Revised Nov. 2008

ACCURACY CERTIFICATES FOR COMPUTATIONAL PROBLEMS WITH CONVEX STRUCTURE

by

Arkadi Nemirovski[1], **Shmuel Onn**[2] **and Uriel G. Rothblum**[3]

November 26, 2008

Abstract

The goal of the current 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 essential increase of the computational effort. Further, we show that (computable) certificates exist whenever an algorithm is capable to produce solutions of guaranteed accuracy.

Key words: convexity, certificates, computation in convex structures, convex minimization, variational inequalities, saddle points, convex Nash equilibrium

[1]Supported in part by NSF grant DMI 0619977

[2]Supported in part by a grant from ISF - the Israel Science Foundation, by a VPR grant at the Technion and by the Fund for the Promotion of Research at the Technion

[3]Supported in part by a grant from ISF - the Israel Science Foundation, by a VPR grant at the Technion and by the Fund for the Promotion of Research at the Technion

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 $F : X \to \mathcal{R}\cup\{+\infty\}$ is closed convex function finite on $\mathrm{int}\, X$. Assume that the problem is *black box represented*, that is,

- X is given by a Separation oracle, a routine which, given on input an $x \in \mathcal{R}^n$, reports whether $x \in \mathrm{int}\, X$, and if it is not the case, returns a *separator* – a vector e such that $\langle e, y\rangle < \langle e, x\rangle$ whenever $y \in \mathrm{int}\, X$. Besides, we are given a "simple" solid $\mathbf{B}$ known to contain X, simplicity meaning that it is easy to minimize linear forms over $\mathbf{B}$ (as it is the case when $\mathbf{B}$ is an ellipsoid, a parallelotope or a simplex);
- F is given by a First order oracle, a routine which, given on input a point $x \in \mathrm{int}\, X$, 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 $\widehat{x}$ such that

$$\widehat{x} \in X \ \&\ \epsilon_{\mathrm{opt}}(\widehat{x}) := F(\widehat{x}) - \mathrm{Opt} \le \epsilon \tag{2}$$

(why we quantify the accuracy in terms of F, and not "in the argument" – by the distance from $\widehat{x}$ to the set of minimizers of F on X – this will be explained later). In the outlined "computational environment", an algorithm for achieving our goal is a routine which, given on input an $\epsilon > 0$, after finitely many calls to the oracles terminates and outputs a vector $\widehat{x} \in X$ satisfying (2). This should be so for *every* $\epsilon > 0$ and *every* CMP (1) from a given in advance family $\mathcal{F}$ of CMP's we are interested in, and, besides, the solution process should be "non-anticipative" – the subsequent 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 $\widehat{x}$ built upon termination.

The issue of primary interest in this paper is how could an algorithm *certify* the target relation (2), and here is the proposed answer. Assume that when solving CMP (1), the algorithm has queried the oracles at the points $x^1, ..., x^\tau$, 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 $x^t \in \mathrm{int}\, X$, $J_\tau = \{1, ..., \tau\}\backslash I_\tau$ is the set of indices of *non-productive* 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 $\mathrm{int}\, X$ reported by the Separation oracle (this is so when the step t is non-productive). 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 $I_\tau \neq \emptyset$, 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

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

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

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

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

The above observation indeed is immediate. Given $x \in X$, we have $F(x) \geq F(x^t) + \langle e_t, x - x^t \rangle$ whenever $t \in I_\tau$ (by convexity of F) and $0 \geq \langle e_t, x - x^t \rangle$ whenever $t \in J_\tau$. Summing up these inequalities, the weights being ξ_t, we get $F(x) \geq \sum_{t \in I_\tau} \xi_t F(x^t) - \sum_{t=1}^{\tau} \xi_t \langle e_t, x^t - x \rangle$. Taking minima over $x \in X$, we get $\mathrm{Opt} \geq \sum_{t \in I_\tau} \xi_t F(x^t) - \sup_{x \in X} \sum_{t=1}^{\tau} \xi_t \langle e_t, x^t - x \rangle$. The right hand side in this inequality is $\geq F_*(\xi | P_\tau, \mathbf{B})$ due to $X \subset \mathbf{B}$, so that $F_*(\xi | P_\tau, \mathbf{B})$ indeed is a lower bound on Opt.

Note that given 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 easy to verify whether ξ is as stated above (i.e., nonnegative with $\sum_{t \in I_\tau} \xi_t = 1$), and if its is the case, $F_*(\xi | P_\tau, \mathbf{B})$ is easy to compute[4]. 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 $\epsilon_{\mathrm{opt}}(\widehat{x}) \leq F(\widehat{x}) - F_*(\xi | P_\tau, \mathbf{B})$.

A close inspection shows that for some important families of CMP's, like the family $\mathcal{F}_{\mathrm{c}}(\mathbf{B})$ of *all* problems (1) with solid feasible domains $X \subset \mathbf{B}$ and finite continuous on X convex objectives F, the outlined way to certify the accuracy guarantees is *a necessity rather than an option* (see Proposition 2.2 below). In other words, an algorithm for solving CMP's from $\mathcal{F}_{\mathrm{c}}(\mathbf{B})$ which is "enough intelligent" 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 $\widehat{x}$ with an accuracy certificate ξ satisfying (4). Some of the existing "intelligent" algorithms (different variants of subgradient and bundle methods, see, e.g., [11, 2] and references therein) produce such certificates explicitly. Other black-box oriented intelligent algorithms, like 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, like polynomial time interior point methods (ipm's) for linear, conic quadratic and semidefinite programming, do produce explicit accuracy certificates, but of a type different from the one just explained. Specifically, ipm's work with "well-structured" convex programs rather than with black-box represented ones, and heavily exploit the specific and known in advance problem's structure (usually, the one of a primal-dual pair of conic problems on a simple cone) to build in parallel approximate solutions and "structure-specific" accuracy certificates for them (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 CMP's – with computationally cheap rules for on-line building of explicit accuracy certificates justifying the standard polynomial time complexity bounds of the algorithms.

For example, given $R < \infty$, consider the family $\mathcal{F}$ of all problems (1) with solids X contained in $\mathbf{B} = \{x \in \mathcal{R}^n : \|x\|_2 \leq R\}$ and convex and bounded on X objectives F. It is known that the Ellipsoid method is capable to solve a problem from this family within (any) given accuracy $\epsilon > 0$ it 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 oracle's answer; here $r(X)$ is the largest of the radii of Euclidean balls contained in

[4] since 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 $\mathbf{B}$.

X, and $V(F) = \max_X F - \min_X F$ is the variation of the objective on the feasible 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 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, and thus cannot equip the method with termination rules which guarantee finite termination with approximate solution of the required accuracy on every problem from the family. A byproduct of the main result of this paper is that the Ellipsoid method can be equipped with on-line rules for building accuracy certificates which 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 these goals might cause. First, a "practically oriented" reader could say that we are looking for a remedy for an illness which by itself is of no practical interest, since 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 than 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 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, while the direct applied potential of polynomial time cutting plane algorithms indeed is minor, these algorithms definitely are of a major theoretical importance, since these are the algorithms underlying the most general theoretical results on polynomial time solvability of convex programming problems. And 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-existing: there are situations when we are interested in a high-accuracy solution to a convex program with small (just few tens) number of variables a large number of constraints, and in these situations the methods like the Ellipsoid algorithm are the methods of choice. And 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 Section 5.2.

The second 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 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 be, by itself means nothing; and 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 so. All 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 on line.* Here is one of precise versions of this vague (what does "in general" mean?) claim. Consider the family $\mathcal{F}$ of all problems (1) with common domain $X = \{x \in \mathcal{R}^2 : x^T x \leq 1\}$ and objectives F which 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 that for a not too small C_d, this family does *not* allow for a non-anticipative 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 which is at the distance $\leq 1/4$ from the optimal set of the problem. Note that there are many algorithms, say, the simplest gradient descent, which, as applied to a whatever problem from the family, converge to the optimal set. The above theorem says that this convergence is completely "unobservable" on line: it is impossible to convert, in a non-anticipating reliable fashion, the accumulated so far information into a point which is at the distance $\leq 1/4$ from the optimal set, and this is so 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 $O(1)\ln(/\epsilon)$ calls to the (usual) First order oracle, each call accompanied by just $O(1)$ arithmetic operations.

Of course, there are cases in convex minimization where the inaccuracy in the argument is "observable on line". The simplest example here is when the objectives in the family of CMP's in question are restricted to be *strictly* convex with a given modulus of strict convexity $\sigma(\cdot)$: $F(x/2+y/2) \leq F(x)/2+F(y)/2-\sigma(\|x-y\|)$ 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 $\epsilon_{\rm opt}(\cdot)$, and similarly for all other known to us cases when a desired inaccuracy in the argument can be certified in an on line 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) on line bounds on inaccuracy in the argument in convex minimization are more or less straightforward consequences of on line 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 CMP's to other black-box represented problems "with convex structure", specifically, variational inequalities with monotone operators and 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 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; the role of certificates is reduces here to quantifying the quality of such a solution, which sometimes can be considered as 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 (e.g., to the best of our understanding, no "certificate-free" ways for building approximate solutions to, say, black-box represented convex-concave game problems are known).

The rest of this paper is organized as follows. In Section 2, we present a more detailed account of accuracy certificates for convex minimization. In Section 3, after recalling basic facts on variational inequalities with monotone operators and convex Nash equilibrium problems and introducing relevant accuracy measures, we extend to this class of problems the notion of a certificate and demonstrate (Proposition 3.4) that a certificate allows both to generate an approximate solution and to quantify its quality. In Section 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 CMP's, variational inequalities with monotone operators and convex Nash Equilibrium problems can be augmented with computationally cheap techniques for "on line" building of accuracy certificates, and that these certificates are fully compatible with theoretical efficiency estimates of the methods (meaning that the number of steps until a prescribed accuracy is certified is as stated in the theoretical efficiency estimate). Note that as far as the existing literature is concerned, the only known to us relevant study [5] deals with the particular case of an LP program solved by the Ellipsoid algorithm and demonstrates that in this case the method can be augmented by a technique for producing feasible solutions to the dual problem with the duality gap certifying the standard polynomial time efficiency estimate of the Ellipsoid method. The technique used in [5] seems to be quite different from the one we propose, and it is unclear whether it can be extended beyond the LP case. Section 5 presents two instructive applications of the certificate machinery.

To make the paper more readable, all proofs are moved to Section 6.

2 Accuracy certificates in convex minimization

Here we present in details the definition of accuracy certificates for black-box represented CMP's (1) and the related results already mentioned in Introduction.

Recall that when solving a black-box CMP (1), an algorithm after τ steps has at its disposal the *execution protocol* $P_\tau = \{I_\tau, J_\tau, \{x^t, e_t\}_{t=1}^\tau\}$, where $x^1, ..., x^\tau$ are the subsequent points where 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., non-productive ($x^t \notin \operatorname{int} X$) query points, and e_t is either the separator of x^t and $\operatorname{int} X$ returned by the Separation oracle (step t is non-productive), 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

$$\begin{array}{ll} (a) & \xi_t \geq 0 \text{ for each } t = 1, \dots, \tau \\ (b) & \sum_{t \in I_\tau} \xi_t = 1 \end{array} \tag{5}$$

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 can define the quantity

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

– the *residual of* the ξ on $\mathbf{B}$, and *the approximate solution induced by* ξ

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

which clearly is a strictly feasible (i.e., belonging to $\operatorname{int} X$) solution to the CMP (1) underlying the protocol. Given, in addition to $\mathbf{B}$, P_τ and ξ, the values of F at the points x^t, $t \in I_\tau$, we can also define the quantity

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

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 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 $F_*(\xi|P_\tau, \mathbf{B}) \leq \text{Opt}$; consequently, for every feasible solution $\widehat{x}$ of the given CMP it holds:*

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

(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 $\widehat{x}^\tau = \widehat{x}^\tau[\xi]$ are strictly feasible solutions of the given CMP, with*

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

It was mentioned in Introduction that for certain problem classes, the only way to guarantee that a feasible approximate solution $\widehat{x}$ to (1) build upon termination satisfies $\epsilon_{\text{opt}}(\widehat{x}) \leq \epsilon$ is the ability to assign the corresponding execution protocol P_τ, τ being the termination step, admits a certificate ξ satisfying (4). In order 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 that 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 which, 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 $\widehat{x}$ to the problem along with an upper bound ϵ on $\epsilon_{\text{opt}}(\widehat{x})$; here both τ and ϵ can depend on the particular problem the algorithms is applied to. Adding, if necessary, one more step, we can assume w.l.o.g. that the approximate solution returned by the algorithm upon termination always is one of the search points generated by the algorithm in 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 $\widehat{x}$ along with a finite valid upper bound ϵ on $\epsilon_{\rm opt}(\widehat{x})$, then $\widehat{x} \in \operatorname{int} X$ 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

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

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

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

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

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 [13] or Proposition 3.1 in the survey [9]).

One of our coming 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. The most convenient for our purposes way to quantify inaccuracy of an $x \in X$ as an approximate solution to this VIP is to use the proximity measure

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

when X and Φ are evident from the context we shall use the abbreviated notation $\epsilon_{\rm vi}(x)$. By definition, we set $\epsilon_{\rm vi}(x|X, \Phi) = \infty$ when $x \notin X$.

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

Elementary properties of weak solutions to VIP's 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* $\mathrm{Dom}\,\Phi \supseteq \operatorname{int} X$. *Then*

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

(ii) *The function $\epsilon_{\rm vi}(x|X, \Phi)$ is a closed convex nonnegative function on X finite everywhere on* $\operatorname{int} X$ *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*

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

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

$$\begin{array}{rcl} X_* & = & \{x \in X : \langle \eta, y - x \rangle \geq 0 \,\forall (y \in \mathrm{Dom}\,\widetilde{\Phi} \cap X, \eta \in \widetilde{\Phi}(y)), \\ \epsilon_{\mathrm{vi}}(x|X, \Phi) & = & \sup_y \left\{ \langle \eta, x - y \rangle : y \in \mathrm{Dom}\,\widetilde{\Phi} \cap X, \eta \in \widetilde{\Phi}(y) \right\} \forall x \in X. \end{array} \tag{13}$$

In particular, weak solutions and proximity measure 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 Section 6.

Convex Nash equilibrium problems and Nash operators. Let $X_i \subset \mathcal{R}^{n_i}$, $i = 1, ..., m$, be solids, and let $X = X_1 \times X_2 \times ... \times X_m \subset \mathcal{E} = \mathcal{R}^{n_1} \times ... \times \mathcal{R}^{n_m} = \mathcal{R}^{n_1 + ... + n_m}$. A point $x \in \mathcal{E}$ is an ordered tuple $(x_1, ..., x_m)$ with $x_i \in \mathcal{R}^{n_i}$; for such a point and for $i \in \{1, ..., m\}$, 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 $\mathcal{D} = \mathcal{D}_1 \times ... \times \mathcal{D}_m$, where for every $i = 1, \ldots, m$, $\mathrm{int}\, X_i \subseteq \mathcal{D}_i \subseteq X_i$. 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/to approximate such an equilibrium. The standard interpretation of Nash equilibrium is as follows: there are m players, i-th choosing a point $x_i \in \mathcal{D}_i$, and incurring cost $F_i(x)$, where $x = (x_1, ..., x_m)$ is comprised of choices of all m players. Every player is interested to minimize his cost, and Nash equilibria are exactly the tuples $x = (x_1, ..., x_m) \in \mathcal{D}$ 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; the corresponding proximity measure is

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

This is nothing but the sum, over the players, of the maximal gain a player i can get by deviating from his choice x_i when the remaining players stick to their choices.

We intend to consider *convex NEP's*, 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 $x_i \in \mathcal{D}_i$, is concave in $x^i \in \mathcal{D}_1 \times ... \times \mathcal{D}_{i-1} \times \mathcal{D}_{i+1} \times ... \times \mathcal{D}_m$ and, in addition, the function $F(x) = \sum_{i=1}^{m} F_i(x)$ is convex on $\mathcal{D}$.

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

$$\Phi(x) = (\Phi_1(x), \Phi_2(x), ..., \Phi_m(x)) : \mathrm{int}\, X \to \mathcal{E},$$

where for every $x \in \mathrm{int}\, X$ 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 Section 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 NEP's are equipped with two proximity measures: $\epsilon_{\rm N}(\cdot)$ (this measure relates to the problem "as it is") and $\epsilon_{\rm vi}(\cdot)$ (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 are about to consider instructive examples of convex NEP's.

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, CMP's are exactly convex single-player NEP's, 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
$$\epsilon_{\rm N}(x) = F(x) - \inf_{y \in \mathrm{Dom}\,(F)} F(y) = \epsilon_{\rm opt}(x),$$
where $\epsilon_{\rm opt}(x)$ is the accuracy measure defined in Introduction for CMP's. Further, 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 $y \in \mathrm{int}\, X$ and therefore
$$\epsilon_{\rm N}(x) = \epsilon_{\rm opt}(x) = F(x) - \inf_{y \in \mathrm{Dom}\,(F)} F(y) \ge \sup_{y \in \mathrm{int}\, X} \langle \Phi(y), x - y \rangle = \epsilon_{\rm vi}(x). \qquad (15)$$
Note that the "gap" between $\epsilon_{\rm vi}(x)$ and $\epsilon_{\rm N}(x)$ can be large, as is seen in the case of $X = [0, D] \subset \mathcal{R}$, $F(x) = \int_0^x \min[L, \frac{\epsilon}{D-s}]ds$, where $\epsilon \le LD$. Indeed, in this case $\epsilon_{\rm vi}(D) = \max_{0<y<D} F'(y)(D - y) = \epsilon$, while $\epsilon_{\rm N}(D) = F(D) = \epsilon \ln\left(\frac{LD}{\epsilon}\right) + \epsilon$.
It is worthy of 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*
$$\epsilon_{\rm N}(x) \le m \left[\epsilon_{\rm vi}(x) \ln\left(\frac{LD}{\epsilon_{\rm vi}(x)}\right) + \epsilon_{\rm vi}(x)\right]. \qquad (16)$$

B. In the case of $m = 2$ and $F(x) \equiv 0$, a convex Nash equilibrium problem is given by a function $\phi(x_1, x_2) = F_1(x_1, x_2)$ which 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 $\epsilon_{\rm N}(x)$ admits a transparent interpretation, namely, as follows: $\phi(\cdot, \cdot)$ specifies a "primal-dual pair" of optimization problems
$$\begin{array}{ll} (P) & \mathrm{Opt}(P) = \inf_{x_1 \in \mathcal{D}_1} \overline{\phi}(x_1),\ \overline{\phi}(x_1) = \sup_{x_2 \in \mathcal{D}_2} \phi(x_1, x_2); \\ (D) & \mathrm{Opt}(D) = \sup_{x_2 \in \mathcal{D}_2} \underline{\phi}(x_2),\ \underline{\phi}(x_2) = \inf_{x_1 \in \mathcal{D}_1} \phi(x_1, x_2); \end{array}$$

The optimal value in the primal problem is $\geq$ 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 $\epsilon_{\rm N}(x_1,x_2)$ is nothing but the *duality gap* $\overline{\phi}(x_1)-\underline{\phi}(x_2)$, and

$$\epsilon_{\rm N}(x_1,x_2)=\overline{\phi}(x_1)-\underline{\phi}(x_2)\geq[\overline{\phi}(x_1)-{\rm Opt}(P)]+[{\rm Opt}(D)-\underline{\phi}(x_2)];$$

here the inequality holds due to weak duality (and holds as equality whenever ${\rm Opt}(P)={\rm Opt}(D)$). 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 $\epsilon_{\rm N}(x_1,x_2)$. Note that in the case in question, similar to the case of $m=1$, we have

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

Indeed, for $x=(x_1,x_2)\in{\rm Dom}\,\phi=\mathcal{D}_1\times\mathcal{D}_2$ and $y=(y_1,y_2)\in{\rm Dom}\,\Phi={\rm int}\,X_1\times{\rm int}\,X_2$, 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

$$\begin{array}{l}\langle\Phi(y),x-y\rangle=\langle\Phi_1(y),x_1-y_1\rangle+\langle\Phi_2(y),x_2-y_2\rangle\\ \leq[\phi(x_1,y_2)-\phi(y_1,y_2)]+[\phi(y_1,y_2)-\phi(y_1,x_2)]=\phi(x_1,y_2)-\phi(y_1,x_2)\leq\overline{\phi}(x_1)-\underline{\phi}(x_2)=\epsilon_{\rm N}(x),\end{array}$$

and (17) follows.

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

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

is finite. Clearly, a bounded operator Φ is semi-bounded with ${\rm Var}_X(\Phi)\leq\|\Phi\|{\rm Diam}(X)$; there exist, however, semi-bounded operators which are unbounded. E.g., the Nash operator Φ of a convex Nash equilibrium problem with bounded $F_i(\cdot)$, $i=1,...,m$, is clearly semi-bounded with ${\rm Var}_X(\Phi)\leq\sum_i[\sup_{{\rm int}\,X}F_i-\inf_{{\rm int}\,X}F_i]$; note that 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 ${\rm int}\,X_i$. Moreover, in the case of $m=1$ (convex minimization) the Nash operator can be semi-bounded for certain *un*bounded objectives F, most notably, when F is a *ϑ-self-concordant barrier for X* [16, Chapter 2]; for such a barrier, ${\rm Var}_X(F)\leq\vartheta$ [16, 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 CMP's to the case of algorithms that are aimed at solving VIP's. Specifically,

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

- 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 $x^1, ..., x^\tau$ are the subsequent query points generated at the first τ steps of the solution process, e_t is either the separator returned by Separation oracle queried at x^t (this is so when $x^t \notin \mathrm{int}\, X$, a "non-productive 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. non-productive, steps $t = 1, ..., \tau$.

- 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* $\epsilon_{\mathrm{cert}}(\xi|P_\tau, \mathbf{B})$ of such a certificate on a solid $\mathbf{B} \supseteq X$ is defined by (6) and the *approximate solution* $\widehat{x}^\tau[\xi]$ *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* $\mathrm{Dom}\,(\Phi) \supseteq \mathrm{int}\, X$*. Let, further, $P_\tau = \{I_\tau, J_\tau, \{(e_t, x^t)\}_{t=1}^\tau\}$ be an execution protocol, $\xi = \{\xi_t\}_{t=1}^\tau$ be a certificate for P_τ and $\widehat{x}^\tau = \widehat{x}^\tau[\xi]$ be the induced approximate solution. Then $\widehat{x}^\tau \in \mathrm{int}\, X$ and*

$$\epsilon_{\mathrm{vi}}(\widehat{x}^\tau) \leq \epsilon_{\mathrm{cert}}(\xi|P_\tau, \mathbf{B}); \tag{18}$$

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

$$\epsilon_{\mathrm{N}}(\widehat{x}^\tau) \leq \epsilon_{\mathrm{cert}}(\xi|P_\tau, \mathbf{B}). \tag{19}$$

Necessity of accuracy certificates for VIP's. We have seen that an accuracy certificate for an execution protocol associated with a VIP allows both to build a strictly feasible approximate solution $\widehat{x}$ for the VIP and to bound from above the inaccuracy $\epsilon_{\mathrm{vi}}(\widehat{x})$ (and $\epsilon_{\mathrm{N}}(\widehat{x})$ in the case of a Nash VIP) of this solution. It turns out that when solving VIP's from certain natural families, this mechanism is to some extent unavoidable. Specifically, assume that when solving VIP (11), all our a priori information is that X is a solid contained in a given solid $\mathbf{B} \subset \mathcal{R}^n$, and $\Phi : \mathrm{int}\, X \to \mathcal{R}^n$ is a monotone operator bounded by a given constant $L < \infty$ (i.e., $\|\Phi(x)\|_2 \leq L$ for all $x \in \mathrm{int}\, X$). The only sources of further information on the VIP are a Separation oracle representing X and a Φ-oracle which, given on input $x \in \mathrm{int}\, X$, returns $\Phi(x)$. Now consider an algorithm for approximate (within ϵ_{vi}-accuracy $\leq \epsilon$) solving VIP's 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 $x^1, x^2, ...,$ calls Separation oracle to check whether $x^t \in \mathrm{int}\, X$, and if it is the case, calls the Φ-oracle to compute $\Phi(x^t)$. After a finite number of steps τ, the algorithm terminates and outputs an approximate solution $\widehat{x} \in \mathrm{int}\, X$ which must satisfy the relation $\epsilon_{\mathrm{vi}}(\widehat{x}|X, \Phi)$. Of course, we assume the algorithm to be non-anticipative, meaning that $x^1, ..., x^\tau$, τ and $\widehat{x}$ are determined solely on the basis of information accumulated so far. Adding, if necessary, one extra step, we can assume w.l.o.g. that $\widehat{x}^\tau \in \{x^1, ..., x^\tau\}$, and, in particular, that upon termination, the set $I_\tau = \{t \leq \tau : x^t \in \mathrm{int}\, X\}$ is nonempty. We have the following result (cf. Proposition 2.2):

Proposition 3.5 *Let (X, Φ) be a VIP compatible with the our a priori information (so that $X \subset \mathbf{B}$ is a solid and $\Phi : \mathrm{int}\, X \to \mathcal{R}^n$ is a monotone operator with $\|\Phi(x)\|_2 \leq L$ for every $x \in \mathrm{int}\, X$), 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*

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

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

Comment. Proposition 3.4 states that when solving the VIP defined by (X, Φ), a certificate ξ with $\epsilon_{\rm cert}(\xi|P_\tau, \mathbf{B}) \leq \epsilon$, 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 to build an approximate solution $\widehat{x} \in \{x_t : t \in I_\tau\}$ such that $\epsilon_{\rm vi}(\widehat{x}^\tau|X, \Phi) \leq \epsilon$. Proposition 3.5, in turn, says that whenever an algorithm is capable to solve every VIP (X, Φ) with $X \subset \mathbf{B}$ and a monotone and bounded by L operator Φ within $\epsilon_{\rm vi}(\cdot)$-accuracy $\leq \epsilon$ and the resulting approximate solution $\widehat{x} \in {\rm int}\, X$ belongs to the trajectory $x^1, ..., x^\tau$ upon termination, then the corresponding execution protocol P_τ admits a certificate ξ with $\epsilon_{\rm cert}(\xi|P_\tau, \mathbf{B}) \leq \sqrt{LD\epsilon}$. Observing that $\epsilon_{\rm vi}(x|X, \Phi) = \sup_{y \in {\rm int}\, X} \langle \Phi(y), x - y\rangle \leq LD$ whenever $x \in {\rm int}\, X$, the only interesting case is the one when $\epsilon \ll LD$, and in this case $\sqrt{LD\epsilon} \gg \epsilon$, so 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 it was already mentioned, the Ellipsoid algorithm[5] is "an intelligent" method which, 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 VIP's 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 algorithm

A generic Cutting Plane algorithm works with a vector field $\Phi : {\rm Dom}\, \Phi \to \mathcal{R}^n$ defined on a convex set ${\rm Dom}\, \Phi \subset \mathcal{R}^n$ and with a solid X, ${\rm int}\, X \subset {\rm Dom}\, \Phi$. 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 $x^t \in {\rm int}\, Q_t$, $t = 1, 2, ...$ The algorithm is as follows:

Algorithm 4.1 *Initialization:* *Choose a solid $Q_1 \supset X$ and a point $x^1 \in {\rm int}\, Q_1$.*
Step t, $t = 1, 2, ...$: Given x^t, Q_t,

1. *Call Separation oracle, x^t being the input. If the oracle reports that $x^t \in {\rm int}\, X$ (productive step), go to 2. 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 3.*
2. *Call Φ-oracle to compute $e_t = \Phi(x^t)$. If $e_t = 0$, terminate, otherwise go to 3.*
3. *Set*
$$\widehat{Q}_{t+1} = \{x \in Q_t : \langle e_t, x - x^t\rangle \leq 0\}.$$
Choose, as Q_{t+1}, a solid which contains the solid $\widehat{Q}_{t+1}$. Choose $x^{t+1} \in {\rm int}\, Q_{t+1}$ 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 $\rho(Q_t) \to 0$, $t \to \infty$.

[5]This method was developed in [22] and independently and slightly later – in [19]; for its role in the theory of Convex Optimization see, e.g., [10, 8, 14, 1] and references therein.

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,

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

2. All localizers Q_t are ellipsoids represented as the images of the unit Euclidean ball under affine mappings:

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

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

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

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

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

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

The ellipsoid method is converging, with

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

Other implementations. Aside 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

- the Center of Gravity method [12, 18]: $Q_1 = X$, x^t is the center of gravity of Q_t, $Q_{t+1} = \widehat{Q}_{t+1}$; here $p(n) = \text{const}$. This method, however, is of academic interest only, since it requires finding centers of gravity of general-type polytopes, which is NP-hard;
- the Inscribed Ellipsoid algorithm [20] 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 $Q_{t+1} = \widehat{Q}_{t+1}$; here again $p(n) = \text{const}$;
- Circumscribed Simplex algorithm [4, 21], where Q_t are simplexes, 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 specific and rare property as follows: *for every* $e \neq 0$, *the set* $\widehat{\mathbf{C}}_e = \{x \in \mathbf{C} : \langle e, x\rangle \leq 0\}$ *can be covered by an affine image of* $\mathbf{C}$ *under an efficiently computable affine mapping which 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 [6] 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 0, and 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[6].

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^+:

$$Q_t^+ = \{[sx; s] : 0 \leq s \leq 1, x \in Q_t\} \qquad ^7.$$

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 $\widehat{Q}_{t+1}^+ = \{[sx; s] : 0 \leq s \leq 1, x \in \widehat{Q}_{t+1}\}$ of the origin in E^+ and the sets $\widehat{Q}_{t+1} \subset E$, the sets Q_t^+ and the cuts are linked by the relation

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

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

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

$$\text{Polar}\,(Q_{t+1}^+) \subset \text{Polar}\,(\widehat{Q}_{t+1}^+) = \text{Polar}\,(Q_t^+) + \mathcal{R}_+ e_t^+ \tag{25}$$

4.3 Building the certificates: the algorithm

Equipped with Proposition 4.1, we can build certificates in a "backward fashion", namely, 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: Since $e_\tau = 0$ 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

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

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

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

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

[6]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.

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

and

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

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 $\|h\|_2 \leq \frac{1}{2\rho(Q_{t+1})}$ (see below). In the general case, we can apply the "ellipsoidal construction" to the *Fritz John ellipsoid* of Q_{t+1} – an ellipsoid $\widetilde{Q} \supset Q_{t+1}$ with $\rho(\widetilde{Q}) \leq n\rho(Q_{t+1})$, see [7]. Note that for all 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.

2. *Observe that both the vectors*

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

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

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

The certificate for the protocol P_τ is well defined only for those τ for which the set $I_\tau = \{t \leq \tau : x^t \in \text{int}\, X\}$ 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

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

and we also set

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

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 really easy. Indeed,

- to find h in rule 1, it suffices to build the singular value decomposition $B_{\tau+1} = UDV$ (U, V are orthogonal, D is diagonal with positive diagonal entries) and to set $h = \frac{1}{2\sigma_{i_*}} U e_{i_*}$, 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 $\sigma_{i_*} \leq |\text{Det}(B_{\tau+1})|^{1/n} = \rho(Q_{\tau+1})$, so that $\|h\|_2 \geq \frac{1}{2\rho(Q_{\tau+1})}$ (i.e., we ensure (27) with $\chi = 1$). Besides this, $\max_{Q_{t+1}} \langle h, x\rangle - \min_{Q_{t+1}} \langle h, x\rangle = \max_{\|u\|_2\leq 1, \|v\|_2 \leq 1} \langle B_{t+1}^T h, u - v\rangle = \frac{1}{2} \max_{\|u\|_2\leq 1, \|v\|_2 \leq 1} \langle V^T e_{i_*}, u - v\rangle = 1$, as required in rule 1;
- we clearly have Polar $(Q_t^+) = \{[e; s] : \|B_t^T e\|_2 + \langle x^t, e\rangle + s \leq 1\}$, so that given $[g; a] \in$ Polar (Q_{t+1}^+), to find a representation $[g; a] = [f; b] + r e_t^+$ with $[f; b] \in$ Polar (Q_t^+) and $r \geq 0$ reduces to solving the problem (known in advance to be solvable) of finding $r \geq 0$ such that $\|B_t^T(g - r e_t)\|_2 + \langle x^t, g\rangle + a \leq 1$. 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 $r_* = \frac{p^T q}{q^T q}$.

From the just presented remarks it follows that with the Ellipsoid method in the role of Algorithm 4.1, the computational cost of building 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 updatings $(x^t, B_t) \mapsto (x^{t+1}, B_{t+1})$, $t = 1, ..., \tau$, with 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,... the associated computational expenses basically do not affect the complexity of the Ellipsoid algorithm, see Discussion below.

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* $\Phi : \operatorname{int} X \to \mathcal{R}^n$ *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* $\epsilon_{\mathrm{cert}}(\xi|P_\tau, Q_1) = 0$*, otherwise*

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

and, besides this,

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

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

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

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

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

the certificate ξ *is well defined, and*

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

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

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

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

$$\exp\{\frac{\tau}{2n(n-1)}\} \ge \frac{32R^2}{r^2}, \tag{35}$$

and in this case

$$\epsilon_{\mathrm{cert}}(\xi|P_\tau, Q_1) \le \frac{32R^2 \mathrm{Var}_X(\Phi)}{r^2} \exp\{-\frac{\tau}{2n(n-1)}\}, \tag{36}$$

provided that Φ *is semi-bounded 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 which is contained in the centered at the origin Euclidean ball $\mathbf{B} = Q_1$ of radius R, and let $r = r(X)$ be the largest of the radii of Euclidean balls contained in X. Further, let $\Phi : \mathrm{Dom}\,\Phi \to \mathcal{R}^n$, $\mathrm{Dom}\,\Phi \supset \mathrm{int}\,X$, be a semi-bounded on $\mathrm{int}\,X$ operator:*

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

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

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

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

$$\epsilon_{\mathrm{cert}}(\xi|P_\tau, \mathbf{B}) \le \epsilon,$$

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

- *when Φ is monotone, we have*
$$\epsilon_{\mathrm{vi}}(\widehat{x}|X, \Phi) \le \epsilon;$$
- *when Φ is the Nash operator of a convex Nash equilibrium problem on X (in particular, in the case of convex minimization), we have*
$$\epsilon_{\mathrm{N}}(\widehat{x}) \le \epsilon.$$

Discussion. Corollary 4.1 states that solving a convex Nash equilibrium problem (or a variational inequality) associated with a semi-bounded, monotone operator Φ within accuracy ϵ "costs" $O(1)n^2 \ln\left(\frac{R\mathrm{Var}_X(\Phi)}{r\epsilon}\right)$ steps of the Ellipsoid algorithm, with (at most) one call to the Separation and the 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 $\epsilon_{\mathrm{cert}}(\xi|P_\tau, \mathbf{B}) \le \epsilon$ is built, we know that we already have in 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, while ensuring a desired quality of the resulting approximate solution, requires no a priori information on the problem* (aside of 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,..., 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 $e \neq 0$ *and* a nonnegative number a such that $X \subseteq \{y : \langle e, y - x\rangle \leq -a\}$. In this situation, all our results remain valid when the residual of a certificate on a solid $\mathbf{B} \supset X$ is redefined as

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

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 $\widehat{Q}_{t+1}$ in Algorithm 4.1 are defined as

$$\widehat{Q}_{t+1} = \{x \in Q_t : \langle e_t, x - x^t\rangle \leq -\alpha_t\}, \tag{38}$$

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

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

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

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

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

for a productive step t (a_t is used in (37), α_t – in (38) and (39)). With this modification, all 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

$$\epsilon_{\text{opt}}(x^\tau_{\text{bst}}) \leq \epsilon_{\text{cert}}(\xi|P_\tau, \mathbf{B})$$

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

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

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 followup paper). The first of examples is of academic nature, while the second is of a practical flavour.

5.1 Recovering Dual Solutions to Large LP Problems

Consider a Linear Programming program with box constraints

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

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 $\operatorname{int} X = \{x \in \mathcal{R}^n : \widehat{A}x < \widehat{b}\}$. Assume that we have in our disposal 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 $a_{j(u)}^T x \le b_{j(u)}$ from the system $\widehat{A}x \le \widehat{b}$ 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 in 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" can we 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 $\widehat{A}$ such that $a_{j(x^t)}^T x^t \ge b_{j(x^t)}$. Setting

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

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

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

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

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

implies that

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

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

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

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

$$\widehat{x}^\tau \in X, \ \widehat{\lambda} \ge 0, \ c + \widehat{A}^T\widehat{\lambda} = 0, c^T\widehat{x} + \widehat{b}^T\widehat{\lambda} \le \epsilon_{\rm cert}(\xi|P_\tau, \mathbf{B}).$$

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

This example is of significant interest when the LP problem (40) has a huge number of constraints which are "well organized" in the sense that it is relatively easy to find a constraint, if any, which is not strictly satisfied at a given point (for examples, see., e.g., [8]). By Corollary 4.1, the latter property allows, given $\epsilon > 0$, to build in $O(n^2 \ln \left(\frac{n\|c\|_2}{r(X)\epsilon}\right))$ steps, with a single call to Separation oracle for X and $O(n^2)$ additional operations per step, a protocol P_τ and associated certificate ξ such that $\epsilon_{\rm cert}(\xi|P_\tau, \mathbf{B}) \le \epsilon$. 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, since the latter require Newton-type iterations which in high dimensions (many thousands of decision variables) become prohibitively expensive, unless the problem possesses a favourable sparsity pattern (which not always is the case). As a result, there is a growing interest, initiated by the breakthrough paper [17], 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 [17], Mirror Prox algorithm from [15], NERML algorithm [2] and many others) is the necessity to solve at every step an auxiliary problem of the form

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

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 which "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 $\{u \in \mathcal{R}^N_+ : \sum_i u_i = 1\}$ 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 incorporate at a cost even small (few, or perhaps few tens) systems of these constraints, it would extend significantly the scope of the "master methods". Well, 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

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

(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: since ω fits U, it is easy to find $u_x \in \mathrm{Argmin}_{u \in U} \phi(u, x)$, 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 (42) satisfies the Slater condition (so that (43) is solvable) and that we have in 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

$$\min_{x\in X} F(x),\;\; X=\{x\in\mathcal{R}^k_+ : \|x\|\leq 2L\} \tag{44}$$

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 to get a high accuracy solution to (44), even when L is large, at a relatively low (like few hundreds) 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); and 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,* $P_\tau=\{I_\tau,J_\tau,\{x^t,e_t\}_{\tau=1}^t\}$ *be the execution protocol upon termination, and let* ξ *be an accuracy certificate for this protocol. Let us set*

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

where $u_{x^t}\in \mathrm{Argmin}_{u\in U}\left[f(u)+\langle x^t,Au-b\rangle\right]$ *are the vectors underlying the answers of the First order oracle for as queried at the points* x_t*,* $t\in I_\tau$*. Then* $\widehat{u}\in U$ *and*

$$(a)\quad \|[A\widehat{u}-b]^+\|_q\leq \epsilon_{\rm cert}(\xi|P_\tau,X),\qquad (b)\quad f(\widehat{u})-\mathrm{Opt}\leq \epsilon_{\rm cert}(\xi|P_\tau,X), \tag{45}$$

where $[A\widehat{u}-b]^+$ *is the "vector of constraint violations" obtained from* $A\widehat{u}-b$ *by replacing the negative components with 0, and* $q=\frac{p}{p+1}$*. Thus,* $\widehat{u}$ *is a nearly feasible and nearly optimal solution to* (42)*, provided that* $\epsilon_{\rm cert}(\xi|P_\tau,\mathbf{B})$ *is small.*

6 Proofs

6.1 Proofs for Section 2

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

$$F_*(\xi|P_\tau,\mathbf{B})=\underbrace{\sum\nolimits_{t\in I_\tau}\xi_t F(x^t)}_{\underset{(*)}{\geq} F(\widehat{x}^\tau)}-\epsilon_{\rm cert}(\xi|P_\tau,\mathbf{B})\geq \underbrace{\min\nolimits_{t\in I_\tau}F(x^t)}_{=F(x^\tau_{\rm bst})}-\epsilon_{\rm cert}(\xi|P_\tau,\mathbf{B}), \tag{46}$$

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

Proof of Proposition 2.2. Let the algorithm in question as applied to a problem $\min_{x\in X}F(x)$ compatible with our a priori information terminate in τ steps, generating execution protocol $P_\tau=\{I_\tau,J_\tau,\{(x^t,e_t)\}_{t=1}^\tau\}$, a feasible solution $\widehat{x}$ and a valid upper bound ϵ on $\epsilon_{\rm opt}(\widehat{x}$. With our assumptions on the algorithm, $\widehat{x}$ is one of the points $x^1,...,x^\tau$, say, $\widehat{x}=x^s$. We first claim that $s\in I_\tau$ (and this $I_\tau\neq\emptyset$ and $\widehat{x}\in\mathrm{int}\,X$). Indeed, our a priori information on X, augmented by the information accumulated in course of running the algorithm on the problem, does not contradict the assumption that $X = X^\delta := \{x \in \mathbf{B} : \langle e_t, x - x^t\rangle \leq -\delta, t \in J_\tau\}$, provided that $\delta > 0$ is small enough, specifically, such that $\langle e_t, x^{t'} - x^t\rangle \leq -\delta$ 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 $\widehat{x} = x^s$ is infeasible, so that the corresponding inaccuracy $\epsilon_{\rm opt}(widehatx)$ is $+\infty$, which is not the case – we know that this inaccuracy is $\leq \epsilon < \infty$.

Now let us set

$$\bar{X} = \{x \in \mathbf{B} : \langle e_t, x - x^t\rangle \leq 0,\ t \in J_\tau\},\ \ \bar{F}(x) = \max_{t\in I_\tau}[F(x^t) + \langle F'(x^t), x - x^t\rangle]. \tag{47}$$

Observe that our a priori information on the problem $\int_{x\in X} F(x)$ ("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

$$\mathrm{Opt} = \mathrm{Opt}_* \equiv \min_{x\in\bar{X}} \bar{F}(x), \tag{48}$$

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

$$F(\widehat{x}) - \mathrm{Opt}_* \leq \epsilon. \tag{49}$$

It remains to note that

$$\begin{array}{rcl}
\mathrm{Opt}_* & = & \min\limits_{x}\left\{\max\limits_{t\in I_\tau}[F(x^t) + \langle e_t, x - x^t] : \langle e_t, x - x^t\rangle \leq 0, t \in J_\tau, x \in \mathbf{B}\right\} \\
& \underset{(a)}{=} & \min\limits_{x,s}\left\{ s:\ \begin{array}{l} s \geq F(x^t) + \langle e_t, x - x^t\rangle,\ t \in I_\tau \\ 0 \geq \langle e_t, x - x^t\rangle,\ t \in J_\tau \\ x \in \mathbf{B}\end{array}\right\} \\
& \underset{(b)}{=} & \max\limits_{\xi\geq 0}\left\{\min\limits_{x\in\mathbf{B},s}\left[s + \sum\limits_{t\in I_\tau}\xi_t[F(x^t) - s] + \sum_{t=1}^{\tau}\xi_t\langle e_t, x^t - x\rangle\right]\right\} \\
& \underset{(c)}{=} & \max\limits_{\xi}\left\{\sum_{t\in I_\tau}\xi_t F(x^t) + \min\limits_{x\in\mathbf{B}}\sum_{t=1}^{\tau}\xi_t\langle e_t, x - x^t\rangle : \xi \geq 0, \sum\limits_{t\in I_\tau}\xi_t = 1\right\} \\
& \underset{(d)}{=} & \max\limits_{\xi}\left\{\sum_{t\in I_\tau}\xi_t F(x^t) - \epsilon_{\rm cert}(\xi|P_\tau, \mathbf{B}) : \xi \geq 0, \sum_{t\in I_\tau}\xi_t = 1\right\}
\end{array}$$

In this chain, (a) is evident, (b) is given by the Lagrange Duality Theorem (see, e.g., [3, Chapter 5]) as applied to the optimization problem in the right hand side of (a) [8]; (b) is evident. 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 evident (c) and (d). We see that

$$\mathrm{Opt}_* = \max_{\xi}\left\{\sum_{t\in I_\tau}\xi_t F(x^t) - \epsilon_{\rm cert}(\xi|P_\tau, \mathbf{B}) : \xi \geq 0, \sum_{t\in I_\tau}\xi_t = 1\right\},$$

so that there exists a certificate ξ^* for the protocol P_τ such that $\mathrm{Opt}_* = \sum_{t\in I_\tau}\xi^*_t F(x^t) - \epsilon_{\rm cert}(\xi^*|P_\tau, \mathbf{B})$. Substituting this representation into (49), we arrive at (4). ■

[8] 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, since $\mathbf{B}$ is bounded and $I_\tau \neq \emptyset$, and satisfies the Slater condition – namely, every point s, x with $x \in \mathrm{int}\, X$ and $s > F(x)$ strictly satisfies the constraints.

6.2 Proofs for Section 3

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 $X_* \neq \emptyset$, assume that X_* is empty, and let us lead this assumption to a contradiction. Since 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 point $y_i \in \mathrm{Dom}\,\Phi \cap X$, $i = 1, ..., I$, such that $\bigcup\limits_{i=1}^{I} X_{y_i} = \emptyset$, or, equivalently, the function $g(x) = \max\limits_{1\leq i\leq I} \langle F(y_i), x - y_i\rangle$ is positive everywhere on X. Applying the von Neumann Lemma, we see that there exist $\lambda_i \geq 0$, $\sum_{i=1}^{I} \lambda_i = 1$ such that

$$\sum_i \lambda_i \langle F(y_i), x - y_i\rangle > 0\,\forall x \in X. \tag{50}$$

On the other hand, setting $\bar{y} = \sum_i \lambda_i y_i$, we get $\bar{y} \in \mathrm{Dom}\,\Phi \cap X$, and by monotonicity of F we get $\langle F(\bar{y}), \bar{y} - y_i\rangle \geq \langle F(y_i), \bar{y} - y_i\rangle$. Taking 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 (50). (i) is proved.

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

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

$$\forall (x \in X, z \in X \cap \mathrm{Dom}\,\widetilde{\Phi}, \phi \in \widetilde{\Phi}(z)) : \langle \phi, x - z\rangle \leq \Theta(x) := \sup_{y\in \mathrm{int}\,X} \langle \Phi(y), x - y\rangle \tag{51}$$

(note that $\Theta(x) \leq \epsilon_{\rm vi}(x)$). To prove (51), 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 $x_{t\theta} \in \mathrm{int}\,X \subset \mathrm{Dom}\,\Phi$. Since $\widetilde{\Phi}$ is a monotone extension of Φ, we have

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

Thus,

$$\langle \phi, x_{t\theta} - z\rangle \leq \frac{1-\theta}{\theta}\Theta(x) + t\psi(t), \quad \psi(t) = \langle F(x_{t\theta}), e\rangle. \tag{52}$$

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

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

Since $\psi(t) \le \psi(1)$ when $0 < t < 1$, (52) implies that

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

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

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

whence $\langle \phi, x - z\rangle \le \frac{1}{\theta}\Theta(x)$. This inequality holds true for every $\theta \in (0,1)$, and the conclusion in (51) follows. ■

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

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

where (a), (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 $x^i \in X^i = X_1 \times .. \times X_{i-1} \times X_{i+1} \times ... \times X_m$ and are concave in $x^i \in X^i$ for every $x_i \in X_i$, and $F(x) = \sum_i F_i(x)$ 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

$$\begin{array}{l}
\frac{1}{2}\langle \Phi(y), y - x_*\rangle = \langle \Phi(\bar{x}+\Delta), \Delta\rangle = \sum_i \underbrace{\langle \Phi_i(\bar{x}+\Delta), \Delta_i\rangle}_{\underset{(a)}{\ge} F_i(\bar{x}+\Delta) - F_i(\bar{x}^i+\Delta^i, \bar{x}_i)} \\
\ge \sum_i [F_i(\bar{x}+\Delta) - F_i(\bar{x}^i+\Delta^i, \bar{x}_i) + \underbrace{F_i(\bar{x}-\Delta) - F_i(\bar{x}^i-\Delta^i, \bar{x}_i)}_{= F(x_*^i, (x_*)_i) - F_i(x_*^i, \bar{x}_i) \underset{(b)}{\le} 0}] \\
= F(\bar{x}+\Delta) + F(\bar{x}-\Delta) - \sum_i [\underbrace{F_i(\bar{x}^i+\Delta^i, \bar{x}_i) + F_i(\bar{x}^i-\Delta^i, \bar{x}_i)}_{\underset{(c)}{\le} 2F_i(\bar{x})}] \\
\ge F(\bar{x}+\Delta) + F(\bar{x}-\Delta) - 2F(\bar{x}) \underset{(d)}{\ge} 0
\end{array}$$

where (a) is due to convexity of $F_i(y^i, \cdot)$, (b) due to the fact that $F_i(x_*^i, x_i)$ attains it 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, which is the same, $F_m(x_m^*, (x_*)_m + \bar{s}[v - (x_*)_m]) > F_m(x_*^m, v)$. Since $F_m(x)$ is continuous in $x \in X$, we can find, first, $\bar{v} \in \operatorname{int} X_m$ close enough to v and, second, a small enough neighbourhood U (in X^m) of the point x_*^m such that

$$\forall (u \in U) : F_m(u, (x_*)_m + \bar{s}[\bar{v} - (x_*)_m]) - F_m(u, \bar{v}) \geq \epsilon > 0. \tag{53}$$

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

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

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

$$\begin{array}{l} \langle \Phi_i(x[\rho, \delta]), x_i[\rho, \delta] - (x_*)_i \rangle = \frac{\rho}{1-\rho} \langle \Phi_i(x[\rho, \delta]), \bar{u}_i - x_i[\rho, \delta] \rangle \\ \underset{(a)}{\leq} \frac{\rho}{1-\rho} [F_i(x^i[\rho, \delta], \bar{u}_i) - F_i(x[\rho, \delta])] \leq \frac{\rho}{1-\rho} M, \end{array} \tag{54}$$

where $M/2$ is an upper bound on $|F_j(x)|$ over $j = 1, ..., m$ 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 (53),

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

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

$$\langle \Phi_m(x[\rho, \delta_\rho]), x_m[\rho, \delta_\rho] - (x_*)_m \rangle = \delta_\rho \langle \Phi_m(x[\rho, \delta_\rho]), \bar{v} - (x_*)_m \rangle \leq -\delta_\rho \epsilon \leq -(1 - \bar{s})\epsilon.$$

Combining this relation with (54), we get

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

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

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 $f_i'(t) = \langle \Phi_i(x^i, y_i + t(x_i - y_i)), x_i - y_i \rangle \leq LD$. Besides this, setting $z = (x^i, y_i + t(x_i - y_i))$, we have $f_i'(t) = \frac{1}{1-t} \langle \Phi(z), x - z \rangle \leq \frac{\epsilon_{\text{vi}}(x)}{1-t}$. Therefore

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

(we have used the evident fact that $\epsilon_{\text{vi}}(x) \leq LD$). We see that for every i one has $F_i(x) - \inf_{y_i \in \operatorname{int} X_i} F_i(x^i, y_i) \leq \epsilon_{\text{vi}}(x) \ln \left(\frac{LD}{\epsilon_{\text{vi}}(x)} \right) + \epsilon_{\text{vi}}(x)$, and (16) follows. ■

Proof of Proposition 3.4. The inclusion $\widehat{x}^\tau \in \operatorname{int} X$ is immediate, since $\widehat{x}^\tau$ 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

$$\epsilon_{\text{cert}}(\xi|P_\tau, \mathbf{B}) = \max_{x\in\mathbf{B}} \left[\sum_{t=1}^{\tau} \xi_t \langle e_t, x^t - x\rangle\right] \geq \max_{x\in X} \left[\sum_{t\in I_\tau} \xi_t \langle \Phi(x^t), x^t - x\rangle\right]. \tag{55}$$

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 $t = 1, \ldots, \tau$, and therefore

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

proving (18).

Next assume that Φ is the Nash operator of a convex NEP on $X = X_1 \times \ldots \times X_m$. Let $F_i(x^i, x_i)$ be the corresponding cost functions, $\mathcal{D} = \mathcal{D}_1 \times \ldots \times \mathcal{D}_m$ 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. Since 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 $t \in I_\tau$,

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

The resulting inequality combines with (55) to imply that

$$\epsilon_{\text{cert}}(\xi|P_\tau, \mathbf{B}) \geq \sup_{x\in\mathcal{D}} \sum_{t\in I_\tau} \xi_t \left[F(x^t) - \sum_{i=1}^m F_i([x^t]^i, x_i)\right]. \tag{56}$$

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

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

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

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

Therefore (56) implies that

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

which verifies (19). ■

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 $Y = \{x^1, ..., x^N\} \subset \mathcal{R}^n$, let $\Phi : Y \to \mathcal{R}^n$ 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 $x \in \mathcal{R}^n \backslash Y$. 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 [13]; to make the paper self-contained, we present a proof.

Proof. For $j = 1, \ldots, N$, let $F_j = \Phi(x_j)$, let $\Omega = \{\xi \in \mathcal{R}^N : \xi \geq 0, \sum_{j=1}^N \xi_j = 1\}$, and 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 $\xi \in \Omega$, $\|F_\xi\|_2 \leq L$ and, 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 $F \in \mathcal{R}^N$

$$\langle F, x \rangle - \langle F, x_\xi \rangle - \langle F_\xi, x \rangle + \sum_{j=1}^N \xi_j \langle F_j, x_j \rangle \geq \langle F - F_\xi, x - x_\xi \rangle. \tag{57}$$

The assertion of our lemma 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

$$\begin{array}{rll}
A & = \max\limits_{\|F\|_2 \leq L} \min\limits_j \langle F - F_j, x - x_j \rangle & \\
& = \max\limits_{\|F\|_2 \leq L} \min\limits_{\xi \in \Omega} [\sum\limits_{j=1}^N \xi_j \langle F - F_j, x - x_j \rangle] & \\
& = \min\limits_{\xi \in \Omega} \max\limits_{\|F\|_2 \leq L} [\sum\limits_{j=1}^N \xi_j \langle F - F_j, x - x_j \rangle] & \left[\begin{array}{l}\text{“max” and “min” can be exchanged as the} \\ \text{bracketed term is bilinear in } \xi \text{ and } F \text{ and} \\ \{F : \|F\|_2 \leq L\}, \Omega \text{ are convex compact sets}\end{array}\right] \\
& = \min\limits_{\xi \in \Omega} \max\limits_{\|F\|_2 \leq L} [\langle F, x \rangle - \langle F, x_\xi \rangle - \langle F_\xi, x \rangle + \sum\limits_{j=1}^N \xi_j \langle F_j, x_j \rangle] & \\
& \geq \min\limits_{\xi \in \Omega} \max\limits_{\|F\|_2 \leq L} \langle F - F_\xi, x - x_\xi \rangle & [\text{ by (57) }] \\
& \geq \min\limits_{\xi \in \Omega} \langle F_\xi - F_\xi, x - x_\xi \rangle = 0. & [\text{ as } \|F_\xi\|_2 \leq L \text{ for each } \xi \in \Omega\ ] \blacksquare
\end{array}$$

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 $\Phi : Y \to \mathcal{R}^n$ 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 $\Phi : Y \to \mathcal{R}^n$ with $Y \subset X$, let us look at the set $\mathcal{Y}$ 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:

$$[(Y', \Phi') \succeq (Y'', \Phi'')] \Leftrightarrow [Y' \supset Y'' \,\&\, \Phi'|_{Y''} = \Phi'']$$

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 $x \in X \backslash \widehat{Y}$. For every point $y \in \widehat{Y}$, the set $\mathcal{F}_y = \{F : \|F\|_2 \le L, \langle F - \widehat{\Phi}(y), x - y \rangle \ge 0\}$ 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 $\mathcal{Y}$ which is $\succ (\widehat{Y}, \widehat{\Phi})$, a contradiction which 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 $I_\tau \neq \emptyset$, since the approximate solution $\widehat{x}$ generated by the algorithm as applied to (X, Φ) belongs to $\operatorname{int} X$, on one hand, and is a point from the trajectory $x^1, ..., x^\tau$, on the other hand. Let us set $\widehat{X} = \{x \in \mathbf{B} : \langle e_t, x - x^t \rangle \le 0, t \in J_\tau\}$. As in the proof of Proposition 2.2, setting

$$\delta = \max_{x \in \widehat{X}} \min_{t \in I_\tau} \langle \Phi(x_t), x^t - x \rangle, \tag{58}$$

there exists a certificate ξ for P_τ such that

$$\epsilon_{\rm cert}(\xi | P_\tau, \mathbf{B}) = \delta;$$

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

$$\max_{\xi \ge 0, \sum_{t \in I_\tau} \xi_t = 1} \left\{ \min_{x \in \mathbf{B}} \left[\sum_{t \in I_\tau} \xi_t \langle e_t, x - x^t \rangle + \sum_{t \in J_\tau} \xi_t \langle e_t, x - x^t \rangle \right] \right\}. \tag{59}$$

The conclusion in Proposition 3.5 is trivially true when $\delta \le 0$, so that let us assume that $\delta > 0$. Let $Y = \{x^t : t \in I_\tau\}$. 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, $x_* \notin Y$ (since otherwise $\delta = \min_{y \in Y} \langle \Phi(y), y - x_* \rangle \le 0$), and since $\widehat{x} \in \{x^1, ..., x^\tau\} \cap \operatorname{int} X$, we have

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

In particular,

$$D \ge \|\widehat{x} - x_*\|_2 \ge \frac{\delta}{L}. \tag{60}$$

Next, set

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

and note that $\|F\|_2 \le \frac{\delta}{D} \le L$ 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

$$\langle \widehat{\Phi}(x_*), \widehat{x} - x_* \rangle = \langle F, \widehat{x} - x_* \rangle = \frac{\delta \|\widehat{x} - x_*\|_2^2}{D \|\widehat{x} - x_*\|} = \frac{\delta \|\widehat{x} - x_*\|_2}{D} \ge \frac{\delta^2}{LD}, \tag{61}$$

(we have used (60)), we have that $\epsilon_{\rm vi}(\widehat{x} | \widehat{X}, \widehat{\Phi}) \ge \frac{\delta^2}{LD}$.

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 $\frac{\delta^2}{LD} \le \epsilon_{\rm vi}(\widehat{x} | \widehat{X}, \widehat{\Phi}) \le \epsilon$. Thus, $\epsilon_{\rm cert}(\xi | P_\tau, \Phi) = \delta \le \sqrt{LD\epsilon}$. ■

6.3 Proofs for Section 4

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* $\mathrm{int}\, P \cap \mathrm{int}\, Q \neq \emptyset$*. Then* $\mathrm{Polar}\,(P \cap Q) = \mathrm{Polar}\,(Q) + P_*$*, where $P_* = \{z \in E^+ : \langle z, u\rangle \leq 0 \,\forall u \in P\}$.*

Proof of Lemma: By standard facts on polars, $\mathrm{Polar}\,(P \cap Q) = \mathrm{cl}\;(\mathrm{Conv}\,(\mathrm{Polar}\,(P) \cup \mathrm{Polar}\,(Q)))$. In our case, $\mathrm{Polar}\,(P) = P_*$; since P_* is a cone and $\mathrm{Polar}\,(Q)$ contains the origin, the convex hull of the union of P_* and $\mathrm{Polar}\,(Q)$ is dense in the arithmetic sum of these two sets, that is, $\mathrm{Polar}\,(P \cap Q) = \mathrm{cl}\,(P_* + \mathrm{Polar}\,(Q))$. It follows that all we should verify in order to prove Lemma is that the set $P_* + \mathrm{Polar}\,(Q)$ is closed. But this is immediate: let $u_i \in P_*$ and $v_i \in \mathrm{Polar}\,(Q)$ be such that $u_i + v_i \to w$, $i \to \infty$; we should prove that $w \in P_* + \mathrm{Polar}\,(Q)$. 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 which 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

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

that is, the sequence $\{\min_{x \in B}\langle u_i, x\rangle\}_{i=1}^{\infty}$ is below bounded. Since the sequence $\{\max_{x \in B}\langle u_i, x\rangle\}_{i=1}^{\infty}$ is above bounded by 1, 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 to equality in (25): By (24), $\widehat{Q}_{t+1}^+ = Q_t^+ \cap P$, where P is the cone $P = \{z \in E^+ : \langle e_t^+, z\rangle \leq 0\}$. Since $x^t \in \mathrm{int}\, Q_t$ and $e_t \neq 0$, the interiors of P and Q_t^+ have a nonempty intersection, so that by Lemma 6.3 $\mathrm{Polar}\,(\widehat{Q}_{t+1}^+) = \mathrm{Polar}\,(Q_t^+) + P_* = \mathrm{Polar}\,(Q_t^+) + \mathcal{R}_+ e_t^+$. ■

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 form 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 get

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

multiplying both sides by x^+, we get

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

where the concluding inequality is given by the inclusions $\phi^+, \psi^+ \in \mathrm{Polar}\,(Q_1^+)$ 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 which 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 $t \in J_\tau \equiv \{1,...,\tau\}\backslash I_\tau$, 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 $y = \frac{(r-\epsilon)x + \epsilon\bar{x}}{r}$, 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 weighted sum of inequalities (P_t), the weights being $[\lambda_t + \mu_t]\epsilon$, we get

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

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

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

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 Euclidean ball B of radius r which is contained in X. Consider first the case when $\langle h, \bar{x} - x^{\tau+1} \rangle \geq 0$. Multiplying both sides in $(28.a)$ by $x^+ = [\bar{x} + re; 1]$ with $e \in \mathcal{R}^n$, $\|e\|_2 \leq 1$, we get

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

The resulting inequality holds true for all unit vectors e; maximizing the left hand side over these e, we get $D_\tau \geq \frac{r\|h\|_2 - 1}{D(Q_1)}$. Recalling that $\|h\|_2 \geq \frac{1}{2\chi\rho(Q_{\tau+1})}$, we arrive at (31). We have established (ii) in the case of $\langle h, \bar{x} - x^{\tau+1} \rangle \geq 0$; 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 semi-bounded on $\operatorname{int} X$, we have $W_\tau \leq \operatorname{Var}_X(\Phi)$ for all τ. ∎

6.4 Proofs for Section 5

Proof of Proposition 5.1. In the situation of Proposition, setting $\epsilon = \epsilon_{\rm cert}(\xi|P_\tau, X)$, $\bar{x} = \sum_{t\in I_\tau} \xi_t x_t$ and $u_t = u_{x_t}$, $t \in I_\tau$, we have $e_t = b - Au_t$, $t \in I_\tau$, and

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

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

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

Multiplying t-th inequality by ξ_t, summing up and taking into account that $\sum_{t\in I_\tau} \xi_t f(u_t) \geq f(\widehat{u})$, we get

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

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

$$2L\|[A\widehat{u} - b]^+\|_q + f(\widehat{u}) - f(u_*) + \langle b - Au_*, \bar{x}\rangle \leq \epsilon, \tag{62}$$

where u_* is an optimal solution to (42). Taking into account that $\bar{x} \in X$ and thus $\bar{x} \geq 0$, whence $\langle b - Au_*, \bar{x}\rangle \geq 0$, we arrive at (45.$b$). Further, we have $f(u_*) \leq f(u) + \langle x_*, Au - b$ for all $u \in U$ due to the origin of x_*, whence $f(\widehat{u}) \geq f(u_*) + \langle x_*, b - A\widehat{u}\rangle$, and the latter quantity is $\geq f(u_*) - L\|[A\widehat{u} - b]^+\|_q$ due to $x_* \geq 0$ and $\|x_*\|_p \leq L$. Thus, $f(\widehat{u}) \geq f(u_*) - L\|[A\widehat{u} - b]^+\|_q$, which combines with (62) to imply (45.a. ■

References

[1] A. Ben-Tal and A. Nemirovski, Lectures on Modern Convex Optimization, SIAM, Philadelphia, 2001.

[2] A. Ben-Tal, A. Nemirovski, Non-Euclidean restricted memory level method for large-scale convex optimization. *Math. Progr.* 102, 407–456, 2005.

[3] S. Boyd and L. Vandenberghe, Convex Optimization, Cambridge University Press, 2004.

[4] V.P. Bulatov and L.O. Shepot'ko, Method of centers of orthogonal simpexes for solving convex programming problems (in Russian), In: *Methods of Optimization and Their Application*, Nauka, Novosibirsk 1982 .

[5] B.P. Burrell, M.J. Todd, Ellipsoid method generates dual variables, *Mathematics of Operations Research* **10:4** (1985), 688–700.

[6] D. Gabelev (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 32000, Israel E-print: http://www2.isye.gatech.edu/ nemirovs/Dima.pdf

[7] Fritz John, Extremum problems with inequalities as subsidiary conditions, in: *Studies and Essays Presented to R. Courant on his 60th Birthday, January 8, 1948*, Interscience Publishers, Inc., New York, N. Y., 1948, pp. 187–204.

[8] M. Grötschel, L. Lovasz and A. Schrijver, *The Ellipsoid Method and Combinatorial Optimizatoon.* Springer-Verlag, 1986.

[9] P.T. Harker and J.-S. Pang, Finite-dimensional variational inequality and nonlinear complementarity problems: a survey of theory, algorithms and applications, *Mathematical Programming* 48:161-220, 1990.

[10] L.G. Khachiyan, A polynomial algorithm in linear programming, *Soviet Mathematics Doklady*, 20:191–194, 1979.

[11] C. Lemarechal, A. Nemirovski, Yu. Nesterov, New variants of bundle methods. *Math. Progr.* 69:1, 111-148, 1995.

[12] A.Yu. Levin, On an algorithm for the minimization of convex functions (in Russian), *Doklady Akad. Nauk SSSR* **160:6** (1965), 1244–1247. (English translation: Soviet Mathematics Doklady, 6:286–290, 1965.)

[13] G.J. Minty, Monotone non-linear operators in Hilber space, *Duke Mathematics Journal*, 29:341-346, 1962.

[14] A. Nemirovski, Polynomial time methods in Convex Programming, in: J. Renegar, M. Shub and S. Smale, Eds., *The Mathematics of Numerical Analysis*, AMS-SIAM Summer Seminar on Mathematics in Applied Mathematics, July 17– August 11, 1995, Park City, Utah. Lectures in Applied Mathematics, AMS, Providence, 32:543-589, 1996.

[15] A. Nemirovski, 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. on Optim.* 15, 229–251, 2004.

[16] Y. Nesterov and A. Nemirovskii, *Interior-Point Polynomial Algorithms in Convex Programming*, SIAM, Philadelphia, 1994.

[17] Yu. Nesterov, Smooth minimization of non-smooth functions, *Math. Progr.* 103:1, 127–152, 2005.

[18] D.J. Newman, Location of maximum on unimodal surfaces, *Journ. of the Assoc. for Computing Machinery* 12:11-23, 1965.

[19] N.Z. Shor, Cutting plane method with space dilation for the solution of convex programming problems (in Russian), *Kibernetika*, 1:94-95, 1977.

[20] S.P. Tarasov, L.G. Khachiyan and I.I. Erlikh, The method of inscribed ellipsoids, *Soviet Mathematics Doklady* 37:226–230, 1988.

[21] B. Yamnitsky and L. Levin, An Old Linear Programming Algorithm Runs in Polynomoial Time, In: 23rd Annual Symposium on Foundations of Computer Science, IEEE, New York, 327-328, 1982.

[22] D. Yudin and A. Nemirovskii, Informational complexity and effective methods of solution for convex extremal problems (in Russian), *Ekonomika i Matematicheskie Metody* 12:2 (1976), 357–369 (translated into English as *Matekon*: Transl. Russian and East European Math. Economics **13** 25–45, Spring '77).

Shmuel Onn
Technion - Israel Institute of Technology, 32000 Haifa, Israel
email: onn@ie.technion.ac.il
http://ie.technion.ac.il/~onn

Arkadi Nemirovski
Georgia Institute of Technology, Atlanta, Georgia xxxxx, USA
email: nemirovs@isye.gatech.edu
http://www2.isye.gatech.edu/~nemirovs/

Uriel G. Rothblum
Technion - Israel Institute of Technology, 32000 Haifa, Israel
email: rothblum@ie.technion.ac.il
*http://ie.technion.ac.il/*rothblum.phtml