MATHEMATICS OF OPERATIONS RESEARCH
Vol. 35, No. 1, February 2010, pp. 52–78
ISSN 0364-765X | EISSN 1526-5471 | 10 | 3501 | 0052

DOI 10.1287/moor.1090.0427

Accuracy Certificates for Computational Problems with Convex Structure

Arkadi Nemirovski
School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332,
nemirovs@isye.gatech.edu, http://isye.gatech.edu/~nemirovs

Shmuel Onn, Uriel G. Rothblum
Faculty of Industrial Engineering and Management, Technion—Israel Institute of Technology, Technion City, Haifa 32000, Israel
{onn@ie.technion.ac.il, http://ie.technion.ac.il/onn.html; rothblum@ie.technion.ac.il, http://ie.technion.ac.il/rothblum.html}

The goal of this paper is to introduce the notion of certificates, which verify the accuracy of solutions of computational problems with convex structure. Such problems include minimizing convex functions, variational inequalities with monotone operators, computing saddle points of convex-concave functions, and solving convex Nash equilibrium problems. We demonstrate how the implementation of the ellipsoid method and other cutting plane algorithms can be augmented with the computation of such certificates without significant increase of the computational effort. Further, we show that (computable) certificates exist for any algorithm that is capable of producing solutions of guaranteed accuracy.

Key words: convexity; certificates; computation in convex structures; convex minimization; variational inequalities; saddle points; convex Nash equilibrium
MSC2000 subject classification: Primary: 52B55; secondary: 52A41, 52A27, 52A20
OR/MS subject classification: Primary: Analysis of algorithms; secondary: computational complexity
History: Received April 3, 2007; revised November 26, 2008, and September 3, 2009. Published online in *Articles in Advance* December 29, 2009.

1. Introduction.

To motivate the goals of this paper, let us start with a convex minimization problem (CMP) in the form of

$$\mathrm{Opt} = \min_{x \in X} F(x), \tag{1}$$

where X is a solid (compact convex set with a nonempty interior) in $\mathcal{R}^n$ and $F\colon X \to \mathcal{R} \cup \{+\infty\}$ is a closed convex function that is finite valued on $\operatorname{int} X$ ($\operatorname{int} X$ standing for the interior of X). Assume that the problem is *black box represented*, that is:

- X is given by a separation oracle, a routine that, given on input $x \in \mathcal{R}^n$ reports whether $x \in \operatorname{int} X$ and, if that is not the case, returns a *separator*—a vector e such that $\langle e, y\rangle < \langle e, x\rangle$ whenever $y \in \operatorname{int} X$. In addition, we are given a solid $B \subseteq \mathcal{R}^n$ that contains X and is "simple" in the sense that minimizing a linear form over $\mathbf{B}$ is "easy" (as is the case when $\mathbf{B}$ is an ellipsoid, a parallelotope, or a simplex);
- F is given by a first-order oracle, a routine that, given on input $x \in \operatorname{int} X$, returns the value $F(x)$ and a subgradient $F'(x)$ of F at x.

Our goal is to solve (1) within a given accuracy $\epsilon > 0$, specifically, to find $\hat{x}$ such that

$$\hat{x} \in X \qquad \text{and} \qquad \epsilon_{\mathrm{opt}}(\hat{x}) := F(\hat{x}) - \mathrm{Opt} \le \epsilon \tag{2}$$

(reasons for quantifying the accuracy in terms of F and not in terms of "the argument"—by the distance from $\hat{x}$ to the set of minimizers of F on X—will be explained later). In the outlined "computational environment," an algorithm for achieving our goal is a routine that, given on input $\epsilon > 0$, executes finitely many calls to the oracles and then terminates and outputs a vector $\hat{x} \in X$ satisfying (2). This should be so for *every* $\epsilon > 0$ and *every* CMP (1) from a family $\mathcal{F}$ of CMPs that is given in advance. Further, the search points (for deterministic algorithms) or their distributions (for randomized ones) should be uniquely defined by the answers of the oracles at the preceding steps and similarly for (the approximate solution) $\hat{x}$ built upon termination. The latter property will be referred to as "nonanticipative."

The issue of primary interest in this paper is how an algorithm can *certify* the target relation (2); here is the proposed answer. Assume that when solving CMP (1), the algorithm has queried the oracles at the points $x^1, \ldots, x^\tau$, where τ is the number of steps before termination. Part of the information acquired by the algorithm in this run forms the *execution protocol* $P_\tau = \{I_\tau, J_\tau, \{x^t, e_t\}_{t=1}^\tau\}$, where I_τ is the set of indices t of *productive* steps—those with $x^t \in \operatorname{int} X$, $J_\tau = \{1, \ldots, \tau\} \setminus I_\tau$ is the set of indices of *nonproductive* steps and e_t is either a subgradient of F at x^t reported by the first-order oracle (this is so when the step t is productive) or e_t is the

separator of x^t and $\operatorname{int} X$ reported by the separation oracle (this is so when the step t is nonproductive). The complete information acquired in the run is obtained by augmenting the execution protocol by the values $F(x^t)$ of F at the productive steps. Now, assume the run in question is *productive*, meaning that $I_\tau \neq \varnothing$, and let us assign steps t, $1 \le t \le \tau$ with nonnegative weights ξ_t such that $\sum_{t\in I_\tau} \xi_t = 1$. Consider the quantity

$$F_*(\xi \mid P_\tau, \mathbf{B}) = \sum_{t\in I_\tau} \xi_t F(x^t) - \overbrace{\max_{x\in\mathbf{B}} \sum_{t=1}^{\tau} \xi_t \langle e_t, x^t - x\rangle}^{\epsilon_{\mathrm{cert}}(\xi \mid P_\tau, \mathbf{B})}. \tag{3}$$

An immediate observation is that $F_*(\xi \mid P_\tau, \mathbf{B})$ *is a lower bound on* Opt and, therefore, *the relation*

$$F(\hat{x}) - F_*(\xi \mid P_\tau, \mathbf{B}) \le \epsilon \tag{4}$$

is a sufficient condition for (2)*, provided* $\hat{x}$ *is feasible.*

The above observation is indeed immediate. Given $x \in X$, we have $F(x) \ge F(x^t) + \langle e_t, x - x^t\rangle$ whenever $t \in I_\tau$ (by convexity of F) and $0 \ge \langle e_t, x - x^t\rangle$ whenever $t \in J_\tau$. Using the ξ_t's as weights, the weighted sum of the above inequalities yields $F(x) \ge \sum_{t\in I_\tau} \xi_t F(x^t) - \sum_{t=1}^{\tau} \xi_t \langle e_t, x^t - x\rangle$. Taking minima over $x \in X$, we get $\mathrm{Opt} \ge \sum_{t\in I_\tau} \xi_t F(x^t) - \sup_{x\in X} \sum_{t=1}^{\tau} \xi_t \langle e_t, x^t - x\rangle$. The right-hand side in this inequality is $\ge F_*(\xi \mid P_\tau, \mathbf{B})$ due to $X \subset \mathbf{B}$. So, $F_*(\xi \mid P_\tau, \mathbf{B})$ is, indeed, a lower bound on Opt.

Consider a τ-step execution protocol P_τ, augmented by the values of F at the points x^t, $t \in I_\tau$, and a vector $\xi \in \mathcal{R}^\tau$. It is clearly easy to verify whether ξ is as stated above (i.e., nonnegative with $\sum_{t\in I_\tau} \xi_t = 1$). Further, if the answer is positive, $F_*(\xi \mid P_\tau, \mathbf{B})$ is easy to compute.[1] It follows that *a certificate* ξ *associated with a* τ*-step execution protocol* P_τ, that is, a collection of τ nonnegative weights ξ_t with $\sum_{t\in I_\tau} \xi_t = 1$, can be considered as a "simple proof" (as is expected of a certificate) of an inaccuracy bound, specifically, of the bound $\epsilon_{\mathrm{opt}}(\hat{x}) \le F(\hat{x}) - F_*(\xi \mid P_\tau, \mathbf{B})$.

A close inspection (see Proposition 2.2) shows that for some important families of CMPs, such as the family $\mathcal{F}_{\mathrm{c}}(\mathbf{B})$ of *all* problems (1) with feasible domains $X \subset \mathbf{B}$ that are solids as well as convex objective functions F that are finite and continuous on X, the outlined way to certify the accuracy guarantees is *a necessity rather than an option*. In other words, an algorithm for solving CMPs from $\mathcal{F}_{\mathrm{c}}(\mathbf{B})$ that is "intelligent enough" to solve every problem from the family within a given accuracy $\epsilon > 0$ always acquires upon termination enough information to equip the resulting approximate solution $\hat{x}$ with an accuracy certificate ξ satisfying (4). Some of the existing "intelligent" algorithms (for example, different variants of subgradient and bundle methods; see, e.g., Lemarechal et al. [13], Ben-Tal and Nemirovski [3], and references therein) produce such certificates explicitly. Other black-box-oriented intelligent algorithms such as the ellipsoid method and other polynomial-time versions of the cutting plane scheme provide accuracy guarantees without building the certificates explicitly. Finally, the most advanced (from the applied perspective) convex optimization techniques such as polynomial-time interior point methods (IPMs) for linear, conic quadratic, and semidefinite programming do produce explicit accuracy certificates, though of a type different from the one just explained. Specifically, IPMs work with "well-structured" convex programs rather than with black box-represented ones, and they heavily exploit the problem's structure (typically, a primal-dual pair of conic problems on a simple cone) to build in parallel approximate solutions and "structure-specific" accuracy certificates (for example, in the conic case, by generating primal-dual feasible pairs of solutions to a primal-dual pair of conic problems, with the dual feasible solution in the role of certificate and the duality gap in the role of the certified accuracy). The primary goal of this paper is to equip the algorithms from the second of the above three groups—the polynomial-time cutting plane algorithms for black box-represented CMPs—with computationally cheap rules for online building of explicit accuracy certificates. A side product of our analysis is a justification of the standard polynomial-time complexity bounds of the algorithms.

Example. Given $R < \infty$, consider the family $\mathcal{F}$ of all problems (1) with solids X contained in $\mathbf{B} = \{x \in \mathcal{R}^n : \|x\|_2 \le R\}$ and convex and bounded on X objectives F. It is known (Nemirovski and Yudin [18]) that the ellipsoid method is capable to solve a problem from this family within (any) given accuracy $\epsilon > 0$ in no more than $N(\epsilon, r(X), V(F)) = 4n^2 \ln(R/r(X) + V(F)/\epsilon)$ calls to separation and first-order oracles, with every call accompanied by just $O(n^2)$ operations to process said oracle's answer; here, $r(X)$ is the largest of the radii of Euclidean balls contained in X, and $V(F) = \max_X F - \min_X F$ is the variation of the objective on the feasible

[1] This is because the only potentially nontrivial part of the computation, that is, the minimization over $x \in \mathbf{B}$ in (3), is trivial due to the simplicity of $\mathbf{B}$.

domain of the problem. The method, however, does not produce accuracy certificates explicitly; as a result, with the traditional implementation of the ellipsoid method and in the absence of a priori upper bounds on $1/r(X)$ and $V(F)$, we at no point in time can conclude that a solution of the required accuracy ϵ is already built. Thus, we cannot equip the method with termination rules that guarantee finite termination with an approximate solution of the required accuracy on every problem from the family. A byproduct from the main result of this paper is that the ellipsoid method can be equipped with online rules for building accuracy certificates that do certify that accuracy ϵ is achieved after at most $O(1)N(\epsilon, r(X), V(F))$ steps, thus eliminating the severe theoretical drawback of the traditional implementation of the method. It should be added that the accuracy certificates turn out to be computationally cheap, e.g., we can keep the per-step computational effort in the new algorithm within, say, factor 1.1 in the per-step effort of the usual implementation.

After outlining, albeit in a yet-incomplete form, the goals of our paper, we find it prudent to address two easily predictable objections that these goals might cause. First, a "practically oriented" reader could say that we are looking for a remedy for an illness that, by itself, is of no practical interest because we address a very restricted group of convex minimization algorithms with seemingly minor applied importance. Indeed, convex problems arising in applications always have a lot of structure, and it does not make much sense to treat them as black box-oriented ones. Moreover, there typically is enough structure to make the problem amenable for powerful polynomial-time interior point methods, the techniques of which outperform dramatically, especially in the large-scale case, the black box-oriented cutting plane schemes we are interested in. Thus, who cares? Our answer here is twofold. First, though the direct applied potential of polynomial-time cutting plane algorithms is indeed minor, these algorithms are definitely of a major theoretical importance because these are the algorithms underlying the most general theoretical results on polynomial-time solvability of convex programming problems. As a matter of fact, the inability of the known polynomial-time cutting plane algorithms to produce explicit accuracy certificates restricts significantly the "theoretical power" of these algorithms, so that our goal seems to be quite meaningful academically. Second, the applied role of the algorithms in question is minor but not non existent: There are situations when we are interested in a high-accuracy solution to a convex program with a small number (just a few tens) of variables and a large number of constraints, and in these situations, methods such as the ellipsoid algorithm are the methods of choice. There are quite respectful and important applications where the problems of this type arise systematically and, moreover, the ability of a cutting plane algorithm to produce an accuracy certificate is instrumental. For an instructive example, see §5.2.

The other reason for a potential criticism is our specific choice of the accuracy measure—residual in the objective $F(x) - \mathrm{Opt}$. Of course, this is a natural accuracy measure but not as natural and universal as the inaccuracy in argument, which, in the case of problems (1), becomes the distance from a candidate solution to the set $X_*(F)$ of minimizers of F on X. First, the inaccuracy in argument is indeed a universal accuracy measure—it is applicable to all computational problems while the residual in the objective is adjusted to the specific form (1) of a convex optimization problem. Second, even when speaking solely on problems (1), there are situations (e.g., parameter estimation via maximum likelihood) when what we want is closeness of an approximate solution to the optimal set, and closeness to optimality in terms of the objective, whatever tight it may be, by itself means nothing. In fact, we are not aware of a *single* application of optimization where the opposite is true—small residual in the objective is a meaningful quality measure while small distance to the set of optimal solutions is not. All of this being said, we stick to the residual in terms of the objective, the "excuse" (in our opinion, quite sufficient) being the fact that *in convex optimization, inaccuracy in argument is, in general, completely unobservable online.* Here is one of the precise versions of this vague (what does "in general" mean?) claim. Consider the family $\mathcal{F}$ of all problems (1) with common domain $X = \{x \in \mathcal{R}^2 : x^T x \le 1\}$ and objectives F that are restrictions on X of C^∞ convex functions on the plane equal to $x^T x$ outside of X and with all partial derivatives up to a given order d bounded in X in absolute value by a given constant C_d. Let the first-order oracle report not only the value and the gradient but also all other derivatives of the objective at a query point. It turns out (for justification, see §6.1) that for a not-too-small C_d, this family does *not* allow for a nonanticipative solution algorithm with the following properties: As applied to any problem from the family, (a) the algorithm terminates after finitely many calls to the oracle (the number of calls can depend on the problem and is not required to be bounded uniformly in the problems) and (b) upon termination, the algorithm outputs an approximate solution that is at the distance $<1/4$ from the optimal set of the problem. Note that there are many algorithms, say, the simplest gradient descent, that, as applied to a (whatever) problem from the family, converge to the optimal set. The above result says that this convergence is completely "unobservable" online: It is impossible to convert, in a nonanticipating reliable fashion, the accumulated-thus-far information into a point that is at the distance $\le 1/4$

from the optimal set, and this is true independently of how long we are ready to wait. This is in striking contrast to what we can achieve for the same family of problems with the residual in the objective in the role of the inaccuracy in argument: a (whatever) accuracy $\epsilon > 0$ can be guaranteed in just $O(1)\ln(1/\epsilon)$ calls to the (usual) first-order oracle, each call accompanied by just $O(1)$ arithmetic operations.

REMARK. Of course, there are cases in convex minimization where the inaccuracy in the argument is "observable online." The simplest example here is when the objectives in the family of CMPs in question are restricted to be *strictly* convex with a given modulus of strict convexity $\sigma(\cdot)$: $F(x/2 + y/2) \le F(x)/2 + F(y)/2 - \sigma(\|x - y\|)$ for all $x, y \in X$, where $\sigma(s) > 0$ for $s > 0$. Note, however, that in this case, inaccuracy in the argument can be straightforwardly bounded in terms of $\epsilon_{\text{opt}}(\cdot)$; the same holds true in all other cases known to us, cases when a desired inaccuracy in the argument can be certified in an online fashion. To the best of our knowledge, all "explicit" (that is, those expressed in terms of the family of problems and independent of a particular choice of an instance in this family) online bounds on inaccuracy in the argument in convex minimization are more or less straightforward consequences of online bounds on the residual in the objective. With this observation in mind, our choice of accuracy measure seems to be quite natural.

Up to now, we were speaking about accuracy certificates for black box-represented convex minimization problems. It turns out that the "certificate machinery" can be extended from CMPs to other black box-represented problems "with convex structure," specifically, to variational inequalities with monotone operators and to convex Nash equilibrium problems (the latter include, in particular, the problems of approximating saddle points of convex-concave functions). We shall see that, in the case of variational inequalities and Nash equilibrium problems, the certificates play a more important role than in the CMP case. Indeed, in the CMP case, building approximate solutions is by itself easy—take the best (with the smallest value of the objective) among the feasible solutions generated so far. Thus, the role of certificates reduces here to quantifying the quality of such a solution, which sometimes can be considered a luxury rather than a necessity. In contrast to this, in the case of black box-represented variational inequalities with monotone operators and convex Nash equilibrium problems, the certificates are responsible not only for quantifying the quality of approximate solutions but for the very building of these solutions. For example, to the best of our understanding, no "certificatefree" ways are known for building approximate solutions to black box-represented convex-concave game problems.

The rest of this paper is organized as follows. In §2, we present a more detailed account of accuracy certificates for convex minimization. We start §3.1 by recalling basic facts on variational inequalities with monotone operators and convex Nash equilibrium problems. We then introduce relevant accuracy measures, extend the notion of a certificate to these classes of problems, and demonstrate (Proposition 3.4) that a certificate allows us to both generate an approximate solution and quantify its quality. In §4, we demonstrate—and this is the central result of the paper—that all known black box-oriented polynomial-time algorithms (in particular, the ellipsoid method) for CMPs, variational inequalities with monotone operators, and convex Nash equilibrium problems can be augmented with computationally cheap techniques for "online" building of accuracy certificates. In particular, we show that the use of these certificates is fully compatible with theoretical efficiency estimates of the methods (meaning that the number of steps until a prescribed accuracy is certified is the same as the number of steps predicted by the theoretical efficiency estimates). The only relevant study known to us in the existing literature is Burrell and Todd [6], which deals with the particular case of an LP solved by the ellipsoid algorithm. It is demonstrated in this study that the method can be augmented by a technique that produces feasible solutions to the dual problem; the duality gap then certifies the standard polynomial-time efficiency estimate of the ellipsoid method. The technique seems to be quite different from the one we propose, and it is unclear whether it can be extended beyond the particular LP case. Section 5 presents two instructive applications of the certificate machinery.

To make the paper more readable, all proofs are included to §6.

2. Accuracy certificates in convex minimization. Here, we present in detail the definition of accuracy certificates for black box-represented CMPs (1) and the related results already mentioned in §1.

Consider an algorithm for solving a black box CMP (1). After τ steps, the algorithm has at its disposal the *execution protocol* $P_\tau = \{I_\tau, J_\tau, \{x^t, e_t\}_{t=1}^\tau\}$, where $x^1, \dots, x^\tau$ are the subsequent points at which the separation and the first-order oracles were queried, I_τ and J_τ are the sets of indices of productive ($x^t \in \operatorname{int} X$), resp., nonproductive ($x^t \notin \operatorname{int} X$) query points, and e_t is either the separator of x^t and $\operatorname{int} X$ returned by the separation oracle (step t is nonproductive) or the subgradient of F at x^t returned by the first-order oracle (step t is

productive). A *certificate* for execution protocol P_τ is, by definition, a collection $\xi = \{\xi_t\}_{t=1}^\tau$ of weights such that

$$\begin{aligned} &\text{(a)} \quad \xi_t \geq 0 \quad \text{for each } t = 1, \ldots, \tau \\ &\text{(b)} \quad \sum_{t \in I_\tau} \xi_t = 1. \end{aligned} \tag{5}$$

Note that certificates exist only for protocols with nonempty sets I_τ.

Given a solid $\mathbf{B}$ known to contain X, an execution protocol P_τ, and a certificate ξ for this protocol, we define the *residual* of ξ on $\mathbf{B}$,

$$\epsilon_{\text{cert}}(\xi \mid P_\tau, \mathbf{B}) \equiv \max_{x \in \mathbf{B}} \sum_{t=1}^{\tau} \xi_t \langle e_t, x^t - x \rangle, \tag{6}$$

and *the approximate solution induced by* ξ *as*

$$\hat{x}^\tau[\xi] \equiv \sum_{t \in I_\tau} \xi_t x^t. \tag{7}$$

Clearly, $\hat{x}^\tau[\xi]$ is strictly feasible for the CMP (1) underlying the protocol, i.e., $\hat{x}^\tau[\xi] \in \operatorname{int} X$. If, in addition, the values of F at the points x^t, $t \in I_\tau$ are given, then we can also define the quantity

$$\begin{aligned} F_*(\xi \mid P_\tau, \mathbf{B}) &\equiv \min_{x \in \mathbf{B}} \left[\sum_{t \in I_\tau} \xi_t [F(x^t) + \langle e_t, x - x^t \rangle] + \sum_{t \in J_\tau} \xi_t \langle e_t, x - x^t \rangle \right] \\ &= \sum_{t \in I_\tau} \xi_t F(x^t) - \epsilon_{\text{cert}}(\xi \mid P_\tau, \mathbf{B}). \end{aligned} \tag{8}$$

Note that the quantities defined by (6) and (8) are easy to compute (recall that $\mathbf{B}$ is assumed to be simple, meaning that it is easy to optimize linear forms over $\mathbf{B}$). The role of the just-defined quantities in certifying the accuracy of approximate solutions to (1) stems from the following simple observation.

PROPOSITION 2.1. *Let P_τ be a τ-point execution protocol associated with a CMP (1), ξ be a certificate for P_τ, and $\mathbf{B} \supset X$ be a solid.*

(i) *One has $F_*(\xi \mid P_\tau, \mathbf{B}) \leq \text{Opt}$; consequently, for every feasible solution $\hat{x}$ of the given CMP,*

$$\epsilon_{\text{opt}}(\hat{x}) \leq F(\hat{x}) - F_*(\xi \mid P_\tau, \mathbf{B}). \tag{9}$$

(ii) *Let x^τ_{bst} be the best—with the smallest value of F—of the search points x^t generated at the productive steps $t \in I_\tau$. Then, both x^τ_{bst} and $\hat{x}^\tau = \hat{x}^\tau[\xi]$ are strictly feasible solutions of the given CMP, with*

$$\epsilon_{\text{opt}}(x^\tau_{bst}) \leq \epsilon_{\text{cert}}(\xi \mid P_\tau, \mathbf{B}) \qquad \text{and} \qquad \epsilon_{\text{opt}}(\hat{x}^\tau) \leq \epsilon_{\text{cert}}(\xi \mid P_\tau, \mathbf{B}). \tag{10}$$

It was mentioned in §1 that for certain problem classes, the only way to guarantee that a feasible approximate solution $\hat{x}$ to (1), built upon termination, satisfies $\epsilon_{\text{opt}}(\hat{x}) \leq \epsilon$ is the ability to assign the corresponding execution protocol P_τ, τ being the termination step, with a certificate ξ satisfying (4). To formulate the underlying statement precisely, consider the situation as follows. We intend to solve the CMP (1) and our a priori information on the problem is that its feasible domain X is a solid contained in a given solid $\mathbf{B} \subset \mathcal{R}^n$ (and, perhaps, containing another given convex set) and the objective F belongs to a "wide enough" family of convex functions on X, specifically, is either (a) convex and continuous, or (b) convex and piecewise linear, or (c) convex and Lipschitz continuous, with a given constant L, on X. All remaining information on the problem can be obtained solely from a separation oracle representing X and a first-order oracle representing F. Now, consider a solution algorithm that, as applied to every problem (1) compatible with our a priori information, in a finite number $\tau < \infty$ of steps terminates and returns a feasible solution $\hat{x}$ to the problem along with an upper bound ϵ on $\epsilon_{\text{opt}}(\hat{x})$; here, both τ and ϵ can depend on the particular problem to which the algorithm is applied. Adding, if necessary, one more step, we can assume w.l.o.g. that the approximate solution returned by the algorithm upon termination is always one of the search points generated by the algorithm in the course of the solution process. The aforementioned "necessity of accuracy certificates" in convex minimization can now be stated as follows.

PROPOSITION 2.2. *If our hypothetical algorithm as applied to a (whatever) CMP (1) compatible with our a priori information terminates after finitely many steps τ and returns a feasible solution $\hat{x}$ along with a finite valid upper bound ϵ on $\epsilon_{\text{opt}}(\hat{x})$, then $\hat{x} \in \operatorname{int} X$ and the associated execution protocol P_τ admits a certificate ξ that satisfies (4).*

3. Certificates for variational inequalities with monotone operators and for convex Nash equilibrium problems.

3.1. Monotone and Nash operators.

3.1.1. Monotone operators and variational inequalities. Let $\Phi(x)\colon \mathrm{Dom}(\Phi) \to \mathcal{R}^n$ be a *monotone* operator, meaning that $\mathrm{Dom}(\Phi) \subset \mathcal{R}^n$ is a convex set and

$$\langle \Phi(x') - \Phi(x''), x' - x'' \rangle \geq 0 \quad \forall\, x', x'' \in \mathrm{Dom}(\Phi).$$

Let $X \subset \mathcal{R}^n$ be a solid and Φ be a monotone operator with $\mathrm{Dom}(\Phi) \supseteq \mathrm{int}\, X$. The pair (X, Φ) defines the *variational inequality problem* (VIP):

$$\text{find } x_* \in X\colon \langle \Phi(x), x - x_* \rangle \geq 0 \quad \forall\, x \in \mathrm{Dom}(\Phi) \cap X. \tag{11}$$

In the literature, the just-defined x_* are called *weak* solutions to the VIP in question as opposed to *strong* solutions $x^* \in X \cap \mathrm{Dom}(\Phi)$: $\langle \Phi(x^*), x - x^* \rangle \geq 0\ \forall\, x \in X$. It is immediately seen that when Φ is monotone, then strong solutions are weak ones; under mild regularity assumptions (e.g., continuity of Φ), the inverse is true as well (see Minty [15] or Proposition 3.1 in Harker and Pang [10]).

One of our goals is to approximate a weak solution of a VIP defined by a pair (X, Φ) where X is a solid and Φ is a monotone operator. For our purposes, the most convenient way to quantify inaccuracy of an $x \in X$ as an approximate solution to this VIP is to use the proximity measure

$$\epsilon_{\mathrm{vi}}(x \mid X, \Phi) = \sup_{y \in \mathrm{Dom}(\Phi) \cap X} \langle \Phi(y), x - y \rangle. \tag{12}$$

When X and Φ are evident from the context, we shall use the abbreviated notation $\epsilon_{\mathrm{vi}}(x)$. By definition, we set $\epsilon_{\mathrm{vi}}(x \mid X, \Phi) = \infty$ when $x \notin X$. This measure is known as the *dual gap function* for a VI; it was introduced for the first time by Auslender [1] and has since then been used by many authors (see, e.g., Facchinei and Pang [7, pp. 122–123], Lemarechal et al. [13], Nemirovskii [19]).

Remark. By definition of (weak) solution, solutions of (11) are exactly the points $x \in X$ where $\epsilon_{\mathrm{vi}}(x \mid x, \Phi) \leq 0$. In fact, the left-hand side of this inequality is nonnegative everywhere on X so that $\epsilon_{\mathrm{vi}}(x \mid X, \Phi)$ is zero if and only if x solves (11) and is positive otherwise, which makes ϵ_{vi} a legitimate proximity measure for (11). The reasons for this particular choice of proximity are similar to those in favor of ϵ_{opt} in the CMP case: in our context, this measure works.

Elementary properties of weak solutions to VIPs with monotone operators and of the just-defined proximity measure are summarized in the following.

Proposition 3.1. *Consider VIP* (11) *where* $X \subset \mathcal{R}^n$ *is a solid and* F *is a monotone operator with* $\mathrm{Dom}\,\Phi \supseteq \mathrm{int}\, X$. *Then,*

(i) *The set* X_* *of (weak) solutions to* (11) *is a nonempty compact subset of* X.

(ii) *The function* $\epsilon_{\mathrm{vi}}(x \mid X, \Phi)$ *is a closed convex nonnegative function on* X, *finite everywhere on* $\mathrm{int}\, X$, *and equal to* 0 *exactly at* X_*.

(iii) *Let* $\widetilde{\Phi}$ *be a monotone extension of* Φ, *that is,* $\widetilde{\Phi}$ *associates with a point* x *of a convex set* $\mathrm{Dom}\,\widetilde{\Phi} \supseteq \mathrm{Dom}\,\Phi$, *a nonempty set* $\widetilde{\Phi}(x)$ *such that*

$$\langle \xi - \xi', x - x' \rangle \geq 0 \quad \forall\, (x, x' \in \mathrm{Dom}\,\widetilde{\Phi}, \xi \in \widetilde{\Phi}(x), \xi' \in \widetilde{\Phi}(x'))$$

and $\Phi(x) \in \widetilde{\Phi}(x)$ *whenever* $x \in \mathrm{Dom}\,\Phi$. *Then,*

$$\begin{aligned} X_* &= \{x \in X\colon \langle \eta, y - x \rangle \geq 0\ \forall\, (y \in \mathrm{Dom}\,\widetilde{\Phi} \cap X, \eta \in \widetilde{\Phi}(y))\}, \\ \epsilon_{\mathrm{vi}}(x \mid X, \Phi) &= \sup_y \{\langle \eta, x - y \rangle\colon y \in \mathrm{Dom}\,\widetilde{\Phi} \cap X, \eta \in \widetilde{\Phi}(y)\} \quad \forall\, x \in X. \end{aligned} \tag{13}$$

In particular, weak solutions and proximity measures remain intact when (a) *replacing the original monotone operator with its maximal monotone extension, and* (b) *replacing* Φ *with its restriction onto* $\mathrm{int}\, X$ (*so that the original operator becomes a monotone extension of the new one*).

To make the paper self-contained, we present the proof of these well-known facts in §6.

3.1.2. Convex Nash equilibrium problems and Nash operators. Let $X_i \subset \mathcal{R}^{n_i}$, $i = 1, \ldots, m$ be solids, and let $X = X_1 \times X_2 \times \cdots \times X_m \subset \mathcal{E} = \mathcal{R}^{n_1} \times \cdots \times \mathcal{R}^{n_m} = \mathcal{R}^{n_1+\cdots+n_m}$. A point $x \in \mathcal{E}$ is an ordered tuple $(x_1, \ldots, x_m)$ with $x_i \in \mathcal{R}^{n_i}$; for such a point and for $i \in \{1, \ldots, m\}$, we denote by x^i the projection of x onto the orthogonal complement of $\mathcal{R}^{n_i}$ in $\mathcal{E}$, and write $x = (x^i, x_i)$.

The *Nash equilibrium problem* (NEP) on X is specified by a collection of m real-valued functions $F_i(x) = F_i(x^i, x_i)$ with common domain $\mathcal{D} = \mathcal{D}_1 \times \cdots \times \mathcal{D}_m$, where for every $i = 1, \ldots, m$, $\operatorname{int} X_i \subseteq \mathcal{D}_i \subseteq X_i$. A *Nash equilibrium* associated with these data is a point $x_* \in \mathcal{D}$ such that for every $i = 1, \ldots, m$ the function $F_i([x_*]^i, x_i)$ of x_i attains its minimum over $x_i \in \mathcal{D}_i$ at $[x_*]_i$, and the Nash equilibrium problem is to find such an equilibrium. The standard interpretation of Nash equilibrium is as follows: There are m players, the ith choosing a point $x_i \in \mathcal{D}_i$ and incurring cost $F_i(x)$, where $x = (x_1, \ldots, x_m)$ is comprised of choices of all m players. Every player is interested in minimizing his cost, and Nash equilibria are exactly the tuples $x = (x_1, \ldots, x_m) \in \mathcal{D}$ of choices of the players, where every one of the players has no incentive to deviate from his choice x_i. There is a natural way to quantify the inaccuracy of a point $x \in \mathcal{D}$ as an approximate Nash equilibrium and the corresponding proximity measure is

$$\epsilon_{\mathrm{N}}(x) = \sum_{i=1}^{m} \Big[F_i(x^i, x_i) - \inf_{y_i \in \mathcal{D}_i} F_i(x^i, y_i) \Big], \tag{14}$$

that is, the sum over the players of the maximal gain that each player i can get by deviating from his choice x_i when the remaining players stick to their choices (cf. the Nikaido-Isoda function (Nikaido and Isoda [23]) and the Ky Fan function for corresponding games).

We intend to consider *convex NEPs*, meaning that $\mathcal{D}$ is convex and, for every $i = 1, \ldots, m$, the function $F_i(x) = F_i(x^i, x_i)$ is convex in $x_i \in \mathcal{D}_i$ and concave in $x^i \in \mathcal{D}_1 \times \cdots \times \mathcal{D}_{i-1} \times \mathcal{D}_{i+1} \times \cdots \times \mathcal{D}_m$, and in addition, the function $F(x) = \sum_{i=1}^m F_i(x)$ is convex on $\mathcal{D}$. Our interest in convex NEPs stems from the fact that, as is well-known and will be explained later in this paper, such a problem reduces to a variational inequality with monotone operator and, as such, is efficiently solvable. To the best of our knowledge, this is the only generic case of NEP with this property. That being said, the reader should be aware that practically speaking, convexity, in minimization or in Nash equilibrium, is a rare commodity. Nevertheless, there exist meaningful situations when an NEP is convex. Here is an example.

NEP with pairwise interactions. Consider an NEP where every pair of m participating players is playing a zero sum matrix game in mixed strategies. The rules are that a common mixed strategy is selected by each player and applied in all the games he plays with the remaining players, and the loss of a player is the sum of his losses in those games. More exactly, $\mathcal{D}_i$ is the simplex of dimension n_i, meaning that ith player chooses at random a point from the n_i-element set $S_i = \{1, \ldots, n_i\}$ of his pure strategies. When players choose pure strategies $k_i \in S_i$, $i = 1, \ldots, m$, the ith player pays to jth one ($i \neq j$) the amount $A^{ij}_{k_i, k_j}$, so that the total loss of the ith player is $\sum_{j \neq i} A^{ij}_{k_i, k_j}$; here, A^{ij} are "matrices of pairwise interactions" and we assume that $A^{ij} = -[A^{ji}]^T$ for all $i \neq j$ (i.e., every pair of distinct players is playing a zero-sum game). Allowing for mixed strategies with independent outcomes over the players, the expected loss of ith player becomes

$$F_i(x) = \sum_{j \neq i} x_i^T A^{ij} x_j,$$

where $x_i = (x_{i,1}, \ldots, x_{i,n_i}) \in \mathcal{D}_i$ is the mixed strategy of the ith player—the vector of probabilities according to which ith player chooses his pure strategies $1, \ldots, n_i$. The resulting Nash equilibrium problem is convex—indeed, $F_i(x)$ is linear (and thus convex) in x_i and is linear (and thus concave) in x^i, and $\sum_i F_i(x) \equiv 0$ is convex. We could modify the situation by augmenting the losses $x_i^T A^{ij} x_j$ of ith player to his adversaries by a self-incurred loss $x_i^T A^{ii} x_i$; here, an entry $A^{ii}_{k,l}$ of the matrix A^{ij} can be thought of as the price ith player should pay for switching from a pure strategy k in a current round of the repeated game to a pure strategy l in the next round. Assuming the (symmetrization of the) matrices A^{ii} are positive semidefinite and, as before, $A^{ij} = -[A^{ji}]^T$ for $i \neq j$, we have a convex Nash problem, now with $\sum_i F_i(x)$ convex quadratic.

Given a convex NEP, we associate with it the *Nash operator*

$$\Phi(x) = (\Phi_1(x), \Phi_2(x), \ldots, \Phi_m(x)) \colon \operatorname{int} X \to \mathcal{E},$$

where for every $x \in \operatorname{int} X$ and $i = 1, \ldots, m$, $\Phi_i(x)$ is a subgradient of the convex function $F_i(x^i, \cdot)$ at the point x_i. We need the following well-known fact (to make the paper self-contained, we present its proof in §6).

PROPOSITION 3.2. *Consider a convex NEP. Then,*

(i) *the Nash operator of the problem is monotone.*

(ii) *when $\mathcal{D} = X$ and the functions F_i are continuous on X, $1 \le i \le m$, the Nash equilibria are exactly the weak solutions to the VIP associated with X and the Nash operator of the problem.*

Note that convex NEPs are equipped with two proximity measures: $\epsilon_{\mathrm{N}}(\cdot)$ (this measure relates to the problem "as it is") and $\epsilon_{\mathrm{vi}}(\cdot)$ (this measure relates to the VIP associated with the Nash operator induced by the NEP). These measures usually do not coincide, and we would say that the first of them usually makes more "applied sense" than the second. We now consider instructive examples of convex NEPs.

Example A. In the case of $m = 1$, a convex NEP requires to minimize a convex function $F(x) = F_1(x)$ over $\mathrm{Dom}(F)$, where $\mathrm{int}\, X \subseteq \mathrm{Dom}\, F \subseteq X$. Thus, CMPs are exactly convex single-player NEPs, the Nash operator $\Phi = \Phi_1$ being (a section of) the subgradient field of F over $\mathrm{int}\, X$. The (weak) solutions to the VIP defined by (X, Φ) are just the minimizers of the lower semicontinuous extension of F from $\mathrm{int}\, X$ onto X. Further, for $x \in \mathrm{Dom}(F)$, one has

$$\epsilon_{\mathrm{N}}(x) = F(x) - \inf_{y \in \mathrm{Dom}(F)} F(y) = \epsilon_{\mathrm{opt}}(x),$$

where $\epsilon_{\mathrm{opt}}(x)$ is the accuracy measure defined in §1 for CMPs. If we view a convex NEP with $m = 1$ as a VIP with $\mathrm{Dom}\, \Phi = \mathrm{int}\, X$, the convexity of F assures that $F(x) - F(y) \ge \langle \Phi(y), x - y \rangle$ for all $x \in \mathrm{Dom}\, F$ and $y \in \mathrm{int}\, X$. Therefore,

$$\epsilon_{\mathrm{N}}(x) = \epsilon_{\mathrm{opt}}(x) = F(x) - \inf_{y \in \mathrm{Dom}(F)} F(y) \ge \sup_{y \in \mathrm{int}\, X} \langle \Phi(y), x - y \rangle = \epsilon_{\mathrm{vi}}(x). \tag{15}$$

Note that the "gap" between $\epsilon_{\mathrm{vi}}(x)$ and $\epsilon_{\mathrm{N}}(x)$ can be large, as is seen in the case of $X = [0, D] \subset \mathcal{R}$, $F(x) = \int_0^x \min[L, \epsilon/(D-s)]\, ds$, where $\epsilon \le LD$. Indeed, in this case, $\epsilon_{\mathrm{vi}}(D) = \max_{0<y<D} F'(y)(D-y) = \epsilon$ while $\epsilon_{\mathrm{N}}(D) = F(D) = \epsilon \ln(LD/\epsilon) + \epsilon$.

It is worth mentioning that the latter example is nearly "as extreme" as possible due to the following observation.

PROPOSITION 3.3. *Let the Nash operator Φ of a convex m-player Nash equilibrium problem on a solid X of the Euclidean diameter D be bounded so that $\|\Phi(x)\|_2 \le L < \infty$ for all $x \in \mathrm{int}\, X$. Then, for every $x \in \mathrm{int}\, X$, one has*

$$\epsilon_{\mathrm{N}}(x) \le m \left[\epsilon_{\mathrm{vi}}(x) \ln \frac{LD}{\epsilon_{\mathrm{vi}}(x)} + \epsilon_{\mathrm{vi}}(x) \right]. \tag{16}$$

Example B. In the case of $m = 2$ and $F(x) \equiv 0$, a convex Nash equilibrium problem is given by a function $\phi(x_1, x_2) = F_1(x_1, x_2)$ that is convex in $x_1 \in \mathcal{D}_1$ and is concave in $x_2 \in \mathcal{D}_2$, and $F_2(x_1, x_2) = -\phi(x_1, x_2) = -F_1(x_1, x_2)$. When $\mathcal{D}_i = X_i$, $i = 1, 2$ and $\phi(x)$ is continuous on X, the weak solutions to the VIP defined by (X, Φ) are exactly the saddle points (min in $x_1 \in X_1$, max in $x_2 \in X_2$) of ϕ. When $x = (x_1, x_2) \in \mathcal{D}$, the proximity measure $\epsilon_{\mathrm{N}}(x)$ admits a transparent interpretation (as follows): $\phi(\cdot, \cdot)$ specifies a "primal-dual pair" of optimization problems

$$(P)\ \ \mathrm{Opt}(P) = \inf_{x_1 \in \mathcal{D}_1} \bar{\phi}(x_1), \quad \bar{\phi}(x_1) = \sup_{x_2 \in \mathcal{D}_2} \phi(x_1, x_2),$$

$$(D)\ \ \mathrm{Opt}(D) = \sup_{x_2 \in \mathcal{D}_2} \underline{\phi}(x_2), \quad \underline{\phi}(x_2) = \inf_{x_1 \in \mathcal{D}_1} \phi(x_1, x_2).$$

The optimal value in the primal problem is $\ge$ the optimal value in the dual problem ("weak duality"), and the optimal values are equal to each other under mild regularity assumptions, e.g., when the convex-concave function ϕ is bounded. Note that now the quantity $\epsilon_{\mathrm{N}}(x_1, x_2)$ is nothing but the *duality gap* $\bar{\phi}(x_1) - \underline{\phi}(x_2)$, and

$$\epsilon_{\mathrm{N}}(x_1, x_2) = \bar{\phi}(x_1) - \underline{\phi}(x_2) \ge [\bar{\phi}(x_1) - \mathrm{Opt}(P)] + [\mathrm{Opt}(D) - \underline{\phi}(x_2)];$$

here, the inequality holds due to weak duality (and holds as equality whenever $\mathrm{Opt}(P) = \mathrm{Opt}(D)$). In particular, for $(x_1, x_2) \in \mathcal{D}$, the sum of nonoptimalities in terms of the respective objectives, of x_1 as a candidate solution to (P), and of x_2 as a candidate solution to (D) does not exceed $\epsilon_{\mathrm{N}}(x_1, x_2)$. Note that in the case in question, similar to the case of $m = 1$, we have

$$x \in \mathrm{Dom}\, \phi \ \Rightarrow\ \epsilon_{\mathrm{vi}}(x) \le \epsilon_{\mathrm{N}}(x). \tag{17}$$

Indeed, for $x = (x_1, x_2) \in \mathrm{Dom}\,\phi = \mathcal{D}_1 \times \mathcal{D}_2$ and $y = (y_1, y_2) \in \mathrm{Dom}\,\Phi = \mathrm{int}\,X_1 \times \mathrm{int}\,X_2$, taking into account that $\Phi_1(y)$ is a subgradient of convex function $\phi(\cdot, y_2)$ at y_1 and $\Phi_2(y)$ is a subgradient of the convex function $-\phi(y_1, \cdot)$ at y_2, we have

$$\begin{aligned}\langle \Phi(y), x - y\rangle &= \langle \Phi_1(y), x_1 - y_1\rangle + \langle \Phi_2(y), x_2 - y_2\rangle \\ &\le [\phi(x_1, y_2) - \phi(y_1, y_2)] + [\phi(y_1, y_2) - \phi(y_1, x_2)] = \phi(x_1, y_2) - \phi(y_1, x_2) \\ &\le \bar{\phi}(x_1) - \underline{\phi}(x_2) = \epsilon_{\mathrm{N}}(x)\end{aligned}$$

and (17) follows.

3.1.3. (Semi)bounded operators. Let $X \subset \mathcal{R}^n$ be a solid and Φ: $\mathrm{Dom}(\Phi) \to \mathcal{R}^n$ be an operator with $\mathrm{Dom}(\Phi) \supseteq \mathrm{int}\,X$. Φ is called *bounded* on X if the quantity $\|\Phi\| = \sup\{\|\Phi(x)\|_2 : x \in \mathrm{int}\,X\}$ is finite. Φ is called *semibounded* on X, if the quantity

$$\mathrm{Var}_X(\Phi) = \sup\{\langle \Phi(x), y - x\rangle : x \in \mathrm{int}\,X,\ y \in X\}$$

is finite. Clearly, a bounded operator Φ is semibounded with $\mathrm{Var}_X(\Phi) \le \|\Phi\|\,\mathrm{Diam}(X)$. There exist, however, semibounded operators that are unbounded. For example, the Nash operator Φ of a convex Nash equilibrium problem with bounded $F_i(\cdot)$, $i = 1, \ldots, m$ is clearly semibounded with $\mathrm{Var}_X(\Phi) \le \sum_i [\sup_{\mathrm{int}\,X} F_i - \inf_{\mathrm{int}\,X} F_i]$; at the same time Φ is unbounded, unless every one of the functions $F_i(x^i, \cdot)$ is Lipschitz continuous, with constant bounded uniformly in x^i on $\mathrm{int}\,X_i$. Moreover, in the case of $m = 1$ (convex minimization), the Nash operator can be semibounded for certain *un*bounded objectives F, most notably, when F is a *ϑ-self-concordant barrier for X* (Nesterov and Nemirovskii [21, Chapter 2]). For such a barrier, $\mathrm{Var}_X(F) \le \vartheta$ (Nesterov and Nemirovskii [21, Proposition 2.3.2]).

Convention. From now on, speaking about variational inequalities (or Nash equilibrium problems), we always assume the monotonicity of the underlying operator (resp., convexity of the NEP).

3.2. Accuracy certificates for VIP and NEP. The notions of an execution protocol and a certificate for such a protocol can be naturally extended from the case when the algorithms in questions are aimed at solving CMPs to the case of algorithms that are aimed at solving VIPs. Specifically,

(1) Given a simple solid $\mathbf{B} \subset \mathcal{R}^n$, we are interested in solving VIPs (X, Φ) (which may represent Nash equilibrium problems) given by solids $X \subset \mathbf{B}$ and by monotone operators Φ with $\mathrm{Dom}\,\Phi \supseteq \mathrm{int}\,X$. We assume that X is given by a separation oracle and Φ is given by a Φ-oracle that, given on input a point $x \in \mathrm{int}\,X$, returns the value $\Phi(x)$.

(2) A τ-point execution protocol P_τ generated by an algorithm as applied to (X, Φ) is $\{I_\tau, J_\tau, \{(x^t, e_t)\}_{t=1}^\tau\}$, where $x^1, \ldots, x^\tau$ are the subsequent query points generated at the first τ steps of the solution process, e_t is either the separator returned by the separation oracle queried at x^t (this is so when $x^t \notin \mathrm{int}\,X$, a "nonproductive step") or the vector $\Phi(x^t)$ returned by the Φ-oracle (this is so when $x^t \in \mathrm{int}\,X$, "a productive step"), and I_τ, J_τ are the set of productive, resp. nonproductive, steps $t = 1, \ldots, \tau$.

(3) A certificate ξ for a protocol P_τ is, exactly as above, a collection of nonnegative weights $\{\xi_t\}_{t=1}^\tau$ with $\sum_{t \in I_\tau} \xi_t = 1$. The *residual* $\epsilon_{\mathrm{cert}}(\xi \mid P_\tau, \mathbf{B})$ of such a certificate on a solid $\mathbf{B} \supseteq X$ is defined by (6) and the *approximate solution* $\hat{x}^\tau[\xi]$ *induced by* ξ is defined by (7).

The role of the just-introduced notions in our context stems from the following.

Proposition 3.4. *Let $X \subseteq \mathbf{B}$ be solids in $\mathcal{R}^n$ and Φ be a monotone operator with $\mathrm{Dom}(\Phi) \supseteq \mathrm{int}\,X$. Further, let $P_\tau = \{I_\tau, J_\tau, \{(e_t, x^t)\}_{t=1}^\tau\}$ be an execution protocol, $\xi = \{\xi_t\}_{t=1}^\tau$ be a certificate for P_τ, and $\hat{x}^\tau = \hat{x}^\tau[\xi]$ be the induced approximate solution. Then, $\hat{x}^\tau \in \mathrm{int}\,X$ and*

$$\epsilon_{\mathrm{vi}}(\hat{x}^\tau) \le \epsilon_{\mathrm{cert}}(\xi \mid P_\tau, \mathbf{B}); \tag{18}$$

in particular, $\epsilon_{\mathrm{cert}}(\xi \mid P_\tau, \mathbf{B}) \ge 0$. Further, when Φ is the Nash operator of a convex NEP, then

$$\epsilon_{\mathrm{N}}(\hat{x}^\tau) \le \epsilon_{\mathrm{cert}}(\xi \mid P_\tau, \mathbf{B}). \tag{19}$$

3.2.1. Necessity of accuracy certificates for VIPs. We have seen that an accuracy certificate for an execution protocol associated with a VIP allows both to build a strictly feasible approximate solution $\hat{x}$ for the VIP and to bound from above the inaccuracy $\epsilon_{\mathrm{vi}}(\hat{x})$ (and $\epsilon_{\mathrm{N}}(\hat{x})$ in the case of a Nash VIP) of this solution. It turns out that when solving VIPs from certain natural families, this mechanism is to some extent unavoidable. Specifically, assume that when solving VIP (11), per our a priori information, X is a solid contained in a given solid $\mathbf{B} \subset \mathcal{R}^n$ and Φ: $\operatorname{int} X \to \mathcal{R}^n$ is a monotone operator bounded by a given constant $L < \infty$ (i.e., $\|\Phi(x)\|_2 \le L$ for all $x \in \operatorname{int} X$). The only sources of further information on the VIP are a separation oracle representing X and a Φ-oracle that, given on input $x \in \operatorname{int} X$, returns $\Phi(x)$. Now, consider an algorithm for approximate (within ϵ_{vi}-accuracy $\le \epsilon$) solving VIPs in the just-described environment. As applied to a VIP (X, Φ) compatible with our a priori information and given on input a required tolerance $\epsilon > 0$, the algorithm generates subsequent search points $x^1, x^2, \ldots$ and calls the separation oracle to check whether $x^t \in \operatorname{int} X$. If that is the case, the algorithm calls the Φ-oracle to compute $\Phi(x^t)$. After a finite number of steps τ, the algorithm terminates and outputs an approximate solution $\hat{x} \in \operatorname{int} X$ that must satisfy the relation $\epsilon_{\mathrm{vi}}(\hat{x} \mid X, \Phi) \le \epsilon$. Of course, we assume that $x^1, \ldots, x^\tau$, τ, and $\hat{x}$ are determined solely on the basis of information accumulated thus far, that is, the algorithm is nonanticipative. Adding, if necessary, one extra step, we can assume w.l.o.g. that $\hat{x}^\tau \in \{x^1, \ldots, x^\tau\}$ and, in particular, that upon termination, the set $I_\tau = \{t \le \tau\colon\ x^t \in \operatorname{int} X\}$ is nonempty. We have the following result (cf. Proposition 2.2).

PROPOSITION 3.5. *Let (X, Φ) be VIP compatible with our a priori information (so that $X \subset \mathbf{B}$ is a solid and Φ: $\operatorname{int} X \to \mathcal{R}^n$ is a monotone operator with $\|\Phi(x)\|_2 \le L$ for every $x \in \operatorname{int} X$), let τ be the number of steps of our hypothetical algorithm as applied to (X, Φ), and let P_τ be the associated execution protocol. Then, there exists a certificate ξ for P_τ such that*

$$\epsilon_{\mathrm{cert}}(\xi \mid P_\tau, \Phi) \le \sqrt{LD\epsilon}, \tag{20}$$

where D is the Euclidean diameter of $\mathbf{B}$.

3.2.2. Comment. Proposition 3.4 states that when solving the VIP defined by (X, Φ), a certificate ξ with $\epsilon_{\mathrm{cert}}(\xi \mid P_\tau, \mathbf{B}) \le \epsilon$, where $P_\tau = \{I_\tau, J_\tau, \{x^t, e_t\}_{t=1}^\tau\}$ is the execution protocol at step τ, P_τ and ξ allow one to build an approximate solution $\hat{x} \in \{x_t\colon\ t \in I_\tau\}$ such that $\epsilon_{\mathrm{vi}}(\hat{x}^\tau \mid X, \Phi) \le \epsilon$. Proposition 3.5, in turn, says that whenever an algorithm is capable of solving every VIP (X, Φ) with $X \subset \mathbf{B}$ and is a monotone and bounded by L operator Φ within $\epsilon_{\mathrm{vi}}(\cdot)$-accuracy $\le \epsilon$, and the resulting approximate solution $\hat{x} \in \operatorname{int} X$ belongs to the trajectory $x^1, \ldots, x^\tau$ upon termination, then the corresponding execution protocol P_τ admits a certificate ξ with $\epsilon_{\mathrm{cert}}(\xi \mid P_\tau, \mathbf{B}) \le \sqrt{LD\epsilon}$. Observing that $\epsilon_{\mathrm{vi}}(x \mid X, \Phi) = \sup_{y \in \operatorname{int} X} \langle \Phi(y), x - y \rangle \le LD$ whenever $x \in \operatorname{int} X$, the only interesting case is the one when $\epsilon \ll LD$ and, in this case, $\sqrt{LD\epsilon} \gg \epsilon$ so that there is a "gap" between the sufficiency and the necessity results stated by Propositions 3.4 and 3.5. We do not know whether this gap reflects the essence of the matter or just a weakness of Proposition 3.5.

4. Ellipsoid algorithm with certificates. As was already mentioned, the ellipsoid algorithm[2] is "an intelligent" method that, however, does not produce accuracy certificates explicitly. Our local goal is to equip the method with a "computationally cheap" mechanism for producing certificates with residuals converging to 0, as $\tau \to \infty$, at the rate justifying the usual polynomial-time theoretical efficiency estimate of the ellipsoid algorithm. In fact, the technique to be described is applicable to a general *cutting plane* scheme for solving black box-represented convex minimization problems and VIPs with monotone operators, and it makes sense to present the technique in question in this general context, thus making the approach more transparent and extending its scope.

4.1. Generic cutting plane algorithms. A generic cutting plane algorithm works with a vector field Φ: $\operatorname{Dom} \Phi \to \mathcal{R}^n$ defined on a convex set $\operatorname{Dom} \Phi \subset \mathcal{R}^n$ and with a solid X, $\operatorname{int} X \subset \operatorname{Dom} \Phi$. The algorithm, as applied to (X, Φ), builds a sequence of search points $x^t \in \mathcal{R}^n$ along with a sequence of *localizers* Q_t, solids such that $x^t \in \operatorname{int} Q_t$, $t = 1, 2, \ldots$. The algorithm is as follows.

Algorithm 4.1

Initialization: Choose a solid $Q_1 \supset X$ and a point $x^1 \in \operatorname{int} Q_1$.
Step t, $t = 1, 2, \ldots$: Given x^t, Q_t,

[2] This method was developed in Yudin and Nemirovskii [27] and independently and slightly later, in Shor [24]. For its role in the theory of convex optimization, see, e.g., Khachiyan [12], Grötschel et al. [9], Nemirovski [16], Ben-Tal and Nemirovski [2], and references therein.

(i) Call separation oracle, x^t being the input. If the oracle reports that $x^t \in \operatorname{int} X$ (*productive step*), go to (ii). Otherwise (*nonproductive step*), the oracle reports a separator $e_t \neq 0$ such that $\langle e_t, x - x^t\rangle \leq 0$ for all $x \in X$. Go to (iii).

(ii) Call Φ-oracle to compute $e_t = \Phi(x^t)$. If $e_t = 0$, terminate; otherwise, go to (iii).

(iii) Set

$$\hat{Q}_{t+1} = \{x \in Q_t\colon \langle e_t, x - x^t\rangle \leq 0\}.$$

Choose, as Q_{t+1}, a solid that contains the solid $\hat{Q}_{t+1}$. Choose $x^{t+1} \in \operatorname{int} Q_{t+1}$ and loop to step $t+1$.

For a solid $B \subset \mathcal{R}^n$, let $\rho(B)$ be the radius of Euclidean ball in $\mathcal{R}^n$ with the same n-dimensional volume as the one of B. A cutting plane algorithm is called *converging* on (X, Φ) if, for the associated localizers Q_t, one has $\rho(Q_t) \to 0$, $t \to \infty$.

4.1.1. An implementation: The ellipsoid algorithm. The ellipsoid algorithm is, historically, the first "polynomial-time" implementation of the cutting plane scheme. In this algorithm,

(i) The initial localizer Q_1 is centered at the origin Euclidean ball $\mathbf{B}$ of radius R known to contain X.

(ii) All localizers Q_t are ellipsoids represented as the images of the unit Euclidean ball under affine mappings

$$Q_t = \{x = B_t u + x^t\colon u^T u \leq 1\} \qquad [B_t\colon n \times n \text{ nonsingular}]$$

so that the search points x^t are the centers of the ellipsoids.

(iii) Q_{t+1} is the ellipsoid of the smallest volume containing the half-ellipsoid $\hat{Q}_{t+1}$. The corresponding updating $(B_t, x^t) \mapsto (B_{t+1}, x^{t+1})$ is given by

$$\begin{aligned} q_t &= B_t^T e_t, \qquad p_t = \frac{B_t^T q_t}{\sqrt{q_t^T B_t B_t^T q_t}} \\ B_{t+1} &= \alpha(n) B_t + (\gamma(n) - \alpha(n)) B_t p_t p_t^T \\ x^{t+1} &= x^t - \frac{1}{n+1} B_t p_t, \end{aligned} \tag{21}$$

where $n > 1$ is the dimension of x and

$$\alpha(n) = \frac{n}{\sqrt{n^2-1}}, \qquad \gamma(n) = \frac{n}{n+1}. \tag{22}$$

The ellipsoid method is converging with

$$\rho(Q_{t+1}) = \kappa(n)\rho(Q_t), \quad \kappa(n) = \alpha^{\frac{n-1}{n}}(n)\gamma^{\frac{1}{n}}(n) = \frac{n}{(n+1)^{\frac{n+1}{2n}}(n-1)^{\frac{n-1}{2n}}} \leq \exp\left\{-\frac{1}{2n(n+1)}\right\}. \tag{23}$$

4.1.2. Other implementations. Besides of the ellipsoid algorithm, there are several other implementations of the cutting plane scheme with "rapid convergence": $\rho(Q_t) \leq p(n)\rho(Q_1)\exp\{-t/p(n)\}$ with a fixed polynomial $p(n) > 0$. The list includes:

(1) the center of gravity method (Levin [14], Newman [22]): $Q_1 = X$, x^t is the center of gravity of Q_t, $Q_{t+1} = \hat{Q}_{t+1}$; here, $p(n) = \text{const}$. This method, however, is of academic interest only because it requires finding centers of gravity of general type polytopes, which is NP-hard.

(2) the inscribed ellipsoid algorithm (Tarasov et al. [25]) where Q_1 is a box containing X, x^t is the center of the ellipsoid of the (nearly) largest volume contained in Q_t, and $Q_{t+1} = \hat{Q}_{t+1}$; here, again, $p(n) = \text{const}$.

(3) the circumscribed simplex algorithm (Bulatov and Shepot'ko [5], Yamnitsky and Levin [26]), where Q_t are simplexes and x^t are the barycenters of Q_t; here, $p(n) = O(1)n^3$. The ellipsoid and the circumscribed simplex algorithms are examples of the *stationary* cutting plane scheme—one where $Q_t = x^t + B_t\mathbf{C}$ for a fixed solid $\mathbf{C}$, $0 \in \operatorname{int}\mathbf{C}$. In order for such a scheme to be converging, $\mathbf{C}$ should possess a specific and rare property as follows. *For every $e \neq 0$, the set $\widehat{\mathbf{C}}_e = \{x \in \mathbf{C}\colon \langle e, x\rangle \leq 0\}$ can be covered by an affine image of $\mathbf{C}$ under an efficiently computable affine mapping that reduces volumes by factor $\exp\{-1/p(n)\}$ for an appropriate polynomial $p(n) > 0$.* For a long time, the only known solids with this property were ellipsoids and simplexes. Recently, it was discovered (Gabelev [8]) that the required property is shared by all compact cross-sections of *symmetric* cones by hyperplanes; here, a symmetric cone is, by definition, a finite direct product of irreducible factors, which are the Lorentz cones and the cones of positive semidefinite symmetric real/Hermitian/quaternion

matrices. Among the associated converging stationary cutting plane algorithms, the ellipsoid method, associated with the Lorentz cone, is the fastest in terms of the guaranteed rate of convergence of $\rho(Q_t)$ to zero. The circumscribed simplex method, associated with the direct product of nonnegative rays (which are nothing but cones of positive semidefinite real 1×1 matrices), is the slowest one.[3]

4.2. Building the certificates: Preliminaries. To equip a generic cutting plane algorithm with certificates, let us treat our original "universe" $\mathcal{R}^n$ as the hyperplane $E = \{(x, t) \in \mathcal{R}^{n+1}: t = 1\}$ in $E^+ = \mathcal{R}^{n+1}$, and let us associate with the localizers Q_t the sets Q_t^+, which are convex hulls of the sets Q_t (treated as subsets in E) and the origin in E^+:

$$Q_t^+ = \{[sx; s]: 0 \le s \le 1, x \in Q_t\}.^4$$

Let us further associate with vectors $e_t \in \mathcal{R}^n$ the *cuts* $e_t^+ = [e_t; -\langle e_t, x^t\rangle] \in \mathcal{R}^{n+1}$. Observe that the convex hulls $\hat{Q}_{t+1}^+ = \{[sx; s]: 0 \le s \le 1, x \in \hat{Q}_{t+1}\}$ of the origin in E^+ and the sets $\hat{Q}_{t+1} \subset E$, the sets Q_t^+, and the cuts are linked by the relation

$$\hat{Q}_{t+1}^+ = \{z \in Q_t^+: \langle e_t^+, z\rangle \le 0\} \subset Q_{t+1}^+. \tag{24}$$

Recall that the polar of a closed convex set $P \subset E^+$, $0 \in P$ is the set $\mathrm{Polar}(P) = \{z \in E^+: \langle z, p\rangle \le 1 \ \forall p \in P\}$. An immediate observation is as follows.

PROPOSITION 4.1. *Assume that $e_t \ne 0$, so that Q_{t+1} is well-defined. Then,*

$$\mathrm{Polar}(Q_{t+1}^+) \subset \mathrm{Polar}(\hat{Q}_{t+1}^+) = \mathrm{Polar}(Q_t^+) + \mathcal{R}_+ e_t^+. \tag{25}$$

4.3. Building the certificates: The algorithm. Equipped with Proposition 4.1, we can build certificates in a "backward fashion" as follows.

Algorithm 4.2
Given an iteration number τ, we build a certificate for the corresponding protocol $P_\tau = \{I_\tau, J_\tau, \{(x^t, e_t)\}_{t=1}^\tau\}$ as follows.

At a terminal (i.e., with $e_\tau = 0$) step τ: Because $e_\tau = 0$ can happen at a productive step only, we set here $\xi_t = 0$, $t < \tau$, $\xi_\tau = 1$, which, due to $e_\tau = 0$, results in a certificate for P_τ with

$$\epsilon_{\mathrm{cert}}(\xi \mid P_\tau, Q_1) = 0.$$

At a nonterminal (i.e., with $e_\tau \ne 0$) step τ:

(i) We choose a "nearly most narrow stripe" containing $Q_{\tau+1}$, namely, find a vector $h \in \mathcal{R}^n$ such that

$$\max_{x \in Q_{\tau+1}} \langle h, x\rangle - \min_{x \in Q_{\tau+1}} \langle h, x\rangle \le 1 \tag{26}$$

and

$$\|h\|_2 \ge \frac{1}{2\chi\rho(Q_{\tau+1})}, \quad \chi \le 4n. \tag{27}$$

Note that such an h always exists and, for all known converging cutting plane algorithms, can be easily found.

Indeed, when Q_{t+1} is an ellipsoid, one can easily find h satisfying (26) with $\|h\|_2 \le 1/(2\rho(Q_{t+1}))$ (see the following). In the general case, we can apply the "ellipsoidal construction" to the *Fritz John ellipsoid* of Q_{t+1}—an ellipsoid $\tilde{Q} \supset Q_{t+1}$ with $\rho(\tilde{Q}) \le n\rho(Q_{t+1})$ (see John [11]). Note that for all the aforementioned "rapidly converging" implementations of the cutting plane Scheme, except for the center-of-gravity one (which, in any case, is not implementable), the Fritz John ellipsoids of the localizers are readily available.

(ii) Observe that both the vectors

$$h^+ = [h; -\langle h, x^{\tau+1}\rangle] \in E^+, \quad h^- = -h^+$$

[3] It should be stressed that we are speaking about the theoretical worst case-oriented complexity bounds. "In reality," the circumscribed simplex algorithm seems to be faster than the ellipsoid one.

[4] Here, and in what follows, we use "MATLAB notation": for matrices $v^1, \dots, v^k$ with a common number of columns (e.g., for column vectors), $[v^1; \dots; v^k]$ stands for the matrix obtained by writing v^2 beneath v^1, v^3 beneath v^2, and so on.

clearly belong to $\mathrm{Polar}(Q_{\tau+1}^+)$. Applying Proposition 4.1 recursively, we build representations

$$\begin{aligned} &\text{(a)}\quad h^+ = \sum_{t=1}^{\tau} \lambda_t e_t^+ + \phi^+ && [\lambda_t \ge 0,\ \phi^+ \in \mathrm{Polar}(Q_1^+)] \\ &\text{(b)}\quad h^- = -h^+ = \sum_{t=1}^{\tau} \mu_t e_t^+ + \psi^+ && [\mu_t \ge 0,\ \psi^+ \in \mathrm{Polar}(Q_1^+)]. \end{aligned} \tag{28}$$

The certificate for the protocol P_τ is well-defined only for those τ for which the set $I_\tau = \{t \le \tau\colon x^t \in \mathrm{int}\, X\}$ is nonempty and, moreover, the quantity $d_\tau \equiv \sum_{t\in I_\tau}[\lambda_t + \mu_t]$ is positive. In this case, the certificate is given by

$$\xi_t = \frac{\lambda_t + \mu_t}{d_\tau}, \quad 1 \le t \le \tau$$

and we also set

$$\begin{aligned} D_\tau &= \sum_{t\in I_\tau} [\lambda_t + \mu_t]\|e_t\|_2 = d_\tau \sum_{t\in I_\tau} \xi_t \|e_t\|_2 \\ W_\tau &= \max_{t\in I_\tau} \max_{x\in X} \langle e_t, x - x^t\rangle. \end{aligned}$$

Note: The quantities D_τ, W_τ are used solely in the convergence analysis and Algorithm 4.2 is *not* required to compute these quantities.

Implementation of Algorithm 4.2 in the case of ellipsoid method in the role of Algorithm 4.1 is very easy. Indeed,

- to find h in rule (i), it suffices to build the singular value decomposition $B_{\tau+1} = UDV$ (U, V are orthogonal and D is diagonal with positive diagonal entries) and set $h = 1/(2\sigma_{i_*})Ue_{i_*}$, where e_i are the standard basic orths and i_* is the index of the smallest diagonal entry σ_{i_*} in D. We clearly have $\sigma_{i_*} \le |\mathrm{Det}(B_{\tau+1})|^{1/n} = \rho(Q_{\tau+1})$, so that $\|h\|_2 \ge 1/(2\rho(Q_{\tau+1}))$ (i.e., we ensure (27) with $\chi = 1$). Besides this,

$$\max_{Q_{t+1}} \langle h, x\rangle - \min_{Q_{t+1}} \langle h, x\rangle = \max_{\|u\|_2\le 1,\ \|v\|_2\le 1} \langle B_{t+1}^T h, u - v\rangle = \tfrac{1}{2} \max_{\|u\|_2\le 1,\ \|v\|_2\le 1} \langle V^T e_{i_*}, u - v\rangle = 1$$

as required in rule (i);

- we clearly have $\mathrm{Polar}(Q_t^+) = \{[e; s]\colon \|B_t^T e\|_2 + \langle x^t, e\rangle + s \le 1\}$ so that, given $[g; a] \in \mathrm{Polar}(Q_{t+1}^+)$, to find a representation $[g; a] = [f; b] + re_t^+$ with $[f; b] \in \mathrm{Polar}(Q_t^+)$ and $r \ge 0$ reduces to solving the problem (known in advance to be solvable) of finding $r \ge 0$ such that $\|B_t^T(g - re_t)\|_2 + \langle x^t, g\rangle + a \le 1$. A solution r_* can be found at the cost of just $O(n^2)$ arithmetic operations. Specifically, if the vectors $p = B_t^T g$ and $q = B_t^T e_t$ have nonpositive inner products, $r_* = 0$ is a solution; otherwise, a solution is given by $r_* = p^T q/(q^T q)$.

From the just-presented remarks, it follows that with the ellipsoid method in the role of Algorithm 4.1, the computational cost of building a certificate for P_τ is $O(n^3) + O(\tau n^2)$ arithmetic operations (a.o.), provided that the subsequently generated data (search points x^t, matrices B_t, and vectors e_t) are stored in the memory. Note that the cost of carrying out τ steps of the ellipsoid algorithm is *at least* $O(\tau n^2)$ (this is the total complexity of updating $(x^t, B_t) \mapsto (x^{t+1}, B_{t+1})$, $t = 1, \ldots, \tau$ with the computational expenses of the oracles excluded). It follows that when certificates are built along a "reasonably dense" subsequence of steps, e.g., at steps $2, 4, 8, \ldots$ the associated computational expenses basically do not affect the complexity of the ellipsoid algorithm (see §4.4.1 for discussion).

4.4. The main result.

THEOREM 4.1. *Let Algorithms 4.1–4.2 be applied to (X, Φ), where $X \subset \mathcal{R}^n$ is a solid and $\Phi\colon \mathrm{int}\, X \to \mathcal{R}^n$ is a vector field, and let $r = r(X)$ be the largest of the radii of Euclidean balls contained in X.*

(i) *Let τ be an iteration number such that $\xi = \{\xi_t\}_{t=1}^\tau$ is well-defined. Then, ξ is a certificate for the corresponding execution protocol $P_\tau = \{I_\tau, J_\tau, \{(x^t, e_t)\}_{t=1}^\tau\}$. If τ is the terminal iteration number (i.e., $e_\tau = 0$), then $\epsilon_{\mathrm{cert}}(\xi \mid P_\tau, Q_1) = 0$; otherwise,*

$$\epsilon_{\mathrm{cert}}(\xi \mid P_\tau, Q_1) \le \frac{2}{d_\tau} \tag{29}$$

and

$$\epsilon_\tau \equiv \frac{2}{D_\tau} < r \ \Rightarrow\ \epsilon_{\mathrm{cert}}(\xi \mid P_\tau, Q_1) \le \frac{\epsilon_\tau}{r - \epsilon_\tau} W_\tau. \tag{30}$$

(ii) *Let $D(Q_1)$ be the Euclidean diameter of Q_1. Then, for every nonterminal iteration number τ, one has*

$$D_\tau \geq D^{-1}(Q_1)\left(\frac{r}{2\chi\rho(Q_{\tau+1})} - 1\right). \tag{31}$$

(iii) *Whenever τ is a nonterminal iteration number such that*

$$\rho(Q_{\tau+1}) \leq \frac{r^2}{16\chi D(Q_1)}, \tag{32}$$

the certificate ξ is well-defined and

$$\epsilon_{\text{cert}}(\xi \mid P_\tau, Q_1) \leq \frac{16\chi D(Q_1) W_\tau}{r^2}\rho(Q_{\tau+1}), \tag{33}$$

so that when Φ is semibounded on $\operatorname{int} X$: $\operatorname{Var}_X(\Phi) \equiv \sup_{x\in\operatorname{int} X,\, y\in X}\langle\Phi(x), y-x\rangle < \infty$, *we have*

$$\epsilon_{\text{cert}}(\xi \mid P_\tau, Q_1) \leq \frac{16\chi D(Q_1)\operatorname{Var}_X(\Phi)}{r^2}\rho(Q_{\tau+1}). \tag{34}$$

In particular, in the case of the ellipsoid algorithm (where $\chi = 1$ and $\rho(Q_{t+1}) \leq R\exp\{-t/(2n(n-1))\}$, R being the radius of the ball $\mathbf{B} = Q_1$), the certificate ξ is well-defined provided that

$$\exp\left\{\frac{\tau}{2n(n+1)}\right\} \geq \frac{32R^2}{r^2} \tag{35}$$

and, in this case,

$$\epsilon_{\text{cert}}(\xi \mid P_\tau, Q_1) \leq \frac{32R^2\operatorname{Var}_X(\Phi)}{r^2}\exp\left\{-\frac{\tau}{2n(n+1)}\right\} \tag{36}$$

provided that Φ is semibounded on X.

Combining Proposition 3.4 and Theorem 4.1, we arrive at the following result.

COROLLARY 4.1. *Let $X \subset \mathcal{R}^n$ be a solid that is contained in the Euclidean ball $\mathbf{B} = Q_1$ of radius R, centered at the origin, and let $r = r(X)$ be the largest of the radii of Euclidean balls contained in X. Further, let* Φ: $\operatorname{Dom}\Phi \to \mathcal{R}^n$, $\operatorname{Dom}\Phi \supset \operatorname{int} X$ *be semibounded on* $\operatorname{int} X$ *operator*:

$$\operatorname{Var}_X(\Phi) \equiv \sup_{x\in\operatorname{int} X,\, y\in X}\langle\Phi(x), y-x\rangle < \infty.$$

For appropriately chosen absolute constant $O(1)$ and every $\epsilon \in (0, \operatorname{Var}_X(\Phi)]$ with

$$\tau(\epsilon) = \left\lfloor O(1)n^2 \ln\left(\frac{R\operatorname{Var}_X(\Phi)}{r\epsilon}\right)\right\rfloor,$$

the ellipsoid algorithm, augmented with Algorithm 4.2 for building certificates as applied to (X, Φ), in $\tau \leq \tau(\epsilon)$ steps generates a protocol P_τ and a certificate ξ for this protocol such that

$$\epsilon_{\text{cert}}(\xi \mid P_\tau, \mathbf{B}) \leq \epsilon,$$

along with the corresponding solution $\hat{x} = \sum_{t\in I_\tau}\xi_t x^t \in \operatorname{int} X$. As a result,

- *when Φ is monotone, we have*

$$\epsilon_{\text{vi}}(\hat{x} \mid X, \Phi) \leq \epsilon;$$

- *when Φ is the Nash operator of a convex Nash equilibrium problem on X (in particular, in the case of convex minimization), we have*

$$\epsilon_{\text{N}}(\hat{x}) \leq \epsilon.$$

4.4.1. Discussion. Corollary 4.1 states that solving a convex Nash equilibrium problem (or a variational inequality) associated with a semibounded, monotone operator Φ within accuracy ϵ "costs" $O(1)n^2\ln(R\,\mathrm{Var}_X(\Phi)/(r\epsilon))$ steps of the ellipsoid algorithm, with (at most) one call to the separation and Φ-oracles and $O(1)n^2$ additional a.o. per step; note that this is nothing but the standard theoretical complexity bound for the ellipsoid algorithm. The advantage offered by the certificates is that whenever a certificate ξ with $\epsilon_{\mathrm{cert}}(\xi \mid P_\tau, \mathbf{B}) \le \epsilon$ is built, we know that we already have at our disposal a strictly feasible approximate solution of accuracy ϵ. Thus, in order to get a strictly feasible solution of a prescribed accuracy $\epsilon > 0$, we can run the ellipsoid algorithm and apply from time to time Algorithm 4.2 in order to generate a strictly feasible approximate solution and to bound from above its nonoptimality, terminating the solution process when this bound becomes $\le\epsilon$. Note that *this implementation, though ensuring a desired quality of the resulting approximate solution, requires no a priori information on the problem* (aside from its convexity and the knowledge of a ball $\mathbf{B}$ containing the feasible domain X). Moreover, if Algorithm 4.2 is invoked along a "reasonably dense, but not too dense" sequence of time instants, e.g., at steps $2, 4, 8, \ldots$, the number of steps upon termination will be *at most* the one given by the theoretical complexity bound of the ellipsoid algorithm, and the computational effort to run Algorithm 4.2 will be a small fraction of the computational effort required to run the ellipsoid algorithm itself.

4.5. Incorporating "deep cuts." The certificates we use can be sharpened by allowing for "deep cuts." Specifically, assume that the separation oracle provides more information than we have postulated, namely, that in the case of $x \notin \operatorname{int} X$, this oracle reports a vector $e \ne 0$ *and* a nonnegative number a such that $X \subseteq \{y\colon \langle e, y - x\rangle \le -a\}$. In this situation, all of our results remain valid when the residual of a certificate on a solid $\mathbf{B} \supset X$ is redefined as

$$\epsilon_{\mathrm{cert}}(\xi \mid P_\tau, \mathbf{B}) = \max_{x\in\mathbf{B}} \sum_{t=1}^{\tau} \xi_t[\langle e_t, x^t - x\rangle - a_t], \tag{37}$$

where $[e_t; a_t]$ is the output of the separation oracle at a nonproductive step t and $a_t = 0$ for productive steps t, and the sets $\hat{Q}_{t+1}$ in Algorithm 4.1 are defined as

$$\hat{Q}_{t+1} = \{x \in Q_t\colon \langle e_t, x - x^t\rangle \le -\alpha_t\}, \tag{38}$$

where $\alpha_t = a_t$ for all t. With the ellipsoid method in use, the latter modification allows one to replace the updating formulae (21), (22) with

$$\begin{aligned} q_t &= B_t^T e_t, \quad p_t = \frac{B_t^T q_t}{\sqrt{q_t^T B_t B_t^T q_t}}, \quad \mu_t = \frac{\alpha_t}{\sqrt{q_t^T B_t B_t^T q_t}}, \\ B_{t+1} &= \left[(1-\mu_t)^2\frac{n^2}{n^2-1}\right]^{1/2}\left[B_t + \left[1 - \sqrt{\frac{(1-\mu_t)(n-1)}{(1+\mu_t)(n+1)}}\right]B_t p_t p_t^T\right] \\ x^{t+1} &= x^t - \frac{1+n\mu_t}{n+1}B_t p_t. \end{aligned} \tag{39}$$

New updating formulae, while still ensuring that the ellipsoid Q_{t+1} contains $\hat{Q}_{t+1}$, result in larger reduction in the volumes of subsequent ellipsoids.

In the case of convex minimization, we can allow for "deep cuts" also at productive steps by setting

$$a_t = F(x^t) - \min_s \{F(x^s)\colon s \le t,\ x^s \in \operatorname{int} X\}, \qquad \alpha_t = a_t/2$$

for a productive step t (a_t is used in (37), α_t is used in (38) and (39)). With this modification, all of our CMP-related results remain valid (modulo changes in absolute constant factors in Theorem 4.1), provided that the approximate solutions we are considering are the best-found-so-far solutions x^τ_{bst}. Moreover, in this situation, the relation

$$\epsilon_{\mathrm{opt}}(x^\tau_{\mathrm{bst}}) \le \epsilon_{\mathrm{cert}}(\xi \mid P_\tau, \mathbf{B})$$

(cf. (10)) remains valid whenever $\mathbf{B}$ is a solid containing the "leftover" part

$$X_\tau = \{x \in X\colon \langle e_t, x - x^t\rangle \le -a_t\ \forall\, t \in I_\tau\}$$

of the feasible domain. Thus, using "deep cuts" in the cutting plane scheme and when defining the residual allows one to accelerate somehow the algorithms and to obtain somehow improved accuracy bounds.

The proofs of the just-outlined "deep cut" modifications of our results are obtained from the proofs presented in this paper by minor and straightforward modifications.

5. Application examples. In this section, we present two instructive examples of application of the certificate machinery (more examples will be given in our forthcoming follow-up paper). The first example is of an academic nature and the second is of a practical flavor.

5.1. Recovering dual solutions to large LP problems. Consider a linear programming (LP) problem with box constraints

$$\min_{x\in X} F(x)\equiv\langle c,x\rangle,\quad X=\{x\in\mathcal{R}^n:\ \hat{A}x\equiv[A;I;-I]x\le\hat{b}=[b;\mathbf{1};\mathbf{1}]\},\tag{40}$$

where $\mathbf{1}$ is the all-one vector. Note that X is contained in the box $\mathbf{B}=\{x\in\mathcal{R}^n:\ -\mathbf{1}\le x\le\mathbf{1}\}$. Assuming that X has a nonempty interior and A has no all-zero rows, we have $\operatorname{int}X=\{x\in\mathcal{R}^n:\ \hat{A}x<\hat{b}\}$. Assume that we have at our disposal an oracle, which, given on input a point $u\in\mathcal{R}^n$, reports whether $u\in\operatorname{int}X$ and, if it is not the case, returns a constraint $a_{j(u)}^Tx\le b_{j(u)}$ from the system $\hat{A}x\le\hat{b}$, which is *not* strictly satisfied at $x=u$. This oracle can be considered as a separation oracle for X, the separator in the case of $x\notin\operatorname{int}X$ being given by the vector $a_{j(x)}$.

Assume that we are solving (40) by an iterative oracle-based method and after a number τ of steps have at our disposal the corresponding execution protocol $P_\tau=\{I_\tau,J_\tau,\{(x^t,e_t)\}_{t=1}^\tau\}$ along with a certificate ξ with small residual on the box $\mathbf{B}$. Let us look what kind of "LP information" we can extract from this certificate. For $t\in J_\tau$, the vectors e_t are (the transposes of) certain rows $a_{j(x^t)}^T$ of the matrix $\hat{A}$ such that $a_{j(x^t)}^Tx^t\ge b_{j(x^t)}$. Setting

$$\lambda_j^\tau=\sum_{t\in J_\tau:\,j(x^t)=j}\xi_t,\quad j=1,\dots,m,$$

where m is the number of rows in $\hat{A}$, we obtain a nonnegative vector λ^τ such that

$$\sum_{t\in J_\tau}\xi_te_t=\hat{A}^T\lambda^\tau,\qquad\sum_{t\in J_\tau}\xi_t\langle e_t,x^t\rangle\ge\hat{b}^T\lambda^\tau.$$

Now, when $t\in I_\tau$, we have $e_t=c$. Setting $\hat{x}=\sum_{t\in I_\tau}\xi_tx^t$ and taking into account the outlined remarks, the relation

$$\epsilon_{\text{cert}}(\xi\mid P_\tau,\mathbf{B})=\max_{x\in\mathbf{B}}\left\{\sum_{t=1}^\tau\xi_t\langle e_t,x^t-x\rangle\right\}$$

implies that

$$\epsilon_{\text{cert}}(\xi\mid P_\tau,\mathbf{B})\ge[c^T\hat{x}+\hat{b}^T\lambda^\tau]+\max_{x\in\mathbf{B}}\langle-x,[c+\hat{A}^T\lambda^\tau]\rangle=\underbrace{[c^T\hat{x}+\hat{b}^T\lambda^\tau]}_{\text{"duality gap"}}+\underbrace{\|c+\hat{A}^T\lambda^\tau\|_1}_{\text{"dual infeasibility"}}\tag{41}$$

(note that $\mathbf{B}$ is the unit box). Now, $\hat{A}^T=[A^T,I,-I]$; decomposing λ^τ accordingly as: $\lambda^\tau=[\lambda_A^\tau;\lambda_+^\tau;\lambda_-^\tau]$, we can easily build nonnegative $\Delta\lambda_+,\Delta\lambda_-\in\mathcal{R}^n$ such that the vector $\hat{\lambda}=[\lambda_A^\tau;\lambda_+^\tau+\Delta\lambda_+;\lambda_-^\tau+\Delta\lambda_-]\ge0$ satisfies the relation $c+\hat{A}^T\hat{\lambda}=0$ and $\|\Delta\lambda_+\|_1+\|\Delta\lambda_-\|_1=\|c+\hat{A}^T\lambda^\tau\|_1$, whence

$$\hat{b}^T\hat{\lambda}\le\|c+\hat{A}^T\lambda^\tau\|_1+\hat{b}^T\lambda^\tau$$

(note that the entries in $\hat{b}$ corresponding to nonzero entries in $\lambda^\tau-\hat{\lambda}$ are ±1). Invoking (41), we arrive at

$$\hat{x}^\tau\in X,\quad\hat{\lambda}\ge0,\quad c+\hat{A}^T\hat{\lambda}=0,\quad c^T\hat{x}+\hat{b}^T\hat{\lambda}\le\epsilon_{\text{cert}}(\xi\mid P_\tau,\mathbf{B}).$$

In other words, in the case in question, a certificate for P_τ straightforwardly induces a pair $(\hat{x},\hat{\lambda})$ of feasible solutions to the problem of interest and its LP dual such that the duality gap, evaluated at this pair, does not exceed $\epsilon_{\text{cert}}(\xi\mid P_\tau,\mathbf{B})$.

This example is of significant interest when the LP problem (40) has a huge number of constraints that are "well-organized" in the sense that it is relatively easy to find a constraint, if any, that is not strictly satisfied at a given point (for examples, see Grötschel et al. [9]). By Corollary 4.1, the latter property allows, given $\epsilon>0$, to build in $O(n^2\ln(n\|c\|_2/(r(X)\epsilon)))$ steps, with a single call to the separation oracle for X and $O(n^2)$ additional operations per step, a protocol P_τ and associated certificate ξ such that $\epsilon_{\text{cert}}(\xi\mid P_\tau,\mathbf{B})\le\epsilon$. With the just-presented construction, we can easily convert (P_τ,ξ) into a pair of feasible solutions to the primal and the dual problems with duality gap $\le\epsilon$. Note that, in our situation, the dual problem has a huge number of variables, which makes it impossible to solve the dual by the standard optimization techniques. In fact, with huge m, it is even impossible to write down efficiently a candidate solution to the dual as a vector; with the outlined approach, this difficulty is circumvented by allowing ourselves to indicate indices and values of *only nonzero* entries in the "sparse" dual solution we get.

5.2. Auxiliary problems in large-scale minimization via first-order methods. As a matter of fact, numerous "simply looking" large-scale problems are beyond the grasp of modern polynomial-time interior point methods because the latter require Newton-type iterations that, in high dimensions (many thousands of decision variables), become prohibitively expensive unless the problem possesses a favorable sparsity pattern (which not always is the case). As a result, there is a growing interest, initiated by the breakthrough paper Nesterov [20], in solving well-structured convex programs by computationally cheap first-order optimization techniques. A common feature of the state-of-the-art first-order algorithms (smoothing technique from Nesterov [20], Mirror Prox algorithm from Nemirovski [17], NERML algorithm (Ben-Tal and Nemirovski [3]), and many others) is the necessity to solve at every step an auxiliary problem of the form

$$\mathrm{Opt} = \min_{u\in U}\{f(u) \equiv \omega(u) + e^T u\colon\ Au \le b\}, \tag{42}$$

where $U \subset \mathcal{R}^N$ is a simple, although perhaps extremely large scale, convex compact set, and $\omega(\cdot)$ is a continuous convex function on U that "fits" U, meaning that it is easy to minimize over U a function of the form $\omega(u)+g^T u$, $g \in \mathcal{R}^N$ (an instructive example is the one when U is a large-scale simplex $\{u \in \mathcal{R}^N_+\colon\ \sum_i u_i = 1\}$ and $\omega(u) = \sum_i u_i \ln u_i$ is the entropy). Thus, the only difficulty when solving (42) can come from the general type constraints $Au \le b$. If we knew how to incorporate at a moderate cost even small (few or perhaps few tens) systems of these constraints, it would extend significantly the scope of the "master methods." A natural way to incorporate a small system of general type linear constraints is to pass from the problem (42) to its partial Lagrange dual problem:

$$\min_{x\ge 0} F(x), \quad F(x) = -\min_{u\in U}\Big\{\underbrace{f(u) + \langle x, Au - b\rangle}_{\phi(u,x)}\Big\} \tag{43}$$

(the standard dual is $\max_{x\ge 0}[-F(x)]$; we represent it in the equivalent form (43) in order to get a CMP). Note that it is easy to equip F with the first-order oracle: Because ω fits U, it is easy to find $u_x \in \arg\min_{u\in U}\phi(u,x)$, which gives us both the value $F(x) = -\phi(u_x, x)$ and a subgradient $F'(x) = -(Au_x - b)$ of F at x. Assuming that (42) satisfies the Slater condition (so that (43) is solvable) and that we have at our disposal an upper bound L on the norm $\|x_*\|_p$ of an optimal solution x_* to (43) (in the applications we are speaking about, both assumptions are quite realistic), we can therefore "reduce" the situation to solving a low-dimensional black box-represented CMP:

$$\min_{x\in X} F(x), \quad X = \{x \in \mathcal{R}^k_+\colon\ \|x\| \le 2L\}, \tag{44}$$

with easy-to-implement separation and first-order oracles. If the dimension k of the resulting problem (i.e., the number of general type linear constraints in (42)) is indeed small, say, ≤ 10, it makes full sense to solve (44) by a polynomial-time cutting plane algorithm, say, by the ellipsoid method, which allows one to obtain a high accuracy solution to (44), even when L is large, in a relatively low (say, a few hundreds) number of steps. Note, however, that as far as our ultimate goals are concerned, we need not only a high-accuracy approximation to the optimal value of (42) (which is readily given by a high-accuracy solution to (44)) but also a high-accuracy approximate solution to (42). The question is how to extract such a solution from the information accumulated when solving (44). It turns out that this can be easily done via the certificate machinery.

Proposition 5.1. *Let (43) be solved by a black box-oriented method, $P_\tau = \{I_\tau, J_\tau, \{x^t, e_t\}_{t=1}^\tau\}$ be the execution protocol upon termination, and ξ be an accuracy certificate for this protocol. Set*

$$\hat{u} = \sum_{t\in I_\tau} \xi_t u_{x^t}, \tag{45}$$

where $u_{x^t} \in \arg\min_{u\in U}[f(u) + \langle x^t, Au - b\rangle]$ underline the answers of the first-order oracle for F (defined by (43)) at the points x^t. Then, $\hat{u} \in U$ and

$$\text{(a)}\ \ \|[A\hat{u} - b]^+\|_q \le \epsilon_{\mathrm{cert}}(\xi \mid P_\tau, X), \quad \text{(b)}\ \ f(\hat{u}) - \mathrm{Opt} \le \epsilon_{\mathrm{cert}}(\xi \mid P_\tau, X), \tag{46}$$

where $[A\hat{u} - b]^+$ is the "vector of constraint violations" obtained from $A\hat{u} - b$ by replacing the negative components with 0, and $q = p/(p+1)$.

Proposition 5.1 shows that the vector $\hat{u}$, defined by (45), is nearly feasible and nearly optimal for (42), provided that $\epsilon_{\mathrm{cert}}(\xi \mid P_\tau, \mathbf{B})$ is small.

6. Proofs.

6.1. Proof of "impossibility claim" in §1. We restrict ourselves with presenting the sketch of the proof and omit boring and straightforward technicalities. Our class of objectives clearly contains a function f_0, which attains

its minimum on a segment Δ_0 of the length $1/2$ and is strongly convex outside of this segment. Now, assume that there exists a nonanticipative algorithm $\mathcal{B}$ for $\mathcal{F}$ with properties (a) and (b), and let us lead this assumption to a contradiction. To this end, consider the construction as follows. The first search point x_1 generated by $\mathcal{B}$ depends solely on $\mathcal{B}$—it is built before any specific information on the objective becomes known. If this point happens to belong to Δ_0, we perturb the function f_0 slightly to get a function $f_1 \in \mathcal{F}$ with exactly the same properties as f_0 but with the segment Δ_1 of minimizers not passing through x_1. If $x_1 \notin \Delta_0$, we set $f_1 \equiv f_0$, $\Delta_1 = \Delta_0$. In both cases, we report to $\mathcal{B}$ the derivatives of f_1 at x_1. With this information, the method generates a new search point x_2. If it does not belong to Δ_1, we set $f_2 = f_1$, $\Delta_2 = \Delta_1$; otherwise, we perturb slightly f_1 to get a function $f_2 \in \mathcal{F}$ with the same properties as f_1 and such that x_2 does not belong to the segment Δ_2 of minimizers of f_2 and that $f_2 \equiv f_1$ in a neighborhood of x_1; such a perturbation clearly exists. We proceed in the same fashion, i.e., after t steps, we have at our disposal functions $f_\tau \in \mathcal{F}$, $\tau \le t$ with the same properties as f_0, and points x_τ such that for every $\tau \le t$, one has

(a_τ): $x_1, \dots, x_\tau$ are the first τ search points generated by $\mathcal{B}$ as applied to f_τ,

(b_τ): no one of the points $x_1, \dots, x_\tau$ belongs to the segment Δ_τ of the minimizers of f_τ,

(c_τ): $f_\tau \equiv f_{\tau'}$ in a neighborhood of x_s whenever $s \le \tau' \le \tau$.

At a step $t+1$, we act as follows.

(1) It might happen that $\mathcal{B}$, as applied to f_t, terminates at the step t and outputs an approximate solution $\bar{x}$. In that case, we could perturb f_t slightly, keeping it in $\mathcal{F}$, to get a function f that coincides with f_t in a neighborhood of $x_1, \dots, x_t$ and attains its minimum at a single point, specifically, at that one of the endpoints of Δ_t, which is at the distance $\ge 1/4$ from $\bar{x}$ (this clearly is possible because no one of the points $x_1, \dots, x_t$ belongs to the set Δ_t of minimizers of f_t by (b_t), and this set is a segment of the length $\ge 1/2$). Now, because $x_1, \dots, x_t, \bar{x}$ is the search trajectory and the result generated by $\mathcal{B}$ as applied to f_t, on one hand, and on the other hand to $f = f_t$ in a neighborhood of $x_1, \dots, x_t$ by (a_t), while the method is nonanticipative, $x_1, \dots, x_t$ and $\bar{x}$ are the search trajectory and the result generated by $\mathcal{B}$ as applied to f. By construction, the distance from $\bar{x}$ to the set of minimizers of f is $\ge 1/4$, which contradicts property (a). Thus, in fact, option (1) is impossible.

We see that $\mathcal{B}$, as applied to f_t, does *not* terminate at step t and generates a search point x_{t+1}. It is possible that x_{t+1} does not belong to the segment Δ_t of minimizers of f_t, in which case we set $f_{t+1} = f_t$ and proceed to the next step of our construction; note that we have ensured the validity of (a_τ), (b_τ), and (c_τ) for all $\tau \le t+1$. Now, consider the case when x_{t+1} does belong to the segment Δ_t of minimizers of f_t. Note that, in this case, x_{t+1} is distinct from $x_1, \dots, x_t$ by (b_t) and, therefore, we can perturb f_t slightly to get a function $f_{t+1} \in \mathcal{F}$ with the following properties: (i) f_{t+1} coincides with f_t in a neighborhood of $x_1, \dots, x_t$; (ii) same as for f_t, the set Δ_{t+1} of minimizers of f_{t+1} is a segment of length $\ge 1/2$, and f_{t+1} is strongly convex outside of this segment; (iii) no one of the points $x_1, \dots, x_{t+1}$ belongs to Δ_{t+1}. Note that $x_1, \dots, x_{t+1}$ are the first $t+1$ points generated by $\mathcal{B}$ as applied to f_t (by (a_t) and the definition of x_{t+1}); because $\mathcal{B}$ is nonanticipative, (i) implies that $x_1, \dots, x_{t+1}$ are the first $t+1$ search points generated by $\mathcal{B}$ as applied to f_{t+1} so that (a_{t+1}) takes place. The validity of (b_{t+1}) and (c_{t+1}) is readily given by (a_t)–(c_t) and (i)–(iii).

We see that we can proceed with our construction and get a sequence of functions $f_t \in \mathcal{F}$ and points x_t, $t = 1, 2, \dots$, satisfying for every τ the relations (a_τ)–(c_τ). Besides this, the perturbation converting f_t in f_{t+1} can be made arbitrarily small, in particular, such that $f_t(\cdot)$ along with their derivatives of every fixed order converge uniformly as $t \to \infty$ to a certain function f that clearly belongs to $\mathcal{F}$. By (c_τ), the derivatives of f at x_τ, for every τ, are the same as the derivatives at x_τ of every one of the functions f_τ, $f_{\tau+1}, \dots$. This combines with (a_τ) and with nonanticipativity of $\mathcal{B}$ to imply that $x_1, x_2, \dots$ is the search trajectory generated by $\mathcal{B}$ as applied to f. We see that *as applied to $f \in \mathcal{F}$, $\mathcal{B}$ does not terminate*. Indeed, if the algorithm as applied to f were terminating at certain step t, it would terminate at this very step as applied to f_t (by nonanticipativity and due to the fact that, by our construction, the derivatives of f_t and f along the trajectory $x_1, \dots, x_t$ of $\mathcal{B}$ as applied to f_t coincide with each other). Thus, $\mathcal{B}$ does not possess property (b), which is a desired contradiction. □

6.2. Proof of Proposition 2.1. (i) was already proved in §1. To prove (ii), note that both x^τ_{bst} and $\widehat{x}^\tau$ belong to $\operatorname{int} X$ (as convex combinations of the belonging to $\operatorname{int} X$ points x^t, $t \in I_\tau$). Next, from (8),

$$F_*(\xi \mid P_\tau, \mathbf{B}) = \underbrace{\sum_{t \in I_\tau} \xi_t F(x^t)}_{\overset{\ge}{(*)} F(\widehat{x}^\tau)} - \epsilon_{\mathrm{cert}}(\xi \mid P_\tau, \mathbf{B}) \ge \underbrace{\min_{t \in I_\tau} F(x^t)}_{= F(x^\tau_{\mathrm{bst}})} - \epsilon_{\mathrm{cert}}(\xi \mid P_\tau, \mathbf{B}), \tag{47}$$

with $(*)$ given by (5.b) and the convexity of F. Thus, (10) follows from (9). □

6.3. Proof of Proposition 2.2. Consider an algorithm (applied to a problem $\min_{x\in X} F(x)$) that is compatible with our a priori information. Suppose the algorithm terminates in τ steps and generates execution protocol $P_\tau = \{I_\tau, J_\tau, \{(x^t, e_t)\}_{t=1}^\tau\}$, a feasible solution $\hat{x}$, and a valid upper bound ϵ on $\epsilon_{\text{opt}}(\hat{x})$. With our assumptions on the algorithm, $\hat{x}$ is one of the points $x^1, \dots, x^\tau$, say, $\hat{x} = x^s$. We first claim that $s \in I_\tau$ (and thus $I_\tau \neq \varnothing$ and $\hat{x} \in \operatorname{int} X$). Indeed, our a priori information on X, augmented by the information accumulated in the course of running the algorithm on the problem, does not contradict the assumption that $X = X^\delta := \{x \in \mathbf{B}\colon \langle e_t, x - x^t\rangle \le -\delta, t \in J_\tau\}$, provided that $\delta > 0$ is small enough, specifically, such that $\langle e_t, x^{t'} - x^t\rangle \le -\delta$ whenever $t \in J_\tau$ and $t' \in I_\tau$. When $s \in J_\tau$ and $\delta > 0$, we clearly have $x^s \notin X^\delta$. Thus, when $\delta > 0$ is small enough, our a priori and accumulated information is compatible with the assumption that $X = X^\delta$ while, under this assumption, the resulting approximate solution $\hat{x} = x^s$ is infeasible so that the corresponding inaccuracy $\epsilon_{\text{opt}}(\hat{x})$ is $+\infty$, which is not the case—we know that this inaccuracy is $\le \epsilon < \infty$.

Now, let us set

$$\bar{X} = \{x \in \mathbf{B}\colon \langle e_t, x - x^t\rangle \le 0,\ t \in J_\tau\}, \qquad \bar{F}(x) = \max_{t\in I_\tau}[F(x^t) + \langle F'(x^t), x - x^t\rangle]. \tag{48}$$

Observe that our a priori information on the problem $\min_{x\in X} F(x)$ ("X is a solid contained in $\mathbf{B}$ and F is either (a) convex continuous, or (b) convex piecewise linear, or (c) convex Lipschitz continuous, with a given constant L, function on X") plus the additional information acquired from the oracles when solving the problem allows for X to be exactly $\bar{X}$ and for F to be exactly the restriction of $\bar{F}$ on $\bar{X}$. That is, it may happen that

$$\text{Opt} = \text{Opt}_* \equiv \min_{x\in\bar{X}} \bar{F}(x) \tag{49}$$

and, in this situation, $\hat{x}$ still should be an ϵ-solution, that is, we should have

$$F(\hat{x}) - \text{Opt}_* \le \epsilon. \tag{50}$$

It remains to note that

$$\begin{aligned}
\text{Opt}_* &= \min_x \left\{ \max_{t\in I_\tau}[F(x^t) + \langle e_t, x - x^t\rangle]\colon \langle e_t, x - x^t\rangle \le 0,\ t \in J_\tau,\ x \in \mathbf{B} \right\} \\
&\underset{\text{(a)}}{=} \min_{x,s} \left\{ s\colon \begin{array}{l} s \ge F(x^t) + \langle e_t, x - x^t\rangle,\ t \in I_\tau \\ 0 \ge \langle e_t, x - x^t\rangle,\ t \in J_\tau \\ x \in \mathbf{B} \end{array} \right\} \\
&\underset{\text{(b)}}{=} \max_{\xi\ge 0} \left\{ \min_{x\in\mathbf{B}, s} \left[s + \sum_{t\in I_\tau} \xi_t[F(x^t) - s] + \sum_{t=1}^\tau \xi_t \langle e_t, x^t - x\rangle \right] \right\} \\
&\underset{\text{(c)}}{=} \max_{\xi} \left\{ \sum_{t\in I_\tau} \xi_t F(x^t) + \min_{x\in\mathbf{B}} \sum_{t=1}^\tau \xi_t \langle e_t, x - x^t\rangle\colon \xi \ge 0,\ \sum_{t\in I_\tau} \xi_t = 1 \right\} \\
&\underset{\text{(d)}}{=} \max_{\xi} \left\{ \sum_{t\in I_\tau} \xi_t F(x^t) - \epsilon_{\text{cert}}(\xi \mid P_\tau, \mathbf{B})\colon \xi \ge 0,\ \sum_{t\in I_\tau} \xi_t = 1 \right\}.
\end{aligned}$$

In this chain, (a) is evident, (b) is given by the Lagrange duality theorem (see, e.g., Boyd and Vandenberghe [4, Chapter 5]) as applied to the optimization problem in the right-hand side of (a).[5] Further, the same Lagrange duality theorem states that the optimal value in the optimization problem in (a), and thus in (b), is achieved, which makes (c) and (d) evident. We see that

$$\text{Opt}_* = \max_{\xi} \left\{ \sum_{t\in I_\tau} \xi_t F(x^t) - \epsilon_{\text{cert}}(\xi \mid P_\tau, \mathbf{B})\colon \xi \ge 0, \sum_{t\in I_\tau} \xi_t = 1 \right\}$$

so that there exists a certificate ξ^* for the protocol P_τ such that $\text{Opt}_* = \sum_{t\in I_\tau} \xi_t^* F(x^t) - \epsilon_{\text{cert}}(\xi^* \mid P_\tau, \mathbf{B})$. Substituting this representation into (50), we arrive at (4). □

[5] Applicability of the Lagrange duality theorem follows from the fact that this convex problem in the right-hand side of (a) clearly is below bounded because $\mathbf{B}$ is bounded and $I_\tau \neq \varnothing$ and satisfies the Slater condition—namely, every point s, x with $x \in \operatorname{int} X$ and $s > F(x)$ strictly satisfies the constraints.

6.4. Proof of Proposition 3.1. (i) By definition, X_* is the solution set of a system of nonstrict linear inequalities and, thus, is convex and closed; it is bounded along with X. To prove that $X_* \neq \varnothing$, assume that X_* is empty and let us lead this assumption to a contradiction. Because X is empty, the intersection of closed subsets $X_y = \{x \in X: \langle \Phi(y), y - x\rangle \geq 0\}$ of the compact set X over all $y \in \mathrm{Dom}\,\Phi \cap X$ is empty, meaning that there exist finitely many points $y_i \in \mathrm{Dom}\,\Phi \cap X$, $i = 1, \ldots, I$ such that $\bigcap_{i=1}^I X_{y_i} = \varnothing$ or, equivalently, the function $g(x) = \max_{1\leq i\leq I}\langle F(y_i), x - y_i\rangle$ is positive everywhere on X. Applying the von Neumann lemma, we see that there exists $\lambda_i \geq 0$, $\sum_{i=1}^I \lambda_i = 1$ such that

$$\sum_i \lambda_i \langle F(y_i), x - y_i\rangle > 0 \quad \forall\, x \in X. \tag{51}$$

On the other hand, setting $\bar{y} = \sum_i \lambda_i y_i$, we obtain $\bar{y} \in \mathrm{Dom}\,\Phi \cap X$ and, by monotonicity of F, we obtain $\langle F(\bar{y}), \bar{y} - y_i\rangle \geq \langle F(y_i), \bar{y} - y_i\rangle$. Taking the weighted sum of these inequalities with the weights λ_i, we get $0 \geq \sum_i \lambda_i \langle F(y_i), \bar{y} - y_i\rangle$, which contradicts (51). (i) is proved.

(ii) $\epsilon_{\mathrm{vi}}(x)$ clearly is convex and closed. If $x \in \mathrm{Dom}\,\Phi \cap X \supset \mathrm{int}\,X$, we have $\langle F(y), x - y\rangle \leq \langle F(x), x - y\rangle$, $y \in X \cap \mathrm{Dom}\,\Phi$, whence $\epsilon_{\mathrm{vi}}(x) = \sup_{y\in X\cap \mathrm{Dom}\,\Phi} \leq \sup_{y\in X\cap \mathrm{Dom}\,\Phi}\langle F(x), x - y\rangle \in [0, \infty]$. We see that $\epsilon_{\mathrm{vi}}(\cdot)$ is a closed convex function with domain contained in X and containing $X \cap \mathrm{Dom}\,\Phi$. We also see that $\epsilon_{\mathrm{vi}}(\cdot)$ is nonnegative on the latter set and, in particular, on $\mathrm{int}\,X$; because $\epsilon_{\mathrm{vi}}(\cdot)$ is convex on X and is nonnegative on $\mathrm{int}\,X$, $\epsilon_{\mathrm{vi}}(\cdot)$ is nonnegative. Finally, by definition, weak solutions to (11) are exactly the points $x \in X$ where $\epsilon_{\mathrm{vi}}(x) \leq 0$, and the latter set, as we just have seen, is exactly the set of zeros of $\epsilon_{\mathrm{vi}}(\cdot)$.

(iii) Let $\widetilde{\Phi}$ be a monotone extension of Φ and let $\widetilde{X}_*$, $\tilde{\epsilon}_{\mathrm{vi}}$ be the right-hand sides in (13). It is immediately seen that the proofs of (i) and (ii) are applicable to multivalued monotone operators so that $\widetilde{X}_*$ is exactly the set of zeros of $\tilde{\epsilon}_{\mathrm{vi}}(\cdot)$ on X and, therefore, all we need in order to verify (13) is to prove that $\tilde{\epsilon}_{\mathrm{vi}}(\cdot) \equiv \epsilon_{\mathrm{vi}}(\cdot)$. To prove the latter relation, observe that $\tilde{\epsilon}_{\mathrm{vi}}(\cdot) \geq \epsilon_{\mathrm{vi}}(\cdot)$ because $\widetilde{\Phi}$ extends Φ. Thus, all we need is to prove that whenever $x \in X$, one has $\tilde{\epsilon}_{\mathrm{vi}}(x) \leq \epsilon_{\mathrm{vi}}(x)$. Recalling the definition of $\tilde{\epsilon}_{\mathrm{vi}}$, we see that it suffices to prove that

$$\forall\,(x \in X,\ z \in X \cap \mathrm{Dom}\,\widetilde{\Phi},\ \phi \in \widetilde{\Phi}(z)):\quad \langle \phi, x - z\rangle \leq \Theta(x) := \sup_{y\in \mathrm{int}\,X}\langle \Phi(y), x - y\rangle \tag{52}$$

(note that $\Theta(x) \leq \epsilon_{\mathrm{vi}}(x)$). To prove (52), let us fix x, z, ϕ satisfying the premise of this statement and let e be such that $z + e \in \mathrm{int}\,X$. For $t \in (0, 1]$, $\theta \in (0, 1)$, let $x_{t\theta} = x + \theta(z + te - x)$; note that the latter point is a convex combination of the points z, $z + e$, x belonging to X, and the weight of the point $z + e \in \mathrm{int}\,X$ in this combination is positive so that $x_{t\theta} \in \mathrm{int}\,X \subset \mathrm{Dom}\,\Phi$. Because $\widetilde{\Phi}$ is a monotone extension of Φ, we have

$$\begin{aligned}
\langle \phi, x_{t\theta} - z\rangle &\leq \langle F(x_{t\theta}), x_{t\theta} - z\rangle = \langle F(x_{t\theta}), (1-\theta)x - (1-\theta)z + \theta t e\rangle \\
&= \left\langle F(x_{t\theta}), \frac{1-\theta}{\theta}[x - \theta(z + te) - (1-\theta)x] + te\right\rangle = \frac{1-\theta}{\theta}\langle F(x_{t\theta}), x - x_{t\theta}\rangle + t\langle F(x_{t\theta}), e\rangle \\
&\leq \frac{1-\theta}{\theta}\Theta(x) + t\langle F(x_{t\theta}), e\rangle.
\end{aligned}$$

Thus,

$$\langle \phi, x_{t\theta} - z\rangle \leq \frac{1-\theta}{\theta}\Theta(x) + t\psi(t), \quad \psi(t) = \langle F(x_{t\theta}), e\rangle. \tag{53}$$

We claim that $\psi(t) \leq \psi(1)$ when $0 < t < 1$. Indeed, we have $x_{1\theta} = x_{t\theta} + \theta(1-t)e$, whence

$$\psi(1) - \psi(t) = \langle F(x_{1\theta}) - F(x_{t\theta}), e\rangle = \frac{1}{\theta(1-t)}\langle F(x_{1\theta}) - F(x_{t\theta}), x_{1\theta} - x_{t\theta}\rangle \geq 0.$$

Because $\psi(t) \leq \psi(1)$ when $0 < t < 1$, (53) implies that

$$\langle \phi, x_{t\theta} - z\rangle \leq \frac{1-\theta}{\theta}\Theta(x) + t\psi(1), \quad 0 < t < 1;$$

passing to limits as $t \to +0$, we get

$$\langle \phi, (1-\theta)(x - z)\rangle = \langle \phi, x_{t\theta} - z\rangle \leq \frac{1-\theta}{\theta}\Theta(x),$$

whence $\langle \phi, x - z\rangle \leq (1/\theta)\Theta(x)$. This inequality holds true for every $\theta \in (0, 1)$ and the conclusion in (52) follows. □

6.5. Proof of Proposition 3.2. (i) Let $x, y \in \operatorname{int} X$ and let us prove that $\langle \Phi(x) - \Phi(y), x - y\rangle \geq 0$ or, equivalently, that $\langle \Phi(\bar{x}+\Delta) - \Phi(\bar{x}-\Delta), \Delta\rangle \geq 0$, where $\bar{x} = \frac{1}{2}(x+y)$ and $\Delta = \frac{1}{2}(x-y)$. We have

$$\begin{aligned}
\langle \Phi(\bar{x}+\Delta) - \Phi(\bar{x}-\Delta), \Delta\rangle &= \sum_i \Big[\underbrace{\langle \Phi_i(\bar{x}+\Delta), \Delta_i\rangle}_{\underset{(a)}{\geq} F_i(\bar{x}+\Delta) - F_i(\bar{x}^i+\Delta^i, \bar{x}_i)} + \underbrace{\langle \Phi_i(\bar{x}-\Delta), -\Delta_i\rangle}_{\underset{(a)}{\geq} F_i(\bar{x}-\Delta) - F_i(\bar{x}^i-\Delta^i, \bar{x}_i)} \Big] \\
&\geq \sum_i [F_i(\bar{x}+\Delta) - F_i(\bar{x}^i+\Delta^i, \bar{x}_i) + F_i(\bar{x}-\Delta) - F_i(\bar{x}^i-\Delta^i, \bar{x}_i)] \\
&= F(\bar{x}+\Delta) + F(\bar{x}-\Delta) - \sum_i \Big[\underbrace{F_i(\bar{x}^i+\Delta^i, \bar{x}_i) + F_i(\bar{x}^i-\Delta^i, \bar{x}_i)}_{\underset{(b)}{\leq} 2F_i(\bar{x})} \Big] \\
&\geq F(\bar{x}+\Delta) + F(\bar{x}-\Delta) - 2F(\bar{x}) \underset{(c)}{\geq} 0,
\end{aligned}$$

where (a) and (b) are due to the fact that $F_i(u^i, u_i)$ are convex in u_i and concave in u^i, and (c) is due to the convexity of $F = \sum_i F_i$.

(ii) Assume that the functions F_i in a convex Nash equilibrium problem are continuous on X. Then, of course, $F_i(x^i, x_i)$ are convex in $x_i \in X_i$ for every $x^i \in X^i = X_1 \times \cdots \times X_{i-1} \times X_{i+1} \times \cdots \times X_m$ and are concave in $x^i \in X^i$ for every $x_i \in X_i$. Also, $F(x) = \sum_i F_i(x)$ is convex on X.

(A) Let x_* be a Nash equilibrium and let us prove that x_* is a weak solution to the Nash VIP. Let $y \in \operatorname{int} X$, $\Delta = \frac{1}{2}(y - x_*)$, and $\bar{x} = \frac{1}{2}(y + x_*)$ so that $\bar{x} \in \operatorname{int} X$. We have

$$\begin{aligned}
\tfrac{1}{2}\langle \Phi(y), y - x_*\rangle = \langle \Phi(\bar{x}+\Delta), \Delta\rangle &= \sum_i \underbrace{\langle \Phi_i(\bar{x}+\Delta), \Delta_i\rangle}_{\underset{(a)}{\geq} F_i(\bar{x}+\Delta) - F_i(\bar{x}^i+\Delta^i, \bar{x}_i)} \\
&\geq \sum_i \Big[F_i(\bar{x}+\Delta) - F_i(\bar{x}^i+\Delta^i, \bar{x}_i) + \underbrace{F_i(\bar{x}-\Delta) - F_i(\bar{x}^i-\Delta^i, \bar{x}_i)}_{= F(x_*^i, (x_*)_i) - F_i(x_*^i, \bar{x}_i) \underset{(b)}{\leq} 0} \Big] \\
&= F(\bar{x}+\Delta) + F(\bar{x}-\Delta) - \sum_i \Big[\underbrace{F_i(\bar{x}^i+\Delta^i, \bar{x}_i) + F_i(\bar{x}^i-\Delta^i, \bar{x}_i)}_{\underset{(c)}{\leq} 2F_i(\bar{x})} \Big] \\
&\geq F(\bar{x}+\Delta) + F(\bar{x}-\Delta) - 2F(\bar{x}) \underset{(d)}{\geq} 0,
\end{aligned}$$

where (a) is due to convexity of $F_i(y^i, \cdot)$, (b) is due to the fact that $F_i(x_*^i, x_i)$ attains its minimum in $x_i \in X_i$ at the point $(x_*)_i$, (c) is due to the concavity of $F_i(x^i, \bar{x}_i)$ in $x^i \in X^i$, and (d) is due to the convexity of F. We see that $\langle \Phi(y), y - x_*\rangle \geq 0$ for all $y \in \operatorname{int} X$ so that x_* is a weak solution to the VIP in question.

(B) Now let x_* be a weak solution to the VIP and let us prove that x_* is a Nash equilibrium. Assume on the contrary that, for certain i, the function $F_i(x_*^i, x_i)$ does *not* attain its minimum over $x_i \in X_i$ at the point $(x_*)_i$; w.l.o.g., let it be the case for $i = m$. Let v be a minimizer of the convex continuous function $F_m(x_*^m, \cdot)$ on X_m. Then, the function $f(s) = F_m(x_*^m, (x_*)_m + s[v - (x_*)_m])$ is nonincreasing in $s \in [0, 1]$ and there exists $\bar{s} \in (0, 1)$ such that $f(\bar{s}) - f(1) > 0$ or, equivalently, $F_m(x_m^*, (x_*)_m + \bar{s}[v - (x_*)_m]) > F_m(x_*^m, v)$. Because $F_m(x)$ is continuous in $x \in X$, we can find $\bar{v} \in \operatorname{int} X_m$ close enough to v and also find a small enough convex neighborhood U (in X^m) of the point x_*^m such that

$$\forall u \in U: \quad F_m(u, (x_*)_m + \bar{s}[\bar{v} - (x_*)_m]) - F_m(u, \bar{v}) \geq \epsilon > 0. \tag{54}$$

Let us choose $\bar{u} \in U \cap \operatorname{int} X^m$ and let

$$x[\rho, \delta] = (\underbrace{x_*^m + \rho[\bar{u} - x_*^m]}_{\bar{u}_\rho}, \underbrace{(x_*)_m + \delta[\bar{v} - (x_*)_m]}_{v_\delta})$$

so that $x[\rho, \delta] \in \operatorname{int} X$ for $0 < \rho, \delta \leq 1$. For $1 \leq i < m$ and $0 \leq \rho < 1$, $0 < \delta \leq 1$, we have

$$\begin{aligned}
\langle \Phi_i(x[\rho, \delta]), x_i[\rho, \delta] - (x_*)_i\rangle &= \frac{\rho}{1-\rho}\langle \Phi_i(x[\rho, \delta]), \bar{u}_i - x_i[\rho, \delta]\rangle \\
&\underset{(a)}{\leq} \frac{\rho}{1-\rho}[F_i(x^i[\rho, \delta], \bar{u}_i) - F_i(x[\rho, \delta])] \leq \frac{\rho}{1-\rho} M,
\end{aligned} \tag{55}$$

where $M/2$ is an upper bound on $|F_j(x)|$ over $j = 1, \dots, m$ and $x \in X$; here, (a) is given by the convexity of F_i in $x_i \in X_i$. We further have $\bar{u}_\rho \in U$ whence, by (54),

$$\begin{aligned} -\epsilon &\geq F_m(\bar{u}_\rho, \bar{v}) - F_m(\bar{u}_\rho, (x_*)_m + \bar{s}[\bar{v} - (x_*)_m]) = F_m(x[\rho, 1]) - F_m(x[\rho, \bar{s}]) \\ &= \int_{\bar{s}}^{1} \langle \Phi_m(x[\rho, s]), \bar{v} - (x_*)_m \rangle \, ds. \end{aligned}$$

We see that there exists $\delta = \delta_\rho \in [\bar{s}, 1]$ such that $\langle \Phi_m(x[\rho, \delta_\rho]), \bar{v} - (x_*)_m \rangle \leq -\epsilon$ or, equivalently,

$$\langle \Phi_m(x[\rho, \delta_\rho]), x_m[\rho, \delta_\rho] - (x_*)_m \rangle = \delta_\rho \langle \Phi_m(x[\rho, \delta_\rho]), \bar{v} - (x_*)_m \rangle \leq -\delta_\rho \epsilon \leq -\bar{s}\epsilon.$$

Combining this relation with (55), we get

$$\langle \Phi(x[\rho, \delta_\rho]), x[\rho, \delta_\rho] - x_* \rangle \leq (m-1)\frac{\rho}{1-\rho} M - \bar{s}\epsilon.$$

For small $\rho > 0$, the right-hand side is <0 while the left-hand side is nonnegative for all $\rho \in (0, 1]$ because x_* is a weak solution to the VIP. We got a desired contradiction. □

6.6. Proof of Proposition 3.3. Given $x \in \operatorname{int} X$, $i \leq m$, and $y_i \in \operatorname{int} X$, let $f_i(t) = F_i(x^i, y_i + t(x_i - y_i)) - F_i(x^i, y_i)$. Then, f_i is a continuous convex function on $[0, 1]$ and $f_i'(t) = \langle \Phi_i(x^i, y_i + t(x_i - y_i)), x_i - y_i \rangle \leq LD$. Also setting $z = (x^i, y_i + t(x_i - y_i))$, we have

$$f_i'(t) = \frac{1}{1-t} \langle \Phi(z), x - z \rangle \leq \frac{\epsilon_{\mathrm{vi}}(x)}{1-t}.$$

Therefore,

$$F_i(x^i, x_i) - F_i(x^i, y_i) = f(1) \leq \int_0^1 \min[LD, (1-t)^{-1}\epsilon_{\mathrm{vi}}(x)] \, dt = \epsilon_{\mathrm{vi}}(x) \ln \frac{LD}{\epsilon_{\mathrm{vi}}(x)} + \epsilon_{\mathrm{vi}}(x)$$

(we have used the evident fact that $\epsilon_{\mathrm{vi}}(x) \leq LD$). We see that, for every i, one has

$$F_i(x) - \inf_{y_i \in \operatorname{int} X_i} F_i(x^i, y_i) \leq \epsilon_{\mathrm{vi}}(x) \ln \frac{LD}{\epsilon_{\mathrm{vi}}(x)} + \epsilon_{\mathrm{vi}}(x)$$

and (16) follows. □

6.7. Proof of Proposition 3.4. The inclusion $\hat{x}^\tau \in \operatorname{int} X$ is immediate because $\hat{x}^\tau$ is a convex combination of points $x^t \in \operatorname{int} X$, $t \in I_\tau$. Also, as $e_t = \Phi(x^t)$ when $t \in I_\tau$ and $\langle e_t, x^t - x \rangle \geq 0$ when $x \in X$ and $t \in J_\tau$, we have that

$$\epsilon_{\mathrm{cert}}(\xi \mid P_\tau, \mathbf{B}) = \max_{x \in \mathbf{B}} \left[\sum_{t=1}^{\tau} \xi_t \langle e_t, x^t - x \rangle \right] \geq \max_{x \in X} \left[\sum_{t \in I_\tau} \xi_t \langle \Phi(x^t), x^t - x \rangle \right]. \tag{56}$$

For $x \in X \cap \operatorname{Dom}(\Phi)$, the monotonicity of Φ implies that $\langle \Phi(x), x^t - x \rangle \leq \langle \Phi(x^t), x^t - x \rangle$ for $t = 1, \dots, \tau$ and, therefore,

$$\epsilon_{\mathrm{cert}}(\xi \mid P_\tau, \mathbf{B}) \geq \sup_{x \in X \cap \operatorname{Dom}(\Phi)} \left[\sum_{t \in I_\tau} \xi_t \langle \Phi(x), x^t - x \rangle \right] = \sup_{x \in X \cap \operatorname{Dom}(\Phi)} \langle \Phi(x), \hat{x}^\tau - x \rangle = \epsilon_{\mathrm{vi}}(\hat{x}^\tau),$$

proving (18).

Next, assume that Φ is the Nash operator of a convex NEP on $X = X_1 \times \cdots \times X_m$. Let $F_i(x^i, x_i)$ be the corresponding cost functions, $\mathcal{D} = \mathcal{D}_1 \times \cdots \times \mathcal{D}_m$ be the common domain of these functions, and $F(x)$ be the function $\sum_{i=1}^m F_i(x)$. In particular, $\operatorname{int} X_i \subseteq \mathcal{D}_i \subseteq X_i$ for all i. Because each function $F_i([x^t]^i, \cdot)$ is convex on $\mathcal{D}_i \supseteq \operatorname{int} X_i$ and $\Phi_i(x^t)$, $t \in I_\tau$ is a subgradient of this function at the point $[x^t]_i \in \operatorname{int} X_i$, we have that, for each $x \in \mathcal{D}$ and $t \in I_\tau$,

$$\begin{aligned} \langle \Phi(x^t), x^t - x \rangle &= \sum_{i=1}^m \langle \Phi_i([x^t]^i, [x^t]_i), [x^t]_i - x_i \rangle \\ &\geq \sum_{i=1}^m [F_i([x^t]^i, [x^t]_i) - F_i([x^t]^i, x_i)] = F(x^t) - \sum_{i=1}^m F_i([x^t]^i, x_i). \end{aligned}$$

The resulting inequality combines with (56) to imply that

$$\epsilon_{\text{cert}}(\xi \mid P_\tau, \mathbf{B}) \geq \sup_{x\in\mathcal{D}} \sum_{t\in I_\tau} \xi_t \left[F(x^t) - \sum_{i=1}^m F_i([x^t]^i, x_i) \right]. \tag{57}$$

Because F is convex on $\operatorname{int} X$, we have

$$\sum_{t\in I_\tau} \xi_t F(x^t) \geq F(\hat{x}^\tau) = \sum_i F_i(\hat{x}^\tau).$$

Also, because for every i, the function $F_i(\cdot, x_i)$ is concave on $\operatorname{int}(X_1 \times \cdots \times X_{i-1} \times X_{i+1} \times \cdots \times X_m)$, we have

$$-\sum_{t\in I_\tau} \xi_t F_i([x^t]^i, x_i) \geq -F_i([\hat{x}^\tau]^i, x_i).$$

Therefore, (57) implies that

$$\epsilon_{\text{cert}}(\xi \mid P_\tau, \mathbf{B}) \geq \sup_{x\in\mathcal{D}} \sum_{i=1}^m [F_i(\hat{x}^\tau) - F_i([\hat{x}^\tau]^i, x_i)] = \epsilon_{\text{N}}(\hat{x}^\tau),$$

which proves (19). □

6.8. Proof of Proposition 3.5. 1^0. We start with the following, important by its own right, result on monotone extensions of bounded monotone operators.

LEMMA 6.1. *Let $Y = \{x^1, \ldots, x^N\} \subset \mathcal{R}^n$, let $\Phi\colon Y \to \mathcal{R}^n$ be monotone (i.e., $\langle \Phi(x_i) - \Phi(x_j), x_i - x_j\rangle \geq 0$ for all i, j) and bounded by L (i.e., $\|\Phi(x_i)\|_2 \leq L$ for all i), and let $x \in \mathcal{R}^n \setminus Y$. Then, Φ can be extended to a monotone and bounded by L mapping of $Y \cup \{x\}$.*

This statement is very close to one of the results of Minty [15]; to make the paper self-contained, we present a proof.

PROOF. For $j = 1, \ldots, N$, let $F_j = \Phi(x_j)$ and let $\Omega = \{\xi \in \mathcal{R}^N\colon \xi \geq 0, \sum_{j=1}^N \xi_j = 1\}$. For $\xi \in \Omega$, let $F_\xi = \sum_{j=1}^N \xi_j F_j$ and $x_\xi = \sum_{j=1}^N \xi_j x_j$. For each j, $\|F_j\|_2 \leq L$ and, for each $\xi \in \Omega$, $\|F_\xi\|_2 \leq L$. From the monotonicity of Φ on Y,

$$0 \leq \sum_j \sum_s \xi_j \xi_s \langle F_j - F_s, x_j - x_s\rangle = \sum_j \xi_j \langle F_j, x_j\rangle + \sum_s \xi_s \langle F_s, x_s\rangle - 2\langle F_\xi, x_\xi\rangle,$$

assuring that $\sum_j \xi_j \langle F_j, x_j\rangle \geq \langle F_\xi, x_\xi\rangle$. Consequently, for each $F \in \mathcal{R}^N$,

$$\langle F, x\rangle - \langle F, x_\xi\rangle - \langle F_\xi, x\rangle + \sum_{j=1}^N \xi_j \langle F_j, x_j\rangle \geq \langle F - F_\xi, x - x_\xi\rangle. \tag{58}$$

The assertion of Lemma 6.1 is equivalent to the statement that the quantity

$$A \equiv \max_{\|F\|_2\leq L} \min_j \langle F - F_j, x - x_j\rangle$$

is nonnegative; given this fact, the required monotone extension of F from Y onto $Y \cup \{x\}$ is given by $\Phi(x) = F_*$, where F_* is the maximizer in the maximization yielding A. We have

$$\begin{aligned}
A &= \max_{\|F\|_2\leq L} \min_j \langle F - F_j, x - x_j\rangle \\
&= \max_{\|F\|_2\leq L} \min_{\xi\in\Omega} \left[\sum_{j=1}^N \xi_j \langle F - F_j, x - x_j\rangle \right] \\
&= \min_{\xi\in\Omega} \max_{\|F\|_2\leq L} \left[\sum_{j=1}^N \xi_j \langle F - F_j, x - x_j\rangle \right] \qquad \left[\begin{array}{l} \text{Here ``max'' and ``min'' can be exchanged as} \\ \text{the bracketed term is bilinear in } \xi \text{ and } F \text{ and} \\ \{F\colon \|F\|_2 \leq L\},\ \Omega \text{ are convex compact sets.} \end{array} \right] \\
&= \min_{\xi\in\Omega} \max_{\|F\|_2\leq L} \left[\langle F, x\rangle - \langle F, x_\xi\rangle - \langle F_\xi, x\rangle + \sum_{j=1}^N \xi_j \langle F_j, x_j\rangle \right] \\
&\geq \min_{\xi\in\Omega} \max_{\|F\|_2\leq L} \langle F - F_\xi, x - x_\xi\rangle \qquad [\text{by (58)}] \\
&\geq \min_{\xi\in\Omega} \langle F_\xi - F_\xi, x - x_\xi\rangle = 0. \qquad [\text{as } \|F_\xi\|_2 \leq L \text{ for each } \xi \in \Omega]. \quad \square
\end{aligned}$$

LEMMA 6.2. *Let L be a positive real, $X \subset \mathcal{R}^n$ be a solid, $Y \subset X$ be a nonempty set, and Φ: $Y \to \mathcal{R}^n$ be monotone and bounded by L. Then, there exists a bounded-by-L monotone extension of Φ from Y onto the entire X.*

PROOF. Given a monotone and bounded-by-L mapping Φ: $Y \to \mathcal{R}^n$ with $Y \subset X$, let us look at the set $\mathcal{Y}$ of all pairs (Y', Φ') with $Y \subset Y' \subset X$ and Φ' being a monotone and bounded-by-L mapping on Y', which coincides with Φ on Y. Introducing partial order on these pairs

$$[(Y', \Phi') \succeq (Y'', \Phi'')] \quad \Leftrightarrow \quad [Y' \supset Y'' \text{ and } \Phi'\big|_{Y''} = \Phi'']$$

and applying Zorn lemma, there exists a $\succeq$-maximal pair $(\widehat{Y}, \widehat{\Phi})$. All we need to prove is that $\widehat{Y} = X$. Indeed, assume otherwise and let $x \in X \setminus \widehat{Y}$. For every point $y \in \widehat{Y}$, the set $\mathcal{F}_y = \{F\colon \|F\|_2 \le L, \langle F - \widehat{\Phi}(y), x - y\rangle \ge 0\}$ is compact. Lemma 6.1 shows that each finite family of sets of this type has a nonempty intersection whence by standard compactness arguments, the intersection of all these sets is nonempty. Let F be a point in the intersection. Extending $\widehat{\Phi}$ from $\widehat{Y}$ to $Y^+ = \widehat{Y} \cup \{x\}$ by setting $\widehat{\Phi}(x) = F$, we get a pair in $\mathcal{Y}$ that is $\succ (\widehat{Y}, \widehat{\Phi})$, a contradiction that proves that $\widehat{Y} = X$. □

2^0. In the situation described in Proposition 3.5, let $P_\tau = \{I_\tau, J_\tau, \{x^t, e_t\}_{t=1}^\tau\}$. Note that $I_\tau \ne \varnothing$ because the approximate solution $\hat{x}$ generated by the algorithm as applied to (X, Φ) belongs to $\operatorname{int} X$ on one hand, and is a point from the trajectory $x^1, \dots, x^\tau$ on the other hand. Let us set $\widehat{X} = \{x \in \mathbf{B}\colon \langle e_t, x - x^t\rangle \le 0,\ t \in J_\tau\}$. As in the proof of Proposition 2.2, setting

$$\delta = \max_{x \in \widehat{X}} \min_{t \in I_\tau} \langle \Phi(x_t), x^t - x\rangle, \tag{59}$$

there exists a certificate ξ for P_τ such that

$$\epsilon_{\text{cert}}(\xi \mid P_\tau, \mathbf{B}) = \delta.$$

In fact, this certificate is given by a solution to the problem

$$\max_{\xi \ge 0,\ \sum_{t \in I_\tau} \xi_t = 1} \left\{ \min_{x \in \mathbf{B}} \left[\sum_{t \in I_\tau} \xi_t \langle e_t, x - x^t\rangle + \sum_{t \in J_\tau} \xi_t \langle e_t, x - x^t\rangle \right] \right\}. \tag{60}$$

The conclusion in Proposition 3.5 is trivially true when $\delta \le 0$ so let us assume that $\delta > 0$. Let $Y = \{x^t\colon t \in I_\tau\}$. The function $x \to \min_{y \in Y} \langle \Phi(y), y - x\rangle$ is continuous and thus attains its maximum over $\widehat{X}$, say, at x_*. Evidently, $x_* \notin Y$ (because, otherwise, $\delta = \min_{y \in Y} \langle \Phi(y), y - x_*\rangle \le 0$) and, because $\hat{x} \in \{x^1, \dots, x^\tau\} \cap \operatorname{int} X$, we have

$$\delta \le \langle \Phi(x^\tau), \hat{x} - x_*\rangle \le \|\Phi(\hat{x})\|_2 \|\hat{x} - x_*\|_2 \le L\|\hat{x} - x_*\|_2.$$

In particular,

$$D \ge \|\hat{x} - x_*\|_2 \ge \frac{\delta}{L}. \tag{61}$$

Next, set

$$F \equiv \delta \frac{\hat{x} - x_*}{D\|\hat{x} - x_*\|_2}$$

and note that $\|F\|_2 \le \delta/D \le L$ and, for each $y \in Y$,

$$\langle F, y - x_*\rangle \le \|F\|_2 \|y - x_*\|_2 \le \frac{\delta}{D} D = \delta \le \langle \Phi(y), y - x_*\rangle$$

(the last inequality follows from the definition of δ). We see that extending Φ from the set Y onto the set $Y \cup \{x_*\}$ by mapping x_* to F preserves monotonicity and boundedness by L. Lemma 6.2 assures that we can further extend the resulting mapping from $Y \cup \{x_*\}$ onto the entire $\widehat{X}$ to get a monotone and bounded-by-L operator $\widehat{\Phi}$ on the entire $\widehat{X}$. As

$$\langle \widehat{\Phi}(x_*), \hat{x} - x_*\rangle = \langle F, \hat{x} - x_*\rangle = \frac{\delta\|\hat{x} - x_*\|_2^2}{D\|\hat{x} - x_*\|} = \frac{\delta\|\hat{x} - x_*\|_2}{D} \ge \frac{\delta^2}{LD} \tag{62}$$

(we have used (61)), we have that $\epsilon_{\text{vi}}(\hat{x} \mid \widehat{X}, \widehat{\Phi}) \ge \delta^2/(LD)$.

Now, the a priori information on (X, Φ) and the information accumulated by the algorithm in question upon termination do not contradict the assumption that $X = \widehat{X}$ and $\Phi = \widehat{\Phi}$, whence $\delta^2/(LD) \le \epsilon_{\text{vi}}(\hat{x} \mid \widehat{X}, \widehat{\Phi}) \le \epsilon$. Thus, $\epsilon_{\text{cert}}(\xi \mid P_\tau, \Phi) = \delta \le \sqrt{LD\epsilon}$. □

6.9. Proof of Proposition 4.1. The inclusion in (25) is evident due to the inclusion in (24) ("the larger is the set, the less is its polar"). The equality in (25) is given by the following simple statement.

LEMMA 6.3. *Let P, Q be two closed convex sets in E^+ containing the origin and such that P is a cone, and let* $\operatorname{int} P \cap \operatorname{int} Q \neq \varnothing$. *Then,* $\operatorname{Polar}(P \cap Q) = \operatorname{Polar}(Q) + P_*$, *where* $P_* = \{z \in E^+ : \langle z, u\rangle \leq 0 \ \forall u \in P\}$.

PROOF OF LEMMA 6.3. By standard facts on polars, $\operatorname{Polar}(P \cap Q) = \operatorname{cl}(\operatorname{Conv}(\operatorname{Polar}(P) \cup \operatorname{Polar}(Q)))$. In our case, $\operatorname{Polar}(P) = P_*$; because P_* is a cone and $\operatorname{Polar}(Q)$ contains the origin, the convex hull of the union of P_* and $\operatorname{Polar}(Q)$ is dense in the arithmetic sum of these two sets, that is, $\operatorname{Polar}(P \cap Q) = \operatorname{cl}(P_* + \operatorname{Polar}(Q))$. It follows that all we should verify in order to prove Lemma 6.3 is that the set $P_* + \operatorname{Polar}(Q)$ is closed. However, this is immediate: Let $u_i \in P_*$ and $v_i \in \operatorname{Polar}(Q)$ be such that $u_i + v_i \to w$, $i \to \infty$. We should prove that $w \in P_* + \operatorname{Polar}(Q)$. To this end, it is clearly enough to verify that the sequences $\{u_i\}$ and $\{v_i\}$ are bounded. The latter is nearly evident: By assumption, there exists a ball B of a positive radius that is contained in both P and Q. The quantities $s_i = \min_{x \in B}\langle u_i + w_i, x\rangle$ form a bounded sequence due to $u_i + v_i \to w$, and $\langle u_i, x\rangle \leq 1$, $\langle v_i, x\rangle \leq 1$ for all $x \in B$ due to $B \subset P \cap Q$. Therefore, we have

$$\min_{x \in B}\langle u_i, x\rangle = \min_{x \in B}[\langle u_i + v_i, x\rangle - \langle v_i, x\rangle] \geq s_i - 1,$$

that is, the sequence $\{\min_{x \in B}\langle u_i, x\rangle\}_{i=1}^{\infty}$ is below bounded. Because the sequence $\{\max_{x \in B}\langle u_i, x\rangle\}_{i=1}^{\infty}$ is above bounded by one, we conclude that the sequence $\{\max_{x \in B}\langle u_i, x\rangle - \min_{x \in B}\langle u_i, x\rangle\}_{i=1}^{\infty}$ is bounded. Recalling that B is a ball of positive radius, we see that the sequence $\{u_i\}_{i=1}^{\infty}$ is bounded. Via the boundedness of the sequence $\{u_i + v_i\}_{i=1}^{\infty}$, this implies that $\{v_i\}_{i=1}^{\infty}$ is bounded as well. □

FROM LEMMA 6.3 TO EQUALITY IN (25). We have $\hat{Q}_{t+1}^+ = Q_t^+ \cap P$, where P is the cone $P = \{z \in E^+ : \langle e_t^+, z\rangle \leq 0\}$. Because $x^t \in \operatorname{int} Q_t$ and $e_t \neq 0$, the interiors of P and Q_t^+ have a nonempty intersection so that by Lemma 6.3 $\operatorname{Polar}(\hat{Q}_{t+1}^+) = \operatorname{Polar}(Q_t^+) + P_* = \operatorname{Polar}(Q_t^+) + \mathcal{R}_+ e_t^+$. □

6.10. Proof of Theorem 4.1. (i) When τ is terminal, the validity of (i) is evident. Now, let τ be nonterminal with well-defined ξ. The fact that ξ is a certificate for P_τ is evident from the construction; all we need is to verify (29) and (30). Let $x \in Q_1$ and let $x^+ = [x; 1]$ so that $x^+ \in Q_1^+$. Summing up relations (28.a–b) and dividing both sides of the resulting equality by d_τ, we obtain

$$\sum_{t=1}^{\tau} \xi_t e_t^+ = -d_\tau^{-1}(\phi^+ + \psi^+).$$

Multiplying both sides by x^+, we obtain

$$\sum_{t=1}^{\tau} \xi_t \langle e_t, x - x^t\rangle = \sum_{t=1}^{\tau} \xi_t \langle e_t^+, x^+\rangle = -d_\tau^{-1}\langle \phi^+ + \psi^+, x^+\rangle \geq -2d_\tau^{-1},$$

where the concluding inequality is given by the inclusions $\phi^+, \psi^+ \in \operatorname{Polar}(Q_1^+)$, and $x^+ \in Q_1^+$. The resulting inequality holds true for every $x \in Q_1$ and (29) follows.

To prove (30), let τ be such that $\epsilon \equiv \epsilon_\tau < r$ and let $\bar{x}$ be the center of Euclidean ball B of the radius r that is contained in X. Observe that

$$t \in I_\tau \ \Rightarrow \ \langle e_t, x - x^t\rangle \leq W_\tau \ \ \forall x \in B,$$

whence $\langle e_t, \bar{x} - x^t\rangle \leq W_\tau - r\|e_t\|_2$. Recalling what is e_t for $t \in J_\tau \equiv \{1, \ldots, \tau\} \backslash I_\tau$, we get the relations

$$(P_t): \quad \langle e_t, \bar{x} - x^t\rangle \leq \begin{cases} W_\tau - r\|e_t\|_2, & t \in I_\tau \\ 0, & t \in J_\tau. \end{cases}$$

Now, let $x \in Q_1$ and let $y = ((r - \epsilon)x + \epsilon\bar{x})/r$ so that $y \in Q_1$. By (29), we have

$$\sum_{t=1}^{\tau} [\lambda_t + \mu_t]\langle e_t, x^t - y\rangle \leq 2.$$

Multiplying this inequality by r and adding the weighted sum of inequalities (P_t), the weights being $[\lambda_t + \mu_t]\epsilon$, we obtain

$$\sum_{t=1}^{\tau} [\lambda_t + \mu_t]\langle e_t, \underbrace{rx^t - ry + \epsilon\bar{x} - \epsilon x^t}_{(r-\epsilon)(x^t - x)}\rangle \leq 2r + \epsilon W_\tau d_\tau - r\epsilon D_\tau.$$

The right-hand side in this inequality, by the definition of ϵ, is $\epsilon W_\tau d_\tau$ and we arrive at the relation

$$(r-\epsilon)\sum_{t=1}^{\tau}[\lambda_t+\mu_t]\langle e_t, x^t-x\rangle \le \epsilon W_\tau d_\tau \quad\Leftrightarrow\quad \sum_{t=1}^{\tau}\xi_t\langle e_t, x^t-x\rangle \le \frac{\epsilon W_\tau}{r-\epsilon}.$$

This relation holds true for every $x\in Q_1$ and (30) follows. (i) is proved.

(ii) As above, let $\bar{x}$ be the center of a Euclidean ball B of radius r that is contained in X. Consider first the case when $\langle h, \bar{x}-x^{\tau+1}\rangle \ge 0$. Multiplying both sides in (28.a) by $x^+=[\bar{x}+re;1]$ with $e\in\mathcal{R}^n$, $\|e\|_2\le 1$, we obtain

$$\begin{aligned}
\langle h, re\rangle &\le \langle h, re\rangle + \langle h, \bar{x}-x^{\tau+1}\rangle = \langle h^+, x^+\rangle = \sum_{t=1}^{\tau}\lambda_t\langle e_t, \bar{x}+re-x^t\rangle + \langle \phi^+, x^+\rangle\\
&\le \sum_{t=1}^{\tau}\lambda_t\langle e_t, \bar{x}+re-x^t\rangle + 1 \quad [\text{because } x^+\in Q_1^+,\ \phi^+\in \mathrm{Polar}(Q_1^+)]\\
&\le 1+\sum_{t\in I_\tau}\lambda_t\langle e_t, \bar{x}+re-x^t\rangle \quad [\text{because } \bar{x}+re\in X \text{ and } e_t \text{ separates } x^t \text{ and } X \text{ for } t\in J_\tau]\\
&\le 1+\sum_{t\in I_\tau}\lambda_t\|e_t\|_2 D(Q_1) \le 1+D_\tau D(Q_1).
\end{aligned}$$

The resulting inequality holds true for all unit vectors e; maximizing the left-hand side over these e, we obtain $D_\tau \ge (r\|h\|_2-1)/(D(Q_1))$. Recalling that $\|h\|_2\ge 1/(2\chi\rho(Q_{\tau+1}))$, we arrive at (31). We have established (ii) in the case of $\langle h, \bar{x}-x^{\tau+1}\rangle\ge 0$; in the opposite case, we can use the same reasoning with $-h$ in the role of h and with (28.b) in the role of (28.a). (ii) is proved.

(iii) This is an immediate consequence of (i), (ii), and the evident fact that whenever Φ is semibounded on $\mathrm{int}\, X$, we have $W_\tau \le \mathrm{Var}_X(\Phi)$ for all τ. □

6.11. Proof of Proposition 5.1. Setting $\epsilon=\epsilon_{\mathrm{cert}}(\xi\mid P_\tau, X)$, $\bar{x}=\sum_{t\in I_\tau}\xi_t x_t$, and $u_t=u_{x_t}$, $t\in I_\tau$, we have $e_t=b-Au_t$, $t\in I_\tau$ and

$$\begin{aligned}
\forall x\in X:\quad & \sum_{t\in I_\tau}\xi_t\langle e_t, x_t-x\rangle \le \sum_{t=1}^{\tau}\xi_t\langle e_t, x_t-x\rangle \le \epsilon\\
\Rightarrow\quad & \sum_{t\in I_\tau}\xi_t\langle b-Au_t, x_t\rangle + \max_{x\in X}\Big\langle \underbrace{-\sum_{t\in I_\tau}\xi_t(b-Au_t)}_{=A\hat{u}-b}, x\Big\rangle \le \epsilon\\
\Rightarrow\quad & 2L\|[A\hat{u}-b]^+\|_q + \sum_{t\in I_\tau}\xi_t\langle b-Au_t, x_t\rangle \le \epsilon. \qquad (*)
\end{aligned}$$

Further, by the definition of u_t we have for every $u\in U$:

$$f(u_t)+\langle Au_t-b, x_t\rangle \le f(u)+\langle Au-b, x_t\rangle, \quad t\in I_\tau.$$

Multiplying tth inequality by ξ_t, summing up and taking into account that $\sum_{t\in I_\tau}\xi_t f(u_t)\ge f(\hat{u})$, we obtain

$$f(\hat{u})-f(u)+\langle b-Au, \bar{x}\rangle \le \sum_{t\in I_\tau}\xi_t\langle b-Au_t, x_t\rangle.$$

This inequality combines with $(*)$ to imply that

$$2L\|[A\hat{u}-b]^+\|_q + f(\hat{u})-f(u_*)+\langle b-Au_*, \bar{x}\rangle \le \epsilon, \qquad (63)$$

where u_* is an optimal solution to (42). Taking into account that $\bar{x}\in X$ and thus $\bar{x}\ge 0$, whence $\langle b-Au_*, \bar{x}\rangle\ge 0$, we arrive at (46.b). Further, we have $f(u_*)\le f(u)+\langle x_*, Au-b\rangle$ for all $u\in U$ due to the origin of x_*, whence $f(\hat{u})\ge f(u_*)+\langle x_*, b-A\hat{u}\rangle$, and the latter quantity is $\ge f(u_*)-L\|[A\hat{u}-b]^+\|_q$ due to $x_*\ge 0$ and $\|x_*\|_p\le L$. Thus, $f(\hat{u})\ge f(u_*)-L\|[A\hat{u}-b]^+\|_q$, which combines with (63) to imply (46.a). □

Acknowledgments. The authors gratefully acknowledge the support of several institutions. Specifically, the research of the first author was partly supported by the National Science Foundation (DMI Grant 0619977); the research of the second and the third authors was partly supported by grants of the Israel Science Foundation, Technion VPR grants, and the Fund for the Promotion of Research at Technion.

References

[1] Auslender, A. 1973. Resolution numerique d'inegalites variationnelles. *RAIRO—Oper. Res.* **7** 67–72.
[2] Ben-Tal, A., A. Nemirovski. 2001. *Lectures on Modern Convex Optimization*. SIAM, Philadelphia.
[3] Ben-Tal, A., A. Nemirovski. 2005. Non-Euclidean restricted memory level method for large-scale convex optimization. *Math. Programming* **102** 407–456.
[4] Boyd, S., L. Vandenberghe. 2004. *Convex Optimization*. Cambridge University Press, Cambridge, UK.
[5] Bulatov, V. P., L. O. Shepot'ko. 1982. Method of centers of orthogonal simplexes for solving convex programming problems (in Russian). *Methods of Optimization and Their Application*. Nauka, Novosibirsk.
[6] Burrell, B. P., M. J. Todd. 1985. Ellipsoid method generates dual variables. *Math. Oper. Res.* **10** 688–700.
[7] Facchinei, F., J.-S. Pang. 2003. *Finite-Dimensional Variational Inequalities and Complementarity Problems*. Springer-Verlag, New York.
[8] Gabelev, D. 2003. Polynomial time cutting plane algorithms associated with symmetric cones. M.Sc. thesis in OR and Systems Analysis, Faculty of Industrial Engineering and Management, Technion—Israel Institute of Technology, Technion City, Haifa, Israel, http://www2.isye.gatech.edu/∼nemirovs/Dima.pdf.
[9] Grötschel, M., L. Lovasz, A. Schrijver. 1986. *The Ellipsoid Method and Combinatorial Optimization*. Springer-Verlag, New York.
[10] Harker, P. T., J.-S. Pang. 1990. Finite-dimensional variational inequality and nonlinear complementarity problems: A survey of theory, algorithms and applications. *Math. Programming* **48** 161–220.
[11] John, F. 1948. *Extremum Problems with Inequalities as Subsidiary Conditions*. (Studies and essays presented to R. Courant on his 60th birthday, January 8). Interscience Publishers, Inc., New York, 187–204.
[12] Khachiyan, L. G. 1979. A polynomial algorithm in linear programming. *Soviet Math. Doklady* **20** 191–194.
[13] Lemarechal, C., A. Nemirovski, Yu. Nesterov. 1995. New variants of bundle methods. *Math. Programming* **69** 111–148.
[14] Levin, A. Yu. 1965. On an algorithm for the minimization of convex functions (in Russian). *Doklady Akad. Nauk SSSR* **160** 1244–1247. (English translation: *Soviet Math. Doklady* **6** 286–290.)
[15] Minty, G. J. 1962. Monotone non-linear operators in Hilbert space. *Duke Math. J.* **29** 341–346.
[16] Nemirovski, A. 1996. Polynomial time methods in convex programming. J. Renegar, M. Shub, S. Smale, eds. *The Mathematics of Numerical Analysis*, *AMS-SIAM Summer Seminar on Mathematics in Applied Mathematics*, *Lectures in Applied Mathematics*, Vol. 32. AMS, Providence, RI, 543–589.
[17] Nemirovski, A. 2004. Prox-method with rate of convergence $O(1/t)$ for variational inequalities with Lipschitz continuous monotone operators and smooth convex-concave saddle point problems. *SIAM J. Optim.* **15** 229–251.
[18] Nemirovski, A., D. Yudin. 1979. *Problem Complexity and Method Efficiency in Optimization*. Nauka Publishers, Moscow (in Russian). (English translation: John Wiley & Sons, 1983.)
[19] Nemirovskii, A. 1981. Efficient iterative algorithms for variational inequalities with monotone operators (in Russian). *Ekonomika i Matematicheskie Metody* **17** 344–359.
[20] Nesterov, Y. U. 2005. Smooth minimization of non-smooth functions. *Math. Programming* **103** 127–152.
[21] Nesterov, Y. U., A. Nemirovskii. 1994. *Interior-Point Polynomial Algorithms in Convex Programming*. SIAM, Philadelphia.
[22] Newman, D. J. 1965. Location of maximum on unimodal surfaces. *J. Assoc. Comput. Machinery* **12** 11–23.
[23] Nikaido, H., K. Isoda. 1955. Note on noncooperative convex games. *Pacific J. Math.* **5** 807–815.
[24] Shor, N. Z. 1977. Cutting plane method with space dilation for the solution of convex programming problems (in Russian). *Kibernetika* **1** 94–95.
[25] Tarasov, S. P., L. G. Khachiyan, I. I. Erlikh. 1988. The method of inscribed ellipsoids. *Soviet Math. Doklady* **37** 226–230.
[26] Yamnitsky, B., L. Levin. 1982. An old linear programming algorithm runs in polynomoial time. *Proc. 23rd Annual Sympos. Foundations Comput. Sci.*, IEEE, New York, 327–328.
[27] Yudin, D., A. Nemirovskii. 1976. Informational complexity and effective methods of solution for convex extremal problems (in Russian). *Ekonomika i Matematicheskie Metody* **12** 357–369. (English translation: *Matekon: Transl. Russian East European Math. Economics* **13** (1977), 25–45.)