Primal-dual methods for solving infinite-dimensional games

P. Dvurechensky, Yu. Nesterov, and V. Spokoiny*

June 2014

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.

1 Introduction

In the last years we can observe an increasing interest to the primal-dual subgradient methods. This line of research, started in [4], leads to the special methods, which allow 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 level 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 dimensional. 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-dimensiomal space of phase variables. It was shown, that an appropriate optimization process in the latter space can generate also

*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 projects No. 13-01-12007 ofi_m and 14-01-00722-a

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

nearly optimal sequence of controls (functions of time). Moreover, this technique was supported by the worst-case complexity analysis.

In this paper we do the next step in this direction. We consider an infinite-dimensional saddle-point problem, which variables (controls) must satisfy sone 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 Section 2 we show how to form the conjugate problem for 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. In Section 3 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 linear operators from infinite to finite-dimensional space. For the future applications, we derive some bounds for their norms. In Section 4 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 schemes and derive the upper bounds on the quality of primal and conjugate solutions. In the last Section 5 we consider differential game with objective function satisfying strong convexity assumption. For this case, we obtain better complexity bounds.

2 Duality for saddle-point problems

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

$$\min_{x\in X}\left[\max_{y\in Y}\left\{\Phi(x,y) : Gy = g\right\} : Hx = h\right] = \max_{y\in Y}\left[\min_{x\in X}\left\{\Phi(x,y) : Hx = h\right\} : Gy = g\right], \qquad (2.1)$$

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, Y are closed convex bounded sets, and function $\Phi(\cdot, y)$ is convex for any fixed y, 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 Lagrange multipliers for the constraints.

Lemma 2.1. *Problem* (2.1) *is equivalent to the the following one:*

$$\min_{\lambda}\max_{\mu}\left\{\min_{x\in X}\max_{y\in Y}\left[\Phi(x,y) - \langle\mu, Hx\rangle + \langle\lambda, Gy\rangle\right] + \langle\mu, h\rangle - \langle\lambda, g\rangle\right\}, \qquad (2.2)$$

which we call conjugate problem to (2.1)*. Here* $\lambda \in \mathbb{R}^k, \mu \in \mathbb{R}^l$*.*

Proof. Let us consider the inner problem in the right-hand side of (2.1). For each $y \in Y$ it is a problem of minimization of convex function over convex set with linear constraint. Hence it is equivalent to

$$f(y) \stackrel{\text{def}}{=} \min_{x\in X} \max_{\mu} \{\Phi(x,y) + \langle \mu, h - Hx\rangle\} \tag{2.3}$$

Due to convexity in x, concavity in μ, and the fact that X is convex and compact, we can swap min and max:

$$f(y) = \max_{\mu} \min_{x\in X} \{\Phi(x,y) + \langle \mu, h - Hx\rangle\}$$

Note that $f(y)$ in (2.3) is a concave function of y. So the outer problem in the right hand side of (2.1) is a problem of maximization of concave function over convex set with linear constraint. Hence it is equivalent to $\max\limits_{y\in Y} \min\limits_{\lambda} \{f(y) + \langle \lambda, Gy - g\rangle\}$. Using the same reasoning as above, we conclude that $\max\limits_{y\in Y} \min\limits_{\lambda} \{f(y) + \langle \lambda, Gy - g\rangle\} = \min\limits_{\lambda} \max\limits_{y\in Y} \{f(y) + \langle \lambda, Gy - g\rangle\}$. Hence we have

$$(2.1) = \min_{\lambda} \max_{y\in Y} \max_{\mu} \min_{x\in X} \{\Phi(x,y) + \langle \mu, h - Hx\rangle + \langle \lambda, Gy - g\rangle\}$$

Swapping two operations of maximization we get

$$(2.1) = \min_{\lambda} \max_{\mu} \max_{y\in Y} \min_{x\in X} \{\Phi(x,y) + \langle \mu, h - Hx\rangle + \langle \lambda, Gy - g\rangle\}$$

Since $\Phi(x,y) + \langle \mu, h - Hx\rangle + \langle \lambda, Gy - g\rangle$ is convex in x and concave in y and since X and Y are convex compact sets, we can swap $\max\limits_{y\in Y}$ and $\min\limits_{x\in X}$ and obtain (2.1)=(2.2). □

Note that the saddle function in the inner problem in (2.2)

$$\psi(\lambda,\mu) \stackrel{\text{def}}{=} \min_{x\in X} \max_{y\in Y} [\Phi(x,y) - \langle \mu, Hx\rangle + \langle \lambda, Gy\rangle] \tag{2.4}$$

is well defined for every λ, μ since its objective function is convex-concave and the sets X, Y are convex and compact. Let us study the properties of function $\psi(\lambda,\mu)$.

Lemma 2.2. *Let (x^*, y^*) be a saddle-point of problem (2.4) with some $\lambda \in \mathbb{R}^k$ and $\mu \in \mathbb{R}^l$ being fixed. Then 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, 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).*

Proof. Since the saddle point in problem (2.4) do exist for any λ, μ, we have

$$\psi(\lambda,\mu) = \min_{x\in X} \left[\tilde{\psi}(x,\lambda) - \langle \mu, Hx\rangle\right], \tag{2.5}$$

$$\psi(\lambda,\mu) = \max_{y\in Y} \left[\hat{\psi}(y,\mu) + \langle \lambda, Gy\rangle\right], \tag{2.6}$$

where

$$\tilde{\psi}(x,\lambda) = \max_{y\in Y} [\Phi(x,y) + \langle \lambda, Gy\rangle], \tag{2.7}$$

$$\hat{\psi}(y,\mu) = \min_{x\in X} [\Phi(x,y) - \langle \mu, Hx\rangle]. \tag{2.8}$$

In the first case, since we take the minimum of linear functions in μ, the result is concave in μ. So $\psi(\lambda,\mu)$ is concave in μ for any fixed λ. Similarly, from the second case we get that $\psi(\lambda,\mu)$ is convex in λ for any fixed μ.

Let us fix some λ and μ_0. Denote by (x_0^*, y_0^*) the saddle point of problem (2.4) for these values, and by (x^*, y^*) the saddle-point of this problem with λ and arbitrary μ. Then

$$\min_{x\in X}\left[\tilde{\psi}(x,\lambda)-\langle\mu,Hx\rangle\right]=\tilde{\psi}(x^*,\lambda)-\langle\mu,Hx^*\rangle,\ \min_{x\in X}\left[\tilde{\psi}(x,\lambda)-\langle\mu_0,Hx\rangle\right]=\tilde{\psi}(x_0^*,\lambda)-\langle\mu_0,Hx_0^*\rangle$$

We have the following:

$$\begin{aligned}
&\psi(\lambda,\mu)-\psi(\lambda,\mu_0)+\langle Hx_0^*,\mu-\mu_0\rangle=\\
&\tilde{\psi}(x^*,\lambda)-\langle Hx^*,\mu\rangle-\tilde{\psi}(x_0^*,\lambda)+\langle Hx_0^*,\mu_0\rangle+\langle Hx_0^*,\mu-\mu_0\rangle=\\
&\tilde{\psi}(x^*,\lambda)-\langle Hx^*,\mu\rangle-(\tilde{\psi}(x_0^*,\lambda)-\langle Hx_0^*,\mu\rangle)=\\
&\min_{x\in X}\{\tilde{\psi}(x,\lambda)-\langle Hx,\mu\rangle\}-(\tilde{\psi}(x_0^*,\lambda)-\langle Hx_0^*,\mu\rangle)\leq 0.
\end{aligned}$$

So, by definition, vector $-Hx^*$ is a supergradient of $\psi(\lambda,\mu)$ with respect to μ. In the same way we prove that Gy^* is subgradient of $\psi(\lambda,\mu)$ with respect to λ. The sets X and Y are bounded, the supergradients and subgradients are bounded too. □

Thus, we started from the saddle-point problem (2.1) and got the equivalent problem (2.2), which also has 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 (2.4) efficiently (or in a closed form) for every λ,μ, then we can apply a standard method (e.g. [4]) for solving the conjugate problem. 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. Below we will apply this dualisation approach for more difficult infinite-dimensional problems from differential games theory.

3 Differential games

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

$$\dot{x}(t)=A_x(t)x(t)+B(t)u(t),\quad \dot{y}(t)=A_y(t)y(t)+C(t)v(t),\quad (x(0),y(0))=(x_0,y_0). \tag{3.1}$$

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 $A_x(t), A_y(t), B(t), andC(t)$ are continuous and have appropriate sizes. The system is considered on the time interval $[0,\theta]$. Controls are restricted in the following way $u(t)\in P\subset\mathbb{R}^p$, $v(t)\in Q\subset\mathbb{R}^q\quad\forall t\in[0,\theta]$. We assume that P,Q are closed convex sets.

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

$$F(u,v)+\Phi(x(\theta),y(\theta))\stackrel{\text{def}}{=}\int_0^\theta \tilde{F}(\tau,u(\tau),v(\tau))d\tau+\Phi(x(\theta),y(\theta)). \tag{3.2}$$

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 $v(\cdot) \in L_2([0,\theta],\mathbb{R}^q)$,
- the saddle-point in this class of strategies exists,
- 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 (3.1). It is the unique solution of the following matix Cauchy problem

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

Here E is 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 (3.1), then we can express $x(\theta)$ as a result of application of the linear operator $\mathcal{B} : L_2([0,\theta],\mathbb{R}^p) \to \mathbb{R}^n$:

$$x(\theta) = V_x(\theta,0)x_0 + \int_0^\theta V_x(\theta,\tau)B(\tau)u(\tau)d\tau \overset{\text{def}}{=} \tilde{x}_0 + \mathcal{B}u \tag{3.3}$$

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

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

Note that 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 found 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 (3.1), the operator $\mathcal{C} : L_2([0,\theta],\mathbb{R}^q) \to \mathbb{R}^m$ defined by the formula $\mathcal{C}v = \int_0^\theta V_y(\theta,\tau)C(\tau)v(\tau)d\tau$, and the vector $\tilde{y}_0 = V_y(\theta,0)y_0$. Adjoint operator $\mathcal{C}^*$ also can be computed using the solution of some ODE.

So our main problem of interest has the following form:

$$\min_{u\in P}\left[\max_{v\in Q}\{F(u,v) + \Phi(x,y) : y = \tilde{y}_0 + \mathcal{C}v\} : x = \tilde{x}_0 + \mathcal{B}u\right] \tag{3.4}$$

The goal of this paper is to introduce a computational method for finding its approximate solution.

Remark 3.1. In the same way, we can treat the problems 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.

3.1 Estimating the norms of operators $\mathcal{B}$ and $\mathcal{C}$

Let us consider the problem of estimating the norms of operators $\mathcal{B}, \mathcal{C}$. This is an important problem since we need these operators to be bounded, and their norms play a significant role in the estimates for the rate of convergence of the methods below. We will work with the operator $\mathcal{B}$ (3.3), since the norm of $\mathcal{C}$ can be estimated in a similar way. The following argument was used in [2]; it is presented here for the reader convenience.

By definition we have

$$\|\mathcal{B}\| = \sup_{u \in L^2([0,\theta],\mathbb{R}^p)} \{\|\mathcal{B}u\| : \quad \|u\|_{L^2([0,\theta],\mathbb{R}^p)} = 1\}. \tag{3.5}$$

As it was shown above the conjugate operator $\mathcal{B}^*$ transforms $\mu \in \mathbb{R}^n$ into the function $B^T(\tau)V_x^T(\theta,\tau)\mu \in L^2([0,\theta],\mathbb{R}^p)$. Let us define a matrix

$$R = \int_0^\theta V_x(\theta,\tau)B(\tau)B^T(\tau)V_x^T(\theta,\tau)d\tau = \mathcal{B}\mathcal{B}^*, \tag{3.6}$$

which is symmetric and positive semidefinite.

Definition 3.1. The system with dynamics given by the first differential equation in (3.1) and initial value $x(0) = 0$ is called reachable on $[0,\theta]$ if for any $\hat{x} \in \mathbb{R}^n$ there exists a control such that $x(\theta) = \hat{x}$.

The teachability is closely related to the properties of matrix R (see Corollary 2.3 in [1]).

Theorem 3.1. *The system with dynamics given by the first differential equation in* (3.1) *and initial value* $x(0) = 0$ *is reachable on* $[0;\theta]$ *if and only if* R *is positive definite.*

We also need the following

Lemma 3.1. *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 equation (3.5),

$$\|\mathcal{B}\| = \left[\inf_{u \in L^2([0,\theta],\mathbb{R}^p)} \{\|u\|_{L^2([0,\theta],\mathbb{R}^p)} : \quad \|\mathcal{B}u\| = 1\}\right]^{-1}$$

Using the teachability property, we have $\text{Im}\mathcal{B}(L^2([0,\theta],\mathbb{R}^p)) = \mathbb{R}^n$ and

$$\inf_{u \in L^2([0,\theta],\mathbb{R}^p)} \{\|u\|_{L^2([0,\theta],\mathbb{R}^p)} : \quad \|\mathcal{B}u\| = 1\} = \inf_{u \in L^2([0,\theta],\mathbb{R}^p), x \in \mathbb{R}^n, \|x\|=1} \{\|u\|_{L^2([0,\theta],\mathbb{R}^p)} : \quad \mathcal{B}u = x\}$$

From Theorem 3.1, we have that

$$\inf_{u \in L^2([0,\theta],\mathbb{R}^p)} \{\|u\|_{L^2([0,\theta],\mathbb{R}^p)} : \quad \mathcal{B}u = x\} = \|\mathcal{B}^*(\mathcal{B}\mathcal{B}^*)^{-1}x\| = \langle(\mathcal{B}\mathcal{B}^*)^{-1}x, x\rangle^{1/2}.$$

Hence

$$\inf_{u\in L^2([0,\theta],\mathbb{R}^p),}\{\|u\|_{L^2([0,\theta],\mathbb{R}^p)}:\quad \|\mathcal{B}u\|=1\}=\inf_{\|x\|=1}\langle(\mathcal{B}\mathcal{B}^*)^{-1}x,x\rangle^{1/2}=\lambda_{\min}^{1/2}((\mathcal{B}\mathcal{B}^*)^{-1}).$$

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

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

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

where $A_x(t)$ is continuous with time.

Recall the following well known result.

Theorem 3.2. *[1] Assume that the matrix $A_x(t)$ is time-independent, and that there exists a matrix $M=M^T\succ 0$ such that $A_x^TM+MA_x\prec 0$. Then the equilibrium $x=0$ is globally exponentially stable.*

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

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

Hence

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

and

$$\tfrac{d}{dt}\langle Mx(t),x(t)\rangle=2\langle Mx(t),\dot{x}(t)\rangle=2\langle Mx(t),Ax(t)+Bu(t)\rangle$$

$$\langle(A^TM+MA)x(t),\dot{x}(t)\rangle+2\langle Mx(t),Bu(t)\rangle$$
$$\leq-\nu\langle Mx(t),x(t)\rangle+2\langle Mx(t),Bu(t)\rangle\leq\tfrac{1}{\nu}\langle MBu(t),Bu(t)\rangle.$$

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

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

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

So we assume that both operators $\mathcal{B}$ and $\mathcal{C}$ has bounded norms.

4 Convex-concave problem

In this section we consider the problem (3.4) satisfying two assumptions.

A1 The sets P and Q are bounded.

A2 In (3.2) 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 operators $\mathcal{B}, \mathcal{C}$ are bounded, $x(\theta), y(\theta)$ are also bounded and we can equivalently reformulate the problem (3.4) in the following way:

$$\min_{u\in P, x\in X}\left[\max_{v\in Q, y\in Y}\{F(u,v)+\Phi(x,y): y=\tilde{y}_0+\mathcal{C}v\}: x=\tilde{x}_0+\mathcal{B}u\right]=$$
$$\max_{v\in Q, y\in Y}\left[\min_{u\in P, x\in X}\{F(u,v)+\Phi(x,y): x=\tilde{x}_0+\mathcal{B}u\}: y=\tilde{y}_0+\mathcal{C}v\right], \tag{4.1}$$

where the sets X and Y are closed convex and bounded.

Lemma 4.1. *Let Assumptions* **A1**, **A2** *be true. Then the problem* (4.1) *is equivalent to the problem*

$$\begin{aligned}\min_\lambda \max_\mu\{ \quad & \min_{u\in P}\max_{v\in Q}[F(u,v)-\langle\mu,\mathcal{B}u\rangle+\langle\lambda,\mathcal{C}v\rangle] \\ & +\min_{x\in X}\max_{y\in Y}[\Phi(x,y)+\langle\mu,x\rangle-\langle\lambda,y\rangle]-\langle\mu,\tilde{x}_0\rangle+\langle\lambda,\tilde{y}_0\rangle\},\end{aligned} \tag{4.2}$$

which we call the conjugate problem to (4.1)*. Here* $\lambda\in\mathbb{R}^n$ *and* $\mu\in\mathbb{R}^m$*.*

Proof. Let us consider the inner problem in (4.1). Due to **A2**, for each $v\in Q$ and $y\in Y$ this is a problem of minimization of convex function over convex set with linear constraints. Hence, it is equivalent to

$$\chi(v,y)\overset{\text{def}}{=}\min_{u\in P, x\in X}\max_{\mu}\{F(u,v)+\Phi(x,y)+\langle\mu,x-\tilde{x}_0-\mathcal{B}u\rangle\} \tag{4.3}$$

Due to assumptions **A1**, **A2**, using the fact that any closed convex bounded set in Hilbert space is compact in 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 [3] we can swap min and max:

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

Note that $\chi(v,y)$ in (4.3) is a concave function of v an y. So the outer problem in (4.1) is a problem of maximization of concave function over a convex set with linear constraints. Hence it is equivalent to: $\max_{v\in Q,y\in Y}\min_{\lambda}\{\chi(v,y)+\langle\lambda,\mathcal{C}v+\tilde{y}_0-y\rangle\}$. Using the same arguments as above, we conclude that $\max_{v\in Q,y\in Y}\min_{\lambda}\{\chi(v,y)+\langle\lambda,\mathcal{C}v+\tilde{y}_0-y\rangle\}=\min_{\lambda}\max_{v\in Q,y\in Y}\{\chi(v,y)+\langle\lambda,\mathcal{C}v+\tilde{y}_0-y\rangle\}$. Hence, we have

$$(4.1)=\min_{\lambda}\max_{v\in Q,y\in Y}\max_{\mu}\min_{u\in P,x\in X}\{F(u,v)+\Phi(x,y)+\langle\lambda,\mathcal{C}v+\tilde{y}_0-y\rangle+\langle\mu,x-\tilde{x}_0-\mathcal{B}u\rangle\}$$

Swapping two operations of maximization, we get.

$$(4.1) = \min_{\lambda} \max_{\mu} \max_{v \in Q, y \in Y} \min_{u \in P, x \in X} \{F(u,v) + \Phi(x,y) + \langle \lambda, \mathcal{C}v + \tilde{y}_0 - y \rangle + \langle \mu, x - \tilde{x}_0 - \mathcal{B}u \rangle\}$$

Since function $F(u,v) + \Phi(x,y) + \langle \lambda, \mathcal{C}v + \tilde{y}_0 - y \rangle + \langle \mu, x - \tilde{x}_0 - \mathcal{B}u \rangle$ is convex in u, x and concave in v, y, and since P, Q, X, and Y are convex compacts, we can swap $\max\limits_{v \in Q, y \in Y}$ and $\min\limits_{u \in P, x \in X}$, and obtain (4.1)=(4.2). □

We assume that problems

$$\psi_1(\lambda, \mu) = \min_{u \in P} \max_{v \in Q} \left[F(u,v) - \langle \mu, \mathcal{B}u \rangle + \langle \lambda, \mathcal{C}v \rangle\right], \tag{4.4}$$

$$\psi_2(\lambda, \mu) = \min_{x \in X} \max_{y \in Y} \left[\Phi(x,y) + \langle \mu, x \rangle - \langle \lambda, y \rangle\right] \tag{4.5}$$

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 the closed convex bounded set in Hilbert space is compact in weak topology, and since $F(u,v)$ is upper semi-continuous in v and lower semi-continuous in u, by Corollary 3.3 in [3], we conclude that the saddle point in the problems (4.4), (4.5) do exist for all $\lambda \in \mathbb{R}^n, \mu \in \mathbb{R}^m$.

Note that the problem (4.4) has the following form

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

$$= \int_0^\theta \left\{\min_{u \in P} \max_{v \in Q} \left[\tilde{F}(\tau, u(\tau), v(\tau)) - \langle \mathcal{B}^* \mu(\tau), u(\tau) \rangle + \langle \mathcal{C}^* \lambda(\tau), v(\tau) \rangle\right] d\tau\right\}, \tag{4.6}$$

and it can be solved pointwise.

Lemma 4.2. *Let the assumptions* **A1** *and* **A2** *be true, and* (u^*, v^*) *be a saddle point of the problem* (4.4) *for some fixed* $\lambda \in \mathbb{R}^n, \mu \in \mathbb{R}^m$*. Then the function* $\psi_1(\cdot, \mu)$ *is convex for any fixed* $\mu \in \mathbb{R}^m$*, and its subgradient* $\nabla_\lambda \psi_1(\lambda, \mu) = \mathcal{C}v^*$ *is bounded in* (λ, μ)*. Similarly, function* $\psi_1(\lambda, \cdot)$ *is concave for any fixed* $\lambda \in \mathbb{R}^n$*, and its supergradient* $\nabla_\mu \psi_1(\lambda, \mu) = -\mathcal{B}u^*$ *is bounded in* (λ, μ)*.*

Proof. Since the saddle point in problem (4.4) exists for any λ, μ, we have

$$\psi_1(\lambda, \mu) = \min_{u \in P} \left[\tilde{\psi}_1(u, \lambda) - \langle \mathcal{B}u, \mu \rangle\right], \tag{4.7}$$

$$\psi_1(\lambda, \mu) = \max_{v \in Q} \left[\hat{\psi}_1(v, \mu) + \langle \mathcal{C}v, \lambda \rangle\right], \tag{4.8}$$

where

$$\tilde{\psi}_1(u, \lambda) = \max_{v \in Q} \left[F(u,v) + \langle \mathcal{C}v, \lambda \rangle\right], \tag{4.9}$$

$$\hat{\psi}_1(v, \mu) = \min_{u \in P} \left[F(u,v) - \langle \mathcal{B}u, \mu \rangle\right]. \tag{4.10}$$

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 (4.4) for λ, μ_0, and by (u^*, v^*) the saddle point of this problem for λ, μ, where μ is arbitrary. Then

$$\min_{u\in P}\left[\tilde{\psi}_1(u,\lambda) - \langle \mathcal{B}u, \mu\rangle\right] = \tilde{\psi}_1(u^*,\lambda) - \langle \mathcal{B}u^*, \mu\rangle, \quad \min_{u\in P}\left[\tilde{\psi}_1(u,\lambda) - \langle \mathcal{B}u, \mu_0\rangle\right] = \tilde{\psi}_1(u_0^*,\lambda) - \langle \mathcal{B}u_0^*, \mu_0\rangle$$

Note that

$$\begin{aligned}
&\psi_1(\lambda,\mu) - \psi_1(\lambda,\mu_0) + \langle \mathcal{B}u_0^*, \mu-\mu_0\rangle = \\
&\tilde{\psi}_1(u^*,\lambda) - \langle \mathcal{B}u^*, \mu\rangle - \tilde{\psi}_1(u_0^*,\lambda) + \langle \mathcal{B}u_0^*, \mu_0\rangle + \langle \mathcal{B}u_0^*, \mu-\mu_0\rangle = \\
&\tilde{\psi}_1(u^*,\lambda) - \langle \mathcal{B}u^*, \mu\rangle - (\tilde{\psi}_1(u_0^*,\lambda) - \langle \mathcal{B}u_0^*, \mu\rangle) = \\
&\min_{u\in P}\{\tilde{\psi}_1(u,\lambda) - \langle \mathcal{B}u, \mu\rangle\} - (\tilde{\psi}_1(u_0^*,\lambda) - \langle \mathcal{B}u_0^*, \mu\rangle) \le 0.
\end{aligned}$$

So, by definition, the vector $-\mathcal{B}u^*$ is supergradient of $\psi_1(\lambda,\mu)$ with respect to μ. In the same way we prove that $\mathcal{C}v^*$ is subgradient of $\psi_1(\lambda,\mu)$ with respect to λ. Since P and Q are bounded, we have $\|u(t)\| \le \sqrt{\theta}\max_{u\in P}\|u\| \stackrel{\text{def}}{=} \sqrt{\theta}\,\|P\|$. Similarly, $\|v(t)\| \le \sqrt{\theta}\,\|Q\|$. Then since the norms $\|\mathcal{B}\|$ and $\|\mathcal{C}\|$ are bounded we have $\|\nabla_\lambda\psi_1(\lambda,\mu)\| \le \sqrt{\theta}\,\|\mathcal{C}\|\,\|Q\|$, and $\|\nabla_\mu\psi_1(\lambda,\mu)\| \le \sqrt{\theta}\,\|\mathcal{B}\|\,\|P\|$. □

In a similar way we can prove the following statement.

Lemma 4.3. *Let the assumptions* **A1** *and* **A2** *be true, and* (x^*, y^*) *be a saddle point of the problem* (4.5) *for some given* $\lambda \in \mathbb{R}^n, \mu \in \mathbb{R}^m$. *Then the function* $\psi_2(\cdot,\mu)$ *is convex for any fixed* $\mu \in \mathbb{R}^m$, *and its subgradient* $\nabla_\lambda\psi_2(\lambda,\mu) = -y^*$ *is bounded in* (λ,μ). *Similarly, function* $\psi_2(\lambda,\cdot)$ *is concave for any fixed* $\lambda \in \mathbb{R}^n$, *and its supergradient* $\nabla_\mu\psi_2(\lambda,\mu) = x^*$ *which is bounded in* (λ,μ).

Combining Lemmas 4.2 and 4.3, we get that the function $\psi(\lambda,\mu) \stackrel{\text{def}}{=} \psi_1(\lambda,\mu) + \psi_2(\lambda,\mu) - \langle \mu, \tilde{x}_0\rangle + \langle \lambda, \tilde{y}_0\rangle$ is convex in λ and concave in μ with partial 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

$$\|\psi'_\lambda(\lambda,\mu)\| \le L_\lambda \stackrel{\text{def}}{=} \sqrt{\theta}\,\|\mathcal{C}\|\,\|Q\| + \|Y\| + \|\tilde{y}_0\|, \quad \left\|\psi'_\mu(\lambda,\mu)\right\| \le L_\mu \stackrel{\text{def}}{=} \sqrt{\theta}\,\|\mathcal{B}\|\,\|P\| + \|X\| + \|\tilde{x}_0\|. \tag{4.11}$$

Here for any set $S \subset \mathbb{R}^k$, we use notation $\|S\| = \max_{s\in S}\|s\|$.

4.1 Example of the problem (4.4)

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:

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

where $y_0 = (2,0)^T$, and the norm 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 (4.4) using (4.6):

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

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

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

4.2 Estimating the norms of λ^*, μ^*

Let us compute the estimates for the norms of the components λ and μ of the solution of conjugate problem.

Lemma 4.4. *Assume that* $\Delta_{ux} \stackrel{\text{def}}{=} \max\limits_{u,\tilde{u}\in P, x,\tilde{x}\in X, v\in Q, y\in Y} \{-F'(\tilde{u}, u-\tilde{u}|v) - \Phi'(\tilde{x}, x-\tilde{x}|y)\} < +\infty$, *and* $\Delta_{vy} \stackrel{\text{def}}{=} \max\limits_{v,\tilde{v}\in Q, y,\tilde{y}\in Y, u\in P, x\in X} \{F'(u|\tilde{v}, v-\tilde{v}) + \Phi'(x|\tilde{y}, y-\tilde{y})\} < +\infty$. *If*

$$\mathfrak{B}_r(0) \subset \{s = \mathcal{B}u + \tilde{x}_0 - x : u \in P, x \in X\}, \quad \mathfrak{B}_r(0) \subset \{s = \mathcal{C}v + \tilde{y}_0 - y : v \in Q, y \in Y\}, \quad (4.12)$$

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

Proof. Consider the function $\psi(\lambda^*,\mu)$. It is concave and it achieves its maximum at μ^*. We also know that

$$\partial_\mu \psi(\lambda^*,\mu) = \{x_\mu - \tilde{x}_0 - \mathcal{B}u_\mu : u_\mu \in P, x_\mu \in X, \tilde{\psi_1}(u_\mu,\lambda^*) + \tilde{\psi_2}(x_\mu,\lambda^*) + \langle x_\mu - \tilde{x}_0 - \mathcal{B}u_\mu, \mu\rangle = \psi(\lambda^*,\mu)\}$$

From (4.9), and a similar definition of $\tilde{\psi_2}(x_\mu,\lambda^*)$, we get that

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

From this equalities and the optimality conditions for the problem defining $\psi(\lambda^*,\mu)$, we get

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

or

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

Hence

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

Using the first inclusion in (4.12) and the fact that $0\in\partial_\mu\psi(\lambda^*,\mu^*)$, we get that

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

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

Let us consider an example of situation when the inclusions (4.12) hold. If the set X has nonempty interior and there exists some $\bar{u}\in P$ such that $\mathcal{B}\bar{u}+\tilde{x}_0=\bar{x}\in \mathrm{int}X$, then there exists some $r>0$ such that $\bar{x}+\mathfrak{B}_r(0)\subset X$. Then we have

$$\mathfrak{B}_r(0)=\mathcal{B}\bar{u}+\tilde{x}_0-\bar{x}+\mathfrak{B}_r(0)\subset\{s=\mathcal{B}u+\tilde{x}_0-x: u\in P, x\in X\}.$$

Similar arguments can be used for the second inclusion in (4.12).

Let us consider another example. Since the problem (3.4) does not have any constraints for x and y, and the sets X and Y can be introduced due to boundedness of the norms of 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 $P\subset\mathfrak{B}_{r_0}(\bar{u})$. Then for every $u\in P$ there exists some $\tilde{u}\in\mathfrak{B}_{r_0}(0):\quad u=\bar{u}+\tilde{u}$. Hence for every $u\in P$, $x_u(\theta)=\tilde{x}_0+\mathcal{B}\bar{u}+\mathcal{B}\tilde{u}$ and $\|x_u(\theta)-\tilde{x}_0-\mathcal{B}\bar{u}\|\le\|\mathcal{B}\tilde{u}\|\le\|\mathcal{B}\|r_0$. So if we choose $X=\mathfrak{B}_{\|\mathcal{B}\|r_0}(\tilde{x}_0+\mathcal{B}\bar{u})$ we will be in the situation of previous example and can take $r=\|\mathcal{B}\|r_0$ in the conditions of Lemma 4.4. For the second inclusion in (4.12) we can apply similar arguments.

4.3 Algorithm description

We assume that we are given some norm in the space $\mathbb{R}^n$ and prox-function $d_\lambda(\lambda)$ with prox-center λ_0, which is strongly convex with convexity parameter σ_λ in the given norm. For μ we introduce the similar assumptions.

Since (λ^*,μ^*) is the saddle point, by 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 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 variational inequality

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

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 [4] for finding an approximate solution of the finite-dimensional problem (4.2).

Let us choose some $\kappa \in (0, 1)$. As in Section 4 in [4], we consider a space of $z \stackrel{\text{def}}{=} (\lambda, \mu)$ with the norm

$$\|z\| = \sqrt{\kappa \sigma_\lambda \|\lambda\|^2 + (1 - \kappa) \sigma_\mu \|\mu\|^2}, \tag{4.13}$$

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 (4.13). We define $W = \mathbb{R}^n \times \mathbb{R}^m$.

The conjugate norm for (4.13) is

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

So we have a uniform upper bound for the answers of the oracle $\|g\|_*^2 \leq L^2 \stackrel{\text{def}}{=} \frac{L_\lambda^2}{\kappa \sigma_\lambda} + \frac{L_\mu^2}{(1-\kappa)\sigma_\mu}$, where L_λ and L_μ are defined in (4.11).

The SDA method for solving (4.2) is the following

1. Initialization: Set $s_0 = 0$. Choose z_0, $\gamma > 0$.

2. Iteration ($k \geq 0$):

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

In accordance to Lemma 3 in [4], for $k \geq 1$ it satisfies inequalities

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

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

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

Since the saddle point in the problem (4.1) do exist, there exists a saddle point (λ^*, μ^*) in the conjugate problem (4.2). According to Theorem 1 in [4] the method (M1) generates bounded sequence $\{z_i\}_{i \geq 0}$. Hence the sequences $\{\lambda_i\}_{i \geq 0}, \{\mu_i\}_{i \geq 0}$ are also bounded. So we can choose D_λ, D_μ such that $d_\lambda(\lambda_i) \leq D_\lambda$, $d_\mu(\mu_i) \leq D_\mu$ for all $i \geq 0$. Also, the pair (λ^*, μ^*) is an interior solution: $B_{r/\sqrt{\kappa \sigma_\lambda}}(\lambda^*) \subseteq W_\lambda \stackrel{\text{def}}{=} \{\lambda : d_\lambda(\lambda) \leq D_\lambda\}$, and $B_{r/\sqrt{(1-\kappa)\sigma_\mu}}(\mu^*) \subseteq W_\mu \stackrel{\text{def}}{=} \{\mu : d_\mu(\mu) \leq D_\mu\}$ for some $r > 0$. Then we have $z^* = (\lambda^*, \mu^*) \in \mathcal{F}_D \stackrel{\text{def}}{=} \{z \in W : d(z) \leq D\}$ with $D = \kappa D_\lambda + (1 - \kappa) D_\mu$ and $B_r(z^*) \subseteq \mathcal{F}_D$.

Let us introduce a gap function

$$\delta_k(D) = \max_z \left\{ \sum_{i=0}^{k} \langle g_i, z_i - z \rangle : z \in \mathcal{F}_D \right\}. \tag{4.14}$$

From Theorem 2 in [4] (equation (4.6)) we have

$$\frac{1}{k+1}\delta_k(D) \le \frac{\hat{\beta}_{k+1}}{k+1}\left(\gamma D + \frac{L^2}{2\sigma_0\gamma}\right). \tag{4.15}$$

Denote

$$\hat{u}_{k+1} = \frac{1}{k+1}\sum_{i=0}^{k} u_i, \quad \hat{v}_{k+1} = \frac{1}{k+1}\sum_{i=0}^{k} v_i, \quad \hat{x}_{k+1} = \frac{1}{k+1}\sum_{i=0}^{k} x_i, \quad \hat{y}_{k+1} = \frac{1}{k+1}\sum_{i=0}^{k} y_i, \tag{4.16}$$

where (u_i, v_i), (x_i, y_i) are the saddle points at the point (λ_i, μ_i) in (4.4) and (4.5) respectively. Note that for all $u \in P$, $v \in Q$, $x \in X$, and $y \in Y$ we have

$$\begin{aligned} &F(u, v_i) + \Phi(x, y_i) + \langle \mu_i, x - \tilde{x}_0 - \mathcal{B}u \rangle + \langle \lambda_i, \mathcal{C}v_i + \tilde{y}_0 - y_i \rangle \ge \psi(\lambda_i, \mu_i) \ge \\ &F(u_i, v) + \Phi(x_i, y) + \langle \mu_i, x_i - \tilde{x}_0 - \mathcal{B}u_i \rangle + \langle \lambda_i, \mathcal{C}v + \tilde{y}_0 - y \rangle. \end{aligned} \tag{4.17}$$

We define a function

$$\begin{aligned} \phi(u, x, v, y) = \min\nolimits_\lambda \max\nolimits_\mu \{ \quad & F(u, v) + \Phi(x, y) + \langle \mu, x - \tilde{x}_0 - \mathcal{B}u \rangle + \langle \lambda, \mathcal{C}v + \tilde{y}_0 - y \rangle : \\ & d_\lambda(\lambda) \le D_\lambda, d_\mu(\mu) \le D_\mu \}. \end{aligned} \tag{4.18}$$

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

$$\min_{u \in P, x \in X} \max_{v \in Q, y \in Y} \phi(u, x, v, y) \tag{4.19}$$

Let us introduce two auxiliary functions:

$$\xi(u, x) = \max_{v \in Q, y \in Y} \phi(u, x, v, y) \tag{4.20}$$

$$\eta(v, y) = \min_{u \in P, x \in X} \phi(u, x, v, y) \tag{4.21}$$

Note that $\xi(u, x)$ is convex, $\eta(v, y)$ is concave, and $\xi(u, x) \ge \phi(u^*, x^*, v^*, y^*) \ge \eta(v, y) \quad \forall u \in P, v \in Q, x \in X, y \in Y$, where $\phi(u^*, x^*, v^*, y^*)$ is the solution to (4.19).

Theorem 4.1. *Let the assumptions* **A1** *and* **A2** *be true. Then the points* (4.16) *generated by method* (M1) *satisfy:*

$$\xi(\hat{u}_{k+1}, \hat{x}_{k+1}) - \eta(\hat{v}_{k+1}, \hat{y}_{k+1}) \le \frac{\hat{\beta}_{k+1}}{k+1}\left(\gamma D + \frac{L^2}{2\gamma}\right) \tag{4.22}$$

$$\begin{aligned} \|\tilde{x}_0 + \mathcal{B}\hat{u}_{k+1} - \hat{x}_{k+1}\| &\le \tfrac{\hat{\beta}_{k+1}\sqrt{\sigma_\mu}}{r(k+1)}\left(\gamma D + \tfrac{L^2}{2\gamma}\right), \\ \|\tilde{y}_0 + \mathcal{C}\hat{v}_{k+1} - \hat{y}_{k+1}\| &\le \tfrac{\hat{\beta}_{k+1}\sqrt{\sigma_\lambda}}{r(k+1)}\left(\gamma D + \tfrac{L^2}{2\gamma}\right). \end{aligned} \tag{4.23}$$

Proof. From inequalities (4.17), by convexity of $F(\cdot, v)$, $\Phi(\cdot, y)$ we have

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

This gives us

$$\xi(\hat{u}_{k+1}, \hat{x}_{k+1}) \leq \frac{1}{k+1}\sum_{i=0}^{k}\psi(\lambda_i, \mu_i) + \max_{\mu}\left\{\frac{1}{k+1}\sum_{i=0}^{k}\langle \mu - \mu_i, x_i - \tilde{x}_0 - \mathcal{B}u_i\rangle : d_\mu(\mu) \leq D_\mu\right\} +$$
$$+ \max_{v\in Q, y\in Y}\min_{\lambda}\left\{\langle \lambda - \frac{1}{k+1}\sum_{i=0}^{k}\lambda_i, \mathcal{C}v + \tilde{y}_0 - y\rangle : d_\lambda(\lambda) \leq D_\lambda\right\}. \tag{4.24}$$

Since method (M1) generates points λ_i, which satisfy $d_\lambda(\lambda_i) \leq D_\lambda$, and there exist $v_1 \in Q$ 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 $\max_{v\in Q, y\in Y}\min_{\lambda}\left\{\langle \lambda - \frac{1}{k+1}\sum_{i=0}^{k}\lambda_i, \mathcal{C}v + \tilde{y}_0 - y\rangle : d_\lambda(\lambda) \leq D_\lambda\right\} = 0.$

Similarly, we have

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

Finally we have the following

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

Combining this with (4.15), we get (4.22).

Let us prove that (4.16) is also a nearly feasible solution. Obviously

$$\frac{1}{(1-\kappa)\sigma_\mu}\left\|\mathcal{B}\hat{u}_{k+1} + \tilde{x}_0 - \hat{x}_{k+1}\right\|^2 \leq \frac{1}{(1-\kappa)\sigma_\mu}\left\|\mathcal{B}\hat{u}_{k+1} + \tilde{x}_0 - \hat{x}_{k+1}\right\|^2 + \frac{1}{\kappa\sigma_\lambda}\left\|\mathcal{C}\hat{v}_{k+1} + \tilde{y}_0 - \hat{y}_{k+1}\right\|^2$$
$$\frac{1}{\kappa\sigma_\lambda}\left\|\mathcal{C}\hat{v}_{k+1} + \tilde{y}_0 - \hat{y}_{k+1}\right\|^2 \leq \frac{1}{(1-\kappa)\sigma_\mu}\left\|\mathcal{B}\hat{u}_{k+1} + \tilde{x}_0 - \hat{x}_{k+1}\right\|^2 + \frac{1}{\kappa\sigma_\lambda}\left\|\mathcal{C}\hat{v}_{k+1} + \tilde{y}_0 - \hat{y}_{k+1}\right\|^2$$

On the other hand, from the proof of the third item in Theorem 1 in [4] we have

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

This in combination with (4.15) gives us (4.23). $\square$

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

5 Strongly convex-concave problem

In this section we consider the problem (3.4) safisfying the following assumptions.

A3 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:

$$\|\nabla_u F(u, v_1) - \nabla_u F(u, v_2)\| \leq L_{uv} \|v_1 - v_2\| \tag{5.1}$$

$$\|\nabla_v F(u_1, v) - \nabla_v F(u_2, v)\| \leq L_{vu} \|u_1 - u_2\| \tag{5.2}$$

A4 $\Phi(\cdot, y)$ is strongly convex for any fixed y with constant $\sigma_{\Phi x}$ which doesn't depend on y and $\Phi(x, \cdot)$ is strongly concave for any fixed x with constant $\sigma_{\Phi y}$ which doesn't depend on x. Also assume that:

$$\|\nabla_x \Phi(x, y_1) - \nabla_x \Phi(x, y_2)\| \leq L_{xy} \|y_1 - y_2\|, \tag{5.3}$$

$$\|\nabla_y \Phi(x_1, y) - \nabla_y \Phi(x_2, y)\| \leq L_{yx} \|x_1 - x_2\|, \tag{5.4}$$

and

$$\|\nabla_x \Phi(x_1, y) - \nabla_x \Phi(x_2, y)\| \leq L_{xx} \|x_1 - x_2\|, \tag{5.5}$$

$$\|\nabla_y \Phi(x, y_1) - \nabla_y \Phi(x, y_2)\| \leq L_{yy} \|y_1 - y_2\|. \tag{5.6}$$

Note that assumptions **A3**, **A4** imply that the level sets of functions $F(u, v), \Phi(x, y)$ are closed convex and bounded. Similarly to the proof of Lemma 4.1, we get that the conjugate problem for (3.4) is

$$\begin{aligned} \min_\lambda \max_\mu \{ \quad & \min_{u \in P} \max_{v \in Q} [F(u, v) - \langle \mu, \mathcal{B}u \rangle + \langle \lambda, \mathcal{C}v \rangle] \\ & + \min_x \max_y [\Phi(x, y) + \langle \mu, x \rangle - \langle \lambda, y \rangle] - \langle \mu, \tilde{x}_0 \rangle + \langle \lambda, \tilde{y}_0 \rangle \ \}. \end{aligned} \tag{5.7}$$

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

We assume that problems

$$\psi_1(\lambda, \mu) = \min_{u \in P} \max_{v \in Q} [F(u, v) - \langle \mu, \mathcal{B}u \rangle + \langle \lambda, \mathcal{C}v \rangle], \tag{5.8}$$

$$\psi_2(\lambda, \mu) = \min_x \max_y [\Phi(x, y) + \langle \mu, x \rangle - \langle \lambda, y \rangle] \tag{5.9}$$

are simple, which means that they can be solved efficiently or in a closed form (see example in Section 4). Note that the conjugate problem is finite-dimensional. Using assumptions **A3**, **A4**, the fact that closed convex bounded set in Hilbert space is compact in 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 [3], we conclude that the saddle points in the problems (5.8), (5.9) do exist for all $\lambda \in \mathbb{R}^n$ and $\mu \in \mathbb{R}^m$.

Lemma 5.1. *Let Assumption* **A3** *be true. Then* $\psi_1(\lambda,\mu)$ *in* (5.8) *is smooth with partial gradients satisfying the following Lipschitz condition:*

$$\|\nabla_\mu\psi_1(\lambda_1,\mu_1)-\nabla_\mu\psi_1(\lambda_2,\mu_2)\| \le \frac{\|\mathcal{B}\|^2}{\sigma_{F_u}}\|\mu_1-\mu_2\| + \frac{\|\mathcal{B}\|\,\|\mathcal{C}\|\,L_{uv}}{\sigma_{F_u}\sigma_{F_v}}\|\lambda_1-\lambda_2\|\,, \tag{5.10}$$

$$\|\nabla_\lambda\psi_1(\lambda_1,\mu_1)-\nabla_\lambda\psi_1(\lambda_2,\mu_2)\| \le \frac{\|\mathcal{C}\|^2}{\sigma_{F_v}}\|\lambda_1-\lambda_2\| + \frac{\|\mathcal{B}\|\,\|\mathcal{C}\|\,L_{vu}}{\sigma_{F_u}\sigma_{F_v}}\|\mu_1-\mu_2\|\,. \tag{5.11}$$

Proof. From Lemma 4.2 we know that $\psi'_{1\mu}(\lambda,\mu)=-\mathcal{B}u^*(\lambda,\mu)$. From the strong convexity of the $F(\cdot,v)$, we have for any $t\in[0,1]$

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

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 (4.7) 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 (4.9) is unique and the function $\tilde{\psi}_1(u,\lambda)$ is smooth with respect to u. Denote u_i the optimal point in (4.7) for some λ,μ_i, $i=1,2$. From the first order optimality conditions for (4.7) 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:

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

Finally we have

$$\|\mathcal{B}u_1-\mathcal{B}u_2\|^2\le\|\mathcal{B}\|^2\|u_1-u_2\|^2\le\frac{\|\mathcal{B}\|^2}{\sigma_{F_u}}\langle\mathcal{B}^*(\mu_1-\mu_2),u_1-u_2\rangle\le\frac{\|\mathcal{B}\|^2}{\sigma_{F_u}}\|\mu_1-\mu_2\|\,\|\mathcal{B}(u_1-u_2)\| \tag{5.12}$$

Denote by (u_i,v_i) the saddle point in (5.8) 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 (4.8) and using strong concavity of $\hat{\psi}_1(v,\mu)$ we have:

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

This gives us

$$\|v_1-v_2\|\le\frac{\|\mathcal{C}\|}{\sigma_{F_v}}\|\lambda_1-\lambda_2\|\,.$$

From the first order optimality conditions for (4.10) we get:

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

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

$$\langle \nabla_u F(u_2, v_2) - \nabla_u F(u_2, v_1), u_1 - u_2 \rangle \geq \sigma_{F_u} \left\| u_1 - u_2 \right\|^2.$$

From this, using (5.1) and the estimate for $\|v_1 - v_2\|$, we get

$$\|u_1 - u_2\| \leq \frac{\|\mathcal{C}\| \, L_{uv}}{\sigma_{F_u} \sigma_{F_v}} \|\lambda_1 - \lambda_2\|.$$

Finally we have

$$\|\mathcal{B}u_1 - \mathcal{B}u_2\| \leq \|\mathcal{B}\| \, \|u_1 - u_2\| \leq \frac{\|\mathcal{C}\| \, \|\mathcal{B}\| \, L_{uv}}{\sigma_{F_u} \sigma_{F_v}} \|\lambda_1 - \lambda_2\| \tag{5.13}$$

Combining (5.12) and (5.13), we get (5.10). Estimate (5.11) can be proved in the same manner. □

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

Lemma 5.2. *Let Assumption* **A4** *be true. Then function* $\psi_2(\lambda, \mu)$ *in* (5.9) *is smooth with partial gradients satisfying the following Lipschitz condition:*

$$\|\nabla_\mu \psi_2(\lambda_1, \mu_1) - \nabla_\mu \psi_2(\lambda_2, \mu_2)\| \leq \frac{1}{\sigma_{\Phi x}} \|\mu_1 - \mu_2\| + \frac{L_{xy}}{\sigma_{\Phi x} \sigma_{\Phi y}} \|\lambda_1 - \lambda_2\| \tag{5.14}$$

$$\|\nabla_\lambda \psi_2(\lambda_1, \mu_1) - \nabla_\lambda \psi_2(\lambda_2, \mu_2)\| \leq \frac{1}{\sigma_{\Phi y}} \|\lambda_1 - \lambda_2\| + \frac{L_{yx}}{\sigma_{\Phi x} \sigma_{\Phi y}} \|\mu_1 - \mu_2\| \tag{5.15}$$

Combining Lemma 5.1 and 5.2, we get that the function $\psi(\lambda, \mu) \stackrel{\text{def}}{=} \psi_1(\lambda, \mu) + \psi_2(\lambda, \mu) - \langle \mu, \tilde{x}_0 \rangle + \langle \lambda, \tilde{y}_0 \rangle$ is convex in λ and concave in μ with partial gradients

$$\nabla_\lambda \psi(\lambda, \mu) = \mathcal{C}v^* + \tilde{y}_0 - y^*, \tag{5.16}$$

$$\nabla_\mu \psi(\lambda, \mu) = x^* - \tilde{x}_0 - \mathcal{B}u^*, \tag{5.17}$$

where (u^*, v^*) is the saddle point for problem (5.8), and (x^*, y^*) is the saddle point for problem (5.9).

5.1 Estimating the norms of λ^*, μ^*

Let us compute bounds for the norms of components λ, μ of the solution of the conjugate problem (5.7).

Lemma 5.3. *Let Assumptions* **A3**,**A4** *be true. Assume that* $P \subset \mathfrak{B}_{r_1}(u_0)$ *and* $Q \subset \mathfrak{B}_{r_2}(v_0)$. *Then* $\|\mu^*\| \leq L_{xx}\|\mathcal{B}\|r_1 + L_{xy}\|\mathcal{C}\|r_2 + \|\nabla_x \Phi(\mathcal{B}u_0 + \tilde{x}_0, \mathcal{C}v_0 + \tilde{y}_0)\|$ *and* $\|\lambda^*\| \leq L_{yx}\|\mathcal{B}\|r_1 + L_{yy}\|\mathcal{C}\|r_2 + \|\nabla_y \Phi(\mathcal{B}u_0 + \tilde{x}_0, \mathcal{C}v_0 + \tilde{y}_0)\|$.

Proof. Consider the function $\psi(\lambda^*, \mu)$. From Lemmas 5.1 and 5.2, we know that this function is concave and smooth in μ with gradient given by (5.17). Also, 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 problem (5.9), which can be rewritten as

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

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

Finally we have the following sequence of inequalities

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

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

5.2 Algorithm description

Let us introduce prox-function $d_\lambda(\lambda) = \frac{\sigma_\lambda}{2}\|\lambda\|^2$, where we use Euclidian norm. Function $d_\lambda(\lambda)$ is strongly convex in this norm with convexity parameter σ_λ. For the variable μ we introduce prox-function $d_\mu(\mu) = \frac{\sigma_\mu}{2}\|\mu\|^2$, which is strongly convex with convexity parameter σ_μ with respect to Euclidian norm. These prox-functions are differentiable everywhere.

For any $\lambda_1, \lambda_2 \in \mathbb{R}^n$ we can define 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 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 $\mathbb{R}^n$. Then we have $\omega_\lambda(\bar{\lambda}, \lambda) = d_\lambda(\lambda)$. For μ we introduce the similar settings.

In the same way as this was done in Section 4.3, we conclude that finding the saddle point (λ^*, μ^*) for conjugate problem (5.7) is equivalent to solving variational inequality

$$\langle g(\lambda, \mu), (\lambda - \lambda^*, \mu - \mu^*)\rangle \ge 0, \quad \forall \lambda, \mu, \tag{5.18}$$

where

$$g(\lambda, \mu) = (\nabla_\lambda \psi(\lambda, \mu), -\nabla_\mu \psi(\lambda, \mu)). \tag{5.19}$$

Let us choose some $\kappa \in (0, 1)$. Consider a space of $z \overset{\text{def}}{=} (\lambda, \mu)$ with the norm

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

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 $W = \mathbb{R}^n \times \mathbb{R}^m$, 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

$$\|g\| = \sqrt{\frac{1}{\kappa\sigma_\lambda}\|g_\lambda\|^2 + \frac{1}{(1-\kappa)\sigma_\mu}\|g_\mu\|^2}$$

Lemma 5.4. *Let Assumptions* **A3**, **A4** *be true, and* $\kappa = \frac{\sigma_\mu}{\sigma_\mu + \sigma_\lambda}$. *Then operator* $g(z)$ *defined in* (5.19) *is Lipschitz continuous:*

$$\|g(z_1) - g(z_2)\| \le L \|z_1 - z_2\| \tag{5.20}$$

with

$$L = \frac{\sigma_\lambda + \sigma_\mu}{\sigma_\mu \sigma_\lambda}\sqrt{2\left(\frac{\|\mathcal{C}\|^2}{\sigma_{F_v}} + \frac{1}{\sigma_{\Phi_y}} + \frac{\|\mathcal{B}\|\|\mathcal{C}\|L_{vu}}{\sigma_{F_u}\sigma_{F_v}} + \frac{L_{yx}}{\sigma_{\Phi_x}\sigma_{\Phi_y}}\right)\left(\frac{\|\mathcal{B}\|\|\mathcal{C}\|L_{uv}}{\sigma_{F_u}\sigma_{F_v}} + \frac{L_{xy}}{\sigma_{\Phi_x}\sigma_{\Phi_y}} + \frac{\|\mathcal{B}\|^2}{\sigma_{F_u}} + \frac{1}{\sigma_{\Phi_x}}\right)}. \tag{5.21}$$

Proof. Denote

$$\begin{aligned}
&c = \|\lambda_1 - \lambda_2\|, && d = \|\mu_1 - \mu_2\|, \\
&\alpha_1 = \frac{\|\mathcal{C}\|^2}{\sigma_{F_v}} + \frac{1}{\sigma_{\Phi_y}}, && \alpha_2 = \frac{\|\mathcal{B}\|\|\mathcal{C}\|L_{uv}}{\sigma_{F_u}\sigma_{F_v}} + \frac{L_{xy}}{\sigma_{\Phi_x}\sigma_{\Phi_y}}, \\
&\beta_1 = \frac{\|\mathcal{B}\|\|\mathcal{C}\|L_{vu}}{\sigma_{F_u}\sigma_{F_v}} + \frac{L_{yx}}{\sigma_{\Phi_x}\sigma_{\Phi_y}}, && \beta_2 = \frac{\|\mathcal{B}\|^2}{\sigma_{F_u}} + \frac{1}{\sigma_{\Phi_x}}.
\end{aligned}$$

Then from equations (5.10), (5.11), (5.14), (5.15) we have:

$$\begin{aligned}
\|\nabla_\lambda \psi(\lambda_1, \mu_1) - \nabla_\lambda \psi(\lambda_2, \mu_2)\|^2 &\le (\alpha_1 c + \beta_1 d)^2, \\
\|\nabla_\mu \psi(\lambda_1, \mu_1) - \nabla_\mu \psi(\lambda_2, \mu_2)\|^2 &\le (\alpha_2 c + \beta_2 d)^2.
\end{aligned}$$

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

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

Using the above expressions we obtain

$$\|g(z_1) - g(z_2)\|^2 = \frac{1}{\kappa\sigma_\lambda}\|\nabla_\lambda \psi(\lambda_1, \mu_1) - \nabla_\lambda \psi(\lambda_2, \mu_2)\|^2 + \frac{1}{(1-\kappa)\sigma_\mu}\|\nabla_\mu \psi(\lambda_1, \mu_1) - \nabla_\mu \psi(\lambda_2, \mu_2)\|^2$$

$$\le \frac{1}{\sigma}(\alpha_1 c + \beta_1 d)^2 + \frac{1}{\sigma}(\alpha_2 c + \beta_2 d)^2 \le \frac{2}{\sigma}(\alpha_1 c + \beta_1 d)(\alpha_2 c + \beta_2 d)$$

$$\le \frac{2}{\sigma}(\alpha_1\alpha_2 c^2 + \beta_1\beta_2 d^2 + (\alpha_1\beta_2 + \alpha_2\beta_1)cd) \le \frac{1}{\sigma}((2\alpha_1\alpha_2 + \alpha_1\beta_2 + \alpha_2\beta_1)c^2 + (2\beta_1\beta_2 + \alpha_1\beta_2 + \alpha_2\beta_1)d^2)$$

$$\le \frac{2}{\sigma}(\sqrt{\alpha_1\alpha_2(\alpha_1+\beta_1)(\alpha_2+\beta_2)}c^2 + \sqrt{\beta_1\beta_2(\alpha_1+\beta_1)(\alpha_2+\beta_2)}d^2)$$

$$\le \frac{2(\alpha_1+\beta_1)(\alpha_2+\beta_2)}{\sigma^2}(\kappa\sigma_\lambda c^2 + (1-\kappa)\sigma_\mu d^2) = \frac{2(\alpha_1+\beta_1)(\alpha_2+\beta_2)}{\sigma^2}\|z_1 - z_2\|^2.$$

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

$$
\begin{aligned}
L &= \sqrt{\tfrac{2(\alpha_1+\beta_1)(\alpha_2+\beta_2)}{\sigma^2}} \\
&= \tfrac{\sigma_\lambda+\sigma_\mu}{\sigma_\mu\sigma_\lambda}\sqrt{2\left(\tfrac{\|\mathcal{C}\|^2}{\sigma_{Fv}}+\tfrac{1}{\sigma_{\Phi y}}+\tfrac{\|\mathcal{B}\|\|\mathcal{C}\|L_{vu}}{\sigma_{Fu}\sigma_{Fv}}+\tfrac{L_{yx}}{\sigma_{\Phi x}\sigma_{\Phi y}}\right)\left(\tfrac{\|\mathcal{B}\|\|\mathcal{C}\|L_{uv}}{\sigma_{Fu}\sigma_{Fv}}+\tfrac{L_{xy}}{\sigma_{\Phi x}\sigma_{\Phi y}}+\tfrac{\|\mathcal{B}\|^2}{\sigma_{Fu}}+\tfrac{1}{\sigma_{\Phi x}}\right)}.
\end{aligned}
$$

□

In accordance to [5] for solving (5.18), we can use the following method:

1. Initialization: Fix $\beta = \frac{L}{\sigma_0}$. Set $s_{-1} = 0$.
2. Iteration ($k \geq 0$):

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

Here

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

Similarly to [4], we can prove that method (M2) generates bounded sequence $\{z_i\}_{i\geq 0}$. Hence the sequences $\{\lambda_i\}_{i\geq 0}, \{\mu_i\}_{i\geq 0}$ are also bounded. Also since the saddle point in the problem (3.4) exists, there exists a saddle point (λ^*, μ^*) for conjugate problem (5.7). These arguments allow us to choose D_λ, D_μ such that $d_\lambda(\lambda_i) \leq D_\lambda$, $d_\mu(\mu_i) \leq D_\mu$ for all $i \geq 0$, which also ensure that (λ^*, μ^*) is an interior solution:

$$B_{r/\sqrt{\kappa\sigma_\lambda}}(\lambda^*) \subseteq W_\lambda \overset{\text{def}}{=} \{\lambda : d_\lambda(\lambda) \leq D_\lambda\},$$

$$B_{r/\sqrt{(1-\kappa)\sigma_\mu}}(\mu^*) \subseteq W_\mu \overset{\text{def}}{=} \{\mu : d_\mu(\mu) \leq D_\mu\}$$

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

From Theorem 1 in [5], similarly to the proof of Theorem 2 in [5], we get the following lemma.

Lemma 5.5. *Assume that operator $g(z)$ is Lipschitz continuous on W with constant L. Let the sequence $\{z_i\}_{i\geq 0}$ be generated by method* (M2)*. Then for any $k \geq 0$ we have*

$$\delta_k(D) \leq \frac{LD}{\sigma_0}, \tag{5.22}$$

where $\delta_k(D)$ is defined in (4.14).

Theorem 5.1. *Let Assumptions* **A3** *and* **A4** *be true, and* $\kappa = \frac{\sigma_\mu}{\sigma_\mu + \sigma_\lambda}$, *L be defined in* (5.21). *Let the points* $z_i = (\lambda_i, \mu_i), i \geq 0$ *be generated by method* (M2). *Let points in* (4.16) *be defined by points* (u_i, v_i), (x_i, y_i) *which are the saddle points at the point* (λ_i, μ_i) *in* (5.8) *and* (5.9) *respectively. Then for functions* $\xi(u, x), \eta(v, y)$ *defined in* (4.20) *and* (4.21) *we have:*

$$\xi(\hat{u}_{k+1}, \hat{x}_{k+1}) - \eta(\hat{v}_{k+1}, \hat{y}_{k+1}) \leq \frac{LD}{k+1} \tag{5.23}$$

Also the following is true:

$$\|\mathcal{B}\hat{u}_{k+1} + \tilde{x}_0 - \hat{x}_{k+1}\| \leq \frac{LD\sqrt{\sigma_\mu}}{r(k+1)}, \quad \|\mathcal{C}\hat{v}_{k+1} + \tilde{y}_0 - \hat{y}_{k+1}\| \leq \frac{LD\sqrt{\sigma_\lambda}}{r(k+1)}. \tag{5.24}$$

Proof. Similarly to the proof of Theorem 4.1, we conclude that

$$\xi(\hat{u}_{k+1}, \hat{x}_{k+1}) - \eta(\hat{v}_{k+1}, \hat{y}_{k+1}) \leq \frac{\delta_k(D)}{k+1},$$

$$\|\mathcal{B}\hat{u}_{k+1} + \tilde{x}_0 - \hat{x}_{k+1}\| \leq \frac{\delta_k(D)\sqrt{\sigma_\mu}}{r(k+1)},$$

$$\|\mathcal{C}\hat{v}_{k+1} + \tilde{y}_0 - \hat{y}_{k+1}\| \leq \frac{\delta_k(D)\sqrt{\sigma_\lambda}}{r(k+1)}.$$

Using Lemma 5.5 and the fact that $\sigma_0 = 1$, these inequalities prove the statement of the theorem. □

References

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

[2] *Devolder, Olivier; Glineur, Franc,ois; Nesterov, Yurii* Double Smoothing Technique for Large-Scale Linearly Constrained Convex Optimization. In: SIAM Journal on Optimization, Vol. 22, no. 2, p. 702-727 (2012).

[3] *Maurice, Sion.* On general minimax theorems. // Pacific J. Math. Volume 8, Number 1 (1958), 171-176. http://projecteuclid.org/euclid.pjm/1103040253

[4] *Nesterov, Yurii.* Primal-dual subgradient methods for convex problems. In: Mathematical Programming, Vol. 120, no. 1, p. 221-259 (August 2009).

[5] *Nesterov, Yurii.* Dual extrapolation and its applications for solving variational inequalities and related problems. Journal Mathematical Programming: Series B Volume 109 Issue 2, January 2007 Pages 319 - 344