J Optim Theory Appl (2015) 166:23–51
DOI 10.1007/s10957-015-0771-3

Primal-Dual Methods for Solving Infinite-Dimensional Games

Pavel Dvurechensky[1] · **Yurii Nesterov**[2] · **Vladimir Spokoiny**[3]

Received: 15 July 2013 / Accepted: 7 December 2013 / Published online: 24 June 2015

Abstract In this paper, we show that the infinite-dimensional differential games with simple objective functional can be solved in a finite-dimensional dual form in the space of dual multipliers for the constraints related to the end points of the trajectories. The primal solutions can be easily reconstructed by the appropriate dual subgradient schemes. The suggested schemes are justified by the worst-case complexity analysis.

Keywords Convex optimization · Primal-dual optimization methods · Saddle-point problems · Differential games

Mathematics Subject Classification 90C06 · 90C25 · 90C60 · 91A23 · 49N70

1 Introduction

In the last years, we can observe an increasing interest to the primal-dual subgradient methods. This line of research, started in [1], leads to the special methods, which allow

Communicated by Boris S. Mordukhovich.

✉ Yurii Nesterov
yurii.nesterov@uclouvain.be

Pavel Dvurechensky
p.dvurechensky@phystech.edu

Vladimir Spokoiny
spokoiny@wias-berlin.de

1 PreMoLab, Moscow Institute of Physics and Technology, Institutskii st.,9, Dolgoprudny, Russia

2 CORE, Universite catholique de Louvain, 1348 Louvain-la-Neuve, Belgium

3 Weierstrass Institute and Humboldt University, Mohrenstr. 39, 10117 Berlin, Germany

one to reconstruct approximate solution to a *conjugate* problem. In order to do this, methods need to get an access to the internal variables of the oracle. Therefore, all these methods are problem specific.

This approach is very interesting, when the primal and conjugate problems have different levels of complexity. For example, we can have a primal minimization problem of very high dimension, with very simple objective function and basic feasible set, and a small number of linear equations. Introducing Lagrange multipliers for these linear constraints, we can pass to the conjugate (dual) problem,[1] which has good chances to be simple in view of its small dimension. The only delicate problem is the reconstruction of the primal variables from the minimization process, which we run in the conjugate space.

In [2], this approach was applied to the problems of optimal control with convex constraints for the end point of the trajectory. These constraints were treated by linear operators from infinite-dimensional space of variables (control) to a finite-dimensional space of phase variables. It was shown that an appropriate optimization process in the latter space can generate also nearly optimal sequence of controls (functions of time). Moreover, this technique was supported by the worst-case complexity analysis.

In this paper, we move further in this direction. We consider an infinite-dimensional saddle-point problem, which variables (controls) must satisfy some linear equality constraints. We show that these constraints can be dualized by *finite-dimensional* multipliers. Moreover, it appears that the dual counterpart of our problem is again a saddle-point problem, but in a finite dimension (we call this problem *conjugate*). We show how to reconstruct the infinite-dimensional primal strategies from a special finite-dimensional scheme, which solves the conjugate problem.

The paper is organized as follows. In the Sect. 2, we consider the basic formulation of differential games with convex–concave objective and with trajectories of the players governed by the systems of linear differential equations. We treat the end points of the trajectories as an image of linear operators from infinite- to finite-dimensional space. For the future applications, we derive some bounds for their norms. In the Sect. 3, we write down an equivalent conjugate saddle-point problem in the finite-dimensional space of dual multipliers and derive some bounds on the size of the optimal conjugate solutions. In the end of this section, we present a numerical scheme and derive the upper bounds on the quality of primal and conjugate solutions. In the Sect. 4, we consider a differential game with an objective function satisfying strong convexity assumption. For this case, we obtain better complexity bounds. In the last Sect. 5, we show how to form the conjugate problem for the initial finite-dimensional convex–concave saddle-point problem with equality constraints. It seems that this transformation is new even in this simplest situation. Therefore, we devote to it a separate section.

[1] Since the objective and the feasible set of our problem are simple, very often this can be done in an explicit form.

2 Differential Games

Consider two moving objects with dynamics given by the following equations:

$$\dot{x}(t) = A_x(t)x(t) + B(t)u(t),\ \dot{y}(t) = A_y(t)y(t) + C(t)v(t),$$
$$(x(0), y(0)) = (x_0, y_0). \tag{1}$$

Here $x(t) \in \mathbb{R}^n$, $y(t) \in \mathbb{R}^m$ are the phase vectors of these objects, $u(t)$ is the control of the first object (pursuer), and $v(t)$ is the control of the second object (evader). Matrices $A_x(t)$, $A_y(t)$, $B(t)$, and $C(t)$ are continuous and have appropriate sizes. The system is considered on the time interval $[0, \theta]$. Controls are restricted in the following way $u(t) \in P \subseteq \mathbb{R}^p$, $v(t) \in Q \subseteq \mathbb{R}^q \quad \forall t \in [0, \theta]$. We assume that P, Q are closed, convex sets.

The goal of the pursuer is to minimize the value of the functional:

$$F(u, v) + \Phi(x(\theta), y(\theta)) := \int_0^\theta \tilde{F}(\tau, u(\tau), v(\tau))d\tau + \Phi(x(\theta), y(\theta)). \tag{2}$$

The goal of the evader is the opposite. We need to find an optimal guaranteed result for each object, which leads to the problem of finding the saddle point of the above functional. We assume the following:

- $u(\cdot) \in L^2([0, \theta], \mathbb{R}^p)$, and $v(\cdot) \in L^2([0, \theta], \mathbb{R}^q)$ (for the notation simplification we denote $L^2([0, \theta], \mathbb{R}^p)$ by L^2_p and $L^2([0, \theta], \mathbb{R}^q)$ by L^2_q),
- the saddle point in this class of strategies exists,
- the function $F(u, v)$ is upper semi-continuous in v and lower semi-continuous in u,
- $\Phi(x, y)$ is continuous.

Denote by $V_x(t, \tau)$ the transition matrix of the first system in (1). It is the unique solution of the following matrix Cauchy problem

$$\frac{\mathrm{d}V_x(t, \tau)}{\mathrm{d}t} = A_x(t)V_x(t, \tau), \quad t \geq \tau, \quad V_x(\tau, \tau) = E.$$

Here E is the identity matrix. If the matrix $A_x(t)$ is constant, then
$V_x(t, \tau) = e^{(t-\tau)A}$.

If we solve the first differential equation in (1), then we can express $x(\theta)$ as a result of the application of the linear operator $\mathcal{B} : L^2_p \to \mathbb{R}^n$:

$$x(\theta) = V_x(\theta, 0)x_0 + \int_0^\theta V_x(\theta, \tau)B(\tau)u(\tau)d\tau := \tilde{x}_0 + \mathcal{B}u. \tag{3}$$

Below, we will use the conjugate operator $\mathcal{B}^*$ for the operator $\mathcal{B}$. Let us find it explicitly. Let μ be a n-dimensional vector. Then,

$$\langle \mu, \mathcal{B}u \rangle = \langle \mu, \int_0^\theta V_x(\theta,\tau)B(\tau)u(\tau)d\tau \rangle = \int_0^\theta \langle \mu, V_x(\theta,\tau)B(\tau)u(\tau)\rangle d\tau$$
$$= \int_0^\theta \langle B^T(\tau)V_x^T(\theta,\tau)\mu, u(\tau)\rangle d\tau = \langle \mathcal{B}^*\mu, u\rangle.$$

Note that the vector $\zeta(t) = V_x^T(\theta, t)\mu$ is the solution of the following Cauchy problem:

$$\dot{\zeta}(t) = -A_x^T(t)\zeta(t), \quad \zeta(\theta) = \mu, \quad t \in [0,\theta].$$

So we can solve this ODE and find $\mathcal{B}^*\mu$ using the obtained solution $\zeta(t)$ as $\mathcal{B}^*\mu(t) = B^T(t)\zeta(t)$.

In the same way, we introduce the transition matrix $V_y(t,\tau)$ of the second system in (1), the operator $\mathcal{C} : L_q^2 \to \mathbb{R}^m$ defined by the formula $\mathcal{C}v := \int_0^\theta V_y(\theta,\tau)C(\tau)v(\tau)d\tau$, and the vector $\tilde{y}_0 := V_y(\theta,0)y_0$. The adjoint operator $\mathcal{C}^*$ also can be computed using the solution of some ODE.

So below we will study differential game problem in the following form:

$$\min_{u\in\mathcal{U}} \left[\max_{v\in\mathcal{V}} \{F(u,v) + \Phi(x,y) : y = \tilde{y}_0 + \mathcal{C}v\} : x = \tilde{x}_0 + \mathcal{B}u \right], \tag{4}$$

where

$$\mathcal{U} := \{u(\cdot) \in L_p^2 : u(t) \in P \quad \forall t \in [0,\theta]\}, \mathcal{V} := \{v(\cdot) \in L_q^2 : v(t) \in Q \quad \forall t \in [0,\theta]\}$$

are sets of admissible strategies of the players and $u \in \mathcal{U}$, $v \in \mathcal{V}$ mean $u(\cdot) \in \mathcal{U}$, $v(\cdot) \in \mathcal{V}$. Our goal is to introduce a computational method for finding an approximate solution of the problem (4).

Remark 2.1 In the same way, we can treat a problem with objective functional of the form $\int_0^\theta \tilde{F}(\tau, u(\tau), v(\tau))d\tau + \sum_{i=0}^k \Phi(x(t_i), y(t_i))$, or constraints of the type $\mathcal{B}u \in T$ and $\mathcal{C}v \in S$, where T and S are closed, convex sets.

2.1 Estimating the Norms of the Operators $\mathcal{B}$ and $\mathcal{C}$

Let us assume that $\mathbb{R}^n$ and $\mathbb{R}^m$ are endowed with the Euclidean norm $\|\cdot\|_2$. Consider the problem of estimating the norms of operators $\mathcal{B}, \mathcal{C}$. This is an important problem since below we need these operators to be bounded, and also their norms play a significant role in the estimates for the rate of convergence of the methods we introduce. Let us study the operator $\mathcal{B}$ (3), since the norm of $\mathcal{C}$ can be estimated in a similar way. The following argument was used in [2], and it is presented here for the reader convenience.

By definition, we have

$$\|\mathcal{B}\|_2 := \sup_{u\in L_p^2}\{\|\mathcal{B}u\|_2 : \quad \|u\|_{L_p^2} = 1\}. \tag{5}$$

As it was shown above, the conjugate operator $\mathcal{B}^*$ transforms $\mu \in \mathbb{R}^n$ into the function $B^T(\tau)V_x^T(\theta,\tau)\mu \in L_p^2$. Let us define a matrix

$$R := \int_0^\theta V_x(\theta,\tau)B(\tau)B^T(\tau)V_x^T(\theta,\tau)d\tau = \mathcal{B}\mathcal{B}^*, \tag{6}$$

which is symmetric and positive semi-definite.

Definition 2.1 The system with the dynamics given by the first differential equation in (1) and the initial value $x(0) = 0$ is called reachable on $[0,\theta]$ iff for any $\hat{x} \in \mathbb{R}^n$ there exists a control such that $x(\theta) = \hat{x}$.

The reachability is closely related to the properties of the matrix R (see Corollary 2.3 in [3]).

Lemma 2.1 *The system with the dynamics given by the first differential equation in* (1) *and the initial value* $x(0) = 0$ *is reachable on* $[0,\theta]$ *if and only if* R *is positive definite.*

We also need the following

Lemma 2.2 *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* $\pi_b(f) = f + A^*(AA^*)^{-1}(b - Af)$.

From the definition (5),

$$\|\mathcal{B}\|_2 = \left[\inf_{u\in L_p^2}\{\|u\|_{L_p^2} : \quad \|\mathcal{B}u\|_2 = 1\}\right]^{-1}.$$

Using the reachability property, we have $\mathrm{Im}\mathcal{B}(L_p^2) = \mathbb{R}^n$ and

$$\inf_{u\in L_p^2}\{\|u\|_{L_p^2} : \quad \|\mathcal{B}u\|_2 = 1\} = \inf_{u\in L_p^2, x\in\mathbb{R}^n, \|x\|_2=1}\{\|u\|_{L_p^2} : \quad \mathcal{B}u = x\}.$$

From the Lemma 2.2, we have that

$$\inf_{u\in L_p^2}\{\|u\|_{L_p^2} : \quad \mathcal{B}u = x\} = \|\mathcal{B}^*(\mathcal{B}\mathcal{B}^*)^{-1}x\|_{L_p^2} = \langle(\mathcal{B}\mathcal{B}^*)^{-1}x, x\rangle^{1/2}.$$

Hence,

$$\inf_{u\in L_p^2}\{\|u\|_{L_p^2}:\quad \|\mathcal{B}u\|_2=1\}$$
$$=\inf_{\|x\|_2=1}\langle(\mathcal{B}\mathcal{B}^*)^{-1}x,x\rangle^{1/2}=\lambda_{\min}^{1/2}((\mathcal{B}\mathcal{B}^*)^{-1}).$$

Finally, $\|\mathcal{B}\|_2=\lambda_{\min}^{-1/2}((\mathcal{B}\mathcal{B}^*)^{-1})=\lambda_{\max}^{1/2}(\mathcal{B}\mathcal{B}^*)$, where $\mathcal{B}\mathcal{B}^*=R$.

Also we can get a time-independent estimate of $\|\mathcal{B}\|_2$ in the case when $x=0$ is an exponentially stable equilibrium of the system with dynamics

$$\dot{x}(t)=A_x x(t),\quad t\geq 0,$$

where A_x is a matrix.

Recall the following well known result.

Theorem 2.1 [3] *Assume that there exists a matrix $M=M^T\succ 0$ such that $A_x^T M+MA_x\prec 0$. Then, the equilibrium $x=0$ is globally exponentially stable.*

So we can consider a case when there exists some $\nu>0$ and $M=M^T\succ 0$ such that $A_x^T M+MA_x\preceq -\nu M$. Let us also assume that the matrix $B(t)$ is time independent. Then, we have that $\mathcal{B}u$ is the position at the moment θ of the point of the unique trajectory defined by the linear system

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

Hence

$$\|x(\theta)\|_2^2=\langle x(\theta),x(\theta)\rangle\leq\frac{\langle Mx(\theta),x(\theta)\rangle}{\lambda_{\min}(M)},$$

and

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\langle Mx(t),x(t)\rangle &=2\langle Mx(t),\dot{x}(t)\rangle=2\langle Mx(t),Ax(t)+Bu(t)\rangle\\&=\langle(A^TM+MA)x(t),\dot{x}(t)\rangle+2\langle Mx(t),Bu(t)\rangle\\&\leq-\nu\langle Mx(t),x(t)\rangle+2\langle Mx(t),Bu(t)\rangle\leq\frac{1}{\nu}\langle MBu(t),Bu(t)\rangle.\end{aligned}$$

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

$$\begin{aligned}\langle Mx(\theta),x(\theta)\rangle&=\int_0^\theta\frac{\mathrm{d}}{\mathrm{d}t}\langle Mx(t),x(t)\rangle dt\leq\frac{1}{\nu}\int_0^\theta\langle MBu(t),Bu(t)\rangle dt\\&\leq\frac{1}{\nu}\lambda_{\max}(M)\int_0^\theta\|Bu(t)\|_2^2dt\leq\frac{1}{\nu}\lambda_{\max}(M)\|B\|_2^2\|u\|_{L_p^2}^2.\end{aligned}$$

Finally, we have $\|\mathcal{B}\|_2^2\leq\frac{\lambda_{\max}(M)}{\nu\lambda_{\min}(M)}\|B\|_2^2$.

From now on, we assume that both operators $\mathcal{B}$ and $\mathcal{C}$ have bounded norms.

3 Convex–Concave Problem

In this section, we consider the problem (4) satisfying two assumptions.

A1 The sets P and Q are bounded.
A2 In (2), the functional $F(\cdot, v)$ is convex for any fixed v, $F(u, \cdot)$ is concave for any fixed u, $\Phi(\cdot, y)$ is convex for any fixed y, and $\Phi(x, \cdot)$ is concave for any fixed x.

From **A1**, since the norms of the operators $\mathcal{B}, \mathcal{C}$ are bounded, $x(\theta)$, $y(\theta)$ are also bounded and we can equivalently reformulate the problem (4) in the following way:

$$\min_{u\in\mathcal{U}, x\in X}\left[\max_{v\in\mathcal{V}, y\in Y}\{F(u,v)+\Phi(x,y) : y=\tilde{y}_0+\mathcal{C}v\} : x=\tilde{x}_0+\mathcal{B}u\right]$$
$$=\max_{v\in\mathcal{V}, y\in Y}\left[\min_{u\in\mathcal{U}, x\in X}\{F(u,v)+\Phi(x,y) : x=\tilde{x}_0+\mathcal{B}u\} : y=\tilde{y}_0+\mathcal{C}v\right], \quad (7)$$

where the sets X and Y are closed, convex and bounded. Let us introduce the spaces of dual variables $\lambda \in \mathbb{R}^m$ and $\mu \in \mathbb{R}^n$ corresponding to the linear constraints in the problem (7), and some norms $\|\cdot\|_\lambda$ and $\|\cdot\|_\mu$ in these spaces. We define the norms in the dual space in the standard way

$$\|s_\lambda\|_{\lambda,*} := \max\{\langle s_\lambda, \lambda\rangle : \|\lambda\|_\lambda \le 1\}, \quad \|s_\mu\|_{\mu,*} := \max\{\langle s_\mu, \mu\rangle : \|\mu\|_\mu \le 1\}.$$

In the simple case, both the primal and the dual norms are Euclidean.

Lemma 3.1 *Let the Assumptions* **A1**, **A2** *be true. Also assume that the function $F(u, v)$ is upper semi-continuous in v and lower semi-continuous in u, the function $\Phi(x, y)$ is continuous, and that the sets P and Q are convex and closed. Then, the problem* (7) *is equivalent to the problem*

$$\min_\lambda \max_\mu \left\{\min_{u\in\mathcal{U}} \max_{v\in\mathcal{V}} [F(u,v) - \langle\mu, \mathcal{B}u\rangle + \langle\lambda, \mathcal{C}v\rangle] \right.$$
$$\left. + \min_{x\in X}\max_{y\in Y}[\Phi(x,y) + \langle\mu, x\rangle - \langle\lambda, y\rangle] - \langle\mu, \tilde{x}_0\rangle + \langle\lambda, \tilde{y}_0\rangle\right\}, \quad (8)$$

which we call the conjugate problem to (7).

Proof Let us consider the inner problem in (7). Due to **A2**, for each $v \in \mathcal{V}$ and $y \in Y$, this is a problem of minimization of a convex function over a convex set with linear constraints. Hence, it is equivalent to

$$\chi(v, y) := \min_{u\in\mathcal{U}, x\in X} \max_\mu \{F(u,v) + \Phi(x,y) + \langle\mu, x - \tilde{x}_0 - \mathcal{B}u\rangle\}. \quad (9)$$

Due to assumptions **A1**, **A2**, using the fact that any closed, convex, and bounded set in Hilbert space is compact in the weak topology, and taking into account that $F(u, v)$ is upper semi-continuous in v and lower semi-continuous in u, by Corollary 3.3 in [4], we can swap min and max:

$$\chi(v, y) = \max_{\mu} \min_{u \in \mathcal{U}, x \in X} \{F(u, v) + \Phi(x, y) + \langle \mu, x - \tilde{x}_0 - \mathcal{B}u \rangle\}.$$

Note that $\chi(v, y)$ in (9) is a concave function of v an y. So the outer problem in (7) is a problem of maximization of a concave function over a convex set with linear constraints.
Hence, it is equivalent to: $\max_{v \in \mathcal{V}, y \in Y} \min_{\lambda} \{\chi(v, y) + \langle \lambda, \mathcal{C}v + \tilde{y}_0 - y \rangle\}$. Using the same argument as above, we conclude that

$$\begin{aligned} &\max_{v \in \mathcal{V}, y \in Y} \min_{\lambda} \{\chi(v, y) + \langle \lambda, \mathcal{C}v + \tilde{y}_0 - y \rangle\} \\ &\quad = \min_{\lambda} \max_{v \in \mathcal{V}, y \in Y} \{\chi(v, y) + \langle \lambda, \mathcal{C}v + \tilde{y}_0 - y \rangle\}. \end{aligned}$$

Denote $F(u, v) + \Phi(x, y) + \langle \lambda, \mathcal{C}v + \tilde{y}_0 - y \rangle + \langle \mu, x - \tilde{x}_0 - \mathcal{B}u \rangle$ by $\Psi(u, v, x, y, \lambda, \mu)$. Hence, we have

$$(7) = \min_{\lambda} \max_{v \in \mathcal{V}, y \in Y} \max_{\mu} \min_{u \in \mathcal{U}, x \in X} \{\Psi(u, v, x, y, \lambda, \mu)\}.$$

Swapping two operations of maximization, we get.

$$(7) = \min_{\lambda} \max_{\mu} \max_{v \in \mathcal{V}, y \in Y} \min_{u \in \mathcal{U}, x \in X} \{\Psi(u, v, x, y, \lambda, \mu)\}.$$

Since the function $\Psi(u, v, x, y, \lambda, \mu)$ is convex in u, x and concave in v, y, and since $\mathcal{U}$, $\mathcal{V}$ are convex weakly compact sets, and X and Y are convex compacts, we can swap $\max_{v \in \mathcal{V}, y \in Y}$ and $\min_{u \in \mathcal{U}, x \in X}$, and obtain (7)=(8). □

We assume that the problems

$$\psi_1(\lambda, \mu) := \min_{u \in \mathcal{U}} \max_{v \in \mathcal{V}} [F(u, v) - \langle \mu, \mathcal{B}u \rangle + \langle \lambda, \mathcal{C}v \rangle], \quad (10)$$

$$\psi_2(\lambda, \mu) := \min_{x \in X} \max_{y \in Y} [\Phi(x, y) + \langle \mu, x \rangle - \langle \lambda, y \rangle] \quad (11)$$

are rather simple so that they can be solved efficiently or in a closed form. Note that the conjugate problem is finite dimensional. By assumptions **A1**, **A2**, since a closed, convex, and bounded set in Hilbert space is compact in the weak topology, and since $F(u, v)$ is upper semi-continuous in v and lower semi-continuous in u, by Corollary 3.3 in [4], we conclude that the saddle point in the problems (10), (11) does exist for all $\lambda \in \mathbb{R}^m$, $\mu \in \mathbb{R}^n$.

Note that the problem (10) has the following form

$$\min_{u\in\mathcal{U}}\max_{v\in\mathcal{V}}\left[\int_0^\theta \left\{\tilde F(\tau,u(\tau),v(\tau)) - \langle \mathcal{B}^*\mu(\tau),u(\tau)\rangle + \langle \mathcal{C}^*\lambda(\tau),v(\tau)\rangle\right\} d\tau\right]$$
$$= \int_0^\theta \left\{\min_{u\in\mathcal{U}}\max_{v\in\mathcal{V}}\left[\tilde F(\tau,u(\tau),v(\tau)) - \langle \mathcal{B}^*\mu(\tau),u(\tau)\rangle + \langle \mathcal{C}^*\lambda(\tau),v(\tau)\rangle\right] d\tau\right\}, \tag{12}$$

and it can be solved pointwise.

Lemma 3.2 *Let the assumptions* **A1** *and* **A2** *be true, and* (u^*, v^*) *be a saddle point of the problem* (10) *for some fixed* $\lambda \in \mathbb{R}^m$, $\mu \in \mathbb{R}^n$. *Then, the function* $\psi_1(\cdot,\mu)$ *is convex for any fixed* $\mu \in \mathbb{R}^n$, *and its subgradient* $\nabla_\lambda \psi_1(\lambda,\mu) = \mathcal{C}v^*$ *is bounded in* (λ,μ). *Similarly, the function* $\psi_1(\lambda,\cdot)$ *is concave for any fixed* $\lambda \in \mathbb{R}^m$, *and its supergradient* $\nabla_\mu \psi_1(\lambda,\mu) = -\mathcal{B}u^*$ *is bounded in* (λ,μ).

Proof Since the saddle point in the problem (10) exists for any λ, μ, we have

$$\psi_1(\lambda,\mu) = \min_{u\in\mathcal{U}}\left[\tilde\psi_1(u,\lambda) - \langle \mathcal{B}u,\mu\rangle\right], \tag{13}$$

$$\psi_1(\lambda,\mu) = \max_{v\in\mathcal{V}}\left[\hat\psi_1(v,\mu) + \langle \mathcal{C}v,\lambda\rangle\right], \tag{14}$$

where

$$\tilde\psi_1(u,\lambda) := \max_{v\in\mathcal{V}}\left[F(u,v) + \langle \mathcal{C}v,\lambda\rangle\right], \tag{15}$$

$$\hat\psi_1(v,\mu) := \min_{u\in\mathcal{U}}\left[F(u,v) - \langle \mathcal{B}u,\mu\rangle\right]. \tag{16}$$

In the first case, since we take the minimum of linear functions of μ, the result is concave in μ. So $\psi_1(\lambda,\mu)$ is concave with respect to μ for any fixed λ. Similarly, we get that $\psi_1(\lambda,\mu)$ is convex in λ for any fixed μ. Let us fix λ, μ_0. Denote by (u_0^*, v_0^*) the saddle point of the problem (10) for λ, μ_0, and by (u^*, v^*) the saddle point of this problem for λ, μ, where μ is arbitrary. Then,

$$\min_{u\in\mathcal{U}}\left[\tilde\psi_1(u,\lambda) - \langle \mathcal{B}u,\mu\rangle\right] = \tilde\psi_1(u^*,\lambda) - \langle \mathcal{B}u^*,\mu\rangle,$$
$$\min_{u\in\mathcal{U}}\left[\tilde\psi_1(u,\lambda) - \langle \mathcal{B}u,\mu_0\rangle\right] = \tilde\psi_1(u_0^*,\lambda) - \langle \mathcal{B}u_0^*,\mu_0\rangle.$$

Note that

$$\begin{aligned}
&\psi_1(\lambda,\mu) - \psi_1(\lambda,\mu_0) + \langle \mathcal{B}u_0^*, \mu-\mu_0\rangle \\
&\quad = \tilde\psi_1(u^*,\lambda) - \langle \mathcal{B}u^*,\mu\rangle - \tilde\psi_1(u_0^*,\lambda) + \langle \mathcal{B}u_0^*,\mu_0\rangle + \langle \mathcal{B}u_0^*,\mu-\mu_0\rangle \\
&\quad = \tilde\psi_1(u^*,\lambda) - \langle \mathcal{B}u^*,\mu\rangle - (\tilde\psi_1(u_0^*,\lambda) - \langle \mathcal{B}u_0^*,\mu\rangle) \\
&\quad = \min_{u\in\mathcal{U}}\{\tilde\psi_1(u,\lambda) - \langle \mathcal{B}u,\mu\rangle\} - (\tilde\psi_1(u_0^*,\lambda) - \langle \mathcal{B}u_0^*,\mu\rangle) \le 0.
\end{aligned}$$

So, by definition, the vector $-\mathcal{B}u^*$ is a supergradient of $\psi_1(\lambda, \mu)$ with respect to μ. In the same way, we prove that $\mathcal{C}v^*$ is a subgradient of $\psi_1(\lambda, \mu)$ with respect to λ. Since P and Q are bounded, we have $\|u(t)\|_{L_p^2} \le \sqrt{\theta}\max_{z\in P}\|z\|_2 = \sqrt{\theta}\mathrm{diam}_2 P$. Here we introduced notation for any set S $\mathrm{diam}_\alpha S := \max\{\|z\|_\alpha : z \in S\}$, where the index α denotes some norm. Similarly, $\|v(t)\|_{L_q^2} \le \sqrt{\theta}\mathrm{diam}_2 Q$. Then, since the norms $\|\mathcal{B}\|_{\mu,L_p^2}$ and $\|\mathcal{C}\|_{\lambda,L_q^2}$, defined by

$$\|\mathcal{B}\|_{\mu,L_p^2} := \max_{\mu\in\mathbb{R}^n, u\in L_p^2}\{\langle \mathcal{B}u, \mu\rangle : \|\mu\|_\mu = 1, \|u\|_{L_p^2} = 1\},$$

$$\|\mathcal{C}\|_{\lambda,L_q^2} := \max_{\lambda\in\mathbb{R}^m, v\in L_q^2}\{\langle \mathcal{C}v, \lambda\rangle : \|\lambda\|_\lambda = 1, \|v\|_{L_q^2} = 1\},$$

are bounded, and then, we have

$$\|\nabla_\lambda\psi_1(\lambda, \mu)\|_{\lambda,*} \le \sqrt{\theta}\,\|\mathcal{C}\|_{\lambda,L_q^2}\,\mathrm{diam}_2 Q,\ \left\|\nabla_\mu\psi_1(\lambda, \mu)\right\|_{\mu,*} \le \sqrt{\theta}\,\|\mathcal{B}\|_{\mu,L_p^2}\,\mathrm{diam}_2 P.$$

□

In a similar way, we can prove the following statement.

Lemma 3.3 *Let the assumptions* **A1** *and* **A2** *be true, and* (x^*, y^*) *be a saddle point of the problem* (11) *for some given* $\lambda \in \mathbb{R}^m$, $\mu \in \mathbb{R}^n$. *Then, the function* $\psi_2(\cdot, \mu)$ *is convex for any fixed* $\mu \in \mathbb{R}^n$, *and its subgradient* $\nabla_\lambda\psi_2(\lambda, \mu) = -y^*$ *is bounded in* (λ, μ). *Similarly, the function* $\psi_2(\lambda, \cdot)$ *is concave for any fixed* $\lambda \in \mathbb{R}^m$, *and its supergradient* $\nabla_\mu\psi_2(\lambda, \mu) = x^*$ *is bounded in* (λ, μ).

Combining the Lemmas 3.2 and 3.3, we get that the function $\psi(\lambda, \mu) := \psi_1(\lambda, \mu) + \psi_2(\lambda, \mu) - \langle \mu, \tilde{x}_0\rangle + \langle \lambda, \tilde{y}_0\rangle$ is convex in λ and concave in μ with the partial subgradients $\psi'_\lambda(\lambda, \mu) = \mathcal{C}v^* + \tilde{y}_0 - y^*$, and $\psi'_\mu(\lambda, \mu) = x^* - \tilde{x}_0 - \mathcal{B}u^*$ satisfying the bounds

$$\begin{aligned}\left\|\psi'_\lambda(\lambda, \mu)\right\|_{\lambda,*} &\le L_\lambda := \sqrt{\theta}\,\|\mathcal{C}\|_{\lambda,L_q^2}\,\mathrm{diam}_2 Q + \mathrm{diam}_{\lambda,*}Y + \|\tilde{y}_0\|_{\lambda,*},\\ \left\|\psi'_\mu(\lambda, \mu)\right\|_{\mu,*} &\le L_\mu := \sqrt{\theta}\,\|\mathcal{B}\|_{\mu,L_p^2}\,\mathrm{diam}_2 P + \mathrm{diam}_{\mu,*}X + \|\tilde{x}_0\|_{\mu,*}.\end{aligned} \tag{17}$$

3.1 Example of the Problem (10)

Let us consider an example with $n = 2, m = 2, \theta = 1, p = q = 1, P = [-1, 1]$, $Q = [-1, 1]$, and

$$A_x(t) = A_y(t) = \begin{pmatrix} 0 & 1\\ 0 & 0\end{pmatrix},\ B(t) = C(t) = \begin{pmatrix} 0\\ 1\end{pmatrix},\ x(0) = \begin{pmatrix} 0\\ 0\end{pmatrix},\ y(0) = \begin{pmatrix} 1\\ 1\end{pmatrix}.$$

The objective functional will be as follows:

$$J(u, v) = \int_0^1\left(\frac{(u(t))^2}{2} - \frac{(v(t))^2}{2}\right)dt + \frac{1}{2}\|x(1) - y(1)\|_2^2 - \|y(1) - y_0\|_2^2,$$

where $y_0 = (2, 0)^T$, and the norm in the second and third terms is Euclidean. This functional satisfies Assumptions **A1** and **A2**.

Note that

$$V_x(t,\tau) = V_y(t,\tau) = \begin{pmatrix} 1 & t-\tau \\ 0 & 1 \end{pmatrix},$$
$$(\mathcal{B}^*\mu)(t) = B^T(t)V_x^T(1,t)\mu = (1-t)\mu_1 + \mu_2, \quad (\mathcal{C}^*\lambda)(t) = (1-t)\lambda_1 + \lambda_2.$$

Also we can explicitly solve (10) using (12):

$$\psi_1(\lambda,\mu) = \int_0^1 (f((1-t)\mu_1 + \mu_2) - f((1-t)\lambda_1 + \lambda_2))dt,$$

where the function $f(\rho) : \mathbb{R} \to \mathbb{R}$ is defined by

$$f(\rho) := \begin{cases} -\frac{\rho^2}{2} & |\rho| \le 1 \\ \frac{1}{2} - |\rho| & |\rho| > 1 \end{cases}.$$

3.2 Estimating the Norms of λ^*, μ^*

Let us compute the estimates for the norms of the components λ and μ of the solution of the conjugate problem. Denote for any function of two variables $\Psi(z,w)$ by $\Psi'(\tilde{z}, z-\tilde{z}|w)$ the directional derivative of the function $\Psi(\cdot, w)$ for some fixed w at the point $\tilde{z}$ in the direction $z-\tilde{z}$. Similarly, $\Psi'(z|\tilde{w}, w-\tilde{w})$ denotes the directional derivative of the function $\Psi(z,\cdot)$ for some fixed z at the point $\tilde{w}$ in the direction $w-\tilde{w}$. Also by $\mathfrak{B}_r^\alpha(0)$ we denote a ball of radius r in the norm $\|\cdot\|_\alpha$ with center at the origin.

Lemma 3.4 *Assume that*

$$\Delta_{ux} := \max_{u,\tilde{u}\in\mathcal{U},x,\tilde{x}\in X,v\in\mathcal{V},y\in Y} \{-F'(\tilde{u}, u-\tilde{u}|v) - \Phi'(\tilde{x}, x-\tilde{x}|y)\} < +\infty,$$
$$\Delta_{vy} := \max_{v,\tilde{v}\in\mathcal{V},y,\tilde{y}\in Y,u\in\mathcal{U},x\in X} \{F'(u|\tilde{v}, v-\tilde{v}) + \Phi'(x|\tilde{y}, y-\tilde{y})\} < +\infty.$$

If

$$\mathfrak{B}_r^{\mu,*}(0) \subseteq \{s = \mathcal{B}u + \tilde{x}_0 - x : u \in \mathcal{U}, x \in X\},$$
$$\mathfrak{B}_r^{\lambda,*}(0) \subseteq \{s = \mathcal{C}v + \tilde{y}_0 - y : v \in \mathcal{V}, y \in Y\}, \tag{18}$$

then $\|\mu^*\|_\mu \le \frac{\Delta_{ux}}{r}$ *and* $\|\lambda^*\|_\lambda \le \frac{\Delta_{vy}}{r}$.

Proof Consider the function $\psi(\lambda^*,\mu)$. It is concave and achieves its maximum at the point μ^*. From the representation (13) and similar representation for $\psi_2(\lambda,\mu)$, we conclude that the function $\psi(\lambda^*,\mu)$ is a minimum of a set of functions concave (linear) in the variable μ and convex in the variable (u,x). Hence, the set of minimizers

(u_μ, x_μ) is convex, and by the theorem about the subdifferential of a function which is a minimum of concave functions, we have

$$\partial_\mu \psi(\lambda^*, \mu) = \Big\{x_\mu - \tilde{x}_0 - \mathcal{B}u_\mu : u_\mu \in \mathcal{U}, x_\mu \in X, \\ \tilde{\psi}_1(u_\mu, \lambda^*) + \tilde{\psi}_2(x_\mu, \lambda^*) + \langle x_\mu - \tilde{x}_0 - \mathcal{B}u_\mu, \mu\rangle = \psi(\lambda^*, \mu)\Big\}.$$

From (15), and a similar definition of $\tilde{\psi}_2(x_\mu, \lambda^*)$, we get that

$$\tilde{\psi}_1'(u_\mu, u - u_\mu|\lambda^*) = F'(u_\mu, u - u_\mu|v^*(u_\mu, \lambda^*)),$$
$$\tilde{\psi}_2'(x_\mu, x - x_\mu|\lambda^*) = \Phi'(x_\mu, x - x_\mu|v^*(x_\mu, \lambda^*)).$$

From this equalities and the optimality conditions for the problem defining $\psi(\lambda^*, \mu)$, we get for all $u \in \mathcal{U}$, $x \in X$

$$F'(u_\mu, u - u_\mu|v^*(u_\mu, \lambda^*)) + \Phi'(x_\mu, x - x_\mu|v^*(x_\mu, \lambda^*)) \\ + \langle x - x_\mu - \mathcal{B}u + \mathcal{B}u_\mu, \mu\rangle \le 0,$$

or

$$\langle x - \mathcal{B}u - \tilde{x}_0, \mu\rangle \\ \le -F'(u_\mu, u - u_\mu|v^*(u_\mu, \lambda^*)) - \Phi'(x_\mu, x - x_\mu|v^*(x_\mu, \lambda^*)) + \langle \mu, \psi_\mu'(\lambda^*, \mu)\rangle.$$

Hence, for any μ,

$$\max_{u \in \mathcal{U}, x \in X} \langle x - \mathcal{B}u - \tilde{x}_0, \mu\rangle \le \Delta_{ux} + \langle \mu, \psi_\mu'(\lambda^*, \mu)\rangle.$$

Using the first inclusion in (18) and the fact that $0 \in \partial_\mu \psi(\lambda^*, \mu^*)$, we get that

$$\|\mu^*\|_\mu \le \frac{\Delta_{ux}}{r}.$$

The estimate for $\|\lambda^*\|_\lambda$ is proved in the same manner. □

Let us consider an example of sufficient conditions for the inclusion (18). If the set X has a nonempty interior and there exists some $\bar{u} \in \mathcal{U}$ such that $\mathcal{B}\bar{u} + \tilde{x}_0 = \bar{x} \in \text{int}X$, then there exists some $r > 0$ such that $\bar{x} + \mathfrak{B}_r^{\mu,*}(0) \subseteq X$. Then, we have

$$\mathfrak{B}_r^{\mu,*}(0) = \mathcal{B}\bar{u} + \tilde{x}_0 - \bar{x} + \mathfrak{B}_r^{\mu,*}(0) \subseteq \{s = \mathcal{B}u + \tilde{x}_0 - x : u \in \mathcal{U}, x \in X\}.$$

Similar arguments can be used for the second inclusion in (18).

Let us consider another example. Since the problem (4) does not have any constraints for x and y, and the sets X and Y can be introduced due to the boundedness of the norms of the operators $\mathcal{B}, \mathcal{C}$, we can apply the following reasoning. Assume that there exists some $\bar{u} \in P$ and $r_0 > 0$ such that $P \subseteq \mathfrak{B}_{r_0}^2(\bar{u})$. Then, for every $u \in P$, there exists some $\tilde{u} \in \mathfrak{B}^2_{r_0}(0)$: $\quad u = \bar{u} + \tilde{u}$. Let us define functions $\bar{u}(t) \equiv \bar{u}$, $\tilde{u}(t) := u(t) - \bar{u}(t)$, where $u(t) \in \mathcal{U}$ is arbitrary. Then, for $x_u(\theta) := \tilde{x}_0 + \mathcal{B}\bar{u}(t) + \mathcal{B}\tilde{u}(t)$, we have $\|x_u(\theta) - \tilde{x}_0 - \mathcal{B}\bar{u}(t)\|_{\mu,*} \le \|\mathcal{B}\tilde{u}(t)\|_{\mu,*} \le \|\mathcal{B}\|_{\mu,L^2_p} r_0\sqrt{\theta}$. So if we choose $X = \mathfrak{B}_{\|\mathcal{B}\|_{\mu,L^2_p} r_0\sqrt{\theta}}(\tilde{x}_0 + \mathcal{B}\bar{u}(t))$, we will be in the situation of the previous example and can take $r = \|\mathcal{B}\|_{\mu,L^2_p} r_0\sqrt{\theta}$ in the conditions of the Lemma 3.4. For the second inclusion in (18), we can apply a similar argument.

3.3 Algorithm Description

Note that, in the case when the dimensions n and m are rather small, then linearly converging cutting plane algorithms such as ellipsoids method, outer simplex method or inscribed ellipsoids method could be used to solve the conjugate problem and reconstruct the approximate solution of the initial problem. But their rate of convergence will depend on the dimensions n and m. Below we use primal-dual subgradient method to solve the conjugate problem and reconstruct the approximate solution of the initial problem. This allows us to construct a method with dimension-independent convergence rate.

We assume that we are given some prox function $d_\lambda(\lambda)$ with prox center λ_0, which is strongly convex with convexity parameter σ_λ in the given norm $\|\cdot\|_\lambda$. For μ, we introduce the similar assumptions.

Since (λ^*, μ^*) is the saddle point, by the definition, we have the following inequalities:

$$\psi(\lambda^*, \mu) \le \psi(\lambda^*, \mu^*) \le \psi(\lambda, \mu^*) \quad \forall \lambda, \mu.$$

From the convexity of the function $\psi(\lambda, \mu)$ with respect to λ, by the definition of partial subgradient $\psi'_\mu(\lambda, \mu)$ at the point (λ, μ), we have the following:

$$\psi(\lambda^*, \mu) \ge \psi(\lambda, \mu) + \langle \psi'_\lambda(\lambda, \mu), \lambda^* - \lambda \rangle \quad \forall \lambda, \mu.$$

Similarly, using concavity of $\psi(\lambda, \mu)$ with respect to μ, we have:

$$\psi(\lambda, \mu^*) \le \psi(\lambda, \mu) + \langle \psi'_\mu(\lambda, \mu), \mu^* - \mu \rangle \quad \forall \lambda, \mu$$

Finally, from the above inequalities, we have:

$$\langle \psi'_\lambda(\lambda, \mu), \lambda - \lambda^* \rangle + \langle -\psi'_\mu(\lambda, \mu), \mu - \mu^* \rangle \ge 0 \quad \forall \lambda, \mu.$$

Hence, (λ^*, μ^*) is a weak solution to the following variational inequality

$$\langle g(\lambda, \mu), (\lambda - \lambda^*, \mu - \mu^*) \rangle \ge 0, \quad \forall \lambda, \mu,$$

where $g(\lambda, \mu) := (\psi'_\lambda(\lambda, \mu), -\psi'_\mu(\lambda, \mu))$.

All of this allows us to apply the method of Simple Dual Averages (SDA) from [1] for finding an approximate solution of the finite-dimensional problem (8).

Let us choose some $\kappa \in]0, 1[$. As in Sect. 4 in [1], we consider a space of $z := (\lambda, \mu)$ with the norm

$$\|z\|_z := \sqrt{\kappa\sigma_\lambda \|\lambda\|_\lambda^2 + (1-\kappa)\sigma_\mu \|\mu\|_\mu^2}, \tag{19}$$

an oracle $g(z) := (g_\lambda(z), -g_\mu(z))$, a new prox function $d(z) := \kappa d_\lambda(\lambda) + (1-\kappa)\, d_\mu(\mu)$, which is strongly convex with constant $\sigma_0 = 1$ with respect to the norm (19). We define $W := \mathbb{R}^m \times \mathbb{R}^n$.

The conjugate norm for (19) is

$$\|g\|_{z,*} := \sqrt{\frac{1}{\kappa\sigma_\lambda}\|g_\lambda\|_{\lambda,*}^2 + \frac{1}{(1-\kappa)\sigma_\mu}\|g_\mu\|_{\mu,*}^2}.$$

So we have a uniform upper bound for the answers of the oracle $\|g(\lambda,\mu)\|_{z,*}^2 \le L^2 := \frac{L_\lambda^2}{\kappa\sigma_\lambda} + \frac{L_\mu^2}{(1-\kappa)\sigma_\mu}$, where L_λ and L_μ are defined in (17).

The SDA method for solving (8) is the following

1. Initialization: Set $s_0 = 0$. Choose $z_0, \gamma > 0$.
2. Iteration ($k \ge 0$):

$$\begin{aligned} &\text{Compute } g_k = g(z_k). \text{ Set } s_{k+1} = s_k + g_k.\\ &\beta_{k+1} = \gamma\hat{\beta}_{k+1}. \text{ Set } z_{k+1} = \pi_{\beta_{k+1}}(-s_{k+1}). \end{aligned} \tag{M1}$$

Here the sequence $\hat{\beta}_{k+1}$ is defined by relations $\hat{\beta}_0 = \hat{\beta}_1 = 1$, $\hat{\beta}_{i+1} = \hat{\beta}_i + \frac{1}{\hat{\beta}_i}$, for $i \ge 1$. In accordance with the Lemma 3 in [1], for $k \ge 1$, it satisfies the inequalities

$$\sqrt{2k-1} \le \hat{\beta}_k \le \frac{1}{1+\sqrt{3}} + \sqrt{2k-1}.$$

The mapping $\pi_\beta(s)$ is defined in the following way

$$\pi_\beta(s) := \arg\min_{z\in W} \left\{-\langle s, z\rangle + \beta d(z)\right\}.$$

Since the saddle point in the problem (7) does exist, there exists a saddle point (λ^*, μ^*) in the conjugate problem (8). According to the Theorem 1 in [1], the method (M1) generates a bounded sequence $\{z_i\}_{i\ge 0}$. Hence, the sequences $\{\lambda_i\}_{i\ge 0}, \{\mu_i\}_{i\ge 0}$ are also bounded. So we can choose D_λ, D_μ such that $d_\lambda(\lambda_i) \le D_\lambda$, $d_\mu(\mu_i) \le D_\mu$ for all $i \ge 0$ and also, the pair (λ^*, μ^*) is an interior solution: $\mathfrak{B}^\lambda_{r/\sqrt{\kappa\sigma_\lambda}}(\lambda^*) \subseteq W_\lambda := \{\lambda : d_\lambda(\lambda) \le D_\lambda\}$, and $\mathfrak{B}^\mu_{r/\sqrt{(1-\kappa)\sigma_\mu}}(\mu^*) \subseteq W_\mu := \left\{\mu : d_\mu(\mu) \le D_\mu\right\}$ for some $r > 0$. Then, we have $z^* := (\lambda^*, \mu^*) \in \mathcal{F}_D := \{z \in W : d(z) \le D\}$ with $D := \kappa D_\lambda + (1-\kappa) D_\mu$ and $\mathfrak{B}^z_r(z^*) \subseteq \mathcal{F}_D$.

Let us introduce a gap function

$$\delta_k(D) := \max_z \left\{ \sum_{i=0}^{k} \langle g_i, z_i - z \rangle : z \in \mathcal{F}_D \right\}. \tag{20}$$

From the Theorem 2 in [1] (equation (4.6)), we have

$$\frac{1}{k+1}\delta_k(D) \le \frac{\hat{\beta}_{k+1}}{k+1}\left(\gamma D + \frac{L^2}{2\gamma}\right). \tag{21}$$

Denote

$$\left(\hat{u}_{k+1}, \hat{v}_{k+1}, \hat{x}_{k+1}, \hat{y}_{k+1}\right) := \frac{1}{k+1}\sum_{i=0}^{k}(u_i, v_i, x_i, y_i), \tag{22}$$

where (u_i, v_i), (x_i, y_i) are the saddle points at the point (λ_i, μ_i) in (10) and (11), respectively. Note that for all $u \in \mathcal{U}$, $v \in \mathcal{V}$, $x \in X$, and $y \in Y$, we have

$$\begin{aligned} &F(u, v_i) + \Phi(x, y_i) + \langle \mu_i, x - \tilde{x}_0 - \mathcal{B}u \rangle + \langle \lambda_i, \mathcal{C}v_i + \tilde{y}_0 - y_i \rangle \ge \psi(\lambda_i, \mu_i) \\ &\ge F(u_i, v) + \Phi(x_i, y) + \langle \mu_i, x_i - \tilde{x}_0 - \mathcal{B}u_i \rangle + \langle \lambda_i, \mathcal{C}v + \tilde{y}_0 - y \rangle. \end{aligned} \tag{23}$$

We define a function

$$\begin{aligned} \phi(u, x, v, y) := \min_{\lambda} \max_{\mu} \{ & F(u, v) + \Phi(x, y) + \langle \mu, x - \tilde{x}_0 - \mathcal{B}u \rangle \\ & + \langle \lambda, \mathcal{C}v + \tilde{y}_0 - y \rangle : d_\lambda(\lambda) \le D_\lambda, d_\mu(\mu) \le D_\mu \}. \end{aligned} \tag{24}$$

Since $d_\lambda(\lambda^*) \le D_\lambda$, $d_\mu(\mu^*) \le D_\mu$, and the conjugate problem is equivalent to the initial one, we conclude that the initial problem is equivalent to the problem

$$\min_{u \in \mathcal{U}, x \in X} \max_{v \in \mathcal{V}, y \in Y} \phi(u, x, v, y). \tag{25}$$

Let us introduce two auxiliary functions:

$$\xi(u, x) := \max_{v \in \mathcal{V}, y \in Y} \phi(u, x, v, y), \tag{26}$$

$$\eta(v, y) := \min_{u \in \mathcal{U}, x \in X} \phi(u, x, v, y). \tag{27}$$

Note that $\xi(u, x)$ is convex, $\eta(v, y)$ is concave, and $\xi(u, x) \ge \phi(u^*, x^*, v^*, y^*) \ge \eta(v, y)$ for all $u \in \mathcal{U}$, $v \in \mathcal{V}$, $x \in X$, $y \in Y$, where $\phi(u^*, x^*, v^*, y^*)$ is the solution to (25).

Theorem 3.1 *Let the assumptions* **A1** *and* **A2** *be true. Then, the points* (22) *generated by the method* (M1) *satisfy:*

$$\xi(\hat{u}_{k+1}, \hat{x}_{k+1}) - \eta(\hat{v}_{k+1}, \hat{y}_{k+1}) \le \frac{\hat{\beta}_{k+1}}{k+1}\left(\gamma D + \frac{L^2}{2\gamma}\right), \tag{28}$$

$$\left\|\tilde{x}_0 + \mathcal{B}\hat{u}_{k+1} - \hat{x}_{k+1}\right\|_{\mu,*} \le \frac{\hat{\beta}_{k+1}\sqrt{\sigma_\mu}}{r(k+1)}\left(\gamma D + \frac{L^2}{2\gamma}\right),$$

$$\left\|\tilde{y}_0 + \mathcal{C}\hat{v}_{k+1} - \hat{y}_{k+1}\right\|_{\lambda,*} \le \frac{\hat{\beta}_{k+1}\sqrt{\sigma_\lambda}}{r(k+1)}\left(\gamma D + \frac{L^2}{2\gamma}\right). \tag{29}$$

Proof From the inequalities (23), by the convexity of $F(\cdot, v)$, $\Phi(\cdot, y)$, we have

$$\begin{aligned}
&F(\hat{u}_{k+1}, v) + \Phi(\hat{x}_{k+1}, y) + \langle \mu, \hat{x}_{k+1} - \tilde{x}_0 - \mathcal{B}\hat{u}_{k+1}\rangle + \langle \lambda, \mathcal{C}v + \tilde{y}_0 - y\rangle \\
&\le \frac{1}{k+1}\sum_{i=0}^{k}\psi(\lambda_i, \mu_i) + \frac{1}{k+1}\sum_{i=0}^{k}\langle \mu - \mu_i, x_i - \tilde{x}_0 - \mathcal{B}u_i\rangle \\
&\quad + \langle \lambda - \frac{1}{k+1}\sum_{i=0}^{k}\lambda_i, \mathcal{C}v + \tilde{y}_0 - y\rangle.
\end{aligned}$$

This gives us

$$\begin{aligned}
\xi(\hat{u}_{k+1}, \hat{x}_{k+1}) &\le \frac{1}{k+1}\sum_{i=0}^{k}\psi(\lambda_i, \mu_i) \\
&\quad + \max_{\mu}\left\{\frac{1}{k+1}\sum_{i=0}^{k}\langle \mu - \mu_i, x_i - \tilde{x}_0 - \mathcal{B}u_i\rangle : d_\mu(\mu) \le D_\mu\right\} \\
&\quad + \max_{v\in\mathcal{V}, y\in Y}\ \min_{\lambda}\left\{\langle \lambda - \frac{1}{k+1}\sum_{i=0}^{k}\lambda_i, \mathcal{C}v + \tilde{y}_0 - y\rangle : d_\lambda(\lambda) \le D_\lambda\right\}.
\end{aligned} \tag{30}$$

Since the method (M1) generates points λ_i, which satisfy $d_\lambda(\lambda_i) \le D_\lambda$, and there exist $v_1 \in \mathcal{V}$ and y_1 such that $\mathcal{C}v_1 + \tilde{y}_0 = y_1$, we conclude that $(\frac{1}{k+1}\sum_{i=0}^{k}\lambda_i, v_1, y_1)$ is the saddle point of the third term, and

$$\max_{v\in\mathcal{V}, y\in Y}\ \min_{\lambda}\left\{\langle \lambda - \frac{1}{k+1}\sum_{i=0}^{k}\lambda_i, \mathcal{C}v + \tilde{y}_0 - y\rangle : d_\lambda(\lambda) \le D_\lambda\right\} = 0.$$

Similarly, we have

$$\begin{aligned}
\eta(\hat{v}_{k+1}, \hat{y}_{k+1}) &\ge \frac{1}{k+1}\sum_{i=0}^{k}\psi(\lambda_i, \mu_i) \\
&\qquad - \max_{\lambda}\left\{\frac{1}{k+1}\sum_{i=0}^{k}\langle \lambda_i - \lambda, \mathcal{C}v_i + \tilde{y}_0 - y_i\rangle : d_\lambda(\lambda) \le D_\lambda\right\}.
\end{aligned}$$

Finally, we have the following

$$
\begin{aligned}
&\xi(\hat{u}_{k+1}, \hat{x}_{k+1}) - \eta(\hat{v}_{k+1}, \hat{y}_{k+1}) \\
&\quad \le \frac{1}{k+1}\left(\max_{\mu} \left\{ \sum_{i=0}^{k} \langle \mu - \mu_i, x_i - \tilde{x}_0 - \mathcal{B}u_i \rangle : d_{\mu}(\mu) \le D_{\mu} \right\} \right. \\
&\qquad \left. + \max_{\lambda} \left\{ \sum_{i=0}^{k} \langle \lambda_i - \lambda, \mathcal{C}v_i + \tilde{y}_0 - y_i \rangle : d_{\lambda}(\lambda) \le D_{\lambda} \right\} \right) \\
&\quad \le \frac{1}{k+1} \max_{z} \left\{ \sum_{i=0}^{k} \langle g_i(z), z_i - z \rangle : d(z) \le \kappa D_{\lambda} + (1-\kappa) D_{\mu} \right\} = \frac{1}{k+1} \delta_k(D).
\end{aligned}
$$

Combining this with (21), we get (28).

Let us prove that (22) is also a nearly feasible solution. Obviously,

$$
\begin{aligned}
\frac{1}{(1-\kappa)\sigma_{\mu}} \left\| \mathcal{B}\hat{u}_{k+1} + \tilde{x}_0 - \hat{x}_{k+1} \right\|_{\mu,*}^2 &\le \frac{1}{(1-\kappa)\sigma_{\mu}} \left\| \mathcal{B}\hat{u}_{k+1} + \tilde{x}_0 - \hat{x}_{k+1} \right\|_{\mu,*}^2 \\
&\quad + \frac{1}{\kappa\sigma_{\lambda}} \left\| \mathcal{C}\hat{v}_{k+1} + \tilde{y}_0 - \hat{y}_{k+1} \right\|_{\lambda,*}^2, \\
\frac{1}{\kappa\sigma_{\lambda}} \left\| \mathcal{C}\hat{v}_{k+1} + \tilde{y}_0 - \hat{y}_{k+1} \right\|_{\lambda,*}^2 & \\
&\le \frac{1}{(1-\kappa)\sigma_{\mu}} \left\| \mathcal{B}\hat{u}_{k+1} + \tilde{x}_0 - \hat{x}_{k+1} \right\|_{\mu,*}^2 \\
&\quad + \frac{1}{\kappa\sigma_{\lambda}} \left\| \mathcal{C}\hat{v}_{k+1} + \tilde{y}_0 - \hat{y}_{k+1} \right\|_{\lambda,*}^2.
\end{aligned}
$$

On the other hand, from the proof of the third item in the Theorem 1 in [1], we have

$$
\begin{aligned}
\left[\frac{1}{r(k+1)} \delta_k(D) \right]^2 \ge \left\| \hat{s}_{k+1} \right\|_{z,*}^2 &= \left\| \frac{1}{k+1} \sum_{i=0}^{k} (g_{\lambda}(z_i), -g_{\mu}(z_i)) \right\|_{z,*}^2 \\
&= \left\| \frac{1}{k+1} \sum_{i=0}^{k} (\mathcal{C}v_i + \tilde{y}_0 - y_i, \mathcal{B}u_i + \tilde{x}_0 - x_i) \right\|_{z,*}^2 \\
&= \frac{1}{(1-\kappa)\sigma_{\mu}} \left\| \mathcal{B}\hat{u}_{k+1} + \tilde{x}_0 - \hat{x}_{k+1} \right\|_{\mu,*}^2 \\
&\quad + \frac{1}{\kappa\sigma_{\lambda}} \left\| \mathcal{C}\hat{v}_{k+1} + \tilde{y}_0 - \hat{y}_{k+1} \right\|_{\lambda,*}^2.
\end{aligned}
$$

This in combination with (21) gives us (29). □

So we conclude that $(\hat{u}_{k+1}, \hat{v}_{k+1}, \hat{x}_{k+1}, \hat{y}_{k+1})$ is a nearly optimal and nearly feasible point with an error of $O\left(\frac{1}{\sqrt{k+1}}\right)$.

Let us make some notes on how to choose parameters γ, D, and r of the method. Recall the example introduced in the Sect. 3.1, but let us choose the following quality functional

$$J(u, v) = \Phi(x(1), y(1)) = \frac{1}{2}\|x(1) - y(1)\|_2^2 - \|y(1) - y_0\|_2^2,$$

where $y_0 = (2, 0)^T$ and norms are Euclidean. Hence, we have $F(u, v) \equiv 0$ and the assumptions **A1** and **A2** hold. Let us use the method described in Sect. 2.1 to estimate the norms of the operators $\mathcal{B}$ and $\mathcal{C}$ in this example. We choose Euclidean norm as $\|\cdot\|_\lambda, \|\cdot\|_\mu$. Since the operators are similar, we consider operator $\mathcal{B}$. To estimate $\|\mathcal{B}\|_2$, we need to calculate the maximum eigenvalue of the matrix

$$\int_0^1 V_x(1, t)B(t)B^T(t)V_x^T(1, t)dt = \begin{pmatrix} \frac{1}{3} & \frac{1}{2} \\ \frac{1}{2} & 1 \end{pmatrix}.$$

We get that $\|\mathcal{B}\|_2 = \sqrt{\lambda_{max}} \approx 1.12$. Next, we need to introduce sets X and Y. We know that $x(\theta) = \tilde{x}_0 + \mathcal{B}u$ and $u(t) \in [-1, 1]$ for all $t \in [0, 1]$. Hence, $\|x(\theta) - \tilde{x}_0\|_2 \leq \|\mathcal{B}u\|_2 \leq \|\mathcal{B}\|_2 = 1.12$ and we can take $X = \mathfrak{B}^2_{1.12}(\tilde{x}_0)$. Similarly, $Y = \mathfrak{B}^2_{1.12}(\tilde{y}_0)$. Also we can find the bound for $\|\psi'_\lambda(\lambda, \mu)\|_2 = \|x - \tilde{x}_0 + \mathcal{B}u\|_2 \leq \|\mathcal{B}u\|_2 + \|x - \tilde{x}_0\|_2 \leq 2\|\mathcal{B}\|_2 = L_\lambda = 2.24$. Similarly, $\|\psi'_\mu(\lambda, \mu)\|_2 \leq 2\|\mathcal{C}\|_2 = L_\mu = 2.24$. Next, we need to find D_λ, D_μ, and D. From the second example after Lemma 3.4 since $\|\mathcal{B}\|_2 = \|\mathcal{C}\|_2 = 1.12$ and $P = Q = [-1, 1]$, we conclude that we can take $r = 1.12$ in both inclusions (18). The estimation of Δ_{ux}, Δ_{vy} gives $\Delta_{ux} \leq 10, \Delta_{vy} \leq 10$. Hence, $\|\lambda^*\|_2 \leq 9$, $\|\mu^*\|_2 \leq 9$.

Using the method above, we can find a nearly feasible nearly optimal solution. We choose the prox functions $d_\lambda(\lambda) := \frac{\|\lambda\|_2^2}{2}$, $d_\mu(\mu) := \frac{\|\mu\|_2^2}{2}$, $\kappa = 1/2$. We choose $\gamma = L/\sqrt{2D}$ to minimize the right-hand side of the error estimations in the Theorem 3.1. Then, we have $d(z) = \frac{\|\lambda\|_2^2}{4} + \frac{\|\mu\|_2^2}{4}$, $\|z^*\| \leq 9$.

From the Theorem 2 in [1], we get that for the sequence $\{z_k\}$ generated by SDA method, it holds that

$$\|z_k - z^*\|_z^2 \leq 2d(z^*) + L^2 = \|z^*\|_z^2 + L^2.$$

Then,

$$d(z_k) = \frac{\|z_k\|_z^2}{2} \leq \frac{1}{2}(\|z^*\|_z + \sqrt{\|z^*\|_z^2 + L^2})^2 \leq 182.$$

So $\|\lambda_k\|_2^2/2 \leq 364$, and we can take $D_\lambda = 364$. In the same way, we get that $D_\mu = 364$ and also that $D = 364$. Finally, we have that $\mathfrak{B}^z_{12.5}(z^*) \subseteq \mathcal{F}_D$ and $r = 12.5$.

4 Strongly Convex–Concave Problem

In this section, we consider the problem (4), satisfying the following assumptions.

A3 The function $F(\cdot, v)$ is strongly convex for any fixed v with constant σ_{F_u} which does not depend on v, and function $F(u, \cdot)$ is strongly concave for any fixed u with constant σ_{F_v} which does not depend on u. Also assume that:

$$\|\nabla_u F(u, v_1) - \nabla_u F(u, v_2)\|_{L_p^2} \leq L_{uv} \|v_1 - v_2\|_{L_q^2}, \tag{31}$$

$$\|\nabla_v F(u_1, v) - \nabla_v F(u_2, v)\|_{L_q^2} \leq L_{vu} \|u_1 - u_2\|_{L_p^2}. \tag{32}$$

A4 $\Phi(\cdot, y)$ is strongly convex for any fixed y with respect to the norm $\|\cdot\|_{\mu,*}$ with constant $\sigma_{\Phi x}$ which does not depend on y and $\Phi(x, \cdot)$ is strongly concave for any fixed x with respect to the norm $\|\cdot\|_{\lambda,*}$ with constant $\sigma_{\Phi y}$ which does not depend on x. Also we assume that:

$$\|\nabla_x \Phi(x, y_1) - \nabla_x \Phi(x, y_2)\|_{\mu} \leq L_{xy} \|y_1 - y_2\|_{\lambda,*}, \tag{33}$$

$$\left\|\nabla_y \Phi(x_1, y) - \nabla_y \Phi(x_2, y)\right\|_{\lambda} \leq L_{yx} \|x_1 - x_2\|_{\mu,*}, \tag{34}$$

and

$$\|\nabla_x \Phi(x_1, y) - \nabla_x \Phi(x_2, y)\|_{\mu} \leq L_{xx} \|x_1 - x_2\|_{\mu,*}, \tag{35}$$

$$\left\|\nabla_y \Phi(x, y_1) - \nabla_y \Phi(x, y_2)\right\|_{\lambda} \leq L_{yy} \|y_1 - y_2\|_{\lambda,*}. \tag{36}$$

Note that the assumptions **A3**, **A4** imply that the level sets of the functions $F(u, v)$, $\Phi(x, y)$ are closed, convex, and bounded. Similarly to the proof of the Lemma 3.1, we get that the conjugate problem for (4) is

$$\min_{\lambda} \max_{\mu} \left\{ \min_{u \in \mathcal{U}} \max_{v \in \mathcal{V}} [F(u, v) - \langle \mu, \mathcal{B}u \rangle + \langle \lambda, \mathcal{C}v \rangle] \right.$$
$$\left. + \min_{x} \max_{y} [\Phi(x, y) + \langle \mu, x \rangle - \langle \lambda, y \rangle] - \langle \mu, \tilde{x}_0 \rangle + \langle \lambda, \tilde{y}_0 \rangle \right\}. \tag{37}$$

Here $\lambda \in \mathbb{R}^m$ and $\mu \in \mathbb{R}^n$.

We assume that the problems

$$\psi_1(\lambda, \mu) := \min_{u \in \mathcal{U}} \max_{v \in \mathcal{V}} [F(u, v) - \langle \mu, \mathcal{B}u \rangle + \langle \lambda, \mathcal{C}v \rangle], \tag{38}$$

$$\psi_2(\lambda, \mu) := \min_{x} \max_{y} [\Phi(x, y) + \langle \mu, x \rangle - \langle \lambda, y \rangle] \tag{39}$$

are simple, which means that they can be solved efficiently or in a closed form (see the example in the Sect. 3). Note that the conjugate problem is finite dimensional. Using the assumptions **A3**, **A4**, the fact that closed, convex, and bounded set in Hilbert space is compact in the weak topology, the fact that $F(u, v)$ is upper semi-continuous in v and lower semi-continuous in u, and Corollary 3.3 in [4], we conclude that the saddle points in the problems (38), (39) do exist for all $\lambda \in \mathbb{R}^m$ and $\mu \in \mathbb{R}^n$.

Lemma 4.1 *Let the Assumption* **A3** *be true. Then, the function* $\psi_1(\lambda, \mu)$ *in* (38) *is smooth with the partial gradients satisfying the following Lipschitz condition:*

$$\left\|\nabla_\mu \psi_1(\lambda_1, \mu_1) - \nabla_\mu \psi_1(\lambda_2, \mu_2)\right\|_{\mu,*} \le \frac{\|\mathcal{B}\|^2_{\mu, L^2_p}}{\sigma_{F_u}} \|\mu_1 - \mu_2\|_\mu + \frac{\|\mathcal{B}\|_{\mu, L^2_p} \|\mathcal{C}\|_{\lambda, L^2_q} L_{uv}}{\sigma_{F_u}\sigma_{F_v}} \|\lambda_1 - \lambda_2\|_\lambda, \tag{40}$$

$$\|\nabla_\lambda \psi_1(\lambda_1, \mu_1) - \nabla_\lambda \psi_1(\lambda_2, \mu_2)\|_{\lambda,*} \le \frac{\|\mathcal{C}\|^2_{\lambda, L^2_q}}{\sigma_{F_v}} \|\lambda_1 - \lambda_2\|_\lambda + \frac{\|\mathcal{B}\|_{\mu, L^2_p} \|\mathcal{C}\|_{\lambda, L^2_q} L_{vu}}{\sigma_{F_u}\sigma_{F_v}} \|\mu_1 - \mu_2\|_\mu. \tag{41}$$

Proof From the strong convexity of the function $F(\cdot, v)$, we have for any $t \in [0, 1]$

$$\begin{aligned}
&\tilde{\psi}_1(tu_1 + (1-t)u_2, \lambda) = \max_{v \in \mathcal{V}} [F(tu_1 + (1-t)u_2, v) + \langle \mathcal{C}v, \lambda\rangle] \\
&\le \max_{v \in \mathcal{V}} \left[tF(u_1, v) + (1-t)F(u_2, v) - t(1-t)\frac{\sigma_{Fu}}{2} \|u_1 - u_2\|^2_{L^2_p} + \langle \mathcal{C}v, \lambda\rangle\right] \\
&\le t \max_{v \in \mathcal{V}} [F(u_1, v) + \langle \mathcal{C}v, \lambda\rangle] + (1-t) \max_{v \in \mathcal{V}} [F(u_2, v) + \langle \mathcal{C}v, \lambda\rangle] \\
&- t(1-t)\frac{\sigma_{Fu}}{2} \|u_1 - u_2\|^2_{L^2_p} = t\tilde{\psi}_1(u_1, \lambda) + (1-t)\tilde{\psi}_1(u_2, \lambda) \\
&- t(1-t)\frac{\sigma_{Fu}}{2} \|u_1 - u_2\|^2_{L^2_p}
\end{aligned}$$

So, by definition, the function $\tilde{\psi}_1(u, \lambda)$ is strongly convex with constant σ_{F_u}. This means that the optimal point u^* in (13) is unique and that $\psi_1(\lambda, \mu)$ is smooth with respect to μ. Hence, $\nabla_\mu \psi_1(\lambda, \mu) = -\mathcal{B}u^*(\lambda, \mu)$. Since $F(u, \cdot)$ is strongly concave for any fixed u, we get that the solution v^* of (15) is unique and the function $\tilde{\psi}_1(u, \lambda)$ is smooth with respect to u. Denote by u_i the optimal point in (13) for some λ, μ_i, $i = 1, 2$. From the first-order optimality conditions for (13), we have

$$\langle \nabla_u \tilde{\psi}_1(u_1, \lambda) - \mathcal{B}^* \mu_1, u_2 - u_1\rangle \ge 0,$$
$$\langle \nabla_u \tilde{\psi}_1(u_2, \lambda) - \mathcal{B}^* \mu_2, u_1 - u_2\rangle \ge 0.$$

Adding these inequalities and using strong convexity of $\tilde{\psi}_1(u, \lambda)$, we continue as follows:

$$\langle \mathcal{B}^*(\mu_1 - \mu_2), u_1 - u_2\rangle \ge \langle \nabla_u \tilde{\psi}_1(u_1, \lambda) - \nabla_u \tilde{\psi}_1(u_2, \lambda), u_1 - u_2\rangle \ge \sigma_{F_u} \|u_1 - u_2\|^2_{L^2_p}$$

Finally, we have

$$\|\mathcal{B}u_1 - \mathcal{B}u_2\|^2_{\mu,*} \le \|\mathcal{B}\|^2_{\mu,L^2_p} \|u_1 - u_2\|^2_{L^2_p} \le \frac{\|\mathcal{B}\|^2_{\mu,L^2_p}}{\sigma_{F_u}} \langle \mathcal{B}^*(\mu_1 - \mu_2), u_1 - u_2 \rangle$$
$$\le \frac{\|\mathcal{B}\|^2_{\mu,L^2_p}}{\sigma_{F_u}} \|\mu_1 - \mu_2\|_\mu \|\mathcal{B}(u_1 - u_2)\|_{\mu,*} \tag{42}$$

Denote by (u_i, v_i) the saddle point in (38) for some λ_i, μ and $i = 1, 2$. Similarly to the previous case, we conclude that $\hat{\psi}_1(v, \mu)$ is strongly concave in v with constant σ_{F_v} and smooth with respect to v. As we did this above, from the first-order optimality conditions for (14) and using the strong concavity of $\hat{\psi}_1(v, \mu)$, we have:

$$\langle \mathcal{C}^*(\lambda_1 - \lambda_2), v_1 - v_2 \rangle \ge \sigma_{F_v} \|v_1 - v_2\|^2_{L^2_q}.$$

This gives us

$$\|v_1 - v_2\|_{L^2_q} \le \frac{\|\mathcal{C}\|_{\lambda,L^2_q}}{\sigma_{F_v}} \|\lambda_1 - \lambda_2\|_\lambda.$$

From the first-order optimality conditions for (16), we get:

$$\langle \nabla_u F(u_2, v_2) - \nabla_u F(u_1, v_1), u_1 - u_2 \rangle \ge 0.$$

Using that $F(\cdot, v)$ is strongly convex for any fixed v, this gives us

$$\langle \nabla_u F(u_2, v_2) - \nabla_u F(u_2, v_1), u_1 - u_2 \rangle \ge \sigma_{F_u} \|u_1 - u_2\|^2_{L^2_p}.$$

From this, using (31) and the estimate for $\|v_1 - v_2\|_{L^2_q}$, we get

$$\|u_1 - u_2\|_{L^2_p} \le \frac{\|\mathcal{C}\|_{\lambda,L^2_q} L_{uv}}{\sigma_{F_u}\sigma_{F_v}} \|\lambda_1 - \lambda_2\|_\lambda.$$

Finally ,we have

$$\|\mathcal{B}u_1 - \mathcal{B}u_2\|_{\mu,*} \le \|\mathcal{B}\|_{\mu,L^2_p} \|u_1 - u_2\|_{L^2_p} \le \frac{\|\mathcal{C}\|_{\lambda,L^2_q} \|\mathcal{B}\|_{\mu,L^2_p} L_{uv}}{\sigma_{F_u}\sigma_{F_v}} \|\lambda_1 - \lambda_2\|_\lambda. \tag{43}$$

Combining (42) and (43), we get (40). Estimate (41) can be proved in the same manner. □

In a similar way, we can prove the following statement.

Lemma 4.2 *Let the Assumption* **A4** *be true. Then, the function* $\psi_2(\lambda, \mu)$ *in* (39) *is smooth with the partial gradients satisfying the following Lipschitz condition:*

$$\left\|\nabla_\mu \psi_2(\lambda_1, \mu_1) - \nabla_\mu \psi_2(\lambda_2, \mu_2)\right\|_{\mu,*} \leq \frac{1}{\sigma_{\Phi x}} \|\mu_1 - \mu_2\|_\mu + \frac{L_{xy}}{\sigma_{\Phi x}\sigma_{\Phi y}} \|\lambda_1 - \lambda_2\|_\lambda, \tag{44}$$

$$\|\nabla_\lambda \psi_2(\lambda_1, \mu_1) - \nabla_\lambda \psi_2(\lambda_2, \mu_2)\|_{\lambda,*} \leq \frac{1}{\sigma_{\Phi y}} \|\lambda_1 - \lambda_2\|_\lambda + \frac{L_{yx}}{\sigma_{\Phi x}\sigma_{\Phi y}} \|\mu_1 - \mu_2\|_\mu. \tag{45}$$

Combining Lemma 4.1 and 4.2, we get that the function $\psi(\lambda, \mu) := \psi_1(\lambda, \mu) + \psi_2(\lambda, \mu) - \langle \mu, \tilde{x}_0 \rangle + \langle \lambda, \tilde{y}_0 \rangle$ is convex in λ and concave in μ with partial gradients

$$\nabla_\lambda \psi(\lambda, \mu) = \mathcal{C}v^* + \tilde{y}_0 - y^*, \tag{46}$$

$$\nabla_\mu \psi(\lambda, \mu) = x^* - \tilde{x}_0 - \mathcal{B}u^*, \tag{47}$$

where (u^*, v^*) is the saddle point for the problem (38), and (x^*, y^*) is the saddle point for the problem (39).

4.1 Estimating the Norms of λ^*, μ^*

Let us find the bounds for the norms of the components λ, μ of the solution of the conjugate problem (37).

Lemma 4.3 *Let the Assumptions* **A3**, **A4** *be true. Assume that* $P \subseteq \mathfrak{B}^2_{r_1}(u_0)$ *and* $Q \subseteq \mathfrak{B}^2_{r_2}(v_0)$*, where* $u_0 \in \mathbb{R}^p$, $v_0 \in \mathbb{R}^q$*. Then,*

$$\|\mu^*\|_\mu \leq L_{xx}\|\mathcal{B}\|_{\mu,L^2_p} r_1\sqrt{\theta} + L_{xy}\|\mathcal{C}\|_{\lambda,L^2_q} r_2\sqrt{\theta} + \|\nabla_x \Phi(\mathcal{B}u_0 + \tilde{x}_0, \mathcal{C}v_0 + \tilde{y}_0)\|_\mu,$$

$$\|\lambda^*\|_\lambda \leq L_{yx}\|\mathcal{B}\|_{\mu,L^2_p} r_1\sqrt{\theta} + L_{yy}\|\mathcal{C}\|_{\lambda,L^2_q} r_2\sqrt{\theta} + \|\nabla_y \Phi(\mathcal{B}u_0 + \tilde{x}_0, \mathcal{C}v_0 + \tilde{y}_0)\|_\lambda.$$

Proof Consider the function $\psi(\lambda^*, \mu)$. From Lemmas 4.1 and 4.2, we know that this function is concave and smooth in μ with the gradient given by (47). Also, the function $\psi(\lambda^*, \mu)$ achieves its maximum at $\mu = \mu^*$. From the corresponding optimality condition, we get $0 = \nabla_\mu \psi(\lambda^*, \mu^*) = x^* - \mathcal{B}u^* - \tilde{x}_0$. Hence, $x^* = \mathcal{B}u^* + \tilde{x}_0$. Similarly, we have $y^* = \mathcal{C}v^* + \tilde{y}_0$.

We can find also x^*, y^* from the problem (39), which can be rewritten as

$$\psi_2(\lambda^*, \mu^*) = \min_x\{\langle \mu^*, x \rangle + \tilde{\psi}_2(x, \lambda^*)\}, \quad \tilde{\psi}_2(x, \lambda^*) := \max_y\{\Phi(x, y) - \langle \lambda^*, y \rangle\}.$$

Since $\Phi(x, y)$ is strongly convex in y, from the optimality condition for the first problem, we obtain that $\mu^* = -\nabla_x \tilde{\psi}_2(x^*, \lambda^*) = -\nabla_x \Phi(x^*, y^*)$.

Let us introduce functions $u_0(t) \equiv u_0,\ v_0(t) \equiv v_0$ Finally, we have the following sequence of inequalities

$$\begin{aligned}
\|\mu^*\|_\mu &= \|\nabla_x \Phi(x^*, y^*)\|_\mu \\
&\le \|\nabla_x \Phi(\mathcal{B}u^* - \tilde{x}_0, \mathcal{C}v^* + \tilde{y}_0) - \nabla_x \Phi(\mathcal{B}u_0(t) + \tilde{x}_0, \mathcal{C}v_0(t) + \tilde{y}_0)\|_\mu \\
&\quad + \|\nabla_x \Phi(\mathcal{B}u_0(t) + \tilde{x}_0, \mathcal{C}v_0(t) + \tilde{y}_0)\|_\mu \le \\
&\le L_{xx}\|\mathcal{B}u^* - \mathcal{B}u_0(t)\|_{\mu,*} + L_{xy}\|\mathcal{C}v^* - \mathcal{C}v_0(t)\|_{\lambda,*} \\
&\quad + \|\nabla_x \Phi(\mathcal{B}u_0(t) + \tilde{x}_0, \mathcal{C}v_0(t) + \tilde{y}_0)\|_\mu \\
&\le L_{xx}\|\mathcal{B}\|_{\mu,L_p^2} r_1\sqrt{\theta} + L_{xy}\|\mathcal{C}\|_{\lambda,L_q^2} r_2\sqrt{\theta} + \|\nabla_x \Phi(\mathcal{B}u_0(t) + \tilde{x}_0, \mathcal{C}v_0(t) + \tilde{y}_0)\|_\mu.
\end{aligned}$$

The proof of the second inequality in the statement of the lemma is similar. □

4.2 Algorithm Description

In this section, we assume that the norms $\|\cdot\|_\lambda$ and $\|\cdot\|_\mu$ are Euclidian. Let us introduce the prox function $d_\lambda(\lambda) := \frac{\sigma_\lambda}{2}\|\lambda\|_\lambda^2$. The function $d_\lambda(\lambda)$ is strongly convex in this norm with the convexity parameter σ_λ. For the variable μ, we introduce the prox function $d_\mu(\mu) := \frac{\sigma_\mu}{2}\|\mu\|_\mu^2$, which is strongly convex with the convexity parameter σ_μ with respect to the norm $\|\cdot\|_\mu$. These prox functions are differentiable everywhere.

For any $\lambda_1, \lambda_2 \in \mathbb{R}^m$ we can define the Bregman distance:

$$\omega_\lambda(\lambda_1, \lambda_2) := d_\lambda(\lambda_2) - d_\lambda(\lambda_1) - \langle \nabla d_\lambda(\lambda_1), \lambda_2 - \lambda_1 \rangle.$$

Using the explicit expression for $d_\lambda(\lambda)$, we get $\omega_\lambda(\lambda_1, \lambda_2) = \frac{\sigma_\lambda}{2}\|\lambda_1 - \lambda_2\|^2$. Let us choose $\bar{\lambda} = 0$ as the center of the space $\mathbb{R}^m$. Then, we have $\omega_\lambda(\bar{\lambda}, \lambda) = d_\lambda(\lambda)$. For μ, we introduce the similar settings.

In the same way as it was done in the Section 3.3, we conclude that finding the saddle point (λ^*, μ^*) for the conjugate problem (37) is equivalent to solving the variational inequality

$$\langle g(\lambda, \mu), (\lambda - \lambda^*, \mu - \mu^*) \rangle \ge 0, \quad \forall \lambda, \mu, \tag{48}$$

where

$$g(\lambda, \mu) := (\nabla_\lambda \psi(\lambda, \mu), -\nabla_\mu \psi(\lambda, \mu)). \tag{49}$$

Let us choose some $\kappa \in]0, 1[$. Consider a space of $z := (\lambda, \mu)$ with the norm

$$\|z\|_z := \sqrt{\kappa\sigma_\lambda \|\lambda\|_\lambda^2 + (1-\kappa)\sigma_\mu \|\mu\|_\mu^2},$$

an oracle $g(z) := (\nabla_\lambda \psi(\lambda, \mu), -\nabla_\mu \psi(\lambda, \mu))$, a new prox function

$$d(z) := \kappa d_\lambda(\lambda) + (1-\kappa) d_\mu(\mu)$$

which is strongly convex with constant $\sigma_0 = 1$. We define $W := \mathbb{R}^m \times \mathbb{R}^n$, the Bregman distance

$$\omega(z_1, z_2) := \kappa\omega_\lambda(\lambda_1, \lambda_2) + (1-\kappa)\omega_\lambda(\mu_2, \mu_2)$$

which has an explicit form of $\omega(z_1, z_2) = d(z_1 - z_2)$, and center $\bar{z} = (0, 0)$. Then, we have $\omega(\bar{z}, z) = d(z)$. Note that the norm in the dual space is defined as

$$\|g\|_{z,*} := \sqrt{\frac{1}{\kappa\sigma_\lambda}\|g_\lambda\|_{\lambda,*}^2 + \frac{1}{(1-\kappa)\sigma_\mu}\|g_\mu\|_{\mu,*}^2}.$$

Lemma 4.4 *Let the Assumptions* **A3**, **A4** *be true, and* $\kappa = \frac{\sigma_\mu}{\sigma_\mu + \sigma_\lambda}$. *Then, the operator* $g(z)$ *defined in* (49) *is Lipschitz continuous:*

$$\|g(z_1) - g(z_2)\|_{z,*} \le L\,\|z_1 - z_2\|_z \tag{50}$$

with

$$L = \frac{\sigma_\lambda + \sigma_\mu}{\sigma_\mu\sigma_\lambda}\sqrt{2\left(\frac{\|\mathcal{C}\|_{\lambda,L_q^2}^2}{\sigma_{F_v}} + \frac{1}{\sigma_{\Phi_y}} + \frac{\|\mathcal{B}\|_{\mu,L_p^2}\,\|\mathcal{C}\|_{\lambda,L_q^2}\,L_{vu}}{\sigma_{F_u}\sigma_{F_v}} + \frac{L_{yx}}{\sigma_{\Phi x}\sigma_{\Phi_y}}\right)}$$
$$\sqrt{\left(\frac{\|\mathcal{B}\|_{\mu,L_p^2}\,\|\mathcal{C}\|_{\lambda,L_q^2}\,L_{uv}}{\sigma_{F_u}\sigma_{F_v}} + \frac{L_{xy}}{\sigma_{\Phi_x}\sigma_{\Phi y}} + \frac{\|\mathcal{B}\|_{\mu,L_p^2}^2}{\sigma_{F_u}} + \frac{1}{\sigma_{\Phi_x}}\right)}. \tag{51}$$

Proof Denote

$$c := \|\lambda_1 - \lambda_2\|_\lambda, \qquad d := \|\mu_1 - \mu_2\|_\mu,$$
$$\alpha_1 := \frac{\|\mathcal{C}\|_{\lambda,L_q^2}^2}{\sigma_{F_v}} + \frac{1}{\sigma_{\Phi_y}}, \qquad \alpha_2 := \frac{\|\mathcal{B}\|_{\mu,L_p^2}\,\|\mathcal{C}\|_{\lambda,L_q^2}\,L_{uv}}{\sigma_{F_u}\sigma_{F_v}} + \frac{L_{xy}}{\sigma_{\Phi_x}\sigma_{\Phi_y}},$$
$$\beta_1 := \frac{\|\mathcal{B}\|_{\mu,L_p^2}\,\|\mathcal{C}\|_{\lambda,L_q^2}\,L_{vu}}{\sigma_{F_u}\sigma_{F_v}} + \frac{L_{yx}}{\sigma_{\Phi_x}\sigma_{\Phi_y}},\ \beta_2 := \frac{\|\mathcal{B}\|_{\mu,L_p^2}^2}{\sigma_{F_u}} + \frac{1}{\sigma_{\Phi_x}}.$$

Then, from Eqs. (40), (41), (44), (45) we have:

$$\|\nabla_\lambda\psi(\lambda_1, \mu_1) - \nabla_\lambda\psi(\lambda_2, \mu_2)\|_{\lambda,*}^2 \le (\alpha_1 c + \beta_1 d)^2,$$
$$\left\|\nabla_\mu\psi(\lambda_1, \mu_1) - \nabla_\mu\psi(\lambda_2, \mu_2)\right\|_{\mu,*}^2 \le (\alpha_2 c + \beta_2 d)^2.$$

Since $\kappa = \frac{\sigma_\mu}{\sigma_\mu + \sigma_\lambda}$, we get

$$\kappa\sigma_\lambda = (1-\kappa)\sigma_\mu = \frac{\sigma_\mu\sigma_\lambda}{\sigma_\mu + \sigma_\lambda} := \sigma.$$

Using the above expressions we obtain

$$\begin{aligned}
\|g(z_1) - g(z_2)\|_{z,*}^2 &= \frac{1}{\kappa\sigma_\lambda}\,\|\nabla_\lambda\psi(\lambda_1,\mu_1) - \nabla_\lambda\psi(\lambda_2,\mu_2)\|_{\lambda,*}^2 \\
&+ \frac{1}{(1-\kappa)\sigma_\mu}\,\left\|\nabla_\mu\psi(\lambda_1,\mu_1) - \nabla_\mu\psi(\lambda_2,\mu_2)\right\|_{\mu,*}^2 \\
&\le \frac{1}{\sigma}\,(\alpha_1 c + \beta_1 d)^2 + \frac{1}{\sigma}\,(\alpha_2 c + \beta_2 d)^2 \le \frac{2}{\sigma}(\alpha_1 c + \beta_1 d)(\alpha_2 c + \beta_2 d) \\
&\le \frac{2}{\sigma}(\alpha_1\alpha_2 c^2 + \beta_1\beta_2 d^2 + (\alpha_1\beta_2 + \alpha_2\beta_1)cd) \\
&\le \frac{1}{\sigma}((2\alpha_1\alpha_2 + \alpha_1\beta_2 + \alpha_2\beta_1)c^2 + (2\beta_1\beta_2 + \alpha_1\beta_2 + \alpha_2\beta_1)d^2) \\
&\le \frac{2}{\sigma}(\sqrt{\alpha_1\alpha_2(\alpha_1+\beta_1)(\alpha_2+\beta_2)}c^2 + \sqrt{\beta_1\beta_2(\alpha_1+\beta_1)(\alpha_2+\beta_2)}d^2) \\
&\le \frac{2(\alpha_1+\beta_1)(\alpha_2+\beta_2)}{\sigma^2}(\kappa\sigma_\lambda c^2 + (1-\kappa)\sigma_\mu d^2) \\
&= \frac{2(\alpha_1+\beta_1)(\alpha_2+\beta_2)}{\sigma^2}\,\|z_1 - z_2\|_z^2\,.
\end{aligned}$$

Thus, we get that $g(z)$ is Lipschitz continuous with

$$L = \sqrt{\frac{2(\alpha_1+\beta_1)(\alpha_2+\beta_2)}{\sigma^2}},$$

which is (51). □

In accordance with [5] for solving (48), we can use the following method:

1. Initialization: Fix $\beta = L$. Set $s_{-1} = 0$.
2. Iteration ($k \ge 0$):

$$\begin{aligned}
&\text{Compute } x_k = T_\beta(\bar{z}, s_{k-1}), \\
&\text{Compute } z_k = T_\beta(x_k, -g(x_k)), \\
&\text{Set } s_k = s_{k-1} - g(z_k).
\end{aligned} \tag{M2}$$

Here

$$T_\beta(z,s) := \arg\max_{x\in W}\{\langle s, x - z\rangle - \beta\omega(z,x)\}.$$

Similarly to [1], we can prove that the method (M2) generates a bounded sequence $\{z_i\}_{i\ge 0}$. Hence, the sequences $\{\lambda_i\}_{i\ge 0}, \{\mu_i\}_{i\ge 0}$ are also bounded. Also, since the saddle point in the problem (4) exists, there exists a saddle point (λ^*, μ^*) for the conjugate problem (37). These arguments allow us to choose D_λ, D_μ such that $d_\lambda(\lambda_i) \le D_\lambda$, $d_\mu(\mu_i) \le D_\mu$ for all $i \ge 0$, which also ensure that (λ^*, μ^*) is an interior solution:

$$\begin{aligned}
\mathfrak{B}^\lambda_{r/\sqrt{\kappa\sigma_\lambda}}(\lambda^*) \subseteq W_\lambda := \{\lambda : d_\lambda(\lambda) \le D_\lambda\}\,, \\
\mathfrak{B}^\mu_{r/\sqrt{(1-\kappa)\sigma_\mu}}(\mu^*) \subseteq W_\mu := \left\{\mu : d_\mu(\mu) \le D_\mu\right\}
\end{aligned}$$

for some $r > 0$. Then we have $z^* := (\lambda^*, \mu^*) \in \mathcal{F}_D := \{z \in W : d(z) \leq D\}$ with $D := \kappa D_\lambda + (1-\kappa) D_\mu$ and $\mathfrak{B}_r^z(z^*) \subseteq \mathcal{F}_D$.

From the Theorem 1 in [5], using the relation $\sigma_0 = 1$, similarly to the proof of the Theorem 2 in [5], we get the following lemma.

Lemma 4.5 *Assume that the operator $g(z)$ is Lipschitz continuous on W with the constant L. Let the sequence $\{z_i\}_{i \geq 0}$ be generated by the method* (M2). *Then, for any $k \geq 0$, we have*

$$\delta_k(D) \leq LD, \tag{52}$$

where $\delta_k(D)$ is defined in (20).

Theorem 4.1 *Let the Assumptions* **A3** *and* **A4** *be true, $\kappa = \frac{\sigma_\mu}{\sigma_\mu + \sigma_\lambda}$, and L be defined in* (51). *Let the points $z_i = (\lambda_i, \mu_i), i \geq 0$ be generated by the method* (M2). *Let the points in* (22) *be defined by points (u_i, v_i), (x_i, y_i) which are the saddle points at the points (λ_i, μ_i) in* (38) *and* (39), *respectively. Then, for functions $\xi(u, x)$, $\eta(v, y)$ defined in* (26) *and* (27), *we have:*

$$\xi(\hat{u}_{k+1}, \hat{x}_{k+1}) - \eta(\hat{v}_{k+1}, \hat{y}_{k+1}) \leq \frac{LD}{k+1}. \tag{53}$$

Also the following is true:

$$\left\| \mathcal{B}\hat{u}_{k+1} + \tilde{x}_0 - \hat{x}_{k+1} \right\|_{\mu,*} \leq \frac{LD\sqrt{\sigma_\mu}}{r(k+1)}, \quad \left\| \mathcal{C}\hat{v}_{k+1} + \tilde{y}_0 - \hat{y}_{k+1} \right\|_{\lambda,*} \leq \frac{LD\sqrt{\sigma_\lambda}}{r(k+1)}. \tag{54}$$

Proof Similarly to the proof of the Theorem 3.1, we conclude that

$$\begin{aligned} \xi(\hat{u}_{k+1}, \hat{x}_{k+1}) - \eta(\hat{v}_{k+1}, \hat{y}_{k+1}) &\leq \frac{\delta_k(D)}{k+1}, \\ \left\| \mathcal{B}\hat{u}_{k+1} + \tilde{x}_0 - \hat{x}_{k+1} \right\|_{\mu,*} &\leq \frac{\delta_k(D)\sqrt{\sigma_\mu}}{r(k+1)}, \\ \left\| \mathcal{C}\hat{v}_{k+1} + \tilde{y}_0 - \hat{y}_{k+1} \right\|_{\lambda,*} &\leq \frac{\delta_k(D)\sqrt{\sigma_\lambda}}{r(k+1)}. \end{aligned}$$

Lemma 4.5 and these inequalities prove the statement of the theorem. □

Let us make some notes on how to choose parameters D and r of the method. Recall the example introduced in the Sect. 3.1 with $y_0 = (2, 0)^T$. We see that the assumptions **A3** and **A4** hold.

We choose Euclidian norms in the spaces $\mathbb{R}^n$, $\mathbb{R}^m$. As we already know from the Sect. 3.3 $\|\mathcal{B}\|_2 = \|\mathcal{C}\|_2 = 1.12$. Also we know that $L_{uv} = L_{vu} = 0$, $L_{xx} = L_{xy} = L_{yx} = L_{yy} = 1$, $\sigma_{F_u} = \sigma_{F_v} = \sigma_{\Phi_x} = \sigma_{\Phi_y} = 1$. Hence, $L = 6.5$.

Next, we need to find D_λ, D_μ, and D. From the Lemma 4.3 since $\|\mathcal{B}\|_2 = \|C\|_2 = 1.12$, $P = Q = [-1, 1]$, we can take $r_1 = r_2 = 1$, $u_0 = 0$, $v_0 = 0$. Hence, $\|\lambda^*\|_2 \leq 8.4$, $\|\mu^*\|_2 \leq 4.5$.

Using the method above, we can find nearly feasible nearly optimal solution. We choose prox functions $d_\lambda(\lambda) = \frac{\|\lambda\|_2^2}{2}, d_\mu(\mu) = \frac{\|\mu\|_2^2}{2}$, $\kappa = 1/2$. Then, we have $d(z) = \frac{\|\lambda\|_2^2}{4} + \frac{\|\mu\|_2^2}{4}$, $\|z^*\|_z \le 6.7$

Making a similar argument as in the proof of the Theorem 2 in [1], we get that for the sequence $\{z_k\}$ generated by the method (M2), it holds that

$$\|z_k - z^*\|_z^2 \le 2d(z^*) = \|z^*\|_z.$$

Then,

$$d(z_k) = \frac{\|z_k\|_z^2}{2} \le \frac{1}{2}(\|z^*\|_z + \|z^*\|_z)^2 \le 90.$$

So $\|\lambda_k\|_2^2/2 \le 180$ and we can take $D_\lambda = 180$. In the same way, we get that $D_\mu = 180$ and also that $D = 180$. Finally, we have that $\mathfrak{B}^z_{7.4}(z^*) \subseteq \mathcal{F}_D$ and $r = 7.4$.

5 Duality for Finite-dimensional Saddle-point Problems

In this section, our goal is to solve the following saddle-point problem:

$$\begin{aligned}\min_{x\in X}&\left[\max_{y\in Y}\{\Phi(x,y) : Gy = g\} : Hx = h\right]\\ &= \max_{y\in Y}\left[\min_{x\in X}\{\Phi(x,y) : Hx = h\} : Gy = g\right],\end{aligned} \tag{55}$$

where $G \in \mathbb{R}^{k\times m}$, $g \in \mathbb{R}^k$, $H \in \mathbb{R}^{l\times n}$, $h \in \mathbb{R}^l$, $X \subset \mathbb{R}^n$, $Y \subset \mathbb{R}^m$ are closed, convex, and bounded sets, and the function $\Phi(\cdot, y)$ is convex for any fixed y, and the function $\Phi(x, \cdot)$ is concave for any fixed x. In this problem, we can pass to a dual formulation by introducing new variables as the Lagrange multipliers for the constraints.

Similarly to the proof of the Lemma 3.1, we get that the conjugate problem for (55) is given by the following

Lemma 5.1 *The problem* (55) *is equivalent to the following one:*

$$\min_{\lambda}\max_{\mu}\left\{\min_{x\in X}\max_{y\in Y}[\Phi(x,y) - \langle\mu, Hx\rangle + \langle\lambda, Gy\rangle] + \langle\mu, h\rangle - \langle\lambda, g\rangle\right\}, \tag{56}$$

which is called the conjugate problem to the problem (55)*. Here* $\lambda \in \mathbb{R}^k$, $\mu \in \mathbb{R}^l$.

Note that the function in the inner problem in the relation (56)

$$\psi(\lambda, \mu) := \min_{x\in X}\max_{y\in Y}[\Phi(x,y) - \langle\mu, Hx\rangle + \langle\lambda, Gy\rangle] \tag{57}$$

is well defined for every λ, μ since its objective function is convex–concave and the sets X, Y are convex and compact.

In the same way as it was done in the Sect. 3, we get the following

Lemma 5.2 *Let* (x^*, y^*) *be a saddle point in the problem* (57) *for some fixed* $\lambda \in \mathbb{R}^k$ *and* $\mu \in \mathbb{R}^l$. *Then, the function* $\psi(\cdot, \mu)$ *is convex for any fixed* $\mu \in \mathbb{R}^l$, *and its subgradient* $\nabla_\lambda \psi(\lambda, \mu) = Gy^*$ *is a bounded function of* (λ, μ). *Similarly, the function* $\psi(\lambda, \cdot)$ *is concave for any fixed* $\lambda \in \mathbb{R}^k$, *and its supergradient* $\nabla_\mu \psi(\lambda, \mu) = -Hx^*$ *is a bounded function of* (λ, μ).

Thus, we started from the saddle-point problem (55) and obtained the equivalent problem (56), which also has a saddle-point structure, and its objective function is also convex–concave. Moreover, we know for this function the partial sub- and supergradients. Hence, if we can solve the problem (57) efficiently (or in a closed form) for every λ, μ, then we can apply a standard method (e.g., [1] as it was done above) for solving the conjugate problem. If the method [1] is used, the primal approximate solution can be reconstructed by averaging saddle points in the problem (57) obtained on each iteration of the method. This approach can be efficient if the sets X and Y are simple and the dimensions of linear constraints are smaller than the dimensions of x and y.

6 Conclusions

In this paper, we have shown that infinite-dimensional differential games can be transformed into finite-dimensional dual form, which is just a usual convex–concave saddle-point problem. The latter problem can be solved by a standard finite-dimensional scheme, which allows reconstruction of the infinite-dimensional primal solution. This approach is feasible when the objective function is simple enough and allows to compute its conjugate function in a closed form. Such situation is quite common since in many applications the main source of complexity is the presence of the linear constraints for the solutions of ordinary differentiable equation (ODE). In this paper, we did not discuss the inaccuracy issues related to numerical errors in the solutions of ODE and discretization of control. However, this can be easily done for a particular ODE solver applying the standard analysis of the minimization schemes with inexact gradient.

This paper continues the line of research, started in [2], which exploits a significant asymmetry in the dimension of primal and dual variables. The authors are going to study applicability of this technique to other infinite-dimensional problems. However, the main challenge for the future research remains the presence of constraints on the phase variables.

Acknowledgments The research presented in this paper was partially supported by the Laboratory of Structural Methods of Data Analysis in Predictive Modeling, MIPT, through the RF government Grant, ag.11.G34.31.0073 and by RFBR, Research Project No. 13-01-12007 ofi_m.

References

1. Nesterov, Y.: Primal-dual subgradient methods for convex problems. Math. Program. **120**(1), 221–259 (2009)

2. Devolder, O., Glineur, F., Nesterov, Y.: Double smoothing technique for large-scale linearly constrained convex optimization. SIAM J. Optim. **22**(2), 702–727 (2012)
3. Antsaklis, P.J., Michel, A.N.: Linear Systems. Birkhauser Book, Boston (2006)
4. Maurice, S.: On general minimax theorems. Pac. J. Math. **8**(1), 171–176 (1958)
5. Nesterov, Y.: Dual extrapolation and its applications for solving variational inequalities and related problems. J. Math. Program. Ser. B **109**(2), 319–344 (2007)