2010/34

Double smoothing technique for infinite-dimensional optimization problems with applications to optimal control

Olivier Devolder, François Glineur
and Yu. Nesterov

CORE

DISCUSSION PAPER

Center for Operations Research
and Econometrics

Voie du Roman Pays, 34
B-1348 Louvain-la-Neuve
Belgium
http://www.uclouvain.be/core

CORE DISCUSSION PAPER
2010/34

Double smoothing technique for infinite-dimensional optimization problems with applications to optimal control.

Olivier DEVOLDER [1], François GLINEUR[2]
and Yu. NESTEROV [3]

July 2010

Abstract
In this paper, we propose an efficient technique for solving some infinite-dimensional problems over the sets of functions of time. In our problem, besides the convex point-wise constraints on state variables, we have convex coupling constraints with finite-dimensional image. Hence, we can formulate a finite-dimensional dual problem, which can be solved by efficient gradient methods. We show that it is possible to reconstruct an approximate primal solution. In order to accelerate our schemes, we apply double-smoothing technique. As a result, our method has complexity $O(1/\varepsilon \ln 1/\varepsilon)$ gradient iterations, where ε is the desired accuracy of the solution of the primal-dual problem. Our approach covers, in particular, the optimal control problems with trajectory governed by a system of ordinary differential equations. The additional requirement could be that the trajectory crosses in certain moments of time some convex sets.

Keywords: convex optimization, optimal control, fast gradient methods, complexity bounds, smoothing technique.

---

[1] Université catholique de Louvain, CORE, B-1348 Louvain-la-Neuve, Belgium. E-mail: Olivier.devolder@uclouvain.be.
[2] Université catholique de Louvain, CORE, B-1348 Louvain-la-Neuve, Belgium. E-mail: francois.glineur@uclouvain.be. This author is also member of ECORE, the association between CORE and ECARES.
[3] Université catholique de Louvain, CORE, B-1348 Louvain-la-Neuve, Belgium. E-mail: yurii.nesterov@uclouvain.be. This author is also member of ECORE, the association between CORE and ECARES.

The first author is a F.R.S.-FNRS Research Fellow.

This paper presents research results of the Belgian Program on Interuniversity Poles of Attraction initiated by the Belgian State, Prime Minister's Office, Science Policy Programming. The scientific responsibility is assumed by the authors.

1 Introduction

In this paper, we are interested in a specific class of convex infinite-dimensional problems with decision variables being functions of time. These problems are characterized by the joint presence of convex bounds on some finite-dimensional characteristics of the decision variables, and of the point-wise time constraints. The key assumption on the problem structure is that, when the coupling constraints are dropped, we can easily optimize the remaining part of the problem separately, for each moment of time. Hence, the first step in our approach is dualization of the difficult convex constraints. Since the number of coupling constraints is finite, the Lagrangian dual problem is a non-smooth convex problem in finite-dimension. We assume that the dual objective function value can be computed for each value of the Lagrangian multipliers. Thus, our primary goal is to find efficiently an approximate solution to the dual problem. At the same time, we are able to reconstruct a nearly feasible optimal primal solution.

In order to satisfy these two goals, we develop a new double-smoothing approach, which is a variant of the smoothing techniques [10, 11, 12]. Using the problem structure, we transform the dual objective function into a smooth strongly convex function with Lipschitz continuous gradient. These modifications allow us to minimize the dual function by an optimal gradient scheme with complexity bound $O\left(\frac{1}{\epsilon}\ln\left(\frac{1}{\epsilon}\right)\right)$iterations, where ϵ is the desired accuracy. We present some applications of our technique to some optimal control problems and to large-scale problems in finite dimension.

The structure of this paper is as follows. In Section 2, we describe the infinite-dimensional primal problem setting and derive the corresponding dual problem. By Danskin theorem, we show that in general the dual objective function is non-smooth. In Section 3, we apply to this function two regularizations which make it smooth and strongly convex (we explain the importance of both properties). In Section 4, we recall the optimal method [9] for smooth and strongly convex functions and describe its rate of convergence. In Section 5, this optimal scheme is applied to our modified dual objective function. From the dual minimization sequence, it is possible to reconstruct a nearly feasible and optimal primal solution. The accuracy of the primal and dual solutions can be adjusted by special parameters. In Section 6, we show that the approximate primal solutions obtained by the double smoothing algorithm converge in a weak sense to the optimal solution of the primal infinite-dimensional problem. This result can be used for a constructive proof of the strong duality for the primal-dual problem. In the last two sections we consider the applications of double-smoothing technique to the optimal control problems and to the large-scale convex optimization problems in finite dimension.

2 Problem formulation and dual approach

Consider the following infinite-dimensional convex optimization problem:

$$\begin{aligned} \inf_u \{ \int_0^T F(t,u(t))dt : \quad & \int_0^T A_i(t)u(t)dt \in Q_i, i = 1,\dots,N, \qquad (*) \\ & u(t) \in Q(t) \text{ a.e in } [0,T]\}, \end{aligned} \tag{1}$$

where $T < \infty$, and the following assumptions are satisfied.

Assumption 1

1. *For matrices* $A_i(t) \in \mathbb{R}^{n_i \times m}$, $t \in [0,T]$, *we have* $\int\limits_0^T \|A_i(t)\|_2^2 \, dt < \infty$, $i = 1, \ldots, N$.
2. *Sets* $Q_i \subset \mathbb{R}^{n_i}$, $i = 1, \ldots, N$, *are convex, closed and bounded.*
3. *All sets* $Q(t) \subset \mathbb{R}^m$, $t \in [0,T]$, *are convex, closed and the graph* $Q := \cup_{t \in [0,T]} Q(t)$ *is bounded.*
4. *Function* $F(t,u) : [0,T] \times Q \to \mathbb{R}$ *is convex in* u *for any* $t \in [0,T]$, *and continuously differentiable in* (t,u).

We measure the size of the control function $u(t) \in \mathbb{R}^m$, $t \in [0,T]$, belonging to $L^2([0,T], \mathbb{R}^m)$, by the standard L_2-norm $\|u\|_2^2 = \int\limits_0^T \|u(t)\|_2^2 \, dt$.

Without convex coupling constraints $(1)^*$, we can solve problem (1) in a pointwise way, minimizing the objective separately for every $t \in [0,T]$. Hence, it is natural to dualize these constraints and pass to a finite-dimensional Lagrangian dual problem. For each value of dual variables, the dual function can be computed by a point-wise minimization in $u(t) \in Q(t)$. Assuming that this operation is feasible, our primary goal is to show that the dual problem can be solved efficiently. After that, we will see that the dual optimization scheme can be used for constructing an approximate optimal primal solution.

Denote by $\mathcal{A}_i$, the linear operators defining the convex coupling constraints:

$$\mathcal{A}_i : L^2([0,T], \mathbb{R}^m) \to \mathbb{R}^{n_i}, \quad u \to \int\limits_0^T A_i(t) u(t) dt, \quad i = 1, \ldots, N.$$

For any $z_i \in R^{n_i}$ we have $\langle \mathcal{A}_i u, z_i \rangle = \langle \int\limits_0^T A_i(t) u(t) dt, z_i \rangle = \int\limits_0^T \langle u(t), A_i^T(t) z_i \rangle dt$. Thus,

$$\mathcal{A}^* z_i \;\; = \;\; A_i^T(t) z_i, \quad t \in [0,T]. \tag{2}$$

In view of Assumption 1.1, the operators $\mathcal{A}_i$ are bounded and therefore continuous:

$$\begin{aligned} \|\mathcal{A}_i u\|_2^2 \;\; &= \;\; \| \int\limits_0^T A_i(t) u(t) dt \|_2^2 \;\le\; \left(\int\limits_0^T \|A_i(t) u(t)\|_2 \, dt \right)^2 \\ &\le \;\; \left(\int\limits_0^T \|A_i(t)\|_2 \cdot \|u(t)\|_2 \, dt \right)^2 \;\le \int\limits_0^T \|A_i(t)\|_2^2 \, dt \cdot \int\limits_0^T \|u(t)\|_2^2 \, dt \;<\; +\infty. \end{aligned}$$

Since Q_i is a convex set, inclusion $\mathcal{A}_i u \in Q_i$ is valid if and only if

$$\langle \mathcal{A}_i u, z^i \rangle \le \sigma_{Q_i}(z^i) \quad \forall z^i \in \mathbb{R}^{n_i}$$

where $\sigma_{Q_i}(z) = \sup_{x \in Q_i} \langle x, z \rangle$ is the support function of Q_i. Therefore, problem (1) can be rewritten in the following form:

$$\begin{aligned} \inf_{u \in L^2([0,T], \mathbb{R}^m)} \{ \int\limits_0^T F(t, u(t)) dt : \;\; & \langle \mathcal{A}_i u, z^i \rangle \le \sigma_{Q_i}(z^i) \quad \forall z^i \in \mathbb{R}^{n_i}, \; i = 1, .., N, \\ & u(t) \in Q(t) \text{ a.e. in } [0,T] \}. \end{aligned} \tag{3}$$

Denote $\overline{U} = \{u \in L^2([0,T],\mathbb{R}^m) : u(t) \in Q(t) \text{ a.e. in } [0,T]\}$, $\bar{n} = \sum\limits_{i=1}^{N} n_i$, and $z = (z^1,\ldots,z^N)^T \in \mathbb{R}^{\bar{n}}$. Dualizing the constraints $\mathcal{A}_i u \in Q_i$, we obtain a dual form of the primal problem:

$$\begin{aligned} P^* &= \inf_{u\in\overline{U}} \left[\int_0^T F(t,u(t))dt + \sup_{z\in\mathbb{R}^{\bar{n}}} \textstyle\sum_{i=1}^N (\langle \mathcal{A}_i u, z^i\rangle - \sigma_{Q_i}(z^i))\right] \geq \\ D^* &\stackrel{\text{def}}{=} \sup_{z\in\mathbb{R}^{\bar{n}}} \left[-\textstyle\sum_{i=1}^N \sigma_{Q_i}(z^i) + \inf\limits_{u\in\overline{U}} \left(\textstyle\sum_{i=1}^N \langle \mathcal{A}_i u, z^i\rangle + \int_0^T F(t,u(t))dt\right)\right]. \end{aligned}$$

Thus, the Lagrangian dual problem (in minimization form) is given by

$$-D^* = \theta^* \stackrel{\text{def}}{=} \inf_{z\in\mathbb{R}^{\bar{n}}} \textstyle\sum_{i=1}^N \sigma_{Q_i}(z^i) + \phi(z) = \inf_{z\in\mathbb{R}^{\bar{n}}} \theta(z) \tag{4}$$

where $\phi : \mathbb{R}^{\bar{n}} \to \mathbb{R}$ is defined by:

$$\begin{aligned} \phi(z) &= \sup_{u\in\overline{U}} [\int_0^T -F(t,u(t))dt - \textstyle\sum_{i=1}^N \langle \mathcal{A}_i u, z^i\rangle] \\ &= \sup_{u\in\overline{U}} [\int_0^T -F(t,u(t))dt - \textstyle\sum_{i=1}^N \langle u, \mathcal{A}_i^* z^i\rangle] \\ &\stackrel{(2)}{=} \sup_{u\in\overline{U}} \int_0^T [-F(t,u(t)) - \textstyle\sum_{i=1}^N \langle u(t), A_i(t)^T z^i\rangle] dt. \end{aligned} \tag{5}$$

The dual convex optimization problem (4) is an unconstrained problem in finite dimension. For each $z \in \mathbb{R}^{\bar{n}}$ we can compute its objective function defining $\phi(z)$ in a pointwise way:

$$u(t) = \arg\max_{v\in Q(t)} \left[-F(t,v) - \textstyle\sum_{i=1}^N \langle v, A_i(t)^T z^i\rangle\right], \quad t \in [0,T]. \tag{6}$$

(We assume that functions $F(t,v)$ and convex sets $Q(t)$ are simple enough.) In general, this function $\theta(z)$ is nondifferentiable. Indeed, by standard reasoning, we can guarantee that

$$\partial\phi(z) \supseteq \{(-\mathcal{A}_1 u, \ldots, -\mathcal{A}_N u)^T : \int_0^T -F(t,u(t))dt - \textstyle\sum_{i=1}^N \langle \mathcal{A}_i u, z^i\rangle = \phi(z), u \in \overline{U}\}.$$

(A rigorous application of Danskin Theorem [5, 3] justifies the equality in the above relation.) As the optimization problem (5) can have multiple optimal solutions, the set $\partial\phi(z)$ can contain more than one element and therefore at this point ϕ is not differentiable.

Thus, the dualization of problem (1) results in a finite-dimensional nonsmooth convex problem (4). The classical algorithms for solving such problems are the subgradient-type schemes. Provided that $\theta(z)$ is computable, we can apply them directly to problem (4). However, their convergence is rather slow. In order to get an accuracy ϵ for the objective value, they need $O\left(\frac{1}{\epsilon^2}\right)$ iterations (e.g. [9, 13]).

In our paper we propose another approach based on the smoothing techniques [10, 11, 12]. In the smoothing approach, using the specific structure of the problem, we apply some regularization to the objective function and obtain much faster methods (which are not anymore the pure black-box schemes). In this work, we develop an algorithm which is able to solve the dual problem and to reconstruct from the nearly optimal dual solution, a nearly optimal and feasible primal solution in $O\left(\frac{1}{\epsilon}\ln\left(\frac{1}{\epsilon}\right)\right)$ iterations.

3 Double Smoothing Technique

In convex optimization there are two important class of objective function:

- $F_L^{1,1}(\mathbb{R}^{\bar{n}})$ is the class of convex functions $f : \mathbb{R}^{\bar{n}} \to \mathbb{R}$ which gradient is Lipschitz-continuous with constant $L > 0$.
- $S_{\kappa,L}^{1,1}(\mathbb{R}^{\bar{n}})$, is the class of functions $f \in F_L^{1,1}(\mathbb{R}^{\bar{n}})$ which are strongly convex with parameter $\kappa > 0$.

We will try to solve the dual problem (4) using a new primal-dual smoothing technique. Note that in general its objective function is not differentiable and not strongly convex. However, we can ensure these properties by double primal-dual regularization of θ. The goal of the first regularization is to obtain an objective function with Lipschitz-continuous gradient. In this case, we will be able to apply much more efficient algorithms of smooth convex optimization.

The goal of the second regularization is to obtain a strongly convex dual objective. As we will see later, for reconstructing primal solution, we need to get a dual solution with small value of the gradient of the objective function. Unfortunately, this feature is not guaranteed by the small residual in the objective function. Indeed, consider the unconstrained minimization problem

$$\min_{z\in\mathbb{R}^{\bar{n}}} g(z),$$

with $g \in F_L^{1,1}(\mathbb{R}^{\bar{n}})$ and with optimal solution z_*. If we apply to this problem the optimal scheme [10], we can obtain the following rate of convergence:

$$g(z_k) - g^* \leq O\left(\frac{1}{k^2}\right).$$

Since (e.g. Theorem 2.1.5 in [9]) $\frac{1}{2L}\|\nabla g(z_k)\|_2^2 \leq g(z_k) - g^*$, we have $\|\nabla g(z_k)\|_2 \leq O\left(\frac{1}{k}\right)$.

On the other hand, we can consider the modified function $\tilde{g}(z) = g(z) + \frac{1}{2}\delta\|z\|_2^2$. Note that $\tilde{g} \in S_{\delta,\delta+L}^{1,1}(\mathbb{R}^{\bar{n}})$. Therefore (e.g. Theorem 2.2.3 in [9])

$$\frac{1}{2(\delta+L)}\|\nabla\tilde{g}(z_k)\|_2^2 \quad\leq\quad \tilde{g}(z_k) - \tilde{g}^* \;\leq\; \exp\left(-k\sqrt{\frac{\delta}{\delta+L}}\right)\cdot\frac{\delta+L}{2}\|z_*\|_2^2.$$

Thus, we can get $\|\nabla g(z_k)\|_2 \leq \delta$ in $O\left(\frac{1}{\delta^{1/2}}\ln\frac{1}{\delta}\right)$ iterations.

These results justify why we want to modify the dual objective function in a strongly convex function with Lipschitz-continuous gradient. Let us start from ensuring the smoothness of the dual function.

The dual objective $\theta(z)$ is a sum of two functions. Both of them can be nonsmooth. Consider its first term. For $\rho > 0$, we can approximate $\sigma_{Q_i}(z^i) = \sup_{x\in Q_i}\langle x, z^i\rangle$ by a smooth function

$$\sigma_{\rho,Q_i}(z^i) \quad=\quad \sup_{x\in Q_i}\{\langle x, z^i\rangle - \frac{\rho}{2}\|x\|_2^2\}. \tag{7}$$

This optimization problem has only one optimal solution since its objective is strongly concave. Therefore the function σ_{ρ,Q_i} is differentiable with gradient given by:

$$\nabla_{z^i}\sigma_{\rho,Q_i}(z^i) = x_{\rho,z^i}$$

where $x_{\rho,z^i} \in Q_i$ is this unique optimal solution to (7). Moreover, $\nabla\sigma_{Q_i}(z^i)$ is Lipschitz-continuous with constant $\frac{1}{\rho}$ (e.g. [10]). Applying this smoothing to all $i = 1, ..., N$, we obtain function $\sum_{i=1}^{N} \sigma_{\rho,Q_i}(z^i)$, which gradient

$$\nabla_z \left(\sum_{i=1}^{N} \sigma_{\rho,Q_i}(z^i)\right) \quad = \quad (x_{\rho,z_1}, ..., x_{\rho,z_N})^T.$$

is Lipschitz-continuous with constant $\frac{1}{\rho}$.

Let us smooth now the second term of the dual objective. For $\mu > 0$, we modify the function $\phi(z)$ as follows:

$$\phi_\mu(z) \quad = \quad \sup_{u\in\overline{U}} \left(\int_0^T -F(t, u(t))dt - \sum_{i=1}^{N}\langle\mathcal{A}_i u, z^i\rangle - \frac{\mu}{2}\int_0^T \|u(t)\|_2^2\, dt\right).$$

Since the objective function of this problem is strongly concave in u, it has a unique optimal solution $u_{\mu,z}(t)$, that can be computed independently for each $t \in [0, T]$:

$$u_{\mu,z}(t) \quad = \quad \arg\max_{v\in Q(t)} \left[-F(t, v) - \sum_{i=1}^{N}\langle v, A_i(t)^T z^i\rangle - \frac{\mu}{2}\|v\|_2^2\right]. \qquad (8)$$

By Danskin Theorem, function $\phi_\mu(z)$ is differentiable and

$$\nabla\phi_\mu(z) \quad = \quad (-\mathcal{A}_1 u_{\mu,z}, ..., -\mathcal{A}_N u_{\mu,z})^T.$$

Let us estimate the Lipschitz constant of its gradient.

Theorem 1 *With the assumption 1, the gradient of function ϕ_μ is Lipschitz-continuous with constant $\frac{1}{\mu}\sum_{i=1}^{N}\|\mathcal{A}_i\|^2$.*

Proof:
For $(u, z) \in \overline{U} \times \mathbb{R}^{\bar{n}}$, define the function

$$\Psi_\mu(u, z) \quad = \quad -\int_0^T F(t, u(t))dt - \sum_{i=1}^{N}\langle\mathcal{A}_i u, z^i\rangle - \frac{\mu}{2}\int_0^T \|u(t)\|_2^2\, dt.$$

This function is Frechet differentiable in u and its Frechet derivative is given by:

$$\langle\nabla_u\Psi_\mu(u, z), h\rangle \quad = \quad \int_0^T [-\langle\nabla_u F(t, u(t)), h(t)\rangle - \mu\langle u(t), h(t)\rangle]\, dt - \sum_{i=1}^{N}\langle\mathcal{A}_i^* z^i, h\rangle$$

(we used the fact that F is continuously differentiable). Therefore, by the first-order optimality conditions for problem

$$\phi_\mu(z) = \sup_{u\in\overline{U}} \Psi_\mu(u, z),$$

for all z' and $z'' \in \mathbb{R}^{\bar{n}}$ we have (see [8]):

$$\langle \nabla_u \Psi_\mu(u_{\mu,z'}, z'), u_{\mu,z''} - u_{\mu,z'} \rangle \leq 0, \quad \langle \nabla_u \Psi_\mu(u_{\mu,z''}, z''), u_{\mu,z'} - u_{\mu,z''} \rangle \leq 0.$$

Summing up these two inequalities, we obtain:

$$\begin{aligned} 0 &\leq \langle \nabla_u \Psi_\mu(u_{\mu,z'}, z') - \nabla_u \Psi_\mu(u_{\mu,z''}, z''), u_{\mu,z'} - u_{\mu,z''} \rangle \\ &= \int_0^T \langle -\nabla_u F(t, u_{\mu,z'}(t)) + \nabla_u F(t, u_{\mu,z''}(t)), u_{\mu,z'}(t) - u_{\mu,z''}(t) \rangle dt \\ &\quad -\mu \int_0^T \langle u_{\mu,z'}(t) - u_{\mu,z''}(t), u_{\mu,z'}(t) - u_{\mu,z''}(t) \rangle dt - \textstyle\sum_{i=1}^N \langle \mathcal{A}_i^*(z_i' - z_i''), u_{\mu,z'} - u_{\mu,z''} \rangle. \end{aligned}$$

Since $F(t,.)$ is convex, $\langle \nabla_u F(t, u_{\mu,z'}(t)) - \nabla_u F(t, u_{\mu,z''}(t)), u_{\mu,z'}(t) - u_{\mu,z''}(t) \rangle \geq 0$ for all $t \in [0,T]$. Hence, we conclude that $-\sum_{i=1}^N \langle z_i' - z_i'', \mathcal{A}_i(u_{\mu,z'} - u_{\mu,z''}) \rangle \geq \mu \left\| u_{\mu,z'} - u_{\mu,z''} \right\|_2^2$. Therefore:

$$\begin{aligned} \|\nabla \phi_\mu(z') - \nabla \phi\mu(z'')\|_2^2 &= \sum_{i=1}^N \left\| \mathcal{A}_i(u_{\mu,z'}) - \mathcal{A}_i(u_{\mu,z''}) \right\|_2^2 \leq \sum_{i=1}^N \|\mathcal{A}_i\|_2^2 \cdot \|u_{\mu,z'} - u_{\mu,z''}\|_2^2 \\ &\leq -\tfrac{1}{\mu} \sum_{i=1}^N \|\mathcal{A}_i\|_2^2 \cdot \sum_{i=1}^N \langle z_i' - z_i'', \mathcal{A}_i(u_{\mu,z'} - u_{\mu,z''}) \rangle \\ &\leq \tfrac{1}{\mu} \sum_{i=1}^N \|\mathcal{A}_i\|_2^2 \cdot \sum_{i=1}^N \|z_i' - z_i''\|_2 \cdot \|\mathcal{A}_i(u_{\mu,z'} - u_{\mu,z''})\|_2 \\ &\leq \tfrac{1}{\mu} \sum_{i=1}^N \|\mathcal{A}_i\|_2^2 \cdot \|z' - z''\|_2 \cdot \left[\sum_{i=1}^N \|\mathcal{A}_i(u_{\mu,z'}) - \mathcal{A}_i(u_{\mu,z''})\|_2^2 \right]^{1/2}. \end{aligned}$$

and we conclude that $\|\nabla \phi_\mu(z') - \nabla \phi_\mu(z'')\|_2 \leq \frac{1}{\mu} \sum_{i=1}^N \|\mathcal{A}_i\|_2^2 \cdot \|z' - z''\|_2$. □

Denote $D_i = \max\{\frac{1}{2}\|x\|_2^2 : x \in Q_i\}$ and $D = \max\{\frac{1}{2}\|u\|_2^2 : u \in \overline{U}\}$. Concerning the value of the modified dual objective function, we have:

$$\sigma_{\rho,Q_i}(z^i) \leq \sigma_{Q_i}(z^i) \leq \sigma_{\rho,Q_i}(z^i) + \rho D_i, \quad \forall z^i \in \mathbb{R}^{n_i},$$

$$\phi_\mu(z) \leq \phi(z) \leq \phi_\mu(z) + \mu D, \qquad \forall z \in \mathbb{R}^{\bar{n}}.$$

Therefore, if we define the function $\theta_{\rho,\mu}(z) = \sum_{i=1}^N \sigma_{\rho,Q_i}(z^i) + \phi_\mu(z)$, then

$$\theta_{\rho,\mu}(z) \leq \theta(z) \leq \theta_{\rho,\mu}(z) + \mu D + \rho \hat{D}, \quad \forall z \in \mathbb{R}^{\bar{n}}, \quad \hat{D} \stackrel{\text{def}}{=} \textstyle\sum_{i=1}^N D_i. \tag{9}$$

Finally, in order to obtain a strongly convex dual objective function, we just add the strongly convex function $\frac{1}{2}\|z\|_2^2$ with coefficient $\kappa > 0$ to function $\theta_{\rho,\mu}$. This gives us a new dual objective function:

$$\theta_{\rho,\mu,\kappa}(z) = \sum_{i=1}^N \sigma_{\rho,Q_i}(z^i) + \phi_\mu(z) + \tfrac{\kappa}{2}\|z\|_2^2,$$

which is strongly convex with parameter κ, and which gradient

$$\nabla \theta_{\rho,\mu,\kappa}(z) = (x_{\rho,z_1}, ..., x_{\rho,z_N})^T - (\mathcal{A}_1 u_{\mu,z}, ..., \mathcal{A}_N u_{\mu,z})^T + \kappa z.$$

is Lipschitz-continuous with constant $L(\rho,\mu,\kappa) = \frac{1}{\rho} + \frac{1}{\mu}\sum_{i=1}^N \|\mathcal{A}_i\|_2^2 + \kappa$. This function can be minimized by the optimal method for the class $S_{\kappa,L(\rho,\mu,\kappa)}^{1,1}(\mathbb{R}^{\bar{n}})$.

4 Optimal scheme for $S^{1,1}_{\kappa,L}(\mathbb{R}^{\bar{n}})$.

For the reader convenience, in this section we present the simplest optimal method for minimizing smooth strongly convex functions.

Let function $g : \mathbb{R}^{\bar{n}} \to \mathbb{R}$ be strongly convex with parameter $\kappa > 0$ and its gradient be Lipschitz-continuous with constant $L > \kappa$. Consider the following problem:

$$\min_{y \in \mathbb{R}^{\bar{n}}} g(y). \tag{10}$$

We assume that this problem is solvable. Denote by g^* its optimal value and by y^* the optimal solution.

$$\begin{array}{ll} \textbf{Algorithm } ([9]): & \text{Choose } w_0 = y_0 \in \mathbb{R}^{\bar{n}}. \\[2ex] \underline{\text{Iteration } (k \geq 0):} & \text{Set } y_{k+1} = w_k - \frac{1}{L}\nabla g(w_k), \text{ and} \\[2ex] & w_{k+1} = y_{k+1} + \frac{\sqrt{L}-\sqrt{\kappa}}{\sqrt{L}+\sqrt{\kappa}}(y_{k+1} - y_k). \end{array} \tag{11}$$

By Theorem 2.2.3 in [9] we have:

$$\begin{aligned} g(y_k) - g^* &\leq \left(g(y_0) - g^* + \tfrac{\kappa}{2}\|y_0 - y^*\|_2^2\right) e^{-k\sqrt{\frac{\kappa}{L}}} \\ &\leq 2(g(y_0) - g^*)e^{-k\sqrt{\frac{\kappa}{L}}}. \end{aligned} \tag{12}$$

Since ∇g is Lipschitz-continuous, in view of Theorem 2.1.5 in [9] we have

$$\tfrac{1}{2L}\|\nabla g(y_k)\|_2^2 \leq g(y_k) - g^* \overset{(12)}{\leq} 2(g(y_0) - g^*)e^{-k\sqrt{\frac{\kappa}{L}}}.$$

Therefore,

$$\|\nabla g(y_k)\|_2^2 \leq 4L(g(y_0) - g^*)e^{-k\sqrt{\frac{\kappa}{L}}}. \tag{13}$$

Finally, since g is strongly convex, by Theorem 2.1.8 in [9] we have:

$$\tfrac{\kappa}{2}\|y_k - y^*\|_2^2 \leq g(y_k) - g^* \overset{(12)}{\leq} 2(g(y_0) - g^*)e^{-k\sqrt{\frac{\kappa}{L}}}.$$

Using this inequality and additional arguments, we conclude that

$$\|y_k - y^*\|_2^2 \leq \min\left\{\|y_0 - y^*\|_2^2, \tfrac{4}{\kappa}(g(y_0) - g^*)e^{-k\sqrt{\frac{\kappa}{L}}}\right\}. \tag{14}$$

5 Solving primal-dual problem by optimal method

Denote by z^* the unique optimal solution of the problem

$$\min_{z \in \mathbb{R}^{\bar{n}}} \theta_{\rho,\mu,\kappa}(z), \tag{15}$$

and by z^{**} one of the optimal solutions of the dual problem (4). We assume that the upper bound

$$\|z^{**}\|_2^2 \;\leq\; R \tag{16}$$

is available.

Applying to this problem the method (11) with starting point $z_0 = 0$, we obtain a sequence $\{z_k\}$ such that:

$$\begin{aligned}
\theta_{\rho,\mu,\kappa}(z_k) - \theta_{\rho,\mu,\kappa}(z^*) \;&\leq\; 2(\theta_{\rho,\mu,\kappa}(0) - \theta_{\rho,\mu,\kappa}(z^*))e^{-k\sqrt{\frac{\kappa}{L(\rho,\mu,\kappa)}}},\\
\|\nabla\theta_{\rho,\mu,\kappa}(z_k)\|_2^2 \;&\leq\; 4L(\rho,\mu,\kappa)(\theta_{\rho,\mu,\kappa}(0) - \theta_{\rho,\mu,\kappa}(z^*))e^{-k\sqrt{\frac{\kappa}{L(\rho,\mu,\kappa)}}},\\
\|z_k - z^*\|_2^2 \;&\leq\; \min\left\{\|z^*\|_2^2, \tfrac{4}{\kappa}(\theta_{\rho,\mu,\kappa}(0) - \theta_{\rho,\mu,\kappa}(z^*))e^{-k\sqrt{\frac{\kappa}{L(\rho,\mu,\kappa)}}}\right\}.
\end{aligned} \tag{17}$$

5.1 Convergence of $\theta(z_k)$ to θ^*

Since $\theta_{\rho,\mu,\kappa}(0) = \theta_{\rho,\mu}(0)$ and $\theta_{\rho,\mu,\kappa}(z^*) = \theta_{\rho,\mu}(z^*) + \frac{\kappa}{2}\,\|z^*\|_2^2$, we have

$$\begin{aligned}
\tfrac{\kappa}{2}\,\|z^*\|_2^2 \;&\leq\; \theta_{\rho,\mu,\kappa}(0) - \theta_{\rho,\mu,\kappa}(z^*) \;=\; \theta_{\rho,\mu}(0) - \theta_{\rho,\mu}(z^*) - \tfrac{\kappa}{2}\,\|z^*\|_2^2\,,\\
\|z_k - z^*\|_2^2 \;&\overset{(12)}{\leq}\; \tfrac{2}{\kappa}(\theta_{\rho,\mu}(0) - \theta_{\rho,\mu}(z^*))e^{-k\sqrt{\frac{\kappa}{L(\rho,\mu,\kappa)}}}.
\end{aligned} \tag{18}$$

Note that

$$\theta_{\rho,\mu}(z_k) - \theta_{\rho,\mu}(z^*) \overset{(12)}{\leq} (\theta_{\rho,\mu}(0) - \theta_{\rho,\mu}(z^*))e^{-k\sqrt{\frac{\kappa}{L(\rho,\mu,\kappa)}}} + \tfrac{\kappa}{2}(\|z^*\|_2^2 - \|z_k\|_2^2).$$

On the other hand,

$$\begin{aligned}
\|z^*\|_2^2 - \|z_k\|_2^2 \;&\leq\; \|z^* - z_k\|_2(\|z^*\|_2 + \|z_k\|_2)\\
&\leq\; \|z^* - z_k\|_2(2\|z^*\|_2 + \|z_k - z^*\|_2)\\
&\overset{(14)}{\leq}\; 3\|z^* - z_k\|_2 \cdot \|z^*\|_2\\
&\overset{(18)}{\leq}\; 3 \cdot \|z^*\|_2\sqrt{\tfrac{2}{\kappa}(\theta_{\rho,\mu}(0) - \theta_{\rho,\mu}(z^*))}e^{-\frac{k}{2}\sqrt{\frac{\kappa}{L(\rho,\mu,\kappa)}}}\\
&\overset{(18)}{\leq}\; \tfrac{3\sqrt{2}}{\kappa}(\theta_{\rho,\mu}(0) - \theta_{\rho,\mu}(z^*))e^{-\frac{k}{2}\sqrt{\frac{\kappa}{L(\rho,\mu,\kappa)}}},
\end{aligned}$$

and therefore (since $1 + \sqrt{\frac{3}{2}} < \frac{25}{8}$)

$$\theta_{\rho,\mu}(z_k) - \theta_{\rho,\mu}(z^*) \;\leq\; \tfrac{25}{8}(\theta_{\rho,\mu}(0) - \theta_{\rho,\mu}(z^*))e^{-\frac{k}{2}\sqrt{\frac{\kappa}{L(\rho,\mu,\kappa)}}}.$$

We also have $\theta_{\rho,\mu}(0) \leq \theta(0)$ and

$$\theta_{\rho,\mu}(z^*) \;\geq\; \theta(z^*) - \rho\hat{D} - \mu D \;\geq\; \theta(z^{**}) - \rho\hat{D} - \mu D.$$

Therefore:

$$\theta_{\rho,\mu}(0) - \theta_{\rho,\mu}(z^*) \;\leq\; \theta(0) - \theta(z^{**}) + \rho\hat{D} + \mu D. \tag{19}$$

Finally, since $\theta_{\rho,\mu}(z^*) + \frac{\kappa}{2}\|z^*\|_2^2 \leq \theta_{\rho,\mu}(z^{**}) + \frac{\kappa}{2}\|z^{**}\|_2^2$, we have:

$$\theta_{\rho,\mu}(z^*) \;\leq\; \theta_{\rho,\mu}(z^{**}) + \tfrac{\kappa}{2}\|z^{**}\|_2^2 \overset{(9)}{\leq} \theta(z^{**}) + \tfrac{\kappa}{2}\|z^{**}\|_2^2,$$

and therefore

$$\theta_{\rho,\mu}(z_k) - \theta_{\rho,\mu}(z^*) \overset{(9)}{\geq} \theta(z_k) - \mu D - \rho\hat{D} - \theta(z^{**}) - \tfrac{\kappa}{2}\|z^{**}\|_2^2.$$

In conclusion, we have:

$$\begin{aligned} \theta(z_k) - \theta(z^{**}) \;\leq\; & \mu D + \rho\hat{D} + \tfrac{\kappa}{2}R^2 \\ & + \tfrac{25}{8}\left(\theta(0) - \theta(z^{**}) + \rho\hat{D} + \mu D\right) e^{-\frac{k}{2}\sqrt{\frac{\kappa}{L(\rho,\mu,\kappa)}}}. \end{aligned} \tag{20}$$

Now it is clear how to choose the smoothing parameters. Let us fix some $\epsilon > 0$. In the upper bound for the residual $\theta(z_k) - \theta(z^{**})$, we have four terms. In order to ensure accuracy $\theta(z_k) - \theta(z^{**}) \leq \epsilon$, we force all of these terms to be less or equal than $\frac{\epsilon}{4}$. This leads to the following values:

$$\mu = \mu(\epsilon) = \tfrac{\epsilon}{4D}, \quad \rho = \rho(\epsilon) = \tfrac{\epsilon}{4\hat{D}}, \quad \kappa = \kappa(\epsilon) = \tfrac{\epsilon}{2R^2}. \tag{21}$$

With this choice we get

$$\theta(z_k) - \theta(z^{**}) \;\leq\; \tfrac{3\epsilon}{4} + \tfrac{25}{8}\left(\theta(0) - \theta(z^{**}) + \tfrac{\epsilon}{2}\right) e^{-\frac{k}{2}\sqrt{\frac{\kappa}{L(\rho,\mu,\kappa)}}}. \tag{22}$$

The last term in the estimate (22) defines the number of iterations needed for getting the accuracy ϵ. Clearly, we ensure

$$\tfrac{25}{8}\left(\theta(0) - \theta(z^{**}) + \tfrac{\epsilon}{2}\right) e^{-\frac{k}{2}\sqrt{\frac{\kappa}{L(\rho,\mu,\kappa)}}} \;\leq\; \tfrac{\epsilon}{4}$$

by taking

$$k \;\geq\; \sqrt{\tfrac{L(\rho,\mu,\kappa)}{\kappa}} \ln \frac{25\left(\theta(0) - \theta(z^{**}) + \frac{\epsilon}{2}\right)}{2\epsilon}. \tag{23}$$

It remains to note that

$$\begin{aligned} \tfrac{L(\rho,\mu,\kappa)}{\kappa} \;&=\; 1 + \tfrac{1}{\rho\kappa} + \tfrac{1}{\mu\kappa}\sum_{i=1}^{N}\|\mathcal{A}_i\|_2^2 \\ &\overset{(21)}{=}\; 1 + \tfrac{8}{\epsilon^2}\left[\hat{D} + D\sum_{i=1}^{N}\|\mathcal{A}_i\|_2^2\right] R^2. \end{aligned} \tag{24}$$

Thus, we need at most $k = O(\frac{1}{\epsilon}\ln\frac{1}{\epsilon})$ iterations.

5.2 Convergence of $\|\nabla\theta_{\rho,\mu}(z_k)\|_2$

In our approach, we are able to reconstruct a nearly optimal and feasible primal solution. In Section 5.3, we will see that the accuracy of this solution depends not only on the convergence rate of objective, but also on the rate of convergence of the norm of the gradient. Let us give an upper bound for the number of iterations needed to drop this norm below a certain level.

We have:

$$\begin{aligned}\|\nabla\theta_{\rho,\mu}(z_k)\|_2 &\leq \|\nabla\theta_{\rho,\mu,\kappa}(z_k)-\kappa z_k\|_2 \leq \|\nabla\theta_{\rho,\mu,\kappa}(z_k)\|_2+\kappa\|z_k\|_2\\ &\overset{(17)}{\leq} \|\nabla\theta_{\rho,\mu,\kappa}(z_k)\|_2+2\kappa\left\|z^*\right\|_2.\end{aligned}$$

Note that

$$\begin{aligned}\tfrac{1}{4L(\rho,\mu,\kappa)}\|\nabla\theta_{\rho,\mu,\kappa}(z_k)\|_2^2 &\overset{(17),(18)}{\leq} (\theta_{\rho,\mu}(0)-\theta_{\rho,\mu}(z^*))e^{-k\sqrt{\frac{\kappa}{L(\rho,\mu,\kappa)}}}\\ &\overset{(19)}{\leq} (\theta(0)-\theta(z^{**})+\mu D+\rho\hat{D})e^{-k\sqrt{\frac{\kappa}{L(\rho,\mu,\kappa)}}}\\ &\overset{(21)}{=} (\theta(0)-\theta(z^{**})+\tfrac{\epsilon}{2})e^{-k\sqrt{\frac{\kappa}{L(\rho,\mu,\kappa)}}}.\end{aligned}$$

At the same time,

$$\begin{aligned}\theta(z^{**})+\tfrac{\kappa}{2}\left\|z^{**}\right\|_2^2 &\overset{(9)}{\geq} \theta_{\rho,\mu}(z^{**})+\tfrac{\kappa}{2}\left\|z^{**}\right\|_2^2 \geq \theta_{\rho,\mu}(z^*)+\tfrac{\kappa}{2}\left\|z^*\right\|_2^2\\ &\overset{(9)}{\geq} \theta(z^*)-\mu D-\rho\hat{D}+\tfrac{\kappa}{2}\left\|z^*\right\|_2^2\\ &\geq \theta(z^{**})-\mu D-\rho\hat{D}+\tfrac{\kappa}{2}\left\|z^*\right\|_2^2.\end{aligned}$$

Hence,

$$\|z^*\|_2 \leq \sqrt{\left\|z^{**}\right\|_2^2+\tfrac{2\mu}{\kappa}D+\tfrac{2\rho}{\kappa}\hat{D}} \overset{(21)}{\leq} \kappa^{-1/2}\sqrt{\tfrac{3\epsilon}{2}} \overset{(21)}{=} \sqrt{3}R, \tag{25}$$

and we obtain:

$$\|\nabla\theta_{\rho,\mu}(z_k)\|_2 \leq \sqrt{4L(\rho,\mu,\kappa)(\theta(0)-\theta(z^{**})+\tfrac{\epsilon}{2})}e^{-\frac{k}{2}\sqrt{\frac{\kappa}{L(\rho,\mu,\kappa)}}}+2\sqrt{3}\kappa R.$$

Taking into account (21), we can see that in $k(\epsilon)=O(\frac{1}{\epsilon}\ln\frac{1}{\epsilon})$ iterations, we can ensure

$$\theta(z_k)-\theta(z^{**})\leq\epsilon,\quad \|\nabla\theta_{\rho,\mu}(z_k)\|_2\leq\tfrac{2\epsilon}{R}. \tag{26}$$

5.3 Constructing an approximate primal solution

In this section, given an accuracy $\epsilon>0$, we will see how to obtain from the dual iterate $z_{k(\epsilon)}$, an approximate primal solution $\widehat{u}_{k(\epsilon)}\in\overline{U}$ such that:

$$\left|\int_0^T F(t,\widehat{u}_{k(\epsilon)}(t))dt-D^*\right| \leq 2\left(1+2\sqrt{3}\right)\cdot\epsilon, \tag{27}$$

$$\left[\sum_{i=1}^N\left\|\mathcal{A}_i\widehat{u}_{k(\epsilon)}-\overline{x}_i\right\|_2^2\right]^{1/2} \leq \tfrac{2\epsilon}{R}, \tag{28}$$

where $\overline{x}_i\in Q_i$ for all $i=1,...,N$.

Since $D^*\leq P^*$, inequality (27) implies $\int_0^{t_n} F(t,\widehat{u}_{k(\epsilon)}(t))dt\leq P^*+C_1\epsilon$. Thus, the control function $\hat{u}_{k(\epsilon)}(t)$ satisfying (27), (28) can be seen as a nearly optimal and feasible primal solution with accuracy proportional to ϵ.

Consider $\widehat{u}_{k(\epsilon)} = u_{\mu(\epsilon), z_{k(\epsilon)}}$, the unique optimal solution of the corresponding problem (8). This solution can be obtained in a pointwise way since $\widehat{u}_{k(\epsilon)}(t)$ is defined almost everywhere as a unique optimal solution of the finite-dimensional strongly convex problem. We assume that, for all $t \in [0,T]$, the convex functions $u \to F(t,u)$ and the convex sets $Q(t)$ are simple enough for solving these point-wise problems analytically or very quickly using, for example, real-time embedded convex optimization.

We have:

$$\begin{aligned} \theta_{\rho(\epsilon),\mu(\epsilon)}(z_{k(\epsilon)}) &= \sum_{i=1}^{N} \sigma_{\rho(\epsilon),Q_i}(z^i_{k(\epsilon)}) + \phi_\mu(z_{k(\epsilon)}) \\ &= \sum_{i=1}^{N} \left(\langle x_{\rho(\epsilon),z^i_{k(\epsilon)}}, z^i_{k(\epsilon)} \rangle - \frac{\rho(\epsilon)}{2} \left\| x_{\rho(\epsilon),z^i_{k(\epsilon)}} \right\|_2^2 \right) - \int_0^T F(t, \widehat{u}_{k(\epsilon)}(t))dt \\ &\quad - \sum_{i=1}^{N} \langle \mathcal{A}_i \widehat{u}_{k(\epsilon)}, z^i_{k(\epsilon)} \rangle - \frac{\mu(\epsilon)}{2} \left\| \widehat{u}_{k(\epsilon)} \right\|_2^2 . \end{aligned}$$

Therefore,

$$\begin{aligned} \int_0^T F(t, \widehat{u}_{k(\epsilon)}(t))dt - D^* &= \textstyle\sum_{i=1}^{N} \langle x_{\rho(\epsilon),z^i_{k(\epsilon)}} - \mathcal{A}_i \widehat{u}_{k(\epsilon)}, z^i_{k(\epsilon)} \rangle - \frac{\rho(\epsilon)}{2} \sum_{i=1}^{N} \left\| x_{\rho(\epsilon),z^i_{k(\epsilon)}} \right\|_2^2 \\ &\quad - \frac{\mu(\epsilon)}{2} \left\| \widehat{u}_{k(\epsilon)} \right\|_2^2 - \theta_{\rho(\epsilon),\mu(\epsilon)}(z_{k(\epsilon)}) + \theta(z^{**}). \end{aligned}$$

Since $\theta_{\rho(\epsilon),\mu(\epsilon)}(z_{k(\epsilon)}) - \theta(z^{**}) \leq \theta(z_{k(\epsilon)}) - \theta(z^{**}) \leq \epsilon$, and

$$\begin{aligned} \theta_{\rho(\epsilon),\mu(\epsilon)}(z_{k(\epsilon)}) - \theta(z^{**}) &\overset{(9)}{\geq} \theta(z_{k(\epsilon)}) - \mu(\epsilon) D - \rho(\epsilon) \hat{D} - \theta(z^{**}) \\ &\overset{(21)}{=} \theta(z_{k(\epsilon)}) - \theta(z^{**}) - \tfrac{1}{2}\epsilon \;\geq\; -\tfrac{1}{2}\epsilon, \end{aligned}$$

we have $|\theta_{\rho(\epsilon),\mu(\epsilon)}(z_{k(\epsilon)}) - \theta(z^{**})| \leq \epsilon$. Therefore:

$$\begin{aligned} \left| \int_0^T F(t, \widehat{u}_{k(\epsilon)}(t))dt - D^* \right| &\leq \sum_{i=1}^{N} \left\| x_{\rho(\epsilon),z^i_{k(\epsilon)}} - \mathcal{A}_i \widehat{u}_{k(\epsilon)} \right\|_2 \left\| z^i_{k(\epsilon)} \right\|_2 + \rho(\epsilon)\hat{D} + \mu(\epsilon) D + \epsilon \\ &\overset{(21)}{\leq} \left\| \nabla \theta_{\rho(\epsilon),\mu(\epsilon)}(z_{k(\epsilon)}) \right\|_2 \left\| z_{k(\epsilon)} \right\|_2 + 2\epsilon \\ &\overset{(26)}{\leq} \frac{2\epsilon}{R} \left\| z_{k(\epsilon)} \right\|_2 + 2\epsilon. \end{aligned}$$

On the other hand:

$$\left\| z_{k(\epsilon)} \right\| \leq \left\| z_{k(\epsilon)} - z^* \right\|_2 + \| z^* \|_2 \overset{(17)}{\leq} 2 \| z^* \|_2 \overset{(25)}{\leq} 2\sqrt{3}R.$$

and we obtain:

$$\left| \int_0^T F(t, \widehat{u}_{k(\epsilon)}(t))dt - D^* \right| \leq 2\left(1 + 2\sqrt{3}\right) \cdot \epsilon.$$

Finally, we have:

- $\widehat{u}_{k(\epsilon)} \in \overline{U}$ by construction i.e. $\widehat{u}_{k(\epsilon)}(t) \in Q(t) \quad \forall t \in [0,T]$
- $\sqrt{\sum_{i=1}^{N} \left\| \mathcal{A}_i \widehat{u}_{k(\epsilon)} - x_{\rho(\epsilon),z^i_{k(\epsilon)}} \right\|_2^2} = \|\nabla \theta_{\rho(\epsilon),\mu(\epsilon)}(z_{k(\epsilon)})\|_2 \overset{(26)}{\leq} \frac{2\epsilon}{R}$,
  where $x_{\rho(\epsilon),z^i_{k(\epsilon)}} \in Q_i \quad \forall i = 1, ..., N$.

Therefore, function $\widehat{u}_k$ can be seen as an approximately feasible and optimal solution for the primal infinite-dimensional problem (1).

6 Condition for strong duality

As a simple consequence of the results of the previous sections, we can prove the strong duality between the primal and the dual problem, i.e. that $D^* = P^*$. We can justify this by the double smoothing algorithm, which can construct a sequence $\{u_n\} \subset \overline{U}$, such that u_n converges in a certain sense to the optimal solution of the primal problem (which is therefore solvable).

Let $\{\epsilon_n\}$ be a decreasing sequence of positive scalars such that $\lim_{n\to\infty} \epsilon_n = 0$. For each $n \geq 0$, we can apply $k(\epsilon_n)$ iterations of the double smoothing algorithm with the parameters $\mu(\epsilon_n), \rho(\epsilon_n), \kappa(\epsilon_n)$ defined by (21). Denote by $u_n = \widehat{u}_{k(\epsilon_n)} \in \overline{U}$ the output of the corresponding minimization process.

Theorem 2 *Let the dual problem (4) be solvable. Then there is no duality gap: $P^* = D^* = -\theta(z^{**})$, and the sequence $u_n = \widehat{u}_{k(\epsilon_n)}$ weakly converges to an optimal solution of the primal problem. Hence the problem (1) is solvable.*

Proof:

Note that $u_n \in \overline{U}$ for all $n \geq 0$, and $J(u_n) := \int\limits_0^T F(t, u_n(t))dt \to D^*$ as $n \to \infty$. Moreover, $\text{dist}\,(\mathcal{A}_i u_n, Q_i) \overset{(28)}{\to} 0$ as $n \to \infty$. Since the set $\overline{U}$ is bounded, the whole sequence $\{u_n\}$ is also bounded. Since $L^2([0,T], \mathbb{R}^m)$ is a reflexive Banach space, by Banach Theorem, we can extract a subsequence $\{u_{n_j}\} \subset \{u_n\}$ which converges weakly in $L^2([0,t_n], \mathbb{R}^m)$. Denote by u^* its weak limit ($u_{n_j} \rightharpoonup u^*$).

Let us prove first that J is continuous. Consider a sequence $\overline{u}_k \to \overline{u}$ in $L^2([0,T], \mathbb{R}^m)$. By Corollary 2.17 in [1], we can find a subsequence $\{\overline{u}_{k_j}\} \subset \{\overline{u}_k\}$, which converges to $\overline{u}$ pointwise almost everywhere. As F is continuous and bounded on $[0,T] \times Q$, we obtain, using the pointwise convergence and the Lebesgue dominated convergence theorem, that

$$\begin{aligned} \lim_{j\to\infty} J(u_{k_j}) &= \lim_{j\to\infty} \int\limits_0^T F(t, u_{k_j}(t))dt \;=\; \int\limits_0^T \lim_{j\to\infty} F(t, u_{k_j}(t))dt \\ &= \int\limits_0^T F(,t, \overline{u}(t))dt \;=\; J(\overline{u}). \end{aligned}$$

Suppose that there exists another subsequence $\{v_l\} \subset \{\overline{u}_k\}$ such that $\lim_{l\to\infty} J(v_l) = \alpha$ and $\alpha \neq J(\overline{u})$. Then, using the same arguments as above, we can extract a subsubsequence $\{v_{l_j}\} \subset \{v_l\}$ such that $\lim_{j\to\infty} J(v_{l_j}) = J(\overline{u})$ and we obtain a contradiction. We conclude that all convergent subsequence of $\{J(\overline{u}_k)\}$ converges necessarily to $J(\overline{u})$. Hence the total sequence is converging i.e $\lim_{k\to\infty} J(\overline{u}_k) = J(\overline{u})$ and we have proved that J is continuous.

Further, since $\overline{U}$ is closed and convex, and since J is convex, its continuity implies the weak lower semi-continuity of this functional (see Corollary III.8 in [4]). We conclude that $J(u^*) \leq \liminf_{j\to\infty} J(u_{n_j}) = D^*$.

Finally, since $u_{n_j} \rightharpoonup u^*$, and the operatots $\mathcal{A}_i$ are linear and continuous, we have $\mathcal{A}_i u_{n_j} \to \mathcal{A}_i u^*$ for all $i = 1, ..., N$. Taking into account that the sets Q_i are compact in $\mathbb{R}^n$, the continuity of the distance function implies

$$\text{dist}(\mathcal{A}_i u^*, Q_i) = \lim_{j\to\infty} \text{dist}(\mathcal{A}_i u_{n_j}, Q_i) = 0.$$

Hence, $\mathcal{A}_i u^* \in Q_i$ for all $i = 1, \ldots, N$.

It remains to note that $\overline{U}$ is closed and convex, and therefore (e.g. Theorem III.7 in [4]), it is weakly closed. Since $\{u_{n_j}\} \subset \overline{U}$ and $u_{n_j} \rightharpoonup u^*$, we conclude that $u^* \in \overline{U}$.

Thus, we have proved that u^* is a feasible solution for (1) and $J(u^*) \leq D^*$. Since $J(u) \geq P^* \geq D^*$ for all feasible u, we conclude that $P^* = D^*$, and u^* is the optimal primal solution. □

7 Applications in optimal control

7.1 Class of optimal control problems and reformulation

In this section, we will look at the optimal control problems that can be written in the form (1). In particular, we consider the optimal control problems governed by a system of linear differential equations with convex objective functional, convex constraints on the state variables at finite number of *inspection moments*, and the point-wise convex constraints on the control variables.

Consider the following optimal control problem:

$$\begin{aligned} \inf_u \{ \int_0^T F(t,u(t))dt : \quad & \dot{x}(t) = A(t)x(t) + B(t)u(t), \ x(0) = x_0, \\ & x(t_i) \in Q_i \quad i = 1..., N, \\ & u(t) \in Q(t) \quad \text{a.e in } [0,T] \}, \end{aligned} \tag{29}$$

where $T < \infty$, and $Q(t) \subset \mathbb{R}^m$, $t \in [0,T]$, are closed convex sets with bounded graph $Q \stackrel{\text{def}}{=} \cup_{t\in[0,T]} Q(t)$. We assume that function $F : [0,T] \times Q \to \mathbb{R}$ is bounded, and continuously differentiable and convex in the second argument, $x(t) \in \mathbb{R}^n$ and $u(t) \in \mathbb{R}^m$, $t \in [0,T]$. For measuring the control variables, we use the norm $\|u\|_2^2 = \int_0^T \|u(t)\|_2^2 dt$. We assume that

$$A(t) \in \mathcal{C}([0,T], \mathbb{R}^{n\times n}), \quad B(t) \in \mathcal{C}([0,T], \mathbb{R}^{n\times m}).$$

In problem (29), we have a finite number of inspection moments $t_i \in (0,T]$, and we assume that $Q_i \subset \mathbb{R}^n$, $i = 1, \ldots, N$, are bounded closed convex sets.

Let us rewrite the problem (29) in terms of control u. Denote by $\Phi(t,\tau)$ the transition matrix of the system. It is the unique solution of the following matricial Cauchy problem:

$$\tfrac{d}{dt}\Phi(t,\tau) \;=\; A(t)\Phi(t,\tau), \quad t \geq \tau, \quad \Phi(\tau,\tau) = I.$$

Remark 1 *When the system is time-invariant, i.e. $A(t) = A$, and $B(t) = B$, $t \in [0,T]$, then the transition matrix is the usual matrix exponent:*

$$\Phi(t,\tau) \;=\; e^{(t-\tau)A} = I + \sum_{k=1}^{\infty} \tfrac{A^k (t-\tau)^k}{k!}.$$

From the Optimal Control Theory (e.g [6]), we know that the state trajectory $x(t)$, generated by the system for a control $u(t)$, is given by the following expression:

$$x(t) = \Phi(t,0)x_0 + \int_0^t \Phi(t,\tau)B(\tau)u(\tau)d\tau, \quad t \in [0,T].$$

Therefore, the constraint $x(t_i) \in Q_i$ can be expressed as follows:

$$\mathcal{A}_i(u) \stackrel{\text{def}}{=} \int_0^{t_i} \Phi(t_i,\tau)B(\tau)u(\tau)d\tau \in \overline{Q}_i \stackrel{\text{def}}{=} Q_i - \Phi(t_i,0)x_0, \tag{30}$$

where $\Phi(t_i,0)x_0$ is the value at time t_i of the unique solution of Cauchy problem

$$\dot{x}(t) = A(t)x(t), \quad x(0) = x_0.$$

Remark 2 *At the first glance, it seems that we are restricted to the objective functionals depending only on the control $u(t)$ and not on the state variable $x(t)$. In fact, using the state transition matrix, we can also consider any convex functions depending on some linear functionals of the state. Such a functional can be defined as*

$$\begin{aligned} l(x) &= \int_0^T \langle x(t), a(t)\rangle dt = \int_0^T \langle \int_0^t \Phi(t,\tau)B(\tau)u(\tau)d\tau, a(t)\rangle dt \\ &= \int_0^T \int_0^t \langle \Phi(t,\tau)B(\tau)u(\tau), a(t)\rangle d\tau dt = \int_0^T \int_0^t \langle u(\tau), B(\tau)^T\Phi(t,\tau)^T a(t)\rangle d\tau dt \\ &= \int_0^T \int_\tau^T \langle u(\tau), B(\tau)^T\Phi(t,\tau)^T a(t)\rangle dt d\tau \stackrel{\text{def}}{=} \int_0^T \langle u(\tau), h(\tau)\rangle d\tau, \end{aligned}$$

with $h(\tau) = \int_\tau^T B(\tau)^T\Phi(t,\tau)^T a(t)dt$. Another possibility is as follows:

$$\begin{aligned} l(x) &= \langle x(t_i), a\rangle = \langle \int_0^{t_i} \Phi(t_i,\tau)B(\tau)u(\tau)d\tau, a\rangle \\ &= \int_0^{t_i} \langle \Phi(t_i,\tau)B(\tau)u(\tau), a\rangle d\tau \stackrel{\text{def}}{=} \int_0^{t_i} \langle u(\tau), h(\tau)\rangle d\tau, \end{aligned}$$

with $h(\tau) = B(\tau)^T\Phi(t_i,\tau)^T a$.

Thus, for the linear operator $\mathcal{A}_i : L^2([0,T],\mathbb{R}^m) \to \mathbb{R}^n$, defined by (30), the ith state constraint becomes:

$$\mathcal{A}_i u = \int_0^T A_i(\tau)u(\tau)d\tau \in \bar{Q}_i, \tag{31}$$

where $A_i(\tau) \stackrel{\text{def}}{=} \begin{cases} \Phi(t_i,\tau)B(\tau), & \text{when } \tau \in [0,t_i], \\ 0, & \text{when } \tau \in]t_i,T]. \end{cases}$

Thus, the optimal control problem (29) can be rewritten in the form (1). Hence, we can solve it by the double smoothing technique. This approach assumes that we are able to solve the pointwise problems

$$\max_{u\in Q(t)} \left\{ -F(t,u) - \sum_{i=1}^N \langle u, A_i^T(t)z^i\rangle - \tfrac{\mu}{2}\|u\|_2^2 \right\},$$

where $A_i(t)$ depends directly on the state transition matrix. However, in practice the state transition matrix $\Phi(t_i, t)$ is often not known. Instead, we can compute the function $A_i^T(t)z^i$ as a solution of some ODE.

Indeed, we have (e.g. Theorem 1.2 in [7])

$$\frac{d}{dt}\Phi^T(t_i, t) \quad = \quad -A(t)^T \Phi^T(t_i, t).$$

Therefore $\Phi(t_i, t)^T$ is the state transition matrix of the system $\dot{v}(t) = -A(t)^T v(t)$. Hence, $A_i^T(t)^T z^i = B(t)^T v(t)$, where $v(t)$ is the unique solution of Cauchy problem

$$\dot{v}(t) \quad = \quad -A(t)^T v(t), \quad v(t_i) = z^i, \quad t \in [0, t_i], \tag{32}$$

extended by zero for $t \in [t_i, T]$.

7.2 Evaluation of $\|\mathcal{A}_i\|_2$

In order to solve the primal-dual problem (1), (4) by double smoothing technique, we need to evaluate the norms $\|\mathcal{A}_i\|_2$. Moreover, from the estimates (23), (24), it is clear that these norms are very essential elements of the global complexity bound of our problem. In this section, using the reachability Gramian of the dynamical system, we derive a closed-form representation for the norm $\|\mathcal{A}_i\|_2$. However, this quantity is not easily computable (it needs the knowledge of the transition matrix). Moreover, its dependence in the length of time interval is not very transparent. Therefore, in the next section, we obtain some simple upper bounds for the norms $\|\mathcal{A}_i\|_2$, which can be easily computed by solving Linear Matrix Inequalities (LMI).

Let us derive first the exact expression for $\|\mathcal{A}_i\|_2$. By definition,

$$\|\mathcal{A}_i\|_2 \quad = \sup_{u \in L^2([0,T],\mathbb{R}^m)} \left\{ \|\mathcal{A}_i u\|_2 : \ \|u\|_{L^2([0,T],\mathbb{R}^m)} = 1 \right\}.$$

Since the vector $\mathcal{A}_i u$ does not depend on values of $u(t)$ for $t \in (t_i, T]$, we can consider the restriction of $\mathcal{A}_i$ on $L^2([0, t_i], \mathbb{R}^m)$:

$$u \quad \to \quad \int_0^{t_i} \Phi(t_i, \tau) B(\tau) u(\tau) d\tau.$$

Then

$$\|\mathcal{A}_i\|_2 \quad = \sup_{u \in L^2([0,t_i],\mathbb{R}^m)} \left\{ \|\mathcal{A}_i u\|_2 : \ \|u\|_{L^2([0,t_i],\mathbb{R}^m)} = 1 \right\},$$

and the operator $\mathcal{A}_i^*$ transforms $y \in R^n$ into the function $B(t)^T \Phi(t_i, t)^T y \in L^2([0, t_i]$.

For all $t_i > 0$, $i = 1, \dots, N$, define the reachability Gramians

$$W_r(0, t_i) \quad = \quad \int_0^{t_i} \Phi(t_i, \tau) B(\tau) B(\tau)^T \Phi(t_i, \tau)^T d\tau \ = \ \mathcal{A}_i \mathcal{A}_i^*,$$

which are symmetric positive semidefinite matrices ($\in S_+^n$). Recall the following definition.

Definition 1 *The system*

$$\dot{x}(t) \quad = \quad A(t)x(t) + B(t)u(t) \quad x(0) = 0, \tag{33}$$

is called reachable *on* $[0, \hat{t}]$ *if for any* $\hat{x} \in \mathbb{R}^n$ *there exist a control* $u(t)$ *such that* $x(\hat{t}) = \hat{x}$.

The reachability is closely related with reachability Gramian (e.g. Corollary 2.3 in [2]):

Theorem 3 *The system (33) is reachable on* $[0, t_i]$ *if and only if the Gramian* $W_r(0, t_i)$ *is positive definite.*

Let us come back now to the definition of the norm $\|\mathcal{A}_i\|_2$. We have:

$$
\begin{aligned}
\|\mathcal{A}_i\|_2 &= \sup_{u \in L^2([0,t_i],\mathbb{R}^m)} \left\{ \|\mathcal{A}_i u\|_2 : \ \|u\|_{L^2([0,t_i],\mathbb{R}^m)} = 1 \right\} \\
&= \left[\inf_{u \in L^2([0,t_i],\mathbb{R}^m)} \left\{ \|u\|_{L^2([0,t_i],\mathbb{R}^m)} : \ \|\mathcal{A}_i u\|_2 = 1 \right\} \right]^{-1}.
\end{aligned}
$$

If the system is reachable on $[0, t_i]$, then $\mathrm{Im}\mathcal{A}_i(L^2([0, t_i], \mathbb{R}^m)) = \mathbb{R}^n$, and we have:

$$
\begin{aligned}
& \inf_{u \in L^2([0,t_i],\mathbb{R}^m)} \left\{ \|u\|_{L^2([0,t_i],\mathbb{R}^m)} : \ \|\mathcal{A}_i u\|_2 = 1 \right\} \\
= & \inf_{\substack{x_i \in R^n, \|x_i\|=1 \\ u \in L^2([0,t_i],\mathbb{R}^m)}} \left\{ \|u\|_{L^2([0,t_i],\mathbb{R}^m)} : \ \mathcal{A}_i u = x_i \right\}.
\end{aligned}
$$

Consider now the minimization problem $\min\limits_{\substack{u \in L^2([0,t_i],\mathbb{R}^m), \\ \mathcal{A}_i u = x_i}} \|u\|_2$. We will use the following simple result

Lemma 1 *Let* H *be a Hilbert space and the linear operator* $A : H \to \mathbb{R}^L$ *be nondegenerate:* $AA^* \succ 0$*. Then for any* $b \in \mathbb{R}^L$ *and* $f \in H$*, the Euclidean projection* $\pi_b(f)$ *of* f *onto the subspace* $\mathcal{L}_b = \{g \in H : Ag = b\}$ *is defined as follows:*

$$
\pi_b(f) = f + A^*(AA^*)^{-1}(b - Af).
$$

Thus,

$$
\begin{aligned}
\inf_{\substack{u \in L^2([0,t_i],\mathbb{R}^m), \\ \mathcal{A}_i u = x_i}} \|u\|_2 &= \|\mathcal{A}_i^*(\mathcal{A}_i\mathcal{A}_i^*)^{-1} x_i\|_2 = \langle \mathcal{A}_i^*(\mathcal{A}_i\mathcal{A}_i^*)^{-1} x_i, \mathcal{A}_i^*(\mathcal{A}_i\mathcal{A}_i^*)^{-1} x_i \rangle^{1/2} \\
&= \langle (\mathcal{A}_i\mathcal{A}_i^*)^{-1} x_i, x_i \rangle^{1/2}.
\end{aligned}
$$

Therefore

$$
\begin{aligned}
\inf_{u \in L^2([0,t_i],\mathbb{R}^m)} \left\{ \|u\|_{L^2([0,t_i],\mathbb{R}^m)} : \ \|\mathcal{A}_i u\|_2 = 1 \right\} &= \inf_{\|x_i\|_2 = 1} \langle (\mathcal{A}_i\mathcal{A}_i^*)^{-1} x_i, x_i \rangle^{1/2} \\
&= \lambda_{\min}^{1/2}((\mathcal{A}_i\mathcal{A}_i^*)^{-1}),
\end{aligned}
$$

and we conclude that

$$
\|\mathcal{A}_i\|_2 = \lambda_{\min}^{-1/2}((\mathcal{A}_i\mathcal{A}_i^*)^{-1}) = \lambda_{\max}^{1/2}(\mathcal{A}_i\mathcal{A}_i^*),
$$

where $\mathcal{A}_i\mathcal{A}_i^* = W_r(0, t_i)$ is the reachability Gramian.

7.3 Bounding the growth of norms $\|\mathcal{A}_i\|_2$ with time

In the previous section, we have shown that the norm $\|\mathcal{A}_i\|_2$ is equal to the square root of the maximal eigenvalue of the reachability Gramian on the interval $[0, t_i]$. Simple examples show that this norm can grow exponentially with t_i. However, for the stable systems the situation is much better.

In this section, we derive the bounds for the growth of the norms $\|\mathcal{A}_i\|_2$ from the stability characteristics of the linear time-varying system:

$$\dot{x}(t) \;=\; A(t)x(t), \quad t \geq 0, \tag{34}$$

where the matrix $A(t)$ is continuous in time.

Recall that the state $x = 0$ is always an equilibrium of the system (34). It is the unique equilibrium if $A(t)$ is nonsingular for all $t \geq 0$. The following facts are standard (e.g. [2]).

Theorem 4 *The equilibrium $x = 0$ is stable if and only if the solutions of the linear systems are bounded. That is*

$$\sup_{t\geq\tau} \|\Phi(t,\tau)\|_2 \;\stackrel{\text{def}}{=}\; k(\tau) < \infty, \quad \forall \tau \geq 0.$$

It is uniformly stable if and only if

$$\sup_{\tau\geq 0} k(\tau) \;=\; \sup_{\tau\geq 0}\sup_{t\geq\tau} \|\Phi(t,\tau)\|_2 \;\stackrel{\text{def}}{=}\; k_0 < \infty.$$

Finally, it is is exponentially stable if $\int_0^t \|\Phi(t,\tau)\|_2^2\, d\tau \leq C$ for all $t \geq 0$ where the constant C is independent on t.

Using these stability results, we can obtain some estimates for the growth of $\|\mathcal{A}_i\|_2$.

Theorem 5 *If the equilibrium $x = 0$ is stable and $k_1 \stackrel{\text{def}}{=} \sup_{t\geq 0} \|B(t)\|_2 < \infty$, then*

$$\|\mathcal{A}_i\|_2 \;\leq\; k_1 \left[\int_0^{t_i} k_0^2(\tau) d\tau\right]^{1/2}. \tag{35}$$

Proof:
For all $u \in L^2([0, t_i]\mathbb{R}^m)$, we have

$$\begin{aligned}
\|\mathcal{A}_i u\| \;&=\; \left\|\int_0^{t_i} \Phi(t_i,\tau)B(\tau)u(\tau)d\tau\right\| \;\leq\; \int_0^{t_i} \|\Phi(t_i,\tau)B(\tau)\|_2\, \|u(\tau)\|_2\, d\tau \\
&\leq\; \left[\int_0^{t_i} \|\Phi(t_i,\tau)\|_2^2 \|B(\tau)\|_2^2 d\tau \cdot \int_0^{t_i} \|u(\tau)\|_2^2 d\tau\right]^{1/2} \\
&\leq\; k_1 \left[\int_0^{t_i} k_0(\tau)^2 d\tau\right]^{1/2} \|u\|_2\,.
\end{aligned}$$

Therefore $\|\mathcal{A}_i\|_2 \leq k_1 \left[\int_0^{t_i} k_0(\tau)^2 d\tau\right]^{1/2}$. □

This upper bound depends on the growth of the integral $\int_0^{t_i} k_0(\tau)^2 d\tau$ with respect to t_i, which can be very fast. Moreover, it can happen that function k_0 is not in $L^2([0, t_i])$ and then the bound (35) gives no information. However, if we assume the uniform stability of the equilibrium $x = 0$, then we can get much better bounds.

Theorem 6 *If equilibrium $x = 0$ is uniformly stable and $k_1 \stackrel{\text{def}}{=} \sup_{t \geq 0} \|B(t)\|_2 < \infty$, then*

$$\|\mathcal{A}_i\|_2 \quad \leq \quad k_0 k_1 \sqrt{t_i}.$$

The proof of this theorem is the same as that of Theorem 5. However, now we can ensure a sublinear bound for the growth $\|\mathcal{A}_i\|_2$ with respect to t_i. If we strengthen again the stability assumption, we can obtain an upper bound independent on t_i.

Theorem 7 *Let equilibrium $x = 0$ be exponentially stable and $k_1 = \sup\limits_{t \geq 0} \|B(t)\|_2 < \infty$. Then $\|\mathcal{A}_i\|_2 \leq k_1 \sqrt{C}$.*

Again, this fact can be easily derived from the arguments of the proof of Theorem 5.

In some case, we can obtain a computable upper bound for the norm $\|\mathcal{A}_i\|_2$. Recall the following well-known sufficient condition of global exponential stability.

Theorem 8 *[2] Let the linear system (34) be time-invariant, and there exists a matrix $P = P^T \succ 0$ such that $A^T P + PA \prec 0$. Then equilibrium $x = 0$ is globally exponentially stable.*

Under conditions of this theorem, there exists $\eta_1 > 0$ such that the following LMI

$$A^T P + PA \preceq -\eta_1 P, \quad P = P^T \succ 0,$$

admits a solution. Matrix P and constant η_1 can help us to obtain an explicit upper-bound for the norm $\|\mathcal{A}_i\|_2$. Indeed, by definition, $\mathcal{A}_i u$ is the position at time t_i of the point of unique trajectory defined by the linear system

$$\dot{x}(t) \quad = \quad Ax(t) + Bu(t), \quad x(0) = 0.$$

Therefore,

$$\|x(t_i)\|_2^2 \quad = \quad \langle x(t_i), x(t_i) \rangle \ \leq \ \tfrac{\langle Px(t_i), x(t_i) \rangle}{\lambda_{\min}(P)} \ \stackrel{\text{def}}{=} = \ \tfrac{R(t_i)}{\lambda_{\min}(P)},$$

where $R(t) \stackrel{\text{def}}{=} \langle Px(t), x(t) \rangle$. The derivative of function R can be bounded as follows:

$$\begin{aligned}
\dot{R}(t) &= \langle P, x(t)\dot{x}(t)^T + \dot{x}(t)x(t)^T \rangle \\
&= \langle P, x(t)(Ax(t) + Bu(t))^T + (Ax(t) + Bu(t))x(t)^T \rangle \\
&= \langle P(Ax(t) + Bu(t)), x(t) \rangle + \langle Px(t), Ax(t) + Bu(t) \rangle \\
&= \langle (PA + A^T P)x(t), x(t) \rangle + 2\langle Px(t), Bu(t) \rangle \\
&\leq -\eta_1 \langle Px(t), x(t) \rangle + 2\langle Px(t), Bu(t) \rangle \ \leq \ \tfrac{1}{\eta_1} \langle PBu(t), Bu(t) \rangle.
\end{aligned}$$

Since $x(0) = 0$, we get

$$\begin{aligned} R(t_i) &= \int_0^{t_i} \dot{R}(t)dt \ \le \ \frac{1}{\eta_1} \int_0^{t_i} \langle PBu(t), Bu(t)\rangle dt \\ &\le \ \frac{1}{\eta_1}\lambda_{\max}(P) \int_0^{t_i} \|Bu(t)\|_2^2 \, dt \ \le \ \frac{1}{\eta_1}\lambda_{\max}(P)\, \|B\|_2^2 \, \|u\|_2^2 . \end{aligned}$$

Hence, $\|\mathcal{A}_i u\|_2^2 \le \frac{\lambda_{\max}(P)}{\eta_1 \lambda_{\min}(P)} \|B\|_2^2 \|u\|_2^2$, and therefore $\|\mathcal{A}_i\|_2^2 \le \frac{\lambda_{\max}(P)}{\eta_1 \lambda_{\min}(P)} \|B\|_2^2$.

If we want to obtain the best upper bound for $\|\mathcal{A}_i\|_2$, we need to solve the following optimization problem in the variables η_1, η_2, η_3, and P:

$$\min\left\{ \tfrac{\eta_3}{\eta_1\eta_2} : \quad A^T P + PA \preceq -\eta_1 P, \quad \eta_2 I \preceq P \preceq \eta_3 I, \quad \eta_1, \eta_2, \eta_3 \ge 0 \right\}. \tag{36}$$

This problem is non-convex, but we can find an upper bound for its optimal solution from quasiconvex LMI. Note that

$$\|\mathcal{A}_i\|_2^2 \ \le \ \min\left\{ \tfrac{\eta_3}{\eta_1\eta_2} : \quad A^T P + PA \preceq -\eta_1\eta_3 I, \quad \eta_2 I \preceq P \preceq \eta_3 I, \quad \eta_1, \eta_2, \eta_3 \ge 0 \right\}$$

since the feasible set of the right-hand side is smaller than that of (36). Introducing a new variable $\tau_1 = \eta_1\eta_3$, we get the following problem:

$$\min\left\{ \frac{\eta_3^2}{\tau_1\eta_2} : \quad A^T P + PA \preceq -\tau_1 I, \quad \eta_2 I \preceq P \preceq \eta_3 I, \quad \tau_1, \eta_2, \eta_3 \ge 0 \right\}.$$

Since the objective of this problem is quasiconvex, it can be solved in polynomial time.

8 Discretization and application to large-scale finite-dimensional problems

The double smoothing technique developed in this paper allows us to solve the primal-dual infinite-dimensional problem (1), (4) up to accuracy ϵ in $O\left(\frac{1}{\epsilon}\ln\left(\frac{1}{\epsilon}\right)\right)$ iterations. The complexity analysis of this approach is done in the infinite-dimensional framework. However, the exact computation of the gradient $\nabla\theta_{\rho,\mu,\kappa}(w_k)$ at each iteration needs in general an infinite number of pointwise operations. For some simple problems this computation can be implemented in a finite time. However, in practice this situation is very rear.

For the moment, there are two ways for avoiding this difficulty. The first one is to adapt the optimal scheme for $S^{1,1}_{\mu,L}$ to the case when the gradient is computed with certain accuracy. Under this assumption, we will be able to keep all the analysis in the infinite-dimensional framework, and to obtain the accuracy of ϵ in $O\left(\frac{1}{\epsilon}\ln\left(\frac{1}{\epsilon}\right)\right)$ iterations for the infinite-dimensional problem itself. The discretization will be used only inside the first order orcale, which computes the gradient of the modified dual objective function with given accuracy. The rigorous justification of corresponding variant of the optimal method for $S^{1,1}_{\mu,L}$ is the topic of our forthcoming paper.

The second possibility is to discretize the infinite-dimensional problem (1) from the very beginning. Even in this classical framework, our double smoothing approach keeps all of its strong points. Indeed, the discretized problem will be typically a very large scale finite-dimensional problem with coupling constraints. Since it is often easier to solve a large number of small problems than one large problem with coupling constraints, it is interesting to tackle these problems by dualizing the coupling constraints and to apply the double smoothing scheme to the Lagrangian dual problem. The analysis presented in this paper for infinite-dimensional framework remains valid also in finite dimension. If we assume that the pointwise problems are solvable in a closed-form, the double smoothing technique gives us a possibility to solve the discretized problem with an accuracy of ϵ in $O\left(\frac{1}{\epsilon}\ln\left(\frac{1}{\epsilon}\right)\right)$ iterations using an appropriate first-order method. More generally, our approach can be applied to any kind of large-scale convex problems in finite dimension in the presence of linear coupling constraints and separable point-wise constraints.

References

[1] R.A. Adams and J.F. Fournier. *Sobolev Spaces (Second edition)*. Elsevier Science (2003).

[2] P.J. Antsaklis and A.N. Michel. *Linear Systems*. Birkhauser Book (2006).

[3] P. Bernhards and A. Rapaport. On a theorem of Danskin with an application to a theorem of Von Neumann-Sion. *Nonlinear analysis*, **24**, 1163-1181 (1995).

[4] H. Brezis. *Analyse fonctionnelle: Theorie et Applications*. Masson (1987).

[5] J.M. Danskin. *The theory of Max-Min and its application to weapons allocation problems*. Springer-Verlag (1967).

[6] D.E. Kirk. *Optimal Control: An Introduction*. Dover publication, Inc. (2004).

[7] H. Kwakernaak and R. Sivan. *Linear Optimal Control Systems*. Wiley Interscience (1972).

[8] D.G. Luenberger. *Optimization by vector space methods*. Wiley (1969).

[9] Yu. Nesterov. *Introductory Lectures on Convex Optimization: A Basic Course*. Kluwer Academic Publishers (2004)

[10] Yu. Nesterov. Smooth minimization of non-smooth functions. *Mathematical programming, Serie A*, **103**, 127-152 (2005).

[11] Yu. Nesterov. Excessive gap technique in nonsmooth convex minimization. *Siam Journal of Optimization*, **16**, 235-249 (2005).

[12] Yu. Nesterov. Smoothing technique and its applications in semidefinite optimization. *Mathematical Programming A*, **110**, 245-259 (2007).

[13] Yu. Nesterov. Primal-Dual subgradient methods for convex problems. *Mathematical programming, Serie B*, **120**, 221-259 (2009).

Recent titles

CORE Discussion Papers

2009/78. Julio DÁVILA. The taxation of savings in overlapping generations economies with unbacked risky assets.
2009/79. Elena DEL REY and Miguel Angel LOPEZ-GARCIA. Optimal education and pensions in an endogenous growth model.
2009/80. Hiroshi UNO. Strategic complementarities and nested potential games.
2009/81. Xavier WAUTHY. Market coverage and the nature of product differentiation: a note.
2009/82. Filippo L. CALCIANO. Nash equilibria of games with increasing best replies.
2009/83. Jacques H. DRÈZE, Oussama LACHIRI and Enrico MINELLI. Stock prices, anticipations and investment in general equilibrium.
2009/84. Claire DUJARDIN and Florence GOFFETTE-NAGOT. Neighborhood effect on unemployment? A test *à la* Altonji.
2009/85. Erwin OOGHE and Erik SCHOKKAERT. School accountability: (how) can we reward schools and avoid cream-skimming.
2009/86. Ilke VAN BEVEREN and Hylke VANDENBUSSCHE. Product and process innovation and the decision to export: firm-level evidence for Belgium.
2010/1. Giorgia OGGIONI and Yves SMEERS. Degree of coordination in market-coupling and counter-trading.
2010/2. Yu. NESTEROV. Efficiency of coordinate descent methods on huge-scale optimization problems.
2010/3. Geert DHAENE an Koen JOCHMANS. Split-panel jackknife estimation of fixed-effect models.
2010/4. Parkash CHANDER. Cores of games with positive externalities.
2010/5. Gauthier DE MAERE D'AERTRYCKE and Yves SMEERS. Liquidity risks on power exchanges.
2010/6. Marc FLEURBAEY, Stéphane LUCHINI, Christophe MULLER and Erik SCHOKKAERT. Equivalent income and the economic evaluation of health care.
2010/7. Elena IÑARRA, Conchi LARREA and Elena MOLIS. The stability of the roommate problem revisited.
2010/8. Philippe CHEVALIER, Isabelle THOMAS and David GERAETS, Els GOETGHEBEUR, Olivier JANSSENS, Dominique PEETERS and Frank PLASTRIA. Locating fire-stations: an integrated approach for Belgium.
2010/9. Jean-Charles LANGE and Pierre SEMAL. Design of a network of reusable logistic containers.
2010/10. Hiroshi UNO. Nested potentials and robust equilibria.
2010/11. Elena MOLIS and Róbert F. VESZTEG. Experimental results on the roommate problem.
2010/12. Koen DECANCQ. Copula-based orderings of multivariate dependence.
2010/13. Tom TRUYTS. Signaling and indirect taxation.
2010/14. Asel ISAKOVA. Currency substitution in the economies of Central Asia: How much does it cost?
2010/15. Emanuele FORLANI. Irish firms' productivity and imported inputs.
2010/16. Thierry BRECHET, Carmen CAMACHO and Vladimir M. VELIOV. Model predictive control, the economy, and the issue of global warming.
2010/17. Thierry BRECHET, Tsvetomir TSACHEV and Vladimir M. VELIOV. Markets for emission permits with free endowment: a vintage capital analysis.
2010/18. Pierre M. PICARD and Patrice PIERETTI. Bank secrecy, illicit money and offshore financial centers.
2010/19. Tanguy ISAAC. When frictions favour information revelation.
2010/20. Jeroen V.K. ROMBOUTS and Lars STENTOFT. Multivariate option pricing with time varying volatility and correlations.
2010/21. Yassine LEFOUILI and Catherine ROUX. Leniency programs for multimarket firms: The effect of Amnesty Plus on cartel formation.
2010/22. P. Jean-Jacques HERINGS, Ana MAULEON and Vincent VANNETELBOSCH. Coalition formation among farsighted agents.

Recent titles

CORE Discussion Papers - continued

2010/23. Pierre PESTIEAU and Grégory PONTHIERE. Long term care insurance puzzle.
2010/24. Elena DEL REY and Miguel Angel LOPEZ-GARCIA. On welfare criteria and optimality in an endogenous growth model.
2010/25. Sébastien LAURENT, Jeroen V.K. ROMBOUTS and Francesco VIOLANTE. On the forecasting accuracy of multivariate GARCH models.
2010/26. Pierre DEHEZ. Cooperative provision of indivisible public goods.
2010/27. Olivier DURAND-LASSERVE, Axel PIERRU and Yves SMEERS. Uncertain long-run emissions targets, CO_2 price and global energy transition: a general equilibrium approach.
2010/28. Andreas EHRENMANN and Yves SMEERS. Stochastic equilibrium models for generation capacity expansion.
2010/29. Olivier DEVOLDER, François GLINEUR and Yu. NESTEROV. Solving infinite-dimensional optimization problems by polynomial approximation.
2010/30. Helmuth CREMER and Pierre PESTIEAU. The economics of wealth transfer tax.
2010/31. Thierry BRECHET and Sylvette LY. Technological greening, eco-efficiency, and no-regret strategy.
2010/32. Axel GAUTIER and Dimitri PAOLINI. Universal service financing in competitive postal markets: one size does not fit all.
2010/33. Daria ONORI. Competition and growth: reinterpreting their relationship.
2010/34. Olivier DEVOLDER, François GLINEUR and Yu. NESTEROV. Double smoothing technique for infinite-dimensional optimization problems with applications to optimal control.

Books

J. GABSZEWICZ (ed.) (2006), La différenciation des produits. Paris, La découverte.

L. BAUWENS, W. POHLMEIER and D. VEREDAS (eds.) (2008), *High frequency financial econometrics: recent developments*. Heidelberg, Physica-Verlag.

P. VAN HENTENRYCKE and L. WOLSEY (eds.) (2007), *Integration of AI and OR techniques in constraint programming for combinatorial optimization problems*. Berlin, Springer.

P-P. COMBES, Th. MAYER and J-F. THISSE (eds.) (2008), *Economic geography: the integration of regions and nations*. Princeton, Princeton University Press.

J. HINDRIKS (ed.) (2008), *Au-delà de Copernic: de la confusion au consensus* ? Brussels, Academic and Scientific Publishers.

J-M. HURIOT and J-F. THISSE (eds) (2009), *Economics of cities.* Cambridge, Cambridge University Press.

P. BELLEFLAMME and M. PEITZ (eds) (2010), *Industrial organization: markets and strategies*. Cambridge University Press.

M. JUNGER, Th. LIEBLING, D. NADDEF, G. NEMHAUSER, W. PULLEYBLANK, G. REINELT, G. RINALDI and L. WOLSEY (eds) (2010), *50 years of integer programming, 1958-2008: from the early years to the state-of-the-art*. Berlin Springer.

CORE Lecture Series

C. GOURIÉROUX and A. MONFORT (1995), Simulation Based Econometric Methods.

A. RUBINSTEIN (1996), Lectures on Modeling Bounded Rationality.

J. RENEGAR (1999), A Mathematical View of Interior-Point Methods in Convex Optimization.

B.D. BERNHEIM and M.D. WHINSTON (1999), Anticompetitive Exclusion and Foreclosure Through Vertical Agreements.

D. BIENSTOCK (2001), Potential function methods for approximately solving linear programming problems: theory and practice.

R. AMIR (2002), Supermodularity and complementarity in economics.

R. WEISMANTEL (2006), Lectures on mixed nonlinear programming.