J Optim Theory Appl (2017) 172:402–435
DOI 10.1007/s10957-016-0949-3

Decomposition Techniques for Bilinear Saddle Point Problems and Variational Inequalities with Affine Monotone Operators

Bruce Cox[1] · **Anatoli Juditsky**[2] · **Arkadi Nemirovski**[3]

Received: 13 October 2015 / Accepted: 29 April 2016 / Published online: 13 June 2016

Abstract The majority of first-order methods for large-scale convex–concave saddle point problems and variational inequalities with monotone operators are *proximal* algorithms. To make such an algorithm practical, the problem's domain should be *proximal-friendly*—admit a strongly convex function with easy to minimize linear perturbations. As a by-product, this domain admits a computationally cheap linear minimization oracle (LMO) capable to minimize linear forms. There are, however, important situations where a cheap LMO indeed is available, but the problem domain is not proximal-friendly, which motivates search for algorithms based solely on LMO. For smooth convex minimization, there exists a classical algorithm using LMO—conditional gradient. In contrast, known to us similar techniques for other problems with convex structure (nonsmooth convex minimization, convex–concave saddle point problems, even as simple as bilinear ones, and variational inequalities with monotone operators, even as simple as affine) are quite recent and utilize common approach based on Fenchel-type representations of the associated objectives/vector fields. The goal of this paper was to develop alternative (and seemingly much simpler) *decomposition* techniques based on LMO for *bilinear* saddle point problems and for variational inequalities with *affine* monotone operators.

✉ Anatoli Juditsky
juditsky@imag.fr

Bruce Cox
bruce.cox@usafa.af.mil

Arkadi Nemirovski
nemirovs@isye.gatech.edu

1 US Air Force, Arlington, VA, USA

2 LJK, Université Grenoble Alpes, B.P. 53, 38041 Grenoble Cedex 9, France

3 Georgia Institute of Technology, Atlanta, GA 30332, USA

Keywords Decomposition techniques · Conditional gradients · Variational problems with affine monotone operator · Proximal algorithms

Mathematics Subject Classification 90C06 · 90C52 · 65K10 · 90C39

1 Introduction

This paper is a follow-up to our paper [1] and, same as its predecessor, is motivated by the desire to develop first-order algorithms for solving convex–concave saddle point problem (or variational inequality with monotone operator) on a convex domain X represented by *linear minimization oracle* (LMO) capable to minimize over X, at a reasonably low cost, any linear function. "LMO-representability" of a convex domain X is an essentially weaker assumption than "proximal friendliness" of X (possibility to minimize over X, at a reasonably low cost, any linear perturbation of a properly selected strongly convex function) underlying the vast majority of known first-order algorithms. There are important applications giving rise to LMO-represented domains which are *not* proximal-friendly, most notably

- nuclear norm balls arising in low-rank matrix recovery and in semidefinite optimization; here LMO reduces to approximating the leading pair of singular vectors of a matrix, while all known proximal algorithms require much costly computationally full singular value decomposition,
- total variation balls arising in image reconstruction; here LMO reduces to solving a specific flow problem [2], while a proximal algorithm needs to solve a much more computationally demanding linearly constrained convex quadratic program,
- some combinatorial polytopes [3].

The needs of these applications inspire the current burst of activity in developing LMO-based optimization techniques. In its major part, this activity was focused on smooth (or Lasso-type smooth regularized) convex minimization over LMO-represented domains, where the classical conditional gradient algorithm of Frank-Wolfe [4] and its modifications are applicable (see, e.g., [2,5–12] and references therein). LMO-based techniques for large-scale nonsmooth convex minimization, convex–concave saddle point problems (SP), even bilinear ones, and variational inequalities (VI) with monotone operators, even affine ones, where no classical optimization methods work, have been developed only recently. To the best of our knowledge, the related results reduce to LMO-based techniques for large-scale nonsmooth convex minimization based on Nesterov's smoothing [13–18]. A different approach to nonsmooth convex minimization, based on Fenchel-type representations of convex functions and processing induced by these representations problems dual to the problem of interest, was developed in [17] and was further extended in [1] to convex–concave SPs and VIs with monotone operators. The goal of this paper is to develop an alternative to [1] *decomposition-based* approach to solving convex–concave SPs and monotone VIs on LMO-represented domains. As we shall see, our decomposition approach can, in principle, handle general convex–concave SPs and monotone VIs. Our emphasis in this paper is, however, on *bilinear* SPs and on VIs with *affine* monotone operators—the cases which, on the one hand, are of primary importance in numerous applications,

and, on the other hand, are the cases where our approach is easy to implement and where this approach seems to be more flexible and much simpler than the machinery of Fenchel-type representations developed in [1] (and in fact even covers this machinery, see Sect. 3.4).

The rest of this paper is organized as follows. In Sects. 2 and 3, we present our decomposition-based approach to SP problems, respectively, to VIs with monotone operators, with emphasis on utilizing the approach to handle *bilinear* SPs and *affine* VIs, on *LMO-represented* domains. We illustrate our constructions by applying them to Colonel Blotto-type matrix game (Sect. 2.6.3) and Nash equilibrium with pairwise interactions (Sect. 3.2.2). In both these illustrations, decomposition allows to overcome difficulties coming from potentially huge ambient dimensions of the problems.

Proofs missing in the main body of the paper are relegated to "Appendix."

2 Decomposition of Convex–Concave Saddle Point Problems

Course of Actions: Outline

In the nutshell, out decomposition approach is extremely simple, and it makes sense to present an *informal* outline of it in the SP case.

Given convex and compact sets X_1, X_2, Y_1, Y_2 in Euclidean spaces, consider a convex–concave saddle point "master" problem

$$\min_{[x_1;x_2]\in X_1\times X_2}\ \max_{[y_1;y_2]\in Y_1\times Y_2} \Phi(x_1,x_2;y_1,y_2)$$

along with two "induced" problems

$$\begin{array}{ll}(P) & \min_{x_1\in X_1}\max_{y_1\in Y_1}\left[\phi(x_1,y_1):=\min_{x_2\in X_2}\max_{y_2\in Y_2}\Phi(x_1,x_2;y_1,y_2)\right],\\ (D) & \min_{x_2\in X_2}\max_{y_2\in Y_2}\left[\psi(x_2,y_2):=\min_{x_1\in X_1}\max_{y_1\in Y_1}\Phi(x_1,x_2;y_1,y_2)\right].\end{array}$$

It is easily seen that (P) and (D) are convex–concave problems, and a good approximate solution to the master problem induces straightforwardly equally good approximate solutions to (P) and to (D). More importantly, it turns out that when solving one of the induced problems, say, (P), by an "intelligent," in a certain precise sense, algorithm, information acquired in course of building an ϵ-solution allows to recover straightforwardly an ϵ-solution to the master problem, thus yielding an ϵ-solution to the other induced problem, in our case, to (D).

Now imagine that we want to solve a convex–concave SP problem which "as is" is too complicated for the standard solution techniques (e.g., problem's domain is not proximal-friendly, or is of huge dimension). Our proposed course of actions is to represent the problem of interest as the problem (D) stemming from a master problem built in a way which ensures that the associated problem (P) is amenable to an "intelligent" solution algorithm $\mathcal{B}$. After such a master problem is built, we solve (P) within a desired accuracy ϵ by $\mathcal{B}$ and use the acquired information to build an ϵ-solution to the problem of interest.

2.1 Situation

In this section, we focus on the situation as follows. We are given

1. convex and compact sets X_i in Euclidean spaces $\mathcal{X}_i$ and convex and compact sets Y_i in Euclidean spaces $\mathcal{Y}_i$, $i = 1, 2$;
2. convex and compact sets X, Y such that

$$X \subset X_1 \times X_2 \subset \mathcal{X} := \mathcal{X}_1 \times \mathcal{X}_2,\ Y \subset Y_1 \times Y_2 \subset \mathcal{Y} := \mathcal{Y}_1 \times \mathcal{Y}_2,$$

such that the projections of X onto $\mathcal{X}_i$ are the sets X_i, and projections of Y onto $\mathcal{Y}_i$ are the sets Y_i, $i = 1, 2$. For $x_1 \in X_1$ and $y_1 \in Y_1$ we set

$$X_2[x_1] = \{x_2 : [x_1; x_2] \in X\} \subset X_2 \text{ and } Y_2[y_1] = \{y_2 : [y_1; y_2] \in Y\} \subset Y_2.$$

Similarly,

$$X_1[x_2] = \{x_1 : [x_1; x_2] \in X\},\ x_2 \in X_2,$$

and

$$Y_1[y_2] = \{y_2 : [y_1; y_2] \in Y\},\ y_2 \in Y_2;$$

3. Lipschitz continuous function

$$\Phi(x = [x_1; x_2]; y = [y_1; y_2]) : X \times Y \to \mathbb{R}, \tag{1}$$

which is convex in $x \in X$, and concave in $y \in Y$.

We call the outlined situation a *direct product* one, when $X = X_1 \times X_2$ and $Y = Y_1 \times Y_2$.

2.2 Induced Convex–Concave Functions

We associate with Φ *primal and dual induced* functions:

$$\begin{aligned}
\phi(x_1, y_1) &:= \min_{x_2 \in X_2[x_1]} \max_{y_2 \in Y_2[y_1]} \Phi(x_1, x_2; y_1, y_2) \\
&= \max_{y_2 \in Y_2[y_1]} \min_{x_2 \in X_2[x_1]} \Phi(x_1, x_2; y_1, y_2) : X_1 \times Y_1 \to \mathbb{R}, \\
\psi(x_2, y_2) &:= \min_{x_1 \in X_1[x_2]} \max_{y_1 \in Y_1[y_2]} \Phi(x_1, x_2; y_1, y_2) \\
&= \max_{y_1 \in Y_1[y_2]} \min_{x_1 \in X_1[x_2]} \Phi(x_1, x_2; y_1, y_2) : X_2 \times Y_2 \to \mathbb{R}.
\end{aligned}$$

(the equalities are due to the convexity–concavity and continuity of Φ and convexity and compactness of $X_i[\cdot]$ and $Y_i[\cdot]$).

Recall [19] that a Lipschitz continuous convex–concave function

$$\theta(u, v) : U \times V \to \mathbb{R}$$

with convex compact U, V gives rise to the primal and dual problems

$$\begin{aligned}\mathrm{Opt}(P[\theta,U,V]) &= \min_{u\in U}\left[\overline{\theta}(u) := \max_{v\in V}\theta(u,v)\right],\\ \mathrm{Opt}(D[\theta,U,V]) &= \max_{v\in V}\left[\underline{\theta}(v) := \min_{u\in U}\theta(u,v)\right],\end{aligned}$$

with equal optimal values:

$$\mathrm{SadVal}(\theta,U,V) := \mathrm{Opt}(P[\theta,U,V)] = \mathrm{Opt}(D[\theta,U,V]),$$

same as gives rise to *saddle point residual*

$$\begin{aligned}\epsilon_{\mathrm{sad}}([u;v]|\theta,U,V) &= \overline{\theta}(u) - \underline{\theta}(v)\\ &= \left[\overline{\theta}(u) - \mathrm{Opt}(P[\theta,U,V])\right] + \left[\mathrm{Opt}(D[\theta,U,V]) - \underline{\theta}(v)\right].\end{aligned}$$

Lemma 2.1 *ϕ and ψ are convex–concave on their domains, are lower (upper) semi-continuous in their "convex" ("concave") arguments, and are Lipschitz continuous in the direct product case. Besides this, it holds*

$$\mathrm{SadVal}(\phi,X_1,Y_1) = \mathrm{SadVal}(\Phi,X,Y) = \mathrm{SadVal}(\psi,X_2,Y_2), \tag{2}$$

and whenever $\bar{x} = [\bar{x}_1;\bar{x}_2] \in X$ and $\bar{y} = [\bar{y}_1;\bar{y}_2] \in Y$, one has

$$\begin{aligned}\epsilon_{\mathrm{sad}}([\bar{x}_1;\bar{y}_1]\big|\phi,X_1,Y_1) &\le \epsilon_{\mathrm{sad}}([\bar{x};\bar{y}]\big|\Phi,X,Y),\\ \epsilon_{\mathrm{sad}}([\bar{x}_2;\bar{y}_2]\big|\psi,X_2,Y_2) &\le \epsilon_{\mathrm{sad}}([\bar{x};\bar{y}]\big|\Phi,X,Y).\end{aligned} \tag{3}$$

The strategy for solving SP problems we intend to develop is as follows:

1. We represent the SP problem of interest as the *dual SP problem*

$$\min_{x_2\in X_2}\max_{y_2\in Y_2}\psi(x_2,y_2) \tag{D}$$

induced by *master SP problem*

$$\min_{[x_1;x_2]\in X}\max_{[y_1;y_2]\in Y}\Phi(x_1,x_2;y_1,y_2). \tag{M}$$

The master SP problem is built in such a way that the associated *primal SP problem*

$$\min_{x_1\in X_1}\max_{y_1\in Y_1}\phi(x_1,y_1) \tag{P}$$

admits first-order oracle and can be solved by a traditional first-order method (e.g., a proximal one).

2. We solve (P) to a desired accuracy by first-order algorithm producing *accuracy certificates* [20] and use these certificates to recover an approximate solution of required accuracy to the problem of interest.

We shall see that the outlined strategy (originating from [21][1]) can be easily implemented when the problem of interest is a bilinear SP on the direct product of two LMO-represented domains.

2.3 Regular Sub- and Supergradients

Implementing the outlined strategy requires some "agreement" between the first-order information of the master and the induced SPs, and this is the issue we address now.

Given $\bar{x}_1 \in X_1$, $\bar{y}_1 \in Y_1$, let $\bar{x}_2 \in X_2[\bar{x}_1]$ and $\bar{y}_2 \in Y_2[\bar{y}_1]$ form a saddle point of the function $\Phi(\bar{x}_1, x_2; \bar{y}_1, y_2)$ (min in $x_2 \in X_2[\bar{x}_1]$, max in $y_2 \in Y_2[\bar{y}_1]$). In this situation, we say that $(\bar{x} = [\bar{x}_1; \bar{x}_2], \bar{y} = [\bar{y}_1; \bar{y}_2])$ belongs to the *saddle point frontier* of Φ, and we denote this frontier by $\mathcal{S}$.[2] Let now $\bar{z} = (\bar{x} = [\bar{x}_1; \bar{x}_2], \bar{y} = [\bar{y}_1; \bar{y}_2]) \in \mathcal{S}$, so that the function $\Phi(\bar{x}_1, x_2; \bar{y}_1, \bar{y}_2)$ attains its minimum over $x_2 \in X_2[\bar{x}_1]$ at $\bar{x}_2$, and the function $\Phi(\bar{x}_1, \bar{x}_2; \bar{y}_1, y_2)$ attains its maximum over $y_2 \in Y_2[\bar{y}_1]$ at $\bar{y}_2$. Consider a subgradient G of $\Phi(\cdot; \bar{y}_1, \bar{y}_2)$ taken at $\bar{x}$ along X: $G \in \partial_x \Phi(\bar{x}; \bar{y})$. We say that G is a *regular* subgradient of Φ at $\bar{z}$, if for some $g \in \mathcal{X}_1$ it holds

$$\forall x = [x_1; x_2] \in X : \langle G, x - \bar{x} \rangle \geq \langle g, x_1 - \bar{x}_1 \rangle;$$

every g satisfying this relation is called *compatible* with G. Similarly, we say that a supergradient H of $\Phi(\bar{x}; \cdot)$ taken at $\bar{y}$ along Y is a regular supergradient of Φ at $\bar{z}$, if for some $h \in \mathcal{Y}_1$ it holds

$$\forall y = [y_1; y_2] \in Y : \langle H, y - \bar{y} \rangle \leq \langle h, y_1 - \bar{y}_1 \rangle,$$

and every h satisfying this relation will be called *compatible* with H.

Remark 2.1 Let $X = X_1 \times X_2$, $Y = Y_1 \times Y_2$, meaning that we are in the direct product case. If $\Phi(x; \bar{y})$ is differentiable in x at $x = \bar{x}$, then the partial gradient $\nabla_x \Phi(\bar{x}; \bar{y})$ is a regular subgradient of Φ at $(\bar{x}, \bar{y})$, and $\nabla_{x_1} \Phi(\bar{x}; \bar{y})$ is compatible with this subgradient:

$$\begin{aligned} \forall x = [x_1; x_2] \in X_1 \times X_2 : \\ \langle \nabla_x \Phi(\bar{x}; \bar{y}), x - \bar{x} \rangle &= \langle \nabla_{x_1} \Phi(\bar{x}; \bar{y}), x_1 - \bar{x}_1 \rangle + \underbrace{\langle \nabla_{x_2} \Phi(\bar{x}; \bar{y}), x_2 - \bar{x}_2 \rangle}_{\geq 0} \\ &\geq \langle \nabla_{x_1} \Phi(\bar{x}; \bar{y}), x_1 - \bar{x}_1 \rangle. \end{aligned}$$

Similarly, if $\Phi(\bar{x}; y)$ is differentiable in y at $y = \bar{y}$, then the partial gradient $\nabla_y \Phi(\bar{x}; \bar{y})$ is a regular supergradient of Φ at $(\bar{x}, \bar{y})$, and $\nabla_{y_1} \Phi(\bar{x}; \bar{y})$ is compatible with this supergradient.

[1] In retrospect, a special case of this strategy was used in [22–24].

[2] Note that the saddle point frontier depends on the order of blocks in the x- and the y-variables, and this order will always be clear from the context.

Lemma 2.2 *In the situation of Sect. 2.1, let*

$$\bar{z} = (\bar{x} = [\bar{x}_1; \bar{x}_2], \bar{y} = [\bar{y}_1; \bar{y}_2]) \in \mathcal{S},$$

let G be a regular subgradient of Φ at $\bar{z}$ and let g be compatible with G. Let also H be a regular supergradient of Φ at $\bar{z}$, and h be compatible with H. Then, g is a subgradient in x_1, taken at $(\bar{x}_1, \bar{y}_1)$ along X_1, of the induced function ϕ, and h is a supergradient in y_1, taken at $(\bar{x}_1, \bar{y}_1)$ along Y_1, of the induced function ϕ:

$$\begin{array}{l}(a)\ \phi(x_1, \bar{y}_1) \geq \phi(\bar{x}_1; \bar{y}_1) + \langle g, x_1 - \bar{x}_1\rangle,\\ (b)\ \phi(\bar{x}_1, y_1) \leq \phi(\bar{x}_1; \bar{y}_1) + \langle h, y_1 - \bar{y}_1\rangle,\end{array}$$

for all $x_1 \in X_1$, $y_1 \in Y_1$.

Regular sub- and supergradient fields of induced functions. In the sequel, we say that $\phi'_{x_1}(x_1, y_1)$, $\phi'_{y_1}(x_1, y_1)$ are *regular* sub- and supergradient fields of ϕ, if for every $(x_1, y_1) \in X_1 \times Y_1$ and properly selected $\bar{x}_2$, $\bar{y}_2$ such that the point $\bar{z} = (\bar{x} = [x_1; \bar{x}_2], \bar{y} = [y_1; \bar{y}_2])$ is on the SP frontier of Φ, $\phi'_{x_1}(x_1, y_1)$, $\phi'_{y_1}(x_1, y_1)$ are the sub- and supergradients of ϕ induced, via Lemma 2.2, by regular sub- and supergradients of Φ at $\bar{z}$. Invoking Remark 2.1, we arrive at the following observation:

Remark 2.2 Let $X = X_1 \times X_2$, $Y = Y_1 \times Y_2$, meaning that we are in the direct product case. If Φ is differentiable in x and in y, then regular sub- and supergradient fields of ϕ can be built as follows: given $(x_1, y_1) \in X_1 \times Y_1$, we find $\bar{x}_2$, $\bar{y}_2$ such that the point $\bar{z} = (\bar{x} = [x_1; \bar{x}_2], \bar{y} = [y_1; \bar{y}_2])$ is on the SP frontier of Φ, and set

$$\phi'_{x_1}(x_1, y_1) = \nabla_{x_1}\Phi(x_1, \bar{x}_2; y_1, \bar{y}_2),\ \ \phi'_{y_1}(x_1, y_1) = \nabla_{y_1}\Phi(x_1, \bar{x}_2; y_1, \bar{y}_2). \qquad (4)$$

2.3.1 Existence of Regular Sub- and Supergradients

The notion of regular subgradient deals with Φ as a function of $[x_1; x_2] \in X$ only, the y-argument being fixed, so that the existence/description questions related to regular subgradient deal in fact with a Lipschitz continuous convex function on X. And of course the questions about existence/description of regular supergradients reduce straightforwardly to existence/description of regular subgradients (by swapping the roles of xs and ys and passing from Φ to $-\Phi$). Thus, as far as existence and description of regular sub- and supergradients are concerned, it suffices to consider the situation where

- $\Psi(x_1, x_2)$ is a Lipschitz continuous and convex function on X,
- $\bar{x}_1 \in X_1$, and $\bar{x}_2 \in X_2[\bar{x}_1]$ is a minimizer of $\Psi(\bar{x}_1, x_2)$ over $x_2 \in X_2[\bar{x}_1]$.

What we need to understand is when a subgradient G of Ψ taken at $\bar{x} = [\bar{x}_1; \bar{x}_2]$ along X and some g satisfy the relation

$$\langle G, [x_1; x_2] - \bar{x}\rangle \geq \langle g, x_1 - \bar{x}_1\rangle \quad \forall x = [x_1; x_2] \in X, \qquad (5)$$

and what can be said about the corresponding gs. The answer is as follows:

Lemma 2.3 *With Ψ, $\bar{x}_1$, $\bar{x}_2$ as above, $G \in \partial\Psi(\bar{x})$ satisfies (5) if and only if the following two properties hold:*

(i) G is a "certifying" subgradient of Ψ at $\bar{x}$, meaning that $\langle G, [0; x_2 - \bar{x}_2]\rangle \geq 0$ for all $x_2 \in X_2[\bar{x}_1]$ (the latter relation indeed certifies that $\bar{x}_2$ is a minimizer of $\Psi(\bar{x}_1, x_2)$ over $x_2 \in X_2[\bar{x}_1]$);

(ii) g is a subgradient, taken at $\bar{x}_1$ along X_1, of the convex function

$$\chi_G(x_1) = \min_{x_2 \in X_2[x_1]} \langle G, [x_1; x_2]\rangle.$$

It is easily seen that with Ψ, $\bar{x} = [\bar{x}_1; \bar{x}_2]$ as in Lemma 2.3 (i.e., Ψ is convex and Lipschitz continuous on X, $\bar{x}_1 \in X_1$, and $\bar{x}_2 \in X_2[\bar{x}_1]$ minimizes $\Psi(\bar{x}_1, x_2)$ over $x_2 \in X_2[\bar{x}_1]$) a certifying subgradient G always exists; when Ψ is differentiable at $\bar{x}$, one can take $G = \nabla_x \Psi(\bar{x})$. The function $\chi_G(\cdot)$, however, not necessary admits a subgradient at $\bar{x}_1$; when $\chi_G(\bar{x}_1) \neq \emptyset$, every $g \in \partial\chi_G(\bar{x}_1)$ satisfies (5). In particular,

1. [Direct Product case] When $X = X_1 \times X_2$, representing a certifying subgradient G of Ψ, taken at $[\bar{x}_1; \bar{x}_2 \in \mathrm{Argmin}_{x_2 \in X_2} \Psi(\bar{x}_1, x_2)]$, as $[g; h]$, we have

$$\langle h, x_2 - \bar{x}_2\rangle \geq 0 \quad \forall x_2 \in X_2,$$

   whence $\chi_G(x_1) = \langle g, x_1\rangle + \langle h, \bar{x}_2\rangle$, and thus g is a subgradient of χ_G at $\bar{x}_1$. In particular, in the direct product case and when Ψ is differentiable at $\bar{x}$, (5) is met by $G = \nabla\Psi(\bar{x})$, $g = \nabla_{x_1}\Psi(\bar{x})$;
2. [Polyhedral case] When X is a polyhedral set, for every certifying subgradient G of Ψ the function χ_G is polyhedrally representable with domain X_1, and as such has a subgradient at every point from X_1;
3. [Interior case] When $\bar{x}_1$ is a point from the relative interior of X_1, χ_G definitely has a subgradient at $\bar{x}_1$.

2.4 Main Result, Saddle Point Case

2.4.1 Preliminaries: Execution Protocols, Accuracy Certificates, Residuals

We start with outlining some simple concepts originating from [20]. Let W be a convex and compact set in Euclidean space $\mathcal{W}$, and $M(w) : W \to \mathcal{W}$ be a vector field on W. A *t-step execution protocol* associated with M, W is a collection $\mathcal{I}_t = \{w_i \in W, M(w_i) : 1 \leq i \leq t\}$. A *$t$-step accuracy certificate* is a t-dimensional probabilistic (i.e., with nonnegative entries summing up to 1) vector λ. Augmenting a t-step accuracy protocol $\mathcal{I}_t$ by t-step accuracy certificate λ gives rise to two entities:

$$\begin{aligned} &\text{approximate solution: } w^t = w^t(\mathcal{I}_t, \lambda) := \textstyle\sum_{i=1}^t \lambda_i w_i \in W; \\ &\text{residual: } \mathrm{Res}(\mathcal{I}_t, \lambda_t | W) = \max_{w \in W} \textstyle\sum_{i=1}^t \lambda_i \langle M(w_i), w_i - w\rangle. \end{aligned} \tag{6}$$

When $W = U \times V$, where U is a closed and convex subset of Euclidean space $\mathcal{U}$ and V is a closed and convex subset of Euclidean space $\mathcal{V}$, and M is vector field induced by convex–concave function $\theta(u, v) : U \times V \to \mathbb{R}$, that is,

$$M(u,v) = [M_u(u,v); M_v(u,v)] : U \times V \to \mathcal{U} \times \mathcal{V}$$

with

$$F_u(u,v) \in \partial_u \theta(u,v),\ F_v(u,v) \in \partial_v[-\theta(u,v)] \tag{7}$$

(such a field is always monotone), an execution protocol associated with (M, W) will be called also *protocol associated with* θ, U *and* V, or *protocol associated with the saddle point problem*

$$\min_{u\in U}\max_{v\in V} \theta(u,v). \tag{8}$$

Typical sources of execution protocols are first-order algorithms. For example, when (8) is solved by a first-order algorithm, the latter queries a first-order oracle along subsequently generated search points $w_1 \in W, w_2 \in W, \ldots$; the information returned by the oracle when queried at $w_i = [u_i; v_i]$ includes *at least* a subgradient of $\theta(u, v_i)$ w.r.t. u taken at $u = u_i$ and a subgradient of $-\theta(u_i, v)$ w.r.t. v taken at $v = v_i$; these two subgradients form the value $F(w_i)$ of a vector field $F(\cdot)$ of form (7). As a result, after $t = 1, 2, \ldots$ calls to the oracle, the algorithm produces a t-step execution protocol associated with θ, U, V.

The importance of these notions in our context stems from the following simple observation [20]:

Proposition 2.1 *Let U, V be nonempty, convex and compact domains in Euclidean spaces $\mathcal{U}$, $\mathcal{V}$, $\theta(u,v) : U \times V \to \mathbb{R}$ be a convex–concave function, and M be induced monotone vector field:*

$$\begin{aligned} M(u,v) = [M_u(u,v); M_v(u,v)] &: U \times V \to \mathcal{U} \times \mathcal{V} \\ &[M_u(u,v) \in \partial_u \theta(u,v),\ M_v(u,v) \in \partial_v[-\theta(u,v)]]. \end{aligned}$$

For a t-step execution protocol

$$\mathcal{I}_t = \{w_i = [u_i; v_i] \in W := U \times V,\ M_i = [M_u(u_i, v_i); M_v(u_i, v_i)] : 1 \le i \le t\}$$

associated with θ, U, V, and t-step accuracy certificate λ, it holds

$$\epsilon_{\mathrm{sad}}(w^t(\mathcal{I}_t, \lambda)\big|\theta, U, V) \le \mathrm{Res}(\mathcal{I}_t, \lambda\big|U \times V). \tag{9}$$

Proof For $[u; v] \in U \times V$, we have

$$\begin{aligned} \mathrm{Res}(\mathcal{I}_t, \lambda\big|U \times V) &\ge \sum_{i=1}^t \lambda_i \langle M_i, w_i - [u; v]\rangle \\ &= \sum_{i=1}^t \lambda_i [\underbrace{\langle M_u(u_i, v_i), u_i - u\rangle}_{\ge \theta(u_i,v_i) - \theta(u,v_i)} - \underbrace{\langle M_v(u_i, v_i), v_i - v\rangle}_{\le \theta(u_i,v_i) - \theta(u_i,v)}] \\ &\ge \sum_{i=1}^t \lambda_i [\theta(u_i, v) - \theta(u, v_i)] \ge \theta(u^t, v) - \theta(u, v^t), \end{aligned}$$

where the inequalities are due to the origin of M and convexity–concavity of θ. The resulting inequality holds true for all $[u; v] \in U \times V$, and (9) follows. □

2.4.2 Main Result

Proposition 2.2 *In the situation and notation of Sects. 2.1–2.3, let ϕ be the primal convex–concave function induced by Φ, and let*

$$\mathcal{I}_t = \{[x_{1,i}; y_{1,i}] \in X_1 \times Y_1, [\alpha_i := \phi'_{x_1}(x_{1,i}, y_{1,i}); \beta_i := -\phi'_{y_1}(x_{1,i}, y_{1,i})] : 1 \le i \le t\}$$

be an execution protocol associated with ϕ, X_1, Y_1, where ϕ'_{x_1}, ϕ'_{y_1} are regular sub- and supergradient fields associated with Φ, ϕ. Due to the origin of ϕ, ϕ'_{x_1}, ϕ'_{y_1}, there exist $x_{2,i} \in X_2[x_{1,i}]$, $G_i \in \mathcal{X}$, $y_{2,i} \in Y_2[y_{1,i}]$, and $H_i \in \mathcal{Y}$ such that

$$\begin{array}{ll}
(a) & G_i \in \partial_x \Phi(x_i := [x_{1,i}; x_{2,i}], y_i := [y_{1,i}; y_{2,i}]), \\
(b) & H_i \in \partial_y \left[-\Phi(x_i := [x_{1,i}; x_{2,i}], y_i := [y_{1,i}; y_{2,i}])\right], \\
(c) & \langle G_i, x - [x_{1,i}; x_{2,i}]\rangle \ge \langle \phi'_{x_1}(x_{1,i}, y_{1,i}), x_1 - x_{1,i}\rangle \quad \forall x = [x_1; x_2] \in X, \\
(d) & \langle H_i, y - [y_{1,i}; y_{2,i}]\rangle \ge \langle -\phi'_{y_1}(x_{1,i}, y_{1,i}), y_1 - y_{1,i}\rangle \quad \forall y = [y_1; y_2] \in Y,
\end{array} \tag{10}$$

implying that

$$\mathcal{J}_t = \left\{z_i = [x_i = [x_{1,i}; x_{2,i}]; y_i = [y_{1,i}; y_{2,i}]], F_i = [G_i; H_i] : 1 \le i \le t\right\}$$

is an execution protocol associated with Φ, X, Y. For every accuracy certificate λ, it holds

$$\mathrm{Res}(\mathcal{J}_t, \lambda \big| X \times Y) \le \mathrm{Res}(\mathcal{I}_t, \lambda \big| X_1 \times Y_1). \tag{11}$$

As a result, given an accuracy certificate λ and setting

$$[x^t; y^t] = [[x_1^t; x_2^t]; [y_1^t; y_2^t]] = \sum_{i=1}^{t} \lambda_i \left[[x_{1,i}; x_{2,i}]; [y_{1,i}; y_{2,i}]\right],$$

we ensure that

$$\epsilon_{\mathrm{sad}}([x^t; y^t] \big| \Phi, X, Y) \le \mathrm{Res}(\mathcal{I}_t, \lambda \big| X_1 \times Y_1), \tag{12}$$

whence also, by Lemma 2.1,

$$\begin{array}{l}
\epsilon_{\mathrm{sad}}([x_1^t; y_1^t] \big| \phi, X_1, Y_1) \le \mathrm{Res}(\mathcal{I}_t, \lambda \big| X_1 \times Y_1), \\
\epsilon_{\mathrm{sad}}([x_2^t; y_2^t] \big| \psi, X_2, Y_2) \le \mathrm{Res}(\mathcal{I}_t, \lambda \big| X_1 \times Y_1),
\end{array} \tag{13}$$

where ψ is the dual function induced by Φ.

Proof Let $z := [[u_1; u_2]; [v_1; v_2]] \in X \times Y$. Then,

$$\begin{aligned}\sum_{i=1}^t \lambda_i \langle F_i, z_i - z\rangle &= \sum_{i=1}^t \lambda_i \Big[\underbrace{\langle G_i, [x_{1,i}; x_{2,i}] - [u_1; u_2]\rangle}_{\le \langle \phi'_{x_1}(x_{1,i}, y_{1,i}), x_{1,i} - u_1\rangle \text{ by (10.c)}} \\ &\qquad + \underbrace{\langle H_i, [y_{1,i}; y_{2,i}] - [v_1; v_2]\rangle}_{\le \langle -\phi'_{y_1}(x_{1,i}, y_{1,i}), y_{1,i} - v_1\rangle \text{by (10.d)}} \Big] \\ &\le \sum_{i=1}^t \lambda_i \big[\langle \alpha_i, x_{1,i} - u_1\rangle + \langle \beta_i, y_{1,i} - v_1\rangle\big] \\ &\le \mathrm{Res}(\mathcal{I}_t, \lambda \big| X_1 \times Y_1),\end{aligned}$$

and (11) follows. □

2.5 Application: Solving Bilinear Saddle Point Problems on Domains Represented by Linear Minimization Oracles

2.5.1 Situation

Let W be a nonempty, convex and compact set in $\mathbb{R}^N$, Z be a nonempty, convex and compact set in $\mathbb{R}^M$, and let $\psi : W \times Z \to \mathbb{R}$ be bilinear convex–concave function:

$$\psi(w, z) = \langle w, p\rangle + \langle z, q\rangle + \langle z, Sw\rangle. \tag{14}$$

Our goal is to solve the convex–concave SP problem

$$\min_{w\in W} \max_{z\in Z} \psi(w, z) \tag{15}$$

given by ψ, W, Z.

2.5.2 Simple Observation

We intend to show that ψ can be represented (in fact, in many ways) as the dual function induced by a bilinear convex–concave function Φ; this is the key element of the outlined in Sect. 2.2 strategy for solving (15).

In the situation described in Sect. 2.5.1, let $U \subset \mathbb{R}^n$, $V \subset \mathbb{R}^m$ be convex and compact sets, and let $D \in \mathbb{R}^{m\times N}$, $A \in \mathbb{R}^{n\times M}$, $R \in \mathbb{R}^{m\times n}$. Consider bilinear (and thus convex–concave) function $[U \times W] \times [V \times Z] \to \mathbb{R}$:

$$\Phi(u, w; v, z) = \langle w, p + D^T v\rangle + \langle z, q + A^T u\rangle - \langle v, Ru\rangle \tag{16}$$

(the "convex" argument is (u, w), the "concave" one is (v, z)). Assume that a pair of functions

$$\begin{aligned}&\bar{u}(w, z) : W \times Z \to U, \\ &\bar{v}(w, z) : W \times Z \to V\end{aligned} \tag{17}$$

satisfies

$$\begin{array}{l} \forall (w,z) \in W \times Z : Dw = R\bar{u}(w,z), \\ \forall (w,z) \in W \times Z : \ \ Az = R^T \bar{v}(w,z). \end{array} \tag{18}$$

Denoting $\bar{u} = \bar{u}(w,z)$, $\bar{v} = \bar{v}(w,z)$, we have

$$\begin{array}{ll} (a) & \langle w, D^T \bar{v} \rangle = \langle Dw, \bar{v} \rangle = \langle R\bar{u}, \bar{v} \rangle, \\ (b) & \langle z, A^T \bar{u} \rangle = \langle Az, \bar{u} \rangle = \langle \bar{u}, R^T \bar{v} \rangle = \langle R\bar{u}, \bar{v} \rangle. \end{array} \tag{19}$$

Thus,

$$\begin{array}{l} \nabla_u \Phi(\bar{u}, w; \bar{v}, z) = Az - R^T \bar{v} = 0, \\ \nabla_v \Phi(\bar{u}, w; \bar{v}, z) = Dw - R\bar{u} = 0, \end{array}$$

whence

$$\begin{array}{l} \bar{\psi}(w,z) := \min_{u \in U} \max_{v \in V} \Phi(u, w; v, z) = \Phi(\bar{u}(w,z), w; \bar{v}(w,z), z) \\ \quad = \langle w, p \rangle + \langle z, q \rangle + \langle Dw, \bar{v}(w,z) \rangle \text{ [by (19)]}. \end{array}$$

We have proved

Lemma 2.4 *In the case of* (17), (18), *assuming that*

$$\langle Dw, \bar{v}(w,z) \rangle = \langle z, Sw \rangle \quad \forall (w \in W, z \in Z), \tag{20}$$

ψ is the dual convex–concave function induced by Φ and the domains $U \times W$, $V \times Z$.

Note that there are easy ways to ensure (18) and (20).

Example 2.1 Here $m = M$, $n = N$, and $D = A^T = R = S$. Assuming $U \supset W$, $V \supset Z$ and setting $\bar{u}(w,z) = w$, $\bar{v}(w,z) = z$, we ensure (17), (18) and (20).

Example 2.2 Let $S = A^T D$ with $A \in \mathbb{R}^{K \times M}$, $D \in \mathbb{R}^{K \times N}$. Setting $m = n = K$, $R = I_K$, $\bar{u}(w,z) = Dw$, $\bar{v}(w,z) = Az$ and assuming that $U \supset DW$, $V \supset AZ$, we again ensure (17), (18) and (20).

2.5.3 Implications

Assume that (17), (18) and (20) take place. Renaming the variables according to $x_1 \equiv u$, $y_1 \equiv v$, $x_2 \equiv w$, $y_2 \equiv z$ and setting $X_1 = U$, $X_2 = W$, $Y_1 = V$, $Y_2 = Z$, $X = X_1 \times X_2 = U \times W$, $Y = Y_1 \times Y_2 = V \times Z$, we find ourselves in the direct product case of the situation of Sect. 2.1, and Lemma 2.4 says that the bilinear SP problem of interest (14), (15) is the dual SP problem associated with the bilinear master SP problem

$$\begin{array}{l} \min_{[u;w] \in U \times W} \max_{[v;z] \in V \times Z} \big[\Phi(u, w; v, z) \\ \quad = \langle w, p + D^T v \rangle + \langle z, q + A^T u \rangle - \langle Ru, v \rangle \big]. \end{array} \tag{21}$$

Since Φ is linear in $[w; z]$, the primal SP problem associated with (21) is

$$\min_{\substack{u\equiv x_1\in \\ U=X_1}} \max_{\substack{v\equiv y_1\in \\ V=Y_1}} \left[\phi(u,v) = \min_{w\in W}\langle w, p + D^T v\rangle + \max_{z\in Z}\langle z, q + A^T u\rangle - \langle Ru, v\rangle\right].$$

Assuming that W, Z allow for cheap linear minimization oracles and defining $w_*(\cdot)$, $z_*(\cdot)$ according to

$$w_*(\xi) \in \operatorname*{Argmin}_{w\in W}\langle w, \xi\rangle, \ \ z_*(\eta) \in \operatorname*{Argmin}_{z\in Z}\langle z, \eta\rangle,$$

we have

$$\begin{aligned}
&\phi(u,v) = \langle w_*(p + D^T v), p + D^T v\rangle + \langle z_*(-q - A^T u), q + A^T u\rangle - \langle Ru, v\rangle,\\
&\phi'_u(u,v) := Az_*(-q - A^T u) - R^T v \in \partial_u \phi(u,v),\\
&\phi'_v(u,v) := Dw_*(p + D^T v) - Ru \in -\partial_v[-\phi(u,v)],
\end{aligned}$$

that is, first-order information on the primal SP problem

$$\min_{u\in U}\max_{v\in V}\phi(u,v), \tag{22}$$

is available. Note that since we are in the direct product case, ϕ'_u and ϕ'_v are regular sub- and supergradient fields associated with Φ, ϕ.

Now let

$$\mathcal{I}_t = \{[u_i; v_i] \in U\times V, [\gamma_i := \phi'_u(u_i, v_i); \delta_i := -\phi'_v(u_i, v_i)] : 1 \le i \le t\}$$

be an execution protocol generated by a first-order algorithm as applied to the primal SP problem (22). Let also

$$\begin{aligned}
w_i &= w_*(p + D^T v_i), z_i = z_*(-q - A^T u_i),\\
\alpha_i &= \nabla_w \Phi(u_i, w_i; v_i, z_i) = p + D^T v_i,\\
\beta_i &= -\nabla_z \Phi(u_i, w_i; v_i, z_i) = -q - A^T u_i,
\end{aligned}$$

so that suffices to consider the situation where

$$\mathcal{J}_t = \left\{[[u_i; w_i]; [v_i; z_i]], [\underbrace{[\gamma_i; \alpha_i]}_{\nabla_{[u;w]}\Phi(u_i,w_i;v_i,z_i)}; \underbrace{[\delta_i; \beta_i]}_{-\nabla_{[v;z]}\Phi(u_i,w_i;v_i,z_i)}] : 1 \le i \le t\right\}$$

is an execution protocol associated with the SP problem (21). By Proposition 2.2, for any accuracy certificate λ, it holds

$$\mathrm{Res}(\mathcal{J}_t, \lambda\big|U\times W\times V\times Z) \le \mathrm{Res}(\mathcal{I}_t, \lambda\big|U\times V); \tag{23}$$

whence, setting

$$[[u^t; w^t]; [v^t; z^t]] = \sum_{i=1}^{t} \lambda_i [[u_i; w_i]; [v_i; z_i]] \tag{24}$$

and invoking Proposition 2.1 with Φ in the role of θ,

$$\epsilon_{\text{sad}}([[u^t; w^t]; [v^t; z^t]] \big| \Phi, \underbrace{X_1 \times X_2}_{U \times W}, \underbrace{Y_1 \times Y_2}_{V \times Z}) \le \text{Res}(\mathcal{I}_t, \lambda \big| U \times V); \tag{25}$$

whence, by Lemma 2.1,

$$\epsilon_{\text{sad}}([w^t; z^t] \big| \psi, W, Z) \le \text{Res}(\mathcal{I}_t, \lambda \big| U \times V). \tag{26}$$

We have arrived at the following

Proposition 2.3 *In the situation of Sect. 2.5.1, let (17), (18) and (20) take place. Suppose that a first-order algorithm $\mathcal{B}$ with accuracy certificates is applied to the primal SP problem (22). Then, t-step execution protocol $\mathcal{I}_t$ and accuracy certificate λ^t, generated by $\mathcal{B}$, yield straightforwardly a feasible solution to the SP problem of interest (14)–(15) of the ϵ_{sad}-inaccuracy $\le \text{Res}(\mathcal{I}_t, \lambda^t \big| U \times V)$.*

Note also that when the constructions from Examples 1 and 2 are used, there is a significant freedom in selecting the domain $U \times V$ of the primal problem (we only require U, V to be convex and compact sets "large enough" to ensure the inclusions mentioned in Examples), so that there is no difficulty to enforce U, V to be proximal-friendly. As a result, we can take as $\mathcal{B}$ a proximal first-order method, for example, non-Euclidean restricted memory level algorithm with certificates [17] or mirror descent [1]. The efficiency estimates of these algorithms, as given in [1,17], imply that the resulting procedure for solving the SP of interest (14) – (15) admits nonasymptotic $O(1/\sqrt{t})$ rate of convergence, with explicitly computable factors hidden in $O(\cdot)$. The resulting complexity bound is completely similar to the one achievable with the machinery of Fenchel-type representations [1,17].

We are about to consider a special case where the $O(1/\sqrt{t})$ complexity admits a significant improvement.

2.6 Matrix Game Case

Let $S \in \mathbb{R}^{M \times N}$ admit representation

$$S = A^T D$$

with $A \in \mathbb{R}^{K \times M}$ and $D \in \mathbb{R}^{K \times N}$. Let also $W = \Delta_N = \{w \in \mathbb{R}^N_+ : \sum_i w_i = 1\}$, $Z = \Delta_M$. Our goal is to solve matrix game

$$\min_{w \in W} \max_{z \in Z} [\psi(w, z) = \langle z, Sw \rangle = \langle Az, Dw \rangle]. \tag{27}$$

Let U, V be convex and compact sets such that

$$V \supset AZ, \ U \supset DW, \tag{28}$$

and let us set

$$\begin{aligned}
\Phi(u, w; v, z) &= \langle u, Az\rangle + \langle v, Dw\rangle - \langle u, v\rangle,\\
\bar{u} := \bar{u}(w, z) &= Dw,\\
\bar{v} := \bar{v}(w, z) &= Az,
\end{aligned}$$

implying that

$$\begin{aligned}
\nabla_u \Phi(\bar{u}, w; \bar{v}, z) &= Az - \bar{v} = 0,\\
\nabla_v \Phi(\bar{u}, w; \bar{v}, z) &= Dw - \bar{u} = 0,\\
\Phi(\bar{u}, w; \bar{v}, z) &= \langle \bar{u}, Az\rangle + \langle \bar{v}, Dw\rangle - \langle \bar{u}, \bar{v}\rangle\\
&= \langle Dw, Az\rangle + \langle Az, Dw\rangle - \langle Dw, Az\rangle\\
&= \langle z, A^T Dw\rangle = \psi(w, z).
\end{aligned}$$

It is immediately seen that the function ψ from (27) is nothing but the dual convex–concave function associated with Φ, as in Example 2, while the primal function is

$$\phi(u, v) = \mathrm{Max}(A^T u) + \mathrm{Min}(D^T v) - \langle u, v\rangle; \tag{29}$$

here $\mathrm{Min}(p)$ and $\mathrm{Max}(p)$ stand for the smallest and the largest entries in vector p. Applying the strategy outlined in Sect. 2.2, we can solve the problem of interest (27) applying to the primal SP problem

$$\min_{u\in U}\max_{v\in V}\Big[\phi(u, v) = \mathrm{Min}(D^T v) + \mathrm{Max}(A^T u) - \langle u, v\rangle\Big], \tag{30}$$

an algorithm with accuracy certificates, and using the machinery outlined in previous sections to convert the resulting execution protocols and certificates into approximate solutions to the problem of interest (27).

We intend to consider a special case when the outlined approach allows to reduce a huge, but simple, matrix game (27) to a small SP problem (30)—so small that it can be solved to high accuracy by a cutting plane method (e.g., the ellipsoid algorithm). This is the case when the matrices A, D in (27) are *simple*—the notion to be defined in the next section and illustrated by a generic example ("knapsack-generated" matrices) in Sect. 2.6.2. In Sect. 2.6.3, we show how our approach allows to process, by an algorithm as simple as the ellipsoid method (an extension of) the well-known *Colonel Blotto Game* which may have really huge sizes (in our numerical illustration, $N = M \approx 10^{11}$).

2.6.1 Simple Matrices

Given a $K \times L$ matrix B, we call B *simple* if, given $x \in \mathbb{R}^K$, it is easy to identify the columns $\overline{B}[x]$, $\underline{B}[x]$ of B making the maximal, respectively, the minimal, inner product with x.

When matrices A, D in (27) are simple, the first-order information for the cost function ϕ in the primal SP problem (30) is easy to get. Besides, all we need from the convex and compact sets U, V participating in (30) is to be large enough to ensure that $U \supset DW$ and $V \supset AZ$, which allows to make U and V simple, e.g., Euclidean balls. Finally, when the design dimension $2K$ of (30) is small, we have at our disposal a multitude of linearly converging, with the converging ratio depending solely on K, methods for solving (30), including the ellipsoid algorithm with certificates presented in [20]. We are about to demonstrate that the outlined situation indeed takes place in some meaningful applications.

2.6.2 Example: Knapsack-Generated Matrices[3]

Assume that we are given *knapsack data*, namely

- positive integer *horizon* m,
- nonnegative integer *bounds* $\bar{p}_s$, $1 \le s \le m$,
- positive integer *costs* h_s, $1 \le s \le m$, and positive integer *budget* H, and
- *output functions* $f_s(\cdot) : \{0, 1, \ldots, \bar{p}_s\} \to \mathbb{R}^{r_s}$, $1 \le s \le m$.

Given the outlined data, consider the set $\mathcal{P}$ of all integer vectors $p = [p_1; \ldots; p_m]$ in $\mathbb{R}^m$ satisfying the following restrictions:

$$\begin{array}{ll} 0 \le p_s \le \overline{p}_s,\ 1 \le s \le m & \text{[range restriction]} \\ \sum_{s=1}^m h_s p_s \le H & \text{[budget restriction]} \end{array}$$

and the matrix B of the size $K \times \mathrm{Card}(\mathcal{P})$, $K = \sum_{s=1}^m r_s$, defined as follows: the columns of B are indexed by vectors $p = [p_1; \ldots; p_s] \in \mathcal{P}$, and the column indexed by p is the vector

$$B_p = [f_1(p_1); \ldots; f_m(p_m)].$$

Assuming that m, $\overline{p}_s$ and r_s are moderate, matrix B is simple—given $x \in \mathbb{R}^K$, it is easy to find $\overline{B}[x]$ and $\underline{B}[x]$ by Dynamic Programming.

Indeed, to identify $\overline{B}[x]$, $x = [x_1; \ldots; x_m] \in \mathbb{R}^{r_1} \times \cdots \times \mathbb{R}^{r_m}$ (identification of $\underline{B}[x]$ is completely similar), it suffices to run for $s = m, m-1, \ldots 1$ the backward Bellman recurrence: for $h = 0, 1, \ldots, H$,

$$\begin{aligned} U_s(h) &= \max_{r \in \mathbb{Z}} \left\{ U_{s+1}(h - h_s r) + \langle f_s(r), x_s \rangle : 0 \le r \le \overline{p}_s, 0 \le h - h_s r \right\} \\ A_s(h) &\in \operatorname*{Argmax}_{r \in \mathbb{Z}} \left\{ U_{s+1}(h - h_s r) + \langle f_s(r), x_s \rangle : 0 \le r \le \overline{p}_s, 0 \le h - h_s r \right\} \end{aligned}$$

[3] The construction to follow can be easily extended from "knapsack-generated" matrices to more general "Dynamic Programming-generated" ones, see Sect. 1 in the "Appendix."

with $U_{m+1}(\cdot) \equiv 0$, and then to recover one by one the entries p_s in the index $p \in \mathcal{P}$ of $\overline{B}[x]$ from the forward Bellman recurrence

$$\begin{aligned} H_1 &= H,\ p_1 = A_1(H_1); \\ H_{s+1} &= H_s - h_s p_s,\ p_{s+1} = A_{s+1}(H_{s+1}),\ 1 \le s < m. \end{aligned}$$

2.6.3 Illustration: Attacker versus Defender

The "covering story" we intend to consider is as follows.[4] Attacker and defender are preparing for a conflict to take place on m battlefields. A pure strategy of attacker is a vector $a = [a_1; \ldots; a_m]$, where nonnegative integer a_s, $1 \le s \le m$, is the number of attacking units to be created and deployed at battlefield s; the only restrictions on a, aside of nonnegativity and integrality, are the bounds $a_s \le \overline{a}_s$, $1 \le s \le m$, and the budget constraint $\sum_{s=1}^m h_{sA} a_s \le H_A$ with positive integer h_{sA} and H_A. Similarly, a pure strategy of defender is a vector $d = [d_1; \ldots; d_m]$, where nonnegative integer d_s is the number of defending units to be created and deployed at battlefield s, and the only restrictions on d, aside of nonnegativity and integrality, are the bounds $d_s \le \overline{d}_s$, $1 \le s \le m$, and the budget constraint $\sum_{s=1}^m h_{sD} d_s \le H_D$ with positive integer h_{sD} and H_D. The total loss of defender (the total gain of attacker), the pure strategies of the players being a and d, is

$$S_{a,d} = \sum_{s=1}^m [\Omega^s]_{a_s,d_s},$$

with given $(\overline{a}_s + 1) \times (\overline{d}_s + 1)$ matrices Ω^s. Our goal is to solve in mixed strategies the matrix game, where defender seeks to minimize his total loss, and attacker seeks to maximize it.

Let us denote by $\mathcal{A}$ and $\mathcal{D}$ the sets of pure strategies of attacker, respectively, defender. When representing

$$\Omega^s = \sum_{i=1}^{r_s} f^{is}[g^{is}]^T,\ f^{is} = [f_0^{is}; \ldots; f_{\overline{a}_s}^{is}],\ g^{is} = [g_0^{is}; \ldots; g_{\overline{d}_s}^{is}],\ r_s = \mathrm{Rank}(\Omega^s),$$

and setting

$$\begin{aligned} K &= \sum_{s=1}^m r_s, \\ A_a &= [[f_{a_1}^{1,1}; \ldots; f_{a_1}^{r_1,1}]; [f_{a_2}^{1,2}; \ldots; f_{a_2}^{r_2,2}]; \ldots; [f_{a_m}^{1,m}; \ldots; f_{a_m}^{r_m,m}]] \in \mathbb{R}^K, a \in \mathcal{A}, \\ D_d &= [[g_{d_1}^{1,1}; \ldots; g_{d_1}^{r_1,1}]; [g_{d_2}^{1,2}; \ldots; g_{d_2}^{r_2,2}]; \ldots; [g_{d_m}^{1,m}; \ldots; g_{d_m}^{r_m,m}]] \in \mathbb{R}^K, d \in \mathcal{D}, \end{aligned}$$

[4] This story is a variation of what is called "Colonel Blotto Game" in Game Theory; see, e.g., [25,26] and references therein.

we end up with $K \times M$, $M = \mathrm{Card}(\mathcal{A})$, knapsack-generated matrix A with columns A_a, $a \in \mathcal{A}$, and $K \times N$, $N = \mathrm{Card}(\mathcal{D})$, knapsack-generated matrix D with columns D_d, $d \in \mathcal{D}$, such that

$$S := [S_{a,d}]_{\substack{a \in \mathcal{A} \\ d \in \mathcal{D}}} = A^T D.$$

As a result, solving the attacker vs. defender game in mixed strategies reduces to solving SP problem (27) with knapsack-generated (and thus simple) matrices A, D and thus can be reduced to convex–concave SP (30) on the product of two K-dimensional convex and compact sets. Note that in the situation in question the design dimension $2K$ of (30) will, typically, be rather small (few tens or at most few hundreds), while the design dimensions M, N of the matrix game of interest (27) can be huge.

Numerical illustration. With the data (quite reasonable in terms of the "attacker vs. defender" game)

$$m = 8,\ h_{sA} = h_{sD} = 1, 1 \le s \le m,\ H_A = H_D = 64 = \overline{d}_s = \overline{a}_s,\ 1 \le s \le m$$

and rank 1 matrices Ω_s, $1 \le s \le m$, the design dimensions of the problem of interest (27) are as large as

$$\dim w = \dim z = 97\,082\,021\,465,$$

while the sizes of problem (30) are just

$$\dim u = \dim v = 8,$$

and thus (30) can be easily solved to high accuracy by the ellipsoid method. In our numerical experiment,[5] the outlined approach allowed to solve (27) within ϵ_{sad}-inaccuracy as small as 5.0e-9 in just 1537 steps of the ellipsoid algorithm (110.0 s on a mid range laptop). This performance is quite promising, especially when taking into account huge—nearly 10^{11}—sizes of the matrix game of interest (27).

3 From Saddle Point Problems to Variational Inequalities with Monotone Operators

In what follows, we extend the decomposition approach (developed so far for convex–concave SP problems) to variational inequalities (VIs) with monotone operators, with the primary goal to handle VIs with affine monotone operators on LMO-represented domains.

[5] For implementation details, see Sect. 1.

3.1 Decomposition of Variational Inequalities with Monotone Operators

3.1.1 Preliminaries

Recall that the (Minty's) variational inequality VI(M, W) associated with a convex and compact subset W of Euclidean space $\mathcal{W}$ and a vector field $M : W \to \mathcal{W}$ is

$$\text{find } w \in W : \langle M(w'), w' - w\rangle \geq 0 \quad \forall w' \in W; \tag{VI(M,W)}$$

w satisfying the latter condition is called a *weak* solution to the VI. A natural measure of inaccuracy for an approximate solution $w \in W$ to VI(M, W) is the *dual-gap function*

$$\epsilon_{\mathrm{VI}}(w\big|M, W) = \sup_{w' \in W} \langle M(w'), w - w'\rangle;$$

weak solutions to the VI are exactly the points of W where this (clearly nonnegative everywhere on W) function is zero.

In the sequel, we utilize the following simple fact originating from [20]:

Proposition 3.1 *Let M be monotone on W, let*

$$\mathcal{I}_t = \{w_i \in W,\ M(w_i) : 1 \leq i \leq t\}$$

be a t-step execution protocol associated with (M, W), λ be a t-step accuracy certificate, and $w^t = \sum_{i=1}^t \lambda_i w_i$ be the associated approximate solution. Then,

$$\epsilon_{\mathrm{VI}}(w^t\big|M, W) \leq \mathrm{Res}(\mathcal{I}_t, \lambda\big|W).$$

Proof We have

$$\begin{aligned}
\mathrm{Res}(\mathcal{I}_t, \lambda\big|W) &= \sup_{w' \in W}\left[\sum_{i=1}^t \lambda_i \langle M(w_i), w_i - w'\rangle\right] \\
&\geq \sup_{w' \in W}\left[\sum_{i=1}^t \lambda_i \langle M(w'), w_i - w'\rangle\right] \text{ [since } M \text{ is monotone]} \\
&= \sup_{w' \in W} \langle M(w'), w^t - w'\rangle = \epsilon_{\mathrm{VI}}(w^t\big|M, W).
\end{aligned}$$

□

3.1.2 Situation

Let $\mathcal{X}$, $\mathcal{H}$ be Euclidean spaces, $\Theta \subset \mathcal{X} \times \mathcal{H}$ be convex and compact set, Ξ be the projection of Θ onto $\mathcal{X}$, and H be the projection of Θ onto $\mathcal{H}$. Given $\xi \in \Xi$, $\eta \in H$, we set

$$H_\xi = \{\eta : [\xi; \eta] \in \Theta\},\ \ \Xi_\eta = \{\xi \in \Xi : [\xi; \eta] \in \Theta\}.$$

We denote a point from $\mathcal{X} \times \mathcal{H}$ as $\theta = [\xi; \eta]$ with $\xi \in \mathcal{X}$, $\eta \in \mathcal{H}$. Let, further,

$$\Phi(\xi, \eta) = [\Phi_\xi(\xi, \eta); \Phi_\eta(\xi, \eta)] : \Theta \to \mathcal{X} \times \mathcal{H}$$

be a continuous and monotone vector field.

Let $\xi \in \Xi$, and let $\overline{\eta} = \overline{\eta}(\xi)$ be a somehow selected, as a function of $\xi \in \Xi$, strong solution to the VI given by $(H_\xi, \Phi_\eta(\xi, \eta))$, that is,

$$\overline{\eta}(\xi) \in H_\xi \ \& \ \langle \Phi_\eta(\xi, \overline{\eta}(\xi)), \eta - \overline{\eta}(\xi) \rangle \geq 0 \quad \forall \eta \in H_\xi. \tag{31}$$

Let us call Φ (more precisely, the pair $(\Phi, \overline{\eta}(\cdot))$) *$\eta$-regular*, if for every $\xi \in \Xi$, there exists $\Psi = \Psi(\xi) \in \mathcal{X}$ such that

$$\langle \Psi(\xi), \xi' - \xi \rangle \leq \langle \Phi(\xi, \overline{\eta}(\xi)), [\xi'; \eta'] - [\xi; \overline{\eta}(\xi)] \rangle \quad \forall [\xi'; \eta'] \in \Theta. \tag{32}$$

Similarly, let $\overline{\xi}(\eta)$ be a somehow selected, as a function of $\eta \in H$, strong solution to the VI given by $(\Xi_\eta, \Phi_\eta(\xi, \eta))$, that is,

$$\overline{\xi}(\eta) \in \Xi_\eta \ \& \ \langle \Phi_\xi(\overline{\xi}(\eta), \eta), \xi - \overline{\xi}(\eta) \rangle \geq 0 \quad \forall \xi \in \Xi_\eta. \tag{33}$$

Let us call $(\Phi, \overline{\xi}(\cdot))$ *ξ-regular*, if for every $\eta \in H$, there exists $\Gamma = \Gamma(\eta) \in \mathcal{H}$ such that

$$\langle \Gamma(\eta), \eta' - \eta \rangle \leq \langle \Phi(\overline{\xi}(\eta), \eta), [\xi'; \eta'] - [\overline{\xi}(\eta); \eta] \rangle \quad \forall [\xi'; \eta'] \in \Theta. \tag{34}$$

When $(\Phi, \overline{\eta})$ is η-regular, we refer to the above $\Psi(\cdot)$ as to a *primal* vector field induced by Φ,[6] and when $(\Phi, \overline{\xi})$ is ξ-regular, we refer to the above $\Gamma(\cdot)$ as to a *dual* vector field induced by Φ.

Example 3: Direct product case. This is the case where $\Theta = \Xi \times H$. In this situation, setting $\Psi(\xi) = \Phi_\xi(\xi, \overline{\eta}(\xi))$, we have for $[\xi'; \eta'] \in \Theta$:

$$\begin{aligned} \langle \Phi(\xi, \overline{\eta}(\xi)), [\xi'; \eta'] - [\xi; \overline{\eta}(\xi)] \rangle &= \underbrace{\langle \Phi_\xi(\xi, \overline{\eta}(\xi)), \xi' - \xi \rangle}_{= \langle \Psi(\xi), \xi' - \xi \rangle} + \underbrace{\langle \Phi_\eta(\xi, \overline{\eta}(\xi)), \eta' - \overline{\eta}(\xi) \rangle}_{\geq 0 \quad \forall \eta' \in H_\xi = H} \\ &\geq \langle \Psi(\xi), \xi' - \xi \rangle, \end{aligned}$$

that is, $(\Phi, \overline{\eta}(\cdot))$ is η-regular, with $\Psi(\xi) = \Phi_\xi(\xi, \overline{\eta}(\xi))$. Setting $\Gamma(\eta) = \Phi_\eta(\overline{\xi}(\eta), \eta)$, we get by similar argument

$$\langle \Phi(\overline{\xi}(\eta)), [\xi'; \eta'] - [\overline{\xi}; \eta] \rangle \geq \langle \Gamma(\eta), \eta' - \eta \rangle, \eta, \eta' \in H,$$

that is, $(\Phi, \overline{\xi}(\cdot))$ is ξ-regular, with $\Gamma(\eta) = \Phi_\eta(\overline{\xi}(\eta), \eta)$.

[6] "a primal" instead of "the primal" reflects the fact that Ψ is not uniquely defined by Φ—it is defined by Φ *and* $\overline{\eta}$ *and* by how the values of Ψ are selected when (32) does not specify these values uniquely.

3.1.3 Main Result: Variational Inequality Case

Proposition 3.2 *In the situation of Sect. 3.1.2, let $(\Phi, \overline{\eta}(\cdot))$ be η-regular. Then,*

(i) *Primal vector field $\Psi(\xi)$ induced by $(\Phi, \overline{\eta}(\cdot))$ is monotone on Ξ. Moreover, whenever $\mathcal{I}_t = \{\xi_i \in \Xi, \Psi(\xi_i) : 1 \le i \le t\}$ and $\mathcal{J}_t = \{\theta_i := [\xi_i; \overline{\eta}(\xi_i)], \Phi(\theta_i) : 1 \le i \le t\}$ and λ is a t-step accuracy certificate, it holds s*

$$\epsilon_{\mathrm{VI}}\left(\sum_{i=1}^{t} \lambda_i \theta_i \big| \Phi, \Theta\right) \le \mathrm{Res}(\mathcal{J}_t, \lambda \big| \Theta) \le \mathrm{Res}(\mathcal{I}_t, \lambda \big| \Xi\Big). \tag{35}$$

(ii) *Let $(\Phi, \overline{\xi})$ be ξ-regular, and let Γ be the induced dual vector field. Whenever $\widehat{\theta} = [\widehat{\xi}; \widehat{\eta}] \in \Theta$, we have*

$$\epsilon_{\mathrm{VI}}(\widehat{\eta} \big| \Gamma, H) \le \epsilon_{\mathrm{VI}}(\widehat{\theta} \big| \Phi, \Theta). \tag{36}$$

3.2 Implications

In the situation of Sect. 3.1.2, assume that for properly selected $\overline{\eta}(\cdot)$, $\overline{\xi}(\cdot)$, $(\Phi, \overline{\eta}(\cdot))$ is η-regular, and $(\Phi, \overline{\xi}(\cdot))$ is ξ-regular, induced primal and dual vector fields being Ψ and Γ. In order to solve the *dual VI* $\mathrm{VI}(\Gamma, H)$, we can apply to the *primal VI* $\mathrm{VI}(\Psi, \Xi)$ an algorithm with accuracy certificates; by Proposition 3.2.i, resulting t-step execution protocol $\mathcal{I}_t = \{\xi_i, \Psi(\xi_i) : 1 \le i \le t\}$ and accuracy certificate λ generate an execution protocol

$\mathcal{J}_t = \{\theta_i := [\xi_i; \overline{\eta}(\xi_i)], \Phi(\theta_i) : 1 \le i \le t\}$ such that

$$\mathrm{Res}(\mathcal{J}_t, \lambda \big| \Theta) \le \mathrm{Res}(\mathcal{I}_t, \lambda \big| \Xi),$$

whence, by Proposition 3.1, for the approximate solution

$$\theta^t = [\xi^t, \eta^t] := \sum_{i=1}^{T} \lambda_i \theta_i = \sum_{i=1}^{t} \lambda_i [\xi_i; \overline{\eta}(\xi_i)]$$

it holds

$$\epsilon_{\mathrm{VI}}(\theta^t \big| \Phi, \Theta) \le \mathrm{Res}(\mathcal{I}_t, \lambda \big| \Xi).$$

Invoking Proposition 3.2.ii, we conclude that η^t is a feasible solution to the dual VI $\mathrm{VI}(\Gamma, H)$, and

$$\epsilon_{\mathrm{VI}}(\eta^t \big| \Gamma, H) \le \mathrm{Res}(\mathcal{I}_t, \lambda \big| \Xi). \tag{37}$$

We are about to present two examples well suited for the just outlined approach.

3.2.1 Solving Affine Monotone VI on LMO-Represented Domain

Let H be a convex and compact set in $\mathcal{H} = \mathbb{R}^N$, and let H be equipped with an LMO. Assume that we want to solve the VI $\mathrm{VI}(F, H)$, where

$$F(\eta) = S\eta + s$$

is an affine monotone operator (so that $S + S^T \succeq 0$). Let us set $\mathcal{X} = \mathcal{H}$, select Ξ as a proximal-friendly convex and compact set containing H, and set $\Theta = \Xi \times H$,

$$\Phi(\xi, \eta) = \underbrace{\left[\begin{array}{c|c} S^T & -S^T \\ \hline S & \end{array}\right]}_{\mathcal{S}} \left[\begin{array}{c} \xi \\ \eta \end{array}\right] + \left[\begin{array}{c} \\ s \end{array}\right]$$

(here and in what follows blank fields in matrices/vectors represent zero blocks). We have

$$\mathcal{S} + \mathcal{S}^T = \left[\begin{array}{c|c} S + S^T & \\ \hline & \end{array}\right] \succeq 0,$$

so that Φ is an affine monotone operator with

$$\begin{aligned} \Phi_\xi(\xi, \eta) &= S^T\xi - S^T\eta, \\ \Phi_\eta(\xi, \eta) &= S\xi + s. \end{aligned}$$

Setting $\overline{\xi}(\eta) = \eta$, we ensure that $\overline{\xi}(\eta) \in \Xi$ when $\eta \in H$ and $\Phi_\xi(\overline{\xi}(\eta), \eta) = 0$, implying (33). Since we are in the direct product case, we can set

$$\Gamma(\eta) = \Phi_\eta(\overline{\xi}(\eta), \eta) = S\eta + s = F(\eta);$$

thus, VI(Γ, H) is our initial VI of interest. On the other hand, setting

$$\overline{\eta}(\xi) \in \operatorname*{Argmin}_{\eta \in H} \langle S\xi + s, \eta \rangle,$$

we ensure (31). Since we are in the direct product case, we can set

$$\Psi(\xi) = \Phi_\xi(\xi, \overline{\eta}(\xi)) = S^T[\xi - \overline{\eta}(\xi)];$$

note that the values of Ψ can be straightforwardly computed via calls to the LMO representing H. We can now solve VI(Ψ, Ξ) by a proximal algorithm $\mathcal{B}$ with accuracy certificates and recover, as explained above, approximate solution to the VI of interest VI(F, H). With the non-Euclidean restricted memory level method with certificates [17] or Mirror Descent with certificates (see, e.g., [1]), the approach results in nonasymptotical $O(1/\sqrt{t})$-converging algorithm for solving the VI of interest, with explicitly computable factors hidden in $O(\cdot)$. This complexity bound, completely similar to the one obtained in [1], seems to be the best known under the circumstances.

3.2.2 Solving Skew-Symmetric VI on LMO-Represented Domain

Let H be an LMO-represented convex and compact domain in $\mathcal{H} = \mathbb{R}^N$ and assume that we want to solve VI(F, H), where

$$F(\eta) = 2Q^T P\eta + f : \mathcal{H} \to \mathcal{H}$$

with $K \times N$ matrices P, Q such that the matrix $Q^T P$ is skew-symmetric:

$$Q^T P + P^T Q = 0. \tag{38}$$

Let $\mathcal{X} = \mathbb{R}^K \times \mathbb{R}^K$, and let Ξ_1, Ξ_2 be two convex and compact sets in $\mathbb{R}^K$ such that

$$QH \subset \Xi_1, -PH \subset \Xi_2. \tag{39}$$

Let us set $\Xi = \Xi_1 \times \Xi_2$, and let

$$\Phi(\xi = [\xi_1; \xi_2], \eta) = \left[\begin{array}{c|c|c} & I_K & P \\ \hline -I_K & & Q \\ \hline -P^T & -Q^T & \end{array}\right] \left[\begin{array}{c} \xi_1 \\ \xi_2 \\ \eta \end{array}\right] + \left[\begin{array}{c} \\ \\ f \end{array}\right].$$

Note that Φ is monotone and affine. Setting

$$\overline{\xi}(\eta) = [Q\eta; -P\eta]$$

and invoking (39), we ensure (33); since we are in the direct product case, we can take, as the dual induced vector field,

$$\begin{aligned} \Gamma(\eta) &= \Phi_\eta(\overline{\xi}(\eta), \eta) = -P^T(Q\eta) - Q^T(-P\eta) + f \\ &= [Q^T P - P^T Q]\eta + f \underbrace{=}_{\text{by (38)}} 2Q^T P\eta + f = F(\eta), \end{aligned}$$

so that the dual VI, VI(Γ, H), is our VI of interest.

On the other hand, setting

$$\overline{\eta}(\xi = [\xi_1; \xi_2]) \in \operatorname*{Argmin}_{\eta \in H} \langle f - P^T\xi_1 - Q^T\xi_2, \eta \rangle,$$

we ensure (31). Since we are in the direct product case, we can define primal vector field as

$$\Psi(\xi = [\xi_1; \xi_2]) = \Phi_\xi([\xi_1; \xi_2], \overline{\eta}([\xi_1; \xi_2])) = \left[\begin{array}{c} \xi_2 + P\overline{\eta}(\xi) \\ -\xi_1 + Q\overline{\eta}(\xi) \end{array}\right].$$

Note that LMO for H allows to compute the values of Ψ, and that Ξ can be selected to be proximal-friendly. We can now solve VI(Ψ, Ξ) by a proximal algorithm $\mathcal{B}$ with accuracy certificates and recover, as explained above, approximate solution to the VI of interest VI(F, H). When the design dimension dim Ξ of the primal VI is small, other choices of $\mathcal{B}$, like the ellipsoid algorithm, are possible, and in this case we can end up with linearly converging, with the converging ratio depending solely on dim Ξ, algorithm for solving the VI of interest. We are about to give a related example, which can be considered as multi-player version of the "attacker vs. defender" game.

3.3 Nash Equilibrium with Pairwise Interactions

Consider the situation as follows: there are

- $L \geq 2$ players, ℓth of them selecting a mixed strategy w_ℓ from probabilistic simplex Δ_{N_ℓ} of dimension N_ℓ,
- *encoding matrices* D_ℓ of sizes $m_\ell \times N_\ell$, and *loss matrices* $M^{\ell\ell'}$ of sizes $m_\ell \times m_{\ell'}$ such that

$$M^{\ell\ell} = 0, M^{\ell\ell'} = -[M^{\ell'\ell}]^T,\ 1 \leq \ell, \ell' \leq L.$$

- The loss of ℓth player depends on mixed strategies of the players according to

$$\mathcal{L}_\ell(\eta := [w_1; \ldots; w_L]) = \sum_{\ell'=1}^{L} w_\ell^T E^{\ell\ell'} w_{\ell'},\ E^{\ell\ell'} = D_\ell^T M^{\ell\ell'} D_{\ell'} + \langle g_\ell, \eta \rangle.$$

  In other words, every pair of distinct players ℓ, ℓ' is playing matrix game with matrix $M^{\ell\ell'}$, and the loss of player ℓ, up to a linear in $[w_1; \ldots; w_L]$ function, is the sum, over the pairwise games he is playing, of his losses in these games, the "coupling constraints" being expressed by the requirement that every player uses the same mixed strategy in all pairwise games he is playing.

We have described *convex* Nash equilibrium problem, meaning that for every ℓ, $\mathcal{L}_\ell(w_1, \ldots, w_L)$ is convex (in fact, linear) in w_ℓ, is jointly concave (in fact, linear) in $w^\ell := (w_1, \ldots, w_{\ell-1}, w_{\ell+1}, \ldots, w_L)$, and $\sum_{\ell=1}^{L} \mathcal{L}_\ell(\eta)$ is the linear function $\langle g, \eta \rangle$, $g = \sum_\ell g_\ell$, and thus is convex. It is known (see, for example, [20]) that Nash equilibria in convex Nash problem are exactly the weak solutions to the VI given by monotone operator

$$F(\eta := [w_1; \ldots; w_L]) = [\nabla_{w_1} \mathcal{L}_1(\eta); \ldots; \nabla_{w_L} \mathcal{L}_L(\eta)]$$

on the domain

$$H = \Delta_{N_1} \times \cdots \times \Delta_{N_L}.$$

Let us set

$$Q := \frac{1}{2} \begin{bmatrix} D_1 & & & \\ & D_2 & & \\ & & \ddots & \\ & & & D_L \end{bmatrix},\ P = \begin{bmatrix} M^{1,1} D_1 & M^{1,2} D_2 & \ldots & M^{1,L} D_L \\ M^{2,1} D_1 & M^{2,2} D_2 & \ldots & M^{2,L} D_L \\ \vdots & \vdots & \ddots & \ldots \\ M^{L,1} D_1 & M^{L,2} D_2 & \ldots & M^{L,L} D_L \end{bmatrix}.$$

Then,

$$Q^T P = \frac{1}{2}\begin{bmatrix} D_1^T M^{1,1} D_1 & D_1^T M^{1,2} D_2 & \dots & D_1^T M^{1,L} D_L \\ D_2^T M^{2,1} D_1 & D_2^T M^{2,2} D_2 & \dots & D_2^T M^{1,L} D_L \\ \vdots & \vdots & \ddots & \dots \\ D_L^T M^{L,1} D_1 & D_L^T M^{L,2} D_2 & \dots & D_L^T M^{L,L} D_L \end{bmatrix},$$

so that $Q^T P$ is skew-symmetric due to $M^{\ell\ell'} = -[M^{\ell'\ell}]^T$. Besides this, we clearly have

$$\begin{aligned} &F(\eta := [w_1; \dots; w_L]) = 2Q^T P\eta + f, \\ &f = [\nabla_{w_1}\langle g_1, [w_1; \dots; w_L]\rangle; \dots; \nabla_{w_L}\langle g_L, [w_1; \dots; w_L]\rangle]. \end{aligned}$$

Observe that, if $D_1, \dots, D_L$ are simple, then so are Q and P.

Indeed, for Q this is evident: to find the column of Q which makes the largest inner product with $x = [x_1; \dots; x_L]$, $\dim x_\ell = m_\ell$, it suffices to find, for every ℓ, the column of D_ℓ which makes the maximal inner product with x_ℓ, and then to select the maximal of the resulting L inner products and the corresponding to this maximum column of Q. To maximize the inner product of the same x with columns of P, note that

$$x^T P = \left[\underbrace{\left[\sum_{\ell=1}^L x_\ell^T M^{\ell,1}\right]}_{y_1^T} D_1, \dots, \underbrace{\left[\sum_{\ell=1}^L x_\ell^T M^{\ell,L}\right]}_{y_L^T} D_L\right],$$

so that to maximize the inner product of x and the columns of P means to find, for every ℓ, the column of D_ℓ which makes the maximal inner product with y_ℓ, and then to select the maximal of the resulting L inner products and the corresponding to this maximum column of P.

We see that, if D_ℓ are simple, then we can use the approach from Sect. 3.2.2 to approximate the solution to the VI generated by F on H. Note that, in the case in question, the dual-gap function $\epsilon_{\rm VI}(\eta\big|F, H)$ admits a transparent interpretation in terms of the Nash equilibrium problem we are solving: for $\eta = [w_1; \dots; w_L] \in H$, we have

$$\begin{aligned} &\epsilon_{\rm VI}(\eta\big|F, H) \geq \epsilon_{\rm Nash}(\eta) \\ &:= \sum_{\ell=1}^L \left[\mathcal{L}_\ell(\eta) - \min_{w'_\ell \in \Delta_{N_\ell}} \mathcal{L}_\ell(w_1, \dots, w_{\ell-1}, w'_\ell, w_{\ell+1}, \dots, w_L)\right], \end{aligned} \tag{40}$$

and the right-hand side here is the sum, over the players, of the (nonnegative) incentives for a player ℓ to deviate from his strategy w_ℓ to another mixed strategy when all other players stick to their strategies as given by η. Thus, small $\epsilon_{\rm VI}([w_1; \dots; w_L]\big|\cdot, \cdot)$ means small incentives for the players to deviate from mixed strategies w_ℓ.

Verification of (40) is immediate: denoting $f_\ell = \nabla_{w_\ell}\langle g_\ell, w\rangle$, by definition of $\epsilon_{\rm VI}$ we have for every $\eta' = [w'_1; \ldots; w'_L] \in H$:

$$\begin{aligned}\epsilon_{\rm VI}(\eta\big|F, H) &\geq \langle F(\eta'), \eta - \eta'\rangle = \textstyle\sum_\ell \langle \nabla_{w_\ell}\mathcal{L}_\ell(\eta'), w_\ell - w'_\ell\rangle \\ &= \textstyle\sum_\ell \langle f_\ell, w_\ell - w'_\ell\rangle + \sum_{\ell,\ell'}\langle D_\ell^T M^{\ell\ell'} D_{\ell'} w'_{\ell'}, w_\ell - w'_\ell\rangle.\end{aligned}$$

Since $\sum_{\ell,\ell'}\langle D_\ell^T M^{\ell\ell'} D_{\ell'} z_{\ell'}, z_\ell\rangle = 0$ due to $M^{\ell\ell'} = -[M^{\ell'\ell}]^T$, one get

$$\begin{aligned}&\epsilon_{\rm VI}(\eta\big|F, H) \geq \textstyle\sum_\ell \langle f_\ell, w_\ell - w'_\ell\rangle + \sum_{\ell,\ell'}\langle D_\ell^T M^{\ell\ell'} D_{\ell'} w_{\ell'}, w_\ell - w'_\ell\rangle \\ &= \textstyle\sum_\ell \langle \nabla_{w_\ell}\mathcal{L}(\eta), w_\ell - w'_\ell\rangle = \sum_\ell [\mathcal{L}_\ell(\eta) - \mathcal{L}_\ell(w_1, \ldots, w_{\ell-1}, w'_\ell, w_{\ell+1}, \ldots, w_L)] \\ &\text{[recall that } \mathcal{L}_\ell \text{ is affine in } w_\ell\text{]}\end{aligned}$$

and (40) follows.

3.4 Relation to [1]

Here we demonstrate that the decomposition approach to solving VIs with monotone operators on LMO-represented domains cover the approach, based on Fenchel-type representations, developed in [1]. Specifically, let H be a convex and compact set in Euclidean space $\mathcal{H} = \mathbb{R}^N$, $G(\cdot)$ be a monotone vector field on H, and $\eta \mapsto A\eta + a$ be an affine mapping from $\mathcal{H}$ to Euclidean space $\mathcal{X} = \mathbb{R}^M$. Given a convex and compact set $\Xi \subset \mathcal{X}$, let us set for $\Theta = \Xi \times H$,

$$\Phi(\xi, \eta) = \Big[\Phi_\xi(\xi, \eta) := A\eta + a; \Phi_\eta(\xi, \eta) := G(\eta) - A^T\xi\Big] : \Theta \to \mathcal{X} \times \mathcal{H}, \quad (41)$$

so that Φ clearly is a monotone vector field on Θ. Assume that $\overline{\eta}(\xi) : \Xi \to H$ is a somehow selected strong solution to VI$(\Phi_\eta(\xi, \cdot), H)$, see (31):

$$\forall \xi \in \Xi : \overline{\eta}(\xi) \in H \;\&\; \underbrace{\langle G(\overline{\eta}(\xi)) - A^T\xi, \eta - \overline{\eta}(\xi)\rangle}_{=\langle \Phi_\eta(\xi, \overline{\eta}(\xi)), \eta - \overline{\eta}(\xi)\rangle} \geq 0 \quad \forall \eta \in H; \quad (42)$$

note that required $\overline{\eta}(\xi)$ definitely exists, provided that $G(\cdot)$ is continuous and monotone. Let us also define $\overline{\xi}(\eta)$ as a selection of the point-to-set mapping $\eta \mapsto \operatorname{Argmin}_{\xi\in\Xi} \langle A\eta + a, \xi\rangle$, so that, see (33),

$$\forall \eta \in H : \overline{\xi}(\eta) \in \Xi \;\&\; \underbrace{\langle A\eta + a, \xi - \overline{\xi}(\eta)\rangle}_{=\langle \Phi_\xi(\overline{\xi}(\eta), \eta), \xi - \overline{\xi}(\eta)\rangle} \geq 0, \quad \forall \xi \in \Xi. \quad (43)$$

Observe that with the just defined $\Xi, H, \Theta, \Phi, \overline{\eta}(\cdot), \overline{\xi}(\cdot)$ we are in the direct product case of the situation described in Sect. 3.1.2. Since we are in the direct product case, $(\Phi, \overline{\eta}(\cdot))$ is η-regular, and we can take, as the induced primal vector field associated with $(\Phi, \overline{\eta}(\cdot))$, the vector field

$$\Psi(\xi) = A\overline{\eta}(\xi) + a = \Phi_{\xi}(\xi, \overline{\eta}(\xi)) : \Xi \to \mathcal{X}, \tag{44}$$

and as the induced dual vector field, the field

$$\Gamma(\eta) = G(\eta) - A^T\overline{\xi}(\eta) = \Phi_{\eta}(\overline{\xi}(\eta), \eta) : H \to \mathcal{X}.$$

Note that in terms of [1], relations (44) and (42), modulo notation, form what in the reference is called a *Fenchel-type representation of vector field* $\Psi : \Xi \to \mathcal{X}$, *the data of the representation being* $\mathcal{H}$, A, a, $\overline{\eta}(\cdot)$, $G(\cdot)$, H. On a closer inspection, every such representation of a given monotone vector field $\Psi : \Xi \to \mathcal{X}$ can be obtained in this fashion from some setup of the form (41).

Assume now that Ξ is LMO-representable, and we have at our disposal G-oracle which, given on input $\eta \in H$, returns $G(\eta)$. This oracle combines with LMO for Ξ to induce a procedure which, given on input $\eta \in H$, returns $\Gamma(\eta)$. As a result, we can apply the decomposition machinery presented in Sects. 3.1, 3.2 to reduce solving VI(Ψ, Ξ) to processing VI(Γ, H) by an algorithm with accuracy certificates. It can be easily seen by inspection that this reduction recovers constructions and results presented in [1, Sections 1–4]. The bottom line is that the developed in Sects. 3.1, 3.2 decomposition-based approach to solving VIs with monotone operators on LMO-represented domains essentially covers the developed in [1] approach based on Fenchel-type representations of monotone vector fields.[7]

Acknowledgements A. Juditsky was supported by the CNRS-Mastodons project Titan, and the LabEx PERSYVAL-Lab (ANR-11-LABX-0025). Research of A. Nemirovski was supported by the NSF Grants CMMI-1232623, CCF-1415498, CMMI-1262063.

Appendix

Proof of Lemma 2.1 It suffices to prove the ϕ-related statements. Lipschitz continuity of ϕ in the direct product case is evident. Furthermore, the function $\theta(x_1, x_2; y_1) = \max_{y_2 \in Y_2[y_1]} \Phi(x_1, x_2; y_1, y_2)$ is convex and Lipschitz continuous in $x = [x_1; x_2] \in X$ for every $y_1 \in Y_1$, whence

$$\phi(x_1, y_1) = \min_{x_2 \in X_2[x_1]} \theta(x_1, x_2; y_1)$$

is convex and lower semicontinuous in $x_1 \in X_1$ (note that X is compact). On the other hand,

$$\begin{aligned}\phi(x_1, y_1) &= \max_{y_2 \in Y_2[y_1]} \min_{x_2 \in X_2[x_1]} \Phi(x_1, x_2; y_1, y_2) \\ &= \max_{y_2 \in Y_2[y_1]} \left[\chi(x_1; y_1, y_2) := \min_{x_2 \in X_2[x_1]} \Phi(x_1, x_2; y_1, y_2)\right],\end{aligned}$$

[7] "covers" instead of "is equivalent" stems from the fact that the scope of decomposition is not restricted to the setups of the form of (41).

so that $\chi(x_1; y_1, y_2)$ is concave and Lipschitz continuous in $y = [y_1; y_2] \in Y$ for every $x_1 \in X_1$, whence

$$\phi(x_1, y_1) = \max_{y_2 \in Y_2[y_1]} \chi(x_1; y_1, y_2)$$

is concave and upper semicontinuous in $y_1 \in Y_1$ (note that Y is compact).

Next, we have

$$\begin{aligned}
\mathrm{SadVal}(\phi, X_1, X_2) &= \inf_{x_1 \in X_1} \left[\sup_{y_1 \in Y_1} \left[\sup_{y_2:[y_1;y_2]\in Y} \inf_{x_2:[x_1;x_2]\in X} \Phi(x_1, x_2; y_1, y_2) \right] \right] \\
&= \inf_{x_1 \in X_1} \left[\sup_{[y_1;y_2]\in Y} \inf_{x_2:[x_1;x_2]\in X} \Phi(x_1, x_2; y_1, y_2) \right] \\
&= \inf_{x_1 \in X_1} \left[\inf_{x_2:[x_1;x_2]\in X} \sup_{[y_1;y_2]\in Y} \Phi(x_1, x_2; y_1, y_2) \right] \text{ [by Sion-Kakutani Theorem [19]]} \\
&= \inf_{[x_1;x_2]\in X} \sup_{[y_1;y_2]\in Y} \Phi(x_1, x_2; y_1, y_2) = \mathrm{SadVal}(\Phi, X, Y),
\end{aligned}$$

as required in (2). Finally, let $\bar{x} = [\bar{x}_1; \bar{x}_2] \in X$ and $\bar{y} = [\bar{y}_1; \bar{y}_2] \in Y$. We have

$$\begin{aligned}
\overline{\phi}(\bar{x}_1) - \mathrm{SadVal}(\phi, X_1, Y_1) &= \overline{\phi}(\bar{x}_1) - \mathrm{SadVal}(\Phi, X, Y) \text{ [by (2)]} \\
&= \sup_{y_1 \in Y_1} \phi(\bar{x}_1, y_1) - \mathrm{SadVal}(\Phi, X, Y) \\
&= \sup_{y_1 \in Y_1} \sup_{y_2:[y_1;y_2]\in Y} \inf_{x_2:[\bar{x}_1;x_2]\in X} \Phi(\bar{x}_1, x_2; y_1, y_2) \\
&\quad - \mathrm{SadVal}(\Phi, X, Y) \\
&= \sup_{[y_1;y_2]\in Y} \inf_{x_2:[\bar{x}_1;x_2]\in X} \Phi(\bar{x}_1, x_2; y_1, y_2) - \mathrm{SadVal}(\Phi, X, Y) \\
&= \inf_{x_2:[\bar{x}_1;x_2]\in X} \sup_{y=[y_1;y_2]\in Y} \Phi(\bar{x}_1, x_2; y) - \mathrm{SadVal}(\Phi, X, Y) \\
&\le \sup_{y=[y_1;y_2]\in Y} \Phi(\bar{x}_1, \bar{x}_2; y) - \mathrm{SadVal}(\Phi, X, Y) \\
&= \overline{\Phi}(\bar{x}) - \mathrm{SadVal}(\Phi, X, Y)
\end{aligned}$$

and

$$\begin{aligned}
\mathrm{SadVal}(\phi, X_1, Y_1) - \underline{\phi}(\bar{y}_1) &= \mathrm{SadVal}(\Phi, X, Y) - \underline{\phi}(\bar{y}_1) \text{ [by (2)]} \\
&= \mathrm{SadVal}(\Phi, X, Y) - \inf_{x_1 \in X_1} \phi(x_1, \bar{y}_1) \\
&= \mathrm{SadVal}(\Phi, X, Y) \\
&\quad - \inf_{x_1 \in X_1} \left[\inf_{x_2:[x_1;x_2]\in X} \sup_{y_2:[\bar{y}_1;y_2]\in Y} \Phi(x_1, x_2; \bar{y}_1, y_2) \right] \\
&= \mathrm{SadVal}(\Phi, X, Y) - \inf_{x=[x_1;x_2]\in X} \sup_{y_2:[\bar{y}_1;y_2]\in Y} \Phi(x; \bar{y}_1, y_2) \\
&\le \mathrm{SadVal}(\Phi, X, Y) - \inf_{x=[x_1;x_2]\in X} \Phi(x; \bar{y}_1, \bar{y}_2) \\
&= \mathrm{SadVal}(\Phi, X, Y) - \underline{\Phi}(\bar{y}).
\end{aligned}$$

We conclude that

$$\begin{aligned}\epsilon_{\rm sad}([\bar{x}_1;\bar{y}_1]\big|\phi,X_1,Y_1) &= \left[\overline{\phi}(\bar{x}_1)-{\rm SadVal}(\phi,X_1,Y_1)\right]\\ &\quad+\left[{\rm SadVal}(\phi,X_1,Y_1)-\underline{\phi}(\bar{y}_1)\right]\\ &\le \left[\overline{\Phi}(\bar{x})-{\rm SadVal}(\Phi,X,Y)\right]+\left[{\rm SadVal}(\Phi,X,Y)-\underline{\Phi}(\bar{y})\right]=\epsilon_{\rm sad}([\bar{x};\bar{y}]\big|\Phi,X,Y),\end{aligned}$$

as claimed in (3). □

Proof of Lemma 2.2 For $x_1\in X_1$, we have

$$\begin{aligned}\phi(x_1;\bar{y}_1) &= \min_{x_2:[x_1;x_2]\in X}\max_{y_2:[\bar{y}_1;y_2]\in Y}\Phi(x_1,x_2;\bar{y}_1,y_2)\ge \min_{x_2:[x_1;x_2]\in X}\Phi(x_1,x_2;\bar{y}_1,\bar{y}_2)\\ &\ge \min_{x_2:[x_1;x_2]\in X}\big[\underbrace{\Phi(\bar{x};\bar{y})}_{\phi(\bar{x}_1;\bar{y}_1))}+\langle G,[x_1;x_2]-[\bar{x}_1;\bar{x}_2]\big]\rangle\\ &\quad[\text{since }\Phi(x;\bar{y})\text{ is convex and }G\in\partial_x\Phi(\bar{x};\bar{y})]\\ &\ge \phi(\bar{x}_1;\bar{y}_1)+\langle g,x_1-\bar{x}_1\rangle\ [\text{by definition of } g,G],\end{aligned}$$

as claimed in (a). "Symmetric" reasoning justifies (b). □

Proof of Lemma 2.3 Assume that (5) holds true. Then, G clearly is certifying, implying that

$$\chi_G(\bar{x}_1)=\langle G,[\bar{x}_1;\bar{x}_2]\rangle,$$

and therefore (5) reads

$$\langle G,[x_1;x_2]\rangle\ge\chi_G(\bar{x}_1)+\langle g,x_1-\bar{x}_1\rangle\quad\forall x=[x_1;x_2]\in X,$$

where taking minimum in the left-hand side over $x_2\in X_2[x_1]$,

$$\chi_G(x_1)\ge\chi_G(\bar{x}_1)+\langle g,x_1-\bar{x}_1\rangle\quad\forall x_1\in X_1,$$

as claimed in (ii).

Now assume that (i) and (ii) hold true. By (i), $\chi_G(\bar{x}_1)=\langle G,[\bar{x}_1;\bar{x}_2]\rangle$, and by (ii) combined with the definition of χ_G,

$$\begin{aligned}\forall x=[x_1;x_2]\in X:\langle G,[x_1;x_2]\rangle &\ge \chi_G(x_1)\ge\chi_G(\bar{x}_1)+\langle g,x_1-\bar{x}_1\rangle\\ &=\langle G,\bar{x}\rangle+\langle g,x_1-\bar{x}_1\rangle,\end{aligned}$$

implying (5). □

Dynamic Programming-Generated Simple Matrices

Consider the situation as follows. There exists an evolving in time system $\mathcal{S}$, with state ξ_s at time $s=1,2,\ldots,m$ belonging to a given finite nonempty set Ξ_s. Furthermore,

every pair (ξ, s) with $s \in \{1, \ldots, m\}$, $\xi \in \Xi_s$ is associated with nonempty finite *set of actions* A^s_ξ, and we set

$$\mathcal{S}_s = \{(\xi, a) : \xi \in \Xi_s, a \in A^s_\xi\}.$$

Furthermore, for every s, $1 \le s < m$, a *transition mapping* $\pi_s(\xi, a) : \mathcal{S}_s \to \Xi_{s+1}$ is given. Finally, we are given vector-valued functions ("outputs") $\chi_s : \mathcal{S}_s \to \mathbb{R}^{r_s}$.

A *trajectory* of $\mathcal{S}$ is a sequence $\{(\xi_s, a_s) : 1 \le s \le m\}$ such that $(\xi_s, a_s) \in \mathcal{S}_s$ for $1 \le s \le m$ and

$$\xi_{s+1} = \pi_s(\xi_s, a_s),\ 1 \le s < m.$$

The *output* of a trajectory $\tau = \{(\xi_s, a_s) : 1 \le s \le m\}$ is the block vector

$$\chi[\tau] = [\chi_1(\xi_1, a_1); \ldots; \chi_m(\xi_m, a_m)].$$

We can associate with $\mathcal{S}$ the matrix $D = D[\mathcal{S}]$ with $K = r_1 + \cdots + r_m$ rows and with columns indexed by the trajectories of $\mathcal{S}$; specifically, the column indexed by a trajectory τ is $\chi[\tau]$.

For example, knapsack-generated matrix D associated with knapsack data from Sect. 2.6.2 is of the form $D[\mathcal{S}]$ with system $\mathcal{S}$ as follows:

- Ξ_s, $s = 1, \ldots, m$, is the set of nonnegative integers which are $\le H$;
- A^s_ξ is the set of nonnegative integers a such that $a \le \bar{p}_s$ and $\xi - h_s p_s \ge 0$;
- the transition mappings are $\pi_s(\xi, a) = \xi - a h_s$;
- the outputs are $\chi_s(\xi, a) = f_s(a)$, $1 \le s \le m$.

In the notation of Sect. 2.6.2, vectors $[p_1; \ldots; p_m] \in \mathcal{P}$ are exactly the sequences of actions $a_1, \ldots, a_m$ stemming from the trajectories of the just defined system $\mathcal{S}$.

Observe that matrix $D = D[\mathcal{S}]$ is simple, provided the cardinalities of Ξ_s and A^s_ξ are reasonable. Indeed, given $x = [x_1; \ldots; x_m] \in \mathbb{R}^n = \mathbb{R}^{r_1} \times \cdots \times \mathbb{R}^{r_m}$, we can identify $\overline{D}[x]$ by dynamic programming, running first the backward Bellman recurrence

$$\left.\begin{array}{l} U_s(\xi) = \max\limits_{a \in A^s_\xi} \left\{x_s^T \chi_s(\xi, a) + U_{s+1}(\pi_s(\xi, a))\right\} \\ A_s(\xi) = \operatorname{Argmax}\limits_{a \in A^s_\xi} \left\{x_s^T \chi_s(\xi, a) + U_{s+1}(\pi_s(\xi, a))\right\} \end{array}\right\}, \xi \in \Xi_s,\ s = m, m-1, \ldots, 1$$

(where $U_{m+1}(\cdot) \equiv 0$), and then recovering the (trajectory indexing the) column of D corresponding to $\overline{D}[x]$ by running the forward Bellman recurrence

$$\begin{array}{r} \xi_1 \in \operatorname{Argmax}_{\xi \in \Xi_1} U_1(\xi) \Rightarrow a_1 \in A_1(\xi_1) \Rightarrow \cdots \\ \Rightarrow \xi_{s+1} = \pi_s(\xi_s, a_s) \Rightarrow a_{s+1} \in A_{s+1}(\xi_{s+1}) \Rightarrow \cdots \end{array}, s = 1, 2, \ldots, m-1.$$

Attacker Versus Defender Via Ellipsoid Algorithm

In our implementation,

1. Relation (39) is ensured by specifying U, V as centered at the origin Euclidean balls of radius R, where R is an upper bound on the Euclidean norms of the columns in D and in A (such a bound can be easily obtained from the knapsack data specifying the matrices D, A).
2. We process the monotone vector field associated with the primal SP problem (30), that is, the field

$$F(u,v) = [F_u(u,v) = \overline{A}[u] - v;\ F_v(u,v) = u - \underline{D}[v]]$$

by ellipsoid algorithm with accuracy certificates from [20]. For $\tau = 1, 2, \ldots$, the algorithm generates *search points* $[u_\tau; v_\tau] \in \mathbb{R}^K \times \mathbb{R}^K$, with $[u_1; v_1] = 0$, along with execution protocols $\mathcal{I}^\tau = \{[u_i; v_i], F(u_i, v_i) : i \in I_\tau\}$, where $I_\tau = \{i \le \tau : [u_i; v_i] \in U \times V\}$, augmented by accuracy certificates $\lambda^\tau = \{\lambda_i^\tau \ge 0 : i \in I_\tau\}$ such that $\sum_{i \in I_\tau} \lambda_i^\tau = 1$. From the results of [20], it follows that for every $\epsilon > 0$ it holds

$$\tau \ge N(\epsilon) := O(1)K^2 \ln\left(2\frac{R+\epsilon}{\epsilon}\right) \Rightarrow \mathrm{Res}(\mathcal{I}^\tau, \lambda^\tau \big| U \times V) \le \epsilon. \tag{45}$$

3. When computing $F(u_i, v_i)$ (this computation takes place only at *productive* steps—those with $[u_i; v_i] \in U \times V$), we get, as a by-product, the columns $A^i = \overline{A}[u_i]$ and $D^i = \underline{D}[v_i]$ of matrices A, D, along with the indexes a^i, d^i of these columns (recall that these indexes are pure strategies of attacker and defender and thus, according to the construction of A, D, are collections of m nonnegative integers). In our implementation, we stored these columns, same as their indexes and the corresponding search points $[u_i; v_i]$. As is immediately seen, in the case in question the approximate solution $[w^\tau; z^\tau]$ to the SP problem of interest (27) induced by execution protocol $\mathcal{I}^\tau$ and accuracy certificate λ^τ is comprised of two sparse vectors

$$w^\tau = \sum_{i \in I_\tau} \lambda_i^\tau \delta^D_{d^i},\ \ z^\tau = \sum_{i \in I_\tau} \lambda_i^\tau \delta^A_{a^i}, \tag{46}$$

where δ^D_d is the "dth basic orth" in the simplex Δ_N of probabilistic vectors with entries indexed by pure strategies of defender, and similarly for δ^A_a. Thus, we have no difficulties with representing our approximate solutions,[8] in spite of their huge ambient dimension.

According to our general theory and (45), the number of steps needed to get an ϵ-solution $[w; z]$ to the problem of interest (i.e., a feasible solution with $\epsilon_{\mathrm{sad}}([w; z] \big| \psi, W, Z) \le \epsilon$) does not exceed $N(\epsilon)$, with computational effort per step dominated by the necessity to identify $\overline{A}[u_i]$, $\underline{D}[v_i]$ by dynamic programming.

[8] Note that applying Carathéodory theorem, we could further "compress" the representations of approximate solutions—make these solutions convex combinations of at most $K+1$ of $\delta^D_{d^i}$s and $\delta^A_{a^i}$s.

In fact, we used the outlined scheme with two straightforward modifications.

– First, instead of building the accuracy certificates λ^τ according to the rules from [20], we used the best, given execution protocols $\mathcal{I}^\tau$, accuracy certificates by solving the convex program

$$\min_{\lambda}\left\{\mathrm{Res}(\mathcal{I}^\tau,\lambda\big|U\times V):=\max_{y\in U\times V}\sum_{i\in I_\tau}\lambda_i\langle F(u_i,v_i),[u_i;v_i]-y\rangle:\lambda_i\geq 0,\sum_{i\in I_\tau}\lambda_i=1\right\}. \tag{47}$$

In our implementation, this problem was solved once per $4K^2$ steps. Note that with U, V being Euclidean balls, (47) is a Conic Quadratic Problem and may be solved using, e.g., CVX [27].

– Second, given current approximate solution (46) to the problem of interest, we can compute its saddle point inaccuracy exactly instead of upper-bounding it by $\mathrm{Res}(\mathcal{I}^\tau,\lambda^\tau\big|U\times V)$. Indeed, it is immediately seen that

$$\epsilon_{\mathrm{sad}}([w^\tau;z^\tau]\big|\psi,W,Z)=\mathrm{Max}\left(A^T\left[\sum_{i\in I_\tau}\lambda_i^\tau D^i\right]\right)-\mathrm{Min}\left(D^T\left[\sum_{i\in I_\tau}\lambda_i^\tau A^i\right]\right).$$

In our implementation, we performed this computation each time when a new accuracy certificate was computed, and terminated the solution process when the saddle point inaccuracy became less than a given threshold (1.e-4).

Proof of Proposition 3.2 (i): Let $\xi_1,\xi_2\in\Xi$, and let $\eta_1=\overline{\eta}(\xi_1)$, $\eta_2=\overline{\eta}(\xi_2)$. By (32), we have

$$\begin{aligned}\langle\Psi(\xi_2),\xi_2-\xi_1\rangle&\geq\langle\Phi(\xi_2,\eta_2),[\xi_2-\xi_1;\eta_2-\eta_1]\rangle,\\ \langle\Psi(\xi_1),\xi_1-\xi_2\rangle&\geq\langle\Phi(\xi_1,\eta_1),[\xi_1-\xi_2;\eta_1-\eta_2]\rangle.\end{aligned}$$

Summing inequalities up, we get

$$\langle\Psi(\xi_2)-\Psi(\xi_1),\xi_2-\xi_1\rangle\geq\langle\Phi(\xi_2,\eta_2)-\Phi(\xi_1,\eta_1),[\xi_2-\xi_1;\eta_2-\eta_1]\rangle\geq 0,$$

so that Ψ is monotone.

Furthermore, the first inequality in (35) is due to Proposition 3.1. To prove the second inequality in (35), let $\mathcal{I}_t=\{\xi_i\in\Xi,\Psi(\xi_i):1\leq i\leq t\}$, $\mathcal{J}_t=\{\theta_i:=[\xi_i;\overline{\eta}(\xi_i)],\Phi(\theta_i):1\leq i\leq t\}$, and let λ be t-step accuracy certificate. We have

$$\begin{aligned}&\theta=[\xi;\eta]\in\Theta\Rightarrow\\ &\textstyle\sum_{i=1}^t\lambda_i\langle\Phi(\theta_i),\theta_i-\theta\rangle\leq\sum_{i=1}^t\lambda_i\langle\Psi(\xi_i),\xi_i-\xi\rangle\ \text{[see (32)]}\\ &\quad\leq\mathrm{Res}(\mathcal{I}_t,\lambda\big|\Xi)\\ &\quad\Rightarrow\mathrm{Res}(\mathcal{J}_t,\lambda\big|\Theta)=\textstyle\sup_{\theta=[\xi;\eta]\in\Theta}\sum_{i=1}^t\lambda_i\langle\Phi(\theta_i),\theta_i-\theta\rangle\leq\mathrm{Res}(\mathcal{I}_t,\lambda\big|\Xi).\end{aligned}$$

(i) is proved.

(ii): Let $\eta \in H$. Invoking (34), we have

$$\langle \Gamma(\eta), \widehat{\eta} - \eta \rangle \leq \langle \Phi(\overline{\xi}(\eta), \eta), [\widehat{\xi}; \widehat{\eta}] - [\overline{\xi}(\eta); \eta] \rangle \leq \epsilon_{\mathrm{VI}}(\widehat{\theta} \big| \Phi, \Theta),$$

and (36) follows. □

References

1. Juditsky, A., Nemirovski, A.: Solving variational inequalities with monotone operators on domains given by linear minimization oracles. Math. Program. **152**(1), 1–36 (2013)
2. Harchaoui, Z., Juditsky, A., Nemirovski, A.: Conditional gradient algorithms for norm-regularized smooth convex optimization. Math. Program. **152**(1), 75–112 (2014)
3. Ziegler, G.M.: Lectures on Polytopes, vol. 152. Springer, Berlin (1995)
4. Frank, M., Wolfe, P.: An algorithm for quadratic programming. Naval Res. Logist. Q. **3**(1–2), 95–110 (1956)
5. Demyanov, V., Rubinov, A.: Approximate Methods in optimization problems, vol. 32. Elsevier, Amsterdam (1970)
6. Dunn, J.C., Harshbarger, S.: Conditional gradient algorithms with open loop step size rules. J. Math. Anal. Appl. **62**(2), 432–444 (1978)
7. Freund, R.M., Grigas, P.: New analysis and results for the Frank–Wolfe method. Math. Program. **155**(1–2), 199–230 (2016)
8. Garber, D., Hazan, E.: Faster Rates for the Frank–Wolfe Method Over Strongly-convex Sets. arXiv preprint arXiv:1406.1305 (2014)
9. Harchaoui, Z., Douze, M., Paulin, M., Dudik, M., Malick, J.: Large-scale image classification with trace-norm regularization. In: 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 3386–3393. IEEE (2012)
10. Jaggi, M.: Revisiting Frank–Wolfe: projection-free sparse convex optimization. In: Proceedings of the 30th International Conference on Machine Learning (ICML-13), pp. 427–435 (2013)
11. Jaggi, M., Sulovsk, M., et al.: A simple algorithm for nuclear norm regularized problems. In: Proceedings of the 27th International Conference on Machine Learning (ICML-10), pp. 471–478 (2010)
12. Pshenichny, B.N., Danilin, Y.M.: Numerical Methods in Extremal Problems. Mir Moscow (1978)
13. Argyriou, A., Signoretto, M., Suykens, J.A.K.: Hybrid algorithms with applications to sparse and low rank regularization, chap. 3. In: Suykens, J.A.K., Signoretto,M., Argyriou, A., (eds.) Regularization, Optimization, Kernels, and Support Vector Machines, pp. 53–82. Chapman & Hall/CRC (2014)
14. Pierucci, F., Harchaoui, Z., Malick, J.: A Smoothing Approach for Composite Conditional Gradient with Nonsmooth Loss. Tech. rep., Inria (2014). https://hal.inria.fr/hal-01096630/
15. Tewari, A., Ravikumar, P.K., Dhillon, I.S.: Greedy algorithms for structurally constrained high dimensional problems. In: Advances in Neural Information Processing Systems, pp. 882–890 (2011)
16. Ying, Y., Li, P.: Distance metric learning with eigenvalue optimization. J. Mach. Learn. Res. **13**(1), 1–26 (2012)
17. Cox, B., Juditsky, A., Nemirovski, A.: Dual subgradient algorithms for large-scale nonsmooth learning problems. Math. Program. **148**(1–2), 143–180 (2014)
18. Lan, G., Zhou, Y.: Conditional Gradient Sliding for Convex Optimization (2014). http://www.ise.ufl.edu/glan/files/2015/09/CGS08-31
19. Hiriart-Urruty, J.B., Lemaréchal, C.: Convex Analysis and Minimization Algorithms I: Fundamentals, vol. 305. Springer, Berlin (2013)
20. Nemirovski, A., Onn, S., Rothblum, U.G.: Accuracy certificates for computational problems with convex structure. Math. Oper. Res. **35**(1), 52–78 (2010)
21. Cox, B.: Applications of Accuracy Certificates for Problems with Convex Structure. Ph.D. thesis, Georgia Institute of Technology (2011). https://smartech.gatech.edu/jspui/bitstream/1853/39489/1/cox_bruce_a_201105_phd
22. Gol'stein, E.: Direct-dual block method of linear programming. Autom. Remote Control **57**(11), 1531–1536 (1996)
23. Gol'stein, E., Sokolov, N.: A decomposition algorithm for solving multicommodity production-and-transportation problem. Ekonomika i Matematicheskie Metody **33**(1), 112–128 (1997)

24. Dvurechensky, P., Nesterov, Y., Spokoiny, V.: Primal-dual methods for solving infinite-dimensional games. J. Optim. Theory Appl. **166**(1), 23–51 (2015)
25. Bellman, R.: On "Colonel Blotto" and analogous games. SIAM Rev. **11**(1), 66–68 (1969)
26. Robertson, B.: The Colonel Blotto game. Econ. Theory **29**(1), 1–24 (2006)
27. Grant, M., Boyd, S.: Cvx: Matlab Software for Disciplined Convex Programming, version 2.1 (2015). http://cvxr.com/cvx