\big)D^{u\top}\big(e^{i\omega_k}\big)\Big), \quad (17)$$

$$\sum_{k=1}^{s+1} \alpha_k w(\omega_k) = \int_0^\pi w(\omega)\, \mathrm{d}\Phi^u(\omega) = K. \quad (18)$$

For extremal points of $\mathcal{M}^u$ just s frequencies suffice.

The above fact is a folklore in the literature on input design. The proof is based on Carathéodory's theorem, and is given in [20] for the case $w(\cdot) \equiv 1$,. The general case is readily obtained by introducing the measure $\mathrm{d}\overline{\Phi}^u(\omega) = w(\omega)\,\mathrm{d}\Phi^u(\omega)/K$, and rewriting (13) as

$$M^u = 2\int_0^\pi \frac{K}{w(\omega)} \Re\Big(D^u\big(e^{-i\omega}\big)D^{u\top}\big(e^{i\omega}\big)\Big)\, \mathrm{d}\overline{\Phi}^u(\omega).$$

For extremal points of $\mathcal{M}^u$ the discrete representation (17) is not only a possibility but in some cases a must. Adapting the arguments of [20, Ch. MA3, Th. 2]. we get:

Theorem 2: Let $\mathrm{d}\Phi^{u*}(\cdot)$ be a spectral measure defining an extremal point of $\mathcal{M}^u$, denoted by M^{u*}. Then there exists a $p \times p$ matrix Λ^* such that

$$L(\omega) := \mathrm{tr}\Lambda^* \Re\Big(D^u\big(e^{-i\omega}\big)D^{u\top}\big(e^{i\omega}\big)\Big) + w(\omega) \geq 0 \quad (19)$$

for all $0 \leq \omega \leq \pi$, and the set of the points of increase of $\mathrm{d}\Phi^{u*}(\cdot)$ is a subset of the solutions of $L(\omega) = 0$.

The theorem is a kind of infinite-dimensional Karush-Kuhn-Tucker condition, see Lemma 1.

Corollary 1: Let $w(\cdot)$ be a piece-wise rational function of $e^{i\omega}$. Then any optimal spectral measure $\mathrm{d}\Phi^{u*}(\cdot)$ is discrete, supported by a finite number of frequencies.

Proof of Theorem 2: Adapting the argument in [20, p. 45] we obtain that M^{u*} is in the boundary of the convex, closed set $\mathcal{M}^u$. Thus, there exists a matrix $\tilde\Lambda$ and a number $c \neq 0$ such that $\mathrm{tr}\tilde\Lambda M^u \leq c$ for any matrix $M^u \in \mathcal{M}^u$ and $\mathrm{tr}\tilde\Lambda M^{u*} = c$. Using the definition (13) of M^u we can write the above condition as

$$\int_0^\pi \Big(\mathrm{tr}\tilde\Lambda \Re\Big(D^u\big(e^{-i\omega}\big)D^{u\top}\big(e^{i\omega}\big) - \frac{c}{K} w(\omega)\Big)\, \mathrm{d}\Phi^u(\omega) \leq 0$$

for any $\mathrm{d}\Phi^u(\cdot)$, for which $\int_0^\pi w(\omega)\,\mathrm{d}\Phi^u(\omega) = K$ holds. Since the function $w(\cdot)$ is strictly positive, it follows that $\mathrm{tr}\tilde\Lambda\Re(D^u(e^{-i\omega})D^{u\top}(e^{i\omega}) - \frac{c}{K}w(\omega) \leq 0$ for $\omega \in [0, \pi]$. Setting $\Lambda = -\frac{K}{c}\tilde\Lambda$ we obtain $\int_0^\pi L(\omega)\,\mathrm{d}\Phi^u(\omega) \geq 0$ for any $\mathrm{d}\Phi^u(\cdot)$, with equality for $\mathrm{d}\Phi^{u*}(\omega)$, implying the claim.

Since $M^{u*} \in \mathcal{M}^u$ is in the boundary of $\mathcal{M}^u$ Proposition 1 implies that it can be generated via a *multi-sine* of at most s terms, see (3). The matrix

$$M^u = 2\sum_{k=1}^{s} \alpha_k \Re\Big(D^u\big(e^{-i\omega_k}\big)D^{u\top}\big(e^{i\omega_k}\big)\Big), \quad (20)$$

is linear in the α_k-s, and hence, for a fixed set of ω_k frequencies, the cost function $\mathrm{tr}(M^{-1}P)$ is convex in the α_k-s. Unfortunately, it is a *non-convex* function in the frequencies $\omega_k, k = 1, \ldots, s$.

A sub-optimal solution to this can be obtained by taking a large t, and a set of equidistant frequencies denoted by $\Omega = \{\omega_k : 0 \leq \omega_k \leq \pi, k = 1, \ldots, t\}$, and considering

$$M^u = 2\sum_{k=1}^{t} \alpha_k \Re\Big(D^u\big(e^{-i\omega_k}\big)D^{u\top}\big(e^{i\omega_k}\big)\Big). \quad (21)$$

The convex set of matrices M^u generated by (21), s.t. the energy constraint will be denoted by $\mathcal{M}^u(\Omega)$. With this notation our input design problem (16) reduces to the problem with $M^u \in \mathcal{M}^u$ being replaced by $M^u \in \mathcal{M}^u(\Omega)$.

To capture the effect of the approximation replacing $\mathcal{M}^u$ by $\mathcal{M}^u(\Omega)$ in the primary input design problem (16) note that any $M^u \in \mathcal{M}^u$ is defined by a Riemann-Stieltjes integral, given by (13), in which the integrand is continuously differentiable and the total mass of the measure $\mathrm{d}\Phi^u(\cdot)$ on $[0, \pi]$ is bounded, due to the energy constraint (15) and the condition $w(\omega) \geq w_0 > 0$ for all ω. It readily follows that any $M^u \in \mathcal{M}^u$ can be approximated by an $M^{ud} \in \mathcal{M}^u(\Omega)$ with an error of the order $1/t$, inducing an error of the same order of magnitude in approximating the optimal value.

Letting $d_k := D^u(e^{-i\omega_k})$ and $w_k = w(\omega_k)$ we thus get the following convex optimization problem:

$$\min_{\alpha \geq 0} \ \mathrm{tr}\left[\left(2\Re \sum_{k=1}^{t} \alpha_k d_k \overline{d}_k^\top + M_e\right)^{-1} P\right] \tag{22}$$

$$\text{s.t.} \quad \sum_{k=1}^{t} \alpha_k w_k \leq K. \tag{23}$$

The formulation of the the Karush-Kuhn-Tucker condition for (22) – (23) is of didactic interest in light of Theorem 2:

Lemma 1: Let α^* be an optimal solution of (22) – (23). Then there exists a $p \times p$, symmetric negative definite matrix Λ^*, and Lagrange multipliers $\lambda^* \geq 0$ and $\mu_k^* \leq 0$ such that

$$\mathrm{tr}\left[\Lambda^* \Re\left(d_k \overline{d}_k^\top\right)\right] + \lambda^* w_k + \mu_k^* = 0, \tag{24}$$

for all $k = 1, \dots, t$ and $\alpha_k^* > 0$ implies $\mu_k^* = 0$.

Proof: Let the cost function in (22) be denoted by $F(\alpha)$. Let $\lambda^* \geq 0$ be the Lagrange multiplier corresponding (23) and let $\mu_k^* \leq 0$ be the Lagrange multipliers corresponding to the constraints $\alpha_k \geq 0$. Let

$$M = M(\alpha) = 2\Re \sum_{k=1}^{t} \alpha_k d_k \overline{d}_k^\top + M_e. \tag{25}$$

The gradient of the cost function $F(\alpha)$ is then as follows:

$$\frac{\partial}{\partial \alpha_k} F(\alpha) = 2\mathrm{tr}\left[-M^{-1} \Re\left(d_k \overline{d}_k^\top\right) M^{-1} P\right]. \tag{26}$$

Letting $\Lambda = \Lambda(\alpha) := -2M^{-1} P M^{-1}$ we can write:

$$\frac{\partial}{\partial \alpha_k} F(\alpha) = \mathrm{tr}\left[\Lambda \Re\left(d_k \overline{d}_k^\top\right)\right]. \tag{27}$$

Setting $\Lambda^* = \Lambda(\alpha^*)$, the Karush-Kuhn-Tucker condition, implying also $\mu_k^* \alpha_k^* = 0$, gives the claim. ∎

For a fixed a set of equidistant ω_k-s let us consider a *relaxation* of the problem defined in (22)-(23) with $\gamma > 0$,

$$\min_{\alpha \geq 0} \mathrm{tr}\left[\left(2\Re \sum_{k=1}^{t} \alpha_k d_k \overline{d}_k^\top + M_e\right)^{-1} P\right] + \gamma \sum_{k=1}^{t} \alpha_k w(\omega_k).$$

Let α^* be an optimal solution of this *relaxed problem*, and let $\Omega^+ = \{\omega_k : \alpha_k^* > 0\}$. Then the Karush-Kuhn-Tucker conditions, with minor modifications of Lemma 1, imply for $\omega \in \Omega^+$

$$\mathrm{tr}\left[\Re\left(D^u\left(e^{-i\omega}\right)\overline{D}^u\left(e^{i\omega}\right)^\top\right)\Lambda^*\right] + \gamma w(\omega) = 0. \tag{28}$$

Let us introduce the notation for the vectorized matrices

$$\mathrm{vec}\Re\left(D^u\left(e^{-i\omega}\right)\overline{D}^u\left(e^{i\omega}\right)^\top\right) =: \Delta(\omega). \tag{29}$$

Lemma 2: The relaxed optimization problem has a solution such that the vectors $\{\Delta(\omega_k), \omega_k \in \Omega^+\}$ are linearly independent. In particular, $|\Omega^+| \leq s$.

Proof: Let $I^+ = \{k : \alpha_k^* > 0\}$. If the vectors $\{\Delta(\omega_k), \omega_k \in \Omega^+\}$ are linearly dependent, then there exists a nontrivial linear combination $\sum_{k \in I^+} \beta_k \Re(d_k \overline{d}_k^\top) = 0$, where we can assume $|\beta_k| < \alpha_k^*$ for all $k \in I^+$. Adding and subtracting this linear combination from the optimal one that cost function cannot decrease. Thus $\sum_{k \in I^+} \beta_k w(\omega_k) = 0$. Taking $\alpha_k^* + \lambda \beta_k$ for some appropriate λ we can achieve that $\alpha_k^* + \lambda \beta_k = 0$ for some $k = l$ while ensuring $\alpha_k^* + \lambda \beta_k \geq 0$ for all other $k \in I^+$, thus reducing the size of Ω^+. Repeating this procedure will yield the desired optimal solution. ∎

Let $I \subset \{1, \dots, t\}$ and let $\alpha_I^\top := (\alpha_k, k \in I)$ denote the reduced parameter vector. Enforcing $\alpha_k = 0$ for $k \notin I$, let $L_I(\alpha_I)$ be the restricted cost function of the relaxed problem.

Lemma 3: Assume that $\{\Delta(\omega_k), k \in I\}$ are linearly independent. Then the Hessian of $L_I(\alpha_I)$ is positive definite.

The proof is obtained considering the quadratic form induced by the Hessian for $v \in \mathbb{R}^t$, with $G_k = 2\Re d_k \overline{d}_k^\top$

$$2\mathrm{tr}\left(P^{1/2} M^{-1}\left(\sum_{k=1}^{t} v_k G_k\right) M^{-1}\left(\sum_{l=1}^{t} v_l G_l\right) M^{-1} P^{1/2}\right).$$

Restricting summation to $k, l \in I^+$ gives the claim.

Theorem 3: Let the t-dimensional vector with components $w_k := w(\omega_k)$ be chosen randomly according to a distribution having a density in $\mathbb{R}^t$. Then the relaxed problem has a unique solution w.p.1, and the vectorized matrices $\mathrm{vec}\Re(D^u(e^{-i\omega})\overline{D}^u(e^{i\omega})^\top)$, $\omega \in \Omega^+$ are linearly independent w.p.1. In particular, we have $|\Omega^+| \leq s$.

Proof: Let us take an optimal solution α^* with Ω^+ as defined above. Let $I \subset \{1, \dots, t\}$ be arbitrary and let $\Omega_I := \{\omega_k, k \in I\}$ be the corresponding subset of frequencies. Note that $P(\Omega^+ = \Omega_I) \leq P(\Omega_I \subseteq \Omega^+)$. Express the latter event $\{\Omega_I \subseteq \Omega^+\}$ via the Karush-Kuhn-Tucker condition as

$$\left(\mathrm{vec}\Lambda^*\right)^\top \Delta(\omega) + \gamma w(\omega) = 0 \quad \text{for} \quad \omega \in \Omega_I. \tag{30}$$

Arrange the column-vectors $\{\Delta(\omega_k), k \in I\}$ into a matrix S_I, and define $w_I^\top := (w(\omega_k), k \in I)$. Write (30) as

$$\left(\mathrm{vec}\Lambda^*\right)^\top S_I + \gamma w_I^\top = 0. \tag{31}$$

If $\mathrm{rank} S_I < |I|$, i.e., the vectors $\{\Delta(\omega_k), k \in I\}$ are linearly dependent, then its rows span a proper subspace $L(S_I) \subset \mathbb{R}^{|I|}$. But the marginal distribution of the random vector w_I has a density in $\mathbb{R}^{|I|}$, hence the event $\{w_I \in L(S_I)\}$ has probability 0. Since the number of subsets I is finite, the second claim follows.

To prove unicity, assume the contrary. Then, by convexity, there is an interval of α-s such that the cost function is constant, and optimal along this interval. Consider its midpoint, say $\overline{\alpha}^*$ and let $I := \overline{I}^+ = \{k : \overline{\alpha}_k^* > 0\}$, and $\overline{\Omega}^+ = \{\omega_k : k \in \overline{I}^+\}$. Then by the proven second claim of the theorem the vectors $\{\Delta(\omega_k), k \in I\}$ are linearly independent w.p.1. Hence, by Lemma 3 the Hessian of $L_I(\alpha_I)$ is positive definite. But this is a contradiction, since $L_I(\alpha_I)$ being constant along an interval, its Hessian has a zero eigenvalue. ∎

IV. A Data-Driven Approach

A shortcoming of the cited literature on input design is that the optimal spectral measure of the input is determined under the hypothesis that the true system parameter θ^*, is actually known. To bypass this paradox we present the basics of a data-driven method within a fairly general context, recapitulating and extending the basic idea of [9].

The key idea is the construction of a data-driven virtual off-line estimator, approximating the off-line PE estimator of Section II, obtained with optimal input, with accuracy $O_M(N^{-1})$. To be specific, consider a parametric family of inputs

$$u(\eta) = F\left(\eta, q^{-1}\right)f, \tag{32}$$

where F is a rational, stable filter, such that $|F|^2$ is linearly parameterized by $\eta \in \mathcal{C} \subset \mathbb{R}^r$, where $\mathcal{C}$ is a closed, convex set. E.g., F may be a FIR filter with η denoting the coefficients of its half-spectra. f is an i.i.d. sequence of random variables, independent of e, with finite moments of all order.

Consider the system dynamics (1) with $u = u(\eta)$, $\eta \in \mathcal{C}$:

$$y(\eta) = H^u\left(\theta^*, q^{-1}\right)u(\eta) + H^e\left(\theta^*, q^{-1}\right)e. \tag{33}$$

Define, for $\theta \in D$, the assumed innovation process $\varepsilon(\theta, \eta)$

$$\varepsilon(\theta, \eta) = H^e(\theta)^{-1}\big(y(\eta) - H^u(\theta)u(\eta)\big). \tag{34}$$

Renaming the cost function of the PE estimator defined in (5) as $V_N(\theta, \eta)$ let $\hat{\theta}_N(\eta)$ and $M(\theta^*, \eta)$ denote the corresponding off-line PE estimator and information matrix, respectively. Assume that for $\theta \in D$, as true systems parameter, the optimal input design problem has a unique solution $\eta^*(\theta)$, such that $\eta^*(\cdot)$ is three-times continuously differentiable. Then the *virtual* off-line PE estimator of θ^* is obtained by minimizing the cost function

$$V_N(\theta, *) := V_N\big(\theta, \eta^*(\theta)\big) = \frac{1}{2}\sum_{n=1}^{N} \varepsilon_n^2\big(\theta, \eta^*(\theta)\big). \tag{35}$$

The solution is denoted by $\hat{\theta}_N(*)$. This estimator is virtual in the sense that it is not practical, since $V_N(\theta, *)$ can not be evaluated for two different values of θ. Nevertheless, we proceed with its analysis along the lines of Section II.

Noting that $\varepsilon(\theta^*, \eta) = e_n$ for all η, we get $\frac{\partial}{\partial \eta}\varepsilon(\theta^*, \eta) = 0$ for all η. Setting $\varepsilon_{\theta 0,n}(\theta, \eta) := \frac{\partial}{\partial \theta}\varepsilon_n(\theta, \eta)$, we conclude:

$$\frac{\partial}{\partial \theta}\varepsilon_n\big(\theta, \eta^*(\theta)\big) = \varepsilon_{\theta 0,n}\big(\theta, \eta^*(\theta)\big). \tag{36}$$

It follows that for the Hessian of the asymptotic cost function associated with the above virtual PE method, given as,

$$W(\theta, *) := W\big(\theta, \eta^*(\theta)\big) := \frac{1}{2}\mathbb{E}\varepsilon_n^2\big(\theta, \eta^*(\theta)\big), \tag{37}$$

evaluated at θ^*, we have, with self-explanatory notation,

$$M\big(\theta^*, *\big) = M\big(\theta^*, \eta^*(\theta^*)\big). \tag{38}$$

Once again referring to [13, Th. 2.1] we get by its straightforward extension, in analogy with Theorem 1.

Theorem 4: Let $u(\eta)$ be given by (32). Assume Conditions 1, 3, and let $M(\theta^*, *)$ be non-singular. Then

$$\hat{\theta}_n(*) - \theta^* = -M\big(\theta^*, *\big)^{-1}\sum_{n=1}^{N} \varepsilon_{\theta 0,n}\big(\theta^*, \eta^*(\theta^*)\big)e_n + r_N,$$

where $r_N = O_M(N^{-1})$, implying the strong approximation:

$$\hat{\theta}_n(*) = \hat{\theta}_n\big(\eta^*(\theta^*)\big) + O_M\left(N^{-1}\right). \tag{39}$$

In particular, the asymptotic covariance matrix of $\hat{\theta}_n(*)$ is

$$\Sigma_{\theta\theta}(*) = \sigma^2 M^*\big(\theta^*, *\big)^{-1} = \sigma^2 M^*\big(\theta^*, \eta^*(\theta^*)\big)^{-1}.$$

Thus the virtual estimator $\hat{\theta}_n(*)$ is optimal from the perspective of input design. The asymptotic estimation problem in the spirit of [12] is defined by the algebraic equation

$$\frac{\partial}{\partial \theta}W(\theta, *) = \mathbb{E}\varepsilon_{\theta 0,n}\big(\theta, \eta^*(\theta)\big)\varepsilon_n\big(\theta, \eta^*(\theta)\big) = 0. \tag{40}$$

Following the ideas of [11], extended in [12], a computable recursive PE estimator $\hat{\hat{\theta}}_N(*)$ can be constructed.

The viability of the proposed approach for data-driven input design has been demonstrated, with all technical details included, in [9] for the case of ARMAX systems exited with inputs u generated by a FIR filter.

To conclude this section we briefly describe the extension of the above approach to *closed loop* systems. Consider a class of linear stochastic control systems with excitation v:

$$y^c\big(\theta'\big) = H^u\left(\theta^*, q^{-1}\right)u^c\big(\theta'\big) + H^e\left(\theta^*, q^{-1}\right)e \tag{41}$$

$$u^c\big(\theta'\big) = -K\left(\theta', q^{-1}\right)y^c\big(\theta'\big) + v. \tag{42}$$

Here v is an external excitation independent of e. We consider the practical scenario when the true parameter θ^* is unknown and we use its tentative value θ' in the feedback loop. The feedback loop or $K(\theta, q^{-1})$ is designed by optimizing a performance criterion for any assumed θ showing up in H^u and H^e. Thus if we had a prior estimate $\hat{\theta}$ of θ^* we would set $\theta' = \hat{\theta}$. Write (41) - (42) as

$$y^c\big(\theta'\big) = H^{cu}\big(\theta^*, \theta'\big)\ v + H^{ce}\big(\theta^*, \theta'\big)e, \tag{43}$$

with $H^{cu}(\theta^*, \theta')$ and $H^{ce}(\theta^*, \theta')$ denoting the closed loop filters. The role of the dither v, being independent of e, is thus identified with that of the input u in open loop identification. Thus, pretending the knowledge of θ^* (and knowing θ') we can proceed with any of the open loop input design methods.

Following (32) assume that v is generated by

$$v(\eta) = F\left(\eta, q^{-1}\right)f, \tag{44}$$

yielding the input and output processes $u^c(\theta', \eta)$ and $y^c(\theta', \eta)$. For fixed θ', η the off-line PE estimator of θ^* is obtained via the assumed innovation process defined by

$$\varepsilon^c\big(\theta, \theta', \eta\big) = H^e(\theta)^{-1}\big(y^c\big(\theta', \eta\big) - H^u(\theta)u^c\big(\theta', \eta\big)\big).$$

The off-line PE estimator is then defined as the solution of

$$\min_{\theta \in D}\sum_{n=1}^{N} \varepsilon_n^{c2}\big(\theta, \theta', \eta\big), \tag{45}$$

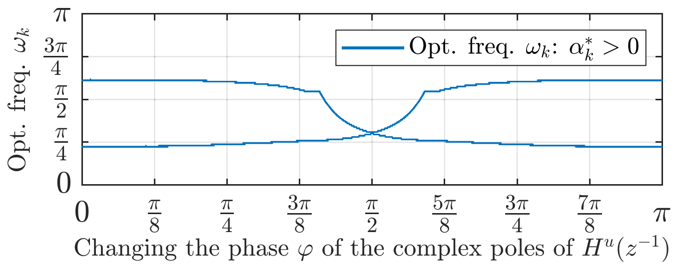

Fig. 1. The two optimal frequencies for $b = 3.0$ and $\omega_c = \frac{3\pi}{4}$.

to be denoted by $\hat{\theta}_N(\theta', \eta)$, which can be analyzed along the lines of Section II. For any θ, θ' let the solution of the input design problem be denoted by $\eta^*(\theta, \theta')$. We redefine the assumed innovation process by *enforcing* $\theta' = \theta$ and $\eta = \eta^*(\theta, \theta')$. Thus we define a virtual closed loop off-line PE estimator, as in [21], which can be considered as the off-line mirror-image of an adaptive control algorithm optimized for the asymptotic covariance matrix of $\hat{\theta}_N(\theta^*, \eta)$, by

$$\min_{\theta \in D} \sum_{n=1}^{N} \varepsilon_n^{c2}\big(\theta, \theta, \eta^*(\theta, \theta)\big). \tag{46}$$

Let the solution be denoted by $\hat{\theta}_N(*, *)$. Noting that $\varepsilon^c(\theta^*, \theta', \eta) = e_n$ for all θ', η, the partial derivatives of $\varepsilon^c(\theta^*, \theta', \eta)$ w.r.t. θ', η are 0, and hence the gradient of the cost function or rather $(\partial/\partial\theta)\varepsilon_n^c(\theta, \theta, \eta^*(\theta, \theta))_{\theta=\theta^*}$ is easily computed. Thus, in analogy with Theorem 4 we get the strong approximation result

$$\hat{\theta}_n(*, *) = \hat{\theta}_n\big(\theta^*, \eta^*(\theta^*)\big) + O_M\Big(N^{-1}\Big), \tag{47}$$

amounting to the fact that $\hat{\theta}_n(*, *)$ is optimal both from control and input design perspective.

V. Experimental Results

We have tested our algorithm for finding the optimal multi-sine on a system modeling a lightly damped oscillator with complex poles $re^{\pm i\varphi}$ and amplification b. Thus we have

$$H^u\Big(z^{-1}\Big) = \frac{b}{1 - 2r\cos(\varphi)z^{-1} + r^2z^{-2}}. \tag{48}$$

Fixing $r = 0.95$, we let the phase vary uniformly in $[0, \pi]$, while b varied in the interval $[3, 10]$. The transfer function H^e is defined by its stable zeros and poles yielding

$$H^e\Big(z^{-1}\Big) = \frac{1 + c_1z^{-1} + c_2z^{-2}}{1 + d_1z^{-1} + d_2z^{-2}} = \frac{1 + 0.6z^{-1} - 0.07z^{-2}}{1 - 0.866z^{-1} + 0.25z^{-2}}.$$

Thus we have $p = 7$ parameters: r, φ, b and c_1, c_2, d_1, d_2, implying $s = p(p+1)/2 = 28$. For the weight function $w(\cdot)$ we use a sigmoid-type functions taking their values between 0.1 and 1.0, setting their medians equal to three possible cut-off frequencies ω_c equal to $\frac{\pi}{4}, \frac{\pi}{2}, \frac{3\pi}{4}$. We let $t = 5s = 140$. We solved the relaxed problem with $P = I$, and $\gamma = 0.1$ In all experimental scenarios a sparse solution was obtained with a maximum of 4 optimal frequencies. In Fig. 1, we present a typical result with two optimal frequencies assuming a moderate SNR (signal-to-noise ratio) $b = 3.0$, and using a weight function with broad band-pass width: $\omega_c = \frac{3\pi}{4}$.

VI. Discussion

A nice project for future research may be the extension of Theorem 2 to input design problems admitting frequency-wise specifications, introduced in [22]. A second problem of interest may be the clarification if the proposed data-driven approach of Section IV is applicable for multi-sine design. Extension of our results to vector-valued multi-sine design along the lines of [23] may be also of interest. Finally, the authors thank to Hakan Hjalmarsson for inspiring this letter, while visiting HUN-REN SZTAKI in 2021.

References

[1] H. Hjalmarsson, "System identification of complex and structured systems," *Eur. J. Control*, vol. 15, nos. 3–4, pp. 275–310, 2009.

[2] A. Wagenmaker and K. Jamieson, "Active learning for identification of linear dynamical systems," in *Proc. Conf. Learn. Theory*, 2020, pp. 3487–3582.

[3] R. Hildebrand, M. Gevers, and G. E. Solari, "Closed-loop optimal experiment design: Solution via moment extension," *IEEE Trans. Autom. Control*, vol. 60, no. 7, pp. 1731–1744, Jul. 2015.

[4] K. Lindqvist and H. Hjalmarsson, "Optimal input design using linear matrix inequalities," *IFAC Proc. Vol.*, vol. 33, no. 15, pp. 1085–1090, 2000.

[5] U. Forssell and L. Ljung, "Closed-loop identification revisited," *Automatica*, vol. 35, no. 7, pp. 1215–1241, 1999.

[6] H. Hjalmarsson and H. Jansson, "Closed loop experiment design for linear time invariant dynamical systems via LMI-s," *Automatica*, vol. 44, no. 3, pp. 623–636, 2008.

[7] H. Hjalmarsson, M. Gevers, and F. De Bruyne, "For model-based control design, closed-loop identification gives better performance," *Automatica*, vol. 32, no. 12, pp. 1659–1673, 1996.

[8] X. Bombois, F. Morelli, H. Hjalmarsson, L. Bako, and K. Colin, "Robust optimal identification experiment design for multisine excitation," *Automatica*, vol. 125, Mar. 2021, Art. no. 109431.

[9] L. Gerencsér, H. Hjalmarsson, and L. Huang, "Adaptive input design for LTI systems," *IEEE Trans. Autom. Control*, vol. 62, no. 5, pp. 2390–2405, May 2017.

[10] L. Gerencsér, "A representation theorem for the error of recursive estimators," *SIAM J. Control Optim.*, vol. 44, no. 6, pp. 2123–2188, 2006.

[11] T. Söderström, "An on-line algorithm for approximate maximum likelihood Identificatioin of linear dynamic systems," Dept. Autom. Control, Lund Inst. Technol., Lund, Sweden, Rep. 7308, 1973.

[12] L. Ljung and T. Söderström, *Theory and Practice of Recursive Identification*. Cambridge, MA, USA: MIT Press, 1983.

[13] L. Gerencsér, "On the martingale approximation of the estimation error of ARMA parameters," *Syst. Control Lett.*, vol. 15, no. 5, pp. 417–423, 1990.

[14] P. E. Caines, *Linear Stochastic Systems*. Philadelphia, PA, USA: SIAM, 2018.

[15] E. J. Hannan and M. Deistler, *The Statistical Theory of Linear Systems*. Philadelphia, PA, USA: SIAM, 2012.

[16] L. Ljung, *System Identification: Theory for the User*, vol. 28. Hoboken, NY, USA: Prentice Hall, 1999,

[17] M. Gevers, A. S. Bazanella, X. Bombois, and L. Miskovic, "Identification and the information matrix: How to get just sufficiently rich?" *IEEE Trans. Autom. Control*, vol. 54, no. 12, pp. 2828–2840, Dec. 2009.

[18] L. Ljung, "Convergence analysis of parametric identification methods," *IEEE Trans. Autom. Control*, vol. 23, no. 5, pp. 770–783, Oct. 1978.

[19] L. Gerencsér, "On a class of mixing processes," *Stochastics*, vol. 26, no. 3, pp. 165–191, 1989.

[20] M. B. Zarrop, *Optimal Experiment Design for Dynamic System Identification*. Berlin, Germany: Springer-Verlag, 1979.

[21] L. Gerencsér, "Closed loop parameter identifiability and adaptive control of a linear stochastic system," *Syst. Control Lett.*, vol. 15, no. 5, pp. 411–416, 1990.

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

[23] E. A. Morelli, "Optimal input design for aircraft stability and control flight testing," *J. Optim. Theory Appl.*, vol. 191, no. 2, pp. 415–439, 2021.