Finite-Time Regret Minimization for Linear Quadratic Adaptive Controllers: an experiment design approach

Kévin Colin, Håkan Hjalmarsson, Xavier Bombois

▶ **To cite this version:**

Kévin Colin, Håkan Hjalmarsson, Xavier Bombois. Finite-Time Regret Minimization for Linear Quadratic Adaptive Controllers: an experiment design approach. 2023. hal-04360490

HAL Id: hal-04360490
https://hal.science/hal-04360490

Preprint submitted on 21 Dec 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Finite-Time Regret Minimization for Linear Quadratic Adaptive Controllers: an experiment design approach ⋆

Kévin Colin [a,b], Håkan Hjalmarsson [a,b], Xavier Bombois [c,d]

[a] *Division of Decision and Control Systems, KTH Royal Institute of Technology, Sweden*

[b] *Centre for Advanced Bio Production AdBIOPRO, KTH Royal Institute of Technology, Sweden*

[c] *Laboratoire Ampère, UMR CNRS 5005, Ecole Centrale de Lyon, Université de Lyon, France*

[d] *Centre National de la Recherche Scientifique (CNRS), France*

Abstract

We tackle the problem of finite-time regret minimization in linear quadratic adaptive control. Regret minimization is a scientific field in both adaptive control and reinforcement learning research communities which studies the so-called trade-off between exploration and exploitation. Even though a large focus has been on linear quadratic adaptive control with theoretical finite-time bound guarantees on the expected regret growth rate, most of the proposed optimal exploration strategies do not take into account the scaling constant associated to the growth rate. Moreover, the exploration strategies are limited to white noise excitation. Using the tools from experiment design, we propose a computational tractable solution for the design of the external excitation chosen as a white noise filtered by a finite impulse response filter which is adapted on-line. The numerical example shows a reduced regret than available strategies in the literature.

Key words: Regret minimization, adaptive control, linear quadratic regulator, experiment design, linear systems, reinforcement learning

1 Introduction

Linear quadratic (LQ) control is a control strategy for linear state systems minimizing a quadratic cost on the states and inputs of the system [4]. However, the controller design requires perfect knowledge of the state matrices in order to yield efficient control cost minimization which is never possible to get as real-life systems are affected by disturbances. A remedy to this problem is to implement a model-based adaptive control policy where the controller is updated online based on the recursive identification of the state matrices using input-state data. Because of the disturbance presence, the model comes with uncertainties which in turn causes control performance degradation.

In many adaptive control problems, it is crucial to excite the system with an additional external excitation in order to guarantee an efficient decrease of the uncertainties. However, the use of an external excitation also causes control performance degradation as it disturbs the signals of the closed-loop. Therefore, a trade-off must be chosen between the control performances degradation due to the uncertainties of the model and the one due to the use of the external excitation. This problem has been formalized in both reinforcement learning and adaptive controller research communities as regret minimization. The regret is a function of both the exploration and exploitation costs and the external excitation is designed in such a way that it minimizes the regret over an infinite or finite-time horizon. Early works on regret minimization can be found in [27] for the multi-armed bandit problem and in [29,28] for minimum variance adaptive control.

Much effort on obtaining bounds for the growth rate of the regret has been done in the literature for linear quadratic adaptive control. In the LQ setting, it was shown that one can achieve at best $\mathcal{O}(\log(t))$ or $\mathcal{O}(\sqrt{t})$ growth rates, depending on the prior knowledge of the

⋆ This work was supported by VINNOVA Competence Center AdBIOPRO, contract [2016-05181] and by the Swedish Research Council through the research environment NewLEADS (New Directions in Learning Dynamical Systems), contract [2016-06079].

Email addresses: kcolin@kth.se (Kévin Colin), hjalmars@kth.se (Håkan Hjalmarsson), xavier.bombois@ec-lyon.fr (Xavier Bombois).

state matrices [22,24,45,2,36,41,35,44]. When both state matrices are unknown, the optimal rate for the regret is $\mathcal{O}(\sqrt{t})$ [22,24] and it is equal to $\mathcal{O}(\log(t))$ when one state matrix is known and under some mild assumptions on the optimal controller matrix. Adaptive control algorithms dealing with this problem can be classified as belonging to two families: optimism in face of uncertainty [1] (OFU) and certainty equivalence (CE) principle based strategies.

The OFU approach designs the adaptive control policy by selecting the one which minimizes the expected regret. The work in [2] gives a regret asymptotically scaling as $\mathcal{O}(\sqrt{t})$ in the LQ problem but the approach is computationally intractable as it requires to solve non-convex problems at each time instant. In pursuit of developing algorithms requiring less computational power, the Thompson sampling strategy was considered in [3] guaranteeing the regret to asymptotically grow as $\mathcal{O}(t^{2/3})$. In that case, the controller is updated at specific time instants based on an element of the uncertainty ellipsoid. The rate $\mathcal{O}(t^{2/3})$ is also attained with the robust algorithm of [13]. The Thompson sampling scheme of [3] was later improved in [36,25] providing a rate of $\mathcal{O}(\sqrt{t})$. This asymptotic rate can also be obtained with the computationally tractable OFU based algorithm of [10].

The CE strategy assumes that the identified model is error-free and designs the controller accordingly. It inspired several works in LQ regret minimization [41,35,44,38,14,24,22]. In [41], the authors developed an episodic algorithm guaranteeing an upper bound for the regret scaling as $\mathcal{O}(\sqrt{t})$ for finite-time horizon. An external excitation is added to the control effort and the properties of this excitation is designed epochs by epochs and the parameter vector is only identified at the end of these epochs when a sufficient number of data have been obtained. In [35,44,22], it is proven that perturbing the input with a white Gaussian noise excitation whose variance decays as $\mathcal{O}(1/\sqrt{t})$ guarantees an asymptotic regret with a growth rate of $\mathcal{O}(\sqrt{t})$ which has been proven to be the the optimal asymptotic rate for LQ problem in [42]. As we argued in [11], this $1/\sqrt{t}$-decaying excitation may not be optimal as it is in essence an open-loop strategy with respect to the external excitation, not accounting for how unmeasured disturbances may excite the system providing useful system information. A closed-loop approach is proposed in [11] referred to as Inverse Fisher Feedback Exploration (IF2E) where the amount of excitation is determined from an estimate of the Fisher information matrix. It is shown that this approach gives the optimal regret rate but with the additional benefit of better finite sample behavior. The IF2E scheme has also been recently extended to the LQ Gaussian setting in [6], a topic which is also treated in [45] for the case where not all the states are measured.

While achieving the optimal growth rate $\mathcal{O}(\sqrt{t})$ is of course very important, in practice, minimizing the scaling constant is also crucial. However, this scaling constant has never been taken into account in the aforementioned works. In experiment design in system identification [7,20,19], this scaling constant is the design objective and the excitation is the decision variable. For linear time invariant (LTI) experiment design problems, the decision variable is often the power spectrum density (PSD) of the excitation. This is to be contrasted with the proposed methods in the aforementioned references where the excitation signal is taken to be white while we know from experiment design works how important it is to optimize the frequency content of the excitation.

Regarding regret minimization, an experiment design approach was considered in [15] for adaptive $\mathcal{H}_2$ control of LTI systems for which the decision variable is an external excitation added to the control effort. In addition to the optimization of the scaling of the regret growth rate, the other advantage of considering a time-invariant experiment design approach as in [15] is the fact that the variance of the external excitation is not enforced to be decaying as $\mathcal{O}(1/\sqrt{t})$ as in, e.g., [44,23] which opens up more degrees of freedom for effective regret minimization. The adaptive control strategy is divided into several intervals of sufficient duration during which the controller is kept constant and only adapted at the end of each epoch using the CE principle. By doing so, the tools from LTI experiment design can be used and the decision variables are the PSDs of the external excitation on each interval. As classically done in experiment design [19], a Taylor expansion up to the second order is performed of the design objective and the obtained expression affinely depends on both the inverse of the Fisher information matrix and the PSDs of the external excitations. Moreover, the Fisher information matrix is also an affine expression of the PSD of the external excitation. Using a linear parametrization of the PSD [21,7], the experiment design problem can be reformulated as a convex semidefinite programming (SDP). This SDP depends on the unknown true system. To solve this problem, the true system will be replaced by the current estimate. To reduce the error induced by this approximation, a receding horizon approach, similar to model predictive control, has also been adopted in [15] for the design of the PSDs. Experiment design is also performed in a receding horizon fashion in [32,31] for model predictive control where it is integrated into the receding horizon control algorithm. Finally, we mention that adaptive input design has been used to address that optimal input design problems typically depend on the to-be-identified, and it has been shown that with such designs the same asymptotic accuracy can be obtained as if the true system was used initially in the design [16].

Consequently, in Section 4 of this paper, we draw on [15] and first propose an extension of their interval-based approach combined with the receding horizon principle to the LQ regret minimization problem, referred here as Adaptive Finite impulse response Fisher Feedback exploration (AF3E). Namely, we will show that, at the be-

[1] It is also referred as bet on the best.

ginning of each interval, we can approximate the regret minimization problem as a SDP. However, the computational power required by AF3E can be very large which prevents its real-life implementation. Hence, we develop two theoretical results which allow to considerably reduce the computation time required for AF3E in Section 5. The first result (Theorem 1) allows us to simplify the objective function of the SDP. It is a direct consequence of the CE strategy and comes from a frequency domain property of the closed-loop transfer functions of the infinite horizon discrete-time LQ control which hitherto seems to have gone unnoticed. The second result, constituting the primary contribution of this paper, shows that, if an excitation signal has to be applied to minimize the regret over a certain amount of intervals, this excitation signal must only be applied in the first of the remaining intervals. This allows a strong simplification of the SDP as all the variables and linear matrix constraints (LMI) constraints related to the external excitation of the future intervals can be removed from the optimization problem. Furthermore, in certain instances, the current interval may not necessitate excitation, provided a specific inequality, verifiable prior to the SDP resolution, is satisfied. This novel result strongly resonates with the observations made in the numerical example of [15] and our recent theoretical results for the LQ regret minimization problem [12] where we show, under some strong approximations, that the optimal exploration strategy focuses all the exploration effort during the first time instant, i.e., future time instants are not excited. The main difference of the proposed scheme is the relaxation of the strong assumptions made in [12] which prevents its real-life implementation.

Despite the approximations made in this paper in order to reformulate the regret minimization as a SDP, we show in a numerical example in Section 7 that the proposed scheme can perform better than $1/\sqrt{t}$-decaying exploration [44], the Thompson sampling strategy [36] and IF2E [11]. Future perspectives are provided in Section 8.

2 Notations

Scalars, vectors and matrices: For any two integers $p \leq m$, the notation $[\![p, m]\!]$ refers to the set of consecutive integers between p and m. The notation j will refer to the complex number satisfying $j^2 = -1$. The set of real-valued matrices of dimension $n \times m$ will be denoted $\mathbb{R}^{n\times m}$. We will often use the notation n_x for the dimension of any vector x. The trace of any square matrix $\mathbf{A}$ will be denoted $\mathrm{tr}(\mathbf{A})$. When $\mathbf{A}$ is positive definite (respectively positive semi-definite), we will write $\mathbf{A} \succ 0$ (respectively $\mathbf{A} \succeq 0$). The identity matrix of dimension $n \times n$ will be denoted by $\mathbf{I}_n$ and $\mathbf{A}^\top$ (respectively $\mathbf{A}^*$) denotes the transpose (respectively the conjugate transpose) of any matrix $\mathbf{A}$. The minimal and maximal eigenvalue of any matrix $\mathbf{A}$ is denoted by $\underline{\lambda}(\mathbf{A})$ and $\overline{\lambda}(\mathbf{A})$ respectively. For a set of m scalars x_k $(k = 1, \cdots, m)$, the notation $\mathcal{T}(x_1, \cdots, x_m)$ refers to the Toeplitz symmetric such that the first row is given by $(x_1, \cdots, x_m)$.

Probability: The notation $x \sim \mathcal{N}(\mu, \mathbf{\Sigma})$ refers to the vector x of random variables which are jointly normally distributed with mean vector μ and covariance matrix $\mathbf{\Sigma}$. The expectation operator will be denoted by $\mathbb{E}$.

System variables: The discrete-time forward operator and the $\mathcal{Z}-$transform variable will be abusively referred by the same notation z and the discrete time variable will be denoted by t. Finally, when dealing with the frequency response of any discrete-time linear time invariant system, we will denote by ω the angular frequency.

Signal: For $\mathcal{L}_2$ summable discrete-time vector-valued signals x, $||x||_{\mathbf{Q}}$ is the weighted $\mathcal{L}_2$ norm with weighting matrix $\mathbf{Q} \succeq 0$ defined by

$$||x||^2_{\mathbf{Q}} = \lim_{T\to+\infty} \frac{1}{T} \sum_{t=1}^{T} \mathbb{E}[x(t)^\top \mathbf{Q} x(t)]$$

3 Context and literature

Consider a LTI discrete-time state system $\mathcal{S}$ with a vector x of n_x states and one input u given by

$$x(t+1) = \mathbf{A}_0 x(t) + \mathbf{B}_0 u(t) + e(t) \tag{1}$$

where e is the process noise vector and $\mathbf{A}_0 \in \mathbb{R}^{n_x\times n_x}$ and $\mathbf{B}_0 \in \mathbb{R}^{n_x\times 1}$ are the state matrices. All the states are assumed to be measured without error [2]. The process noise vector e is assumed to be zero-mean, white and normally distributed with a covariance matrix $\mathbf{\Sigma}_e \succ 0$. Finally, the pair of matrices $\{\mathbf{A}_0, \mathbf{B}_0\}$ is assumed controllable.

We want to control $\mathcal{S}$ with a linear quadratic controller minimizing the infinite-time horizon quadratic control cost $J_\infty(u) = \lim_{T\to+\infty} \sum_{t=1}^{T} \mathbb{E}[J_t(u)]/T$, where the instantaneous control cost $J_t(u)$ is defined by

$$J_t(u) = x(t)^\top \mathbf{Q} x(t) + \mathbf{R} u(t)^2$$

where $\mathbf{Q} \in \mathbb{R}^{n_x\times n_x}$ and $\mathbf{R} \in \mathbb{R}$ are user-defined matrices such that $\mathbf{Q} \succ 0$ and $\mathbf{R} > 0$. With the controllability assumption, there is an optimal gain vector $\mathbf{K}_0 \in \mathbb{R}^{1\times n_x}$ such that the control policy $u(t) = -\mathbf{K}_0 x(t)$ stabilizes the system while minimizing the cost $J_\infty(u)$. The optimal gain is given by

$$\mathbf{K}_0 = \left(\mathbf{R} + \mathbf{B}_0^\top \mathbf{P} \mathbf{B}_0\right)^{-1} \mathbf{B}_0^\top \mathbf{P} \mathbf{A}_0 \tag{2}$$

where $\mathbf{P} \in \mathbb{R}^{n_x\times n_x}$ is the unique positive definite solution to the following discrete-time algebraic Riccati equation (DARE)

[2] This is of course an ideal situation which was considered in many of the prior works such as [44,22,2,9,30]. In practice a measurement equation with added noise is present. The proposed scheme can be extended to that case. The same holds for the multiple inputs case. This will be reported elsewhere.

$$\mathbf{P} - \mathbf{A}_0^\top \mathbf{P}\mathbf{A}_0 = \mathbf{Q} + \mathbf{A}_0^\top \mathbf{P}\mathbf{B}_0(\mathbf{R} + \mathbf{B}_0^\top \mathbf{P}\mathbf{B}_0)^{-1}\mathbf{B}_0^\top \mathbf{P}\mathbf{A}_0 \quad (3)$$

Optimal control requires knowledge of the true state matrices $\mathbf{A}_0$ and $\mathbf{B}_0$ since the optimal gain is a function of these two quantities. Model based adaptive control addresses lack of this information by estimating these matrices from input and state data on-line. The process noise $e(t)$ precludes exact identification and we will refer to the estimates at time t by $\mathbf{A}(t)$ and $\mathbf{B}(t)$ for $\mathbf{A}_0$ and $\mathbf{B}_0$, respectively. In model-based adaptive control, the controller is updated on-line using new observed data. This update is done by following the following steps at each time instant t: (i) a new input/output data sample is acquired, (ii) the model is updated (often recursively), and in some cases also its uncertainty is updated, (iii) the controller, say $\mathbf{K}(t)$, is computed based on the newly identified model and (iv) the following control policy

$$u(t) = -\mathbf{K}(t)x(t) + v(t)$$

is applied and the process is re-started. The signal $v(t) \in \mathbb{R}$ is an user-defined external excitation for identification purposes. As mentioned in the introduction, there are two classes of approaches for the design of the adaptive controller $\mathbf{K}(t)$. In CE, the identified state matrices are considered to be the truth and the controller $\mathbf{K}(t)$ is designed accordingly from (2) and (3), with $\mathbf{A}(t)$ and $\mathbf{B}(t)$ replacing $\mathbf{A}_0$ and $\mathbf{B}_0$, respectively. In OFU, the controller $\mathbf{K}(t)$ is chosen among the controllers that can be designed based on the elements of the uncertainty region for $\{\mathbf{A}_0, \mathbf{B}_0\}$ as the one minimizing the expected control cost.

3.1 Regret minimization and literature

As explained in the introduction, the choice of the adaptive strategy (OFU or CE) and the external excitation v is important as it affects the trade-off between exploration (related to the control performance degradation caused by using the external excitation v) and exploitation (related to the control performance degradation caused by the model errors). The regret is a measure of this trade-off and the optimal trade-off is obtained by minimizing the cumulative regret. As in many works of the literature [44,22,43], we will consider the following cumulative expected regret $r(T)$

$$r(T) = \sum_{t=1}^{T} \mathbb{E}[J_t(u)] - \mathbb{E}[J_t(\tilde{u})] \quad (4)$$

where the expectation $\mathbb{E}$ is taken with respect to e and $\tilde{u}$ is the ideal control effort obtained with $\mathbf{K}(t) = \mathbf{K}_0 \ \forall t$ and without external excitation v.

Both in the asymptotic regime (T becoming large) and for the finite time horizon case the regret $r(T)$ is upper-bounded by $\mathcal{O}(\log(T))$ if $\mathbf{A}_0$ is known or if $\mathbf{B}_0$ is known [22,45]. If both $\mathbf{A}_0$ and $\mathbf{B}_0$ are unknown, then the regret is upper-bounded by $\mathcal{O}(\sqrt{T})$ [22,45,44].

In the literature, several adaptive control strategies have been proposed in order to reach the optimal growth rate of the regret. In this paragraph, we review three of them in some detail as they will be used later in a numerical example.

$1/\sqrt{t}$-decaying exploration. The work in [44,22,45] showed that an excitation of the form

$$v(t) \sim \mathcal{N}\left(0, \frac{a}{\sqrt{t}}\right) \qquad a \geq 0 \quad (5)$$

combined with the CE strategy guarantees an asymptotic cumulative regret growing as $\mathcal{O}(\sqrt{T})$.

The $1/\sqrt{t}$-decaying excitation can be seen as an open-loop strategy concerning the choice of external excitation. In [11,3,36] it is argued that it may be beneficial to use the Fisher information matrix to determine the magnitude of the excitation. There are two schemes in the literature that are based on this idea:

Inverse Fisher feedback exploration (IF2E). We proposed in [11] a zero-mean white noise Gaussian excitation for which the variance is adapted based on a feedback of the minimal eigenvalue of the inverse of the estimated covariance matrix $\mathbf{P}(t)$ of the identified state matrices $\mathbf{A}(t)$ and $\mathbf{B}(t)$ (see [11] for the expression), i.e.,

$$v(t) \sim \mathcal{N}\left(0, \frac{b}{\underline{\lambda}(\mathbf{P}(t)^{-1})}\right) \qquad b \geq 0 \quad (6)$$

The main drawback of $1/\sqrt{t}$-decaying excitation and IF2E is that they employ white noise excitation which from an optimal experiment design perspective may not be the optimal excitation because there is no correlation. Thus, even if the optimal regret rate and the optimal value for a and b can be determined [3], the regret may still be further reduced with a correlated excitation signal.

Thompson sampling exploration. Several Thompson sampling strategies have been developed in [3,36,25] for LQ regret minimization. At specific time instants, a parameter vector is sampled from the uncertainty ellipsoid and the LQ controller is designed based on this sample. This scheme does not add an additional external v. The schemes in [36,25] provide a regret $r(T)$ asymptotically scaling as $\mathcal{O}(\sqrt{T})$ and [36] has the nice advantage to not depend on some hyperparameters.

As mentioned in [39], the approach with Thompson sampling is not suited for reinforcement learning problems which do not require an active exploration, i.e., an exploration which never stops. However, as was observed for *finite-time* LQ regret minimization in [12] or in [15] for *finite-time* $\mathcal{H}_2$ based adaptive control, the optimal exploration strategy only excites the system at the beginning of the experiment. Thus, active exploration strategies may not be optimal for the considered problem and so Thompson sampling approaches may not be adequate in that case.

[3] Notice that the optimal value for a and b depends on the true parameter vector θ_0.

In this contribution, we propose an approach which uses the system information, as measured by the Fisher information matrix, to determine the excitation. Thus we build on IF2E but with the following important improvement which consists in adding correlation in the external excitation. We will call this strategy AF3E for Adaptive FIR Fisher Feedback Exploration.

4 Adaptive FIR Fisher Feedback Exploration (AF3E)

4.1 Pre-requirements

In this paragraph, we mention what is required in order to use the proposed scheme.

First, we assume that we know a model structure $\mathcal{M} = \{\{\mathbf{A}(\theta), \mathbf{B}(\theta)\} \mid \theta \in \mathbb{R}^{n_\theta}\}$ parametrized by a vector θ containing n_θ parameters. Moreover, we assume $\mathcal{M}$ to be full-order, i.e., there exists a true parameter vector $\theta_0 \in \mathbb{R}^{n_\theta}$ such that $\mathbf{A}(\theta_0) = \mathbf{A}_0$ and $\mathbf{B}(\theta_0) = \mathbf{B}_0$.

Remark 1. If the user does not know a suitable model structure, both matrices $\mathbf{A}(\theta)$ and $\mathbf{B}(\theta)$ can be chosen such that each of their entries is independently parametrized. In that case, θ contains $n_\theta = n_x(n_x + 1)$ parameters. □

From this model structure, we introduce the notation $\mathbf{K}(\theta)$ which is the LQ controller row vector obtained by (2) and (3) replacing $\mathbf{A}_0$, $\mathbf{B}_0$ and $\mathbf{P}_0$ by $\mathbf{A}(\theta)$, $\mathbf{B}(\theta)$ and $\mathbf{P}(\theta)$ respectively.

Finally, we will assume that we have an initial estimate θ_{init} which is normally distributed around θ_0 with a covariance matrix $\mathbf{P}_{init}$ which is also assumed known.

Remark 2. If the user does not know an initial estimate $\hat{\theta}_{init}$ and/or its corresponding covariance matrix $\mathbf{P}_{init}$, an initial identification experiment of long duration [4] can be performed in order to compute them so that the approximation $\hat{\theta}_{init} \sim N(\theta_0, \mathbf{P}_{init}^{-1})$ holds. □

4.2 Design by intervals

Under the certainty equivalence strategy, the regret minimization of the adaptive LQ control problem becomes a time-varying experiment design problem where the aim is to compute the optimal sequence $\{v(t)\}_{t=1}^{T}$ such that the regret $r(T)$ is minimized. This optimization problem is non-convex which makes it computationally intractable to be solved.

Inspired from the work in [15], we are going to approximate the regret minimization problem as a convex experiment design optimization problem. The idea in [15] is to divide the experiment interval $[\![1, T]\!]$ into L epochs of equal duration N so that $T = LN$. During each epoch time-invariance and stationarity is ensured by keeping the controller constant and using a realization of a stationary stochastic process as external excitation. We respectively denote by e_k, v_k, u_k, x_k, $\tilde{u}_k$ and $\tilde{x}_k$ the noise, the external excitation, the input, the state, the ideal input and the ideal state of the k-th interval. The variable $\tau \in [\![1, N]\!]$ will be used to index the time instants of each interval. With this notation, we have, e.g., $x((k-1)N + \tau) = x_k(\tau)$. We can also split the cumulative regret $r(T)$ into the sum of sub-regrets r_k obtained during the k-th interval

$$r(T) = \sum_{k=1}^{L} r_k$$
$$r_k = \sum_{\tau=1}^{N} \mathbb{E}[x_k^\top(\tau)\mathbf{Q}x_k(\tau) + \mathbf{R}u_k^2(\tau) - \tilde{x}_k^\top(\tau)\mathbf{Q}\tilde{x}_k(\tau) - \mathbf{R}\tilde{u}_k^2(\tau)]$$

Denote by $\mathbf{K}(\hat{\theta}_k)$ the constant controller used in the feedback of the system (1) during interval k, i.e.,

$$u_k(\tau) = -\mathbf{K}(\hat{\theta}_k)x_k(\tau) + v_k(\tau)$$

At the beginning of each interval $k \geq 2$, the controller is computed using the CE strategy, based on the least-squares estimate $\hat{\theta}_k$ of θ_0 computed at the beginning of the interval k by using all the past input-state data of the previous intervals. Each $\hat{\theta}_k$ with $k \geq 2$ is given by

$$\hat{\theta}_k = \underset{\theta \in \mathbb{R}^{n_\theta}}{\mathbf{arg\,min}} \sum_{l=1}^{k-1} \sum_{\tau=1}^{N} \varepsilon_l(\tau, \theta)^\top \mathbf{\Sigma}_e^{-1} \varepsilon_l(\tau, \theta) \quad (7)$$

with $\varepsilon_l(\tau, \theta) = x_l(\tau + 1) - \mathbf{A}(\theta)x_l(\tau) - \mathbf{B}(\theta)u_l(\tau)$, $\hat{\theta}_1 = \hat{\theta}_{init}$. The initial state $x_1(1) = x(1)$ and noise covariance matrix $\mathbf{\Sigma}_e$ will be assumed to be known for simplicity [5]. For each v_k, we consider a white noise excitation with unit variance, independent from e_k and which is filtered by an arbitrary finite impulse response (FIR) filter $F_k(z)$ of order m allowing us to introduce some correlation in the external excitation v_k. Hence, the decision variables for regret minimization are the filters $\{F_k(z)\}_{k=1}^{L}$.

Two important assumptions will be considered in order to rewrite the regret minimization problem as a convex experiment design problem.

Assumption 1. For the design of the FIR filters $\{F_k(z)\}_{k=1}^{L}$, we will make the assumption that all the identified pairs of matrices $\{\mathbf{A}(\hat{\theta}_k), \mathbf{B}(\hat{\theta}_k)\}_{k=1}^{L}$ are controllable and all the corresponding CE controllers $\{\mathbf{K}(\hat{\theta}_k)\}_{k=1}^{L}$ stabilize the loop. We will explain later in Section 6 how to deal in practice with this assumption.

Assumption 2. The length N of the epochs is assumed to be sufficiently large that transients can be neglected and that the asymptotic theory for parameter estimation in linear dynamical systems (as given in Chapter 9 in [33]) is applicable.

Remark 3. While Assumption 2 may seem restrictive at this point, as we will see in the numerical example in Section 7, the performance of the resulting algorithm is competitive with state-of-the-art methods and this also for small epoch lengths.

[4] Under some mild assumptions [33], the estimate $\hat{\theta}_{init}$ is *asymptotically* normally distributed around θ_0 with covariance $\mathbf{P}_{init}$. Hence, an identification experiment with a large number of data can approximate well the normal assumption $\hat{\theta}_{init} \sim N(\theta_0, \mathbf{P}_{init}^{-1})$ of the estimate.

[5] The initial state $x_1(1) = x(1)$ and the covariance matrix $\mathbf{\Sigma}_e$ can also be estimated together with $\hat{\theta}_k$ [33].

Under Assumption 2 and the filtered white noise choice for v_k, we have that $\hat{\theta}_k$ is asymptotically normally distributed around θ_0 with a covariance matrix equal to the inverse of $\boldsymbol{\mathcal{I}}_k$ where $\boldsymbol{\mathcal{I}}_k$ is the Fisher information matrix satisfying the following additive property $\boldsymbol{\mathcal{I}}_k = \boldsymbol{\mathcal{I}}_{k-1} + \boldsymbol{\mathcal{L}}_{k-1}$ with $\boldsymbol{\mathcal{I}}_1 = \mathbf{P}_{init}^{-1}$. The matrix $\boldsymbol{\mathcal{L}}_k$ corresponds to the additional Fisher information obtained during interval k. Under the assumption of white zero-mean Gaussian noise e with covariance matrix $\boldsymbol{\Sigma}_e \succ 0$, $\boldsymbol{\mathcal{L}}_k$ is given by [33]

$$\boldsymbol{\mathcal{L}}_k = N\mathbb{E}\left[\frac{\partial \varepsilon_k(\tau,\theta)}{\partial\theta}\boldsymbol{\Sigma}_e^{-1}\left(\frac{\partial \varepsilon_k(\tau,\theta)}{\partial\theta}\right)^\top\right]|_{\theta=\theta_0}$$

4.3 Power spectrum parametrization

Recall that the considered decision variables for regret minimization are the filters $\{F_k(z)\}_{k=1}^L$ of order m. Then, the PSD ϕ_{v_k} of the excitation signals v_k have the following form [21]

$$\phi_{v_k}(\omega) = c_0(k) + 2\sum_{q=1}^{m} c_q(k)\cos(q\omega) \tag{8}$$

where $c_q(k)$ $(q = 0, \cdots, m,\ k = 1, \cdots, L)$ are the scalar parameters to be tuned. Spectral factorization techniques compute the FIR filter $F_k(z)$ from $\phi_{v_k}(\omega)$ [40]. Since ϕ_{v_k} is affine with respect to $c_q(k)$, we will show that we can approximate the regret minimization problem as a SDP with $c_q(k)$ $(q = 0, \cdots, m,\ k = 1, \cdots, L)$ as decision variables.

Nevertheless, in order to guarantee the existence of a proper $F_k(z)$ from $\phi_{v_k}(\omega)$, we need to guarantee the non-negativity of the PSD, i.e., $\phi_{v_k}(\omega) \geq 0\ \forall\omega \in]-\pi, \pi]$. With the parametrization (8), a necessary and sufficient condition for $\phi_{v_k}(\omega) \geq 0\ \forall\omega \in]-\pi, \pi]$ when $m > 0$ is the existence of a symmetric matrix $\mathbf{X}(k)$ of dimension $m \times m$ such that the following linear matrix inequality (LMI) is guaranteed [21]

$$\begin{pmatrix} \mathbf{X}(k) - \mathbf{Y}^\top\mathbf{X}(k)\mathbf{Y} & c_{1:m}(k) - \mathbf{Y}^\top\mathbf{X}(k)\mathbf{Z} \\ c_{1:m}^\top(k) - \mathbf{Z}^\top\mathbf{X}(k)\mathbf{Y} & c_0(k) - \mathbf{Z}^\top\mathbf{X}(k)\mathbf{Z} \end{pmatrix} \succeq 0 \tag{9}$$

where $c_{1:m}^\top(t) = (c_1(k), \cdots, c_m(k))$ and

$$\mathbf{Y} = \begin{pmatrix} \mathbf{0}_{1\times(m-1)} & 0 \\ \mathbf{I}_{m-1} & 0_{(m-1)\times 1} \end{pmatrix} \qquad \mathbf{Z} = \begin{pmatrix} 1 \\ \mathbf{0}_{(m-1)\times 1} \end{pmatrix}$$

This comes from an application of the positive real lemma [1], a particular case of the KYP lemma [37]. When $m = 0$, this condition simply becomes $c_0(k) \geq 0$.

In the next paragraph, we show how to reformulate each sub-regret r_k as an explicit expression of the parameters $c_q(k)$ under Assumption 2. For this purpose, we introduce additional notations. For any θ and controller row vector $\mathbf{K}$ stabilizing $\{\mathbf{A}(\theta), \mathbf{B}(\theta)\}$, we introduce the following closed loop transfer function notation where subscript zw denote the transfer function from w to z:

$$\begin{aligned}
\mathbf{T}_{xe}(z,\theta,\mathbf{K}) &= (z\mathbf{I}_{n_x} - (\mathbf{A}(\theta) - \mathbf{B}(\theta)\mathbf{K}))^{-1} \\
\mathbf{T}_{ue}(z,\theta,\mathbf{K}) &= -\mathbf{K}\mathbf{T}_{xe}(z,\theta,\mathbf{K}) \\
\mathbf{T}_{xv}(z,\theta,\mathbf{K}) &= \mathbf{T}_{xe}(z,\theta,\mathbf{K})\mathbf{B}(\theta) \\
\mathbf{T}_{uv}(z,\theta,\mathbf{K}) &= 1 - \mathbf{K}\mathbf{T}_{xv}(z,\theta,\mathbf{K})
\end{aligned}$$

Consequently, the ideal state $\tilde{x}$ and input $\tilde{u}$ are given by $\tilde{x}(t) = \mathbf{T}_{xe}(z,\theta_0,\mathbf{K}_0)e(t)$ and $\tilde{u}(t) = \mathbf{T}_{ue}(z,\theta_0,\mathbf{K}_0)e(t)$.

4.4 Asymptotic approximation

Under Assumption 2, the stability of the closed-loop on each interval k (Assumption 1) and the stationary assumption of both e_k and v_k, each sub-regret r_k can be approximated by $r_k \approx N\mathbb{E}[||x_k||_{\mathbf{Q}}^2 + ||u_k||_{\mathbf{R}}^2 - ||\tilde{x}_k||_{\mathbf{Q}}^2 - ||\tilde{u}_k||_{\mathbf{R}}^2]$. Moreover, the contribution of the transient time dynamics on the above $\mathcal{L}_2$ norm terms is negligible. Hence, we have $||x_k||_{\mathbf{Q}}^2 \approx ||\bar{x}_k||_{\mathbf{Q}}^2$ and $||u_k||_{\mathbf{R}}^2 \approx ||\bar{u}_k||_{\mathbf{R}}^2$ where $\bar{x}_k$ and $\bar{u}_k$ are the state and input obtained from the following transfer function representation with zero initial conditions

$$\begin{aligned}
\bar{x}_k(\tau) &= \mathbf{T}_{xv}(z,\theta_0,\mathbf{K}(\hat{\theta}_k))v_k(\tau) + \mathbf{T}_{xe}(z,\theta_0,\mathbf{K}(\hat{\theta}_k))e_k(\tau) \\
\bar{u}_k(\tau) &= \mathbf{T}_{uv}(z,\theta_0,\mathbf{K}(\hat{\theta}_k))v_k(\tau) + \mathbf{T}_{ue}(z,\theta_0,\mathbf{K}(\hat{\theta}_k))e_k(\tau)
\end{aligned}$$

Recalling that the ideal state $\tilde{x}$ and input $\tilde{u}$ are given by $\tilde{x}_k(t) = \mathbf{T}_{xe}(z,\theta_0,\mathbf{K}_0)e_k(t)$ and $\tilde{u}_k(t) = \mathbf{T}_{ue}(z,\theta_0,\mathbf{K}_0)e_k(t)$ and both e_k and v_k are independent, we can rewrite r_k as follows

$$\begin{aligned}
r_k &\approx N\left(r_k^e(\theta_0,\mathbf{K}(\hat{\theta}_k)) + r_k^v(\theta_0,\mathbf{K}(\hat{\theta}_k))\right) \\
r_k^e(\theta_0,\mathbf{K}(\hat{\theta}_k)) &= \mathbb{E}[||\mathbf{T}_{xe}(z,\theta_0,\mathbf{K}(\hat{\theta}_k))e_k||_{\mathbf{Q}}^2] \\
&\quad + \mathbb{E}[||\mathbf{T}_{ue}(z,\theta_0,\mathbf{K}(\hat{\theta}_k))e_k||_{\mathbf{R}}^2] \\
&\quad - \mathbb{E}[||\mathbf{T}_{xe}(z,\theta_0,\mathbf{K}(\theta_0))e_k||_{\mathbf{Q}}^2] \\
&\quad - \mathbb{E}[||\mathbf{T}_{ue}(z,\theta_0,\mathbf{K}(\theta_0))e_k||_{\mathbf{R}}^2] \qquad (10) \\
r_k^v(\theta_0,\mathbf{K}(\hat{\theta}_k)) &= \mathbb{E}[||\mathbf{T}_{xv}(z,\theta_0,\mathbf{K}(\hat{\theta}_k))v_k||_{\mathbf{Q}}^2] \\
&\quad + \mathbb{E}[||\mathbf{T}_{uv}(z,\theta_0,\mathbf{K}(\hat{\theta}_k))v_k||_{\mathbf{R}}^2] \qquad (11)
\end{aligned}$$

Here, each sub-regret r_k is split into two terms. The term $r_k^v(\theta_0,\mathbf{K}(\hat{\theta}_k))$ is an increasing function with respect to the power $c_0(k)$ of the external excitation v_k, i.e., it is the control performance degradation due to the use of an external excitation. It will be referred as the exploration sub-regret. The term $r_k^e(\theta_0,\mathbf{K}(\hat{\theta}_k))$, called exploitation sub-regret, is the difference between the LQ cost for the rejection of the disturbance e using the controller $\mathbf{K}(\hat{\theta}_k)$ and the optimal controller $\mathbf{K}_0 = \mathbf{K}(\theta_0)$, i.e., it is the control degradation performances due to the uncertainty of $\hat{\theta}_k$. Let us now rewrite both sub-regrets $r_k^v(\theta_0,\mathbf{K}(\hat{\theta}_k))$ and $r_k^e(\theta_0,\mathbf{K}(\hat{\theta}_k))$ so that they depend explicitly on the decision variables $c_q(k)$ of the PSD parametrization (8).

4.5 Rewriting of the exploitation sub-regret

Firstly, observe that the function $\hat{\theta}_k \to r_k^e(\theta_0,\mathbf{K}(\hat{\theta}_k))$ takes its minimum 0 at θ_0 (LQ control property). Hence, its gradient evaluated at θ_0 is 0 and its Hessian matrix

evaluated at θ_0, denoted by $\mathbf{W}(\theta_0)$, is positive semi-definite. By performing a Taylor approximation up to the second order of $\hat{\theta}_k \to r_k^e(\theta_0, \mathbf{K}(\hat{\theta}_k))$ evaluated at θ_0, we get $r_k^e(\theta_0, \mathbf{K}(\hat{\theta}_k)) \approx \mathbb{E}[(\hat{\theta}_k - \theta_0)^\top \mathbf{W}(\theta_0)(\hat{\theta}_k - \theta_0)]/2$. The computation of the Hessian matrix $\mathbf{W}(\theta_0)$ can be done using finite differentiation. Since $\hat{\theta}_k \sim N(\theta_0, \boldsymbol{\mathcal{I}}_k^{-1})$, we get

$$r_k^e(\theta_0, \mathbf{K}(\hat{\theta}_k)) \approx \frac{1}{2} \operatorname{tr}(\mathbf{W}(\theta_0)\boldsymbol{\mathcal{I}}_k^{-1}) \tag{12}$$

where $\boldsymbol{\mathcal{I}}_k$ is the Fisher information matrix. From our assumptions of stationarity and stability during one epoch, the per-epoch Fisher information matrix $\boldsymbol{\mathcal{L}}_k$in (13) is given by (9.54) in [33]. Combined with the PSD parametrization (8), it follows that the per-epoch Fisher information matrices $\boldsymbol{\mathcal{L}}_k$ can be rewritten as

$$\begin{aligned}\boldsymbol{\mathcal{L}}_k = N\boldsymbol{\mathcal{L}}^e(\theta_0, \mathbf{K}(\hat{\theta}_k)) + Nc_0(k)\boldsymbol{\mathcal{L}}_0^v(\theta_0, \mathbf{K}(\hat{\theta}_k)) \\ + N\sum_{q=0}^{m} c_q(k)\left(\boldsymbol{\mathcal{L}}_q^v(\theta_0, \mathbf{K}(\hat{\theta}_k)) + \boldsymbol{\mathcal{L}}_q^v(\theta_0, \mathbf{K}(\hat{\theta}_k))^\top\right)\end{aligned} \tag{13}$$

for some matrix $\boldsymbol{\mathcal{L}}^e(\theta_0, \mathbf{K}(\hat{\theta}_k))$ and some matrices $\{\boldsymbol{\mathcal{L}}_q^v(\theta_0, \mathbf{K}(\hat{\theta}_k))\}_{q=0}^m$. Indeed, as shown in [7], the matrix $\boldsymbol{\mathcal{L}}_k$ defined at the end of Section 4.2 is the sum of a contribution of the noise e (i.e., $N\boldsymbol{\mathcal{L}}^e(\theta_0, \mathbf{K}(\hat{\theta}_k))$) and a term of the form

$$\frac{N}{2\pi}\int_{-\pi}^{\pi} \boldsymbol{\Gamma}(e^{j\omega})\phi_{v_k}(\omega)d\omega \tag{14}$$

for some matrix $\boldsymbol{\Gamma}(e^{j\omega})$ which is positive semi-definite at all $\omega \in]-\pi, \pi]$ (see [7]). The matrices $\{\boldsymbol{\mathcal{L}}_q^v(\theta_0, \mathbf{K}(\hat{\theta}_k))\}_{q=0}^m$ are thus the first $m+1$ elements of the Markov expansion of $\boldsymbol{\Gamma}(e^{j\omega}, \theta_0, \mathbf{K}(\hat{\theta}_k))$. The expression (13) thus provides a linear parametrization of the Fisher information matrix $\boldsymbol{\mathcal{I}}_k$.

Next, we will reformulate the minimization of $r_k^e(\theta_0, \mathbf{K}(\hat{\theta}_k))$ given by (12) into an SDP following the standard procedure in optimal experiment design [21]. Denote by $\tilde{\mathbf{W}}(\theta_0)$ the square root matrix of $\mathbf{W}(\theta_0)/2$. By introducing the matrices $\mathbf{H}(k)$ such that $\mathbf{H}(k) \succeq \tilde{\mathbf{W}}(\theta_0)\boldsymbol{\mathcal{I}}_k^{-1}\tilde{\mathbf{W}}(\theta_0)$ and using Schur complement, minimizing $r_k^e(\theta_0, \mathbf{K}(\hat{\theta}_k)) \approx \operatorname{tr}(\mathbf{W}(\theta_0)\boldsymbol{\mathcal{I}}_k^{-1})/2 = \operatorname{tr}(\tilde{\mathbf{W}}(\theta_0)\boldsymbol{\mathcal{I}}_k^{-1}\tilde{\mathbf{W}}(\theta_0))$ is equivalent to minimizing $\operatorname{tr}(\mathbf{H}(k))$ such that the following positive semidefinite inequality holds

$$\begin{pmatrix} \mathbf{H}(k) & \tilde{\mathbf{W}}(\theta_0) \\ \tilde{\mathbf{W}}(\theta_0) & \boldsymbol{\mathcal{I}}_1 + \sum_{i=1}^{k-1} \boldsymbol{\mathcal{L}}_i \end{pmatrix} \succeq 0 \tag{15}$$

Injecting the linear parametrization (13) of the per-epoch Fisher information matrices $\boldsymbol{\mathcal{L}}_k$ in (18) yields a LMI with respect to $\mathbf{H}(k)$ and $c_q(k)$. Consequently, we get a SDP reformulation of the minimization of the regret caused by the model uncertainties.

4.6 Rewriting of the exploration sub-regret

Let us now consider the exploration sub-regret $r_k^v(\theta_0, \mathbf{K}(\hat{\theta}_k))$ in (11). Using Parseval's theorem, $r_k^v(\theta_0, \mathbf{K}(\hat{\theta}_k))$ can be rewritten as follows $r_k^v(\theta_0, \mathbf{K}(\hat{\theta}_k)) = 1/2\pi \int_{-\pi}^{\pi} \mathbb{E}[\mathcal{D}(e^{j\omega}, \theta_0, \mathbf{K}(\hat{\theta}_k))]\phi_{v_k}(\omega)d\omega$ where the integrand $\mathcal{D}(e^{j\omega}, \theta, \mathbf{K}(\hat{\theta}_k))$ is defined by

$$\begin{aligned}\mathcal{D}(e^{j\omega}, \theta, \mathbf{K}(\hat{\theta}_k)) = \mathbf{T}_{xv}^*(e^{j\omega}, \theta, \mathbf{K}(\hat{\theta}_k))\mathbf{Q}\mathbf{T}_{xv}(e^{j\omega}, \theta, \mathbf{K}(\hat{\theta}_k)) \\ + \mathbf{T}_{uv}^*(e^{j\omega}, \theta, \mathbf{K}(\hat{\theta}_k))\mathbf{R}\mathbf{T}_{uv}(e^{j\omega}, \theta, \mathbf{K}(\hat{\theta}_k))\end{aligned} \tag{16}$$

With the parametrization (8) of the PSDs ϕ_{v_k}, each $r_k^v(\theta_0, \mathbf{K}(\hat{\theta}_k))$ becomes

$$\begin{aligned} r_k^v(\theta_0, \mathbf{K}(\hat{\theta}_k)) &= \beta_0(\theta_0, \mathbf{K}(\hat{\theta}_k))c_0(k) + 2\sum_{q=1}^{m} \beta_q(\theta_0, \mathbf{K}(\hat{\theta}_k))c_q(k) \\ \beta_q(\theta_0, \mathbf{K}(\hat{\theta}_k)) &= \frac{1}{2\pi}\int_{-\pi}^{\pi} \mathbb{E}[\mathcal{D}(e^{j\omega}, \theta_0, \mathbf{K}(\hat{\theta}_k))]\cos(q\omega)d\omega \end{aligned}$$

i.e., we obtain a linear parametrization in terms of $\{c_q(k)\}$ of the regret caused by the exploration. However, we still have the expectation operator in the above expression. As was done in the previous paragraph for the exploitation sub-regret, we could perform another Taylor expansion up to the second order of each $\hat{\theta}_k \to \beta_q(\theta_0, \mathbf{K}(\hat{\theta}_k))$ and evaluated at θ_0 in order to have an approximation of the expected value which depends on $\boldsymbol{\mathcal{I}}_k$. However, as pointed out in [15], we obtain an expression which is not convex in the decision variables. Hence, we will do the following approximation by removing the expectation operator

$$\beta_q(\theta_0, \mathbf{K}(\hat{\theta}_k)) \approx \frac{1}{2\pi}\int_{-\pi}^{\pi} \mathcal{D}(e^{j\omega}, \theta_0, \mathbf{K}(\hat{\theta}_k))\cos(q\omega)d\omega \tag{17}$$

Notice that the less uncertain the estimate $\hat{\theta}_k$, the more accurate the above approximation.

4.7 A SDP formulation of regret minimization

Using the above reformulation of both exploitation and exploration sub-regrets minimization, we show in this paragraph that we can rewrite the regret minimization problem as a SDP which is a convex optimization problem. Before that, we have two observations to make. Firstly, the exploitation sub-regret r_1^e of the first interval does not depend on any ϕ_{v_k}, so the LMI (18) for $k = 1$ can be removed from the optimization problem. Secondly, we can also pose $v_L = 0$ (and thus $r_L^v = 0$) since an excitation signal in the last interval will only increase the regret while having no effect on the model accuracy (the model will indeed not be updated at the end of the L-th interval in the chosen setting). Consequently, we can set $c_q(L) = 0$. Hence, combining both regret reformulations and re-indexing the variables $\mathbf{H}(k)$ such that $\mathbf{H}(k) \leftarrow \mathbf{H}(k+1)$, the final optimization problem of minimizing the cumulative regret $r(T)$ can be ap-

proximately rewritten as

$$\begin{aligned}
&\min_{\substack{c_q(k),\mathbf{X}(k)\\ \mathbf{H}(k)}} \textstyle\sum_{k=1}^{L-1} \left(\mathrm{tr}(\mathbf{H}(k)) + r_k^v\right)\\
&\textbf{Subject to}, \text{ for all } k \in [\![1, L-1]\!],\\
&r_k^v = \beta_0(\theta_0, \mathbf{K}(\hat{\theta}_k))c_0(k) + 2\textstyle\sum_{q=1}^m \beta_q(\theta_0, \mathbf{K}(\hat{\theta}_k))c_q(k)\\
&\begin{pmatrix} \mathbf{H}(k) & \tilde{\mathbf{W}}(\theta_0) \\ \tilde{\mathbf{W}}(\theta_0) & \mathbf{P}_{init}^{-1} + \sum_{i=1}^k \boldsymbol{\mathcal{L}}_i \end{pmatrix} \succeq 0\\
&\boldsymbol{\mathcal{L}}_i = N\boldsymbol{\mathcal{L}}^e(\theta_0, \mathbf{K}(\hat{\theta}_i)) + Nc_0(i)\boldsymbol{\mathcal{L}}_0^v(\theta_0, \mathbf{K}(\hat{\theta}_i))\\
&\quad + N\textstyle\sum_{q=1}^m c_q(i)\boldsymbol{\mathcal{N}}_q^v(\theta_0, \mathbf{K}(\hat{\theta}_i))\\
&\begin{pmatrix} \mathbf{X}(l) - \mathbf{Y}^\top\mathbf{X}(l)\mathbf{Y} & c_{1:m}(l) - \mathbf{Y}^\top\mathbf{X}(l)\mathbf{Z} \\ c_{1:m}^\top(l) - \mathbf{Z}^\top\mathbf{X}(l)\mathbf{Y} & c_0(l) - \mathbf{Z}^\top\mathbf{X}(l)\mathbf{Z} \end{pmatrix} \succeq 0
\end{aligned} \tag{18}$$

where $\boldsymbol{\mathcal{N}}_q^v((\theta_0, \mathbf{K}(\hat{\theta}_i)) = \boldsymbol{\mathcal{L}}_q^v((\theta_0, \mathbf{K}(\hat{\theta}_i)) + (\boldsymbol{\mathcal{L}}_q^v((\theta_0, \mathbf{K}(\hat{\theta}_i)))^\top$. This problem is a SDP for which efficient solvers exist such as MOSEK [5]. However, it cannot be solved in its current form since several terms depend on the unknown true parameter vector θ_0 and the future estimates $\{\hat{\theta}_k\}_{k=2}^L$. We could replace them by the initial (known) estimate $\hat{\theta}_1 = \hat{\theta}_{init}$. As a consequence, the optimal PSDs obtained after solving the corresponding SDP may not be appropriate if $\hat{\theta}_1$ is highly uncertain since all the aforementioned approximations of the regret minimization problem may not hold.

Fortunately, the future estimates $\hat{\theta}_k$ are expected to be less and less uncertain throughout the epochs since they are computed with more and more data. This suggest to consider a receding horizon strategy for the design of the PSDs as proposed in [15]. At the beginning of any interval p, we optimize the PSDs $\phi_{v_p}, \ldots, \phi_{v_{L-1}}$ minimizing the remaining regret $\sum_{k=p}^L r_k$ reformulated as a SDP following the steps done in the previous paragraphs. We replace θ_0 and the future estimates $\{\hat{\theta}_k\}_{k=p+1}^L$ by the estimate $\hat{\theta}_p$ computed at the beginning of interval p. In order to reduce the notation complexity, we will drop the arguments and abusively write β_q, $\tilde{\mathbf{W}}$, $\boldsymbol{\mathcal{L}}^e$ and $\boldsymbol{\mathcal{L}}_q^v$ instead of $\beta_q(\hat{\theta}_p, \mathbf{K}(\hat{\theta}_p))$, $\tilde{\mathbf{W}}(\hat{\theta}_p)$, $\boldsymbol{\mathcal{L}}^e(\hat{\theta}_p, \mathbf{K}(\hat{\theta}_p))$ and $\boldsymbol{\mathcal{L}}_q^v(\hat{\theta}_p, \mathbf{K}(\hat{\theta}_p))$. Then, the SDP to be solved at the beginning of interval p is given by

$$\begin{aligned}
&\min_{\substack{c_q(k),\mathbf{X}(k)\\ \mathbf{H}(k)}} \textstyle\sum_{k=p}^{L-1} \left(\mathrm{tr}(\mathbf{H}(k)) + \beta_0 c_0(k) + 2\sum_{q=1}^m \beta_q c_q(k)\right)\\
&\textbf{Subject to}, \text{ for all } k \in [\![p, L-1]\!],\\
&\begin{pmatrix} \mathbf{H}(k) & \tilde{\mathbf{W}} \\ \tilde{\mathbf{W}} & \boldsymbol{\mathcal{M}}_k + N\sum_{i=p}^k \left(c_0(i)\boldsymbol{\mathcal{L}}_0^v + \sum_{q=1}^m c_q(i)\boldsymbol{\mathcal{N}}_q^v\right) \end{pmatrix} \succeq 0\\
&\begin{pmatrix} \mathbf{X}(k) - \mathbf{Y}^\top\mathbf{X}(k)\mathbf{Y} & c_{1:m}(k) - \mathbf{Y}^\top\mathbf{X}(k)\mathbf{Z} \\ c_{1:m}^\top(k) - \mathbf{Z}^\top\mathbf{X}(k)\mathbf{Y} & c_0(k) - \mathbf{Z}^\top\mathbf{X}(k)\mathbf{Z} \end{pmatrix} \succeq 0
\end{aligned} \tag{19}$$

where, for any $k \geq p$, $\boldsymbol{\mathcal{M}}_k = \boldsymbol{\mathcal{I}}_p + N(k-p+1)\boldsymbol{\mathcal{L}}_e$ and, for any $q \geq 1$, $\boldsymbol{\mathcal{N}}_q^v = \boldsymbol{\mathcal{L}}_q^v + (\boldsymbol{\mathcal{L}}_q^v)^\top$. For $\boldsymbol{\mathcal{I}}_p$, we will consider the inverse of the estimated covariance matrix of $\hat{\theta}_p$ computed as described in [33]. Once the SDP is solved, we excite the system with a realization of the external excitation v_p obtained from the optimized PSD ϕ_{v_p} until the beginning of the next interval $p+1$. We then re-iterate the design process.

Even though it is now theoretically possible to implement AF3E, we can be highly concerned about its required computation time. Indeed, at the beginning of each interval p, we need to have sufficient computational power in order to realize several tasks which should last less than the sampling time: the computation of $\tilde{\mathbf{W}}$, $\boldsymbol{\mathcal{L}}_e$, $\boldsymbol{\mathcal{L}}_q^v$ and β_q, the solving of the SDP (19) and the spectral factorization of the optimal ϕ_{v_p}. Analyzing the complexity of the SDP, we observe that there are $2(L-p)$ constraints and, for the decision variables, there are $L-p$ symmetric matrices $\mathbf{H}(k)$ of dimension $n_\theta \times n_\theta$, $L-p$ symmetric matrices $\mathbf{X}(k)$ of dimension $(m+1)\times(m+1)$ and $(m+1)(L-p)$ scalars decision variable $c_q(k)$. Consequently, the complexity of the SDP is the largest for the first intervals and it might exceed the available processor sampling time. In the next section, we show that we can reduce the complexity of AF3E from two theoretical results which are the main contributions of this paper.

5 Theoretical results on AF3E

The first simplification relates to the exploration penalization coefficients β_q in (17). It comes from a property of the LQ control problem which seems to have gone unnoticed until now.

Theorem 1 (LQ frequency domain identity)**.** Consider any pair of controllable states matrices $\{\mathbf{A}, \mathbf{B}\}$ with n_x states and one input and any LQ matrix $\mathbf{Q} \succ 0$ of dimension $n_x \times n_x$ and scalar $\mathbf{R} > 0$. Denote by $\mathbf{P}$ the positive semidefinite solution of the corresponding DARE and by $\mathbf{K}$ the corresponding infinite horizon discrete-time LQ controller involved in the control policy $u(t) = -\mathbf{K}x(t) + v(t)$ where v is an additional external excitation. Then, the closed-loop transfer function matrices $\mathbf{T}_{xv}(z) = (z\mathbf{I}_{n_x} - (\mathbf{A} - \mathbf{B}\mathbf{K}))^{-1}\mathbf{B}$ and $\mathbf{T}_{uv}(z) = 1 - \mathbf{K}\mathbf{T}_{xv}(z)$ satisfies the following frequency domain property for any ω

$$\mathbf{T}_{xv}^*(e^{j\omega})\mathbf{Q}\mathbf{T}_{xv}(e^{j\omega}) + \mathbf{T}_{uv}^*(e^{j\omega})\mathbf{R}\mathbf{T}_{uv}(e^{j\omega}) = \mathbf{R} + \mathbf{B}^\top\mathbf{P}\mathbf{B}$$

PROOF. See Appendix A. ■

From Theorem 1, we conclude that the integrand $\mathcal{D}(e^{j\omega}, \hat{\theta}_p, \mathbf{K}(\hat{\theta}_p))$ in (16), involved in the computation of the coefficients β_q in (17) with both θ_0 and $\hat{\theta}_k$ replaced by $\hat{\theta}_p$, is constant and equal to $\mathbf{R} + \mathbf{B}(\hat{\theta}_p)^\top\mathbf{P}(\hat{\theta}_p)\mathbf{B}(\hat{\theta}_p)$. Hence, we have $\beta_0 = \mathbf{R} + \mathbf{B}(\hat{\theta}_p)^\top\mathbf{P}(\hat{\theta}_p)\mathbf{B}(\hat{\theta}_p)$ and $\beta_q = 0$ for every $q \geq 1$. This simplifies the objective function of the SDP (19). The next theorem, which is a significant contribution of this paper, provides the largest simplification of the computational complexity of AF3E. It shows that the solution of the SDP (19) has a strong sparsity.

Theorem 2 (The lazy/immediate excitation theorem)**.** Consider the SDP (19) at any interval p with $\beta_1 = \cdots = \beta_m = 0$ (consequence of Theorem 1). The optimal values of the decision variables $c_q(k)$ $(q = 0, \cdots, m)$ and $\mathbf{X}(k)$

are equal to 0 for $k = p+1, \cdots, L-1$. This implies $v_{p+1} = \cdots = v_{L-1} = 0$, i.e., the future intervals are not excited. The excitation v_p (i.e., the excitation for the current interval) is the only one that can be non-zero. This excitation v_p is nevertheless also equal to 0 if

$$\beta_0 > N\overline{\lambda}\left(\sum_{k=p}^{L-1}\mathcal{T}\left(\zeta_0(k),\cdots,\zeta_m(k)\right)\right) \tag{20}$$

where $\mathcal{T}(.)$ is defined as in Section 2 and $\zeta_q(k) = \text{tr}(\mathbf{W}\boldsymbol{\mathcal{M}}_k^{-1}\boldsymbol{\mathcal{L}}_q^v\boldsymbol{\mathcal{M}}_k^{-1})/2 \;\; \forall q \in [\![0,m]\!]$ with $\boldsymbol{\mathcal{M}}_k = \boldsymbol{\mathcal{I}}_p + N(k-p+1)\boldsymbol{\mathcal{L}}_e$.

PROOF. See Appendix B. ■

Theorem 2 implies that we can set all the decision variables related to $v_{p+1}, \cdots, v_L$ (i.e., $c_q(k)$ and $\mathbf{X}(k)$ for all $k \in [\![p+1, L-1]\!]$) equal to 0. Hence, the SDP is equivalent to the following reduced SDP

$$\begin{aligned}
&\min_{\substack{c_q(p),\mathbf{X}(p)\\ \mathbf{H}(k)}} \textstyle\sum_{k=p}^{L-1}\text{tr}(\mathbf{H}(k)) + \beta_0 c_0(p)\\
&\textbf{Subject to},\ \forall k \in [\![p, L-1]\!],\\
&\begin{pmatrix} \mathbf{H}(k) & \tilde{\mathbf{W}} \\ \tilde{\mathbf{W}} & \boldsymbol{\mathcal{M}}_k + N(c_0(p)\boldsymbol{\mathcal{L}}_0^v + \sum_{q=1}^m c_q(p)\boldsymbol{\mathcal{N}}_q^v)\end{pmatrix} \succeq 0 \qquad (21)\\
&\textbf{and}\\
&\begin{pmatrix} \mathbf{X}(p) - \mathbf{Y}^\top\mathbf{X}(p)\mathbf{Y} & c_{1:m}(p) - \mathbf{Y}^\top\mathbf{X}(p)\mathbf{Z} \\ c_{1:m}^\top(p) - \mathbf{Z}^\top\mathbf{X}(p)\mathbf{Y} & c_0(p) - \mathbf{Z}^\top\mathbf{X}(p)\mathbf{Z}\end{pmatrix} \succeq 0
\end{aligned}$$

Hence, only the current interval p might require an excitation (immediate exploration) depending on the validity of the constraint (20). If it it is satisfied, there is no need to explore during interval p (lazy exploration), i.e., $v_p = 0$. Since all the terms in this constraint can be computed, we can check a priori if we have to solve SDP (21) or if we can set v_p equal to 0. It is obviously faster to check this constraint than to solve the SDP. By analyzing (20), we conclude that it is not worth exploring if (i) the exploration coefficient β_0 is large, (ii) the number of remaining time instants $N(L-p)$ is small and (iii) $\boldsymbol{\mathcal{M}}_k$ is large and/or its main eigenvectors are perpendicular to the ones of the Hessian matrix $\mathbf{W}$.

Theorem 2 has some connections with some results in the literature. In the numerical example of [15], it was observed that the the exploration effort is only done during the first intervals. Moreover, by considering stronger assumptions than here, we recently showed in [12] that the optimal exploration strategy must be either focused during the first time instant (immediate exploration) or it must be set to 0 (lazy exploration).

6 Dealing with controllability and stability

Before evaluating the performances of AF3E with a numerical example, we need to come back to the stability and controllability assumption we made for the development of the results (Assumption 1).

Thanks to the receding horizon principle and our approach to deal with the chicken-an-egg issue, we only have to verify the controllability of $\{\mathbf{A}(\hat{\theta}_p), \mathbf{B}(\hat{\theta}_p)\}$ and the stability of the closed-loop with the CE controller $\mathbf{K}(\hat{\theta}_p)$ of the current interval p since the next ones are not present in the final SDP (21) we solve.

For the controllability test, we just have to check that the controllability matrix computed with $\{\mathbf{A}(\hat{\theta}_p), \mathbf{B}(\hat{\theta}_p)\}$ is full rank [4]. For the stability, we can analyze the robustness of the controller $\mathbf{K}(\hat{\theta}_p)$ with respect to an uncertain region containing the true system and that can be built using the fact that $\hat{\theta}_p \sim N(\theta_0, \boldsymbol{\mathcal{I}}_p^{-1})$ (see [8,15]).

If it happens that, for any interval p, the CE controller $\mathbf{K}(\hat{\theta}_p)$ fails the stability test or the controllability matrix obtained from $\{\mathbf{A}(\hat{\theta}_p), \mathbf{B}(\hat{\theta}_p)\}$ is not full rank, then the idea is to do the design of the PSDs using the estimate of the previous interval $p-1$. In other words, the controller used during interval p is $\mathbf{K}(\hat{\theta}_{p-1})$ and θ_0 and the future estimates $\{\hat{\theta}_k\}_{k=p+1}^L$ are replaced by $\hat{\theta}_{p-1}$ in the SDP. The motivation for this choice is that the results of Theorems 1 and 2 remain valid which keeps the scheme simple. Other alternatives will be investigated in the future such that robust stabilizing design of the controllers in the case when the CE controller fails the stability test.

Note that the aforementioned choice for dealing with controllability and stability requires that the initial estimate $\hat{\theta}_{init}$ is such that the initial CE controller $\mathbf{K}(\hat{\theta}_{init})$ stabilizes the loop and the controllability matrix obtained from $\{\mathbf{A}(\hat{\theta}_{init}), \mathbf{B}(\hat{\theta}_{init})\}$ is full rank. If it is not the case for the stability, an additional identification experiment must be done and experiment design tools such as [7] can be exploited in order to design the excitation such that the stability is achieved.

Remark 4. Because the initial controller stabilizes the loop, it will be rare that the CE controllers of the next intervals can destabilize the loop. For the controllability, if $\{\mathbf{A}_0, \mathbf{B}_0\}$ is controllable and the chosen model structure is full-order and linearly parametrized, then the controllability matrix will be most of the times full rank.

By considering both aforementioned tests in our scheme, we can finally construct the algorithm of AF3E. It is given in Algorithm 1.

7 Numerical example

7.1 System and control variables

For the system $\mathcal{S}$ in (1), we consider $n_x = 3$ states with the following state matrices

$$\mathbf{A}_0 = \begin{pmatrix} -0.390 & 0.370 & -0.570 \\ -0.250 & -0.780 & -0.080 \\ 1.320 & 0.250 & -0.130 \end{pmatrix} \quad \mathbf{B}_0 = \begin{pmatrix} 0.210 \\ 0 \\ 0 \end{pmatrix}$$

This system has one real pole at -0.790 and two complex conjugate poles which describe a resonance with resonance frequency equal to 1.883 rad/s (the sampling time is equal to 1s) and a damping ratio of 0.027. The two half power points are 1.831 rad/s and 1.933 rad/s.

Algorithm 1 AF3E

Require: LQ weighting matrix $\mathbf{Q} \succ 0$ and scalar $\mathbf{R} > 0$, FIR order m, initial estimate $\hat{\theta}_{init}$ and corresponding covariance matrix $\mathbf{P}_{init} \succ 0$ such that the pair of matrices $\{\mathbf{A}(\hat{\theta}_{init}), \mathbf{B}(\hat{\theta}_{init})\}$ and the CE controller $\mathbf{K}(\hat{\theta}_{init})$ passes the controllability and stability tests of Section 6 and initial state value $x(1)$.
Set $\bar{\theta} = \hat{\theta}_{init}$ and $\bar{\boldsymbol{\mathcal{I}}} = \mathbf{P}_{init}^{-1}$.
for $p = 1, \ldots, L$ **do**
At the beginning of the interval p (i.e., at $t = (p-1)N + 1$),

- Step 1: if $p > 1$, identify $\hat{\theta}_p$ as in (7), estimate its covariance matrix (see [33]) and set $\boldsymbol{\mathcal{I}}_p$ equal to its inverse.
- Step 2: compute the CE controller $\mathbf{K}(\hat{\theta}_p)$ obtained from (2) and (3) for which $\mathbf{A}_0$, $\mathbf{B}_0$ and $\mathbf{P}_0$ are replaced by $\mathbf{A}(\hat{\theta}_p)$, $\mathbf{B}(\hat{\theta}_p)$ and $\mathbf{P}(\hat{\theta}_p)$ respectively.
- Step 3: if $p > 1$, realize the stability test and controllability test mentioned in Section 6. If both test are successful, set $\bar{\theta} = \hat{\theta}_p$ and $\bar{\boldsymbol{\mathcal{I}}} = \boldsymbol{\mathcal{I}}_p$. Otherwise, set $\bar{\theta} = \hat{\theta}_{p-1}$ and $\bar{\boldsymbol{\mathcal{I}}} = \boldsymbol{\mathcal{I}}_{p-1}$
- Step 3: if $p = L$, set $v_p = 0$. Otherwise, do the following tasks
  - Step 3.a: compute the Hessian matrix $\mathbf{W}(\bar{\theta})$ evaluated at $\bar{\theta}$ of the function $\theta \to r_k^e(\bar{\theta}, \mathbf{K}(\theta))$ defined in (10), the square root matrix $\tilde{\mathbf{W}}$ of $\mathbf{W}(\bar{\theta})/2$, the Fisher matrices $\boldsymbol{\mathcal{L}}^e = \boldsymbol{\mathcal{L}}^e(\bar{\theta}, \mathbf{K}(\bar{\theta}))$ and $\boldsymbol{\mathcal{L}}_q^v = \boldsymbol{\mathcal{L}}_q^v(\bar{\theta}, \mathbf{K}(\bar{\theta}))$ (see [7]), the exploration coefficient $\beta_0 = \mathbf{R} + \mathbf{B}(\bar{\theta})^\top \mathbf{P}(\bar{\theta}) \mathbf{B}(\bar{\theta})$ and $\boldsymbol{\mathcal{M}}_k = \bar{\boldsymbol{\mathcal{I}}} + N(k-p+1)\boldsymbol{\mathcal{L}}_e$.
  - Step 3.b: check if (20) holds. If it is true, set $v_p = 0$. Otherwise, solve the SDP (21), use spectral factorization on the optimized PSD $\phi_{v_p}(\omega)$ and generate a realization of duration N from it for v_p.
- Step 4: apply the control effort $u_p(\tau) = -\mathbf{K}(\bar{\theta}) x_p(\tau) + v_p(\tau)$ to $\mathcal{S}$ until the beginning of the next interval.

end for

The noise covariance will be taken as $\boldsymbol{\Sigma}_e = \mathbf{I}_3$. For the LQ weighting matrices, we choose $\mathbf{Q} = \mathbf{I}_3$ and $\mathbf{R} = 10$. Consequently, the ideal controller is $\mathbf{K}_0 = (-0.081, 0.045, -0.161)$. For the model structure, we consider all the entries of $\mathbf{A}(\theta)$ and $\mathbf{B}(\theta)$ to be independently parametrized, i.e., there are $n_\theta = 12$ parameters to be identified in total. To compute (7), we here use a recursive least-squares algorithm as described in [33]. Such a recursive algorithm is of course not necessary for AF3E, but it is chosen since it is the algorithm used in the methods described in Section 3. We will consider $T = 100000$ for the horizon of regret minimization. The division in intervals will be such that $N = 1000$ and so there are in total $L = T/N = 100$ intervals.

7.2 Initial identification and computation details

For the initial estimate $\hat{\theta}_{init}$ and its corresponding covariance matrix $\mathbf{P}_{init}$, we perform an initial open-loop identification with 200 data and we consider a zero-mean white Gaussian noise of variance 0.1 for the input $u(t)$. The data informativity property, ensuring the uniqueness of the minimizer of the least square identification cost [33,17], is therefore guaranteed which implies that the initial covariance matrix $\mathbf{P}_{init}$ is invertible.

The simulation is performed with MATLAB R2021b on a computer equipped with the processor Intel(R) Core(TM) i5-8365U CPU, 1.60GHz, 4 cores and with 16.0GB of RAM. For AF3E, we solve the SDP in (19) using the interface YALMIP 4.0 [34] combined with the SDP solver of MOSEK 9.3 [5].

7.3 Complexity reduction with Theorems 1 and 2

Before analyzing the regret performances, we illustrate in this paragraph the computational complexity reduction brought by Theorems 1 and 2. For this purpose, we consider one noise realization from $t = 1$ till $t = T$ and we run the simplified AF3E scheme described in Algorithm 1 and the scheme for which we do not take into consideration the results from Theorems 1 and 2, i.e., at the beginning of each interval p, we also compute β_1, $\cdots$, β_q (which we know are equal to 0) and we solve the original SDP (19) with all the decision variables present. The obtained computation times are given in Table 1. Comparing the two columns in the table we see that the computational time is reduced up to a factor of 40 for the most demanding case where $m = 3$.

m	Computation time without Theorems 1 and 2	Computation time with Theorems 1 and 2
0	72.3s	4.1s
1	126.2s	5.7s
2	163.5s	5.6s
3	236.8s	5.9s

Table 1
Computation times obtained for one noise realization from $t = 1$ till $t = T$ with AF3E and different FIR orders m.

7.4 Monte Carlo simulation details

In this section, we compare AF3E with other exploration strategies mentioned in Section 3: Thompson sampling, IF2E and $1/\sqrt{t}$-decaying excitation. For the comparison, we perform 100 Monte-Carlo simulations with different realization for e, the initial zero-mean white noise input sequence as described in Section 7.2 and the white Gaussian noise used to generate the external excitation v for $1/\sqrt{t}$-decaying excitation, IF2E and AF3E. With these 100 simulations, we approximate the expectation operator $\mathbb{E}$ in the expression (4) of the cumulative regret $r(t)$ by computing the average of the 100 simu-

Method	$r(T)$	AF3E with θ_0
$1/\sqrt{t}$-decaying exploration	30393	-
IF2E	30224	-
Thompson sampling	28823	-
AF3E with $m = 0$	26490	24479
AF3E with $m = 1$	26333	24249
AF3E with $m = 2$	26290	23588
AF3E with $m = 3$	24973	22977

Table 2
Cumulative regret $r(T)$ obtained with $N = 1000$ and the different methods listed in descending order of $r(T)$ (first column) and for AF3E without replacing θ_0 in the SDP (21) (second column).

lated control cost degradation $\sum_{t=1}^{T}(J_t(u) - J_t(\tilde{u}))$. For comparison purposes, the hyperparameters a of $1/\sqrt{t}$-decaying excitation (5) and b of IF2E (6) are tuned by choosing the value minimizing the Monte Carlo estimate of the expected cumulative regret $r(T)$ among a log-regularly grid of 500 points between 10^{-3} and 10. This gives the smallest regret that one can get for the considered numerical example with these exploration strategies. For AF3E, we will consider $m = 0, 1, 2$ and 3 for the FIR order of the external excitation v.

7.5 Results and discussion

In Figure 1, we compare the time evolution of the Monte Carlo estimate of the expected cumulative regret obtained with the four explorations strategies. Because some exploration strategies give almost similar regret evolution, we give in the first column of Table 2 the final cumulative regret $r(T)$ for the different methods, listed in descending order of $r(T)$. We observe that the proposed scheme performs better than the optimal $1/\sqrt{t}$-decaying exploration, IF2E and Thompson sampling for the cumulative regret minimization at $t = T$ and for the considered FIR orders m. Increasing the FIR order m seems to improve the regret minimization.

Because we compared AF3E with the optimal IF2E and $1/\sqrt{t}$-decaying excitation, we also give in the second column of Table 2 the regret obtained with AF3E when, for each interval $p = 1, \cdots, L$, we do not replace θ_0 by $\hat{\theta}_p$ in the terms of the SDP. In that case (which is of course unrealistic since θ_0 would be unknown in practice), the design of the PSDs should be more adequate for regret minimization. Note that Theorem 1 does not hold anymore, i.e., the coefficients $\beta_1, \cdots, \beta_m$ are non-zero. Fortunately, it is easy to show that the sparsity of the solution still holds as in Theorem 2 with non-zero $\beta_1, \cdots, \beta_m$ by following the proof as we did in Appendix B. Comparing the first and second column of Table 2, we can observe a relative error of around 10% between the obtained regret and the best regret that AF3E can achieve. This error is reasonable showing that the simple approach of replacing θ_0 by $\hat{\theta}_p$ is a good practice for this numerical example.

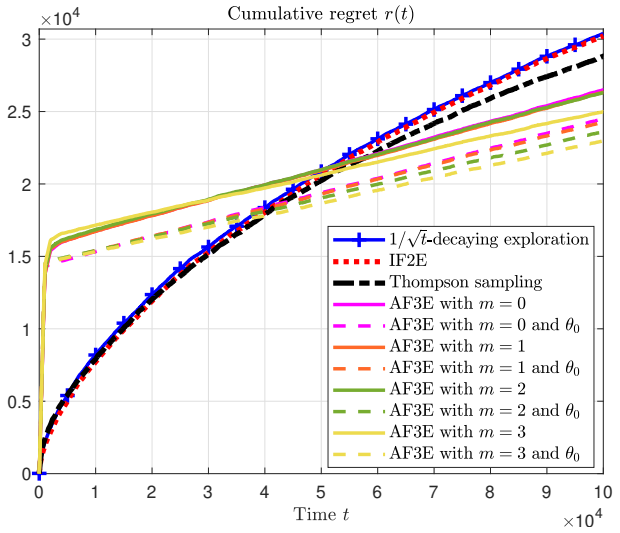

Figure 1. Cumulative regret obtained with $1/\sqrt{t}$-decaying exploration (blue solid line with markers |), IF2E (red dotted line), Thompson sampling (dashed-dotted black line) and AF3E with θ_0 replaced in the SDP by $\hat{\theta}_p$ (solid lines) and AF3E with θ_0 not replaced (dashed lines).

As shown in Figure 1 where we depict the evolution of the average cumulative regret obtained with the different methods, AF3E gives the worst regret performances during the first half of the experiment. Particularly, we can notice a steep increase at the beginning. This jumping phenomenon is also observed when we apply AF3E assuming the knowledge of θ_0. This sharp increase could be explained from the theoretical result in Theorem 2. Indeed, all the exploration effort is focused during the first interval since Theorem 2 implies $v_2 = \cdots = v_{L-1} = 0$. Moreover, the excitation power during the first interval is kept constant during $N = 1000$ time instants while IF2E and $1/\sqrt{t}$-decaying exploration continuously decrease their excitation power. Hence, AF3E proposes an aggressive exploration strategy at the beginning.

In Figure 2, we depict the average of the power of each external excitation v_k applied to the system (i.e., average of the obtained $c_0(k)$ with $k = 1, \cdots, 100$) with AF3E when m is chosen equal to 0. This is done for the case where θ_0 is assumed known and for the case where we do not make this assumption. In the former case, the excitation profile is in line with Theorem 2 since only the first interval is largely excited, the second interval is almost not excited (negligible power) and the remaining intervals are not excited at all. For the realistic scheme, we observe that more intervals are excited. This is due to the fact that θ_0 is replaced by the first estimates $\hat{\theta}_k$ which are the most uncertain. No excitation was applied to the system for $t \geq 9000$ with the realistic AF3E scheme, i.e., the inequality (20) was always satisfy for the last 91 intervals.

Finally, in Figure 3, we depict the average of the magnitude of the PSD of the external during the first interval (i.e., ϕ_{v_1}). By increasing m, we allow the external ex-

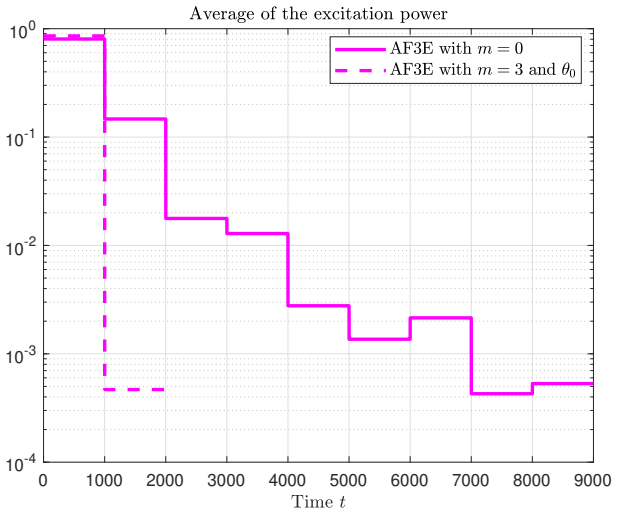

Figure 2. Average of the power of each v_k applied to the system with AF3E and $m = 0$ (solid line) and AF3E with θ_0 not replaced in the SDP (solid line).

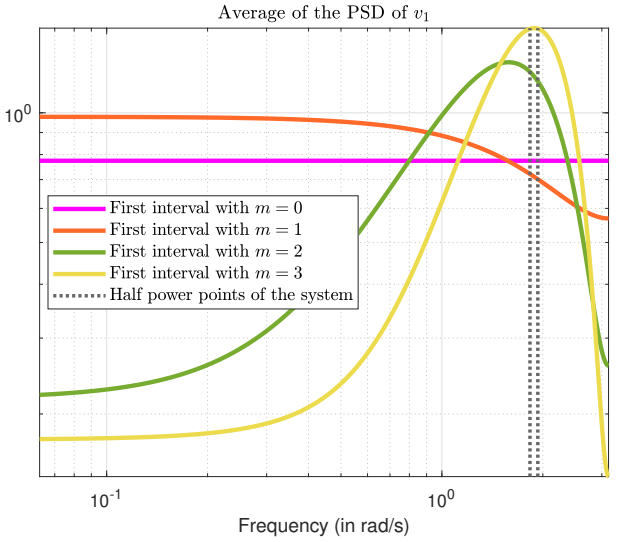

Figure 3. Average of the PSD of v_1 with AF3E and $m = 0, 1, 2$ and 3 and with θ_0 replaced by $\hat{\theta}_p$. The dotted black vertical lines correspond to the half power points of the system.

citation to be more correlated which optimally shaped the excitation frequency content at the resonance with $m = 3$. This is an illustration of the advantage of adding correlation in the external excitation which is something missing in the other exploration strategies.

7.6 Effect of the epoch lengths N

In this paragraph, we study the effect of the epoch length N for AF3E. In Table 3, we give the Monte Carlo estimate of the expected cumulative regret with $N = 500, 1000$ and 2000. It seems that decreasing the epoch length N can improve the regret.

FIR order m	$N = 500$	$N = 1000$	$N = 2000$
0	24822	26490	29154
1	24494	26333	28791
2	24566	26290	28682
3	23052	24973	28010

Table 3
Cumulative regret $r(T)$ obtained with AF3E for different FIR orders m and different epoch lengths N.

8 Conclusion and future directions

Using the tools/approaches developed in the system identification community and the receding horizon principle, we developed a new exploration strategy for regret minimization in linear quadratic problems which (i) does not require the tuning of hyperparameters depending on the true dynamics as is the case in most available methods in the literature and (ii) adds correlation for the external excitation. Moreover, our approach considers the scaling constant of the regret growth rate in the design. Even though the design of the exploration is a semidefinite programming to be solved, we had to develop two theoretical results (Theorems 1 and 2) in order to reduce the computation time. The simulation results show a great reduction of the required computational power and a reduced regret in finite-time compared to the optimal $1/\sqrt{t}$-decaying and IF2E explorations as well as the Thompson sampling approach.

Several perspectives are under investigation in order to improve AF3E. Indeed, as we have seen in the numerical example, it tends to make the regret explode at the beginning and we might be able to get better regret results if we could reduce that effect. The approach we want to follow in the future is the worst-case design approach as done in [15] by using the uncertainties of the identified models. The corresponding SDP to be solved can however become computationally expensive since the number of LMI constraints is higher than in the approach presented in this paper. In the numerical example of [15], it was shown that the optimal external excitation, whose design is divided in several epochs, only excites the system during the first epochs. This could suggest that the SDP for the approach presented in [15] could be simplified in a similar manner as the SDP in this paper (see Theorem 2). We will investigate if using a similar proof approach as we did in Appendix B for Theorem 2 for the robustified approach. Another drawback of the proposed approach is that power constraints, which exist in real-life, are not taken into account. Such constraints can be implemented as LMI constraint which is advantageous [20,15]. Adding this type of constraint might help in reducing the jumping effect of the regret at the beginning of the experiment with the current form of AF3E. Finally, we would like to get more insights on the choice of the epoch length with an theoretical study of AF3E.

References

[1] Yakubovich V. A. Solution of certain matrix inequalities occurring in the theory of automatic control. *Doklady*

Akademii Nauk, SSSR, page 1304–1307, 1962.

[2] Y. Abbasi-Yadkori and C. Szepesvári. Regret bounds for the adaptive control of linear quadratic systems. In Sham M. Kakade and Ulrike von Luxburg, editors, *Proceedings of the 24th Annual Conference on Learning Theory*, volume 19 of *Proceedings of Machine Learning Research*, pages 1–26, Budapest, Hungary, 09–11 Jun 2011. PMLR.

[3] M. Abeille and A. Lazaric. Thompson sampling for linear-quadratic control problems. In *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics*, volume 54, pages 1246–1254. PMLR, 2017.

[4] Brian DO Anderson and John B Moore. *Optimal control: linear quadratic methods*. Courier Corporation, 2007.

[5] MOSEK ApS. *The MOSEK optimization toolbox for MATLAB manual. Version 10.0.46.*, 2023.

[6] Archith Athrey, Othmane Mazhar, Meichen Guo, Bart De Schutter, and Shengling Shi. Regret analysis of learning-based linear quadratic gaussian control with additive exploration. *arXiv preprint arXiv:2311.02679*, 2023.

[7] X. Bombois, G. Scorletti, M. Gevers, P. M.J. Van den Hof, and R. Hildebrand. Least costly identification experiment for control. *Automatica*, 42(10):1651–1662, 2006.

[8] Xavier Bombois, Michel Gevers, Gérard Scorletti, and Brian DO Anderson. Robustness analysis tools for an uncertainty set obtained by prediction error identification. *Automatica*, 37(10):1629–1636, 2001.

[9] M.C. Campi. Achieving optimality in adaptive control: the "bet on the best" approach. In *Proc. 36th IEEE Conf. on Decision and Control*, volume 5, pages 4671–4676, San Diego, California, 1997.

[10] A. Cohen, T. Koren, and Y. Mansour. Learning linear-quadratic regulators efficiently with only $\sqrt{T}$ regret. In *International Conference on Machine Learning*, pages 1300–1309. PMLR, 2019.

[11] K. Colin, M. Ferizbegovic, and H. Hjalmarsson. Regret minimization for linear quadratic adaptive controllers using fisher feedback exploration. *IEEE Control Systems Letters*, 6:2870–2875, 2022.

[12] K. Colin, H. Hjalmarsson, and X. Bombois. Optimal exploration strategies for finite horizon regret minimization in some adaptive control problems. *IFAC-PapersOnLine*, 56(2):2564–2569, 2023.

[13] S. Dean, H. Mania, N. Matni, B. Recht, and S. Tu. Regret bounds for robust adaptive control of the linear quadratic regulator. *Advances in Neural Information Processing Systems*, 31, 2018.

[14] M.K. Shirani Faradonbeh, A. Tewari, and G. Michailidis. Input perturbations for adaptive control and learning. *Automatica*, 117, 2020.

[15] M. Forgione, X. Bombois, and P. M.J. Van den Hof. Data-driven model improvement for model-based control. *Automatica*, 52:118–124, 2015.

[16] László Gerencsér, Håkan Hjalmarsson, and Lirong Huang. Adaptive input design for lti systems. *IEEE Transactions on Automatic Control*, 62(5):2390–2405, 2017.

[17] M. Gevers, A. S. Bazanella, X. Bombois, and L. Miskovic. Identification and the information matrix: how to get just sufficiently rich? *IEEE Transactions on Automatic Control*, 2009.

[18] R. Hildebrand and M. Gevers. Minimizing the worst-case ν-gap by optimal input design. *IFAC Proceedings Volumes*, 36(16):645–650, 2003.

[19] H. Hjalmarsson. System identification of complex and structured systems. *European journal of control*, 15(3-4):275–310, 2009.

[20] H. Jansson. *Experiment design with applications in identification for control*. PhD thesis, Signaler, sensorer och system, 2004.

[21] H. Jansson and H. Hjalmarsson. Input design via LMIs admitting frequency-wise model specifications in confidence regions. *IEEE Transactions on Automatic Control*, 50(10):1534–1549, 2005.

[22] Y. Jedra and A. Proutiere. Optimal best-arm identification in linear bandits. *Advances in Neural Information Processing Systems*, 33:10007–10017, 2020.

[23] Y. Jedra and A. Proutiere. Minimal expected regret in linear quadratic control. In *International Conference on Artificial Intelligence and Statistics*, pages 10234–10321. PMLR, 2022.

[24] Yassir Jedra. *Statistical Learning in Linearly Structured Systems: Identification, Control, and Reinforcement Learning*. PhD thesis, KTH Royal Institute of Technology, 2023.

[25] T. Kargin, S. Lale, K. Azizzadenesheli, A. Anandkumar, and B. Hassibi. Thompson sampling achieves $\tilde{O}(\sqrt{T})$ regret in linear quadratic control. In *Conference on Learning Theory*, pages 3235–3284. PMLR, 2022.

[26] S. Karlin and W. J. Studden. *Tchebycheff Systems: With Applications in Analysis and Statistics*. Interscience, New York, 1966.

[27] T. L. Lai and H. Robbins. Asymptotically efficient adaptive allocation rules. *Advances in Applied Mathematics*, 6:4–22, 1985.

[28] T. L. Lai and C.-Z. Wei. Asymptotically efficient self-tuning regulators. *SIAM Journal on Control and Optimization*, 25(2):466–481, 1987.

[29] T.L. Lai. Asymptotically efficient adaptive control in stochastic regression models. *Advances in Applied Mathematics*, 7:23–45, 1986.

[30] S. Lale, K. Azizzadenesheli, B. Hassibi, and A. Anandkumar. Reinforcement Learning with Fast Stabilization in Linear Dynamical Systems. In *International Conference on Artificial Intelligence and Statistics*, pages 5354–5390. PMLR, 2022.

[31] C. Larsson, C.R. Rojas, X. Bombois, and H. Hjalmarsson. Experimental evaluation of model predictive control with excitation (MPC-X) on an industrial depropanizer. *Journal of Process Control*, 31:1–16, Jul 2015.

[32] C. A. Larsson, M. Annergren, H. Hjalmarsson, C. R Rojas, X. Bombois, A. Mesbah, and P. E. Modén. Model predictive control with integrated experiment design for output error systems. In *2013 European Control Conference (ECC)*, pages 3790–3795. IEEE, 2013.

[33] L. Ljung. *System identification, Theory for the user*. System sciences series. Prentice Hall, Upper Saddle River, NJ, USA, second edition, 1999.

[34] J. Lofberg. Yalmip: A toolbox for modeling and optimization in matlab. In *2004 IEEE international conference on robotics and automation (IEEE Cat. No. 04CH37508)*, pages 284–289. IEEE, 2004.

[35] H. Mania, S. Tu, and B. Recht. Certainty equivalence is efficient for linear quadratic control. In *NeurIPS*, 2019.

[36] Y. Ouyang, M. Gagrani, and R. Jain. Control of unknown linear systems with Thompson sampling. In *2017 55th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pages 1198–1205. IEEE, 2017.

[37] A. Rantzer. On the kalman-yakubovich-popov lemma. *System & Control Letters*, 28(1):7–10, 1996.

[38] A. Rantzer. Concentration bounds for single parameter adaptive control. In *Proceedings of American Control Conference (ACC 2018)*, pages 1862–1866, 2018.

[39] D. J. Russo, B. Van Roy, A. Kazerouni, I. Osband, and Z. Wen. A tutorial on thompson sampling. *Foundations and Trends® in Machine Learning*, 11(1):1–96, 2018.

[40] A. H. Sayed and T. Kailath. A survey of spectral factorization methods. *Numerical linear algebra with applications*, 8(6-7):467–496, 2001.

[41] M. K. S. Shirani Faradonbeh, A. Tewari, and G. Michailidis. Finite-time adaptive stabilization of linear systems. *IEEE Transactions on Automatic Control*, 64(8):3498–3505, 2019.

[42] M. Simchowitz and D. Foster. Naive exploration is optimal for online LQR. In Hal Daumé III and Aarti Singh, editors, *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pages 8937–8948. PMLR, 13–18 Jul 2020.

[43] M. Simchowitz, H. Mania, S. Tu, M. I. Jordan, and B. Recht. Learning without mixing: Towards a sharp analysis of linear system identification. In *Conference On Learning Theory*, pages 439–473. PMLR, 2018.

[44] F. Wang and L. Janson. Exact asymptotics for linear quadratic adaptive control. *Journal of Machine Learning Research*, 22(265):1–112, 2021.

[45] I. Ziemann and H. Sandberg. Regret lower bounds for learning linear quadratic gaussian systems. *arXiv preprint arXiv:2201.01680*, 2022.

A Proof of Theorem 1

First, let us observe that the DARE $\mathbf{P} - \mathbf{A}^\top\mathbf{P}\mathbf{A} = \mathbf{Q} + \mathbf{A}^\top\mathbf{P}\mathbf{B}(\mathbf{R} + \mathbf{B}^\top\mathbf{P}\mathbf{B})^{-1}\mathbf{B}^\top\mathbf{P}\mathbf{A}$ can be recast into

$$\mathbf{P} - \mathbf{L}^\top\mathbf{P}\mathbf{L} = \mathbf{Q} + \mathbf{K}^\top\mathbf{R}\mathbf{K} \tag{A.1}$$

with $\mathbf{L} = \mathbf{A} - \mathbf{B}\mathbf{K}$ by using the fact that $\mathbf{K} = (\mathbf{R} + \mathbf{B}^\top\mathbf{P}\mathbf{B})^{-1}\mathbf{B}^\top\mathbf{P}\mathbf{A}$. Based on this rewriting of the DARE, we are going to show that $\mathcal{D}(e^{j\omega}) = \mathbf{T}^*_{xv}(e^{j\omega})\mathbf{Q}\mathbf{T}^*_{xv}(e^{j\omega}) + \mathbf{T}^*_{uv}(e^{j\omega})\mathbf{R}\mathbf{T}^*_{uv}(e^{j\omega})$ is equal to $\mathbf{R}+\mathbf{B}^\top\mathbf{P}\mathbf{B}$. With $\mathbf{T}_{uv}(z) = 1-\mathbf{K}\mathbf{T}_{xv}(z)$, $\mathcal{D}(\omega)$ becomes

$$\begin{aligned}\mathcal{D}(e^{j\omega}) = \mathbf{R} &+ \mathbf{T}^*_{xv}(e^{j\omega})(\mathbf{Q} + \mathbf{K}^\top\mathbf{R}\mathbf{K})\mathbf{T}_{xv}(e^{j\omega}) \\ &- \mathbf{T}^*_{xv}(e^{j\omega})\mathbf{K}^\top\mathbf{R} - \mathbf{R}\mathbf{K}\mathbf{T}_{xv}(e^{j\omega})\end{aligned}$$

Using (A.1), we get $\mathcal{D}(e^{j\omega}) = \mathbf{R} + \mathbf{T}^*_{xv}(e^{j\omega})(\mathbf{P} - \mathbf{L}^\top\mathbf{P}\mathbf{L})\mathbf{T}_{xv}(e^{j\omega}) - \mathbf{T}^*_{xv}(e^{j\omega})\mathbf{K}^\top\mathbf{R} - \mathbf{R}\mathbf{K}\mathbf{T}_{xv}(e^{j\omega})$ Since $\mathbf{K} = (\mathbf{R} + \mathbf{B}^\top\mathbf{P}\mathbf{B})^{-1}\mathbf{B}^\top\mathbf{P}\mathbf{A}$, then $\mathbf{R}\mathbf{K} = \mathbf{B}^\top\mathbf{P}\mathbf{L}$. Hence, $\mathcal{D}(e^{j\omega})$ becomes

$$\begin{aligned}\mathcal{D}(e^{j\omega}) = \mathbf{R} &+ \mathbf{T}^*_{xv}(e^{j\omega})(\mathbf{P} - \mathbf{L}^\top\mathbf{P}\mathbf{L})\mathbf{T}_{xv}(e^{j\omega}) \\ &- \mathbf{T}^*_{xv}(e^{j\omega})\mathbf{L}^\top\mathbf{P}\mathbf{B} - \mathbf{B}^\top\mathbf{P}\mathbf{L}\mathbf{T}_{xv}(e^{j\omega})\end{aligned} \tag{A.2}$$

From $\mathbf{T}_{xv}(z) = (z\mathbf{I}_{n_x} - \mathbf{L})^{-1}\mathbf{B}$, we have $\mathbf{B} = (z\mathbf{I}_{n_x} - \mathbf{L})\mathbf{T}_{xv}(z)$ which, injected in $\mathcal{D}(e^{j\omega})$ in (A.2) and by factorizing by $(e^{j\omega}\mathbf{I}_{n_x} - \mathbf{L})$, leads to $\mathcal{D}(e^{j\omega}) = \mathbf{R} + \mathbf{T}^*_{xv}(e^{j\omega})\mathbf{H}(e^{j\omega})\mathbf{T}_{xv}(e^{j\omega})$ with $\mathbf{H}(e^{j\omega}) = (e^{j\omega}\mathbf{I}_{n_x} - \mathbf{L})^*\mathbf{P}(e^{j\omega}\mathbf{I}_{n_x} - \mathbf{L})$. Since $\mathbf{T}_{xv}(z) = (z\mathbf{I}_{n_x} - \mathbf{L})^{-1}\mathbf{B}$, we obtain $\mathcal{D}(e^{j\omega}) = \mathbf{R} + \mathbf{B}^\top\mathbf{P}\mathbf{B}$, concluding the proof. ■

B Proof of Theorem 2

B.1 Preliminaries

For the proof of Theorem 2, we require one preliminary result on positive semidefinite symmetric Toeplitz matrices [18,26]. It comes from the theory of Tchebycheff trigonometric moments.

Lemma 1 ([18,26])**.** Consider $m+1$ real-valued scalars x_q $(q = 0, \cdots, m)$. The symmetric Toeplitz matrix $\mathcal{T}(x_0, x_1, \cdots, x_m)$ is positive semidefinite if and only if it exists an infinite number of scalars x_{m+1}, x_{m+2}, $\cdots$ such that $x_0 + 2\sum_{q=1}^{+\infty} x_q \cos(q\omega) \geq 0\ \forall\omega \in \mathbb{R}$. □

B.2 Proof of Theorem 2

The outline of the proof is as follows

- *Part 1*: we define the dual optimization problem of (19) and simplifies it.
- *Part 2*: we show that we can remove many constraints, simplifying the dual problem.
- *Part 3*: by assuming that the optimal solution of the simplified dual problem is unconstrained, we prove that the dual optimum (equal to the primal optimum since the primal problem is strictly feasible according to Slater's condition) is equal to the primal objective function for which $\mathbf{X}(k) = 0\ \forall k \in [\![p, L-1]\!]$, $c_q(k) = 0\ \forall t \in [\![k \in [\![p, L-1]\!]$ and $\forall q \in [\![0, m]\!]$ and this is true only if the constraint (20) in the theorem statement is satisfied.
- *Part 4*: in the case the constraint (20) is not satisfied, we determine the dual problem of the simplified dual problem and we show that we recover the original primal problem (19) for which all $\mathbf{X}(k) = 0\ \forall k \in [\![p+1, L-1]\!]$, $c_q(k) = 0\ \forall k \in [\![p+1, L-1]\!]$ and $\forall q \in [\![0, m]\!]$.

Part 1: Define $L-p$ dual positive semidefinite matrices $\{\bar{\mathbf{\Delta}}(k)\}_{k=p}^{L-1}$ of dimension $2n_\theta \times 2n_\theta$ and $L-p$ dual positive semidefinite matrices $\{\bar{\mathbf{\Psi}}(k)\}_{k=p}^{L-1}$ of dimension $(m+1)\times(m+1)$ structured as follows

$$\bar{\mathbf{\Delta}}(k) = \begin{pmatrix} \mathbf{\Delta}_{11}(k) & \mathbf{\Delta}_{12}(k) \\ \mathbf{\Delta}_{12}^\top(k) & \mathbf{\Delta}_{22}(k) \end{pmatrix} \quad \bar{\mathbf{\Psi}}(k) = \begin{pmatrix} \mathbf{\Psi}(k) & \psi_{1:m}(k) \\ \psi_{1:m}^\top(k) & \psi_0(k) \end{pmatrix} \tag{B.1}$$

where $\mathbf{\Delta}_{11}(k) \in \mathbb{R}^{n_\theta\times n_\theta}$, $\mathbf{\Delta}_{12}(k) \in \mathbb{R}^{n_\theta\times n_\theta}$, $\mathbf{\Delta}_{22}(k) \in \mathbb{R}^{n_\theta\times n_\theta}$, $\mathbf{\Psi}(k) \in \mathbb{R}^{m\times m}$, $\psi_{1:m}(k) = (\psi_1(k), \cdots, \psi_m(k))^\top \in \mathbb{R}^{m\times 1}$ and $\psi_0(k) \in \mathbb{R}$. Consider the Lagrangian function of the SDP (19) with $\beta_1 = \cdots = \beta_m = 0$ given by

$$\begin{aligned}&\sum_{k=p}^{L-1} \{\mathrm{tr}(\mathbf{H}(k)) + \beta_0 c_0(k) \\ &- \mathrm{tr}\left(\bar{\mathbf{\Delta}}(k)\begin{pmatrix} \mathbf{H}(k) & \tilde{\mathbf{W}} \\ \tilde{\mathbf{W}} & \boldsymbol{\mathcal{M}}_k + N(c_0(i)\boldsymbol{\mathcal{L}}_q^0 + \sum_{i=p}^k \sum_{q=1}^m c_q(i)\boldsymbol{\mathcal{N}}_q^v) \end{pmatrix}\right) \\ &- \mathrm{tr}\left(\bar{\mathbf{\Psi}}(k)\begin{pmatrix} \mathbf{X}(k) - \mathbf{Y}^\top\mathbf{X}(k)\mathbf{Y} & c_{1:m}(k) - \mathbf{Y}^\top\mathbf{X}(k)\mathbf{Z} \\ c_{1:m}^\top(k) - \mathbf{Z}^\top\mathbf{X}(k)\mathbf{Y} & c_0(k) - \mathbf{Z}^\top\mathbf{X}(k)\mathbf{Z} \end{pmatrix}\right)\Bigg\}\end{aligned}$$

Setting the gradient of the Lagrangian with respect to each primal variable to 0, we get the following con-

straints, for each $k \in [\![p, L-1]\!]$,

$$\begin{aligned}
&\boldsymbol{\Delta}_{11}(k) = \mathbf{I}_{n_\theta} \qquad (B.2)\\
&2\psi_q(k) = -N\sum_{i=p}^{k} \mathrm{tr}\left(\boldsymbol{\Delta}_{22}(i)\boldsymbol{\mathcal{N}}_q^v\right) \ \forall q \in [\![1, m]\!]\\
&\psi_0(k) = \beta_0 - N\sum_{i=p}^{k} \mathrm{tr}(\boldsymbol{\Delta}_{22}(i)\boldsymbol{\mathcal{L}}_0^v)\\
&\boldsymbol{\Psi}(k) = \mathbf{Y}\boldsymbol{\Psi}(k)\mathbf{Y}^\top + \mathbf{Z}\psi_{1:m}^\top(k)\mathbf{Y}^\top + \mathbf{Y}\psi_{1:m}(k)\mathbf{Z}^\top + \psi_0(k)\mathbf{Z}\mathbf{Z}^\top
\end{aligned}$$

The corresponding dual objective function is given by $-\sum_{k=p}^{L-1}\mathrm{tr}(\boldsymbol{\Delta}_{22}(k)\boldsymbol{\mathcal{M}}_k) - 2\sum_{k=p}^{L-1}\mathrm{tr}(\tilde{\mathbf{W}}\boldsymbol{\Delta}_{12}(k))$. From duality theory, the maximum of this objective function constrained by the above equalities and the positive semidefiniteness inequalities $\bar{\boldsymbol{\Delta}}(k) \succeq 0$ and $\bar{\boldsymbol{\Psi}}(k) \succeq 0$ for all $k \in [\![p, L-1]\!]$ is a lower bound to the objective function of the primal problem (19) with $\beta_1 = \cdots = \beta_m = 0$. Because the primal problem is a strictly feasible SDP, Slater's condition implies that the primal minimum coincides with the dual maximum.

Let us now analyze the effects of the constraints (B.2) combined with $\bar{\boldsymbol{\Delta}}(k) \succeq 0$ on the dual problem. First, both constraints lead to $\boldsymbol{\Delta}_{22}(k) \succeq \boldsymbol{\Delta}_{12}^\top(k)\boldsymbol{\Delta}_{12}(k)$. Secondly, because the dual objective function is monotonically decreasing with respect to $\boldsymbol{\Delta}_{22}(k)$ and $\boldsymbol{\Delta}_{12}(k)$ is not involved in other constraints, the optimum of the dual problem will satisfy $\boldsymbol{\Delta}_{22}(k) = \boldsymbol{\Delta}_{12}^\top(k)\boldsymbol{\Delta}_{12}(k)$. Hence, we can replace all $\boldsymbol{\Delta}_{22}(k)$ by $\boldsymbol{\Delta}_{12}^\top(k)\boldsymbol{\Delta}_{12}(k)$. By dropping the index 12 in $\boldsymbol{\Delta}_{12}(k)$ for ease of notation, we get the following dual problem

$$\max_{\boldsymbol{\Delta}(k), \bar{\boldsymbol{\Psi}}(k)} -\sum_{k=p}^{L-1} \mathrm{tr}\left(\boldsymbol{\Delta}(k)\boldsymbol{\mathcal{M}}_k\boldsymbol{\Delta}^\top(k) + 2\tilde{\mathbf{W}}\boldsymbol{\Delta}(k)\right)$$

Subject to, for all $k \in [\![p, L-1]\!]$,

$$2\psi_q(k) = -N\sum_{i=p}^{k} \mathrm{tr}(\boldsymbol{\Delta}(i)\boldsymbol{\mathcal{N}}_q^v\boldsymbol{\Delta}^\top(i)) \ \forall q \in [\![1, m]\!] \qquad (B.3)$$

$$\psi_0(k) = \beta_0 - N\sum_{i=p}^{k} \mathrm{tr}(\boldsymbol{\Delta}(i)\boldsymbol{\mathcal{L}}_0^v\boldsymbol{\Delta}^\top(i)) \qquad (B.4)$$

$$\bar{\boldsymbol{\Psi}}(k) = \begin{pmatrix} \boldsymbol{\Psi}(k) & \psi_{1:m}(k) \\ \psi_{1:m}^\top(k) & \psi_0(k) \end{pmatrix} \succeq 0 \qquad (B.5)$$

$$\boldsymbol{\Psi}(k) = \mathbf{Y}\boldsymbol{\Psi}(k)\mathbf{Y}^\top + \mathbf{Z}\psi_{1:m}^\top(k)\mathbf{Y}^\top + \mathbf{Y}\psi_{1:m}(k)\mathbf{Z}^\top + \psi_0(k)\mathbf{Z}\mathbf{Z}^\top \qquad (B.6)$$

Part 2: We will now prove that all the above constraints are satisfied if and only the constraints (B.3)-(B.6) defined at $k = L-1$ are satisfied.

First, let us analyze the effect of (B.6) on $\boldsymbol{\Psi}(k)$. Denote by $\boldsymbol{\Psi}_{1:(m-1)}(k)$ its submatrix obtained by removing its last column and its last row. The terms $\mathbf{Y}\boldsymbol{\Psi}(k)\mathbf{Y}^\top$, $\mathbf{Z}\psi_{1:m}(k)\mathbf{Y}^\top$ and $\psi_0(k)\mathbf{Z}\mathbf{Z}^\top$ are given by

$$\mathbf{Y}\boldsymbol{\Psi}(k)\mathbf{Y}^\top = \begin{pmatrix} 0 & \mathbf{0}_{1\times(m-1)} \\ \mathbf{0}_{(m-1)\times 1} & \boldsymbol{\Psi}_{1:(m-1)}(k) \end{pmatrix}$$

$$\mathbf{Y}\Psi_{1:m}(k)\mathbf{Z}^\top = \begin{pmatrix} 0 & \mathbf{0}_{1\times(m-1)} \\ \Psi_{1:(m-1)}(k) & \mathbf{0}_{(m-1)\times(m-1)} \end{pmatrix}$$

$$\Psi_0(k)\mathbf{Z}\mathbf{Z}^\top = \begin{pmatrix} \Psi_0(k) & \mathbf{0}_{1\times(m-1)} \\ \mathbf{0}_{(m-1)\times 1} & \mathbf{0}_{(m-1)\times(m-1)} \end{pmatrix}$$

where $\psi_{1:(m-1)}(k) = (\psi_1(k), \ldots, \psi_{m-1}(k))^\top$. Then, because of the constraint (B.6), the matrix $\boldsymbol{\Psi}(k)$ is equal to the Toeplitz symmetric matrix $\mathcal{T}(\psi_0(k), \cdots, \psi_{m-1}(k))$. Injecting this into $\bar{\boldsymbol{\Psi}}(k)$ whose structure is given by (B.1), we get

$$\bar{\boldsymbol{\Psi}}(k) = \begin{pmatrix} \mathcal{T}(\psi_0(k), \cdots, \psi_{m-1}(k)) & \psi_{1:m}(k) \\ \psi_{1:m}^\top(k) & \psi_0(k) \end{pmatrix}$$

Using a particular similarity transformation consisting of rows and columns permutations, this matrix can be recast into $\mathcal{T}(\psi_0(k), \cdots, \psi_m(k))$. Hence, $\bar{\boldsymbol{\Psi}}(k) \succeq 0$ if and only if $\mathcal{T}(\psi_0(k), \cdots, \psi_m(k)) \succeq 0$. Recall that $\boldsymbol{\mathcal{N}}_q^v = \boldsymbol{\mathcal{L}}_q^v + (\boldsymbol{\mathcal{L}}_q^v)^\top$ for each $q \geq 1$ and so we have $\mathrm{tr}(\boldsymbol{\Delta}(i)\boldsymbol{\mathcal{N}}_q^v\boldsymbol{\Delta}^\top(i)) = 2\,\mathrm{tr}(\boldsymbol{\Delta}(i)\boldsymbol{\mathcal{L}}_q^v\boldsymbol{\Delta}^\top(i))$. Combined with the constraints (B.3)-(B.4), $\mathcal{T}(\psi_0(k), \cdots, \psi_m(k)) \succeq 0$ is equivalent to

$$\beta_0\mathbf{I}_{m+1} \succeq N\sum_{i=p}^{k} \mathcal{T}\left(\mathrm{tr}(\boldsymbol{\Delta}(i)\boldsymbol{\mathcal{L}}_0^v\boldsymbol{\Delta}^\top(i)), \cdots, \mathrm{tr}\left(\boldsymbol{\Delta}(i)\boldsymbol{\mathcal{L}}_m^v\boldsymbol{\Delta}^\top(i)\right)\right) \qquad (B.7)$$

It must be guaranteed for each $k = [\![p, L-1]\!]$. Let us now prove that satisfying (B.7) for all $k = [\![p, L-1]\!]$ is equivalent to only satisfy (B.7) for $k = L-1$. In order to obtain such a result, we will show that, for any square matrix $\mathbf{X}$ of dimension $n_\theta \times n_\theta$, we have

$$\mathcal{T}(\mathrm{tr}(\mathbf{X}\boldsymbol{\mathcal{L}}_0^v\mathbf{X}^\top), \cdots, \mathrm{tr}(\mathbf{X}\boldsymbol{\mathcal{L}}_m^v\mathbf{X}^\top)) \succeq 0 \qquad (B.8)$$

Indeed, if the latter holds, we would get

$$\begin{aligned}
&\sum_{i=p}^{L-1} \mathcal{T}(\mathrm{tr}(\boldsymbol{\Delta}(i)\boldsymbol{\mathcal{L}}_0^v\boldsymbol{\Delta}^\top(i)), \cdots, \mathrm{tr}(\boldsymbol{\Delta}(i)\boldsymbol{\mathcal{L}}_m^v\boldsymbol{\Delta}^\top(i)))\\
&\succeq \sum_{i=p}^{k} \mathcal{T}(\mathrm{tr}(\boldsymbol{\Delta}(i)\boldsymbol{\mathcal{L}}_0^v\boldsymbol{\Delta}^\top(i)), \cdots, \mathrm{tr}(\boldsymbol{\Delta}(i)\boldsymbol{\mathcal{L}}_m^v\boldsymbol{\Delta}^\top(i)))
\end{aligned}$$

for any $k = [\![p, L-1]\!]$ and so satisfying (B.7) for all $k = [\![p, L-1]\!]$ is equivalent to only satisfy (B.7) for $k = L-1$. In order to prove (B.8), we will use Lemma 1. First, let us note that the positive-semidefinite matrix $\boldsymbol{\Gamma}(e^{j\omega})$ in (14) can be expanded as follows

$$\boldsymbol{\mathcal{L}}_0^v + \sum_{q=1}^{+\infty}\left(\boldsymbol{\mathcal{L}}_q^v e^{jq\omega} + (\boldsymbol{\mathcal{L}}_q^v)^\top e^{-jq\omega}\right) \succeq 0 \quad \forall\omega$$

Left- and right-multiplying the latter by any real-valued square matrix $\mathbf{X}$, taking the trace and using the fact that $\mathrm{tr}(\mathbf{X}\boldsymbol{\mathcal{L}}_q^v\mathbf{X}^\top) = \mathrm{tr}(\mathbf{X}(\boldsymbol{\mathcal{L}}_q^v)^\top\mathbf{X}^\top)$, we get

$$\mathrm{tr}(\mathbf{X}\boldsymbol{\mathcal{L}}_0^v\mathbf{X}^\top) + 2\sum_{q=1}^{+\infty}\mathrm{tr}(\mathbf{X}\boldsymbol{\mathcal{L}}_q^v\mathbf{X}^\top)\cos(q\omega) \geq 0 \quad \forall\omega$$

Using Lemma 1, we prove that (B.8) holds. Hence, it is necessary and sufficient to guarantee the constraint (B.7) at $k = L-1$. Recall that this constraint comes from the combination of the three constraints (B.3), (B.4) and (B.5) evaluated at $k = L-1$. Consequently, the original dual problem is equivalent to the following simplified optimization problem

$$\begin{aligned}
&\max_{\{\boldsymbol{\Delta}(k)\}_{k=p}^{L-1},\bar{\boldsymbol{\Psi}}(L-1)} -\textstyle\sum_{k=p}^{L-1}\mathrm{tr}\left(\boldsymbol{\Delta}(k)\boldsymbol{\mathcal{M}}_k\boldsymbol{\Delta}^\top(k) + 2\tilde{\mathbf{W}}\boldsymbol{\Delta}(k)\right)\\
&\textbf{Subject to}, \forall q \in [\![1,m]\!],\\
&2\psi_q(L-1) = -N\textstyle\sum_{k=p}^{L-1}\mathrm{tr}(\boldsymbol{\Delta}(k)\boldsymbol{\mathcal{N}}_q^v\boldsymbol{\Delta}^\top(k))\\
&\textbf{and}\\
&\psi_0(L-1) = \beta_0 - N\textstyle\sum_{k=p}^{L-1}\mathrm{tr}(\boldsymbol{\Delta}(k)\boldsymbol{\mathcal{L}}_0^v\boldsymbol{\Delta}^\top(k))\\
&\bar{\boldsymbol{\Psi}}(L-1) = \begin{pmatrix}\boldsymbol{\Psi}(L-1) & \psi_{1:m}(L-1)\\ \psi_{1:m}^\top(L-1) & \psi_0(L-1)\end{pmatrix} \succeq 0\\
&\boldsymbol{\Psi}(L-1) - \mathbf{Y}\boldsymbol{\Psi}(L-1)\mathbf{Y}^\top - \mathbf{Z}\psi_{1:m}^\top(L-1)\mathbf{Y}^\top\\
&\quad -\mathbf{Y}\psi_{1:m}(L-1)\mathbf{Z}^\top - \psi_0(L-1)\mathbf{Z}\mathbf{Z}^\top = 0
\end{aligned} \tag{B.9}$$

Part 3: Let us consider the case where the optimal solution of the simplified dual optimization problem does not make the inequality constraint $\bar{\boldsymbol{\Psi}}(L-1) \succeq 0$ active, i.e., $\bar{\boldsymbol{\Psi}}(L-1) \succ 0$ at the optimal solution. Then, by using the fact that each $\boldsymbol{\mathcal{M}}_k$ is invertible for all $k \in [\![p, L-1]\!]$, we can complete the square of the objective function as follows

$$\begin{aligned}
&-\sum_{k=p}^{L-1}\mathrm{tr}\left((\boldsymbol{\Delta}(k)+\tilde{\mathbf{W}}\boldsymbol{\mathcal{M}}_k^{-1})\boldsymbol{\mathcal{M}}_k(\boldsymbol{\Delta}(k)+\tilde{\mathbf{W}}\boldsymbol{\mathcal{M}}_k^{-1})^\top\right)\\
&+\sum_{k=p}^{L-1}\mathrm{tr}\left(\tilde{\mathbf{W}}\boldsymbol{\mathcal{M}}_k^{-1}\tilde{\mathbf{W}}\right)
\end{aligned}$$

The optimal unconstrained dual solution $\boldsymbol{\Delta}^{opt}(k)$ is $\boldsymbol{\Delta}^{opt}(k) = -\tilde{\mathbf{W}}\boldsymbol{\mathcal{M}}_k^{-1}$ and the corresponding maximum of the dual objective function is $\sum_{k=p}^{L-1}\mathrm{tr}(\tilde{\mathbf{W}}\boldsymbol{\mathcal{M}}_k^{-1}\tilde{\mathbf{W}})$. Looking back at the primal problem (19) with $\beta_1 = \cdots = \beta_m = 0$, this optimum is reached by setting, for each $k \in [\![p, L-1]\!]$, $\mathbf{H}(k) = \tilde{\mathbf{W}}\boldsymbol{\mathcal{M}}_k^{-1}\tilde{\mathbf{W}}$ and $c_0(k) = 0$ (recall that $\beta_1 = \cdots = \beta_m = 0$). Since $c_0(k)$ is the power of the excitation v_k, this means $v_k = 0$ for all $k \in [\![p, L-1]\!]$. Note that $\mathbf{X}(k) = 0$ is the only solution satisfying the PSD realizability LMI (9) when $c_0(k) = 0$. However, the inequality constraint $\bar{\boldsymbol{\Psi}}(L-1) \succeq 0$ should not be active, i.e., $\bar{\boldsymbol{\Psi}}(L-1) \succ 0$. Recall that the four constraints of the simplified dual problem are equivalent to (B.7) evaluated at $k = L-1$. Hence, replacing $\boldsymbol{\Delta}(k)$ by the unconstrained solution, the following strict inequality must be satisfied

$$\beta_0\mathbf{I}_{m+1} - N\sum_{k=p}^{L-1}\mathcal{T}(\bar{\zeta}_0(k), \cdots, \bar{\zeta}_m(k)) \succ 0$$

with $\bar{\zeta}_q(k) = \mathrm{tr}(\tilde{\mathbf{W}}\boldsymbol{\mathcal{M}}_k^{-1}\boldsymbol{\mathcal{L}}_q^v\boldsymbol{\mathcal{M}}_k^{-1}\tilde{\mathbf{W}})\ \forall q \in [\![0,m]\!]$. Since $\tilde{\mathbf{W}}$ is the square root matrix of $\mathbf{W}/2$, we have $\bar{\zeta}_q(k) = \mathrm{tr}(\mathbf{W}\boldsymbol{\mathcal{M}}_k^{-1}\boldsymbol{\mathcal{L}}_q^v\boldsymbol{\mathcal{M}}_k^{-1})/2\ \forall q \in [\![0,m]\!]$. This inequality constraint is strict if and only if the condition (20) in the theorem statement is guaranteed.

Part 4: Let us now consider the case where the solution of (B.9) is not equal to the unconstrained one, i.e., (20) is not satisfied. The approach is to define the dual of (B.9) and to show that it corresponds to the primal problem (19) with the particular solution structure as in the theorem statement. In order to write the dual problem of (B.9), we define $m+1$ variables μ_q ($q = 0\cdots, m$), a $m \times m$ symmetric matrix $\mathbf{U}$ and a $(m+1)\times(m+1)$ positive semidefinite symmetric matrix $\bar{\boldsymbol{\Xi}}$ structured as follows

$$\bar{\boldsymbol{\Xi}} = \begin{pmatrix}\boldsymbol{\Xi} & \xi_{1:m}\\ \xi_{1:m}^\top & \xi_0\end{pmatrix} \succeq 0$$

where $\boldsymbol{\Xi} \in \mathbb{R}^{m\times m}$, $\xi_{1:m} = (\xi_1, \cdots, \xi_m)^\top \in \mathbb{R}^{m\times 1}$ and $\xi_0 \in \mathbb{R}$. Define the Lagrangian $\mathcal{G}$ of the simplified optimization problem (B.9) by

$$\begin{aligned}
\mathcal{G} =& \sum_{k=p}^{L-1}\mathrm{tr}\left(\boldsymbol{\Delta}(k)\boldsymbol{\mathcal{M}}_k\boldsymbol{\Delta}^\top(k)\right) + 2\sum_{k=p}^{L-1}\mathrm{tr}\left(\tilde{\mathbf{W}}\boldsymbol{\Delta}(k)\right)\\
&+\sum_{q=1}^{m}\left(2\psi_q(L-1) + N\sum_{k=p}^{L-1}\mathrm{tr}(\boldsymbol{\Delta}(t)\boldsymbol{\mathcal{N}}_q^v\boldsymbol{\Delta}^\top(t)))\right)\mu_q\\
&+\left(\psi_0(L-1) - \beta_0 + N\sum_{k=p}^{L-1}\mathrm{tr}(\boldsymbol{\Delta}(t)\boldsymbol{\mathcal{L}}_0^v\boldsymbol{\Delta}^\top(t)))\right)\mu_0\\
&-\mathrm{tr}\left(\bar{\boldsymbol{\Xi}}\begin{pmatrix}\boldsymbol{\Psi}(L-1) & \psi_{1:m}(L-1)\\ \psi_{1:m}^\top(L-1) & \psi_0(L-1)\end{pmatrix}\right)\\
&+\mathrm{tr}(\mathbf{U}(\boldsymbol{\Psi}(L-1) - \mathbf{Y}\boldsymbol{\Psi}(L-1)\mathbf{Y}^\top - \mathbf{Z}\psi_{1:m}^\top(L-1)\mathbf{Y}^\top))\\
&+\mathrm{tr}(\mathbf{U}(-\mathbf{Y}\psi_{1:m}(L-1)\mathbf{Z}^\top - \psi_0(L-1)\mathbf{Z}\mathbf{Z}^\top))
\end{aligned}$$

Setting the gradient of the Lagrangian with respect to the primal variables equal to 0, we get

$$\boldsymbol{\Xi} = \mathbf{U} - \mathbf{Y}^\top\mathbf{U}\mathbf{Y} \qquad \xi_0 = \mu_0 - \mathbf{Z}\mathbf{U}\mathbf{Z}^\top \tag{B.10}$$

$$\xi_{1:m} = \mu_{1:m} - \mathbf{Y}\mathbf{U}\mathbf{Z}^\top \tag{B.11}$$

$$\boldsymbol{\mathcal{J}}_k\boldsymbol{\Delta}(k)^\top = -\tilde{\mathbf{W}} \quad \forall k \in [\![p, L-1]\!] \tag{B.12}$$

with $\mu_{1:m}^\top = (\mu_1, \cdots, \mu_m)$ and $\boldsymbol{\mathcal{J}}_k = \boldsymbol{\mathcal{M}}_k + N(\mu_0\boldsymbol{\mathcal{L}}_0^v + \sum_{q=1}^m\mu_q\boldsymbol{\mathcal{N}}_q^v)$. Combining the constraints (B.10) and (B.11) with $\bar{\boldsymbol{\Xi}} \succeq 0$, then a symmetric matrix $\mathbf{U}$ exists such that μ_0 and $\mu_{1:m}$ satisfies

$$\begin{pmatrix}\mathbf{U} - \mathbf{Y}^\top\mathbf{U}\mathbf{Y} & \mu_{1:m} - \mathbf{Y}^\top\mathbf{U}\mathbf{Z}\\ \mu_{1:m}^\top - \mathbf{Z}^\top\mathbf{U}\mathbf{Y} & \mu_0 - \mathbf{Z}^\top\mathbf{U}\mathbf{Z}\end{pmatrix} \succeq 0 \tag{B.13}$$

which is the necessary and sufficient LMI condition for the realizability of the truncated PSD $\phi(\omega) = \mu_0 + 2\sum_{q=1}^{m} \mu_q \cos(q\omega)$ (see (9)).

Let us analyze now analyze the constraints (B.12). Recall that $\boldsymbol{\mathcal{M}}_k = \boldsymbol{\mathcal{I}}_p + (k-p+1)N\boldsymbol{\mathcal{L}}^e$ and so $\boldsymbol{\mathcal{J}}_k = \boldsymbol{\mathcal{I}}_p + (k-p)N\boldsymbol{\mathcal{L}}^e + N(\boldsymbol{\mathcal{L}}^e + \mu_0\boldsymbol{\mathcal{L}}_0^v + \sum_{q=1}^{m} \mu_q \boldsymbol{\mathcal{N}}_q^v)$. For any $\mu_0, \cdots, \mu_m$ satisfying the LMI (B.13), we have that $N(\boldsymbol{\mathcal{L}}^e + \mu_0\boldsymbol{\mathcal{L}}_0^v + \sum_{q=1}^{m} \mu_q \boldsymbol{\mathcal{N}}_q^v) \succeq 0$ since it corresponds to the additional Fisher information matrix coming from a filtered white noise external excitation of duration N with the PSD $\phi(\omega) = \mu_0 + 2\sum_{q=1}^{m} \mu_q \cos(q\omega)$. Since $\boldsymbol{\mathcal{I}}_p + (k-p)N\boldsymbol{\mathcal{L}}^e \succ 0$, we conclude that $\boldsymbol{\mathcal{J}}_k \succ 0$ for all $k \in [\![p, L-1]\!]$. Consequently, (B.12) is equivalent to $\boldsymbol{\Delta}(k) = -\tilde{\mathbf{W}}\boldsymbol{\mathcal{J}}_k^{-1}$. Hence, with the equalities (B.10)-(B.11) and injecting $\boldsymbol{\Delta}(k) = -\tilde{\mathbf{W}}\boldsymbol{\mathcal{J}}_k^{-1}$ into the Lagrangian $\mathcal{G}$, we get the dual objective function $-\sum_{k=p}^{L-1} \operatorname{tr}\left(\tilde{\mathbf{W}}\boldsymbol{\mathcal{J}}_k^{-1}\tilde{\mathbf{W}}\right) - \beta_0\mu_0$ for the optimization problem (B.9) and it has to be maximized with respect to $\mu_0, \cdots, \mu_m$ under the constraint that it exists a symmetric matrix $\mathbf{U}$ such that (B.13) holds.

Let us rewrite furthermore this dual maximization problem. First, maximizing the dual objective function is equivalent on minimizing its negative counterpart $\sum_{k=p}^{L-1} \operatorname{tr}(\tilde{\mathbf{W}}\boldsymbol{\mathcal{J}}_k^{-1}\tilde{\mathbf{W}}) + \beta_0\mu_0$. Now, by defining $L-p$ symmetric matrices $\{\mathbf{V}(k)\}_{k=p}^{L-1}$ such that $\mathbf{V}(k) \succeq \tilde{\mathbf{W}}\boldsymbol{\mathcal{J}}_k^{-1}\tilde{\mathbf{W}}$, we can use Schur complement in order to get the following equivalent minimization problem for the dual of (B.9)

$$\min_{\mathbf{U},\mu_q,\mathbf{V}(k)} \sum_{k=p}^{L-1} \operatorname{tr}\left(\mathbf{V}(k)\right) + \beta_0\mu_0$$

Subject to

$$\begin{pmatrix} \mathbf{V}(k) & \tilde{\mathbf{W}} \\ \tilde{\mathbf{W}} & \boldsymbol{\mathcal{M}}_k + N(\mu_0\boldsymbol{\mathcal{L}}_0^v + \sum_{q=1}^{m} \mu_q \boldsymbol{\mathcal{N}}_q^v) \end{pmatrix} \succeq 0 \quad \forall k \in [\![p, L-1]\!]$$

$$\begin{pmatrix} \mathbf{U} - \mathbf{Y}^\top\mathbf{U}\mathbf{Y} & \mu_{1:m} - \mathbf{Y}^\top\mathbf{U}\mathbf{Z} \\ \mu_{1:m}^\top - \mathbf{Z}^\top\mathbf{U}\mathbf{Y} & \mu_0 - \mathbf{Z}^\top\mathbf{U}\mathbf{Z} \end{pmatrix} \succeq 0$$

We recover the primal problem (19) with $\beta_1 = \cdots = \beta_m = 0$ and the following particular sparsity for the primal decision variables

- $\mathbf{U} = \mathbf{X}(p)$ and $\mathbf{X}(k) = 0\ \forall k \in [\![p+1, L-1]\!]$,
- $\mu_q = c_q(p)\ \forall q \in [\![0, m]\!]$ and $c_q(k) = 0\ \forall (k,q) \in [\![p+1, L-1]\!] \times [\![0, m]\!]$,
- $\mathbf{V}(k) = \mathbf{H}(k)\ \forall k \in [\![p, L-1]\!]$,

implying $v_{p+1} = \cdots = v_{L-1} = 0$ which concludes the proof. ■