A DOUBLE SMOOTHING TECHNIQUE FOR CONSTRAINED CONVEX OPTIMIZATION PROBLEMS AND APPLICATIONS TO OPTIMAL CONTROL

OLIVIER DEVOLDER, FRANÇOIS GLINEUR, AND YURII NESTEROV *

Abstract. In this paper, we propose an efficient approach for solving a class of convex optimization problems in Hilbert spaces. Our feasible region is a (possibly infinite-dimensional) simple convex set, i.e. we assume that projections on this set are computationally easy to compute. The problem we consider is the minimization of a convex function over this region under the additional constraint $Au \in T$, where A is a linear operator and T is a (finite-dimensional) convex set whose dimension is small as compared to the dimension of the feasible region.

In our approach, we dualize the linear constraints, solve the resulting dual problem with a purely dual gradient method and show how to reconstruct an approximate primal solution. In order to accelerate our scheme, we introduce a novel double smoothing technique that involves regularization of the dual problem to allow the use of a fast gradient method. As a result, we obtain a method with complexity $O(\frac{1}{\epsilon} \ln \frac{1}{\epsilon})$ gradient iterations, where ϵ is the desired accuracy for the primal-dual solution.

Our approach covers, in particular, optimal control problems with trajectory governed by a system of linear differential equations, where the additional constraints can for example force the trajectory to visit some convex sets at certain moments of time.

Key words. convex optimization, fast gradient methods, complexity theory, smoothing technique, optimal control

AMS subject classifications. 90C06, 90C25, 90C60

1. Introduction.

In large-scale convex optimization, first-order methods are the methods of choice due to their cheap iteration cost. In particular, constrained problems can be solved provided that projection on their feasible set is computationally easy to compute. In this work, we assume that both the convex objective function J, defined on the Hilbert space U, and the convex feasible region $S \subset U$ are sufficiently simple so that the problem $\min_{u \in S} J(u)$ can be solved efficiently, or even in closed-form. However, the situation becomes completely different when adding to this problem the constraint $\mathcal{A}u \in T$, based on a linear operator $\mathcal{A} : U \to V^*$, where V is another Hilbert space and T a bounded closed convex set in V^* (the dual space of V). Indeed, the problem may become difficult because projection onto the new feasible set $\{u \in S : \mathcal{A}u \in T\}$ may be computationally very expensive, or even intractable.

A natural approach is therefore to dualize this difficult linear constraint, obtaining the primal-dual pair of problems

$$P^* = \min_{u \in S}\{J(u) + \max_{z \in V}[\langle \mathcal{A}u, z\rangle - \sigma_T(z)]\}, \quad D^* = \max_{z \in V}\{-\sigma_T(z) + \min_{u \in S}[J(u) + \langle \mathcal{A}u, z\rangle]\}$$

where $\sigma_T(z) = \sup_{x \in T}\langle x, z\rangle$ denotes the support function of set T, defined on V.

In this paper, we assume that the dimension of set V (i.e. the size of the linear constraints) is small compared to the dimension of set U, the latter being allowed to be infinite. Thanks to this asymmetry, we are led to consider a purely dual algorithmic scheme, generating its iterates only in the low-dimensional space V. The

*Center for Operations Research and Econometrics (CORE), Université catholique de Louvain (UCL), 34 voie du Roman Pays, 1348 Louvain-la-Neuve, Belgium; E-mail: Olivier.Devolder@uclouvain.be, Francois.Glineur@uclouvain.be, Yurii.Nesterov@uclouvain.be. The first author is a F.R.S.-FNRS Research Fellow. This text 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 rests with its authors.

only operations that we need to be able to perform in the infinite-dimensional or high-dimensional space U is the computation of the value of dual objective function at a given point $z \in V$, which requires solving the optimization subproblem $\Theta(z) = \min_{u \in S}[J(u) + \langle \mathcal{A}u, z\rangle] - \sigma_T(z)$ over the simple set S.

Example 1. *As a first motivation, consider the purely linear infinite-dimensional problem:*

$$P^* = \inf_{u \in L^2([\alpha,\beta]): M_1 \leq u(t) \leq M_2} \int_\alpha^\beta f(t)u(t)dt : \quad \int_\alpha^\beta a_i(t)u(t)dt = b_i \quad \forall i = 1, ..., m$$

where data consists of functions a_i $(1 \leq i \leq m)$ and f in $L^2([\alpha,\beta])$.
It is straightforward to define

$$\begin{aligned} U &= L^2([\alpha,\beta]), \quad V = R^m, \\ S &= \{u \in L^2([\alpha,\beta]) : M_1 \leq u(t) \leq M_2 \ \forall \alpha \leq t \leq \beta\}, \quad T = \{b\} \subset R^m, \\ J(u) &= \int_\alpha^\beta f(t)u(t)dt, \quad \mathcal{A} : U \to V : u \to \left(\int_\alpha^\beta a_1(t)u(t)dt, ..., \int_\alpha^\beta a_m(t)u(t)dt\right)^T \end{aligned}$$

so that the problem fits our formulation $\min_{u \in S} J(u) : \mathcal{A}u \in T$. Dualizing the linear equality constraints, we obtain the dual function:

$$\Theta(z) = \inf_{u \in S} \int_\alpha^\beta f(t)u(t)dt + \sum_{i=1}^m z_i \left(\int_\alpha^\beta a_i(t)u(t)dt - b_i\right) \tag{1.1}$$

$$= -\sum_{i=1}^m z_i b_i + \inf_{u \in S} \int_\alpha^\beta \Big(f(t) + \sum_{i=1}^m z_i a_i(t)\Big)u(t)dt. \tag{1.2}$$

Due to the fact that only pointwise constraints $M_1 \leq u(t) \leq M_2$ are still present in this problem, we can solve it in a pointwise way, minimizing for each value of t separately. Indeed, any solution $u_z \in S$ satisfying $u_z(t) = M_1$ when $f(t) + \sum_{i=1}^m y_i a_i(t) > 0$ and $u_z(t) = M_2$ when $f(t) + \sum_{i=1}^m y_i a_i(t) < 0$ is optimal solution for problem (1.1).
Since we are able to compute the value of $\Theta(z)$ in closed form for any value of z, we can apply a first-order method to the finite-dimensional problem $\max_{z \in V} \Theta(z)$.

Our goal in this work is to show that it is possible to solve the dual problem efficiently and reconstruct from this process a nearly optimal and feasible primal solution. We develop to that effect a new double-smoothing approach, which is a variant of the smoothing techniques described in [20, 21, 22]. This technique uses the problem structure to regularize the dual objective function into a smooth strongly convex function with Lipschitz continuous gradient. These modifications allow us to minimize the dual function with an optimal gradient scheme in $O\left(\frac{1}{\epsilon}\ln\left(\frac{1}{\epsilon}\right)\right)$iterations, where ϵ is the desired accuracy. From the dual minimization sequence, we reconstruct a nearly feasible and optimal primal solution, whose accuracy can be controlled by parameters of our algorithm.

The structure of this paper is as follows. In the Section 2, we recall briefly two standard approaches for solving a non-smooth convex optimization problem with a first-order method: subgradient type schemes and smoothing techniques. We also present the first-order methods that can be used to solve efficiently the smoothed problem obtained by the smoothing technique. In particular, we recall the optimal method [19] for smooth and strongly convex functions and describe its rate of convergence. In Section 3, we present in a more general form our problem class and derive the corresponding dual problem. We show using Danskin's Theorem that the dual objective function is in general non-smooth. Section 4 present two simple examples of problems (one finite-dimensional, the other infinite-dimensional) with separable structure that fit our problem class. The double smoohting is described in Section 5, where we apply two regularizations to the dual objective function in order to make it smooth and strongly convex (we explain the importance of both properties). In Section 6, we study under which regularity conditions strong duality holds and how it is possible to bound the size of the dual optimal set. This bound will be useful in the convergence analysis of our scheme. In Section 7, an optimal first-order method is applied to modified dual objective function and a nearly feasible and optimal primal solution is reconstructed from the dual minimization sequence. Accuracy of the primal and dual solutions can be adjusted by parameters of our algorithm. In Section 8, we consider applications of double-smoothing technique to optimal control problems. We conclude this paper with a comparison between our results and the existing literature.

2. First-order methods in convex optimization. Consider the convex optimization problem $\min_{y \in V} f(y)$ where $f : V \to R$ is a convex function defined on the finite-dimensional space V. If f is non-differentiable, we know that the complexity of a black-box first-order method that does not use the problem structure cannot be better that $O\left(\frac{1}{\epsilon^2}\right)$ iterations, where ϵ is the desired accuracy for the objective function (see [18, 19]). This lower bound is achievable by various first-order methods for non-smooth convex problems, such as subgradient methods (see e.g. [19, 23]). These schemes can therefore be applied directly to the non-smooth convex function, albeit with a relatively slow convergence rate.

When the non-smooth function has a particular saddle-point structure:

$$f(y) = \max_{u \in S} \{g(u) + \langle Au, y \rangle\} \tag{2.1}$$

where $g : U \to R$ is concave on the finite-dimensional space U and $S \subset U$ is closed and convex, another approach can be used. In the smoothing technique developed in [20, 21, 22], this non-smooth function is approximated by a smooth one and an optimal first-order method of smooth convex optimization is applied to the smooth approximation. With this approach, we can solve the original non-smooth problem up to accuracy ϵ in only $O\left(\frac{1}{\epsilon}\right)$ iterations (instead of $O\left(\frac{1}{\epsilon^2}\right)$ with subgradient scheme). We follow in this paper this smoothing approach, and the efficiency of the first-order method of smooth convex optimization used for minimizing the smoothed approximation plays therefore an important role.

In smooth convex optimization, an important class of objective function is $F_L^{1,1}(V)$ the class of convex function $f : V \to R$ with Lipschitz-continuous gradient i.e:

$$\|\nabla f(x) - \nabla f(y)\|_{V^*} \le L \|x - y\|_V \quad \forall x, y \in V$$

for some $L > 0$. The easiest and the most classical numerical scheme that can be applied to such problem is the gradient method. However, it is well known that this method exhibits non optimal complexity of $O(\frac{L}{\epsilon})$ iterations where ϵ is the desired accuracy for the objective function. Several variants of first-order methods for the class $F_L^{1,1}(Q)$ that achieve the lower complexity bound of $O(\sqrt{\frac{L}{\epsilon}})$ iterations have been known since 1984 [24, 25, 19, 20]. These schemes, also called fast-gradient methods, outperform theoretically and very often in practice the classical gradient method.

In the smoothing approach, we want to solve the original non-smooth function with an accuracy ϵ. We construct a smooth approximation of f belonging to $F^{1,1}_{L(\epsilon)}$ with $L(\epsilon) = \Theta(\frac{1}{\epsilon})$. Applying a fast gradient method for $F^{1,1}_{L(\epsilon)}$ to this smoothed function, we can solve the original problem with the desired accuracy in $O(\frac{1}{\epsilon})$ iterations. In our work, the function that we optimize using a first-order method and that has the form (2.1), is the dual objective function. However, our goal is to solve efficiently the primal problem, not the dual one.

In order to reconstruct a good primal solution from the iterates of the numerical scheme applied to the dual problem, we will need to apply a second smoothing to the dual function before we apply the fast gradient method. Its purpose will be to ensure strong convexity of the resulting dual objective function. Let $S^{1,1}_{\kappa,L}(V)$ be the class of functions $f \in F^{1,1}_L(V)$ which are strongly convex with parameter $\kappa > 0$. Fast gradient methods that are optimal for $S^{1,1}_{\kappa,L}(V)$ are also known (see for example [19]), and we will use such a method to minimize the doubly smoothed dual objective function.

For the reader's convenience, we conclude this section with a presentation of the simplest optimal method for minimizing smooth strongly convex functions. Let function $f : V \to R$ be strongly convex with parameter $\kappa > 0$ and its gradient be Lipschitz-continuous with constant $L > \kappa$. Consider the problem: $\min_{y \in V} f(y)$. We assume that this problem is solvable. Denote by f^* its optimal value and by y^* the optimal solution.

Algorithm ([19]): Choose $w_0 = y_0 \in V$.

(2.2) Iteration ($k \geq 0$): Set $y_{k+1} = w_k - \frac{1}{L}\nabla f(w_k)$, and

$$w_{k+1} = y_{k+1} + \frac{\sqrt{L}-\sqrt{\kappa}}{\sqrt{L}+\sqrt{\kappa}}(y_{k+1} - y_k).$$

By Theorem 2.2.3 in [19] we have:

$$f(y_k) - f^* \leq \left(f(y_0) - f^* + \frac{\kappa}{2}\|y_0 - y^*\|^2_V\right) e^{-k\sqrt{\frac{\kappa}{L}}} \leq 2(f(y_0) - f^*)e^{-k\sqrt{\frac{\kappa}{L}}}. \tag{2.3}$$

Since ∇f is Lipschitz-continuous, in view of Theorem 2.1.5 in [19] we have

$$\frac{1}{2L}\|\nabla f(y_k)\|^2_{V^*} \quad \leq \quad f(y_k) - f^* \quad \stackrel{(2.3)}{\leq} \quad 2(f(y_0) - f^*)e^{-k\sqrt{\frac{\kappa}{L}}}.$$

Therefore,

$$\|\nabla f(y_k)\|^2_{V^*} \quad \leq \quad 4L(f(y_0) - f^*)e^{-k\sqrt{\frac{\kappa}{L}}}. \tag{2.4}$$

Finally, since f is strongly convex, by Theorem 2.1.8 in [19] we have:

$$\frac{\kappa}{2}\|y_k - y^*\|^2_V \quad \leq \quad f(y_k) - f^* \quad \stackrel{(2.3)}{\leq} \quad 2(f(y_0) - f^*)e^{-k\sqrt{\frac{\kappa}{L}}}.$$

Using this inequality and additional arguments, we conclude that

$$\|y_k - y^*\|^2_V \quad \leq \quad \min\left\{\|y_0 - y^*\|^2_V, \frac{4}{\kappa}(f(y_0) - f^*)e^{-k\sqrt{\frac{\kappa}{L}}}\right\}. \tag{2.5}$$

3. Problem formulation and dual approach. As described in the introduction, we consider in this work optimization problems of the form:

$$P^* = \inf_{u \in S} J(u) : \mathcal{A}u \in T. \tag{3.1}$$

where S is a bounded, closed, convex set in U, U is a Hilbert space endowed with the Euclidean norm $\|.\|_U = \sqrt{(.|.)_U}$, T is a bounded, closed, convex set in V^*, the dual space of V, V is a finite-dimensional Hilbert space endowed with the Euclidean norm $\|.\|_V = \sqrt{(.|.)_V}$, $J : S \to R$ is a closed and convex functional and $\mathcal{A} : U \to V^*$ is a bounded linear operator. Space U is allowed to be infinite-dimensional, but the approach used in this paper is also efficient for finite-dimensional problems when $\dim V \ll \dim U$.

Remark 1. *Note that problems with multiple linear constraints also belong to problem class (3.1):*

$$P^* = \inf_{u \in S} J(u) : \mathcal{A}_i u \in T_i \quad \forall i = 1, ..., m. \tag{3.2}$$

Indeed, assume that $V = V_1 \times V_2 \times ... \times V_m$ where V_i is a finite-dimensional Hilbert space for $i = 1, ..., m$. For each i, we can consider a bounded linear operator $\mathcal{A}_i : U \to V_i^*$. Let $T = T_1 \times T_2 \times ... \times T_m$ where T_i is a bounded closed convex set in V_i^*. Defining the linear application $\mathcal{A} : U \to V^*$ such that, for all $u \in U$, $\mathcal{A}u : V \to R, z = (z_1, ..., z_m) \to \langle \mathcal{A}u, z\rangle = \sum_{i=1}^m \langle \mathcal{A}_i u, z_i\rangle$, the constraint $\mathcal{A}u \in T$ is clearly equivalent to $\mathcal{A}_i u \in T_i \quad \forall i = 1, ..., m$. Finally, we have:

$$\begin{aligned} \|\mathcal{A}\| &= \max_{\|u\|_U=1, \|z\|_V=1} \langle \mathcal{A}u, z\rangle = \max_{\|u\|_U=1, \sum_{i=1}^m \|z_i\|^2_{V_i}=1} \sum_{i=1}^m \langle \mathcal{A}_i u, z_i\rangle \\ &= \max_{\|u\|_U=1} [\sum_{i=1}^m \|\mathcal{A}_i u\|_{V_i^*}]^{1/2} \le [\sum_{i=1}^m \|\mathcal{A}_i\|^2]^{1/2}. \end{aligned}$$

Our assumptions on J, S and T are motivated by the following classical result (see for example [27]):

Theorem 3.1. *If X is a reflexive Banach space, $M \subset X$ is a bounded, closed, convex set and $F : X \to R$ is a closed, convex function then the optimization problem $\min_{x \in M} F(x)$ is solvable. Furthermore, if in addition X is an Hilbert space and F is strongly convex, then the optimal solution of this problem is unique.*

We conclude that subproblems of the form:

$$\inf_{u \in S}\{J(u) + \langle \mathcal{A}u, z\rangle\} \qquad \inf_{x \in T}\langle x, z\rangle$$

are solvable for each $z \in V$ and that subproblems of the form:

$$\inf_{u \in S}\{J(u) + \langle \mathcal{A}u, z\rangle + \frac{\mu}{2}\|u\|^2_U\}, \quad \inf_{x \in T}\{\langle x, z\rangle + \frac{\rho}{2}\|x\|^2_{V^*}\}$$

have a unique optimal solution for each $z \in V$, $\mu > 0, \rho > 0$.

In our setting, it is natural to dualize the linear constraint $\mathcal{A}u \in T$ and to consider a dual method, working only in the small dimensional space V. Since T is a closed convex set, inclusion $\mathcal{A}u \in T$ is valid if and only if $\langle \mathcal{A}u, z\rangle \le \sigma_T(z) \quad \forall z \in V$, where $\sigma_T(z) = \sup_{x \in T}\langle x, z\rangle$ denotes the support function of T.
After dualization of the linear constraints, we obtain the primal-dual pair of problems:

$$P^* = \inf_{u \in S}[J(u) + \sup_{z \in V}(\langle \mathcal{A}u, z\rangle - \sigma_T(z))], \quad D^* = \sup_{z \in V}[-\sigma_T(z) + \inf_{u \in S}(J(u) + \langle \mathcal{A}u, z\rangle)].$$

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

$$-D^* = \theta^* = \inf_{z\in V}[\sigma_T(z) + \phi(z) := \theta(z)] \geq -P^* \tag{3.3}$$

where $\phi(z) = \sup_{u\in S}[-J(u) - \langle \mathcal{A}u, z\rangle]$. Due to our assumptions, for each value of z, we can compute easily the value of $\theta(z)$ but this function is typically non-differentiable. Indeed, using the Danskin theorem ([11, 3]), we have:

$$\partial\sigma_T(z) = \{x \in T : \langle x, z\rangle = \sigma_T(z)\}$$
$$\partial\phi(z) = \{-\mathcal{A}u : -J(u) - \langle \mathcal{A}u, z\rangle = \phi(z), u \in T\}.$$

As the optimization problems defining $\sigma_T(z)$ and $\phi(z)$ can have multiple optimal solution, we conclude that function $\theta(z)$ can be non-smooth. Thus, the dualization of problem (3.1) results in a nonsmooth convex problem.

As explained in the previous section, instead of relying subgradient-type schemes with relatively slow convergence, we use a smoothing technique [20, 21, 22]. 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). We develop in this paper an algorithm able to solve the dual problem with accuracy ϵ and to reconstruct, from a 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.

4. Examples. Before we go into the details of the double smoothing, we provide two examples of problems with separable structure (one finite-dimensional, the other infinite-dimensional) that belong to our problems class.

4.1. A finite-dimensional example. Consider the case where:

- $U = R^N = R^{N_1} \times R^{N_2} \times ... \times R^{N_m}$ and $S = S_1 \times S_2 \times ... \times S_m$ where $S_i \subset R^{N_i}$
- $V = R^n$ whith $n \ll N$ and T is a bounded closed convex set in R^n
- $J(u) = \sum_{i=1}^m J_i(u_i)$ where $u_i \in R^{N_i}$ and $J_i : R^{N_i} \to R$ is a closed and convex function
- $\mathcal{A} = (A_1 \quad A_2...A_m) \in R^{n\times N}$ where $A_i \in R^{n\times N_i}$

Our problem becomes $\min \sum_{i=1}^m J_i(u_i) : \sum_{i=1}^m A_i u_i \in T, u_i \in S_i \quad \forall i = 1, ..., m$. This problem has a specific structure that we want to exploit. When the coupling constraint $\sum_{i=1}^m A_i u_i \in T$ is dropped, we obtain a separable problem that we can solve separately for each u_i. With this property, it seems natural to dualize the coupling constraint and to consider the dual problem: $\min_{z\in R^n} \theta(z) = \min_{z\in R^n} \sigma_T(z) + \phi(z)$ in the small-dimensional space R^n. For each $z \in R^n$, the dual objective function can be computed in a pointwise way: solving the subproblem $\max_{u\in S}\{-J(u) - \langle Au, z\rangle\}$ is equivalent to solve for each i, the separate subproblems: $\max_{u_i\in S_i}\{-J_i(u_i) - \langle A_i u_i, z\rangle\}$ that we assume to be easy to solve. The same properties are also satisfied by the modified dual objective function that we will obtain after the smoothing: $\max_{u\in S}\{-J(u) - \langle \mathcal{A}u, z\rangle - \frac{\mu}{2}\|u\|_{R^N}^2\}$ since the euclidean norm $\|u\|_{R^N}$ is also separable: $\|u\|_{R^N}^2 = \sum_{i=1}^m \|u_i\|_{R^{N_i}}^2$.

4.2. An infinite-dimensional example. Consider the case where:

- $U = L^2([0, \mathcal{T}], R^m)$ and $S = \{u \in L^2([0, \mathcal{T}], R^m) : u(t) \in S(t) \forall t \in [0, \mathcal{T}]\}$ where $S(t)$ is a closed, convex set in R^m for each $t \in [0, \mathcal{T}]$ and $\cup_{t\in[0,\mathcal{T}]} S(t)$ is bounded.
- $V = R^n$ and T is a bounded, closed, convex set in R^n
- $J(u) = \int_0^{\mathcal{T}} F(t, u(t))dt$ where the function $F : [0, \mathcal{T}] \times R^m \to R$ is convex and continuous

- $\mathcal{A} : U \to R^n, u \to \int_0^{\mathcal{T}} A(t)u(t)dt$ where $\int_0^{\mathcal{T}} \|A(t)\|_2^2 dt < +\infty$ and $A(t) \in R^{n\times m} \quad \forall t \in [0,\mathcal{T}]$

Our problem becomes:

$$\inf \int_0^{\mathcal{T}} F(t,u(t))dt : \int_0^{\mathcal{T}} A(t)u(t)dt \in T, u(t) \in S(t) \quad \forall t \in [0,\mathcal{T}]. \tag{4.1}$$

When the linear coupling constraint is dropped, we also obtain a separable problem that we can solve separately for each $t \in [0,\mathcal{T}]$: $\min_{u(t)\in S(t)} F(t,u(t))$. The dualization of the linear coupling constraint is here also a natural approach. For each $z \in R^m$, the dual objective function can be computed in a pointwise way. Solving the subproblem: $\sup_{u\in S}\{-J(u) - \langle \mathcal{A}u, z\rangle\}$ is equivalent to solve for each value of $t \in [0,\mathcal{T}]$, the subproblems in $S(t) \subset R^m$: $\max_{v\in S(t)}[-F(t,v) - \langle v, A(t)^T z\rangle]$ that we can solve easily or even in closed form. The same separability property will be also satisfied by the smoothed dual objective function that we obtain using the smoothing technique: $\sup_{u\in S}\{-J(u) - \langle \mathcal{A}u, z\rangle - \frac{\mu}{2}\|u\|_U^2\}$ since this infinite-dimensional problem is equivalent to the pointwise subproblems: $\max_{v\in S(t)}\{-F(t,v) - \langle v, A(t)^T z\rangle - \frac{\mu}{2}\|v\|_{R^m}^2\}$.

5. Double Smoothing Technique. We will try to solve the dual problem (3.3) 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. This property gives us a possibility to use nearly optimal dual solutions to reconstruct efficiently a nearly feasible and optimal primal solution.

5.1. First Smoothing. 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. We have seen that the non-smoothness of θ comes from the fact that the optimization problems defining $\sigma_T(z)$ and $\phi(z)$ at a given point z can have multiple optimal solutions. A natural way for obtaining a smooth approximation of θ is to modify these optimization subproblems in order to ensure the uniqueness of optimal solutions for each $z \in V$. For $\rho > 0$, we can approximate $\sigma_T(z) = \sup_{x\in T}\langle x, z\rangle$ by a modified function:

$$\sigma_{\rho,T}(z) = \sup_{x\in T}\{\langle x, z\rangle - \frac{\rho}{2}\|x\|_{V^*}^2\}. \tag{5.1}$$

In the same way, for $\mu > 0$, we modify the function $\phi(z)$ as follows:

$$\phi_\mu(z) = \sup_{u\in S}\{-J(u) - \langle \mathcal{A}u, z\rangle - \frac{\mu}{2}\|u\|_U^2\}. \tag{5.2}$$

The following result can be seen as an easy generalization of Theorem 1 in [20]:

Theorem 5.1. *Let H_1, H_2 be two Hilbert spaces. Assume that the linear operator $A : H_1 \to H_2^*$ is bounded and that the function $G : H_1 \to R$ is closed and strongly convex with parameter κ. Let $Q \subset \mathrm{dom}\,(G)$ be a closed, convex set. Then the function*

$$F(z) = \sup_{u\in Q}\{-G(u) - \langle \mathcal{A}u, z\rangle\} \tag{5.3}$$

is smooth with Lipschitz-continuous gradient $\nabla F(z) = -\mathcal{A}u_z$ where u_z is the unique optimal solution of the optimization problem defining $F(z)$. The Lipschitz constant of *the gradient is given by:* $\frac{\|A\|^2}{\kappa}$ *where* $\|A\| = \sup\{\langle \mathcal{A}u, z\rangle : u \in H_1, \|u\|_{H_1} = 1, z \in H_2, \|z\|_{H_2} = 1\}$.

Choosing now $H_1 = U$, $Q = S$, $H_2 = V$, $A = \mathcal{A}$ and $G(u) = J(u) + \frac{\mu}{2}\|u\|_U^2$, by Theorem 5.1, we conclude that ϕ_μ is smooth and convex with Lipschitz continuous gradient: $\nabla\phi_\mu(z) = -\mathcal{A}u_{\mu,z}$ where $u_{\mu,z}$ denotes the unique optimal solution of the problem (5.2). The Lipschitz constant of this gradient is given by $L(\phi_\mu) = \frac{\|\mathcal{A}\|^2}{\mu}$. On the other hand, if we choose $H_1 = V^*$, $Q = T$, $H_2 = V$, $A = I : V^* \to V^*$ and $G(x) = \frac{\rho}{2}\|x\|_{V^*}^2$, by Theorem 5.1 we conclude that $\sigma_{\rho,T}$ is smooth and convex with Lipschitz-continuous gradient : $\nabla\sigma_{\rho,T}(z) = x_{\rho,z}$ where $x_{\rho,z}$ denotes the unique optimal solution of the problem (5.1). The Lispchitz constant of this gradient is given by $L(\sigma_{\rho,T}) = \frac{1}{\rho}$.

Remark 2. *When the function $J(u)$ is strongly convex with parameter κ, we do not need to apply the first smoothing to $\phi(z)$ which is aready smooth in this case with a Lipschitz-continuous gradient with constant* $\frac{\|A\|^2}{\kappa}$.

Denote $D_T = \max\{\frac{1}{2}\|x\|_{V^*}^2 : x \in T\}$ and $D_S = \max\{\frac{1}{2}\|u\|_U^2 : u \in S\}$. Then:

$$\sigma_{\rho,T}(z) \le \sigma_T(z) \le \sigma_{\rho,T}(z) + \rho D_T \quad \forall z \in V, \quad \phi_\mu(z) \le \phi(z) \le \phi_\mu(z) + \mu D_S \quad \forall z \in V.$$

Therefore, if we define the function $\theta_{\rho,\mu}(z) = \sigma_{\rho,T}(z) + \phi_\mu(z)$ we have $\theta_{\rho,\mu} \in F^{1,1}_{L(\rho,\mu)}(V)$ with $L(\rho,\mu) = \frac{1}{\rho} + \frac{\|\mathcal{A}\|^2}{\mu}$ and

$$\theta_{\rho,\mu}(z) \le \theta(z) \le \theta_{\rho,\mu}(z) + \mu D_S + \rho D_T \quad \forall z \in V. \tag{5.4}$$

Applying a fast-gradient method to the function $\theta_{\rho,\mu}$, we know (see [20]) that we can generate a point $z_\epsilon \in V$ such that $\theta(z_\epsilon) - \theta^* \le \epsilon$ in $O\left(\frac{1}{\epsilon}\right)$ iterations. However, our aim is not only to solve the dual problem efficiently but also to generate a nearly optimal and nearly feasible solution for the primal problem. We will see that a single smoothing is not enough in order to achieve this goal. Let us show how it is possible using a dual iterate z to reconstruct a primal solution with good accuracy. Let $z \in V$, we have:

$$\theta_{\rho,\mu}(z) = \sigma_{\rho,T}(z) + \phi_\mu(z) = \langle x_{\rho,z}, z\rangle - \frac{\rho}{2}\|x_{\rho,z}\|_{V^*}^2 - J(u_{\mu,z}) - \langle \mathcal{A}u_{\mu,z}, z\rangle - \frac{\mu}{2}\|u_{\mu,z}\|_U^2.$$

Let us find the conditions that z must satisfy in order to guarantee that $u_{\mu,z}$ is nearly optimal and nearly feasible for the primal problem. We have:

$$\begin{aligned} J(u_{\mu,z}) - D^* &= \langle x_{\rho,z}, z\rangle - \frac{\rho}{2}\|x_{\rho,z}\|_{V^*}^2 - \theta_{\rho,\mu}(z) - \langle \mathcal{A}u_{\mu,z}, z\rangle - \frac{\mu}{2}\|u_{\mu,z}\|_U^2 + \theta^* \\ &= \langle \nabla\theta_{\rho,\mu}(z), z\rangle + (\theta^* - \theta_{\rho,\mu}(z)) - \frac{\rho}{2}\|x_{\rho,z}\|_{V^*}^2 - \frac{\mu}{2}\|u_{\mu,z}\|_U^2. \end{aligned}$$

Therefore:

$$|J(u_{\mu,z}) - D^*| \le |\langle \nabla\theta_{\rho,\mu}(z), z\rangle| + |\theta_{\rho,\mu}(z) - \theta^*| + \rho D_T + \mu D_S.$$

Furthermore as: $P^* \ge D^*$ and $|\theta_{\rho,\mu}(z) - \theta^*| \le |\theta(z) - \theta^*| + \mu D_S + \rho D_T$, we conclude that:

$$J(u_{\mu,z}) \le P^* + |\langle \nabla\theta_{\rho,\mu}(z), z\rangle| + |\theta^* - \theta(z)| + 2\rho D_T + 2\mu D_S.$$

If we apply the fast-gradient method to the function $\theta_{\rho,\mu}$ with μ and ρ choosen of order $\frac{1}{k}$, we know ([20]) that the kth iterate generated by this algorithm z_k satisfies: $\theta(z_k)-\theta^* \leq \frac{C_1}{k}$. However, the norm of the gradient of the smoothed function $\|\nabla\theta_{\rho,\mu}(z_k)\|_V$ does not decrease at the same rate. Indeed, as $\theta_{\rho,\mu} \in F^{1,1}_{L(\rho,\mu)}(V)$, we have (see Theorem 2.1.5 in [19]):

$$\|\nabla\theta_{\rho,\mu}(z_k)\|^2_{V^*} \leq 2L(\rho,\mu)(\theta_{\rho,\mu}(z_k) - \theta^*_{\rho,\mu}).$$

As the fast-gradient method is applied to the function $\theta_{\rho,\mu}$, we have ([20]) also:

$$\theta_{\rho,\mu}(z_k) - \theta^*_{\rho,\mu} \leq \frac{4L(\rho,\mu)\,\|z_0 - z^*_S\|}{(k+1)(k+2)},$$

where z^*_S denotes any optimal solution of the smoothed dual problem $\min_{z\in V}\theta_{\rho,\mu}(z)$. Therefore

$$\|\nabla\theta_{\rho,\mu}(z_k)\|_{V^*} \leq \frac{2\sqrt{2}L(\rho,\mu)\sqrt{\|z_0 - z^*_S\|}}{\sqrt{(k+1)(k+2)}}.$$

Due to the fact that $L(\rho,\mu)$ is of order k, we cannot guarantee that the norm of the gradient $\|\nabla\theta_{\rho,\mu}(z_k)\|_{V^*}$ is decreasing with respect to k. With a minor modification of the scheme, we can obtain in $2k$ iterations a point $\tilde{z}$ such that $\|\nabla_{\rho,\mu}\theta(\tilde{z})\|_{V^*}$ is of order $O(\frac{1}{\sqrt{k}})$. Indeed if we apply after the k steps of the fast gradient method, k other steps but now using the classical gradient method with constant stepsize $\frac{1}{L(\rho,\mu)}$ i.e.:

$$z_{k+i} = z_{k+i-1} - \frac{1}{L(\rho,\mu)}\nabla\theta_{\rho,\mu}(z_{k+i-1}).$$

We have (see Theorem 2.1.14 in [19]):

$$\frac{1}{2L(\rho,\mu)}\,\|\nabla\theta_{\rho,\mu}(z_{k+i-1})\|^2_{V^*} \leq \theta_{\rho,\mu}(z_{k+i-1}) - \theta_{\rho,\mu}(z_{k+i}) \quad \forall i = 1,...,k.$$

Summing these inequalities:

$$\begin{aligned}\sum_{i=0}^{k-1}\frac{1}{2L(\rho,\mu)}\,\|\theta_{\rho,\mu}(z_{k+i})\|^2_{V^*} \leq \theta_{\rho,\mu}(z_k) - \theta_{\rho,\mu}(z_{2k}) &\leq \theta_{\rho,\mu}(z_k) - \theta^*_{\rho,\mu}\\ &\leq \frac{4L(\rho,\mu)\,\|z_0 - z^*_S\|^2_V}{(k+1)(k+2)}.\end{aligned}$$

If we denote by $\tilde{z}$, the iterate with the smallest norm of the gradient, we conclude that:

$$\|\nabla\theta_{\rho,\mu}(\tilde{z})\|^2_{V^*} = \min_{i=0,\ldots,k-1}\|\nabla\theta_{\rho,\mu}(z_{k+i})\|^2_{V^*} \leq \frac{8L^2(\rho,\mu)\,\|z_0 - z^*_S\|^2_V}{k(k+1)(k+2)}.$$

In conclusion, after $2k$ iterates, we are able mixing fast and classical gradient method, to obtain a point $\tilde{z}$ such that $\|\nabla\theta_{\rho,\mu}(\tilde{z})\|_{V^*} = O(\frac{1}{\sqrt{k}})$. However, this convergence is very slow. Therefore we need at least $k = O(\frac{1}{\epsilon^2})$ iterations in order to have a primal solution $u_{\mu,z_k} \in S$ with accuracy ϵ (i.e. $J(u_{\mu,z_k}) \leq P^* + \epsilon$). This is not better than the result of subgradient approach. Furthermore, the norm of $\|\nabla\theta_{\rho,\mu}(\tilde{z})\|_{V^*}$ gives also an upper-bound for the non admissibility measure of $u_{\mu,\tilde{z}}$. Indeed, we have:

$$d(\mathcal{A}u_{\mu,\tilde{z}},T) \le \|\mathcal{A}u_{\mu,\tilde{z}} - x_{\rho,\tilde{z}}\|_V = \|\nabla\theta_{\rho,\mu}(\tilde{z})\|_V .$$

In order to obtain efficiently a nearly feasible and optimal solution for the primal problem, having a good convergence for $\theta(z_k)-\theta^*$ to zero is not sufficient. We need also to have the same good rate for $\|\nabla\theta_{\rho,\mu}(z_k)\|_V$ to 0. A way to obtain this good property is to apply a second smoothing to the dual objective function, making it also strongly convex.

5.2. Second Smoothing. In order to obtain a strongly convex dual objective function, we just add the strongly convex function $\frac{\kappa}{2}\|z\|_V^2$ to the function $\theta_{\rho,\mu}$. This gives us a new dual objective function:

$$\theta_{\rho,\mu,\kappa}(z) = \sigma_{\rho,T}(z) + \phi_\mu(z) + \frac{\kappa}{2}\|z\|_V^2 ,$$

which is strongly convex with parameter κ. If we denote by $B = B^* : V \to V^*$ (with $B \succ 0$) the duality map between V and its dual space i.e.: $\langle Bz, \overline{z}\rangle = (z|\overline{z})_V \quad \forall z, \overline{z} \in V$, we have: $\nabla\theta_{\rho,\mu,\kappa}(z) = x_{\rho,z} - \mathcal{A}u_{\mu,z} + \kappa Bz$. This gradient is Lipschitz-continuous with constant $L(\rho,\mu,\kappa) = \frac{1}{\rho} + \frac{\|\mathcal{A}\|^2}{\mu} + \kappa$. This function is therefore in $S^{1,1}_{\kappa,L(\rho,\mu,\kappa)}(V)$. Denote by $\theta^*_{\rho,\mu,\kappa}$ the optimal solution of the problem, $\min_{z\in S}\theta_{\rho,\mu,\kappa}(z)$ and by z^*_{DS} the optimal solution of this problem. Applying the fast-gradient method for the class $S^{1,1}_{\kappa,L(\rho,\mu),\kappa}$ to the function $\theta_{\rho,\mu,\kappa}$, we generate a sequence z_k satisfying:

$$\theta_{\rho,\mu,\kappa}(z_k) - \theta^*_{\rho,\mu,\kappa} \overset{(2.3)}{\le} \exp\left(-k\sqrt{\frac{\kappa}{L(\rho,\mu,\kappa)}}\right) 2(\theta_{\rho,\mu,\kappa}(z_0) - \theta^*_{\rho,\mu,\kappa})$$

and

$$\|\nabla\theta_{\rho,\mu,\kappa}(z_k)\|^2_{V^*} \overset{(2.4)}{\le} 4L(\rho,\mu,\kappa)(\theta_{\rho,\mu,\kappa}(z_0) - \theta^*_{\rho,\mu,\kappa}) \exp\left(-k\sqrt{\frac{\kappa}{L(\rho,\mu,\kappa)}}\right)$$

i.e.:

$$\|\nabla\theta_{\rho,\mu,\kappa}(z_k)\|_{V^*} \le 2\sqrt{L(\rho,\mu,\kappa)}\sqrt{\theta_{\rho,\mu,\kappa}(z_0) - \theta^*_{\rho,\mu,\kappa}} \exp\left(\frac{-k}{2}\sqrt{\frac{\kappa}{L(\rho,\mu,\kappa)}}\right)$$

and we have the same rate of convergence for $\|\nabla\theta_{\rho,\mu,\kappa}(z_k)\|_{V^*}$ that for $\theta_{\rho,\mu,\kappa}(z_k) - \theta^*_{\rho,\mu,\kappa}$. This property is crucial in order to obtain a nearly feasible and optimal primal solution in $O\left(\frac{1}{\epsilon}\ln\left(\frac{1}{\epsilon}\right)\right)$ iterations (instead of $O\left(\frac{1}{\epsilon^2}\right)$ with a simple smoothing).

6. Strong duality and norm of dual optimal solutions. Before to apply the fast gradient method to the obtained double smoothed dual function, we study on this section under which condition strong duality i.e $P^* = D^*$ holds and how it is possible to bound the size of the optimal solution set of the dual problem (3.3) . Such bound will be useful in the convergence analysis of our scheme, as we will see in the following section.

THEOREM 6.1. *If there exists $r > 0$ such that*

$$B(0,r) \subset Q := \{s = x - \mathcal{A}u \in V^* : u \in S, x \in T\} \tag{6.1}$$

then:

- *There is no duality gap: $P^* = D^*$*
- *The optimal solution set of the dual problem (3.3) is a nonempty, bounded, closed and convex set in V*

Furthermore if $\Delta(J') := \max_{u,v\in S} J'(v, u - v) < +\infty$, we have the following upper-bound for the norm of the dual optimal solutions:

$$\|z^*\|_V \leq \frac{\Delta(J')}{r}.$$

Proof. Applying Theorem 2.165 in [4] to the primal problem: $\min_{u\in U}\{f(u) = J(u) + I_S(u) : G(u) = Au \in T\}$ (where I_S denotes the indicator function of S), we conclude that, with our assumptions on J, A, S and T and with the regularity condition (6.1), there is no duality gap between this problem and its Lagrangian dual (3.3). Furthermore, as the primal optimal value P^* is assumed to be finite, we have that the optimal solution set of the dual problem is a non-empty, bounded, closed and convex set in V.

It remains to obtain the bound $\|z^*\|_V \leq \frac{\Delta(J')}{r}$. As the subproblems $\sup_{x\in T}\langle x, z\rangle$ and $\sup_{u\in S}\{-J(u) - \langle \mathcal{A}u, z\rangle\}$ are solvable for all $z \in V$,

$$\partial\theta(z) = \{x_z - \mathcal{A}u_z \in V^* : x_z \in T, \langle x_z, z\rangle = \sigma_T(z), u_z \in S, -J(u_z) - \langle \mathcal{A}u_z, z\rangle = \phi(z)\}$$

is non-empty for all $z \in V$. Let $g_z = x_z - \mathcal{A}u_z$ be any element in $\partial\theta(z)$. By the optimality condition of the problems defining $\sigma_T(z)$ and $\phi(z)$:

$$\langle x - x_z, z\rangle \leq 0 \quad \forall x \in T \quad -J'(u_z, u - u_z) - \langle \mathcal{A}(u - u_z), z\rangle \leq 0 \quad \forall u \in S.$$

We have:

$$\langle x_z, z\rangle \geq \langle x, z\rangle \quad \forall x \in T \quad -\langle \mathcal{A}u_z, z\rangle \geq -J'(u_z, u - u_z) + \langle -\mathcal{A}u, z\rangle \quad \forall u \in S.$$

Therefore:

$$\langle g_z, z\rangle = \langle x_z - \mathcal{A}u_z, z\rangle \geq \langle x, z\rangle - J'(u_z, u - u_z) + \langle -\mathcal{A}u, z\rangle = \langle x - \mathcal{A}u, z\rangle - J'(u_z, u - u_z).$$

We obtain:

$$\langle x - Au, z\rangle \leq J'(u_z, u - u_z) + \langle g_z, z\rangle \quad \forall u \in S, \quad \forall x \in T$$

and therefore:

$$\max_{u\in S, x\in T}\langle x - \mathcal{A}u, z\rangle \leq \max_{u\in S} J'(u_z, u - u_z) + \langle g_z, z\rangle \leq \max_{u,v\in S} J'(v, u - v) + \langle g_z, z\rangle$$

As $B(0, r) \subset \{v = x - \mathcal{A}u \in V^* : u \in S, x \in T\}$, we have: $\max_{u\in S, x\in T}\langle x - \mathcal{A}u, z\rangle \geq r\,\|z\|_V$ and therefore:

$$r\,\|z\|_V \leq \max_{u,v\in S} J'(v, u - v) + \langle g_z, z\rangle \quad \forall z \in V. \tag{6.2}$$

Consider now an optimal solution z^* of the dual problem. By the optimality condition of this problem, we have: $0 \in \partial\theta(z^*)$ and therefore $\|z^*\|_V \leq \frac{\Delta(J')}{r}$. □

REMARK 3. *As J is convex, we have: $J(u) \geq J(v) + J'(v, u - v) \quad \forall u, v \in S$ and therefore:*

$$\max_{u,v\in S} J(u) - J(v) = \max_{u\in S} J(u) - \min_{v\in S} J(v) \geq \max_{u,v\in S} J'(v, u - v).$$

The condition $\max_{u,v\in S} J'(v,u-v) < +\infty$ *is therefore weaker than imposing a bounded variation of* J *on* S*:* $\max_{u\in S} J(u) - \min_{v\in S} J(v) < +\infty$.
Indeed, consider the function:

$$J(u) = \int_0^1 \ln(1-u^2(t))dt$$

on $S = \{u \in L^2([0,1]) : -1 \leq u(t) \leq 1 \quad a.e \quad in \quad [0,1]\}$*. Clearly* $\max_{u\in S} J(u) - \min_{v\in S} J(v) = +\infty$ *and this functionl can be seen as as barrier function for* S*. However we have:*

$$J'(v,u-v) = \int_0^1 \frac{2(u(t)-v(t))v(t)}{1-v^2(t)}dt$$

and $\Delta(J') = max_{u,v\in S'} J'(v,u-v) \leq 2$.

Remark 4. *If the primal problem is feasible, it is clear that there exist* $\overline{u} \in S$ *and* $\overline{x} \in T$ *such that* $\mathcal{A}\overline{u} = \overline{x}$ *i.e* $0 \in Q$*. In order to have* $B(0,r) \subset Q$ *with* $r > 0$*, one of the following extra assumption is enough:*

- ***The set*** T ***has a non-empty interior and there exists*** $\overline{u} \in S$ ***such that*** $\mathcal{A}\overline{u} = \overline{x} \in intT$ ***(generalized Slater condition).***
  As $\overline{x} \in intT$*, there exists* $r > 0$ *such that:* $\overline{x} + B(0,r) \subset T$ *and therefore* $B(0,r) = \overline{x} - \mathcal{A}\overline{u} + B(0,r) \subset Q$.
- ***The set*** S ***has a non-empty interior,*** $\mathcal{A} : U \to V^*$ ***is surjective and there exists*** $\overline{u} \in intS$ ***such that*** $\mathcal{A}\overline{u} = \overline{x} \in T$***.***
  Indeed in this case, as $\overline{u} \in intS$*, there exists* $\tilde{r} > 0$ *such that* $B(\overline{u},\tilde{r}) \subset S$*. By the Banach-Schauder theorem, the image of any open subset of* U *by* $\mathcal{A}$ *is an open subset in* V^**. Therefore, there exists* $r > 0$ *such that* $\mathcal{A}\overline{u} + B(0,r) \subset \mathcal{A}(B(\overline{u},\tilde{r})) \subset \mathcal{A}(S)$*. We conclude that* $\overline{x} - \mathcal{A}\overline{u} + B(0,r) = B(0,r) \subset Q$.

7. Solving the primal-dual problem in $O(\frac{1}{\epsilon}\ln\left(\frac{1}{\epsilon}\right))$ iterations . Denote by z^*_{DS} the unique optimal solution of the problem

$$\min_{z\in V} \theta_{\rho,\mu,\kappa}(z), \tag{7.1}$$

and by z^* one of the optimal solutions of the dual problem (3.3). We assume that the upper bound

$$\|z^*\|_V \quad \leq \quad D_D \tag{7.2}$$

is available. This can be ensured by very natural assumptions on S,T and J using theorem 6.1.
If we apply the method (2.2) to double smoothed dual problem with starting point $z_0 = 0$, we obtain a sequence $\{z_k\}$ such that:

$$\begin{aligned} \theta_{\rho,\mu,\kappa}(z_k) - \theta_{\rho,\mu,\kappa}(z^*_{DS}) &\leq 2(\theta_{\rho,\mu,\kappa}(0) - \theta_{\rho,\mu,\kappa}(z^*_{DS}))e^{-k\sqrt{\frac{\kappa}{L(\rho,\mu,\kappa)}}}, \\ \|\nabla\theta_{\rho,\mu,\kappa}(z_k)\|^2_{V^*} &\leq 4L(\rho,\mu,\kappa)(\theta_{\rho,\mu,\kappa}(0) - \theta_{\rho,\mu,\kappa}(z^*_{DS}))e^{-k\sqrt{\frac{\kappa}{L(\rho,\mu,\kappa)}}}, \\ \|z_k - z^*_{DS}\|^2_{V^*} &\leq \min\left\{\|z^*_{DS}\|^2_{V^*}, \tfrac{4}{\kappa}(\theta_{\rho,\mu,\kappa}(0) - \theta_{\rho,\mu,\kappa}(z^*_{DS}))e^{-k\sqrt{\frac{\kappa}{L(\rho,\mu,\kappa)}}}\right\}. \end{aligned} \tag{7.3}$$

7.1. Convergence of $\theta(z_k)$ to θ^*. Since $\theta_{\rho,\mu,\kappa}(0) = \theta_{\rho,\mu}(0)$ and $\theta_{\rho,\mu,\kappa}(z^*_{DS}) = \theta_{\rho,\mu}(z^*_{DS}) + \frac{\kappa}{2}\|z^*_{DS}\|^2_V$, we have

$$\begin{aligned} \tfrac{\kappa}{2}\|z^*_{DS}\|^2_V &\leq \theta_{\rho,\mu,\kappa}(0) - \theta_{\rho,\mu,\kappa}(z^*_{DS}) = \theta_{\rho,\mu}(0) - \theta_{\rho,\mu}(z^*_{DS}) - \tfrac{\kappa}{2}\|z^*_{DS}\|^2_V, \\ \|z_k - z^*_{DS}\|^2_V &\overset{(2.3)}{\leq} \tfrac{4}{\kappa}(\theta_{\rho,\mu}(0) - \theta_{\rho,\mu}(z^*_{DS}))e^{-k\sqrt{\frac{\kappa}{L(\rho,\mu,\kappa)}}}. \end{aligned} \tag{7.4}$$

Note that

$$\theta_{\rho,\mu}(z_k) - \theta_{\rho,\mu}(z^*_{DS}) \overset{(2.3)}{\leq} (\theta_{\rho,\mu}(0) - \theta_{\rho,\mu}(z^*_{DS}))e^{-k\sqrt{\frac{\kappa}{L(\rho,\mu,\kappa)}}} + \frac{\kappa}{2}(\|z^*_{DS}\|^2_V - \|z_k\|^2_V).$$

On the other hand,

$$\begin{aligned}
\|z^*_{DS}\|^2_V - \|z_k\|^2_V &\leq \|z^*_{DS} - z_k\|_V(\|z^*_{DS}\|_V + \|z_k\|_V) \\
&\leq \|z^*_{DS} - z_k\|_V(2\|z^*_{DS}\|_V + \|z_k - z^*_{DS}\|_V) \\
&\overset{(2.5)}{\leq} 3\|z^*_{DS} - z_k\|_V \cdot \|z^*_{DS}\|_V \\
&\overset{(7.4)}{\leq} 3 \cdot \|z^*_{DS}\|_V \sqrt{\frac{4}{\kappa}(\theta_{\rho,\mu}(0) - \theta_{\rho,\mu}(z^*_{DS}))}e^{-\frac{k}{2}\sqrt{\frac{\kappa}{L(\rho,\mu,\kappa)}}} \\
&\overset{(7.4)}{\leq} \frac{6}{\kappa}(\theta_{\rho,\mu}(0) - \theta_{\rho,\mu}(z^*_{DS}))e^{-\frac{k}{2}\sqrt{\frac{\kappa}{L(\rho,\mu,\kappa)}}},
\end{aligned}$$

and therefore

$$\theta_{\rho,\mu}(z_k) - \theta_{\rho,\mu}(z^*_{DS}) \leq 4(\theta_{\rho,\mu}(0) - \theta_{\rho,\mu}(z^*_{DS}))e^{-\frac{k}{2}\sqrt{\frac{\kappa}{L(\rho,\mu,\kappa)}}}.$$

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

$$\theta_{\rho,\mu}(z^*_{DS}) \geq \theta(z^*_{DS}) - \rho D_T - \mu D_S \geq \theta(z^*) - \rho D_T - \mu D_S.$$

Therefore,

$$\theta_{\rho,\mu}(0) - \theta_{\rho,\mu}(z^*_{DS}) \leq \theta(0) - \theta(z^*) + \rho D_T + \mu D_S. \tag{7.5}$$

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

$$\theta_{\rho,\mu}(z^*_{DS}) \leq \theta_{\rho,\mu}(z^*) + \frac{\kappa}{2}\|z^*\|^2_V \overset{(5.4)}{\leq} \theta(z^*) + \frac{\kappa}{2}\|z^*\|^2_V,$$

and therefore

$$\theta_{\rho,\mu}(z_k) - \theta_{\rho,\mu}(z^*_{DS}) \overset{(5.4)}{\geq} \theta(z_k) - \mu D_S - \rho D_T - \theta(z^*) - \frac{\kappa}{2}\|z^*\|^2_V.$$

In conclusion, we have

$$\theta(z_k) - \theta(z^*) \leq \mu D_S + \rho D_T + \frac{\kappa}{2}D^2_D + 4\left(\theta(0) - \theta(z^*) + \rho D_T + \mu D_S\right)e^{-\frac{k}{2}\sqrt{\frac{\kappa}{L(\rho,\mu,\kappa)}}}. \tag{7.6}$$

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:

$$\mu = \mu(\epsilon) = \frac{\epsilon}{4D_S}, \quad \rho = \rho(\epsilon) = \frac{\epsilon}{4D_T}, \quad \kappa = \kappa(\epsilon) = \frac{\epsilon}{2D^2_D}. \tag{7.7}$$

Under this choice we get

$$\theta(z_k) - \theta(z^*) \leq \frac{3\epsilon}{4} + 4\left(\theta(0) - \theta(z^*) + \frac{\epsilon}{2}\right)e^{-\frac{k}{2}\sqrt{\frac{\kappa}{L(\rho,\mu,\kappa)}}}. \tag{7.8}$$

The last term in the estimate (7.8) defines the number of iterations needed for reaching the accuracy ϵ. Clearly, we ensure $4\left(\theta(0) - \theta(z^*) + \frac{\epsilon}{2}\right)e^{-\frac{k}{2}\sqrt{\frac{\kappa}{L(\rho,\mu,\kappa)}}} \leq \frac{\epsilon}{4}$ by taking

$$k \geq \sqrt{\frac{L(\rho,\mu,\kappa)}{\kappa}} \ln \frac{16\left(\theta(0) - \theta(z^*) + \frac{\epsilon}{2}\right)}{\epsilon}. \tag{7.9}$$

It remains to note that

$$\frac{L(\rho,\mu,\kappa)}{\kappa} = 1 + \frac{1}{\rho\kappa} + \frac{1}{\mu\kappa}\|\mathcal{A}\|^2 \overset{(7.7)}{=} 1 + \frac{8}{\epsilon^2}\left[D_T + D_S\|\mathcal{A}\|^2\right]D^2_D. \tag{7.10}$$

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

7.2. Convergence of $\|\nabla\theta_{\rho,\mu}(z_k)\|_{V^*}$. In our approach, we want to able to reconstruct a nearly optimal and feasible primal solution efficiently. In Section 4.1, we have seen that the accuracy of this primal solution depends not only on the rate of convergence for the dual objective function, but also on the rate of convergence of the norm of its gradient. Let us give an upper bound for the number of iterations needed to drop this norm below a certain level.

We have

$$\begin{aligned}\|\nabla\theta_{\rho,\mu}(z_k)\|_{V^*} &\leq \|\nabla\theta_{\rho,\mu,\kappa}(z_k)-\kappa\beta z_k\|_{V^*} \leq \|\nabla\theta_{\rho,\mu,\kappa}(z_k)\|_{V^*}+\kappa\|\beta z_k\|_{V^*}\\ &= \|\nabla\theta_{\rho,\mu,\kappa}(z_k)\|_{V^*}+\kappa\|z_k\|_V \overset{(7.3)}{\leq} \|\nabla\theta_{\rho,\mu,\kappa}(z_k)\|_{V^*}+2\kappa\,\|z^*_{DS}\|_V\,.\end{aligned}$$

Note that

$$\begin{aligned}\tfrac{1}{4L(\rho,\mu,\kappa)}\,\|\nabla\theta_{\rho,\mu,\kappa}(z_k)\|^2_{V^*} &\overset{(7.3),(7.4)}{\leq} (\theta_{\rho,\mu}(0)-\theta_{\rho,\mu}(z^*_{DS}))e^{-k\sqrt{\frac{\kappa}{L(\rho,\mu,\kappa)}}}\\ &\overset{(7.5)}{\leq} (\theta(0)-\theta(z^*)+\mu D_S+\rho D_T)e^{-k\sqrt{\frac{\kappa}{L(\rho,\mu,\kappa)}}}\\ &\overset{(7.7)}{=} (\theta(0)-\theta(z^*)+\tfrac{\epsilon}{2})e^{-k\sqrt{\frac{\kappa}{L(\rho,\mu,\kappa)}}}.\end{aligned}$$

At the same time,

$$\begin{aligned}\theta(z^*)+\tfrac{\kappa}{2}\,\|z^*\|^2_V &\overset{(5.4)}{\geq} \theta_{\rho,\mu}(z^*)+\tfrac{\kappa}{2}\,\|z^*\|^2_V \;\geq\; \theta_{\rho,\mu}(z^*_{DS})+\tfrac{\kappa}{2}\,\|z^*_{DS}\|^2_V\\ &\overset{(5.4)}{\geq} \theta(z^*_{DS})-\mu D_S-\rho D_T+\tfrac{\kappa}{2}\,\|z^*_{DS}\|^2_V\\ &\geq \theta(z^*)-\mu D_S-\rho D_T+\tfrac{\kappa}{2}\,\|z^*_{DS}\|^2_2\,.\end{aligned}$$

Hence,

$$\|z^*_{DS}\|_V \leq \sqrt{\|z^*\|^2_2+\tfrac{2\mu}{\kappa}D_S+\tfrac{2\rho}{\kappa}D_T} \overset{(7.7)}{\leq} \kappa^{-1/2}\sqrt{\tfrac{3\epsilon}{2}} \overset{(7.7)}{=} \sqrt{3}D_D, \tag{7.11}$$

and we obtain:

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

Taking into account (7.7), we can see that in $k(\epsilon)=O(\frac{1}{\epsilon}\ln\frac{1}{\epsilon})$ iterations, we can ensure

$$\theta(z_k)-\theta(z^*)\leq\epsilon, \quad \|\nabla\theta_{\rho,\mu}(z_k)\|_{V^*}\leq\tfrac{2\epsilon}{D_D}. \tag{7.12}$$

7.3. 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 $\widehat{u}_{k(\epsilon)}\in S$ such that

$$\left|J(\widehat{u}_{k(\epsilon)})-D^*\right| \leq 2\left(1+2\sqrt{3}\right)\cdot\epsilon, \tag{7.13}$$

$$d(\mathcal{A}\widehat{u}_{k(\epsilon)},T)\leq\frac{2\epsilon}{D_D} \tag{7.14}$$

Since $D^*\leq P^*$, inequality (7.13) implies $J(\widehat{u}_{k(\epsilon)})dt\leq P^*+2\left(1+2\sqrt{3}\right)\cdot\epsilon$. Thus $\hat{u}_{k(\epsilon)}$ satisfying (7.13), (7.14) 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 optimization problem defining $\phi_{\mu(\epsilon)}(z_{k(\epsilon)})$. We have

$$\begin{array}{rcl} \theta_{\rho(\epsilon),\mu(\epsilon)}(z_{k(\epsilon)}) & = & \sigma_{\rho(\epsilon),Q}(z_{k(\epsilon)}) + \phi_\mu(z_{k(\epsilon)}) \\ & = & \langle x_{\rho(\epsilon),z_{k(\epsilon)}}, z_{k(\epsilon)}\rangle - \frac{\rho(\epsilon)}{2}\left\| x_{\rho(\epsilon),z_{k(\epsilon)}}\right\|^2_{V^*} - J(\widehat{u}_{k(\epsilon)}) \\ & & -\langle \mathcal{A}\widehat{u}_{k(\epsilon)}, z_{k(\epsilon)}\rangle - \frac{\mu(\epsilon)}{2}\left\|\widehat{u}_{k(\epsilon)}\right\|^2_U . \end{array}$$

Therefore,

$$\begin{array}{rcl} J(\widehat{u}_{k(\epsilon)}) - D^* & = & \langle x_{\rho(\epsilon),z_{k(\epsilon)}} - \mathcal{A}\widehat{u}_{k(\epsilon)}, z_{k(\epsilon)}\rangle - \frac{\rho(\epsilon)}{2}\left\| x_{\rho(\epsilon),z_{k(\epsilon)}}\right\|^2_{V^*} \\ & & -\frac{\mu(\epsilon)}{2}\left\|\widehat{u}_{k(\epsilon)}\right\|^2_U - \theta_{\rho(\epsilon),\mu(\epsilon)}(z_{k(\epsilon)}) + \theta(z^*). \end{array}$$

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

$$\begin{array}{rcl} \theta_{\rho(\epsilon),\mu(\epsilon)}(z_{k(\epsilon)}) - \theta(z^*) & \overset{(5.4)}{\geq} & \theta(z_{k(\epsilon)}) - \mu(\epsilon)D_S - \rho(\epsilon)D_T - \theta(z^*) \\ & \overset{(7.7)}{=} & \theta(z_{k(\epsilon)}) - \theta(z^*) - \frac{1}{2}\epsilon \;\geq\; -\frac{1}{2}\epsilon, \end{array}$$

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

$$\begin{array}{rcl} \left|J(\widehat{u}_{k(\epsilon)}) - D^*\right| & \leq & \left\| x_{\rho(\epsilon),z_{k(\epsilon)}} - \mathcal{A}\widehat{u}_{k(\epsilon)}\right\|_{V^*}\left\|z_{k(\epsilon)}\right\|_V + \rho(\epsilon)\hat{D} + \mu(\epsilon)D + \epsilon \\ & \overset{(7.7)}{\leq} & \left\|\nabla\theta_{\rho(\epsilon),\mu(\epsilon)}(z_{k(\epsilon)})\right\|_{V^*}\left\|z_{k(\epsilon)}\right\|_V + 2\epsilon \overset{(7.12)}{\leq} \frac{2\epsilon}{D_D}\left\|z_{k(\epsilon)}\right\|_V + 2\epsilon. \end{array}$$

On the other hand, $\left\|z_{k(\epsilon)}\right\|_V \;\leq\; \left\|z_{k(\epsilon)} - z^*_{DS}\right\|_V + \|z^*_{DS}\|_V \overset{(7.3)}{\leq} 2\|z^*_{DS}\|_V \overset{(7.11)}{\leq} 2\sqrt{3}D_D.$
and we obtain $\left|J(\widehat{u}_{k(\epsilon)}) - D^*\right| \;\leq\; 2\left(1 + 2\sqrt{3}\right)\cdot\epsilon.$

Finally, we have $\widehat{u}_{k(\epsilon)} \in S$ and $\left\|\mathcal{A}\widehat{u}_{k(\epsilon)} - x_{\rho(\epsilon),z_{k(\epsilon)}}\right\|^2_{V^*} = \|\nabla\theta_{\rho(\epsilon),\mu(\epsilon)}(z_{k(\epsilon)})\|^2_{V^*} \overset{(7.12)}{\leq} \left(\frac{2\epsilon}{R}\right)^2,$
where $x_{\rho(\epsilon),z_{k(\epsilon)}} \in T$. Therefore, $\widehat{u}_k$ can be seen as an approximately feasible and optimal solution for the primal problem (3.1).

8. Applications in Optimal Control. In this section, we will look at the optimal control problems (OCP), which can be written in the form (3.1) (more precisely in the form (4.1)). In particular, we consider OCP 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 point-wise convex constraints on the control variables. In order to motivate our choice of problem classes, let us show that OCP with nonlinear system of differential equations are NP-hard.

Consider the following OCP with convex objective function:

$$\text{(8.1)} \qquad \min_{u\in L^2([0,1],R^n)} \left\{\; \|x(1)\|^4_4 + \langle c, x(1)\rangle^2 :\; \dot{x} = -x\cdot\langle x,u\rangle + u,\; x(0) = x_0 \;\right\}.$$

We assume that vector c has integer coefficients.

LEMMA 8.1. *Let* $\|x_0\|^2_2 = 1$. *Then, finding an approximate solution to problem (8.1) with absolute accuracy higher than* $\hat{\epsilon} \overset{\text{def}}{=} \frac{1}{n(1+n^{1/2}\|c\|_1)^2}$ *is NP-hard.*

Proof. In view of the system of ODE in (8.1), we have $\langle \dot{x}, x\rangle \equiv 0$. Hence, by condition of the lemma, $\|x(t)\|_2 \equiv 1$. Note that by an appropriate control u we can move the starting point x_0 to any position at the unit sphere. Hence, the problem (8.1) is equivalent to the following finite-dimensional minimization problem:

$$\text{(8.2)} \qquad \text{Find } \xi_* = \min_{\|y\|_2=1} \left\{ \xi(y) \stackrel{\text{def}}{=} \|y\|_4^4 + \langle c, y\rangle^2 \right\}.$$

Let us show that this problem is equivalent to solving the equation $\langle c, y\rangle = 0$ with Boolean variables (this is a well-known NP-hard problem).

If this equation has Boolean solution y_* with coefficients $y_*^{(i)} = \pm m^{-1/2}$, $i = 1, \dots, n$, then $\xi_* = \frac{1}{m}$. On the other hand, note that for any $y \in R^m$ with unit Euclidean norm we have $\|y\|_4^4 = \frac{1}{m} + \sum\limits_{i=1}^{m} \left((y^{(i)})^2 - \frac{1}{m}\right)^2$. If we manage to find such a point y with $\xi(y) - \xi_* < \hat\epsilon$, then in the case $\xi(y) \geq \frac{1}{m} + \hat\epsilon$ we guarantee the absence of Boolean solutions. If $\xi(y) < \frac{1}{m} + \hat\epsilon$, then $|\langle c, y\rangle| < \hat\epsilon^{1/2}$, $\max\limits_{1\leq i\leq m} \left|(y^{(i)})^2 - \frac{1}{m}\right| < \hat\epsilon^{1/2}$. In this case, we can define the Boolean vector $u^{(i)} = \frac{1}{m^{1/2}} \cdot \text{sign}(y^{(i)})$, $i = 1, \dots, m$. For this vector we have

$$\begin{aligned} |\langle c, u\rangle| &= |\langle c, y\rangle + \langle c, u - y\rangle| \leq |\langle c, y\rangle| + \left| \sum_{i=1}^{m} c^{(i)} \cdot \text{sign}(y^{(i)}) \left(\tfrac{1}{n^{1/2}} - |y^{(i)}|\right) \right| \\ &< \hat\epsilon^{1/2} + \|c\|_1 \max_{1\leq i\leq m} \left| \tfrac{1}{m^{1/2}} - |y^{(i)}| \right| < \hat\epsilon^{1/2}(1 + m^{1/2}\|c\|_1) = \tfrac{1}{m^{1/2}}. \end{aligned}$$

Since vector c has integer coefficients, we conclude that $\langle c, u\rangle = 0$. □

8.1. Class of optimal control problems and reformulation. Consider the following optimal control problem:

$$\text{(8.3)} \qquad \begin{aligned} \inf_u \{ \int_0^1 F(t, u(t))dt : \quad & \dot x(t) = A(t)x(t) + B(t)u(t),\ x(0) = x_0, \\ & x(t_i) \in T_i \quad i = 1..., N, \\ & u(t) \in S(t) \quad \text{a.e in } [0,1] \}, \end{aligned}$$

where $S(t) \subset R^m$, $t \in [0,1]$, are closed convex sets with bounded graph $\overline{S} \stackrel{\text{def}}{=} \cup_{t\in[0,1]} Q(t)$. We assume that function $F : [0,1] \times \overline{S} \to R$ is bounded, and continuously differentiable and convex in the second argument, $x(t) \in R^n$ and $u(t) \in R^m$, $t \in [0,1]$. For measuring the control variables, we use the norm $\|u\|_2^2 = \int\limits_0^1 \|u(t)\|_2^2 dt$. We assume that $A(t) \in \mathcal{C}([0,1], R^{n\times n})$ and $B(t) \in \mathcal{C}([0,1], R^{n\times m})$. In problem (8.3), we have a finite number of inspection moments $t_i \in (0, T]$, and we assume that $T_i \subset R^n$, $i = 1, \dots, N$, are bounded closed convex sets. Let us rewrite the problem (8.3) 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 5. *When the system is time-invariant, i.e.* $A(t) = A$, *and* $B(t) = B$, $t \in [0,1]$, *then the transition matrix is the usual matrix exponent:*

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

From the Optimal Control Theory (e.g [12]), we know that the state trajectory $x(t)$, generated by the system of ODE under the control $u(t)$, is defined by the following expression: $x(t) \;=\; \Phi(t,0)x_0 + \int\limits_0^t \Phi(t,\tau)B(\tau)u(\tau)d\tau, \quad t \in [0,1]$. Therefore, the constraint $x(t_i) \in T_i$ can be expressed as follows:

$$\mathcal{A}_i(u) \;\stackrel{\text{def}}{=}\; \int\limits_0^{t_i} \Phi(t_i,\tau)B(\tau)u(\tau)d\tau \;\in\; \overline{T}_i \stackrel{\text{def}}{=} T_i - \Phi(t_i,0)x_0, \tag{8.4}$$

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 6. *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*

$$\begin{aligned} l(x) &= \int\limits_0^1 \langle x(t), a(t)\rangle dt \;=\; \int\limits_0^1 \langle \int\limits_0^t \Phi(t,\tau)B(\tau)u(\tau)d\tau, a(t)\rangle dt \\ &= \int\limits_0^1 \int\limits_0^t \langle \Phi(t,\tau)B(\tau)u(\tau), a(t)\rangle d\tau dt \;=\; \int\limits_0^1 \int\limits_0^t \langle u(\tau), B(\tau)^T\Phi(t,\tau)^T a(t)\rangle d\tau dt \\ &= \int\limits_0^1 \int\limits_\tau^T \langle u(\tau), B(\tau)^T\Phi(t,\tau)^T a(t)\rangle dt d\tau \;\stackrel{\text{def}}{=}\; \int\limits_0^1 \langle u(\tau), h(\tau)\rangle d\tau, \end{aligned}$$

with $h(\tau) = \int\limits_\tau^1 B(\tau)^T\Phi(t,\tau)^T a(t)dt$. Another possibility is as follows:

$$\begin{aligned} l(x) &= \langle x(t_i), a\rangle \;=\; \langle \int\limits_0^{t_i} \Phi(t_i,\tau)B(\tau)u(\tau)d\tau, a\rangle \\ &= \int\limits_0^{t_i} \langle \Phi(t_i,\tau)B(\tau)u(\tau), a\rangle d\tau \;\stackrel{\text{def}}{=}\; \int\limits_0^{t_i} \langle u(\tau), h(\tau)\rangle d\tau, \end{aligned}$$

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

Thus, for the linear operator $\mathcal{A}_i : L^2([0,1],R^m) \to R^n$, defined by (8.4), the ith state constraint becomes:

$$\mathcal{A}_i u \;=\; \int\limits_0^1 A_i(\tau)u(\tau)d\tau \;\in \bar{T}_i, \tag{8.5}$$

where $A_i(\tau) \stackrel{\text{def}}{=} \begin{cases} \Phi(t_i,\tau)B(\tau), & \text{when } \tau \in [0,t_i], \\ 0, & \text{when } \tau \in]t_i,1]. \end{cases}$

With $J(u) = \int\limits_0^1 F(t,u(t))dt$ and $S = \{u \in L^2([0,1],R^m) : u(t) \in S(t) \text{ a.e. in } [0,1]\}$, the optimal control problem (8.3) can be rewritten in the form (3.2) and therefore in the form (3.1) defining the linear application: $\mathcal{A} : L^2([0,1],R^m) \to R^{N\times n}$ such that for all $u \in L^2([0,1],R^m$ $\mathcal{A}u : R^{N\times n} \to R, \quad z = (z_1,...,z_N) \to \langle \mathcal{A}u, z\rangle = \sum_{i=1}^N \langle \mathcal{A}_i u, z_i\rangle$ and the convex set $T = T_1 \times T_2 \times ... \times T_N \subset R^{N\times n}$. Hence, we can solve it by the double smoothing technique. This approach assumes that we are able to solve the pointwise problems

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

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

$$\dot{v}(t) \quad = \quad -A(t)^T v(t), \quad v(t_i) = z^i, \quad t \in [0, t_i], \tag{8.6}$$

extended by zero for $t \in [t_i, 1]$.

8.2. Evaluation of $\|\mathcal{A}_i\|_2$. In order to solve the primal-dual problem (3.1), (3.3) by double smoothing technique, we need to evaluate the norm $\|\mathcal{A}\|_2 \le [\sum_{i=1}^N \|\mathcal{A}_i\|_2^2]^{1/2}$. Moreover, from the estimates (7.9), (7.10), it is clear that this norms is a very essential element 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,

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

Since the vector $\mathcal{A}_i u$ does not depend on values of $u(t)$ for $t \in (t_i, 1]$, we can consider the restriction of $\mathcal{A}_i$ on $L^2([0,t_i],R^m)$: $u \quad \to \quad \int_0^{t_i} \Phi(t_i,\tau)B(\tau)u(\tau)d\tau$. Then

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

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 $t_i > 0$, $i = 1, \ldots, N$, 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 8.2. *The system*

$$\dot{x}(t) \quad = \quad A(t)x(t) + B(t)u(t) \quad x(0) = 0, \tag{8.7}$$

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

The reachability is closely related to the reachability Gramian (e.g. Corollary 2.3 in [1]):

THEOREM 8.3. *The system (8.7) 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:

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

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

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

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

LEMMA 8.4. *Let* H *be a Hilbert space and the linear operator* $A : H \to R^L$ *be nondegenerate:* $AA^* \succ 0$. *Then for any* $b \in R^L$ *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* $\pi_b(f) = f + A^*(AA^*)^{-1}(b - Af)$. Thus,

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

Therefore,

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

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.

8.3. 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:

$$\dot{x}(t) = A(t)x(t), \quad t \geq 0, \tag{8.8}$$

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

Recall that the state $x = 0$ is always an equilibrium of the system (8.8). It is the unique equilibrium if $A(t)$ is nonsingular for all $t \geq 0$. The following facts are standard (e.g. [1]):

THEOREM 8.5. *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

$$\sup_{\tau\geq 0} k(\tau) = \sup_{\tau\geq 0} \sup_{t\geq\tau} \|\Phi(t,\tau)\|_2 \stackrel{\text{def}}{=} \kappa_0 < \infty.$$

Finally, it is is exponentially stable if $\int\limits_0^t \|\Phi(t,\tau)\|_2^2 \, d\tau \leq C$ 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 8.6. *If the equilibrium $x = 0$ is stable and $k_1 \stackrel{\text{def}}{=} \sup\limits_{t\geq 0} \|B(t)\|_2 < \infty$, then*

$$\|\mathcal{A}_i\|_2 \leq k_1 \left[\int\limits_0^{t_i} k^2(\tau) d\tau \right]^{1/2}. \tag{8.9}$$

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

$$\begin{aligned}
\|\mathcal{A}_i u\| &= \left\| \int\limits_0^{t_i} \Phi(t_i,\tau)B(\tau)u(\tau)d\tau \right\| \leq \int\limits_0^{t_i} \|\Phi(t_i,\tau)B(\tau)\|_2 \, \|u(\tau)\|_2 \, d\tau \\
&\leq \left[\int\limits_0^{t_i} \|\Phi(t_i,\tau)\|_2^2 \|B(\tau)\|_2^2 d\tau \cdot \int\limits_0^{t_i} \|u(\tau)\|_2^2 d\tau \right]^{1/2} \leq k_1 \left[\int\limits_0^{t_i} k^2(\tau) d\tau \right]^{1/2} \|u\|_2 .
\end{aligned}$$

Therefore $\|\mathcal{A}_i\|_2 \leq k_1 \left[\int\limits_0^{t_i} k^2(\tau) d\tau \right]^{1/2}$. □

This upper bound depends on the growth of the integral $\int\limits_0^{t_i} k^2(\tau) d\tau$ with respect to t_i, which can be very fast. Moreover, it can happen that function $k(\cdot)$ is not in $L^2([0,t_i])$ and then the bound (8.9) gives no information. However, if we assume the uniform stability of the equilibrium $x = 0$, then we can get much better bounds.

THEOREM 8.7. *If equilibrium $x = 0$ is uniformly stable and $k_1 \stackrel{\text{def}}{=} \sup_{t\geq 0} \|B(t)\|_2 < \infty$, then*

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

The proof of this theorem is the same as that of Theorem 8.6. 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 8.8. *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 8.6. In some cases, we can obtain a computable upper bound for the norm $\|\mathcal{A}_i\|_2$. Recall the following well-known sufficient condition for the global exponential stability.

THEOREM 8.9. *[1] Let the linear system (8.8) 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

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

Therefore,

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

where $R(t) \stackrel{\mathrm{def}}{=} \langle Px(t), x(t)\rangle$. 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

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

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

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:

$$\text{(8.10)}\quad \min\left\{\tfrac{\eta_3}{\eta_1\eta_2} : \quad A^T P + PA \preceq -\eta_1 P, \quad \eta_2 I \preceq P \preceq \eta_3 I, \quad \eta_1, \eta_2, \eta_3 \geq 0\right\}.$$

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

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

since the feasible set of the right-hand side is smaller than that of (8.10).
Furthermore, if we introduce new variables: $\tilde{P} = \frac{P}{\eta_1\eta_3}$, $\tilde{\eta}_2 = \frac{\eta_2}{\eta_1\eta_3}$, $\tilde{\eta}_3 = \frac{1}{\eta_1}$ we obtain a convex problem that can be solved in polynomial time:

$$\min\left\{\frac{\tilde{\eta}_3^2}{\tilde{\eta}_2} : A^T\tilde{P} + \tilde{P}A \preceq -I, \quad \tilde{\eta}_2 I \leq P \leq \tilde{\eta}_3 I, \quad \tilde{\eta}_2, \tilde{\eta}_3 \geq 0\right\}.$$

9. Comparison with the literature and conclusion. The subject of this paper can be summarized as the development of an efficient first-order method (obtained using the double smoothing technique) in order to solve partially finite (or finite) convex optimization problems with linear constraints.

Partially finite convex problems have been extensively studied in a theoretical way with duality results, weak constraints qualification ([6, 7, 9, 10, 14, 15]) and applications for example to maximum entropy ([5, 8]).

On the other hand, it is not the first time that the smoothing technique is used for solving finite-dimensional convex problems with linear constraint by the first-order methods. As our approach can be also interesting for solving finite-dimensional problems, we briefly mention these papers here, and discuss the differences with our approach.

In [16], the authors consider the case of a conic problem with linear objective function i.e. $J(u) = \langle c^*, u\rangle$ with $c^* \in U^*$, $T = \{b^*\} \subset V^*$ and $S = \mathcal{L} \subset U$ a closed convex cone. Using the rich duality theory for such kind of conic problems, they consider a primal-dual approach. The main idea is to reformulate the primal dual optimality conditions: $\mathcal{A}^* y + s^* - c^* = 0, \mathcal{A}u - b^* = 0, \langle c^*, u\rangle - \langle b^*, y\rangle = 0, (u, y, s^*) \in \mathcal{L} \times V \times \mathcal{L}^*$ as a non-smooth convex problem:

$$\overline{f} = \min_{z\in Z}\{f(z) = \|Ez - e\|_*\} = \min_{z\in Z} \max_{\|w\|\leq 1} \langle Ez - e, w\rangle \tag{9.1}$$

where $\|.\|$ denotes a norm on $U \times V \times R$, $\|.\|_*$ its dual norm, $z = (u, y, s^*)^T$, $e = (c^*, b^*, 0)^T$, E=

$$\begin{pmatrix} 0 & \mathcal{A}^* & I \\ \mathcal{A} & 0 & 0 \\ c^* & -b^* & 0 \end{pmatrix}$$

and $Z = \mathcal{L} \times V \times \mathcal{L}^*$.
Applying the smoothing technique to the function f, they are able to find a primal-dual solution z_ϵ such that $\|Ez_\epsilon - e\|_* \leq \epsilon$ in $O\left(\frac{1}{\epsilon}\right)$. This primal-dual approach does not work in our case for two reasons. First, primal-dual optimality condition cannot be expressed as a linear system $Ez = e$ (subject to a conic constraint). Furthermore, we do not want to work in the primal-dual space but preferably in the dual one due to our asymmetry assumption (the problem (9.1) can be infinite-dimensional in our framework).

The approach considered in [2] is more comparable to what we are doing. Their problem class is composed of problems of the form (3.1) with J a convex function not necessarily smooth, $S = U$ and $T = \mathcal{K} - b$, where $\mathcal{K}$ is a closed convex cone. Dualizing the constraint $\mathcal{A}u + b \in \mathcal{K}$, they obtain a dual problem with conic constraint: $\max_{z\in\mathcal{K}^*} g(z)$ where $g(z) = \inf_u J(u) - \langle \mathcal{A}u + b, z\rangle$. They apply the smoothing technique to the dual objective function and compare different optimal first-order methods of smooth convex optimization for solving the smoothed dual problem. They are able to solve the dual problem with accuracy ϵ in $O\left(\frac{1}{\epsilon}\right)$ iteration. Concerning the reconstruction of a nearly optimal primal solution u_ϵ from a nearly optimal dual solution z_ϵ, they suggest a very easy and natural way. u_ϵ is simply chosen as the minimizer of the optimization subproblem defining the smoothed dual objective function at the point z_ϵ. However, this suggestion is not supported by the analysis of the convergence rates.

In the framework of separable convex problems, smoothing technique has also been applied in [17] to convex problems with linear coupling constraint. Dualizing the coupling constraint, the authors obtain a dual objective function that can be computed in a separable way. Applying a simple smoothing to this dual objective function, they obtain a smooth dual objective function keeping the separability structure. Here also, this approach allows them to solve the dual problem with accuracy ϵ in $O\left(\frac{1}{\epsilon}\right)$ iterations. Concerning the reconstruction of a primal solution, they propose to use averaging of the minimizers of the subproblems defining the smoothed dual objective function at the different dual iterates. They prove that the quality of this primal solution is also of order ϵ. It depends also on the norm of the dual optimal solution (which is typically unknown).

The approach considered in this paper, allows us also to exploit the separability structure of decomposable problems with linear coupling constraint (see the two examples given in Section 4). In our work, we apply a double smoothing that gives us a possibility to reconstruct more easily a nearly optimal primal solution from a nearly optimal dual solution without using averaging. The price that we pay for this simplicity is just a logarithmic term $\log\left(\frac{1}{\epsilon}\right)$ in the complexity. For the level of accuracy we are interested in, the logarithmic factor is not distinguishable from an absolute constant. Furthermore, whereas we use also in our analysis the norm of the dual optimal solution, we provide an explicit upper-bound for this quantity.

More generally, to the best of our knowledge, this work is the first one where application of smoothing technique for solving some infinite dimensional problems is discussed. In particular, we consider in the Section 8, optimal control problems governed by a system of linear differential equations with the constraint that the trajectory crosses in certain moments of time some convex set.

REFERENCES

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

[2] S. Becker, E. Candès and M. Grant. Templates for Convex Cone Problems with Applications to Sparse Signal Recovery. *Submitted for publication, 2010.*

[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] J.F. Bonnans and A. Shapiro. *Perturbation Analysis of Optimization Problems*. Springer. (2000)

[5] J.M. Borwein and A.S. Lewis. Duality Relationship for Entropy-Like Minimization Problems. *SIAM. J. Control and Optimization*, **29(2)**, 325-338 (1991).

[6] J.M. Borwein and A.S. Lewis. Partially finite convex programming, Part I: Quasi relative interiors and duality theory. *Mathematical Programming*, **57**, 15-48 (1992).

[7] J.M. Borwein and A.S. Lewis. Partially finite convex programming, Part II: Explicit Lattice models. *Mathematical Programming*, **57**, 49-83 (1992).

[8] J.M. Borwein and A.S. Lewis Partially finite programming in L_1 and the existence of maximum entopy estimates. *SIAM. J. Optimization*, **3(2)**, 248-267 (1993).

[9] J.M. Borwein and H. Wolkowicz. A simple constraint qualification in infinite dimensional programming. *Mathematical Programming*, **35**, 83-96 (1986).

[10] R.I. Bot, S-M. Grad and G. Warka New regularity conditions for Lagrange and Fenchel-Lagrange duality in infinite dimensional spaces. *Mathematical Inequalities and Applications*,**12(1)**, 171-189 (2009).

[11] J.M. Danskin. *The theory of Max-Min and its application to weapons allocation problems.*

Springer-Verlag (1967).
[12] D.E. Kirk. *Optimal Control: An Introduction.* Dover publication, Inc. (2004).
[13] H. Kwakernaak and R. Sivan. *Linear Optimal Control Systems.* Wiley Interscience (1972).
[14] V. Jeyakumar, N. Dinh and G.M. Lee A New closed cone constraint qualification for convex optimization. *Applied Mathematics Report AMR 04/08, University of New South Wales, Australia* (2004).
[15] V. Jeyakumar and H. Wolkowicz. Generalizations of Slater's constraint qualification for infinite convex programs. *Mathematical Programming*, **57**, 85-101 (1992).
[16] G. Lan, Z. Lu and R. D.C Monteiro. Primal-dual first-order methods with $O(\frac{1}{\epsilon})$ iteration-complexity for cone programming. *to appear Mathematical Programming, 2009, online-first.*
[17] I. Necoara and J.A.K. Suykens. Application of a Smoothing Technique to Decomposition in Convex Optimization. *IEEE Transactions on Automatic Control*, **53**, 2674-2679 (2008).
[18] A. Nemirovskii and D. Yudin.Problem complexity and method efficiency in optimization. *John Wiley* (1983)
[19] Yu. Nesterov. *Introductory Lectures on Convex Optimization: A Basic Course.* Kluwer Academic Publishers (2004)
[20] Yu. Nesterov. Smooth minimization of non-smooth functions. *Mathematical programming, Serie A*, **103**, 127-152 (2005).
[21] Yu. Nesterov. Excessive gap technique in nonsmooth convex minimization. *Siam Journal of Optimization*, **16**, 235-249 (2005).
[22] Yu. Nesterov. Smoothing technique and its applications in semidefinite optimization. *Mathematical Programming A*, **110**, 245-259 (2007).
[23] Yu. Nesterov. Primal-Dual subgradient methods for convex problems. *Mathematical programming, Serie B*, **120**, 221-259 (2009).
[24] Yu. Nesterov. A method for unconstrained convex minimization with the rate of convergence of $O(\frac{1}{k^2})$, *Doklady AN SSSR*, **269**, 543-547 (1983).
[25] Yu. Nesterov.On an approach to the construction of optimal methods of minimization of smooth convex function, *Èkonom. i. Mat. Metody (In Russian)*, **24**, 509-517 (1988).
[26] P. Tseng. On accelerated Proximal Gradient Methods for Convex-Concave Optimization *Submitted to Siam. J. Optim.* (2008).
[27] E. Zeidler. *Nonlinear Functional Analysis and its Application III: Variational Methods and Optimization*, Springer-Verlag (1985).