Approximate Positioning of Final State for Parabolic PDEs with Dirichlet Boundary Conditions

Research Report LNCC/Systems & Control, 2018

G. O. Corrêa, M. M. López-Flores and A. L. Madureira

Contents

1 Introduction

The problem of choosing a control function to steer the solution of a partial-differential, evolution equation (PDEE, for short) in a prescribed manner has given rise to a vast literature (cf. Zuazua (2002)), Trötlzch (2010), Koput & Leugering (2011) and references therein) in which great attention has been devoted to linear PDEEs and problems defined over finite-time intervals. In particular, the basic aim of approximately reaching a prescribed final state from a given initial one gives rise to various optimal control problems for parabolic equations with different types of boundary conditions and control functions acting either on the boundary of the spatial domain or as a source term in its interior.

The control functions to be determined as solutions of such optimization problems are often allowed to depend on both time and space coordinates. On the other hand, having in mind potential applications, interest naturally arises in considering control functions which depend solely on time (their spatial action being defined by the "actuators" used). Accordingly, the final-state approximation problems for parabolic PDEEs considered here will have such control signals as decision variables. Moreover, it is desirable that "finite-dimensional" approximations to the optimal control signals are characterized either as solutions to approximations of the original problems or as approximations to the optimal solutions of those problems.

In this report, the problem of approximate positioning of the final state on a finite time interval is examined for parabolic PDEEs with Dirichlet boundary conditions and point (source) control functions. Minimization of a quadratic cost involving the final-state approximation error is considered with and without a constraint on the maximum magnitude of the control functions. To compute approximate solutions to such control problems, approximate versions of them are tackled which are obtained from finite-dimensional approximations to the control-to-final-state operator.

This report is organized as follows. In Section 2, the basic control problem is introduced. In Section 3, its optimal solution is characterized. In Section 4, approximate solutions to the basic, unconstrained control problem are derived. In Section 5, "peak value" constraints are added to the basic problem and both the original problem and approximate versions of it are discussed, including the use of Lagrangian duality to obtain approximate solutions.

Finally, in Section 6, two simple numerical examples are presented to illustrate the main points previously discussed. Unless otherwise stated, proofs are presented in the Appendix.

2 Background and Problem Formulation

Consider a initial/boundary condition problem for the parabolic equation given ("in its classical form") by

$$\forall\, \boldsymbol{x} \in U,\ \forall\, t \in (0,\infty)\ ,\ \frac{\partial \theta}{\partial t}(\boldsymbol{x},t) - \sum_{i,j=1}^{n} \frac{\partial}{\partial x_i}\left[\ a^{ij}\frac{\partial \theta}{\partial x_j}\right] + \sum_{i=1}^{n} b^i(x)\frac{\partial \theta}{\partial x_i} + d(\boldsymbol{x})\theta(\boldsymbol{x},t) = f(\boldsymbol{x},t) \tag{1}$$

$$\forall \boldsymbol{x} \in U\ ,\ \theta(\boldsymbol{x},0) = g(\boldsymbol{x}) \qquad \text{(initial condition)} \tag{2}$$

$$\forall\, t \in (0,\infty)\ ,\ \forall \boldsymbol{x} \in \partial U\ ,\ \ \theta(\boldsymbol{x},t) = 0 \qquad \text{(boundary conditions)} \tag{3}$$

where $U \in \mathbb{R}^n$ is an open and connected set, $a^{ij} = a^{ji}$, b^i, d, f and g are given functions, $\{A\}_{ij} = a^{ij}$, $A(x) \geq 0$, a.e. in U. The associated (weak) function-space, ordinary differential equation version, is given by

$$\dot{\underline{\theta}}(t) = A[\underline{\theta}(t)] + \underline{f}(t) \quad , \quad t > 0 \quad , \quad \underline{\theta}(0) = g \tag{4}$$

where $g \in L_2(U)$, $\underline{f} : (0,\infty) \to L_2(U)$, $\underline{\theta} : (0,\infty) \to L_2(U)$

$A : H_0^1(U) \to L_2(U)$ is defined by

$$\forall \phi \in H_0^1(U), \forall \psi \in H_0^1(U), \quad \langle A[\phi], \psi\rangle = -\mathbf{B}[\phi,\psi], \tag{5}$$

where

$$\mathbf{B}[\phi,\psi] \triangleq \sum_{i,j} \langle a^{ij}\frac{\partial \phi}{\partial x_i},\ \frac{\partial \psi}{\partial x_j}\rangle + \sum_i \langle b^i \frac{\partial \phi}{\partial x_i}\ ,\ \psi\rangle + \langle d\phi, \psi\rangle, \tag{6}$$

where $\langle\cdot,\cdot\rangle$ denotes the inner product of $L_2(U)$.

For a^{ij}, b^i and d in $L_\infty(U)$, A so defined is the infinitesimal generator of a C_0-semigroup S_A (say). It then follows that whenever $\underline{f} \in L_2(0,\infty; L_2(U))$ and $g \in L_2(U)$ the weak solution of (4) is given by

$$\underline{\theta}(t;\underline{f},g) = S_A(t)[g] + \int_0^t S_A(t-\alpha)[\underline{f}(\alpha)]d\alpha \quad , \quad \forall t \in [0, t_F]. \tag{7}$$

It is now assumed that $f(\boldsymbol{x},t) = f_S(\boldsymbol{x},t) + \boldsymbol{\beta_S}(\boldsymbol{x})^{\mathrm{T}}\boldsymbol{u}(t)$ where $f_S : U \times [0,t_F] \to \mathbb{R}$ and $\boldsymbol{\beta_S} : U \to \mathbb{R}^m$ are given functions and $\boldsymbol{u} : [0,t_F] \to \mathbb{R}^m$ is a control signal to be chosen in such a way as to make $\underline{\theta}(t_F;\underline{f},g)$ "close" to a prescribed $\theta_r \in L_2(U)$.

More specifically, let $\boldsymbol{u} \in L_2(0,t)^m$, $\rho_{\boldsymbol{u}} \in \mathbb{R}_+$ and define the cost functional

$$\mathcal{J}(\boldsymbol{u}) \triangleq \|\underline{\theta}(t_F;f,g) - \theta_r\|^2_{L_2(U)} + \rho_{\boldsymbol{u}}\|\boldsymbol{u}\|^2_{L_2(0,t_F)^m} \tag{8}$$

(from now on, the "space" subindices of norms and inner products will be omitted whenever context information makes then redundant).

A control signal is to be chosen on the basis of the optimization problem

$$\underline{\textit{Prob. I}:} \min_{\boldsymbol{u}\in L_2(0,t_F)^m} \mathcal{J}(\boldsymbol{u}). \tag{9}$$

3 Final State Positioning with Source Control

In this section, the optimal solution to *Prob. I* is explicitly characterized. To this effect, note first that due to the linearity of $\underline{\theta}(\cdot;\underline{f},g)$ on $(\underline{f},g)$,

$$\underline{\theta}(\cdot;\underline{f},g) = \underline{\theta}(\cdot;\underline{f}_S,g) + \underline{\theta}(\cdot;\underline{f}_{\boldsymbol{u}},0), \text{ where } \underline{f}_{\boldsymbol{u}}(t) = \boldsymbol{\beta}^{\mathrm{T}}_{\boldsymbol{S}}(\cdot)\boldsymbol{u}(t), \tag{10}$$

i.e.,

$$\underline{\theta}(\cdot;\underline{f},g) = \underline{\theta}(\cdot;\underline{f}_S,g) + \check{\mathcal{T}}_\theta[\boldsymbol{u}](\cdot), \tag{11}$$

where $\check{\mathcal{T}}_\theta : L_2(0,t_F)^m \to \{\underline{h} : [0,t_F] \to L_2(\mathcal{U})\}$

$$\check{\mathcal{T}}_\theta[\boldsymbol{u}](t) \triangleq \int_0^t S_A(t_F-\alpha)[\underline{f}_{\boldsymbol{u}}(\alpha)]d\alpha, \tag{12}$$

so that $\mathcal{J}(\boldsymbol{u})$ can be rewritten as

$$\mathcal{J}(\boldsymbol{u}) = \|\mathcal{T}_\theta[\boldsymbol{u}] - \theta_{ro}\|^2_{L_2(U)} + \rho_{\boldsymbol{u}}\|\boldsymbol{u}\|^2_{L_2(0,t_F)^m}, \tag{13}$$

where $\theta_{ro} \triangleq \theta_r - \underline{\theta}(t_F; f_S, g)$ and $\mathcal{T}_\theta : L_2(0,t_F)^m \to L_2(U)$ is defined by $\mathcal{T}_\theta[\boldsymbol{u}] = \check{\mathcal{T}}_\theta[\boldsymbol{u}](t_F)$.

The existence of an optimal solution to *Prob. I* can be ascertained by means of a basic result on minimum-distance problems pertaining to closed convex sets (Luenberger, 1963, p. 69), as stated in the next proposition.

Proposition 3.1. *There exists* $\boldsymbol{u}_o \in L_2(0,t_F)^m$ *such that* $\forall \boldsymbol{u} \in L_2(0,t_F)^m$, $\boldsymbol{u} \neq \boldsymbol{u}_o$, $\mathcal{J}(\boldsymbol{u}_o) < \mathcal{J}(\boldsymbol{u})$.

Moreover, $\boldsymbol{u}_o$ *is the unique solution of the linear equation*

$$\rho_{\boldsymbol{u}}\boldsymbol{u}_o + \mathcal{T}^*_\theta \cdot \mathcal{T}_\theta[\boldsymbol{u}_o] - \mathcal{T}^*_\theta[\theta_{ro}] = 0, \tag{14}$$

i.e.,

$$\boldsymbol{u}_o = [\rho_{\boldsymbol{u}} I + \mathcal{T}^*_\theta \cdot \mathcal{T}_\theta]^{-1}\left[\mathcal{T}^*_\theta[\theta_{ro}]\right], \tag{15}$$

where $\mathcal{T}^*_\theta : L_2(U) \to L_2(0,t_F)^m$ *is the adjoint of* $\mathcal{T}_\theta$. ∇

Proof. Let $\mathcal{T}_a : L_2(0,t_F)^m \to L_2(0,t_F)^m \times L_2(U)$ be defined by $\mathcal{T}_a[\boldsymbol{u}] \triangleq (\rho^{1/2}_{\boldsymbol{u}}\boldsymbol{u}, \mathcal{T}_\theta[\boldsymbol{u}])$. Then $\mathcal{J}(\boldsymbol{u}) = \|\mathcal{T}_a[\boldsymbol{u}] - (0,\theta_{ro})\|^2_{X_a}$, where $X_a \triangleq L_2(0,t_F)^m \times L_2(U)$, and *Prob. I* is seen as the problem of finding the minimum-distance approximation to $(0,\theta_{ro}) \in X_a$ in $\mathcal{T}_a[L_2(0,t_F)^m]$ - note that X_a is a Hilbert Space with the inner product

$$\langle (v_1,w_1),(v_2,w_2)\rangle_{X_a} = \langle v_1,v_2\rangle_{L_2(0,t_F)^m} + \langle w_1,w_2\rangle_{L_2(U)}.$$

Moreover, $\mathcal{T}_a[L_2(0,t_F)^m]$ is closed. Indeed, if $\mathcal{T}_a[\boldsymbol{u}_K] \to \boldsymbol{x}_0 = (\hat{\boldsymbol{u}}_o, \hat{\theta}_{ao})$ or, equivalently, $(\rho^{1/2}_{\boldsymbol{u}}\boldsymbol{u}_K, \mathcal{T}_\theta[\boldsymbol{u}_K]) \to (\hat{\boldsymbol{u}}_o, \hat{\theta}_{ao})$ then $\boldsymbol{u}_K \to \rho^{-1/2}_{\boldsymbol{u}}\hat{\boldsymbol{u}}_o$ and (since $\mathcal{T}_\theta$ is continuous) $\mathcal{T}_\theta[\boldsymbol{u}_K] \to \mathcal{T}_\theta[\rho^{-1/2}_{\boldsymbol{u}}\hat{\boldsymbol{u}}_o] = \hat{\theta}_{ao}$. Thus, $\mathcal{T}_a(\rho^{-1/2}_{\boldsymbol{u}}\hat{\boldsymbol{u}}_o) = (\hat{\boldsymbol{u}}_o, \mathcal{T}_\theta[\rho^{-1/2}_{\boldsymbol{u}}\hat{\boldsymbol{u}}_o]) = (\hat{\boldsymbol{u}}_o, \hat{\theta}_{ao}) = \boldsymbol{x}_0 \Rightarrow \boldsymbol{x}_0 \in \mathcal{T}_a[L_2(0,t_F)^m]$.

As $\mathcal{T}_a[L_2(0,t_F)^m]$ is also convex, it follows from Theorem 3.12.1 (Luenberger, 1969, pg. 69) that *Prob. I* has a unique solution $\boldsymbol{u}_o$ (say).

Note now that $\boldsymbol{u}_o$ is a solution to *Prob. I* $\Leftrightarrow \forall \delta \boldsymbol{u} \in L_2(0,t_F)^m$, $\mathcal{J}(\boldsymbol{u}_o) \leq \mathcal{J}(\boldsymbol{u}_o + \delta \boldsymbol{u})$

$$\Leftrightarrow \qquad \forall \delta \boldsymbol{u} \in L_2(0,t_F)^m \ ,$$

$$2\rho_{\boldsymbol{u}} \langle \boldsymbol{u}_o, \delta \boldsymbol{u} \rangle_{L_2(0,t_F)^m} + \rho_{\boldsymbol{u}} \|\delta \boldsymbol{u}\|^2_{L_2(0,t_F)^m} + 2\langle \mathcal{T}_\theta[\boldsymbol{u}_o] - \theta_{ro}, \mathcal{T}_\theta[\delta \boldsymbol{u}] \rangle + \|\mathcal{T}_\theta[\delta \boldsymbol{u}]\|^2_{L_2(U)} \geq 0$$

$$\Leftrightarrow \qquad \forall \delta \boldsymbol{u} \in L_2(0,t_F)^m \quad , \quad \langle \rho_{\boldsymbol{u}} \boldsymbol{u}_o + \mathcal{T}_\theta^* \cdot \mathcal{T}_\theta[\boldsymbol{u}_o] - \mathcal{T}_\theta^*[\theta_{ro}] \ , \ \delta \boldsymbol{u} \rangle_{L_2(0,t_F)^m} \geq 0$$

$$\Leftrightarrow \rho_{\boldsymbol{u}} \boldsymbol{u}_o + \mathcal{T}_\theta^* \cdot \mathcal{T}_\theta[\boldsymbol{u}_o] - \mathcal{T}_\theta^*[\theta_{ro}] = 0.$$

Thus, $\boldsymbol{u}_o$ is the unique solution of the linear equation (14). ■

Remark 3.1. *The final-state error achieved with a given control signal, namely,*

$$\|\underline{\theta}(t_F; \underline{f}_S + \boldsymbol{\beta}_{\boldsymbol{S}}^{\mathrm{T}} \boldsymbol{u}, g) - \theta_r\|_2^2 = \|\mathcal{T}_\theta[\boldsymbol{u}] - \theta_{ro}\|_2^2$$

can be written as

$$\|\mathcal{T}_\theta[\boldsymbol{u}] - \hat{\theta}_{ro}\|_2^2 + \|\theta_{ro} - \hat{\theta}_{ro}\|_2^2,$$

where $\hat{\theta}_{ro}$ denotes the $L_2(U)$–orthogonal projection of θ_{ro} on the closure of $\mathcal{T}_\theta[L_2(0,t_F)^m]$ in $L_2(U)$. Thus, by appropriately choosing control signals, the final-state error can be made arbitrarily close to

$$\inf \left\{ \|\mathcal{T}_\theta[\boldsymbol{u}] - \hat{\theta}_{ro}\|_2^2 \ : \ \boldsymbol{u} \in L_2(0,t_F)^m \right\} + \|\theta_{ro} - \hat{\theta}_{ro}\|_2^2.$$

In fact, this can be done with the optimal $\boldsymbol{u}_o(\rho_{\boldsymbol{u}})$ of Prob. I, *for decreasing values of $\rho_{\boldsymbol{u}}$. Indeed, taking $\varepsilon > 0$ and $\boldsymbol{u}_\varepsilon \in L_2(0,t_F)^m$ such that*

$$\|\mathcal{T}_\theta[\boldsymbol{u}_\varepsilon] - \hat{\theta}_{ro}\|_2^2 \leq \varepsilon \ , \ \textit{the fact that } \mathcal{J}(\boldsymbol{u}_o(\rho_{\boldsymbol{u}}); \rho_{\boldsymbol{u}}) \leq \mathcal{J}(\boldsymbol{u}_\varepsilon; \rho_{\boldsymbol{u}})$$

implies that

$$\rho_{\boldsymbol{u}} \|\boldsymbol{u}_o(\rho_{\boldsymbol{u}})\|^2_{L_2(0,t_F)^m} + \|\mathcal{T}_\theta[\boldsymbol{u}_o(\rho_{\boldsymbol{u}})] - \hat{\theta}_{ro}\|_2^2 \leq \rho_{\boldsymbol{u}} \|\boldsymbol{u}_\varepsilon\|^2_{L_2(0,t_F)^m} + \varepsilon.$$

Thus,

$$\forall \varepsilon > 0 \ , \forall \rho_{\boldsymbol{u}} > 0 \ , \ \|\mathcal{T}_\theta[\boldsymbol{u}_o(\rho_{\boldsymbol{u}})] - \hat{\theta}_{ro}\| \leq \rho_{\boldsymbol{u}} \|\boldsymbol{u}_\varepsilon\|^2_{L_2(0,t_F)^m} + \varepsilon$$

and, hence, $\lim_{\rho_{\boldsymbol{u}} \to 0} \|\mathcal{T}_\theta[\boldsymbol{u}_o(\rho_{\boldsymbol{u}})] - \hat{\theta}_{ro}\|_2^2 = 0.$ ∇

Proposition 3.1 above characterizes the optimal solution $\boldsymbol{u}_o$ in terms of the linear operators $\mathcal{T}_\theta$ and $\mathcal{T}_\theta^*$. To obtain explicit approximations to $\boldsymbol{u}_o$, the question naturally arises of considering finite-dimensional approximations to these operators and the corresponding version of equation (13). This is pursued in the next section.

4 Approximate Solutions

In this section, a sequence $\{\boldsymbol{u}_K\}$ is introduced which is defined on the basis of Galerkin approximations to the operator $\mathcal{T}_\theta$. It is then shown that under appropriate conditions this sequence converges to $\boldsymbol{u}_o$ in the $L_2(0, t_F)^m$–norm.

To this effect, let $\{X_K\}$ be a sequence of finite-dimensional subspaces of $H_0^1(U)$ with approximability property, i.e., such that $\forall \psi \in H_0^1(U)$ there exists a sequence $\{\psi_K\} \subset H_0^1(U)$ such that $\psi_K \in X_K$ and

$$\lim_{K \to \infty} \|\psi - \psi_K\|_{H_0^1(U)} = 0. \tag{16}$$

Let $\mathcal{A}_K : X_K \to X_K$ be such that

$$\forall \phi \in X_K, \forall \psi \in X_K, \quad \langle \mathcal{A}_K[\phi], \psi \rangle = -\mathbf{B}[\phi, \psi]$$

or, equivalently, for an orthonormal basis $\{\phi_1, \ldots, \phi_k\}$ of X_K,

$$\forall \phi \in X_K, \quad \mathcal{A}_K[\phi] = -\sum_{k=1}^{n} \mathbf{B}[\phi, \phi_k]\phi_k \quad \Leftrightarrow \quad \forall \ell = 1, \ldots, n, \quad \mathcal{A}_K[\phi_\ell] = -\sum_{k=1}^{n} \mathbf{B}[\phi_\ell, \phi_k]\phi_k.$$

Let then $\mathbf{A}_K \in \mathbb{R}^{n \times n}$ be defined by $\{\mathbf{A}_K\}_{\ell k} = -\mathbf{B}[\phi_\ell, \phi_k]$.

Let P_K be the orthogonal projection from $L_2(U)$ onto X_K and define

$$\mathcal{T}_\theta^K \ : \ L_2(0, t_F)^m \to X_K \ \text{ by } \ \mathcal{T}_\theta^K[\boldsymbol{u}] \triangleq \left[\int_0^{t_F} S_K(t_F - \tau) \left[P_K \left[\boldsymbol{\beta}_{\boldsymbol{S}}^{\mathrm{T}} \boldsymbol{u}(\tau)\right]\right] d\tau\right],$$

where S_K is the semigroup generated by $\mathcal{A}_K$.

Note that $\mathcal{T}_\theta^K[\boldsymbol{u}] = \sum_{k=1}^{n} c_k(t_F; \boldsymbol{u})\phi_k$, where $\bar{\boldsymbol{c}}_K(t; \boldsymbol{u}) = [c_1(t; \boldsymbol{u}), \ldots, c_n(t; \boldsymbol{u})]$ is given by $\bar{\boldsymbol{c}}_K(t; \boldsymbol{u}) = \int_0^t \exp[\mathbf{A}_K(t-\tau)]\mathbf{M}_{\boldsymbol{\beta}}^K \boldsymbol{u}(\tau)d\tau$, $\boldsymbol{\beta}_{\boldsymbol{S}}^{\mathrm{T}} = [\boldsymbol{\beta}_{\boldsymbol{S}1} \ \cdots \ \boldsymbol{\beta}_{\boldsymbol{S}m}]$ and

$$\mathbf{M}_{\boldsymbol{\beta}}^K \triangleq \begin{bmatrix} \langle \boldsymbol{\beta}_{\boldsymbol{S}1}, \phi_1 \rangle & \cdots & \langle \boldsymbol{\beta}_{\boldsymbol{S}m}, \phi_1 \rangle \\ \vdots & & \vdots \\ \langle \boldsymbol{\beta}_{\boldsymbol{S}1}, \phi_n \rangle & \cdots & \langle \boldsymbol{\beta}_{\boldsymbol{S}m}, \phi_n \rangle \end{bmatrix}.$$

The corresponding version of *Prob. I* is then defined by

$$\underline{\textit{Prob. } I_K :} \min_{\boldsymbol{u} \in L_2(0,t_F)^m} \mathcal{J}_K(\boldsymbol{u}), \tag{17}$$

where

$$\mathcal{J}_K[\boldsymbol{u}] \triangleq \|\mathcal{T}_\theta^K[\boldsymbol{u}] - \theta_{ro}\|_{L_2(U)}^2 + \rho_{\boldsymbol{u}} \|\boldsymbol{u}\|_{L_2(0,t_F)^m}^2$$

Similarly to what happens in the case of *Prob. I*, *Prob. I_K* has a unique solution $\boldsymbol{u}_K$ which is obtained from the optimality condition

$$\rho_{\boldsymbol{u}} \boldsymbol{u}_K + (\mathcal{T}_\theta^K)^*[\mathcal{T}_\theta^K[\boldsymbol{u}_K] - \theta_{ro}] = 0, \tag{18}$$

where the adjoint operator $(\mathcal{T}_\theta^K)^* : L_2(U) \to L_2(0, t_F)^m$ is such that

$$\forall \boldsymbol{u} \in L_2(0, t_F)^m, \ \forall \phi \in L_2(u), \ \langle \phi, \mathcal{T}_\theta^K[\boldsymbol{u}] \rangle = \langle (\mathcal{T}_\theta^K)^*[\phi], \boldsymbol{u} \rangle$$

$$\Leftrightarrow \quad \sum_{k=1}^{n} \langle \phi, \phi_k \rangle c_k(t_F; \boldsymbol{u}) = \bar{\boldsymbol{\phi}}_K^{\mathrm{T}} \bar{\boldsymbol{c}}_K(t_F; \boldsymbol{u}) = \int_0^{t_F} (\mathbf{F}_K(\tau)\bar{\boldsymbol{\phi}}_K)^{\mathrm{T}} \boldsymbol{u}(\tau) d\tau$$

so that $(\mathcal{T}_\theta^K)^*[\phi] = \mathbf{F}_K(\tau)\bar{\boldsymbol{\phi}}_K$, where $\bar{\boldsymbol{\phi}}_K^{\mathrm{T}} \triangleq [\langle \phi, \phi_1 \rangle \cdots \langle \phi, \phi_K \rangle]$ and

$$\mathbf{F}_K(\tau) \triangleq (\mathbf{M}_{\boldsymbol{\beta}}^K)^{\mathrm{T}} \exp[\mathbf{A}_K^{\mathrm{T}}(t_F - \tau)]. \tag{19}$$

To obtain $\boldsymbol{u}_K$ note that it follows from (18) that $\boldsymbol{u}_K$ belongs to the image of $(\mathcal{T}_\theta^K)^*$, i.e., there exists

$$\phi \in L_2(U) \text{ such that } \boldsymbol{u}_K = (\mathcal{T}_\theta^K)^*[\phi] = \mathbf{F}_K \bar{\boldsymbol{\phi}}_K,$$

i.e., there exists $\boldsymbol{\alpha}_K \in \mathbb{R}^n$ such that

$$\boldsymbol{u}_K = \mathbf{F}_K \boldsymbol{\alpha}_K. \tag{20}$$

It then follows from (18) that

$$\rho_{\boldsymbol{u}} \mathbf{F}_K \boldsymbol{\alpha}_K + \mathbf{F}_K \bar{\boldsymbol{c}}_K(t_F; \mathbf{F}_K \boldsymbol{\alpha}_K) - \mathbf{F}_K \bar{\boldsymbol{\theta}}_{ro}^K = 0 \tag{21}$$

a sufficient condition for which being

$$\rho_F \boldsymbol{\alpha}_K + \bar{\boldsymbol{c}}_K(t_F; \mathbf{F}_K \boldsymbol{\alpha}_K) - \bar{\boldsymbol{\theta}}_{ro}^K = \mathbf{0}, \tag{22}$$

where $\bar{\boldsymbol{\theta}}_{ro}^K \triangleq \left[\langle \phi_1, \theta_{ro} \rangle \cdots \langle \phi_{n(K)}, \theta_{ro} \rangle\right]^{\mathrm{T}}$.

Thus, as $\bar{\boldsymbol{c}}_K(t_F; \mathbf{F}_K \boldsymbol{\alpha}_K) = \mathbf{G}_K \boldsymbol{\alpha}_K$, where $\mathbf{G}_K \triangleq \int_0^{t_F} \mathbf{F}_K(\tau)^{\mathrm{T}} \mathbf{F}_K(\tau) d\tau$, (21) can be rewritten as $\rho_{\boldsymbol{u}} \boldsymbol{\alpha}_K + \mathbf{G}_K \boldsymbol{\alpha}_K = \bar{\boldsymbol{\theta}}_{ro}^K$ from which it follows that $\boldsymbol{\alpha}_K = (\rho_{\boldsymbol{u}} \mathbf{I} + \mathbf{G}_K)^{-1} \bar{\boldsymbol{\theta}}_{ro}^K$ and, hence,

$$\boldsymbol{u}_K(\tau) = \mathbf{F}_K(\tau)(\rho_{\boldsymbol{u}} \mathbf{I} + \mathbf{G}_K)^{-1} \bar{\boldsymbol{\theta}}_{ro}^K, \quad \tau \in [0, t_F]. \tag{23}$$

The remaining part of this section is devoted to proving that the sequence $\{\boldsymbol{u}_K\}$ converges to $\boldsymbol{u}_o$ in the sense of the $L_2(0, t_F)^m$– norm.

To this effect, consider the following proposition.

Proposition 4.1. *There exists a real sequence* $\{\eta_{\mathcal{T}}^K : K \in \mathbb{Z}_+\}$*such that*
(a) $\forall \boldsymbol{u} \in L_2(0, t_F)^m, \quad \|\mathcal{T}_\theta[\boldsymbol{u}] - \mathcal{T}_\theta^K[\boldsymbol{u}]\|_{L_2(U)} \leq \eta_{\mathcal{T}}^K \|\boldsymbol{u}\|_{L_2(0,t_F)^m}$.
(b) $\{\eta_{\mathcal{T}}^K\}$ *converges to zero.* ∇

Note now that $\mathcal{J}_K(\boldsymbol{u}) = \rho_{\boldsymbol{u}} \|\boldsymbol{u}\|_{L_2(0,t_F)}^2 + \|\mathcal{T}_\theta[\boldsymbol{u}] - \theta_{ro} - (\mathcal{T}_\theta[\boldsymbol{u}] - \mathcal{T}_\theta^K[\boldsymbol{u}])\|_2^2 \iff$ $\mathcal{J}_K(\boldsymbol{u}) = \mathcal{J}(\boldsymbol{u}) + \|\mathcal{T}_\theta[\boldsymbol{u}] - \mathcal{T}_\theta^K[\boldsymbol{u}]\|_2^2 - 2\langle \mathcal{T}_\theta[\boldsymbol{u}] - \theta_{ro}, \mathcal{T}_\theta[\boldsymbol{u}] - \mathcal{T}_\theta^K[\boldsymbol{u}] \rangle$.

As a result, with $E_{\mathcal{J}}^K(\boldsymbol{u}) \triangleq \mathcal{J}(\boldsymbol{u}) - \mathcal{J}_K(\boldsymbol{u})$, it follows from Proposition 4.1 that

$$|E_{\mathcal{J}}^K(\boldsymbol{u})| \leq (\eta_{\mathcal{T}}^K)^2 \|\boldsymbol{u}\|_{L_2(0,t_F)^m}^2 + 2\|\mathcal{T}_\theta[\boldsymbol{u}] - \theta_{ro}\|_2 \eta_{\mathcal{T}}^K \|\boldsymbol{u}\|_{L_2(0,t_F)^m}. \tag{24}$$

On the other hand,

$\mathcal{J}_K(\boldsymbol{u}_K) \leq \mathcal{J}_K(\boldsymbol{u}_o) = \mathcal{J}(\boldsymbol{u}_o) - E^K_{\mathcal{J}}(\boldsymbol{u}_o) \Longleftrightarrow \mathcal{J}(\boldsymbol{u}_K) - E^K_{\mathcal{J}}(\boldsymbol{u}_K) \leq \mathcal{J}(\boldsymbol{u}_o) - E^K_{\mathcal{J}}(\boldsymbol{u}_o)$

$\Longrightarrow \mathcal{J}(\boldsymbol{u}_K) \leq J(\boldsymbol{u}_o) - E^K_{\mathcal{J}}(\boldsymbol{u}_o) + E^K_{\mathcal{J}}(\boldsymbol{u}_K) \Longrightarrow$

$$\mathcal{J}(\boldsymbol{u}_K) \leq \mathcal{J}(\boldsymbol{u}_o) + |E^K_{\mathcal{J}}(\boldsymbol{u}_o)| + |E^K_{\mathcal{J}}(\boldsymbol{u}_K)|$$

$\Longrightarrow$ (since $\mathcal{J}(\boldsymbol{u}_K) \geq \mathcal{J}(\boldsymbol{u}_o)$)

$$0 \leq \mathcal{J}(\boldsymbol{u}_K) - \mathcal{J}(\boldsymbol{u}_o) \leq |E^K_{\mathcal{J}}(\boldsymbol{u}_o)| + |E^K_{\mathcal{J}}(\boldsymbol{u}_K)|. \tag{25}$$

Note also that, as $\eta^K_{\mathcal{T}} \to 0$ (Proposition 4.1(b)), it follows from (24) that $|E^K_{\mathcal{J}}(\boldsymbol{u}_o)| \to 0$. Moreover, $\{\boldsymbol{u}_K\}$ is a bounded sequence – indeed, $\|\boldsymbol{u}_K\|^2_{L_2(0,t_F)^m} \leq \|\boldsymbol{\theta}_{ro}\|^2_{L_2(U)}\rho^{-1}_{\boldsymbol{u}}$ for, if $\|\boldsymbol{u}_K\|^2 > \rho^{-1}_{\boldsymbol{u}}\|\boldsymbol{\theta}_{ro}\|^2_{L_2(U)}$ then $\mathcal{J}_K(\boldsymbol{u}_K) > \|\boldsymbol{\theta}_{ro}\|^2_{L_2(U)} = \mathcal{J}_K(0)$ in which case $\boldsymbol{u}_K$ would not be optimal for *Prob. I_K*. Thus, as $\mathcal{T}_\theta[\boldsymbol{u}] = \int_0^{t_F} S_A(t_F - \alpha)\{\sum_{i=1}^m \boldsymbol{\beta}_{\boldsymbol{S}i}\boldsymbol{u}(\alpha)\}d\alpha$, $\{\mathcal{T}_\theta[\boldsymbol{u}_K]\}$ is also bounded and, hence, it follows from (24) that (as $\eta^K_{\mathcal{T}} \to 0$) $E^K_{\mathcal{J}}(\boldsymbol{u}_K) \to 0$. Thus,

$$\left\{|E^K_{\mathcal{J}}(\boldsymbol{u}_o)| + |E^K_{\mathcal{J}}(\boldsymbol{u}_K)|\right\} \to 0 \tag{26}$$

which together with (25) implies that $\mathcal{J}(\boldsymbol{u}_K) \to \mathcal{J}(\boldsymbol{u}_o)$. Thus, the following corollary of Proposition 4.1 has been established

Corollary 4.1: $\mathcal{J}(\boldsymbol{u}_K) \to \mathcal{J}(\boldsymbol{u}_o)$. ∇

Moreover, as $\{\boldsymbol{u}_K\}$ is bounded and $\mathcal{J}(\boldsymbol{u}_K) \to \mathcal{J}(\boldsymbol{u}_o)$, the desired convergence of the approximate solutions $\{\boldsymbol{u}_K\}$ can be established, as stated in the following proposition.

Proposition 4.2. *The sequence $\{\boldsymbol{u}_K : K \in \mathbb{Z}_+\}$ of solutions to the approximate problems* Prob. I_K *converges to the solution $\boldsymbol{u}_o$ of* Prob. I *in the sense of the $L_2(0,t_F)^m$–norm.* ∇

Proof. Note first that (since $\boldsymbol{u}_o$ is an optimal solution of *Prob. I*)

$\mathcal{J}(\boldsymbol{u}_K) = \mathcal{J}(\boldsymbol{u}_o + (\boldsymbol{u}_K - \boldsymbol{u}_o)) = \mathcal{J}(\boldsymbol{u}_o) + \rho_{\boldsymbol{u}}\|\boldsymbol{u}_K - \boldsymbol{u}_o\|^2_{L_2(0,t_F)^m} + \|\mathcal{T}_\theta[(\boldsymbol{u}_K - \boldsymbol{u}_o)]\|^2_{L_2(U)}.$

It then follows from (25) that

$$\rho_{\boldsymbol{u}}\|\boldsymbol{u}_K-\boldsymbol{u}_o\|^2_{L_2(0,t_F)^m}+\|\mathcal{T}_\theta[(\boldsymbol{u}_K-\boldsymbol{u}_o)]\|^2_{L_2(U)}\leq|E^K_{\mathcal{J}}(\boldsymbol{u}_o)|+|E^K_{\mathcal{J}}(\boldsymbol{u}_K)|\Rightarrow$$

$$\rho_{\boldsymbol{u}}\|\boldsymbol{u}_K-\boldsymbol{u}_o\|^2_{L_2(0,t_F)^m}\leq|E^K_{\mathcal{J}}(\boldsymbol{u}_o)|+|E^K_{\mathcal{J}}(\boldsymbol{u}_K)|.$$

Thus, in the light of (26), $\boldsymbol{u}_K\to\boldsymbol{u}_o$ in $L_2(0,t_F)^m$.

■

5 Peak-value Constraints on Control Signals

In spite of the fact that the coefficient $\rho_{\boldsymbol{u}}$ can be manipulated with a view to keeping $\|\boldsymbol{u}_o\|_{L_2(0,t_F)^m}$ within acceptable levels, it is also desirable to ensure that $|u_{oi}(t)|$ remains within prescribed bounds. In this connection, a version of *Prob. I* with pointwise (with respect to t) constraints may be formulated as follows

$$\underline{\textit{Prob. II}:}\ \min_{\boldsymbol{u}\in L_2(0,t_F)^m}\mathcal{J}(\boldsymbol{u})$$
$$\text{subject to: }\forall i=1,\cdots,m,\forall t\text{ a.e. in }[0,t_F],u_{ai}(t)\leq u_i(t)\leq u_{b_i}(t),\quad(27)$$

where $u_{ai}\in C^0[0,t_F]$ and $u_{bi}\in C^0[0,t_F],u_{bi}(t)>u_{ai}(t)$.

The existence of an optimal solution to *Prob. II* can be ascertained by means of an argument entirely similar to the one used in connection with *Prob. I*. This leads to the next proposition.

Proposition 5.1. *Let* $\mathcal{I}_{Fi}(t)\triangleq[u_{ai}(t),u_{bi}(t)]$ *and*

$$S_{\boldsymbol{u}F}\triangleq\{\boldsymbol{u}\in L_2(0,t_F)^m:\forall i=1\cdots m,\ \forall t\ a.e.\ in\ [0,t_F],\ u_i(t)\in\mathcal{I}_{Fi}(t)\}$$

There exists $\boldsymbol{u}_c\in S_{\boldsymbol{u}F}$ *such that* $\forall\boldsymbol{u}\in S_{\boldsymbol{u}F}$, $\boldsymbol{u}\neq\boldsymbol{u}_c$, $\mathcal{J}(\boldsymbol{u}_c)<\mathcal{J}(\boldsymbol{u})$.

Moreover, $\boldsymbol{u}_c$ *satisfies the following optimality condition:*
$\forall\delta\boldsymbol{u}$ *such that* $(\boldsymbol{u}_c+\delta\boldsymbol{u})\in S_{\boldsymbol{u}F}$, $\langle\rho_{\boldsymbol{u}}\boldsymbol{u}_c+\mathcal{T}^*_\theta[\mathcal{T}_\theta-\theta_{r0}]\ ,\ \delta_{\boldsymbol{u}}\rangle_{L_2(0,t_F)^m}\geq 0$ *or, equivalently,* $\forall\boldsymbol{u}$ *such that* $(\boldsymbol{u}_c+\delta\boldsymbol{u})\in S_{\boldsymbol{u}F}$, $\forall t$ *a.e. in* $[0,t_F]$ $\left\{(\rho_{\boldsymbol{u}}\boldsymbol{u}_c+\mathcal{T}^*_\theta[\mathcal{T}_\theta[\boldsymbol{u}_c]-\theta_{r0}])^{\mathrm{T}}\delta\boldsymbol{u}\right\}(t)\geq 0.$ ∇

A characterization of the optimal solution which does not involve quantifiers is now presented for which the following "saturation" operators are required: for $i = 1, \cdots, m$ define $P_{Ii} : L_2(0, t_F) \to L_2(0, t_F)$

$$\begin{aligned} P_{Ii}[v](t) &= v(t) \quad \text{if} \quad v(t) \in \mathcal{I}_{Fi}(t) \\ P_{Ii}[v](t) &= u_{ai}(t) \quad \text{if} \quad v(t) < u_{ai}(t) \\ P_{Ii}[v](t) &= u_{bi}(t) \quad \text{if} \quad v(t) > u_{bi}(t) \end{aligned}$$

Proposition 5.2. *Let* $Z_a[\boldsymbol{u}] \triangleq \mathcal{T}_\theta^*[\mathcal{T}_\theta[\boldsymbol{u}] - \theta_{r0}]$. *For* $i = 1, \cdots, m$
$\{\boldsymbol{u}_c\}_i = P_{Ii}[-(1/\rho_{\boldsymbol{u}})\{Z_a[\boldsymbol{u}_c]\}_i]$ *a.e. in* $[0, t_F]$. ∇

In the light of Proposition 5.2, the problem of computing (approximations to) $\boldsymbol{u}_c$ is reduced to (approximately) solving a "system" of equations in $L_2(0, t_F)$ with respect to $\boldsymbol{u} \in L_2(0, t_F)^m$, the $i - th$ one of which is based on P_{Ii}. As the latter is also an involved problem, even if $\mathcal{T}_\theta$ is replaced by an approximation $\mathcal{T}_\theta^K$, motivation arises for bringing duality considerations to bear on approximations of *Prob. II*.

To this effect, let $\mathcal{J}_K(\boldsymbol{u}) \triangleq \rho_{\boldsymbol{u}} \|\boldsymbol{u}\|^2_{L_2(0,t_F)^m} + \|\mathcal{T}_\theta^K[\boldsymbol{u}] - \theta_{ro}\|_2^2$ and consider.

$$\underline{Prob.\ II_K:} \ \min_{\boldsymbol{u} \in S_{\boldsymbol{u}F}} \mathcal{J}_K(\boldsymbol{u}). \tag{28}$$

Approximate solutions to *Prob. II* can be obtained on the basis of *Prob.* II_K, as stated in the following proposition.

Proposition 5.3. (a) $\forall K \in \mathbb{Z}_+$ *there exists* $\boldsymbol{u}_c^K \in S_{\boldsymbol{u}F}$ *such that* $\forall \boldsymbol{u} \in S_{\boldsymbol{u}F}$, $\boldsymbol{u} \neq \boldsymbol{u}_c^K$, $\mathcal{J}_K(\boldsymbol{u}_c^K) < \mathcal{J}_K(\boldsymbol{u})$.
(b) $\boldsymbol{u}_c^K \to \boldsymbol{u}_c$ *in* $L_2(0, t_F)^m$. ∇

A counterpart of Proposition 5.1 for *Prob.* II_K leads to

$$\{\boldsymbol{u}_c^K\}_i = P_{Ii}[-(1/\rho_{\boldsymbol{u}})\{Z_a^K[\boldsymbol{u}]\}_i] \tag{29}$$

where $Z_a^K[\boldsymbol{u}] \triangleq (\mathcal{T}_\theta^K)^*[\mathcal{T}_\theta^K[\boldsymbol{u}] - \theta_{r0}]$. However, instead of directly tackling equation (29) in order to obtain approximations to $\boldsymbol{u}_c^K$, duality considerations pertaining to *Prob.* II_K are now introduced.

To this effect, note first that $u_i(t) \in \mathcal{I}_{Fi}(t) \Leftrightarrow u_{ai}(t) - u_i(t) \leq 0$ and $u_i(t) - u_{bi}(t) \leq 0$, so that a Lagrangian functional for *Prob.* II_K can be defined by

$$Lag_K(\boldsymbol{u}, \boldsymbol{\lambda}) \triangleq \mathcal{J}_K(\boldsymbol{u}) + 2\langle \boldsymbol{\lambda}_a, \boldsymbol{u}_a - \boldsymbol{u} \rangle_{L_2(0,t_F)^m} + 2\langle \boldsymbol{\lambda}_b, \boldsymbol{u} - \boldsymbol{u}_b \rangle_{L_2(0,t_F)^m} \ , \tag{30}$$

where $\boldsymbol{\lambda}_a \in L_2(0, t_F)^m$, $\boldsymbol{\lambda}_b \in L_2(0, t_F)^m$ and $\boldsymbol{\lambda} = (\boldsymbol{\lambda}_a, \boldsymbol{\lambda}_b)$.

The corresponding dual functional and dual problem are given by

$$\varphi_{DK}(\boldsymbol{\lambda}) \triangleq \min\{Lag_K(\boldsymbol{u}, \boldsymbol{\lambda}) : \boldsymbol{u} \in L_2(0, t_F)^m\} \tag{31}$$

and $\underline{Prob.\ II_{DK}:} \max_{\boldsymbol{\lambda} \in S_{\boldsymbol{\lambda}}} \varphi_{DK}(\boldsymbol{\lambda}),$

where

$S_{\boldsymbol{\lambda}} \triangleq \{(\boldsymbol{\lambda}_a, \boldsymbol{\lambda}_b) : \boldsymbol{\lambda}_a \in L_2(0, t_F)^m, \boldsymbol{\lambda}_b \in L_2(0, t_F)^m, \forall t \text{ a.e. in } [0, t_F], \boldsymbol{\lambda}_{ai}(t) \geq 0, \boldsymbol{\lambda}_{bi}(t) \geq 0\}$.

The following proposition is a direct consequence of Theorem 1 in Luenberger (1969), pp. 224.

Proposition 5.4. **(a)** $\sup\{\varphi_{DK}(\boldsymbol{\lambda}) : \boldsymbol{\lambda} \in S_{\boldsymbol{\lambda}}\} = \min\{\mathcal{J}_K(\boldsymbol{u}) : \boldsymbol{u} \in S_{\boldsymbol{u}F}\}$.
(b) *Let* $\boldsymbol{u}_c^K(\boldsymbol{\lambda})$ *be the unique solution of* $\min_{\boldsymbol{u} \in L_2(0,t_F)^m} Lag_K(\boldsymbol{u}, \boldsymbol{\lambda})$. *Then* $\boldsymbol{u}_c^K = \boldsymbol{u}_c^K(\boldsymbol{\lambda}_K)$, *where* $\boldsymbol{\lambda}_K = \arg\max_{\boldsymbol{\lambda} \in S_{\boldsymbol{\lambda}}} \varphi_{DK}(\boldsymbol{\lambda})$. ∇

To rely on Proposition 5.3 to obtain approximate solutions to *Prob.* II_K explicit characterizations of both $\boldsymbol{u}_c^K(\boldsymbol{\lambda})$ and $\varphi_{DK}(\boldsymbol{\lambda})$ are presented in the next proposition.

Proposition 5.5. *For any* $\boldsymbol{\lambda} = (\boldsymbol{\lambda}_a, \boldsymbol{\lambda}_b) \in S_{\boldsymbol{\lambda}}$

$$\boldsymbol{u}_c^K[\boldsymbol{\lambda}] = \boldsymbol{u}_K - \mathbf{F}_K(I - (\rho_{\boldsymbol{u}}\mathbf{I} + \mathbf{G}_K)^{-1})\boldsymbol{\alpha}_{\boldsymbol{\lambda}}^K + (\boldsymbol{\lambda}_a - \boldsymbol{\lambda}_b), \ \textit{and}$$

$$\varphi_{DK}(\boldsymbol{\lambda}) = \|\theta_{ro}\|_{L_2(U)}^2 + \langle \mathcal{T}_\theta^K[\boldsymbol{u}_K], -\theta_{ro}\rangle + \hat{\varphi}_{DK}(\boldsymbol{\lambda}),$$

where

$$\hat{\varphi}_{DK}(\boldsymbol{\lambda}) \triangleq -\rho_{\boldsymbol{u}}^{-1}\langle \boldsymbol{\lambda}_{ab}, \boldsymbol{\lambda}_{ab}\rangle + \rho_{\boldsymbol{u}}^{-1}\langle \boldsymbol{\xi}_{\boldsymbol{\lambda}}^K, (\rho_{\boldsymbol{u}}\mathbf{I} + \mathbf{G}_K)^{-1}\boldsymbol{\xi}_{\boldsymbol{\lambda}}^K\rangle - 2\langle \boldsymbol{\xi}_{\lambda}^K, \boldsymbol{\alpha}_K\rangle + 2\langle \boldsymbol{\lambda}_a, \boldsymbol{u}_a\rangle - 2\langle \boldsymbol{\lambda}_b, \boldsymbol{u}_b\rangle \ ,$$

$$\boldsymbol{\lambda}_{ab} = \boldsymbol{\lambda}_a - \boldsymbol{\lambda}_b, \quad \boldsymbol{\xi}_{\boldsymbol{\lambda}}^K \triangleq \int_0^{t_F} \mathbf{F}_k^{\mathrm{T}}(\tau)(\boldsymbol{\lambda}_a(\tau) - \boldsymbol{\lambda}_b(\tau))d\tau \quad \textit{and} \ \ \mathbf{G}_K\boldsymbol{\alpha}_{\boldsymbol{\lambda}}^K = \boldsymbol{\xi}_{\boldsymbol{\lambda}}^K. \qquad \nabla$$

It follows from Proposition 5.5 that $\boldsymbol{\lambda}_K = \arg\max_{\boldsymbol{\lambda}\in S_{\boldsymbol{\lambda}}} \hat{\varphi}_{DK}(\boldsymbol{\lambda})$ is the solution of a quadratic problem in $L_2(0,t_F)^m$ with non-negativeness constraints on the values $\boldsymbol{\lambda}(t)$ (a.e. in $[0,t_F)$). This suggests that approximate solutions $\hat{\boldsymbol{\lambda}}_K$ for this problem should be sought on the basis of which the corresponding approximate solutions $\boldsymbol{u}_c^K[\hat{\boldsymbol{\lambda}}_K]$ can be readily obtained in the light of Proposition 5.5.

Remark 5.1. *Although $\boldsymbol{u}_c^K[\hat{\boldsymbol{\lambda}}_K]$ may fail to be in the feasible set $S_{\boldsymbol{u}F}$, a closely-related feasible solution $\boldsymbol{u}_K^R[\hat{\boldsymbol{\lambda}}_K]$ can also be obtained on the on the basis of Proposition 5.5. To this effect, let $\boldsymbol{u}_K^o[\boldsymbol{\lambda}] = \boldsymbol{u}_K - \mathbf{F}_K(\mathbf{I}-(\rho_{\boldsymbol{u}}\mathbf{I}+\mathbf{G}_K)^{-1})\boldsymbol{\alpha}_{\boldsymbol{\lambda}}^K$ so that $\boldsymbol{u}_c^K[\boldsymbol{\lambda}] = \boldsymbol{u}_K^o[\boldsymbol{\lambda}]+(\boldsymbol{\lambda}_a-\boldsymbol{\lambda}_b)$; define $\boldsymbol{u}_K^R[\boldsymbol{\lambda}]$ by:*

$$\begin{aligned}
&\forall t\in[0,t_F] \ \ \textit{such that} \ \ \{\boldsymbol{u}_K^o[\boldsymbol{\lambda}(t)]\}_i \in I_{Fi}(t), \ \ \left\{\boldsymbol{u}_K^R[\boldsymbol{\lambda}(t)]\right\}_i = \{\boldsymbol{u}_K^o[\boldsymbol{\lambda}(t)]\}_i\,,\\
&\forall t\in[0,t_F] \ \ \textit{such that} \ \ \{\boldsymbol{u}_K^o[\boldsymbol{\lambda}(t)]\}_i < \boldsymbol{u}_{ai}(t), \ \ \left\{\boldsymbol{u}_K^R[\boldsymbol{\lambda}(t)]\right\}_i = \boldsymbol{u}_{ai}(t),\\
&\forall t\in[0,t_F] \ \ \textit{such that} \ \ \{\boldsymbol{u}_K^o[\boldsymbol{\lambda}(t)]\}_i > \boldsymbol{u}_{bi}(t), \ \ \left\{\boldsymbol{u}_K^R[\boldsymbol{\lambda}(t)]\right\}_i = \boldsymbol{u}_{bi}(t).
\end{aligned}$$

For any $\boldsymbol{\lambda}$, $\boldsymbol{u}_K^R[\boldsymbol{\lambda}] \in S_{\boldsymbol{u}F}$; moreover, due to the so-called KKT optimality conditions for Prob. *II_K, $\boldsymbol{u}_K^R[\boldsymbol{\lambda}_K] = \boldsymbol{u}_c^K[\boldsymbol{\lambda}_K]$ (i.e., at $\boldsymbol{\lambda}=\boldsymbol{\lambda}_K$, $\boldsymbol{u}_K^R$ equals the optimal solution of* Prob. *II_K). In addition, given $\hat{\boldsymbol{\lambda}}_K$, the assessment of $\boldsymbol{u}_K^R[\hat{\boldsymbol{\lambda}}_K]$ as an approximate solution to* Prob. *II_K can be carried out on the basis of the inequality $\varphi_{DK}(\hat{\boldsymbol{\lambda}}_K) \le \mathcal{J}_K(\boldsymbol{u}_c) \le \mathcal{J}(\boldsymbol{u}_K^R[\hat{\boldsymbol{\lambda}}_K])$ so that whenever $\varphi_{DK}(\hat{\boldsymbol{\lambda}}_K)$ and $\mathcal{J}(\boldsymbol{u}_K^R[\hat{\boldsymbol{\lambda}}_K])$ are "close" $\boldsymbol{u}_K^R[\hat{\boldsymbol{\lambda}}_K]$ can be regarded as an "approximate" solution to* Prob. *II_K. In the next section this approach is illustrated in two simple numerical examples.* ∇

6 Actuator Location

It is often the case that the spatial effect of the control signals $\boldsymbol{u}_i$, $i=1,\dots,m$, have a local character due to the functions $\boldsymbol{\beta}_{\boldsymbol{S}i}$ only having non-zero value on "small" subsets of the spatial domain U. In such cases, the "location" of each $\boldsymbol{u}_i$ (i.e., the "centre" of the support of $\boldsymbol{\beta}_{\boldsymbol{S}i}$) may have significant effects on the magnitude of the final-state approximation error attained with the optimal $\boldsymbol{u}$.

More specifically, assume that U is symmetric with respect to $x_{\boldsymbol{a}} \in U$ and let $U_{\boldsymbol{\beta}} \subset U$ be an open and connected set also centred on $x_{\boldsymbol{a}}$. Let $\boldsymbol{\beta}_{\boldsymbol{a}} : U \to \mathbb{R}$ be such that $\forall\, x \in U - U_{\boldsymbol{\beta}}$,

$\boldsymbol{\beta}_{\boldsymbol{a}}(x) = 0$ (i.e., $U_{\boldsymbol{\beta}}$ is the support of $\boldsymbol{\beta}_{\boldsymbol{a}}$) and for a list $\underline{\mathcal{X}}$ of locations $\mathcal{X}_i$, $\underline{\mathcal{X}} = (\mathcal{X}_1, \ldots, \mathcal{X}_m)$, $\mathcal{X}_i \in U$ and such that $U_{\boldsymbol{\beta}} + (\mathcal{X}_i - x_{\boldsymbol{a}}) \subset U$, define $\boldsymbol{\beta}_{\boldsymbol{S}i}(\cdot; \mathcal{X}_i) : U \to \mathbb{R}$ by $\forall\ x \in U$, $\boldsymbol{\beta}_{\boldsymbol{S}i}(x; \mathcal{X}_i) \triangleq \boldsymbol{\beta}_{\boldsymbol{a}}(x - (\mathcal{X}_i - x_{\boldsymbol{a}}))$ – note that $U_{\boldsymbol{\beta}} + (\mathcal{X}_i - x_{\boldsymbol{a}})$ is the support of $\boldsymbol{\beta}_{\boldsymbol{S}i}(\cdot; \mathcal{X}_i)$.

Recall that the approximation error magnitude is given by
$\|\mathcal{T}_\theta^K[\boldsymbol{u}_K] - \theta_{ro}^K\|_2 = \|\underline{c}_K(t_F; \boldsymbol{u}_K) - \overline{\theta}_{ro}^K\|_2$ where $\boldsymbol{u}_K = \mathbf{F}_K \boldsymbol{\alpha}_K$, $\underline{c}_K(t_F; \boldsymbol{u}_K) = \mathbf{G}_K \boldsymbol{\alpha}_K$ and $\boldsymbol{\alpha}_K = (\rho_{\boldsymbol{u}}\mathbf{I} + \mathbf{G}_K)^{-1}\overline{\theta}_{ro}^K$.

Thus, $\|\mathcal{T}_\theta^K[\boldsymbol{u}_K] - \overline{\theta}_{ro}^K]\|_2 = \|\{\mathbf{G}_K(\rho_{\boldsymbol{u}}\mathbf{I} + \mathbf{G}_K)^{-1} - \mathbf{I}\}\overline{\theta}_{ro}^K\|_2 = \|(\mathbf{I} + \rho_{\boldsymbol{u}}^{-1}\mathbf{G}_K)^{-1}\overline{\theta}_{ro}^K\|_2$, since $\mathbf{G}_K(\rho_{\boldsymbol{u}}\mathbf{I} + \mathbf{G}_K)^{-1} = \rho_{\boldsymbol{u}}^{-1}\mathbf{G}_K(\mathbf{I} + \rho_{\boldsymbol{u}}^{-1}\mathbf{G}_K)^{-1} = \mathbf{I} - (\mathbf{I} + \rho_{\boldsymbol{u}}^{-1}\mathbf{G}_K)^{-1}$.

Thus, to choose actuator locations with the purpose of obtaining a good final-state approximation, a natural formulation for the actuator location problem would be:

$$\underline{\textit{Prob. Loc.:}} \quad \min_{\underline{\mathcal{X}} = (\mathcal{X}_1, \ldots, \mathcal{X}_m),\ U_{\boldsymbol{\beta}} + (\mathcal{X}_i - x_{\boldsymbol{a}}) \subset U} \nu(\underline{\mathcal{X}}), \tag{32}$$

where $\nu(\underline{\mathcal{X}}) \triangleq \|\{\mathbf{I} + \rho_F^{-1}\mathbf{G}_K(\mathbf{M}_{\boldsymbol{\beta}}^K(\underline{\mathcal{X}}))\}^{-1}\overline{\theta}_{ro}^K\|_2^2$, $\mathbf{G}_K(\mathbf{M}) \triangleq \displaystyle\int_0^{t_F} \exp[\mathbf{A}_K t]\mathbf{M}\mathbf{M}^{\mathrm{T}} \exp[\mathbf{A}_K^{\mathrm{T}} t]\ dt$,

$$\mathbf{M}_{\boldsymbol{\beta}}^K(\underline{\mathcal{X}}) = \begin{bmatrix} \langle \boldsymbol{\beta}_{\boldsymbol{S}1}(\mathcal{X}_1), \phi_1 & \cdots & \langle \boldsymbol{\beta}_{\boldsymbol{S}m}(\mathcal{X}_m), \phi_1 \rangle \\ \vdots & & \vdots \\ \langle \boldsymbol{\beta}_{\boldsymbol{S}1}(\mathcal{X}_1), \phi_K & \cdots & \langle \boldsymbol{\beta}_{\boldsymbol{S}m}(\mathcal{X}_m), \phi_K \rangle \end{bmatrix}.$$

Remark 6.1. *The problem formulation above hinges upon the approximation error attained with the optimal, unconstrained control signal* $\boldsymbol{u}_k$. *It is also natural to focus on the constrained optimal control signal* $\boldsymbol{u}_K^{\mathrm{c}}$, *in which case* $\nu(\cdot)$ *would be replaced by* $\nu_c(\cdot)$ *in the formulation of* Prob. Loc. *by* $\nu_c(\underline{\mathcal{X}}) = \|\underline{c}_K(t_F; \boldsymbol{u}_K^{\mathrm{c}}) - \overline{\theta}_{ro}^K\|_2$. ∇

Remark 6.2. *Those two choices of cost functional for the actuator location problem are "tuned" to a given final-state target* $\overline{\theta}_{ro}^K$. *Alternatively, if any final-state in a "broad" class may be targeted with the same actuator-location arrangement, a natural choice for the cost-functional of* Prob. Loc. *would be* $\nu_s(\underline{\mathcal{X}}) \triangleq \|(\mathbf{I} + \rho_{\boldsymbol{u}}^{-1}\mathbf{G}_K(\mathbf{M}_{\boldsymbol{\beta}}^K(\underline{\mathcal{X}})))^{-1}\|_s$. *This would be relevant for both* $\boldsymbol{u}_K$ *and* $\boldsymbol{u}_K^{\mathrm{c}}$ *for, in the case of* $\boldsymbol{u}_K$, *it yields an upper bound on* $\nu(\underline{\mathcal{X}})$ *for any* $\overline{\theta}_{ro}^K$ *with euclidean norm smaller or equal to a pre-specified value; whereas, in the case of* $\boldsymbol{u}_K^{\mathrm{c}}$, *as* $\boldsymbol{u}_K(\tau) = \mathbf{F}_K(\tau)\rho_{\boldsymbol{u}}^{-1}(\mathbf{I} + \rho_{\boldsymbol{u}}^{-1}\mathbf{G}_K(\mathbf{M}_{\boldsymbol{\beta}}^K(\underline{\mathcal{X}})))^{-1}\overline{\theta}_{ro}^K$, *making* $\nu_s(\underline{\mathcal{X}})$ *"small" tends to make the values of* $\boldsymbol{u}(\cdot)$ *smaller thereby mitigating the increase in the approximation error* *magnitude due to the enforcement of peak-value constraints.* ∇

The possible effect of actuator locations on the controlled final state is illustrated in Figures 13, 14 and 15 below for the case of the one-dimensional heat equation with one scalar control signal (i.e., $\boldsymbol{u}(t) \in \mathbb{R}$). Three locations are considered: a central one and two others symmetrically situated with respect to the centre of $U = (0, L_x)$ (i.e., $x = L_x/2$) and close to the boundary ∂U. In this case, with the desired final state also symmetric with respect to $x = L_x/2$ and for the approximating subspaces $S_K = span\{\sqrt{2/L}\sin\left((\pi/L_x)x\right), \ldots, \sqrt{2/L}\sin\left((K\pi/L_x)x\right)\}$, it can be shown that $\mathcal{X}_0 = L_x/2$ is a local extremum for $\nu(\cdot)$. It can be observed that the central location yields significantly better approximations for the desired final state than those provided by the two other locations taken into account–this is the case for both $\boldsymbol{u}_K(\mathcal{X})$ and $\boldsymbol{u}_K^c(\mathcal{X})$.

In general, solving *Prob. Loc.* (even for the cost-functional $\nu(\cdot)$) is a difficult task as global optimization techniques are required to obtain a solution on U^m and $\nu(\cdot)$ depends on $\underline{\mathcal{X}}$ in an intricate manner (through the inverse of $(\mathbf{I} + \rho_{\boldsymbol{u}}^{-1}\mathbf{G}_K(\mathbf{M}_{\boldsymbol{\beta}}^K(\underline{\mathcal{X}})))$ with $\mathbf{G}_K(\mathbf{M})$ depending on $\mathbf{M}\mathbf{M}^{\mathrm{T}}$ and $\{\mathbf{M}_{\beta}^K(\underline{\mathcal{X}})\}_{\ell k} = \langle \boldsymbol{\beta}_{\boldsymbol{S}\ell}(\mathcal{X}_\ell), \phi_k \rangle$). Although a grid search would seem feasible in the physically motivated cases of $n-$dimensional spatial domains with $n = 1, 2, 3$, it is noted that with N_g points along each dimension, the number of possible actuator locations arrangement would be $(N_g^n)^m$. To perform a less demanding search, optimization objectives may be weakened so that a randomly-generated sample of possible actuator-location arrangements is examined with the sample size being specified on the basis of probabilistic considerations–this approach has attracted considerable attention in the control literature (cf. Tempo & Ishii (2007) and references therein). The sample size calculations of interest here are presented below.

6.1 Sample Size for Random Search

Let x be a continuous, $n-$dimensional random variable with probability density function (pdf) p_x the support of which is denoted by $S_x \subset \mathbb{R}^n$. Let $f : S_x \to \mathbb{R}_+^c$ be continuous and such that $\forall\ w \in f(S_x)$ the set $\{x \in S_x : f(x) = w\}$ has zero Lebesgue measure. Let f_* be defined as $f_* = \inf\{f(x) : x \in S_x\}$. For a given $\varepsilon \in (0, 1)$ define $\delta_\varepsilon > 0$ by $Pr\{x \in S_x : f(x) \geq f_* + \delta_\varepsilon\} = 1 - \varepsilon$. Note that $\delta_\varepsilon \to 0$ as $\varepsilon \to 0$.

Let $\{x_i : i = 1, \ldots, N\}$ be (a sample of) independent and identically distributed random variables with pdf p_x and define $f_*^N \triangleq \min\{f(x_i) : i = 1, \ldots, N\}$. For a given $\alpha \in (0, 1)$, N is to be chosen so that

$$Pr\{f_*^N < f_* + \delta_\varepsilon\} \geq 1 - \alpha. \tag{33}$$

To this effect, note that

$$\begin{aligned} Pr\{f_*^N < f_* + \delta_\varepsilon\} &= 1 - Pr\{f_*^N \geq f_* + \delta_\varepsilon\} = 1 - Pr\left\{\bigcap_{i=1}^N \{x_i \in S_x : f(x_i) \geq f_* + \delta_\varepsilon\}\right\} \Leftrightarrow \\ &= 1 - \prod_{i=1}^N Pr\{x_i \in S_x : f(x_i) \geq f_* + \delta_\varepsilon\} \\ &= 1 - \{Pr\{x \in S_x : f(x) \geq f_* + \delta_\varepsilon\}\}^N \end{aligned}$$

$\Leftrightarrow Pr\{f_*^N < f_* + \varepsilon\} = 1 - (1 - \varepsilon)^N$.

Thus, (33) holds if and only if

$$\begin{aligned} &1 - (1 - \varepsilon)^N \geq 1 - \alpha \quad \Leftrightarrow \quad \alpha \geq (1 - \varepsilon)^N \quad \Leftrightarrow \quad \log \alpha \geq N \log(1 - \varepsilon) \\ &\Leftrightarrow \quad N \geq N_{\alpha\varepsilon} \triangleq \log \alpha / \log(1 - \varepsilon) = \frac{\log(1/\alpha)}{\log(1/(1 - \varepsilon))}. \end{aligned}$$

Thus, roughly speaking, in the case of a uniform pdf on S_x, for $N \geq N_{\alpha\varepsilon}$ the probability that f_*^N is smaller than "the values of $f(x)$ on $(1 - \varepsilon) \times 100\%$ of S_x" is greater that $(1 - \alpha)$.

In Section 7.3, an example is presented to illustrate the potential of such a random search to choose the locations of two "actuators" in connection with the heat equation on a two-dimensional spatial domain.

7 Examples and Numerical Results

In this section, two simple numerical examples are presented to illustrate the way the results above can be used to characterize control signals which aim at steering a solution of a PDEE over a given interval $[0, t_F]$ towards a prescribed final state.

7.1 A One–dimensional Example

Let $U = (0, L_x)$ and consider the one-dimensional heat equation with homogeneous Dirichlet boundary conditions and single-point control $\boldsymbol{u} : [0, t_F] \to \mathbb{R}$, i.e.,

$$\forall\ x \in U,\ \forall\ t \in (0, \infty)\ ,\ \frac{\partial \theta}{\partial t}(x,t) = k_\alpha \frac{\partial^2 \theta}{\partial x^2}(x,t) + \boldsymbol{\beta}_{\boldsymbol{S}}(x)\boldsymbol{u}(t)$$

$$\begin{array}{lll} \forall x \in U\ , & \theta(x,0) = 0 & \text{(zero initial condition)} \\ \forall\ t \in (0,\infty)\ , & \theta(0,t) = \theta(L_x,t) = 0 & \text{(boundary conditions)} \end{array}$$

with the corresponding weak version given by

$$\begin{aligned} \forall i = 1,2,\ldots, \left\langle \frac{\partial \theta}{\partial t}(\cdot,t), \phi_i \right\rangle &= -k_\alpha \left\langle \frac{\partial \theta}{\partial x}(\cdot,t), \frac{\partial \phi_i}{\partial x} \right\rangle + \langle \boldsymbol{\beta}_{\boldsymbol{S}}, \phi_i \rangle\, \boldsymbol{u}(t) \\ \langle \theta(\cdot,0), \phi_i \rangle &= 0, \end{aligned}$$

where $\phi_k : [0, L_x] \to \mathbb{R}$ is given by $\phi_k(x) = \sqrt{\frac{2}{L_x}} \sin\left(\frac{k\pi x}{L_x}\right)$.

Approximate solutions $\boldsymbol{u}_K$ and $\boldsymbol{u}_K^c$ are sought to the problems

$$\underline{\textit{Prob. I}:}\ \min_{\boldsymbol{u} \in L_2(0,t_F)} \check{\mathcal{J}}(\boldsymbol{u};\rho_F) \quad \text{or} \quad \underline{\textit{Prob. I}_c:}\ \min_{\boldsymbol{u} \in S_{\boldsymbol{u}F}} \check{\mathcal{J}}(\boldsymbol{u};\rho_F),$$

where $\check{\mathcal{J}}(\boldsymbol{u};\rho_F) = \|\boldsymbol{u}\|^2_{L_2(0,t_F)} + \rho_F \|\mathcal{T}_\theta[\boldsymbol{u}] - \theta_{ro}\|_2^2$, θ_{ro} is the final state to be approximately reached and

$$S_{\boldsymbol{u}F} = \left\{\boldsymbol{u} \in L_\infty(0,t_F) : \|\boldsymbol{u}\|_{L_\infty(0,t_F)} \le \mu_{\boldsymbol{u}}\right\}.$$

In this case, $\{\mathbf{A}_K\}_{k\ell} = -\left\langle \sqrt{\frac{2}{L_x}}\left(-\frac{k\pi}{L_x}\right)\cos\left(\frac{k\pi\ (\cdot)}{L_x}\right), \sqrt{\frac{2}{L_x}}\left(-\frac{\ell\pi}{L_x}\right)\cos\left(\frac{\ell\pi\ (\cdot)}{L_x}\right)\right\rangle$, i.e., $\mathbf{A}_K = diag\left\{-k_\alpha\left(\frac{k\pi}{L_x}\right)^2\right\}$ and $\bar{\boldsymbol{\beta}}^{\mathrm{T}}_{\boldsymbol{S}K} = \left[\left\langle \boldsymbol{\beta}_{\boldsymbol{S}}, \sqrt{\frac{2}{L_x}}\sin\left(\frac{1\pi\ (\cdot)}{L_x}\right)\right\rangle \cdots \left\langle \boldsymbol{\beta}_{\boldsymbol{S}}, \sqrt{\frac{2}{L_x}}\sin\left(\frac{K\pi\ (\cdot)}{L_x}\right)\right\rangle\right]$.

The optimal solution of *Prob. I* is given by, $\forall \tau \in [0, t_F]$

$$\boldsymbol{u}(\tau) = \bar{\boldsymbol{\beta}}^{\mathrm{T}}_{\boldsymbol{S}K} \exp\{\mathbf{A}_K^{\mathrm{T}}(t_F - \tau)\}\bar{\boldsymbol{\alpha}}_K,$$

where

$\bar{\boldsymbol{\alpha}}_K = (\mathbf{I} + \rho_F \mathbf{G}_K)^{-1} \rho_F \bar{\boldsymbol{\theta}}_{ro}^K$, $(\bar{\boldsymbol{\theta}}_{ro}^K)^{\mathrm{T}} = \left[\left\langle \theta_{ro}, \sqrt{\frac{2}{L_x}} \sin\left(\frac{1\pi\ (\cdot)}{L_x}\right) \right\rangle \cdots \left\langle \theta_{ro}, \sqrt{\frac{2}{L_x}} \sin\left(\frac{K\pi\ (\cdot)}{L_x}\right) \right\rangle \right]$ and $\mathbf{G}_K = \displaystyle\int_0^{t_F} \exp[\mathbf{A}_K t] \bar{\boldsymbol{\beta}}_{\boldsymbol{S}K} \bar{\boldsymbol{\beta}}_{\boldsymbol{S}K}^{\mathrm{T}} \exp[\mathbf{A}_K t]^{\mathrm{T}} dt$, i.e., $\mathbf{G}_K$ is the unique solution of

$$\mathbf{A}_K \mathbf{G}_K + \mathbf{G}_K \mathbf{A}_K^{\mathrm{T}} = \exp[\mathbf{A}_K t_F] \bar{\boldsymbol{\beta}}_{\boldsymbol{S}K} \bar{\boldsymbol{\beta}}_{\boldsymbol{S}K}^{\mathrm{T}} \exp[\mathbf{A}_K t_F]^{\mathrm{T}} - \bar{\boldsymbol{\beta}}_{\boldsymbol{S}K} \bar{\boldsymbol{\beta}}_{\boldsymbol{S}K}^{\mathrm{T}}.$$

The approximation error on the final state for a given control signal $\boldsymbol{u}$ is given by $\mathcal{T}_\theta[\boldsymbol{u}] - \theta_{ro} = \boldsymbol{e}_K[\boldsymbol{u}] + \check{\boldsymbol{e}}_K[\boldsymbol{u}]$ where $\boldsymbol{e}_K[\boldsymbol{u}] \triangleq \mathcal{T}_\theta^K[\boldsymbol{u}] - \theta_{ro}^K$ (error projection on $span\{\phi_1, \ldots, \phi_K\}$) and $\check{\boldsymbol{e}}_K[\boldsymbol{u}] = \{\mathcal{T}_\theta[\boldsymbol{u}] - \mathcal{T}_\theta^K[\boldsymbol{u}]\} - \{\theta_{ro} - \theta_{ro}^K\}$.

To get an upper bound on $\|\mathcal{T}_\theta[\boldsymbol{u}] - \theta_{ro}\|_2$ note that

$$\|\mathcal{T}_\theta[\boldsymbol{u}] - \theta_{ro}\|_2^2 = \|\boldsymbol{e}_K[\boldsymbol{u}]\|_2^2 + \|\check{\boldsymbol{e}}_K[\boldsymbol{u}]\|_2^2, \tag{34}$$

$$\|\boldsymbol{e}_K[\boldsymbol{u}]\|_2^2 = \|\bar{\boldsymbol{c}}_K(t_F; \boldsymbol{u}) - \bar{\boldsymbol{\theta}}_{ro}^K\|_E^2, \tag{35}$$

$$\|\check{\boldsymbol{e}}_K[\boldsymbol{u}]\|_2 \leq \|\mathcal{T}_\theta[\boldsymbol{u}] - \mathcal{T}_\theta^K[\boldsymbol{u}]\|_2 + \|\theta_{ro} - \theta_{ro}^K\|_2. \tag{36}$$

Note also that $\left\|\mathcal{T}_\theta[\boldsymbol{u}] - \mathcal{T}_\theta^K[\boldsymbol{u}]\right\|_2^2 = \left\| \displaystyle\sum_{k=K+1}^{\infty} \boldsymbol{c}_k(t_F; \boldsymbol{u}) \phi_k \right\|_2^2 = \displaystyle\sum_{k=K+1}^{\infty} \boldsymbol{c}_k(t_F; \boldsymbol{u})^2$, and

$\boldsymbol{c}_k(t_F, \boldsymbol{u}) = \displaystyle\int_0^{t_F} \exp\left[-k_\alpha \left(\frac{k\pi}{L_x}\right)^2 (t_F - \tau) \right] \boldsymbol{\beta}_{\boldsymbol{S}k} \boldsymbol{u}(\tau) d\tau$, where $\boldsymbol{\beta}_{\boldsymbol{S}k} \triangleq \langle \boldsymbol{\beta}_{\boldsymbol{S}}, \phi_K \rangle$, so that (in the light of Cauchy-Schwarz inequality)

$$\Rightarrow \quad \boldsymbol{c}_k(t_F; \boldsymbol{u})^2 \leq |\boldsymbol{\beta}_{\boldsymbol{S}k}|^2 \left\| \exp\left[-k_\alpha \left(\frac{k\pi}{L_x}\right)^2 (t_F - \cdot) \right] \right\|_{L_2(0,t_F)}^2 \|\boldsymbol{u}\|_{L_2(0,t_F)}^2$$

$$\Rightarrow \quad \boldsymbol{c}_k(t_F; \boldsymbol{u})^2 \leq |\boldsymbol{\beta}_{\boldsymbol{S}k}|^2 \frac{1}{k_\alpha \left(\frac{k\pi}{L_x}\right)^2} \left\{ 1 - \exp\left[-k_\alpha \left(\frac{k\pi}{L_x}\right)^2 t_F \right] \right\} \|\boldsymbol{u}\|_{L_2(0,t_F)}^2$$

$$\leq |\boldsymbol{\beta}_{\boldsymbol{S}k}|^2 \frac{1}{k_\alpha \left(\frac{k\pi}{L_x}\right)^2} \|\boldsymbol{u}\|_{L_2(0,t_F)}^2.$$

It then follows that

$$\|\mathcal{T}_\theta[\boldsymbol{u}] - \mathcal{T}_\theta^K[\boldsymbol{u}]\|_2^2 \leq \|\boldsymbol{\beta}_{\boldsymbol{S}} - \hat{\boldsymbol{\beta}}_{\boldsymbol{S}K}\|_2^2 \frac{1}{k_\alpha\{(K+1)\frac{\pi}{L_x}\}^2}\|\boldsymbol{u}\|_{L_2(0,t_F)}^2, \tag{37}$$

where $\hat{\boldsymbol{\beta}}_{\boldsymbol{S}k} \triangleq \sum_{k=1}^K \boldsymbol{\beta}_{\boldsymbol{S}k}\phi_k$.

Thus, combining (34) -(37) gives an upper bound on $\|\mathcal{T}_\theta[\boldsymbol{u}] - \theta_{ro}\|_2^2$ which approaches the squared norm of the approximation error in $span\{\phi_1, \ldots, \phi_K\}$ (i.e., $\|\boldsymbol{e}_K[\boldsymbol{u}]\|_2^2$) as $K \to \infty$.

For the optimal solution of $Prob.\ I_K$, the latter is given by
$\|\boldsymbol{e}_K[\boldsymbol{u}_K]\|_2^2 = \|\bar{\boldsymbol{c}}_K(t_F;\boldsymbol{u}_K) - \bar{\boldsymbol{\theta}}_{ro}^K\|_2^2$ and since

$\bar{\boldsymbol{c}}_K(t_F;\boldsymbol{u}_K) = \int_0^{t_F} \mathbf{H}_K(t_F-\tau)\boldsymbol{u}_K(\tau)d\tau = \int_0^{t_F} \mathbf{H}_K(t_F-\tau)\mathbf{H}_K(t_F-\tau)^{\mathrm{T}}\bar{\boldsymbol{\alpha}}_K d\tau$, where
$\mathbf{H}_K(t) = \exp\left[\mathbf{A}_K t\right]\boldsymbol{\beta}_{\boldsymbol{S}K}$, $\quad \bar{\boldsymbol{c}}_K(t_F;\boldsymbol{u}_K) = \mathbf{G}_K\bar{\boldsymbol{\alpha}}_K \quad \Leftrightarrow$
$\bar{\boldsymbol{c}}_K(t_F;\boldsymbol{u}_K) = \mathbf{G}_K(\mathbf{I}+\rho_F\mathbf{G}_K)^{-1}\rho_F\bar{\boldsymbol{\theta}}_{ro}^K = \{\mathbf{I} - (\mathbf{I}+\rho_F\mathbf{G}_K)^{-1}\}\,\bar{\boldsymbol{\theta}}_{ro}^K$ it follows that

$$\|\boldsymbol{e}_K[\boldsymbol{u}_K]\|_2^2 = \|(\mathbf{I}+\rho_F\mathbf{G}_K)^{-1}\bar{\boldsymbol{\theta}_{ro}}^K\|_2^2. \tag{38}$$

To compute approximate solutions to $Prob.\ I_c$, consider the truncated problem
<u>$Prob.$</u> I_{cK} : $\min\limits_{\boldsymbol{u}\in S_{\boldsymbol{u}F}} \check{\mathcal{J}}_K(\boldsymbol{u};\rho_F)$ and the corresponding dual problem,
<u>$Prob.$</u> D_K : $\max\limits_{\boldsymbol{\lambda}_a,\boldsymbol{\lambda}_b} \varphi_D^K(\boldsymbol{\lambda}_a,\boldsymbol{\lambda}_b;\rho_F)$ subject to $\forall t$ a.e. in $(0,t_F)$, $\boldsymbol{\lambda}_a \geq 0$, $\boldsymbol{\lambda}_b \geq 0$,
where $\varphi_D^K(\boldsymbol{\lambda}_a,\boldsymbol{\lambda}_b) = \inf\{Lag_K(\boldsymbol{u};\boldsymbol{\lambda}_a,\boldsymbol{\lambda}_b) : \boldsymbol{u} \in L_2(0,t_F)\}$,
$Lag_K(\boldsymbol{u};\boldsymbol{\lambda}_a,\boldsymbol{\lambda}_b) = \check{\mathcal{J}}_K(\boldsymbol{u};\rho_F) + 2\langle\boldsymbol{\lambda}_a,\boldsymbol{u}_a-\boldsymbol{u}\rangle + 2\langle\boldsymbol{\lambda}_b,\boldsymbol{u}-\boldsymbol{u}_b\rangle$ and $\boldsymbol{u}_b = \mu_{\boldsymbol{u}}$ and $\boldsymbol{u}_a = -\boldsymbol{u}_b$.

The unique solution to the problem $\min\limits_{\boldsymbol{u}\in L_2(0,t_F)} Lag_K(\boldsymbol{u};\boldsymbol{\lambda}_a,\boldsymbol{\lambda}_b)$ is given by
$\boldsymbol{u}_K^c[\boldsymbol{\lambda}] = \hat{\boldsymbol{u}}_K^c + \boldsymbol{\lambda}_{ab}$, where $\hat{\boldsymbol{u}}_K^c[\boldsymbol{\lambda}](\tau) = \mathbf{H}_K^{\mathrm{T}}(t_F-\tau)\left\{\bar{\boldsymbol{\alpha}}_K - (\mathbf{I}+\rho_F\mathbf{G}_K)^{-1}\rho_F\boldsymbol{\xi}_{\boldsymbol{\lambda}}^K\right\}$,
$\boldsymbol{\lambda} = (\boldsymbol{\lambda}_a,\boldsymbol{\lambda}_b)$, $\boldsymbol{\lambda}_{ab} = \boldsymbol{\lambda}_a - \boldsymbol{\lambda}_b$ and $\boldsymbol{\xi}_{\boldsymbol{\lambda}}^K = \int_0^{t_F}\mathbf{H}_K(t_F-\tau)\boldsymbol{\lambda}_{ab}(\tau)d\tau$.

The corresponding value for the dual functional is given by

$$\varphi_D^K(\boldsymbol{\lambda}_a,\boldsymbol{\lambda}_b) = Lag_K(\boldsymbol{u}_K^c[\boldsymbol{\lambda}];\boldsymbol{\lambda}_a,\boldsymbol{\lambda}_b) = \rho_F\|\theta_{ro}\|_2^2 + \rho_F\langle\mathcal{T}_\theta^K[\boldsymbol{u}_K^c], -\theta_{ro}\rangle + \hat{\varphi}_D^K(\boldsymbol{\lambda}_a,\boldsymbol{\lambda}_b),$$

where

$$\hat{\varphi}_D^K(\boldsymbol{\lambda}_a, \boldsymbol{\lambda}_b) = -\langle \boldsymbol{\lambda}_{ab}, \boldsymbol{\lambda}_{ab}\rangle + \rho_F \left\langle (\mathbf{I} + \rho_F \mathbf{G}_K)^{-1} \boldsymbol{\xi}_{\boldsymbol{\lambda}}^K, \boldsymbol{\xi}_{\boldsymbol{\lambda}}^K \right\rangle_E - 2\langle \boldsymbol{\xi}_{\lambda}^K, \bar{\boldsymbol{\alpha}}_K\rangle_E + 2\langle \boldsymbol{\lambda}_a, \boldsymbol{u}_a\rangle - 2\langle \boldsymbol{\lambda}_b, \boldsymbol{u}_b\rangle.$$

Note that for any non-negative $\boldsymbol{\lambda}_a$ and $\boldsymbol{\lambda}_b$, $\varphi_D^K(\boldsymbol{\lambda}_a, \boldsymbol{\lambda}_b)$ is a lower bound for the optimal value of *Prob.* I_{cK}. If $(\boldsymbol{\lambda}_a^o, \boldsymbol{\lambda}_b^o)$ is optimal $\boldsymbol{u}_K^c \in S_{\boldsymbol{u}F}$. Moreover, $\boldsymbol{\lambda}_a^o(\tau) = 0$ and $\boldsymbol{\lambda}_b^o(\tau) = 0$ (hence, $\boldsymbol{\lambda}_{ab}^o(\tau) = 0$) whenever $\boldsymbol{u}_K^c[\boldsymbol{\lambda}^o](\tau) \in (\boldsymbol{u}_a, \boldsymbol{u}_b)$ so that, in this case, $\hat{\boldsymbol{u}}_K^c[\boldsymbol{\lambda}^o](\tau)$ also belongs to $(\boldsymbol{u}_a, \boldsymbol{u}_b)$. When $\boldsymbol{\lambda}_a^o(\tau) \neq 0$ (respectively $\boldsymbol{\lambda}_b^o(\tau) \neq 0$) $\boldsymbol{u}_K^c(\tau) = \boldsymbol{u}_a$ and $\hat{\boldsymbol{u}}_K^c[\boldsymbol{\lambda}^o](\tau) < \boldsymbol{u}_a$ (respectively, $\boldsymbol{u}_K^c[\boldsymbol{\lambda}^o](\tau) = \boldsymbol{u}_b$ and $\hat{\boldsymbol{u}}_K^c[\boldsymbol{\lambda}^o](\tau) > \boldsymbol{u}_a$). This suggests a heuristic way of obtaining a feasible $\boldsymbol{u}_K^R[\boldsymbol{\lambda}]$, namely,
$\boldsymbol{u}_K^R[\boldsymbol{\lambda}](\tau) = \hat{\boldsymbol{u}}_K^c[\boldsymbol{\lambda}](\tau)$ if $\hat{\boldsymbol{u}}_K^c[\boldsymbol{\lambda}] \in (\boldsymbol{u}_a, \boldsymbol{u}_b)$,
$\boldsymbol{u}_K^R[\boldsymbol{\lambda}](\tau) = \boldsymbol{u}_a$ if $\hat{\boldsymbol{u}}_K^R[\boldsymbol{\lambda}](\tau) \leq \boldsymbol{u}_a$ and $\boldsymbol{u}_K^R[\boldsymbol{\lambda}](\tau) = \boldsymbol{u}_b$ if $\hat{\boldsymbol{u}}_K^c[\boldsymbol{\lambda}](\tau) \geq \boldsymbol{u}_b$.

To obtain approximate solutions to *Prob.* D_K, piecewise linear classes of multipliers are considered, i.e., let $N_{\boldsymbol{\lambda}} \in \mathbb{Z}_+$, $\delta_t = t_F/N_{\boldsymbol{\lambda}}$, $\mathcal{I}_k = [(k-1)\delta_t, k\delta_t]$, $\boldsymbol{\gamma} = [\gamma_1 \cdots \gamma_{N_{\boldsymbol{\lambda}}+1}]$ and define
$\forall k = 1, \ldots, N_{\boldsymbol{\lambda}}, \quad \forall t \in \mathcal{I}_k, \quad \boldsymbol{\lambda}(t; \boldsymbol{\gamma}) = \gamma_k + (1/\delta_t)(\gamma_{k+1} - \gamma_k)\Delta t_k$,
where $\Delta t_k = t - (k-1)\delta_t$ (note that γ_k and γ_{k+1} are respectively the values of $\boldsymbol{\lambda}(t, \boldsymbol{\lambda})$ at the lower and upper extreme points of the interval $\mathcal{I}_k$). Such multipliers can then be written as a function of $\boldsymbol{\gamma}$ as follows:

$$\forall t \in \mathcal{I}_k, \quad \boldsymbol{\lambda}(t; \boldsymbol{\gamma}) = \boldsymbol{h}_{kab}^{\mathrm{T}}(t) \boldsymbol{E}_k \boldsymbol{\gamma},$$

where $\boldsymbol{h}_{kab}^{\mathrm{T}}(t) = [h_{ka}(t) \vdots h_{kb}(t)]$, $\boldsymbol{E}_k^{\mathrm{T}} = [e_k(m_\gamma) \vdots e_{k+1}(m_\gamma)]$, $m_\gamma = N_{\boldsymbol{\lambda}} + 1$, $h_{ka} : \mathcal{I}_k \to \mathbb{R}$, $h_{ka}(t) = 1 - h_{kb}(t)$, $h_{kb} : \mathcal{I}_k \to \mathbb{R}$, $h_{kb}(t) = (1/\delta_t)(t - a_k)$, where $a_k = (k-1)\delta_t$.

As a result, $\boldsymbol{\xi}_{\boldsymbol{\lambda}}^K = \boldsymbol{T}_{\boldsymbol{\xi}\boldsymbol{\gamma}}(\boldsymbol{\gamma}_a - \boldsymbol{\gamma}_b)$, where $\boldsymbol{T}_{\boldsymbol{\xi}\boldsymbol{\gamma}} = \left\{ \sum_{k=1}^{N_{\boldsymbol{\lambda}}} \int_{\mathcal{I}_k} \mathbf{H}_K(t_f - \tau) \boldsymbol{h}_{kab}^{\mathrm{T}}(\tau) d\tau \right\} \boldsymbol{E}_k$ and

$$-\hat{\varphi}_D^K(\boldsymbol{\lambda}_a, \boldsymbol{\lambda}_b) = \boldsymbol{\gamma}_{ab}^{\mathrm{T}} \left(\boldsymbol{P}_{\boldsymbol{\gamma}} - \boldsymbol{T}_{\boldsymbol{\xi}\boldsymbol{\gamma}}^{\mathrm{T}} \rho_F (\mathbf{I} + \rho_F \mathbf{G}_K)^{-1} \boldsymbol{T}_{\boldsymbol{\xi}\boldsymbol{\gamma}} \right) \boldsymbol{\gamma}_{ab} + 2\bar{\boldsymbol{\alpha}}_K^{\mathrm{T}} \boldsymbol{T}_{\boldsymbol{\xi}\boldsymbol{\gamma}} \boldsymbol{\gamma}_{ab} - 2\boldsymbol{r}_{\boldsymbol{\gamma}a}^{\mathrm{T}} \boldsymbol{\gamma}_a + 2\boldsymbol{r}_{\boldsymbol{\gamma}b}^{\mathrm{T}} \boldsymbol{\gamma}_b,$$

where
$\boldsymbol{\gamma}_{ab} \triangleq \boldsymbol{\gamma}_a - \boldsymbol{\gamma}_b, \quad \boldsymbol{P}_{\boldsymbol{\gamma}} \triangleq \sum_{k=1}^{N_{\boldsymbol{\lambda}}} \boldsymbol{E}_k^{\mathrm{T}} \int_{\mathcal{I}_k} \boldsymbol{h}_{kab}(t) \boldsymbol{h}_{kab}^{\mathrm{T}}(t) dt \boldsymbol{E}_k, \ \boldsymbol{r}_{\boldsymbol{\gamma}a}^{\mathrm{T}} = \sum_{k=1}^{N_{\boldsymbol{\lambda}}} \left\{ \left[\int_{\mathcal{I}_k} \boldsymbol{u}_a(t) \boldsymbol{h}_{kab}^{\mathrm{T}}(t) dt \right] \boldsymbol{E}_k \right\},$

and $\boldsymbol{r}^{\mathrm{T}}_{\gamma b} = \sum_{k=1}^{N_{\boldsymbol{\lambda}}} \left\{ \left[\int_{\mathcal{I}_k} \boldsymbol{u}_b(t) \boldsymbol{h}^{\mathrm{T}}_{kab}(t) dt \right] \boldsymbol{E}_k \right\}$.

The problem to be numerically solved is then

$$\underline{Prob.\ D^K_{\boldsymbol{\gamma}}:} \quad \max_{\gamma_a, \gamma_b \in \mathbb{R}^{N_{\boldsymbol{\lambda}}+1}} \varphi^K_D(\boldsymbol{\lambda}_a(\boldsymbol{\gamma}_a), \boldsymbol{\lambda}_b(\boldsymbol{\gamma}_b); \rho_F). \tag{39}$$

Prob. I_K and *Prob.* $D^K_{\boldsymbol{\gamma}}$ were numerically solved for two pairs $(\theta_{ro}, \boldsymbol{\beta_S})$ respectively displayed in Figures 1, 2 and Figures 5, 6, with $\rho_F = 2000$, $K = 5$, $L_x = 1$ or 2, and $N_{\boldsymbol{\lambda}} = 30$. For the first pair $(\theta_{ro}, \boldsymbol{\beta_S})$ the unconstrained problem was solved leading to the approximate solution $\boldsymbol{u}_K(\ \cdot; \rho_F)$ which is plotted in Fig. 3 (dashed blue curve, labeled $\boldsymbol{u}_K$). Table 1 gives the $L_2(0, t_F)$ and $L_\infty(0, t_F)$ norms of $\boldsymbol{u}_K(\ \cdot; \rho_F)$ and the $L_2(U)$ norm of the projection of the final-state, approximation error on $span\{\phi_1, \ldots, \phi_K\}$.

$\check{\mathcal{J}}_K(\boldsymbol{u}_K; \rho_F)$	$\Vert\boldsymbol{u}_K\Vert_2$	$\Vert\boldsymbol{u}_K\Vert_\infty$	$\Vert\mathcal{T}^K_\theta[\boldsymbol{u}_K] - \boldsymbol{\theta}^K_{ro}\Vert_2$
160.5171	10.9233	43.5917	0.1435

Table 1: Unconstrained problem for the first pair $(\theta_{ro}, \boldsymbol{\beta_S})$, $\rho_F = 2000$.

The constrained problem *Prob.* I_{cK} was then solved for the same pair $(\theta_{ro}, \boldsymbol{\beta_S})$ with the prescribed upper bound $\mu_{\boldsymbol{u}}$ on $\Vert\boldsymbol{u}\Vert_\infty$ taken to be $\mu_{\boldsymbol{u}} = 30$. Approximate solutions are then obtained for *Prob.* $D^K_{\boldsymbol{\gamma}}$, say $(\boldsymbol{\gamma}^K_a, \boldsymbol{\gamma}^K_b)$. The corresponding multipliers are denoted by $\boldsymbol{\lambda}^K_a$ and $\boldsymbol{\lambda}^K_b$ on the basis of which a feasible solution for *Prob.* I_{cK} is computed, namely, $\check{\boldsymbol{u}}^R_K = \boldsymbol{u}^R_K[\boldsymbol{\lambda}^K]$ where $\boldsymbol{\lambda}^K = (\boldsymbol{\lambda}^K_a, \boldsymbol{\lambda}^K_b)$. Table 2 below exhibits the results to *Prob.* I_{cK} for the first pair $(\theta_{ro}, \boldsymbol{\beta_S})$.

$\check{\mathcal{J}}_K(\check{\boldsymbol{u}}^R_K; \rho_F)$	$\varphi^K_D(\boldsymbol{\lambda}^K)$	$\Vert\check{\boldsymbol{u}}^R_K\Vert_2$	$\Vert\check{\boldsymbol{u}}^R_K\Vert_\infty$	$\Vert\mathcal{T}^K_\theta[\check{\boldsymbol{u}}^R_K] - \boldsymbol{\theta}^R_{ro}\Vert_2$
168.2210	167.0747	10.5405	30	0.1690

Table 2: Constrained problem for the first pair $(\theta_{ro}, \boldsymbol{\beta_S})$, $\rho_F = 2000$.

Recall that $\varphi^K_D(\boldsymbol{\lambda}^K)$ is a lower bound on the optimal value of *Prob.* I_{cK} and that $\check{\boldsymbol{u}}^R_K$ is a feasible solution for it. Thus, as shown in Table 2, $\check{\mathcal{J}}_K(\check{\boldsymbol{u}}^R_K)$ does not exceed the optimal value of *Prob.* I_{cK} (say $\mathcal{J}^o_{cK}$) by more than 1.15 (or by 0.7% of $\mathcal{J}^o_{cK}$) – thus, $\check{\boldsymbol{u}}^R_K$ can be taken to be an "approximately - optimal" solution to *Prob.* I_{cK}.

Figure 3 displays the plots of $\check{\boldsymbol{u}}_K^R$ and $\boldsymbol{u}_K$. Figure 4 exhibits the plots of θ_{ro}^K (the projection of θ_{ro} on $span\{\phi_1,\ldots,\phi_K\}$, in green), $\hat{\theta}_K \triangleq \mathcal{T}_\theta^K[\boldsymbol{u}_K]$ (dashed blue) and
$\hat{\theta}_K^R \triangleq \mathcal{T}_\theta^K[\check{\boldsymbol{u}}_K^R]$(in red).

To illustrate the role of ρ_F in getting better approximation of the desired final state, numerical results were obtained for the same pair $(\theta_{ro}, \boldsymbol{\beta_S})$ with $\rho_F = 4000$. The results are presented in Tables 3, 4 and Figures 5 and 6.

$\check{\mathcal{J}}_K(\boldsymbol{u}_K;\rho_F)$	$\Vert\boldsymbol{u}_K\Vert_2$	$\Vert\boldsymbol{u}_K\Vert_\infty$	$\Vert\mathcal{T}_\theta^K[\boldsymbol{u}_K]-\boldsymbol{\theta}_{ro}^K\Vert_2$
187.54639	12.3752	46.5118	0.0926

Table 3: Unconstrained problem for the first pair $(\theta_{ro}, \boldsymbol{\beta_S})$, $\rho_F = 4000$.

$\check{\mathcal{J}}_K(\check{\boldsymbol{u}}_K^R;\rho_F)$	$\varphi_D^K(\boldsymbol{\lambda}^K)$	$\Vert\check{\boldsymbol{u}}_K^R\Vert_2$	$\Vert\check{\boldsymbol{u}}_K^R\Vert_\infty$	$\Vert\mathcal{T}_\theta^K[\check{\boldsymbol{u}}_K^R]-\boldsymbol{\theta}_{ro}^R\Vert_2$
211.2104	212.2948	11.9634	30	0.1305

Table 4: Constrained problem for the first pair $(\theta_{ro}, \boldsymbol{\beta_S})$, $\rho_F = 4000$.

Comparing Tables 1 and 3, it can be noted that the increase in ρ_F from 2000 to 4000 brought about a decrease in the $L_2(0,t_F)$–norm of the approximation error on $span\{\phi_1,\ldots,\phi_5\}$ (from 0.1435 to 0.0926) at the expense of increases in both the $L_2(0,t_F)$ and $L_\infty(0,t_F)$ norms of $\boldsymbol{u}_K$ (respectively, from 10.9233 to 12.3752 and from 43.5917 and 46.5118).

Similarly, in the case of constrained problems (Tables 2 and 4) it can be noted that the increase in ρ_F decreased the $L_2(0,t_F)$–norm of the "projected" approximation error obtained under "peak-value" constraint ($\Vert\boldsymbol{u}\Vert_\infty \leq 30$) from 0.1690 (Table 2) to 0.1305 (Table 4). Note also that $\boldsymbol{u}_K^R$ is "approximately optimal" as $|\varphi_D^K(\boldsymbol{\lambda}^K) - \check{\mathcal{J}}_K(\boldsymbol{u}_K^R;4000)|/\varphi_D^K(\boldsymbol{\lambda}^K) \approx 1.09/212.2948 \leq 0.5\times10^{-2}$.

The plots of $\boldsymbol{u}_K$ and $\boldsymbol{u}_K^R$ and those of the corresponding approximations $\hat{\theta}_K$ and $\hat{\theta}_K^R$ of the desired final state are respectively displayed in Figures 5 and 6.

Numerical results were also obtained for the pair $(\theta_{ro}, \boldsymbol{\beta_S})$ shown in Figures 7 and 8. First, an approximate solution $\boldsymbol{u}_K$ was obtained for *Prob.* I_K – see Table 5 for the values of its $L_2(0,t_F)$ and $L_\infty(0,t_F)$ norms and the corresponding values of the cost-functional and the $L_2(0,1)$ norm of the final-state error (projected on $span\{\phi_1,\ldots,\phi_K\}$).

$\check{\mathcal{J}}_K(\boldsymbol{u}_K;\rho_F)$	$\|\boldsymbol{u}_K\|_2$	$\|\boldsymbol{u}_K\|_\infty$	$\|\mathcal{T}_\theta^K[\boldsymbol{u}_K]-\boldsymbol{\theta}_{ro}^K\|_2$
283.5120	13.5254	23.5491	0.2242

Table 5: Unconstrained problem for the second pair $(\theta_{ro},\boldsymbol{\beta_S})$, $\rho_F = 2000$.

A numerical solution $\check{\boldsymbol{u}}_K^R$ was then obtained for *Prob.* I_{cK} with the prescribed upper limit $\mu_{\boldsymbol{u}}$ on the $L_\infty(0,t_F)$–norm of $\boldsymbol{u}$ being set at $\mu_{\boldsymbol{u}} = 18$. This was done along the same lines described above in connection with the first pair $(\theta_{ro},\boldsymbol{\beta_S})$. Table 6 exhibits the corresponding assessment data for $\check{\boldsymbol{u}}_K^R$.

$\check{\mathcal{J}}_K(\check{\boldsymbol{u}}_K^R;\rho_F)$	$\varphi_D^K(\boldsymbol{\lambda}^K)$	$\|\check{\boldsymbol{u}}_K^R\|_2$	$\|\check{\boldsymbol{u}}_K^R\|_\infty$	$\|\mathcal{T}_\theta^K[\check{\boldsymbol{u}}_K^R]-\boldsymbol{\theta}_{ro}^R\|_2$
300.2274	286.3859	12.6191	18.0000	0.2655

Table 6: Constrained problem for the second pair $(\theta_{ro},\boldsymbol{\beta_S})$, $\rho_F = 2000$.

Note that $\check{\mathcal{J}}_K(\check{\boldsymbol{u}}_K^R;\rho_F)$ may only exceed the optimal value $\mathcal{J}_{cK}^o$ of *Prob.* I_{cK} by less than 5% (of $\mathcal{J}_{cK}^o$). Figures 9 and 10 respectively display the plots of $\boldsymbol{u}_K$ (dashed blue) and $\check{\boldsymbol{u}}_K^R$ and those of θ_{ro}^K (the projection of θ_{ro} on $span\{\phi_1,\ldots,\phi_K\}$), $\check{\theta}_K \triangleq \mathcal{T}_\theta^K[\boldsymbol{u}_K]$ (dashed blue) and $\check{\theta}_K^R \triangleq \mathcal{T}_\theta^K[\boldsymbol{u}_K^R]$.

Results were also obtained for the second pair $(\theta_{ro},\boldsymbol{\beta_S})$ with $\rho_F = 4000$, as presented in Tables 7 and 8 and Figures 11 and 12

$\check{\mathcal{J}}_K(\boldsymbol{u}_K;\rho_F)$	$\|\boldsymbol{u}_K\|_2$	$\|\boldsymbol{u}_K\|_\infty$	$\|\mathcal{T}_\theta^K[\boldsymbol{u}_K]-\boldsymbol{\theta}_{ro}^K\|_2$
362.0183	15.3659	26.4600	0.1774

Table 7: Unconstrained problem for the second pair $(\theta_{ro},\boldsymbol{\beta_S})$, $\rho_F = 4000$.

$\check{\mathcal{J}}_K(\check{\boldsymbol{u}}_K^R;\rho_F)$	$\varphi_D^K(\boldsymbol{\lambda}^K)$	$\|\check{\boldsymbol{u}}_K^R\|_2$	$\|\check{\boldsymbol{u}}_K^R\|_\infty$	$\|\mathcal{T}_\theta^K[\check{\boldsymbol{u}}_K^R]-\check{\boldsymbol{\theta}}_{ro}^R\|_2$
387.3645	387.2568	14.7342	18	0.2063

Table 8: Constrained problem for the second pair $(\theta_{ro},\boldsymbol{\beta_S})$, $\rho_F = 4000$.

Again, it can be noted that increasing ρ_F brings about a better approximation to the desired final state. Note also that $|\varphi_D^K(\boldsymbol{\lambda}^K) - \check{\mathcal{J}}_K(\boldsymbol{u}_K^R;4000)|/\varphi_D^K(\boldsymbol{\lambda}^K) \approx 0.11/387.2568 \leq 0.03\times 10^{-2}$ and hence $\check{\boldsymbol{u}}_K^R$ can be regarded as "approximately optimal" for the constrained problem.

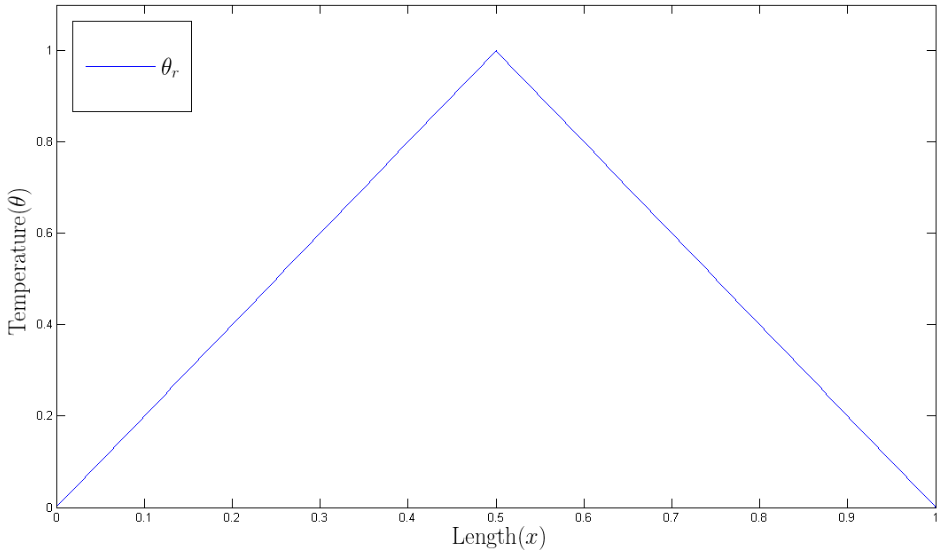

Figure 1: Example 1. θ_{ro}: target final state.

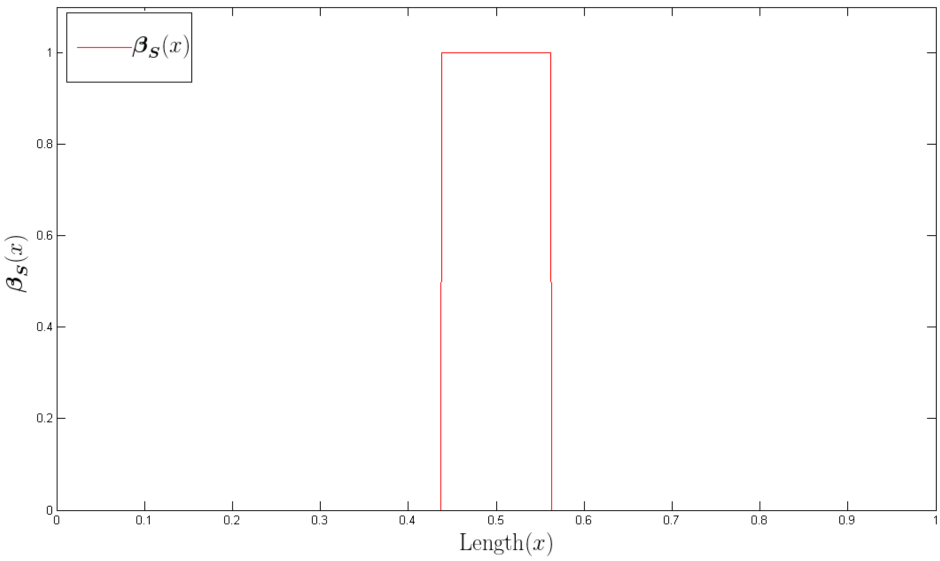

Figure 2: Example 1. $\boldsymbol{\beta_S}$: control-to-state actuator.

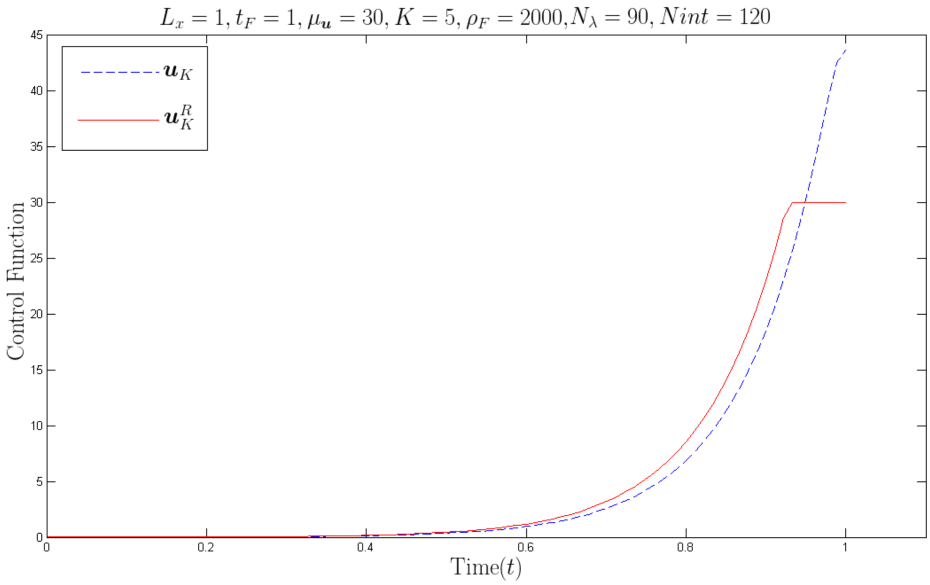

Figure 3: Example 1. Control signals $\boldsymbol{u}_K$ (blue dashed), $\boldsymbol{u}_K^R$ (red solid) for $\rho_F = 2000$.

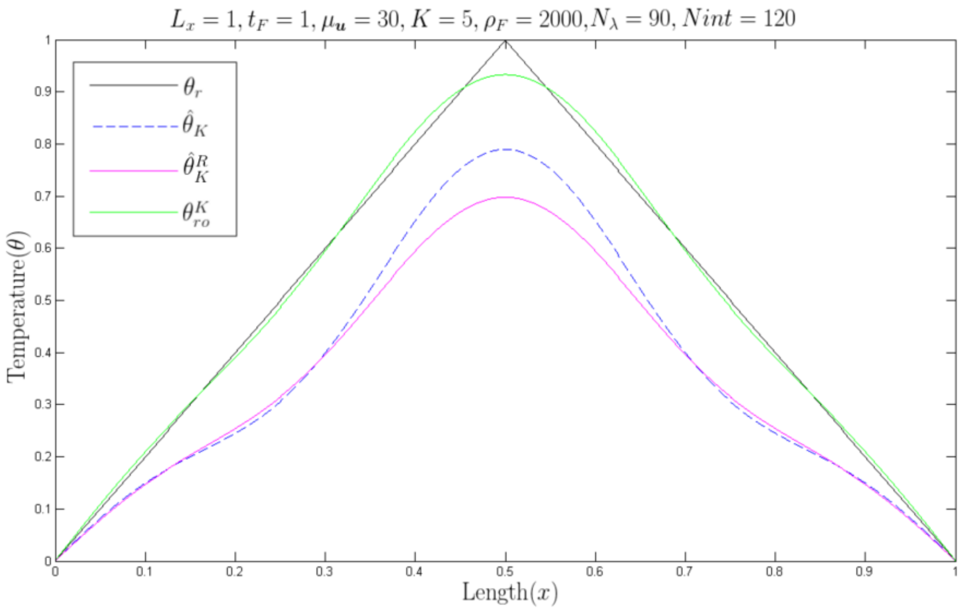

Figure 4: Example 1. Approximations to target final state for $\rho_F = 2000$.

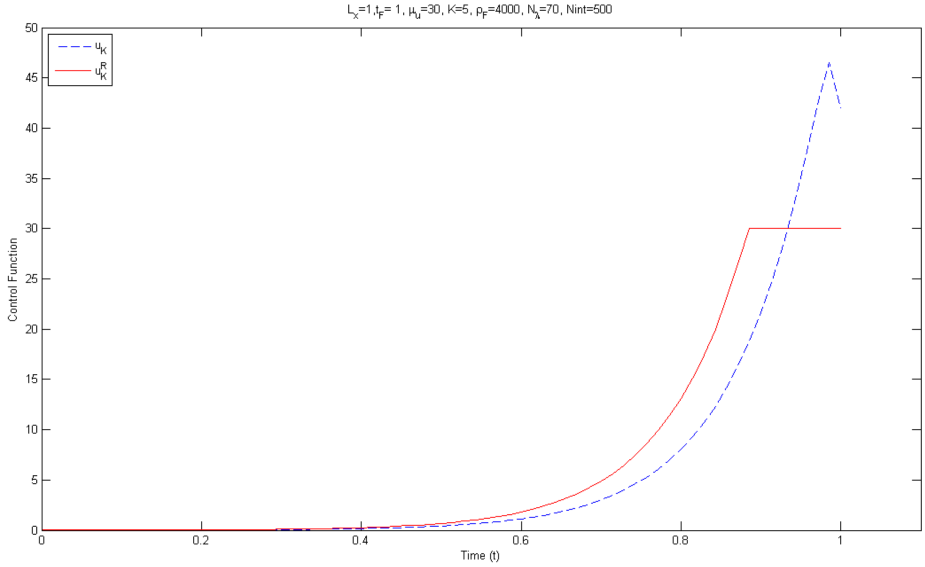

Figure 5: Example 1. Control signals $\boldsymbol{u}_K$ (blue dashed), $\boldsymbol{u}_K^R$ (red solid) for $\rho_F = 4000$.

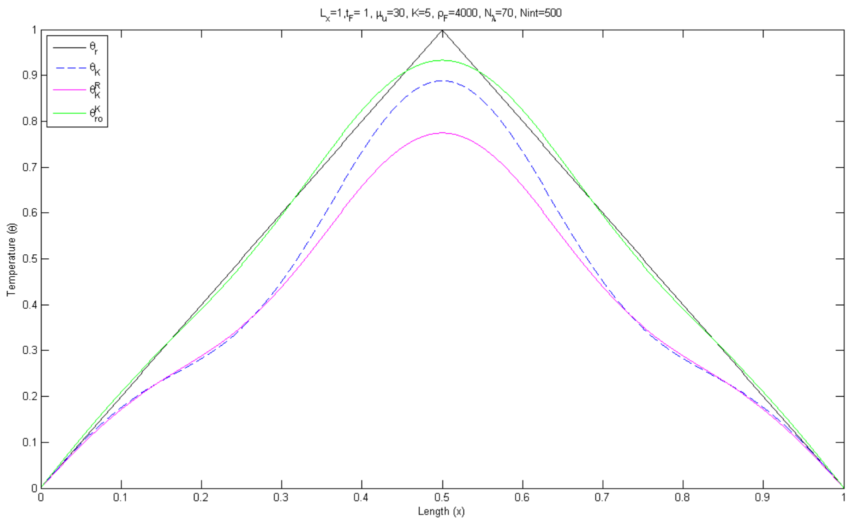

Figure 6: Example 1. Approximations to target final state for $\rho_F = 4000$.

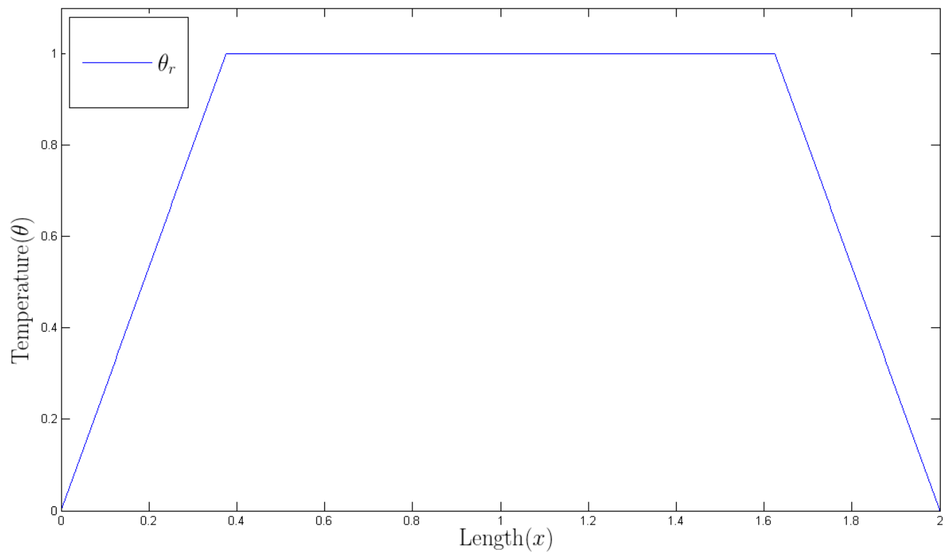

Figure 7: Example 2. θ_{ro}: target final state.

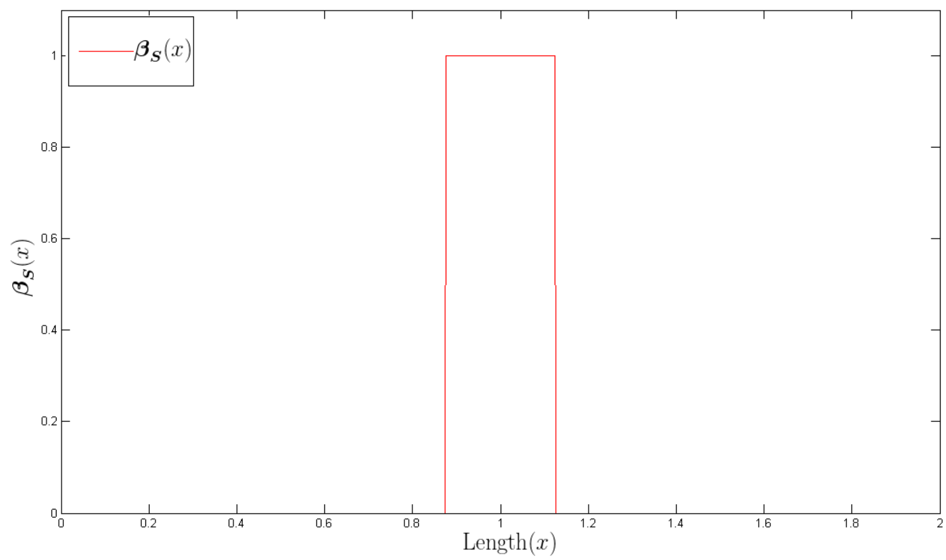

Figure 8: Example 2. $\boldsymbol{\beta_S}$: control-to-state actuator.

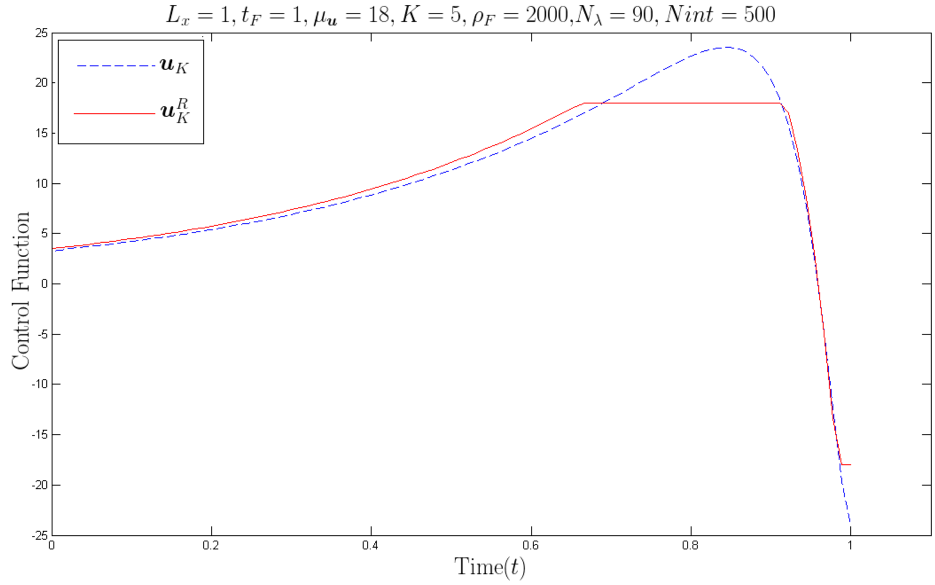

Figure 9: Example 2. Control signals $\boldsymbol{u}_K$ (blue dashed), $\boldsymbol{u}_K^R$ (red solid) for $\rho_F = 2000$.

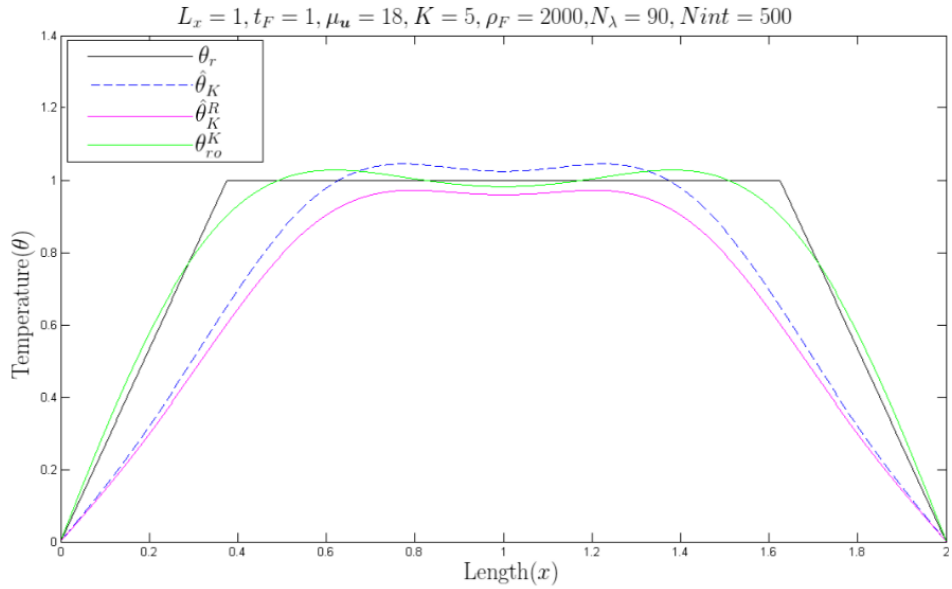

Figure 10: Example 2. Approximations to target final state for $\rho_F = 2000$.

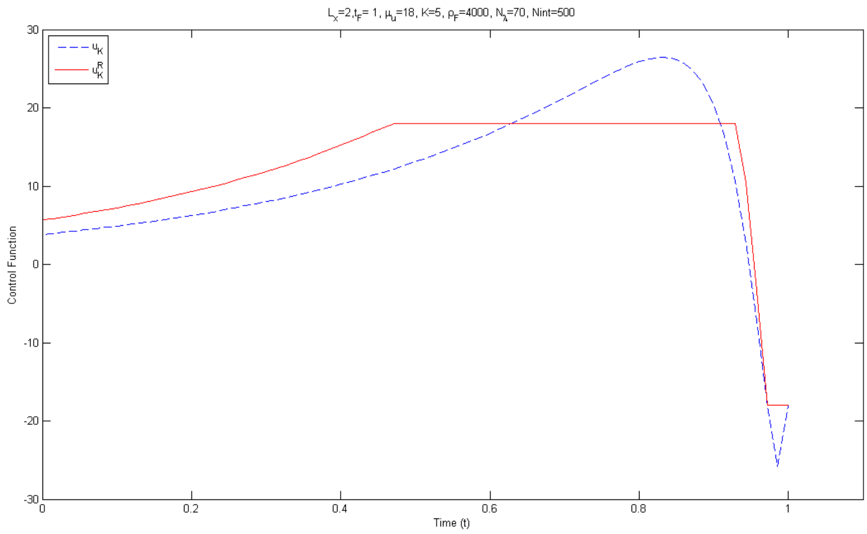

Figure 11: Example 2. Control signals $\boldsymbol{u}_K$ (blue dashed), $\boldsymbol{u}_K^R$ (red solid) for $\rho_F = 4000$.

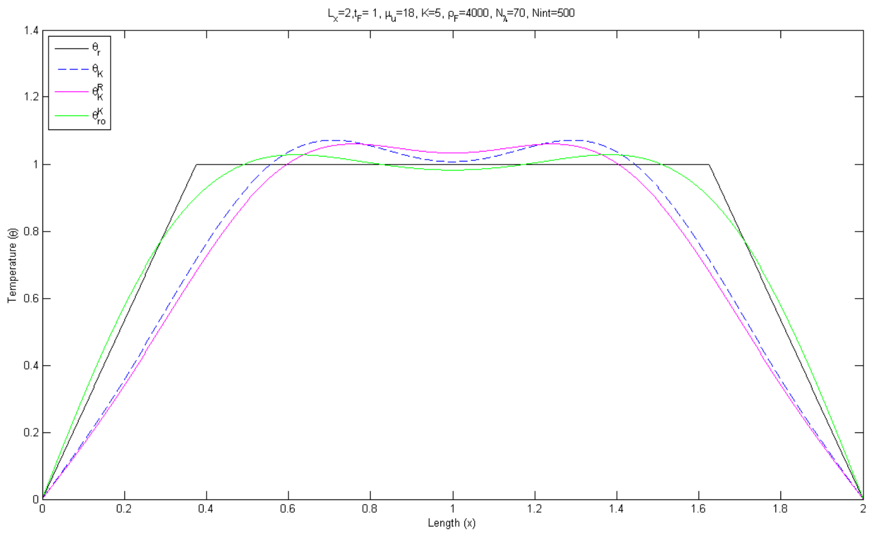

Figure 12: Example 2. Approximations to target final state for $\rho_F = 4000$.

Finally, the effect of the location of the "actuator" $\boldsymbol{\beta_S}$ on the final-state error $\mathcal{T}_\theta^K[\boldsymbol{u}_K^c] - \theta_{ro}$ is illustrated by taking $\boldsymbol{\beta_S}$ to be centered on $\ell_x \in (0,2)$, i.e., by letting $\boldsymbol{\beta_S}$ to be given by $\boldsymbol{\beta_S}(x) = 1, \quad \forall x \in (\ell_x - \delta_{\boldsymbol{\beta}}, \ell_x + \delta_{\boldsymbol{\beta}}), \quad \boldsymbol{\beta_S}(x) = 0$ otherwise, and computing the resulting $\mathcal{T}_\theta^K[\boldsymbol{u}_K^c]$ for several values of ℓ_x (with $\delta_{\boldsymbol{\beta}} = 0.1$), which are displayed in Figures 13–15, respectively for $\ell_x = 3/10$, $\ell_x = 1$ and $\ell_x = 2 - 3/10$.

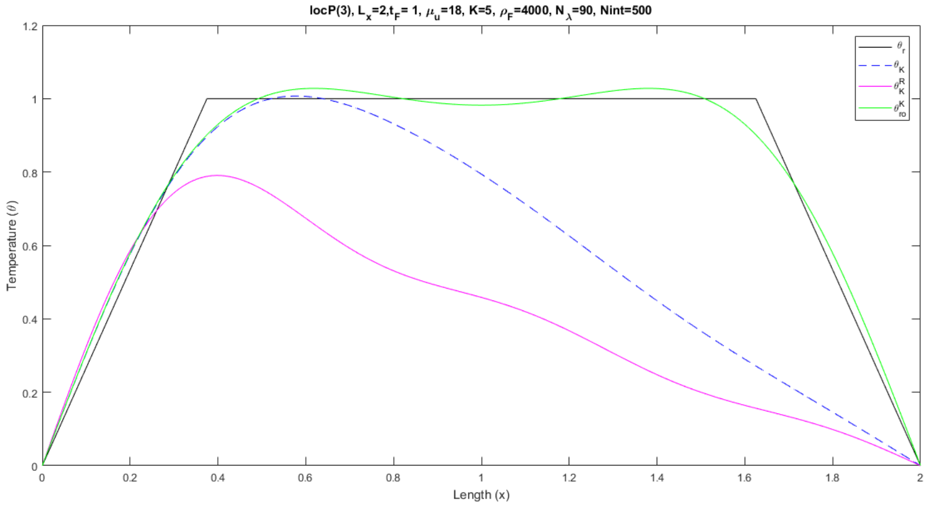

Figure 13: Example 2. Approximations to target final state for $\rho_F = 4000$, $\ell_x = 3/10$.

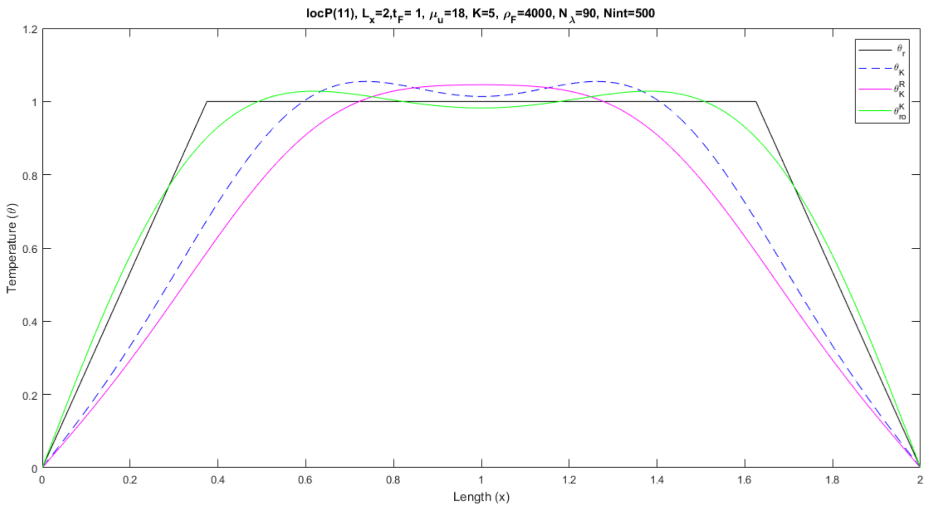

Figure 14: Example 2. Approximations to target final state for $\rho_F = 4000$, $\ell_x = 1$.

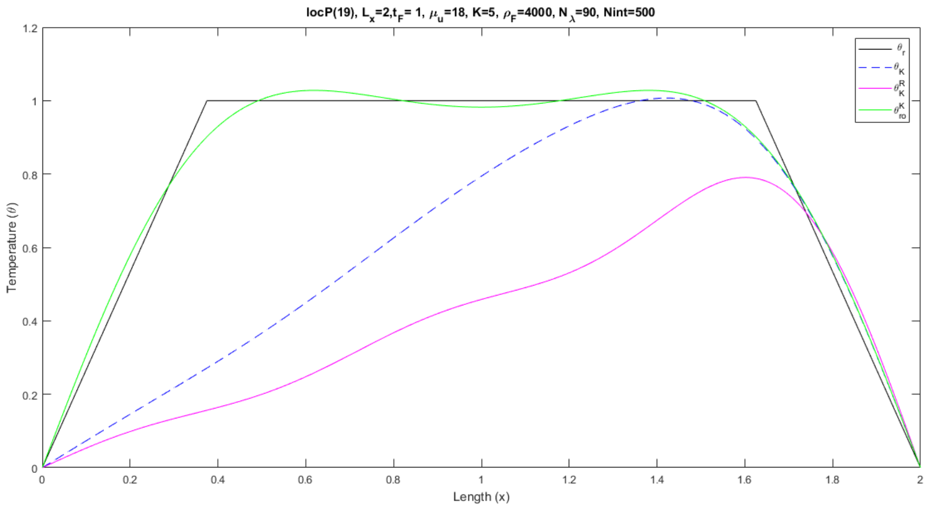

Figure 15: Example 2. Approximations to target final state for $\rho_F = 4000$, $\ell_x = 2 - 3/10$.

7.2 A Two–dimensional Example

An example is now presented of an initial/boundary-value problem defined by the heat equation on a rectangle in $\mathbb{R}^2$. More specifically, let $U = (0, L_x) \times (0, L_y)$, where $L_x, L_y \in \mathbb{R}_+$ and consider the following equation:

$$\forall (x, y) \in U, \quad \frac{\partial \theta}{\partial t}(x, y, t) = k_\alpha \left\{ \frac{\partial^2 \theta}{\partial x^2} + \frac{\partial^2 \theta}{\partial y^2} \right\} (x, y, t) + \boldsymbol{\beta_S}(x, y) \boldsymbol{u}(t)$$

with zero initial conditions, i.e., $\forall (x, y) \in U, \quad \theta(x, y, 0) = 0$ and homogeneous Dirichlet boundary conditions, i.e.,

$$\forall t \in [0, t_F], \quad \forall (x, y) \in \partial U, \quad \theta(x, y, t) = 0,$$

where $\boldsymbol{u} : [0, t_F] \to \mathbb{R}$ and $\boldsymbol{\beta_S} : U \to \mathbb{R}$.

The corresponding weak, "K-th order", Galerkin version is given by $\forall k = 1, \ldots, K$,

$$\left\langle \frac{\partial \theta}{\partial t}(\cdot, \cdot, t), \phi_k \right\rangle = -k_\alpha \left\{ \left\langle \frac{\partial \theta}{\partial x}(\cdot, \cdot, t), \frac{\partial \phi_k}{\partial x} \right\rangle + \left\langle \frac{\partial \theta}{\partial y}(\cdot, \cdot, t), \frac{\partial \phi_k}{\partial y} \right\rangle \right\} + \boldsymbol{\beta_{S}}_k \boldsymbol{u}(t),$$

where $i = 1, \ldots, K_x, \;\; j = 1, \ldots, K_y, \; k(i, j) = (i - 1)K_y + j, \;\; K = K_x K_y,$

$\phi_{k(i,j)}(x,y) = \phi_i^x(x)\phi_j^y(y),\;\; \phi_i^x(x) = \sqrt{\frac{2}{L_x}}\sin\left[\frac{i\pi x}{L_x}\right],\;\; \phi_j^y(y) = \sqrt{\frac{2}{L_y}}\sin\left[\frac{j\pi y}{L_y}\right].$

As in the previous example, control signals $\boldsymbol{u}_K$ and $\boldsymbol{u}_K^c$ are sought by means of the problems

$$\underline{\mathit{Prob.}\ I_K :}\ \min_{\boldsymbol{u}\in L_2(0,t_F)} \check{\mathcal{J}}_K(\boldsymbol{u};\rho_F) \quad \text{and} \quad \underline{\mathit{Prob.}\ I_{cK} :}\ \min_{\boldsymbol{u}\in S_{\boldsymbol{u}F}} \check{\mathcal{J}}_K(\boldsymbol{u};\rho_F),$$

where $\check{\mathcal{J}}_K(\boldsymbol{u};\rho_F) = \|\boldsymbol{u}\|^2_{L_2(0,t_F)} + \rho_F\|\mathcal{T}_\theta^K[\boldsymbol{u}] - \theta_{ro}\|_2^2$, $\mathcal{T}_\theta^K[\boldsymbol{u}] = \sum_{k=1}^K c_k(t_F;\boldsymbol{u})\phi_k$, θ_{ro} is the final state to be "approximately reached" and, as before, $\bar{\boldsymbol{c}}_K(t;\boldsymbol{u}) = [c_1(t;\boldsymbol{u}) \;\cdots\; c_K(t;\boldsymbol{u})]^{\mathrm{T}}$ is given by $\bar{\boldsymbol{c}}_K(t;\boldsymbol{u}) = \displaystyle\int_0^t \mathbf{F}_K(\tau)^{\mathrm{T}}\boldsymbol{u}(\tau)d\tau$ with $\mathbf{F}_K$ as in (19). In this case,

$$\mathbf{A}_K = diag\{a_k : k = k(1,1),\ldots,k(1,K_y),k(2,1),\ldots,k(2,K_y),\ldots,k(K_x,1),\ldots,k(K_x,K_y)\},$$

where $a_{k(i,j)} = -k_\alpha\left\{\left[\frac{i\pi}{L_x}\right]^2 + \left[\frac{j\pi}{L_y}\right]^2\right\}$, and $\mathbf{M}_{\boldsymbol{\beta}}^K = [\langle\boldsymbol{\beta_S},\phi_1\rangle \;\cdots\; \langle\boldsymbol{\beta_S},\phi_k\rangle]^{\mathrm{T}}$,
$S_{\boldsymbol{u}F} = \{\boldsymbol{u}\in L_2(0,t_F) :\ \text{a.e.},\ |\boldsymbol{u}(t)| \le \mu_{\boldsymbol{u}}\}$.

Note that $\check{\mathcal{J}}_K(\boldsymbol{u};\rho_F) = \|\boldsymbol{u}\|^2_{L_2(0,t_F)} + \rho_F\|\mathcal{T}_\theta^K[\boldsymbol{u}] - \theta_{ro}^K\|^2_{L_2(U)} + \|\theta_{ro} - \theta_{ro}^K\|^2_{L_2(U)}$, where θ_{ro}^K is the orthogonal projection of θ_{ro} on the span of $\{\phi_1,\ldots,\phi_K\}$.

The numerical results shown in Tables 9 – 12 were obtained with the following problem data: $k_\alpha = 1$, $L_x = L_y = 1$, $t_F = 1$, $\rho_F = 8000$ and 20000, $\mu_{\boldsymbol{u}} = 100$, $K_x = K_y = 5$, $\theta_{ro}(x,y) = 0\ \forall(x,y)\in\partial U$, $\theta_{ro}(x,y) = 2\ \forall(x,y)\in[L_x/10, 9L_x/10]\times[L_y/10, 9L_y/10]$, the graph of θ_{ro} is the frustum of a rectangular pyramid with $[0,L_x]\times[0,L_y]$ as basis, $\|\theta_{ro}^K\|_2 = 1.7289$ and $\boldsymbol{\beta_S}$ is given by $\begin{cases}\boldsymbol{\beta_S} = 1 & \text{for } (x,y)\in[L_x/4, 3L_x/4]\times[L_y/4, 3L_y/4]\\ \boldsymbol{\beta_S} = 0 & \text{otherwise}\end{cases}$.

$\check{\mathcal{J}}_K(\boldsymbol{u}_K;\rho_F)$	$\|\boldsymbol{u}_K\|_2$	$\|\boldsymbol{u}_K\|_\infty$	$\|\mathcal{T}_\theta^K[\boldsymbol{u}_K] - \boldsymbol{\theta}_{ro}^K\|_2^2$
4978.00	45.6636	192.5735	0.6037

Table 9: Unconstrained problem with $\rho_F = 8000$.

$\check{\mathcal{J}}_K(\boldsymbol{u}_K^c;\rho_F)$	$\varphi_D^K(\boldsymbol{\lambda}^K)$	$\|\boldsymbol{u}_K^c\|_2$	$\|\boldsymbol{u}_K^c\|_\infty$	$\|\mathcal{T}_\theta^K[\boldsymbol{u}_K^c] - \boldsymbol{\theta}_{ro}^K\|_2^2$
5668.10	5485.00	33.0038	100	0.7565

Table 10: Constrained problem with $\rho_F = 8000$.

$\check{\mathcal{J}}_K(\boldsymbol{u}_K;\rho_F)$	$\lVert\boldsymbol{u}_K\rVert_2$	$\lVert\boldsymbol{u}_K^c\rVert_\infty$	$\lVert\mathcal{T}_\theta^K[\boldsymbol{u}_K]-\boldsymbol{\theta}_{ro}^K\rVert_2^2$
8127.40	64.4017	265.37	0.4485

Table 11: Unconstrained problem with $\rho_F = 20000$.

$\check{\mathcal{J}}_K(\boldsymbol{u}_K^c;\rho_F)$	$\varphi_D^K(\boldsymbol{\lambda}^K)$	$\lVert\boldsymbol{u}_K^c\rVert_2$	$\lVert\boldsymbol{u}_K^c\rVert_\infty$	$\lVert\mathcal{T}_\theta^K[\boldsymbol{u}_K^c]-\boldsymbol{\theta}_{ro}^K\rVert_2^2$
12281.00	11195.00	37.8125	100	0.7366

Table 12: Constrained problem with $\rho_F = 20000$.

Similarly to the results in the case of a one-dimensional spatial domain, Tables 9–11 illustrate the effect of increasing ρ_F on the decrease of the approximation errors $\lVert\mathcal{T}_\theta^K[\boldsymbol{u}_K] - \theta_{ro}^K\rVert_2$ (from 0.6037 in Table 9 to 0.4484 in Table 11) and $\lVert\mathcal{T}_\theta^K[\boldsymbol{u}_K^c] - \theta_{ro}^K\rVert_2$ (from 0.7565 in Table 10 to 0.7366 in Table 12). Note that in the latter case, increasing ρ_F from 8000 to 20000 had a small effect on the approximation error - this is due to the fact that the maximum magnitude of $\boldsymbol{u}$ was kept at the same value ($\mu_{\boldsymbol{u}} = 100$).

Again, as observed in the 1D-case, the "relatively small" difference between $\varphi_D^K(\boldsymbol{\lambda}^K)$ and $\check{\mathcal{J}}_K(\boldsymbol{u}_K^c;\rho_F)$ (3.2% for $\rho_F = 8000$ and 8.8% for $\rho_F = 20000$) indicates that $\boldsymbol{u}_K^c$ is "nearly optimal" for the constrained problem - recall that $\varphi_D^K(\boldsymbol{\lambda}^K)$ is a lower bound on $\check{\boldsymbol{u}}_K(\boldsymbol{u};\rho_F)$ for any $\boldsymbol{u} \in S_{\boldsymbol{u}F}$.

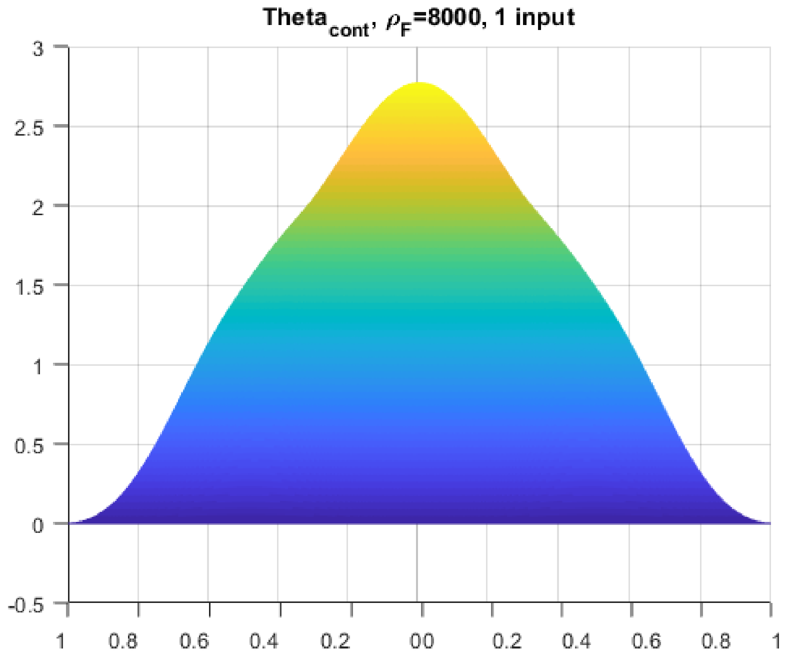

Figure 16: Transversal section of $\mathcal{T}_\theta^K[\boldsymbol{u}_K^c]$ at $\ell_x = \ell_y$ for $\rho_F = 8000$.

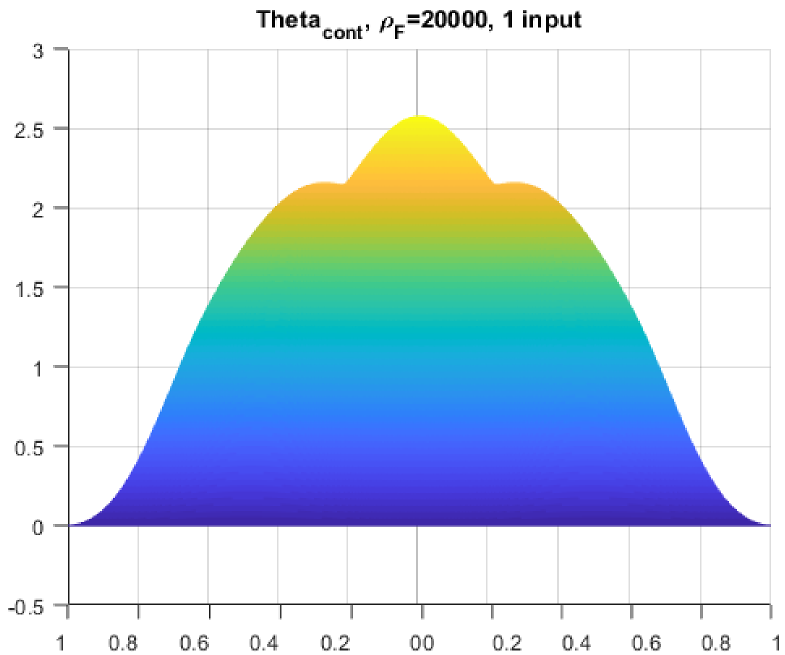

Figure 17: Transversal section of $\mathcal{T}_\theta^K[\boldsymbol{u}_K^c]$ at $\ell_x = \ell_y = 1/2$ for $\rho_F = 20000$.

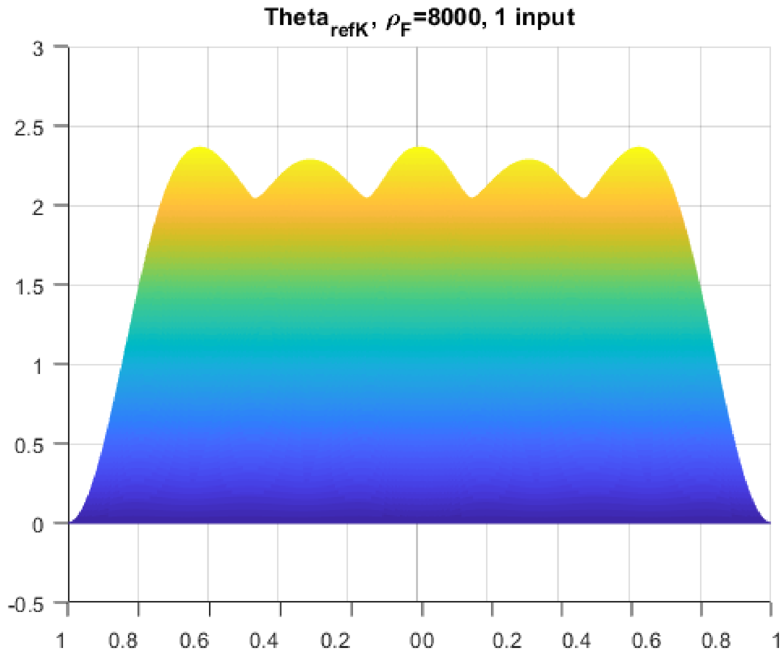

Figure 18: Transversal section of θ_{ro}^{K}.

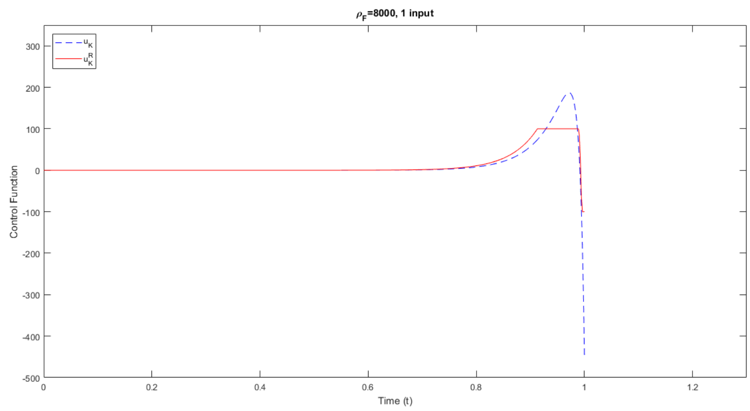

Figure 19: Graphs of $\boldsymbol{u}_K$ and $\boldsymbol{u}_K^c$ for $\rho_F = 8000$.

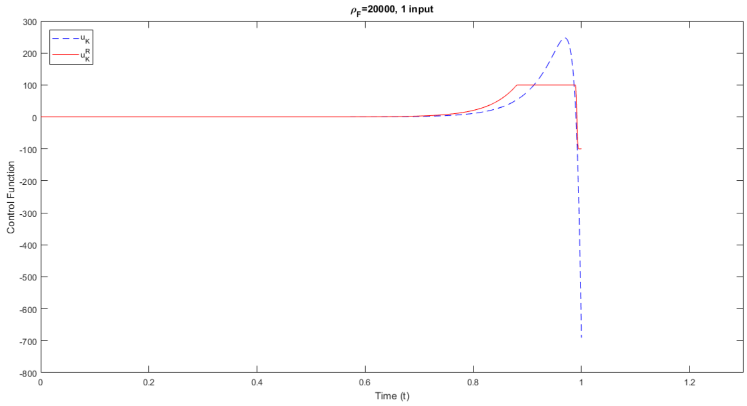

Figure 20: Graphs of $\boldsymbol{u}_K$ and $\boldsymbol{u}_K^c$ for $\rho_F = 20000$.

Figures 16 – 20 respectively display $\mathcal{T}_\theta^K[\boldsymbol{u}_K^c]$, θ_{ro}^K, transversal sections of the first two plot and $\boldsymbol{u}_K$ and $\boldsymbol{u}_K^c$ for $\rho_F = 8000$ and 20000.

7.3 Actuator Location

The initial/boundary value example defined by the heat equation on a rectangle $(0, L_x) \times (0, L_y)$ in $\mathbb{R}^2$ which was introduced above is now slightly modified to involve two scalar control signals ($\boldsymbol{u}(t) \in \mathbb{R}^2$) and numerical results obtained searching the set of their possible "locations" will be presented.

More specifically, let the "source" term in the heat equation be given by

$$\boldsymbol{\beta}_{\boldsymbol{S}}(x, y; \mathcal{X})\boldsymbol{u}(t) = \sum_{i=1}^{2} \boldsymbol{\beta}_{\boldsymbol{S}i}(x, y; \mathcal{X}_i)u_i(t),$$

where $\underline{\mathcal{X}} = (\mathcal{X}_1, \mathcal{X}_2)$, $\mathcal{X}_i = (\mathcal{X}_i^x, \mathcal{X}_i^y) \in \mathbb{R}^2$ and $\boldsymbol{\beta}_{\boldsymbol{S}i}(\cdot)$ is defined by

$$\begin{aligned} \boldsymbol{\beta}_{\boldsymbol{S}i}(x, y; \mathcal{X}_i) &= 1 \qquad && \forall\ (x, y) \in [\mathcal{X}_i^x - \delta_{\boldsymbol{\beta}}, \mathcal{X}_i^x + \delta_{\boldsymbol{\beta}}] \times [\mathcal{X}_i^y - \delta_{\boldsymbol{\beta}}, \mathcal{X}_i^y + \delta_{\boldsymbol{\beta}}] \\ \boldsymbol{\beta}_{\boldsymbol{S}i}(x, y; \mathcal{X}_i) &= 0 && \text{otherwise.} \end{aligned}$$

A location $\underline{\mathcal{X}}$ will be assessed by the approximation error relative to the desired final state θ_{ro}^K achieved by the optimal (unconstrained) control over $(0, t_F)$, i.e., by

$$\nu(\mathcal{X}_1, \mathcal{X}_2) = \|(\mathbf{I} + \rho_F \mathbf{G}_K(\mathbf{M}_{\boldsymbol{\beta}}^K(\mathcal{X}_1, \mathcal{X}_2)))\overline{\theta}_{ro}^K\|_2,$$

where $\mathbf{M}_{\boldsymbol{\beta}}^K$ and $\mathbf{G}_K(\mathbf{M})$ are as in Section 6.

Two searches were carried with the following data: $L_x = L_y = 1, \quad t_F = 1, \quad \rho_F = 8000,$ $\forall\ (x, y) \in (0, L_x) \times (0, L_y), \quad \theta_r(x, y) = 2, \quad K = 5, \quad \delta_{\boldsymbol{\beta}} = 0.1.$

For the first one, a 5×5 grid was defined by $S_{gr} = \{0.1, 0.3, 0.5, 0.7, 0.9\}$ as $S_{gr} \times S_{gr}$ and the set of possible locations $S_{gr}^{\mathcal{X}} \triangleq \{(\mathcal{X}_1, \mathcal{X}_2) : \mathcal{X}_i \in (S_{gr} \times S_{gr}), i = 1, 2\}$ (comprised of 625 "locations") was exhaustively searched. The minimum of $\nu(\cdot, \cdot)$ on $S_{gr}^{\mathcal{X}}$ was found to be 1.2563 and it was attained at the location $\left(\begin{bmatrix} 0.5 \\ 0.7 \end{bmatrix}, \begin{bmatrix} 0.5 \\ 0.3 \end{bmatrix}\right)$.

For $N_{\boldsymbol{a}} > N_{\alpha\varepsilon}$, a search was also carried out on a set of $N_{\boldsymbol{a}}$ pseudo-random samples of a constant pdf on $S_{\mathcal{X}_2} = \{(\mathcal{X}_1, \mathcal{X}_2) : \mathcal{X}_i \in (0, L_x) \times (0, L_y), i = 1, 2\}$, with α and ε set to $\alpha = \varepsilon = 10^{-2}, \quad N_{\alpha\varepsilon} = 2/\log(1/0.99) \approx 454.5454$, so that $N_{\boldsymbol{a}}$ was taken to be 500.

The minimum of $\nu(\cdot, \cdot)$ on the 500 pseudo-random samples was found to be 1.2562 and it was attained at the location $\left(\begin{bmatrix} 0.2854 \\ 0.4170 \end{bmatrix}, \begin{bmatrix} 0.6641 \\ 0.5468 \end{bmatrix}\right)$.

Finally, the case of three scalar control signals was considered in the same setting. With the same values for α, ε and $N_{\boldsymbol{a}}$ a search on pseudo-random samples of a constant pdf on $S_{\mathcal{X}_3} = \{(\mathcal{X}_1, \mathcal{X}_2, \mathcal{X}_3) : \mathcal{X}_i \in (0, L_x) \times (0, L_y), i = 1, 2\}$ was carried out leading to the minimum value 1.1431 for $\nu(\cdot, \cdot, \cdot)$ which was attained at the location

$$\mathcal{X}_1 = \begin{bmatrix} 0.2952 \\ 0.4485 \end{bmatrix}, \quad \mathcal{X}_2 = \begin{bmatrix} 0.7628 \\ 0.2222 \end{bmatrix}, \quad \mathcal{X}_3 = \begin{bmatrix} 0.7064 \\ 0.8012 \end{bmatrix}$$

8 References

[1] K. Ito & F. Kappel, The Trotter-Kato theorem and approximation of PDEs, Mathematics of Computation, vol.67, n.221, pp.21–44, 1989.

[2] P. I. Kogut & G. R. Leugering, *Optimal Control Problems for Partial Differential Equations on Reticulated Domains: Approximation and Asymptotic Analysis*, Birkhäuser, 2011.

[3] G. Luenberger, *Optimization by vector space methods*, John Wiley & Sons, 1969.

[4] K. Morris, Design of finite- dimensional controllers for infinite-dimensional systems by approximation, Journal of Mathematical Systems, Estimation and Control, v.4, n.2, pp.1-30, 1994.

[5] R. Tempo & H. Ishii, Monte Carlo and Las Vegas randomized algorithms for systems and control: An introduction, European Journal of Control, vol.13, n.2–3, pp.189–203, 2007.

[6] F. Tröltzsch, *Optimal Control of Partial Differential Equations: Theory, Methods, and Applications.*, American Mathematical Society, 2010.

[7] E. Zuazua, Controllability of Partial Differential Equations and its Semi-discrete Approximations, Discrete and Continuous Dynamical Systems, vol.8, n.2, pp.469–513, 2002.

9 Appendix

Proof of Proposition 4.1(a): Consider the following auxiliary propositions

Auxiliary Proposition 1: $\mathcal{T}_\theta[\boldsymbol{u}] - \mathcal{T}_\theta^K[\boldsymbol{u}] = E_S^K[\boldsymbol{u}] + E_{\mathcal{T}}^K[\boldsymbol{u}]$ where

$$E_S^K[\boldsymbol{u}] \triangleq \int_0^{t_F} \sum_{i=1}^m \left(S_A(t_F-\tau) - S_K(t_F-\tau)\right)[P_K[\boldsymbol{\beta}_{\boldsymbol{S}i}]]\, u_i(\tau)d\tau \quad \text{and}$$

$$E_{\mathcal{T}}^K[\boldsymbol{u}] \triangleq \int_0^{t_F} \sum_{i=1}^m S_A(t_F-\tau)\,[(\mathbf{I}-P_K)[\boldsymbol{\beta}_{\boldsymbol{S}i}]]\, u_i(\tau)d\tau. \qquad \nabla$$

Auxiliary Proposition 2: $\|E_{\mathcal{T}}^K[\boldsymbol{u}]\|_{L_2(U)} \leq \eta_{\mathcal{T}\boldsymbol{f}}^K \|\boldsymbol{u}\|_{L_2(0,t_F)^m}$ and $\|E_S^K[\boldsymbol{u}]\|_{L_2(U)} \leq \eta_{\mathcal{T}\boldsymbol{g}}^K \|\boldsymbol{u}\|_{L_2(0,t_F)^m}$,
where

$$\eta_{\mathcal{T}\boldsymbol{f}}^K \triangleq \left\{\sum_{i=1}^m \|f_i^K(t_F-\cdot)\|^2_{L_2(0,t_F)}\right\}^{1/2} \quad , \quad \eta_{\mathcal{T}\boldsymbol{g}}^K \triangleq \left\{\sum_{i=1}^m \|g_i^K(t_F-\cdot)\|^2_{L_2(0,t_F)}\right\}^{1/2}$$

$$f_i^K(t_F-\sigma) \triangleq \|S_A(t_F-\sigma)\,[(\mathbf{I}-P_K)[\boldsymbol{\beta}_{\boldsymbol{S}i}]]\,\|_{L_2(U)} \qquad \text{and}$$

$$g_i^K(t_F-\sigma) \triangleq \|\left(S_A(t_F-\sigma) - S_K(t_F-\sigma)\right)[P_K[\boldsymbol{\beta}_{\boldsymbol{S}i}]]_{L_2(U)}\,. \qquad \nabla$$

Proposition 4.1(a) follows immediately from the two statements above, since bringing the second one to bear on the first leads to

$$\mathcal{T}_\theta[\boldsymbol{u}] - \mathcal{T}_\theta^K[\boldsymbol{u}] \leq (\eta_{\mathcal{T}\boldsymbol{f}}^K + \eta_{\mathcal{T}\boldsymbol{g}}^K)\|\boldsymbol{u}\|_{L_2(0,t_F)^m} \quad (\text{e.i., } \eta_{\mathcal{T}}^K = \eta_{\mathcal{T}\boldsymbol{f}}^K + \eta_{\mathcal{T}\boldsymbol{g}}^K).$$

■

Proof of Auxiliary Proposition 1: Note first that

$$\forall \theta \in L_2(U), \ \ \forall \ t \in [0,t_F] \ \ , \ \ P_K\left[S_K(t)[P_K(\theta)]\right] = S_K(t)[P_K(\theta)] \tag{A.1}$$

It then follows that

$$\mathcal{T}_\theta^K[\boldsymbol{u}] = P_K \int_0^{t_F} S_K(t_F-\alpha)\left[P_K[\boldsymbol{\beta}_{\boldsymbol{S}}^{\mathrm{T}}\boldsymbol{u}(\alpha)]\right] d\alpha = \int_0^{t_F} P_K\left[S_K(t_F-\alpha)[P_K[\boldsymbol{\beta}_{\boldsymbol{S}}^{\mathrm{T}}\boldsymbol{u}(\alpha)]]\right] d\alpha$$

so that (in the light of (A.1))

$$\mathcal{T}_\theta^K[\boldsymbol{u}] = \int_0^{t_F} S_K(t_F-\alpha)\left[P_K[\boldsymbol{\beta}_{\boldsymbol{S}}^{\mathrm{T}}\boldsymbol{u}(\alpha)]\right] d\alpha = \int_0^{t_F} \sum_{i=1}^{m} \left(S_K(t_F-\alpha)\left[P_K[\boldsymbol{\beta}_{Si}u_i(\alpha)]\right]\right) d\alpha \tag{A.2}$$

Note now that

$$\mathcal{T}_\theta[\boldsymbol{u}] = \int_0^{t_F} S_A(t_F-\alpha)[\boldsymbol{\beta}_{\boldsymbol{S}}^{\mathrm{T}}\boldsymbol{u}(\alpha)]d\alpha = \int_0^{t_F} S_A(t_F-\alpha)\left[P_K[\boldsymbol{\beta}_{\boldsymbol{S}}^{\mathrm{T}}\boldsymbol{u}(\alpha)]\right] d\alpha + E_{\mathcal{T}}^K[\boldsymbol{u}]$$

where

$$E_{\mathcal{T}}^K[\boldsymbol{u}] \triangleq \int_0^{t_F} S_A(t_F-\alpha)\left[(\mathbf{I}-P_K)[\boldsymbol{\beta}_{\boldsymbol{S}}^{\mathrm{T}}\boldsymbol{u}(\alpha)]\right] d\alpha = \int_0^{t_F} \sum_{i=1}^{m} S_A(t_F-\alpha)\left[(\mathbf{I}-P_K)[\boldsymbol{\beta}_{\boldsymbol{S}i}u_i(\alpha)]\right] d\alpha.$$

As a result,

$$\mathcal{T}_\theta[\boldsymbol{u}] - \mathcal{T}_\theta^K[\boldsymbol{u}] = \int_0^{t_F} \sum_{i=1}^{m} S_A(t_F-\alpha)\left[P_K[\boldsymbol{\beta}_{\boldsymbol{S}i}u_i(\alpha)]\right] d\alpha + E_{\mathcal{T}}^K[\boldsymbol{u}] - \mathcal{T}_\theta^K[\boldsymbol{u}]$$

$$\Rightarrow \text{ (in the light of (A.2)) } \mathcal{T}_\theta[\boldsymbol{u}] - \mathcal{T}_\theta^K[\boldsymbol{u}] = E_S^K[\boldsymbol{u}] + E_{\mathcal{T}}^K[\boldsymbol{u}].$$

To conclude the proof of Auxiliary Proposition 1, it remains to prove that (A.1)) holds. To this effect let $\hat{\theta}_K(t) \triangleq S_K(t)[P_K(\theta)]$ and $\hat{\theta}_K^a(t) = P_K[\hat{\theta}_K(t)]$ and note that:

(i) $\hat{\theta}_K(0) = P_K[\theta]$ and $\hat{\theta}_K^a(0) = P_K[\hat{\theta}_K(0)] = P_K\left[P_K[\theta]\right] = P_K[\theta] = \hat{\theta}_K(0)$,

(ii) $\forall t \in (0, t_F), \dot{\hat{\theta}}_K(t) = \mathcal{A}_K \circ P_K\left[S_K(t)[P_K(\theta)]\right] = \mathcal{A}_K[\hat{\theta}_K(t)]$ and $\dot{\theta}_K^a(t) = P_K[\dot{\hat{\theta}}_K(t)] = P_K[\mathcal{A}_K[\hat{\theta}_K(t)]] = \mathcal{A}_K[\hat{\theta}(t)]$ (since $\mathcal{A}_K = P_K \circ \mathcal{A}_{|X_K} \circ P_K$ and, hence, $P_K \circ \mathcal{A}_K = \mathcal{A}_K$).

Thus, in the light of (i) and (ii), $\forall t \in [0,t]$, $\hat{\theta}_K(t) = \hat{\theta}_K^a(t)$. ■

Proof of Auxiliary Proposition 2: Note first that

$$\|E_{\mathcal{T}}^K[\boldsymbol{u}]\|_{L_2(U)} \leq \sum_{i=1}^{m} \int_0^{t_F} \|S_A(t_F-\alpha)\left[(I-P_K)[\boldsymbol{\beta}_{\boldsymbol{S}i}]\right] u_i(\alpha)\|_{L_2(U)} d\alpha \Rightarrow$$

$$\|E^K_{\mathcal{T}}[\boldsymbol{u}]\|_{L_2(U)} \leq \sum_{i=1}^{m} \int_0^{t_F} f_i^K(t_F-\alpha)|u_i(\alpha)|d\alpha \leq \sum_{i=1}^{m} \|f_i^K(t_F-\cdot)\|_{L_2(0,t_F)}\|u_i\|_{L_2(0,t_F)},$$

where $f_i^K(t_F-\alpha) \triangleq \|S_A(t_F-\alpha)\left[(\mathbf{I}-P_K)[\boldsymbol{\beta}_{Si}]\right\|_{L_2(U)} \quad \Rightarrow$

$$\|E^K_{\mathcal{T}}[\boldsymbol{u}]\|_{L_2(U)} \leq \left\{\sum_{i=1}^{m} \|f_i^K(t_F-\cdot)\|^2_{L_2(0,t_F)}\right\}^{1/2} \|\boldsymbol{u}\|_{L_2(0,t_F)^m}.$$

Proceeding along the same lines, it follows that

$$\|E^K_{S}[\boldsymbol{u}]\|_{L_2(U)} \leq \left\{\sum_{i=1}^{m} \|g_i^K(t_F-\cdot)\|^2_{L_2(0,t_F)}\right\}^{1/2} \|\boldsymbol{u}\|_{L_2(0,t_F)^m}.$$

■

Proof of Proposition 4.1(b): Note first that $\|S_A(t)\| \leq \mu_A e^{\sigma_A t}$ and, hence,

$$\begin{aligned}
&f_i^K(t_F-\alpha) \leq \mu_A e^{\sigma_A(t_F-\alpha)}\|(I-P_K)[\boldsymbol{\beta}_{\boldsymbol{S}i}]\|_{L_2(U)} \Rightarrow \\
&\|f_i^K(t_F-\alpha)\|_{L_2(0,t_F)} = \mu_A\|(I-P_K)[\boldsymbol{\beta}_{\boldsymbol{S}i}]\|_{L_2(U)}\|e^{\sigma_A(t_F-\cdot)}\|_{L_2(0,t_F)}.
\end{aligned}$$

Thus it follows from (16) that $\|f_i^K(t_F-\cdot)\|_{L_2(0,t_F)} \to 0$ as $K \to \infty$ and, hence, $\eta^K_{\mathcal{T}\boldsymbol{f}} \to 0$ as $K \to \infty$.

With respect to $\{\eta^K_{\mathcal{T}\boldsymbol{g}}\}$ note that under the "assumption" that $B[\phi,\psi]$ satisfies Garding's inequality (cf; Theorem 2, Evans (1998), pp.300), it follows from (5) and Theorem 5.2 (Morris, 1994) that

$$\forall \theta \in L_2(U), \quad \|S_K(t)[\theta] - S_A(t)[\theta]\|_{L_2(U)}$$

converges uniformly on $[0,t_F]$ to zero as $K \to \infty$. Hence,

$$\|g_i^K(t_F-\cdot)\|^2_{L_2(0,t_F)} \to 0 \Rightarrow \eta^K_{\mathcal{T}\boldsymbol{g}} \to 0 \text{ as } \ K \to \infty.$$

It then follows that $\eta^K_{\mathcal{T}} = \eta^K_{\mathcal{T}\boldsymbol{f}} + \eta^K_{\mathcal{T}\boldsymbol{g}} \to 0$ as $K \to \infty$. ■

Proof of Proposition 5.1: Once it is established that $S_{\boldsymbol{u}F}$ is convex and closed the argument employed in the proof of Proposition 3.1 also proves Proposition 5.1.

To show that $S_{\boldsymbol{u}F}$ is convex let $\boldsymbol{u}_i \in S_{\boldsymbol{u}F}$, $i = 1, 2$ and define $\boldsymbol{u}(t;\sigma) = \sigma \boldsymbol{u}_1(t) + (1 - \sigma)\boldsymbol{u}_2(t)$, $\sigma \in [0, 1]$. Then $\forall i = 1, \ldots, m$, $\forall t \in [0, t_F]$ a.e $u_i(t, \sigma) = \sigma u_{1i}(t) + (1 - \sigma)u_{2i}(t) \in \mathcal{I}_{Fi}(t)$ (since $u_{1i}(t) \in \mathcal{I}_{Fi}(t)$, $u_{2i}(t) \in \mathcal{I}_{Fi}(t)$ and $\mathcal{I}_{Fi}(t)$ is an interval).

To show that $S_{\boldsymbol{u}F}$ is closed, let $\boldsymbol{u}^\ell \in S_{\boldsymbol{u}F}$ be such that $\boldsymbol{u}^\ell \to \boldsymbol{u}$ in the sense of the $L_2(0, t_F)^m$–norm. Then $\forall i = 1, \ldots m$, $\|u_i - u_i^\ell\|_{L_2(0,t_F)} \to 0$ and, hence,

$$\forall t \ \text{ a.e. in } [0, t_F] \ , \ |u_i(t) - u_i^\ell(t)| \to 0. \tag{A.3}$$

Now let $\mathrm{m}_i(t) \triangleq (1/2)(a_i(t) + b_i(t))$ and $\gamma_i(t) \triangleq (1/2)(b_i(t) - a_i(t))$. Then

$$\begin{aligned} \forall i, \forall \ell,\ \boldsymbol{u}^\ell \in S_{\boldsymbol{u}F} \Rightarrow \forall t \text{ a.e. in } [0, t_F], |u_i^\ell(t) - \mathrm{m}_i(t)| &\le \gamma_i(t) \text{ and} \\ |u_i(t) - u_i^\ell(t)| &= |(u_i(t) - \mathrm{m}_i(t)) - (u_i^\ell(t) - \mathrm{m}_i(t))| \Rightarrow \\ &\ge |u_i(t) - \mathrm{m}_i(t)| - |u_1^\ell(t) - \mathrm{m}_i(t)|. \end{aligned}$$

Thus $\forall t$ a.e. in $[0, t_F]$, $|u_i(t) - u_i^\ell(t)| \ge |u_i(t) - \mathrm{m}_i(t)| - \gamma_i(t) \quad \Rightarrow \quad \forall i = 1, \cdots, m,\ \forall \ell \in \mathbb{Z}_+$

$$|u_i(t) - \mathrm{m}_i(t)| \le \gamma_i(t) + |u_i(t) - u_i^\ell(t)|.$$

Thus in the light of (A.2), $\forall t$ a.e. in $[0, t_F]$, $\forall i = 1, \cdots, m$, $|u_i(t) - \mathrm{m}_i(t)| \le \gamma_i(t) \Rightarrow \boldsymbol{u} \in S_{\boldsymbol{u}F}$.

With respect to the optimality condition, note that

$$\mathcal{J}(\boldsymbol{u} + \Delta \boldsymbol{u}) = \mathcal{J}(\boldsymbol{u}) + 2\rho_{\boldsymbol{u}}\langle \boldsymbol{u}, \Delta \boldsymbol{u}\rangle + \rho_{\boldsymbol{u}}\|\Delta \boldsymbol{u}\|^2_{L_2(0,t_F)^m} + 2\langle \mathcal{T}_\theta[\boldsymbol{u}] - \boldsymbol{\theta}_{ro}, \mathcal{T}_\theta[\Delta \boldsymbol{u}]\rangle + \|\mathcal{T}_\theta[\Delta \boldsymbol{u}]\|_2^2$$

$$\Longleftrightarrow \mathcal{J}(\boldsymbol{u} + \Delta \boldsymbol{u}) = \mathcal{J}(\boldsymbol{u}) + 2\langle \rho_{\boldsymbol{u}}\boldsymbol{u} + Z_a[\boldsymbol{u}], \Delta \boldsymbol{u}\rangle + (\rho_{\boldsymbol{u}}\|\Delta \boldsymbol{u}\|_2^2 + \|\mathcal{T}_\theta[\Delta \boldsymbol{u}]\|^2_{L_2(0,t_F)^m}),$$

where $Z_a[\boldsymbol{u}] \triangleq \mathcal{T}_\theta^*[\mathcal{T}_\theta[\boldsymbol{u}] - \theta_{ro}]$.

Thus $\boldsymbol{u}_c \in S_{\boldsymbol{u}F}$ is optimal if and only if $\forall \Delta \boldsymbol{u} \in L_2(0, t_F)^m$ such that $(\boldsymbol{u}_c + \Delta \boldsymbol{u}) \in S_{\boldsymbol{u}F}$, $\langle \rho_{\boldsymbol{u}}\boldsymbol{u}_c + Z_a[\boldsymbol{u}_c], \Delta \boldsymbol{u}\rangle \ge 0$.

Note now that since

$$\langle \rho_{\boldsymbol{u}} \boldsymbol{u} + Z_a[\boldsymbol{u}], \Delta \boldsymbol{u} \rangle = \int_0^{t_F} (\rho_{\boldsymbol{u}} \boldsymbol{u}(t) + Z_a[\boldsymbol{u}](t))^{\mathrm{T}} \Delta \boldsymbol{u}(t) dt \ ,$$

the condition

"$\forall \Delta \boldsymbol{u}$ such that $(\boldsymbol{u}_c + \Delta \boldsymbol{u}) \in S_{\boldsymbol{u}F}$, $\forall t$ a.e. in $[0, t_F]$, $(\rho_{\boldsymbol{u}} \boldsymbol{u}_c(t) + Z_a[\boldsymbol{u}_c](t))^{\mathrm{T}} \Delta \boldsymbol{u}(t) \geq 0$"
is sufficient for $\boldsymbol{u}_c$ to be optimal. To see that it is also necessary, suppose that there exists $\Delta \boldsymbol{u} \in L_2(0, t_F)^m$ such that $(\boldsymbol{u}_c + \Delta \boldsymbol{u}) \in S_{\boldsymbol{u}F}$ and for some subset S_a of $[0, t_F]$ with non-zero measure, $(\rho_{\boldsymbol{u}} \boldsymbol{u}_c(t) + Z_a[\boldsymbol{u}_c](t))^{\mathrm{T}} \Delta \boldsymbol{u}(t) < 0$ for any $t \in S_a$. Then, defining $\hat{\Delta} \boldsymbol{u}(t) = \Delta \boldsymbol{u}(t)$ for $t \in S_a$ and $\hat{\Delta} \boldsymbol{u}(t) = 0$ otherwise,

$$(\boldsymbol{u}_c + \hat{\Delta} \boldsymbol{u}) \in S_{\boldsymbol{u}F} \text{ and } \langle \rho_{\boldsymbol{u}} \boldsymbol{u}_c + Z_a[\boldsymbol{u}_c], \hat{\Delta} \boldsymbol{u} \rangle = \int_{S_a} (\rho_{\boldsymbol{u}} \boldsymbol{u}_c(t) + Z_a[\boldsymbol{u}_c](t))^{\mathrm{T}} \Delta \boldsymbol{u}(t) dt < 0$$

so that $\boldsymbol{u}_c$ cannot be optimal. ■

Proof of Proposition 5.2: Consider the following optimization problem for $t \in [0, t_F]$: $\min_{\boldsymbol{v} \in \mathbb{R}^m} \|\rho_{\boldsymbol{u}} \boldsymbol{v} + Z_a[\boldsymbol{u}_c](t)\|_2^2$ subject to $\forall i = 1, \cdots, m \ v_i \in \mathcal{I}_{Fi}(t)$.

As $\|\rho_{\boldsymbol{u}}(\boldsymbol{v}+\Delta\boldsymbol{v})+Z_a[\boldsymbol{u}_c](t)\|_2^2 = \|\rho_{\boldsymbol{u}}\boldsymbol{v}+Z_a[\boldsymbol{u}_c](t)\|_2^2+2\langle \rho_{\boldsymbol{u}}\boldsymbol{v}+Z_a[\boldsymbol{u}_c](t), \rho_{\boldsymbol{u}}\Delta\boldsymbol{v}\rangle+\|\rho_{\boldsymbol{u}}\Delta\boldsymbol{v}\|_2^2$ $\boldsymbol{v}_t$ is optimal if and only if $\boldsymbol{v}_{ti} \in \mathcal{I}_{Fi}(t)$ and $\forall \Delta \boldsymbol{v}$ such that $\boldsymbol{v}_{ti} + \Delta \boldsymbol{v}_i \in \mathcal{I}_{Fi}(t)$

$$\langle \rho_{\boldsymbol{u}} \boldsymbol{v}_t + Z_a[\boldsymbol{u}_c](t), \rho_{\boldsymbol{u}} \Delta \boldsymbol{v} \rangle \geq 0 \Leftrightarrow \langle \rho_{\boldsymbol{u}} \boldsymbol{v}_t + Z_a[\boldsymbol{u}_c](t), \Delta \boldsymbol{v} \rangle \geq 0. \tag{A.4}$$

As the solution of both this problem and of *Prob. II* are unique it follows from (A.3) and (A.4) that $\forall t$ a.e. in $[0, t_F]$, $\boldsymbol{u}_c(t) = \boldsymbol{v}_t(t)$.

Now, the problem above is equivalent to the problem

$$\min_{\boldsymbol{v}_i \in \mathbb{R}, i=1,\cdots,m} \sum_{i=1}^{m} (\rho_{\boldsymbol{u}} v_i + \{Z_a[\boldsymbol{u}_c](t)\}_i)^2 \text{ subject to } \forall i = 1, \cdots, m, v_i \in \mathcal{I}_{Fi}(t)$$

which breaks down into m problems (for $i = 1, \cdots m$)

$$\min_{\boldsymbol{v}_i \in \mathbb{R}} (\boldsymbol{v}_i - (1/\rho_{\boldsymbol{u}}) \{-Z_a[\boldsymbol{u}_c](t)\}_i)^2 \text{ subject to } \boldsymbol{v}_i \in \mathcal{I}_{Fi}(t)$$

the solution of which is given by

$$\begin{aligned}
\boldsymbol{v}_i &= -(1/\rho_{\boldsymbol{u}})\left\{Z_a[\boldsymbol{u}_c(t)]\right\}_i \quad \text{if} \quad -(1/\rho_{\boldsymbol{u}})\left\{Z_a[\boldsymbol{u}_c(t)]\right\}_i \in \mathcal{I}_{Fi}(t) \\
\boldsymbol{v}_i &= \boldsymbol{u}_{bi}(t) \quad \text{if} \quad -(1/\rho_{\boldsymbol{u}})\left\{Z_a[\boldsymbol{u}_c(t)]\right\}_i > \boldsymbol{u}_{bi}(t) \\
\boldsymbol{v}_i &= \boldsymbol{u}_{ai}(t) \quad \text{if} \quad -(1/\rho_{\boldsymbol{u}})\left\{Z_a[\boldsymbol{u}_c(t)]\right\}_i < \boldsymbol{u}_{ai}(t).
\end{aligned}$$

■

Proof of Proposition 5.3: Proceeding as in the proof of Proposition 4.2, write

$$\begin{aligned}
\mathcal{J}(\boldsymbol{u}_c^K) = \mathcal{J}(\boldsymbol{u}_c + (\boldsymbol{u}_c^K - \boldsymbol{u}_c)) &= \mathcal{J}(\boldsymbol{u}_c) + 2\langle \rho_{\boldsymbol{u}}\boldsymbol{u}_c + Z_a[\boldsymbol{u}_c], (\boldsymbol{u}_c^K - \boldsymbol{u}_c)\rangle \\
&\quad + \|\rho_{\boldsymbol{u}}(\boldsymbol{u}_c^K - \boldsymbol{u}_c)\|_2^2 + \mathcal{T}_\theta[\boldsymbol{u}_c^K - \boldsymbol{u}_c]\| \qquad \text{(A.5)}
\end{aligned}$$

and note that (as in the derivation of (24))

$$\begin{aligned}
\mathcal{J}_K(\boldsymbol{u}_c^K) &\leq \mathcal{J}_K(\boldsymbol{u}_c) = \mathcal{J}(\boldsymbol{u}_c) - E_{\mathcal{J}}^K(\boldsymbol{u}_c) \Leftrightarrow \\
\mathcal{J}(\boldsymbol{u}_c^K) &- E_{\mathcal{J}}^K(\boldsymbol{u}_c^K) \leq \mathcal{J}(\boldsymbol{u}_c) - E_{\mathcal{J}}^K(\boldsymbol{u}_c) \Rightarrow \qquad \text{(A.6)} \\
\mathcal{J}(\boldsymbol{u}_c^K) &\leq \mathcal{J}(\boldsymbol{u}_c) - E_{\mathcal{J}}^K(\boldsymbol{u}_c) + E_{\mathcal{J}}^K(\boldsymbol{u}_c^K) \Rightarrow \qquad \text{(A.7)} \\
\mathcal{J}(\boldsymbol{u}_c^K) &\leq \mathcal{J}(\boldsymbol{u}_c) + |E_{\mathcal{J}}^K(\boldsymbol{u}_c)| + |E_{\mathcal{J}}^K(\boldsymbol{u}_c^K)|. \qquad \text{(A.8)}
\end{aligned}$$

Combining (A.5) and (A.8) leads to

$$\begin{aligned}
\|\rho_{\boldsymbol{u}}(\boldsymbol{u}_c^K - \boldsymbol{u}_c)\|_2^2 + \|\mathcal{T}_\theta[\boldsymbol{u}_c^K - \boldsymbol{u}_c]\|_2^2 &+ 2\langle \rho_{\boldsymbol{u}}\boldsymbol{u}_c + Z_a[\boldsymbol{u}_c], (\boldsymbol{u}_c^K - \boldsymbol{u}_c)\rangle \\
&\leq |E_{\mathcal{J}}^K(\boldsymbol{u}_c)| + |E_{\mathcal{J}}^K(\boldsymbol{u}_c^K)|
\end{aligned}$$

$\Rightarrow$ (in the light of the optimality condition of Proposition 5.1)

$$\rho_{\boldsymbol{u}}\|\boldsymbol{u}_c^K - \boldsymbol{u}_c\|_2^2 \leq |E_{\mathcal{J}}^K(\boldsymbol{u}_c) + |E_{\mathcal{J}}^K(\boldsymbol{u}_c^K)|.$$

With the same argument used in the proof of Proposition 4.2, the right hand side of the last inequality above is shown to go to zero as $K \to 0$. Hence, $\|\boldsymbol{u}_c^K - \boldsymbol{u}_c\|_2 \to$ as $K \to \infty$. ■

Proof of Proposition 5.5:: The optimality condition satisfied by $\boldsymbol{u}_c^K(\boldsymbol{\lambda})$ is given by

$$\rho_{\boldsymbol{u}}\boldsymbol{u} + (\mathcal{T}_\theta^K)^*[\mathcal{T}_\theta^K[\boldsymbol{u}] - \boldsymbol{\theta}_{ro}] + (\boldsymbol{\lambda}_b - \boldsymbol{\lambda}_a) = 0 \tag{A.9}$$

or, equivalently, taking orthogonal projections $\boldsymbol{u}^1$ and $\boldsymbol{u}^2$ of $\boldsymbol{u}$ on $(\mathcal{T}_\theta^K)^*[L_2(U)]$ and on its orthogonal complement,

$$\rho_{\boldsymbol{u}}\boldsymbol{u}^1 + (\mathcal{T}_\theta^K)^*[\mathcal{T}_\theta^K[\boldsymbol{u}^1 + \boldsymbol{u}^2]] - (\mathcal{T}_\theta^K)^*[\theta_{ro}] - \boldsymbol{\lambda}_{ab}^1 = 0$$

and $\rho_{\boldsymbol{u}}\boldsymbol{u}^2 - \boldsymbol{\lambda}_{ab}^2 = 0$ where $\boldsymbol{\lambda}_{ab} \triangleq \boldsymbol{\lambda}_a - \boldsymbol{\lambda}_b$, $\boldsymbol{\lambda}_{ab}^1$ and $\boldsymbol{\lambda}_{ab}^2$ are the corresponding projections of $\boldsymbol{\lambda}_{ab}$.

Noting further that $\mathcal{T}_\theta^K[\boldsymbol{u}^2] = 0$ ($\boldsymbol{u}^2$ is orthogonal to the range space of $(\mathcal{T}_\theta^K)^*$ and hence is in the null space of $\mathcal{T}_\theta^K$) the equations above can be rewritten as

$$\rho_{\boldsymbol{u}}\boldsymbol{u}^1 + (\mathcal{T}_\theta^K)^*[\mathcal{T}_\theta^K[\boldsymbol{u}^1]] - (\mathcal{T}_\theta^K)^*[\theta_{ro}] - \boldsymbol{\lambda}_{ab}^1 = 0$$

and $\quad \rho_{\boldsymbol{u}}\boldsymbol{u}^2 = \boldsymbol{\lambda}_{ab} - \boldsymbol{\lambda}_{ab}^1$.

Now, $\mathcal{T}_\theta^K[\boldsymbol{u}] = \sum_{k=1}^{K} c_k(t_F;\boldsymbol{u})\phi_k$ and $(\mathcal{T}_\theta^K)^*[\boldsymbol{w}](\tau) = \mathbf{H}_K^{\mathrm{T}}(t_F - \tau)\bar{\boldsymbol{w}}^K$,
where $\{\phi_k; k = 1, \cdots n(K)\}$ is an orthogonal basis for X_K,
$c_k(t_F;\boldsymbol{u}) \triangleq e_k(n(K))^{\mathrm{T}} \int_0^{t_F} \mathbf{H}_K(t_F - \tau)\boldsymbol{u}(\tau)d\tau$, $\mathbf{H}_K(t) \triangleq \phi_K(t)\bar{\boldsymbol{\beta}}_{\boldsymbol{S}K}$,
$\bar{\boldsymbol{w}}^K \triangleq [\langle \boldsymbol{w}, \phi_1, \rangle \cdots \langle \boldsymbol{w}, \phi_{n(K)}\rangle]$ and

$$\bar{\boldsymbol{\beta}}_{\boldsymbol{S}K} \triangleq \begin{bmatrix} \langle (\boldsymbol{\beta}_{\boldsymbol{S}}^{\mathrm{T}})_1, \phi_1\rangle \cdots \langle (\boldsymbol{\beta}_{\boldsymbol{S}}^{\mathrm{T}})_m, \phi_1\rangle \\ \langle (\boldsymbol{\beta}_{\boldsymbol{S}}^{\mathrm{T}})_1, \phi_{n(K)}\rangle \cdots \langle (\boldsymbol{\beta}_{\boldsymbol{S}}^{\mathrm{T}})_m, \phi_{n(K)}\rangle \end{bmatrix} \in \mathbb{R}^{n(K)\times m}.$$

It follows that $\boldsymbol{u}^1 = \mathbf{H}_K^{\mathrm{T}}(t_F - \cdot)\boldsymbol{\alpha}_c^K$ and $\boldsymbol{\lambda}_{ab}^1 = \mathbf{H}_K^{\mathrm{T}}(t_F - \cdot)\boldsymbol{\alpha}_{\boldsymbol{\lambda}}^K$ and, hence, the equation involving $\boldsymbol{u}^1$ above can be written as

$$\mathbf{H}_K^{\mathrm{T}}(t_F - \cdot)\left\{\rho_{\boldsymbol{u}}\boldsymbol{\alpha}_c^K + \bar{\boldsymbol{w}}_a^K[\boldsymbol{\alpha}_c^K] - \bar{\boldsymbol{\theta}}_{ro}^K - \boldsymbol{\alpha}_{\boldsymbol{\lambda}}^K\right\} = 0, \tag{A.10}$$

where $\bar{\boldsymbol{\theta}}_{ro}^K \triangleq [\langle \theta_{ro}, \theta_1 \rangle \cdots \langle \theta_{ro}, \theta_{n(K)} \rangle]^{\mathrm{T}}$ and

$$\bar{\boldsymbol{w}}_a^K[\boldsymbol{\alpha}_c^K] \triangleq \left[\langle \mathcal{T}_\theta^K[\boldsymbol{u}^1], \phi_1 \rangle \cdots \langle \mathcal{T}_\theta^K[\boldsymbol{u}^1], \phi_{n(K)} \rangle\right]^{\mathrm{T}}$$

i.e., $\bar{\boldsymbol{w}}[\boldsymbol{\alpha}_c^K] = [c_1(t_F; \boldsymbol{u}^1) \cdots c_{n(K)}(t_F; \boldsymbol{u}^1)]^{\mathrm{T}} = \int_0^{t_F} \mathbf{H}_K(t-\tau)\mathbf{H}_K^{\mathrm{T}}(t_F-\tau)\boldsymbol{u}^1(\tau)d\tau$
$\Leftrightarrow \bar{\boldsymbol{w}}_a^K[\boldsymbol{\alpha}_c^K] = \mathbf{G}_K \boldsymbol{\alpha}_c^K$ and $\mathbf{G}_K \triangleq \displaystyle\int_0^{t_F} \mathbf{H}_K(t_F-\tau)\mathbf{H}_K^{\mathrm{T}}(t_F-\tau)d\tau.$
A sufficient condition for (A.10) to be satisfied is then given by

$$\rho_{\boldsymbol{u}}\boldsymbol{\alpha}_c^K + \mathbf{G}_K\boldsymbol{\alpha}_c^K = \bar{\boldsymbol{\theta}}_{ro}^K + \boldsymbol{\alpha}_{\boldsymbol{\lambda}}^K \;\Leftrightarrow\; \boldsymbol{\alpha}_c^K = (\rho_{\boldsymbol{u}}\mathbf{I} + \mathbf{G}_K)^{-1}(\bar{\boldsymbol{\theta}}_{ro}^K + \boldsymbol{\alpha}_{\boldsymbol{\lambda}}^K)$$

It then follows that $\boldsymbol{u}_c^K[\boldsymbol{\lambda}]$ is given by (since $\rho_{\boldsymbol{u}}\boldsymbol{u}^2 = \boldsymbol{\lambda}_{ab}^2$)

$$\begin{aligned}
\boldsymbol{u}_c^K[\boldsymbol{\lambda}] &= \mathbf{H}_K^{\mathrm{T}}(t_F-\cdot)\boldsymbol{\alpha}_c^K + \rho_{\boldsymbol{u}}^{-1}(\boldsymbol{\lambda}_{ab} - \mathbf{H}_K^{\mathrm{T}}(t_F-\cdot)\boldsymbol{\alpha}_{\boldsymbol{\lambda}}^K) \Leftrightarrow \\
\boldsymbol{u}_c^K[\boldsymbol{\lambda}](\tau) &= \mathbf{H}_K^{\mathrm{T}}(t_F-\tau)(\boldsymbol{\alpha}_c^K - \rho_{\boldsymbol{u}}^{-1}\boldsymbol{\alpha}_{\boldsymbol{\lambda}}^K) + \rho_{\boldsymbol{u}}^{-1}\boldsymbol{\lambda}_{ab}(\tau) \Leftrightarrow \\
\boldsymbol{u}_c^K[\boldsymbol{\lambda}](\tau) &= \boldsymbol{u}_K(\tau) + \mathbf{H}_K^{\mathrm{T}}(t-\tau)\left\{(\rho_{\boldsymbol{u}}\mathbf{I} + \mathbf{G}_K)^{-1} - \rho_{\boldsymbol{u}}^{-1}\mathbf{I}\right\}\boldsymbol{\alpha}_{\boldsymbol{\lambda}}^K + \rho_{\boldsymbol{u}}^{-1}\boldsymbol{\lambda}_{ab}(\tau).
\end{aligned}$$

With respect to the dual functional $\varphi_{DK}(\boldsymbol{\lambda})$, rewrite Lag_K as

$$\begin{aligned}
Lag_K(\boldsymbol{u}, \boldsymbol{\lambda}) &= \langle \rho_{\boldsymbol{u}}\boldsymbol{u} + (\mathcal{T}_\theta^K)^*[\mathcal{T}_\theta^K[\boldsymbol{u}] - \boldsymbol{\theta}_{ro}] + (\boldsymbol{\lambda}_b - \boldsymbol{\lambda}_a), \boldsymbol{u}\rangle + \langle \mathcal{T}_\theta^K[\boldsymbol{u}] - \boldsymbol{\theta}_{ro}, -\boldsymbol{\theta}_{ro}\rangle \\
&+ \langle \boldsymbol{\lambda}_b - \boldsymbol{\lambda}_a, \boldsymbol{u}\rangle + 2\langle \boldsymbol{\lambda}_a, \boldsymbol{u}_a\rangle - 2\langle \boldsymbol{\lambda}_b, \boldsymbol{u}_b\rangle. \qquad \text{(A.11)}
\end{aligned}$$

Thus, as $\varphi_{DK}(\boldsymbol{\lambda}) = Lag_K(\boldsymbol{u}_c^K[\boldsymbol{\lambda}], \boldsymbol{\lambda})$, it follows from (A.9) and (A.11) that

$$\varphi_{DK}(\boldsymbol{\lambda}) = \langle \mathcal{T}_\theta^K[\boldsymbol{u}_c^K[\boldsymbol{\lambda}]] - \boldsymbol{\theta}_{ro}, -\boldsymbol{\theta}_{ro}\rangle + \langle \boldsymbol{\lambda}_b - \boldsymbol{\lambda}_a, \boldsymbol{u}_c^K[\boldsymbol{\lambda}]\rangle + 2\langle \boldsymbol{\lambda}_a, u_a\rangle - 2\langle \boldsymbol{\lambda}_b, u_b\rangle$$

or equivalently, since $\boldsymbol{u}_c^K[\boldsymbol{\lambda}] = \boldsymbol{u}_K - \boldsymbol{u}_K^\xi + \rho_{\boldsymbol{u}}^{-1}\boldsymbol{\lambda}_{ab}$,
$\varphi_{DK}(\boldsymbol{\lambda}) = \|\boldsymbol{\theta}_{ro}\|_2^2 + \langle \mathcal{T}_\theta^K[\boldsymbol{u}_K], -\boldsymbol{\theta}_{ro}\rangle + \hat{\varphi}_{DK}(\boldsymbol{\lambda}),$
where

$$\hat{\varphi}_{DK}(\boldsymbol{\lambda}) \triangleq \langle \mathcal{T}_\theta^K[\rho_{\boldsymbol{u}}^{-1}\boldsymbol{\lambda}_{ab} - \boldsymbol{u}_K^\xi], -\boldsymbol{\theta}_{ro}\rangle - \langle \boldsymbol{\lambda}_{ab}, \boldsymbol{u}_K - \boldsymbol{u}_K^\xi + \rho_{\boldsymbol{u}}^{-1}\boldsymbol{\lambda}_{ab}\rangle + 2\langle \boldsymbol{\lambda}_a, \boldsymbol{u}_a\rangle - 2\langle \boldsymbol{\lambda}_b, \boldsymbol{u}_b\rangle$$

i.e,

$$\begin{aligned}\hat{\varphi}_{DK}(\boldsymbol{\lambda}) &= -\rho_{\boldsymbol{u}}^{-1}\langle\boldsymbol{\lambda}_{ab},\boldsymbol{\lambda}_{ab}\rangle + \langle\boldsymbol{\lambda}_{ab}, \boldsymbol{u}_K^{\xi} - \boldsymbol{u}_K - \rho_{\boldsymbol{u}}^{-1}(\mathcal{T}_\theta^K)^*[\boldsymbol{\theta}_{ro}]\rangle + \langle\boldsymbol{u}_K^{\xi}, (\mathcal{T}_\theta^K)^*[\boldsymbol{\theta}_{ro}]\rangle \\ &+ 2\langle\boldsymbol{\lambda}_a,\boldsymbol{u}_a\rangle - 2\langle\boldsymbol{\lambda}_b,\boldsymbol{u}_b\rangle,\end{aligned}$$

and $\boldsymbol{u}_K^{\xi}[\boldsymbol{\lambda}] \triangleq \mathbf{H}_K^{\mathrm{T}}(t_F-\cdot)\left\{\rho_{\boldsymbol{u}}^{-1}\mathbf{I} - (\rho_{\boldsymbol{u}}\mathbf{I}+\mathbf{G}_K)^{-1}\right\}\boldsymbol{\alpha}_{\boldsymbol{\lambda}}^K$ or, equivalently

$$\begin{aligned}(\text{as}\ \ \rho_{\boldsymbol{u}}^{-1}\mathbf{I} - (\rho_{\boldsymbol{u}}\mathbf{I}+\mathbf{G}_K)^{-1} &= \rho_{\boldsymbol{u}}^{-1}\left\{\mathbf{I} - \rho_{\boldsymbol{u}}(\rho_{\boldsymbol{u}}\mathbf{I}+\mathbf{G}_K)^{-1}\right\} = \rho_{\boldsymbol{u}}^{-1}\left\{\mathbf{I} - (\mathbf{I}+\rho_{\boldsymbol{u}}^{-1}\mathbf{G}_K)^{-1}\right\} \\ &= \rho_{\boldsymbol{u}}^{-1}\left\{\rho_{\boldsymbol{u}}^{-1}\mathbf{G}_K(\mathbf{I}+\rho_{\boldsymbol{u}}^{-1}\mathbf{G}_K)^{-1}\right\} = \rho_{\boldsymbol{u}}^{-1}(\mathbf{I}+\rho_{\boldsymbol{u}}^{-1}\mathbf{G}_K)^{-1}\rho_{\boldsymbol{u}}^{-1}\mathbf{G}_K)\end{aligned}$$

$$\boldsymbol{u}_K^{\xi}[\boldsymbol{\lambda}] = \mathbf{H}_K^{\mathrm{T}}(t_F-\cdot)\rho_{\boldsymbol{u}}^{-1}(\rho_{\boldsymbol{u}}\mathbf{I}+\mathbf{G}_K)^{-1}\mathbf{G}_K\boldsymbol{\alpha}_{\boldsymbol{\lambda}}^K.$$

Finally, as $(\mathcal{T}_\theta^K)^*[\theta_{ro}] = \mathbf{H}_K^{\mathrm{T}}(t_F-\cdot)\bar{\boldsymbol{\theta}}_{ro}^K$ and $\boldsymbol{u}_K = \mathbf{H}_K^{\mathrm{T}}(t_F-\cdot)\boldsymbol{\alpha}_K$, $\langle\boldsymbol{u}_K^{\xi}[\boldsymbol{\lambda}], (\mathcal{T}_\theta^K)^*[\theta_{ro}]\rangle = \langle\rho_{\boldsymbol{u}}^{-1}(\rho_{\boldsymbol{u}}\mathbf{I}+\mathbf{G}_K)^{-1}\mathbf{G}_K\boldsymbol{\alpha}_{\boldsymbol{\lambda}}^K, \mathbf{G}_K\bar{\boldsymbol{\theta}}_{ro}^K\rangle_E$ and $\langle\boldsymbol{\lambda}_{ab}, \boldsymbol{u}_K^{\xi} - \boldsymbol{u}_K - \rho_{\boldsymbol{u}}^{-1}(\mathcal{T}_\theta^K)^*[\theta_{ro}]\rangle = \langle\boldsymbol{\xi}_{\boldsymbol{\lambda}}^K, \rho_{\boldsymbol{u}}^{-1}(\rho_{\boldsymbol{u}}\mathbf{I}+\mathbf{G}_K)^{-1}\mathbf{G}_K\boldsymbol{\alpha}_{\boldsymbol{\lambda}}^K - \boldsymbol{\alpha}_K - \rho_{\boldsymbol{u}}^{-1}\bar{\boldsymbol{\theta}}_{ro}^K\rangle_E$, where $\quad\boldsymbol{\alpha}_K = (\rho_{\boldsymbol{u}}\mathbf{I}+\mathbf{G}_K)^{-1}\bar{\boldsymbol{\theta}}_{ro}^K, \quad \boldsymbol{\xi}_{\boldsymbol{\lambda}}^K \triangleq \int_0^{t_F}\mathbf{H}_K(t_F-\tau)\boldsymbol{\lambda}_{ab}(\tau)d\tau \quad (\boldsymbol{\xi}_{\boldsymbol{\lambda}}^K = \mathbf{G}_K\boldsymbol{\alpha}_{\boldsymbol{\lambda}}^K)$.

As a result, $\hat{\varphi}_{DK}(\boldsymbol{\lambda})$ is given by

$$\begin{aligned}\hat{\varphi}_{DK}(\boldsymbol{\lambda}) &= -\rho_{\boldsymbol{u}}^{-1}\langle\boldsymbol{\lambda}_{ab},\boldsymbol{\lambda}_{ab}\rangle + \rho_{\boldsymbol{u}}^{-1}\langle\boldsymbol{\xi}_{\lambda}^K, (\rho_{\boldsymbol{u}}\mathbf{I}+\mathbf{G}_K)^{-1}\mathbf{G}_K\boldsymbol{\alpha}_{\boldsymbol{\lambda}}^K\rangle - \langle\boldsymbol{\xi}_{\boldsymbol{\lambda}}^K, \boldsymbol{\alpha}_K + \rho_{\boldsymbol{u}}^{-1}\bar{\boldsymbol{\theta}}_{ro}^K\rangle \\ &+\rho_{\boldsymbol{u}}^{-1}\langle(\rho_{\boldsymbol{u}}\mathbf{I}+\mathbf{G}_K)^{-1}\boldsymbol{\alpha}_{\boldsymbol{\lambda}}^K, \mathbf{G}_K\bar{\boldsymbol{\theta}}_{ro}^K\rangle + 2\langle\boldsymbol{\lambda}_a,\boldsymbol{u}_a\rangle - 2\langle\boldsymbol{\lambda}_b,\boldsymbol{u}_b\rangle.\end{aligned}$$

Now, $\langle(\rho_{\boldsymbol{u}}\mathbf{I}+\mathbf{G}_K)^{-1}\mathbf{G}_K\boldsymbol{\alpha}_{\boldsymbol{\lambda}}^K, \mathbf{G}_K\bar{\theta}_{ro}^K\rangle = \langle\boldsymbol{\xi}_{\boldsymbol{\lambda}}^K, (\rho_{\boldsymbol{u}}\mathbf{I}+\mathbf{G}_K)^{-1}\mathbf{G}_K\bar{\boldsymbol{\theta}}_{ro}^K\rangle \qquad \Rightarrow$

$$\begin{aligned}\hat{\varphi}_{DK}(\boldsymbol{\lambda}) &= -\rho_{\boldsymbol{u}}^{-1}\langle\boldsymbol{\lambda}_{ab},\boldsymbol{\lambda}_{ab}\rangle + \rho_{\boldsymbol{u}}^{-1}\langle\boldsymbol{\xi}_{\lambda}^K, (\rho_{\boldsymbol{u}}\mathbf{I}+\mathbf{G}_K)^{-1}\boldsymbol{\xi}_{\boldsymbol{\lambda}}^K\rangle - \langle\boldsymbol{\xi}_{\boldsymbol{\lambda}}^K, \boldsymbol{\alpha}_K\rangle \\ &+\rho_{\boldsymbol{u}}^{-1}\langle\boldsymbol{\xi}_{\boldsymbol{\lambda}}^K, \left\{(\rho_{\boldsymbol{u}}\mathbf{I}+\mathbf{G}_K)^{-1}\mathbf{G}_K - \mathbf{I}\right\}\bar{\boldsymbol{\theta}}_{ro}^K\rangle + 2\langle\boldsymbol{\lambda}_a,\boldsymbol{u}_a\rangle - 2\langle\boldsymbol{\lambda}_b,\boldsymbol{u}_b\rangle.\end{aligned}$$

Moreover, as $\quad(\rho_{\boldsymbol{u}}\mathbf{I}+\mathbf{G}_K)^{-1}\mathbf{G}_K = (\mathbf{I}+\rho_{\boldsymbol{u}}^{-1}\mathbf{G}_K)^{-1}\rho_{\boldsymbol{u}}^{-1}\mathbf{G}_K = \mathbf{I} - (\mathbf{I}+\rho_{\boldsymbol{u}}^{-1}\mathbf{G}_K)^{-1}$ so that

$$\begin{aligned}\hat{\varphi}_{DK}(\boldsymbol{\lambda}) &= -\rho_{\boldsymbol{u}}^{-1}\langle\boldsymbol{\lambda}_{ab},\boldsymbol{\lambda}_{ab}\rangle + \rho_{\boldsymbol{u}}^{-1}\langle\boldsymbol{\xi}_{\lambda}^K, (\rho_{\boldsymbol{u}}\mathbf{I}+\mathbf{G}_K)^{-1}\boldsymbol{\xi}_{\boldsymbol{\lambda}}^K\rangle - \langle\boldsymbol{\xi}_{\boldsymbol{\lambda}}^K, \boldsymbol{\alpha}_K\rangle \\ &-\rho_{\boldsymbol{u}}^{-1}\langle\boldsymbol{\xi}_{\boldsymbol{\lambda}}^K, (\mathbf{I}+\rho_{\boldsymbol{u}}^{-1}\mathbf{G}_K)^{-1}\bar{\boldsymbol{\theta}}_{ro}^K\rangle + 2\langle\boldsymbol{\lambda}_a,\boldsymbol{u}_a\rangle - 2\langle\boldsymbol{\lambda}_b,\boldsymbol{u}_b\rangle.\end{aligned}$$

Note now that $\rho_{\boldsymbol{u}}^{-1}(\mathbf{I}+\rho_{\boldsymbol{u}}^{-1}\mathbf{G}_K)^{-1}\bar{\boldsymbol{\theta}}_{ro}^K = (\rho_{\boldsymbol{u}}\mathbf{I}+\mathbf{G}_K)^{-1}\bar{\boldsymbol{\theta}}_{ro}^K = \boldsymbol{\alpha}_K$. Thus,

$$\begin{aligned}\hat{\varphi}_{DK}(\boldsymbol{\lambda}) &= -\rho_{\boldsymbol{u}}^{-1}\langle\boldsymbol{\lambda}_{ab},\boldsymbol{\lambda}_{ab}\rangle + \rho_{\boldsymbol{u}}^{-1}\langle\boldsymbol{\xi}_{\lambda}^K,(\rho_{\boldsymbol{u}}\mathbf{I}+\mathbf{G}_K)^{-1}\boldsymbol{\xi}_{\boldsymbol{\lambda}}^K\rangle - 2\langle\boldsymbol{\xi}_{\boldsymbol{\lambda}}^K,\boldsymbol{\alpha}_K\rangle \\ &\quad +2\langle\boldsymbol{\lambda}_a,\boldsymbol{u}_a\rangle - 2\langle\boldsymbol{\lambda}_b,\boldsymbol{u}_b\rangle.\end{aligned}$$

■