Subgradient Methods for Saddle-Point Problems

Angelia Nedić* and Asuman Ozdaglar†

August 19, 2007

Abstract

We consider computing the saddle points of a convex-concave function using subgradient methods. The existing literature on finding saddle points has mainly focused on establishing convergence properties of the generated iterates under some restrictive assumptions. In this paper, we propose a subgradient algorithm for generating approximate saddle points and provide per-iteration convergence rate estimates on the constructed solutions. We then focus on Lagrangian duality, where we consider a convex primal optimization problem and its Lagrangian dual problem, and generate approximate primal-dual optimal solutions as approximate saddle points of the Lagrangian function. We present a variation of our subgradient method under the Slater constraint qualification and provide stronger estimates on the convergence rate of the generated primal sequences. In particular, we provide bounds on the amount of feasibility violation and on the primal objective function values at the approximate solutions. Our algorithm is particularly well-suited for problems where the subgradient of the dual function cannot be evaluated easily (equivalently, the minimum of the Lagrangian function at a dual solution cannot be computed efficiently), thus impeding the use of dual subgradient methods.

Keywords: saddle-point subgradient methods, averaging, approximate primal solutions, primal-dual subgradient methods, convergence rate.

*Department of Industrial and Enterprise Systems Engineering, University of Illinois at Urbana-Champaign, angelia@uiuc.edu

†Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, asuman@mit.edu

1 Introduction

In this paper, we consider a convex-concave function $\mathcal{L} : X \times M \to \mathbb{R}$, where X and M are closed convex sets in $\mathbb{R}^n$ and $\mathbb{R}^m$. We are interested in computing a saddle point (x^*, μ^*) of $\mathcal{L}(x, \mu)$ over the set $X \times M$, where a saddle point is defined as a vector pair (x^*, μ^*) that satisfies

$$\mathcal{L}(x^*, \mu) \leq \mathcal{L}(x^*, \mu^*) \leq \mathcal{L}(x, \mu^*) \qquad \text{for all } x \in X,\ \mu \in M.$$

Saddle point problems arise in a number of areas such as constrained optimization duality, zero-sum games, and general equilibrium theory. It has been long recognized that gradient methods (or more generally subgradient methods for the non-differentiable case) provide efficient decentralized computational means for solving saddle point and optimization problems in many disciplines. For example, as long ago as 1949, Samuelson [27] wrote: "The gradient method may be considered as a decentralized or computational mechanism for achieving optimum allocation of scarce resources."

We propose a subgradient algorithm for generating approximate saddle-point solutions for a convex-concave function. Our algorithm builds on the seminal Arrow-Hurwicz-Uzawa algorithm [1] and the averaging scheme suggested by Nemirovski and Yudin [24]. In contrast to existing work with focus on the convergence of the iterates to a saddle point, we present an algorithm that generates approximate saddle points and provide explicit rate estimates per-iteration. In particular, we establish upper and lower bounds on the function value of the generated solutions. For the case of Lagrangian saddle point problems, we provide a variation on our main algorithm and establish stronger convergence and convergence rate results under the Slater constraint qualification.

Subgradient methods for solving saddle point problems have been the focus of much research since the seminal work of Arrow, Hurwicz, and Uzawa [1]. Arrow, Hurwicz, and Uzawa proposed continuous-time versions of these methods for general convex-concave functions. They provide global stability results under strict convexity assumptions in a series of papers in the collection [1]. Uzawa in [31] focused on a discrete-time version of the subgradient method with a constant stepsize rule and showed that the iterates converge to any given neighborhood of a saddle point provided that the stepsize is sufficiently small.[1] Methods of Arrow-Hurwicz-Uzawa type for finding saddle-points of a general convex-concave function have also been studied by Gol'shtein [10] and Maistroskii [19], who provided convergence results using diminishing stepsize rule and under some stability assumptions on saddle points, which are weaker than strict convexity assumptions. Zabotin [32] has extended the preceding work by establishing convergence without assuming saddle-point stability. Korpelevich [14] considered an "extragradient method" for computing saddle points, which can be viewed as a gradient method with perturbations. Perturbations have also been used more recently by Kallio and Ruszczyński [13] to construct a class of subgradient methods for computing saddle points with an adaptive stepsize rule that uses the "primal-dual gap" information. These methods have been further developed by Kallio and Rosa [12] and used in computing the saddle points

[1]Note that, though not explicitly stated in [31], Uzawa's proof holds under the assumption that the function $\mathcal{L}(x, \mu)$ is strictly convex in x.

of the standard Lagrangian function. Nemirovski and Yudin [24] considered a different approach where they combined the subgradient method with a simple averaging scheme in the context of Lagrangian saddle point problems and provided convergence results with adaptive stepsize rules.

In this paper, we focus on the computation of approximate saddle points (as opposed to asymptotically exact solutions) using subgradient methods with a constant stepsize. We consider constant stepsize rule because of its simplicity and practical relevance, and because our interest is in generating approximate solutions in finite number of iterations. Our first main result can be summarized as follows. Let $\mathcal{L}$ be a convex-concave function with saddle point (x^*, μ^*). Let $\{x_k, \mu_k\}$ be the iterates generated by our subgradient algorithm. Then, under the assumption that the generated iterates are bounded, the function value $\mathcal{L}(\hat{x}_k, \hat{\mu}_k)$ at the time-averaged iterates $\hat{x}_k$ and $\hat{\mu}_k$ converges to $\mathcal{L}(x^*, \mu^*)$ at rate $1/k$ within some error level explicitly given as a function of the stepsize.

Our second set of results focus on Lagrangian saddle point problems and offer a variant of our main algorithm with stronger error estimates for the averaged primal solution under the Slater constraint qualification. These results are obtained without assuming the boundedness of the iterates owing to the fact that we can exploit the special structure of the optimal solution set of the Lagrangian dual problem under the Slater condition. The error estimates are in terms of the amount of constraint violation and the primal objective function value for the generated solutions. This method can be applied in wide range of problems where the standard subgradient methods cannot be used to solve the dual problem because of the difficulties associated with efficiently computing subgradients of the dual function.

In addition to the papers cited above, our work is related to recent work on subgradient methods based on non-Euclidean projections. Beck and Teboulle [4] for example, consider a non-differentiable convex optimization problem with simple constraints and study the Mirror Descent Algorithm (MDA), first proposed by Nemirovski and Yudin [24]. They show that MDA achieves a rate estimate of $1/\sqrt{k}$ for the objective function value of the generated iterates. In related work, Nemirovski [23] proposes a prox-type method with averaging for computing saddle points of continuously differentiable convex-concave functions with Lipschitz continuous gradient and provides a rate estimate of $1/k$ for the saddle function value. Auslender and Teboulle [2, 3] propose subgradient methods based on non-Euclidean projections for solving variational inequalities and non-differentiable convex optimization problems. Under the assumption that the constraint set is compact, they provide a rate estimate of $1/\sqrt{k}$. Our contribution relative to these works is that we provide a rate estimate of $1/k$ for a constant stepsize subgradient method with standard Euclidean projection and averaging for computing approximate saddle points of a general (not necessarily differentiable) convex-concave function. In addition, our approach leads to a natural algorithm for solving non-differentiable convex constrained optimization problems under the Slater constraint qualification. In this case, the rate estimate of $1/k$ holds even when the boundedness assumption on the iterates is relaxed. Finally, our error bounds highlight the explicit tradeoffs between the accuracy of the solutions and computational complexity of the algorithm as a function of the stepsize.

The paper is organized as follows. In Section 2, we introduce the notation and the basic notions that we use throughout the paper and present a formulation of the saddle point problem. In Section 3, we present our subgradient method and provide relations for the iterates of the method. We construct approximate solutions by considering running averages of the generated iterates and provide error estimates for these solutions. In Section 4, we consider application of our subgradient method to Lagrangian duality. In Section 5, we propose a variation on the original algorithm under the Slater condition and present error estimates for the averaged primal solutions. In Section 6, we summarize our results and provide some concluding remarks.

2 Preliminaries

In this section, we formulate the saddle-point problem of interest and present some preliminary results that we use in the subsequent development. We start by introducing the notation and basic terminology that we use throughout the paper.

2.1 Notation and Terminology

We consider the n-dimensional vector space $\mathbb{R}^n$ and the m-dimensional vector space $\mathbb{R}^m$. We denote the m-dimensional nonnegative orthant by $\mathbb{R}^m_+$. We view a vector as a column vector, and we write x^i to denote the i-th component of a vector x. We denote by $x'y$ the inner product of two vectors x and y. We use $\|y\|$ to denote the standard Euclidean norm, $\|y\| = \sqrt{y'y}$. Occasionally, we also use the standard 1-norm and ∞-norm denoted respectively by $\|y\|_1$ and $\|y\|_\infty$, i.e., $\|y\|_1 = \sum_i |y^i|$ and $\|y\|_\infty = \max_i |y^i|$. We write $dist(\bar{y}, Y)$ to denote the standard Euclidean distance of a vector $\bar{y}$ from a set Y, i.e.,

$$dist(\bar{y}, Y) = \inf_{y \in Y} \|\bar{y} - y\|.$$

For a vector $u \in \mathbb{R}^m$, we write u^+ to denote the projection of u on the nonnegative orthant in $\mathbb{R}^m$, i.e., u^+ is the component-wise maximum of the vector u and the zero vector:

$$u^+ = (\max\{0, u^1\}, \cdots, \max\{0, u^m\})' \quad \text{for } u = (u^1, \cdots, u^m)'.$$

For a convex function $F : \mathbb{R}^n \to [-\infty, \infty]$, we denote the domain of F by $\mathrm{dom}(F)$, where

$$\mathrm{dom}(F) = \{x \in \mathbb{R}^n \mid F(x) < \infty\}.$$

We use the notion of a subgradient of a convex function $F(x)$ at a given vector $\bar{x} \in \mathrm{dom}(F)$. A subgradient $s_F(\bar{x})$ of a convex function $F(x)$ at any $\bar{x} \in \mathrm{dom}(F)$ provides a linear underestimate of the function F. In particular, $s_F(\bar{x}) \in \mathbb{R}^n$ is a *subgradient of a convex function* $F : \mathbb{R}^n \to \mathbb{R}$ *at a given vector* $\bar{x} \in \mathrm{dom}(F)$ when the following relation holds:

$$F(\bar{x}) + s_F(\bar{x})'(x - \bar{x}) \le F(x) \qquad \text{for all } x \in \mathrm{dom}(F). \tag{1}$$

The set of all subgradients of F at $\bar{x}$ is denoted by $\partial F(\bar{x})$.

Similarly, for a concave function $q : \mathbb{R}^m \to [-\infty, \infty]$, we denote the domain of q by $\mathrm{dom}(q)$, where

$$\mathrm{dom}(q) = \{\mu \in \mathbb{R}^m \mid q(\mu) > -\infty\}.$$

concave function $q(\mu)$. A subgradient of a concave function is defined through a subgradient of a convex function $-q(\mu)$. In particular, $s_q(\bar{\mu}) \in \mathbb{R}^m$ is a *subgradient of a concave function $q(\mu)$ at a given vector $\bar{\mu} \in \mathrm{dom}(q)$* when the following relation holds:

$$q(\bar{\mu}) + s_q(\bar{\mu})'(\mu - \bar{\mu}) \geq q(\mu) \qquad \text{for all } \mu \in \mathrm{dom}(q). \tag{2}$$

Similarly, the set of all subgradients of q at $\bar{\mu}$ is denoted by $\partial q(\bar{\mu})$.

2.2 Saddle-Point Problem

We consider the following saddle-point problem

$$\min_{x \in X} \max_{\mu \in M} \mathcal{L}(x, \mu), \tag{3}$$

where X is a closed convex set in $\mathbb{R}^n$, M is a closed convex set in $\mathbb{R}^m$, and $\mathcal{L}$ is a convex-concave function defined over $X \times M$. In particular, $\mathcal{L}(\cdot, \mu) : X \to \mathbb{R}$ is convex for every $\mu \in M$, and $\mathcal{L}(x, \cdot) : M \to \mathbb{R}$ is concave for every $x \in X$. For any given $(\bar{x}, \bar{\mu}) \in X \times M$, the subdifferential set of $\mathcal{L}(\bar{x}, \bar{\mu})$ with respect to x is denoted by $\partial_x \mathcal{L}(\bar{x}, \bar{\mu})$, while the subdifferential set of $\mathcal{L}(\bar{x}, \bar{\mu})$ with respect to μ is denoted by $\partial_\mu \mathcal{L}(\bar{x}, \bar{\mu})$. We assume that these *subdifferential sets are nonempty for all* $(\bar{x}, \bar{\mu}) \in X \times M$. We use $\mathcal{L}_x(\bar{x}, \bar{\mu})$ and $\mathcal{L}_\mu(\bar{x}, \bar{\mu})$ to denote a subgradient of $\mathcal{L}$ with respect to x and a subgradient of $\mathcal{L}$ with respect to μ at any $(\bar{x}, \bar{\mu}) \in X \times M$.

A solution to the problem of Eq. (3) is a vector pair $(x^*, \mu^*) \in X \times M$ such that

$$\mathcal{L}(x^*, \mu) \leq \mathcal{L}(x^*, \mu^*) \leq \mathcal{L}(x, \mu^*) \qquad \text{for all } x \in X \text{ and } \mu \in M. \tag{4}$$

Such a vector pair (x^*, μ^*) is also referred to as *saddle point* of the function $\mathcal{L}(x, \mu)$ over the set $X \times M$.

3 Subgradient Algorithm for Approximate Saddle-Points

To generate approximate solutions to problem (3), we consider a subgradient method motivated by Arrow-Hurwicz-Uzawa algorithm [1]. In particular, the method has the following form:

$$x_{k+1} = \mathcal{P}_X \left[x_k - \alpha \mathcal{L}_x(x_k, \mu_k) \right] \qquad \text{for } k = 0, 1, \ldots, \tag{5}$$

$$\mu_{k+1} = \mathcal{P}_M [\mu_k + \alpha \mathcal{L}_\mu(x_k, \mu_k)] \qquad \text{for } k = 0, 1, \ldots, \tag{6}$$

where $\mathcal{P}_X$ and $\mathcal{P}_M$ denote the projection on sets X and M respectively. The vectors $x_0 \in X$ and $\mu_0 \in M$ are initial iterates, and the scalar $\alpha > 0$ is a constant stepsize. The vectors $\mathcal{L}_x(x_k, \mu_k)$ and $\mathcal{L}_\mu(x_k, \mu_k)$ denote subgradients of $\mathcal{L}$ at (x_k, μ_k) with respect to x and μ, correspondingly.

Our focus on a constant stepsize is motivated by the fact that we are interested in quantifying the progress of the algorithm in finite number of iterations. An adaptive diminishing stepsize α_k that varies with each iteration may be used if the interest is in establishing the convergence properties of the iterates to a saddle point as the number of iterations k goes to infinity.

3.1 Basic Relations

In this section, we establish some basic relations that hold for the sequences $\{x_k\}$ and $\{\mu_k\}$ obtained by the algorithm in Eqs. (5)–(6). These properties are important in our construction of approximate primal solutions, and in particular, in our analysis of error estimates of these solutions.

We start with a lemma providing relations that hold under minimal assumptions. The relations given in part (b) of this lemma have been known and used extensively to analyze dual subgradient approaches (for example, see Shor [29], Polyak [26], Demyanov and Vasilev [9], Correa and Lemaréchal [8], Nedić and Bertsekas [20], [21], Nedić and Ozdaglar [22]). The proofs are included here for completeness.

Lemma 1 *(Basic Iterate Relations)* Let the sequences $\{x_k\}$ and $\{\mu_k\}$ be generated by the subgradient algorithm (5)–(6). We then have:

(a) For any $x \in X$ and for all $k \geq 0$,

$$\|x_{k+1} - x\|^2 \leq \|x_k - x\|^2 - 2\alpha \left(\mathcal{L}(x_k, \mu_k) - \mathcal{L}(x, \mu_k)\right) + \alpha^2 \left\|\mathcal{L}_x(x_k, \mu_k)\right\|^2.$$

(b) For any $\mu \in M$ and for all $k \geq 0$,

$$\|\mu_{k+1} - \mu\|^2 \leq \|\mu_k - \mu\|^2 + 2\alpha \left(\mathcal{L}(x_k, \mu_k) - \mathcal{L}(x_k, \mu)\right) + \alpha^2 \|\mathcal{L}_\mu(x_k, \mu_k)\|^2.$$

Proof.
(a) By using the nonexpansive property of the projection operation and relation (5) we obtain for any $x \in X$ and all $k \geq 0$,

$$\begin{aligned} \|x_{k+1} - x\|^2 &= \|\mathcal{P}_X[x_k - \alpha \mathcal{L}_x(x_k, \mu_k)] - x\|^2 \\ &\leq \|x_k - \alpha \mathcal{L}_x(x_k, \mu_k) - x\|^2 \\ &= \|x_k - x\|^2 - 2\alpha \mathcal{L}_x'(x_k, \mu_k)(x_k - x) + \alpha^2 \left\|\mathcal{L}_x(x_k, \mu_k)\right\|^2. \end{aligned}$$

Since the function $\mathcal{L}(x, \mu)$ is convex in x for each $\mu \in M$, and since $\mathcal{L}_x(x_k, \mu_k)$ is a subgradient of $\mathcal{L}(x, \mu_k)$ with respect to x at $x = x_k$ [cf. the definition of a subgradient in Eq. (1)], we obtain for any x,

$$\mathcal{L}_x'(x_k, \mu_k)(x - x_k) \leq \mathcal{L}(x, \mu_k) - \mathcal{L}(x_k, \mu_k),$$

or equivalently

$$-\mathcal{L}_x'(x_k, \mu_k)(x_k - x) \leq -\left(\mathcal{L}(x_k, \mu_k) - \mathcal{L}(x, \mu_k)\right).$$

Hence, for any $x \in X$ and all $k \geq 0$,

$$\|x_{k+1} - x\|^2 \leq \|x_k - x\|^2 - 2\alpha \left(\mathcal{L}(x_k, \mu_k) - \mathcal{L}(x, \mu_k)\right) + \alpha^2 \left\|\mathcal{L}_x(x_k, \mu_k)\right\|^2.$$

(b) Similarly, by using the nonexpansive property of the projection operation and relation (6) we obtain for any $\mu \in M$,

$$\|\mu_{k+1} - \mu\|^2 = \left\|[\mu_k + \alpha \mathcal{L}_\mu(x_k, \mu_k)]^+ - \mu\right\|^2 \leq \|\mu_k + \alpha \mathcal{L}_\mu(x_k, \mu_k) - \mu\|^2.$$

Therefore,

$$\|\mu_{k+1} - \mu\|^2 \leq \|\mu_k - \mu\|^2 + 2\alpha(\mu_k - \mu)'\mathcal{L}_\mu(x_k, \mu_k) + \alpha^2\|\mathcal{L}_\mu(x_k, \mu_k)\|^2 \qquad \text{for all } k.$$

Since $\mathcal{L}_\mu(x_k, \mu_k)$ is a subgradient of the concave function $\mathcal{L}(x_k, \mu)$ at $\mu = \mu_k$ [cf. Eq. (2)], we have for all μ,

$$(\mu_k - \mu)'\mathcal{L}_\mu(x_k, \mu_k) = \mathcal{L}(x_k, \mu_k) - \mathcal{L}(x_k, \mu).$$

Hence, for any $\mu \in M$ and all $k \geq 0$,

$$\|\mu_{k+1} - \mu\|^2 \leq \|\mu_k - \mu\|^2 + 2\alpha\left(\mathcal{L}(x_k, \mu_k) - \mathcal{L}(x_k, \mu)\right) + \alpha^2\|\mathcal{L}_\mu(x_k, \mu_k)\|^2.$$

■

We establish some additional properties of the iterates x_k and μ_k under the assumption that the subgradients used by the method are bounded. This is formally stated in the following.

Assumption 1 *(Subgradient Boundedness)* The subgradients $\mathcal{L}_x(x_k, \mu_k)$ and $\mathcal{L}_\mu(x_k, \mu_k)$ used in the method defined by (5)–(6) are uniformly bounded, i.e., there is a constant $L > 0$ such that

$$\|\mathcal{L}_x(x_k, \mu_k)\| \leq L, \qquad \|\mathcal{L}_\mu(x_k, \mu_k)\| \leq L \qquad \text{for all } k \geq 0.$$

This assumption is satisfied for example when the sets X and M are compact and the function $\mathcal{L}$ is continuous over $X \times M$. Also, the assumption is satisfied when the function $\mathcal{L}$ is affine in (x, μ).

Under the preceding assumption, we provide a relation for the iterates x_k and μ_k and an arbitrary pair $(x, \mu) \in X \times M$. This relation plays a crucial role in our subsequent analysis.

Lemma 2 Let the sequences $\{x_k\}$ and $\{\mu_k\}$ be generated by the subgradient algorithm (5)–(6). Let Subgradient Boundedness assumption hold (cf. Assumption 1), and let $\hat{x}_k$ and $\hat{\mu}_k$ be the iterate averages given by

$$\hat{x}_k = \frac{1}{k}\sum_{i=0}^{k-1} x_i, \qquad \hat{\mu}_k = \frac{1}{k}\sum_{i=0}^{k-1} \mu_i.$$

We then have for all $k \geq 1$,

$$\frac{1}{k}\sum_{i=0}^{k-1} \mathcal{L}(x_i, \mu_i) - \mathcal{L}(x, \hat{\mu}_k) \leq \frac{\|x_0 - x\|^2}{2\alpha k} + \frac{\alpha L^2}{2} \qquad \text{for any } x \in X, \tag{7}$$

$$-\frac{\|\mu_0 - \mu\|^2}{2\alpha k} - \frac{\alpha L^2}{2} \leq \frac{1}{k}\sum_{i=0}^{k-1} \mathcal{L}(x_i, \mu_i) - \mathcal{L}(\hat{x}_k, \mu) \qquad \text{for any } \mu \in M. \tag{8}$$

Proof. We first show the relation in Eq. (7). By using Lemma 1(a) and the boundedness of the subgradients $\mathcal{L}_x(x_i, \mu_i)$ [cf. Assumption 1], we have for any $x \in X$ and $i \geq 0$,

$$\|x_{i+1} - x\|^2 \leq \|x_i - x\|^2 - 2\alpha \left(\mathcal{L}(x_i, \mu_i) - \mathcal{L}(x, \mu_i)\right) + \alpha^2 L^2.$$

Therefore,

$$\mathcal{L}(x_i, \mu_i) - \mathcal{L}(x, \mu_i) \leq \frac{1}{2\alpha} \left(\|x_i - x\|^2 - \|x_{i+1} - x\|^2\right) + \frac{\alpha L^2}{2}.$$

By adding these relations over $i = 0, \ldots, k-1$, we obtain for any $x \in X$ and $k \geq 1$,

$$\sum_{i=0}^{k-1} \left(\mathcal{L}(x_i, \mu_i) - \mathcal{L}(x, \mu_i)\right) \leq \frac{1}{2\alpha} \left(\|x_0 - x\|^2 - \|x_k - x\|^2\right) + \frac{k\alpha L^2}{2},$$

implying that

$$\frac{1}{k} \sum_{i=0}^{k-1} \mathcal{L}(x_i, \mu_i) - \frac{1}{k} \sum_{i=0}^{k-1} \mathcal{L}(x, \mu_i) \leq \frac{\|x_0 - x\|^2}{2\alpha k} + \frac{\alpha L^2}{2}.$$

Since the function $\mathcal{L}(x, \mu)$ is concave in μ for any fixed $x \in X$, there holds

$$\mathcal{L}(x, \hat{\mu}_k) \geq \frac{1}{k} \sum_{i=0}^{k-1} \mathcal{L}(x, \mu_i) \qquad \text{with } x \in X \text{ and } \hat{\mu}_k = \frac{1}{k} \sum_{i=0}^{k-1} \mu_i.$$

Combining the preceding two relations, we obtain for any $x \in X$ and $k \geq 1$,

$$\frac{1}{k} \sum_{i=0}^{k-1} \mathcal{L}(x_i, \mu_i) - \mathcal{L}(x, \hat{\mu}_k) \leq \frac{\|x_0 - x\|^2}{2\alpha k} + \frac{\alpha L^2}{2},$$

thus establishing relation (7).

Similarly, by using Lemma 1(b) and the boundedness of the subgradients $\mathcal{L}_\mu(x_i, \mu_i)$, we have for any $\mu \in M$ and $i \geq 0$,

$$\|\mu_{i+1} - \mu\|^2 \leq \|\mu_i - \mu\|^2 + 2\alpha \left(\mathcal{L}(x_i, \mu_i) - \mathcal{L}(x_i, \mu)\right) + \alpha^2 L^2.$$

Hence,

$$\frac{1}{2\alpha} \left(\|\mu_{i+1} - \mu\|^2 - \|\mu_i - \mu\|^2\right) - \frac{\alpha L^2}{2} \leq \mathcal{L}(x_i, \mu_i) - \mathcal{L}(x_i, \mu).$$

By adding these relations over $i = 0, \ldots, k-1$, we obtain for all $\mu \in M$ and $k \geq 1$,

$$\frac{1}{2\alpha} \left(\|\mu_k - \mu\|^2 - \|\mu_0 - \mu\|^2\right) - \frac{k\alpha L^2}{2} \leq \sum_{i=0}^{k-1} \left(\mathcal{L}(x_i, \mu_i) - \mathcal{L}(x_i, \mu)\right),$$

implying that

$$-\frac{\|\mu_0 - \mu\|^2}{2\alpha k} - \frac{\alpha L^2}{2} \leq \frac{1}{k} \sum_{i=0}^{k-1} \mathcal{L}(x_i, \mu_i) - \frac{1}{k} \sum_{i=0}^{k-1} \mathcal{L}(x_i, \mu).$$

Because the function $\mathcal{L}(x,\mu)$ is convex in x for any fixed $\mu \in M$, we have

$$\frac{1}{k}\sum_{i=0}^{k-1}\mathcal{L}(x_i,\mu) \geq \mathcal{L}(\hat{x}_k,\mu) \qquad \text{with } \mu \in M \ \text{ and } \ \hat{x}_k = \frac{1}{k}\sum_{i=0}^{k-1} x_i.$$

Combining the preceding two relations, we obtain for all $\mu \in M$ and $k \geq 1$,

$$-\frac{\|\mu_0-\mu\|^2}{2\alpha k} - \frac{\alpha L^2}{2} \leq \frac{1}{k}\sum_{i=0}^{k-1}\mathcal{L}(x_i,\mu_i) - \mathcal{L}(\hat{x}_k,\mu),$$

thus showing relation (8). ■

The relations in the previous lemma will be key in providing *approximate saddle points* with per-iteration performance bounds in the following section. In particular, using this lemma, we establish a relation between the averaged values $\frac{1}{k}\sum_{i=0}^{k-1}\mathcal{L}(x_i,\mu_i)$ and the saddle point value $\mathcal{L}(x^*,\mu^*)$. We also show a relation between the function value $\mathcal{L}(\hat{x}_k,\hat{\mu}_k)$ at the iterate averages $\hat{x}_k$ and $\hat{\mu}_k$, and the saddle point value $\mathcal{L}(x^*,\mu^*)$.

3.2 Approximate Saddle Points

In this section, we focus on constructing approximate saddle points using the information generated by the subgradient method given in Eqs. (5)-(6). We use a simple averaging scheme to generate approximate solutions. In particular, we consider the running averages $\hat{x}_k$ and $\hat{\mu}_k$ generated by

$$\hat{x}_k = \frac{1}{k}\sum_{i=0}^{k-1} x_i, \qquad \hat{\mu}_k = \frac{1}{k}\sum_{i=0}^{k-1}\mu_i \qquad \text{for } k \geq 1.$$

These averages provide approximate solutions for the saddle-point problem, as indicated by the following result.

Proposition 1 Let Subgradient Boundedness Assumption hold (cf. Assumption 1). Let $\{x_k\}$ and $\{\mu_k\}$ be the sequences generated by method (5)–(6), and let $(x^*,\mu^*) \in X \times M$ be a saddle point of $\mathcal{L}(x,\mu)$. We then have:

(a) For all $k \geq 1$,

$$-\frac{\|\mu_0-\mu^*\|^2}{2\alpha k} - \frac{\alpha L^2}{2} \leq \frac{1}{k}\sum_{i=0}^{k-1}\mathcal{L}(x_i,\mu_i) - \mathcal{L}(x^*,\mu^*) \leq \frac{\|x_0-x^*\|^2}{2\alpha k} + \frac{\alpha L^2}{2}.$$

(b) The averages $\hat{x}_k$ and $\hat{\mu}_k$ satisfy the following relation for all $k \geq 1$:

$$-\frac{\|\mu_0-\mu^*\|^2+\|x_0-\hat{x}_k\|^2}{2\alpha k} - \alpha L^2 \leq \mathcal{L}(\hat{x}_k,\hat{\mu}_k) - \mathcal{L}(x^*,\mu^*) \leq \frac{\|x_0-x^*\|^2+\|\mu_0-\hat{\mu}_k\|^2}{2\alpha k} + \alpha L^2,$$

where L is the subgradient bound of Assumption 1.

Proof. (a) Our proof is based on Lemma 2. In particular, by letting $x = x^*$ and $\mu = \mu^*$ in Eqs. (7) and (8), respectively, we obtain for any $k \geq 1$,

$$\frac{1}{k}\sum_{i=0}^{k-1} \mathcal{L}(x_i, \mu_i) - \mathcal{L}(x^*, \hat{\mu}_k) \leq \frac{\|x_0 - x^*\|^2}{2\alpha k} + \frac{\alpha L^2}{2},$$

$$-\frac{\|\mu_0 - \mu^*\|^2}{2\alpha k} - \frac{\alpha L^2}{2} \leq \frac{1}{k}\sum_{i=0}^{k-1} \mathcal{L}(x_i, \mu_i) - \mathcal{L}(\hat{x}_k, \mu^*).$$

By convexity of the sets X and M, we have $\hat{x}_k \in X$ and $\hat{\mu}_k \in M$ for all $k \geq 1$. Therefore, by the saddle-point relation [cf. Eq. (4)], we have

$$\mathcal{L}(x^*, \hat{\mu}_k) \leq \mathcal{L}(x^*, \mu^*) \leq \mathcal{L}(\hat{x}_k, \mu^*).$$

Combining the preceding three relations, we obtain for all $k \geq 1$,

$$-\frac{\|\mu_0 - \mu^*\|^2}{2\alpha k} - \frac{\alpha L^2}{2} \leq \frac{1}{k}\sum_{i=0}^{k-1} \mathcal{L}(x_i, \mu_i) - \mathcal{L}(x^*, \mu^*) \leq \frac{\|x_0 - x^*\|^2}{2\alpha k} + \frac{\alpha L^2}{2}. \tag{9}$$

(b) Since $\hat{x}_k \in X$ and $\hat{\mu}_k \in M$ for all $k \geq 1$, we use Lemma 2 with $x = \hat{x}_k$ and $\mu = \hat{\mu}_k$ to obtain for all $k \geq 1$,

$$\frac{1}{k}\sum_{i=0}^{k-1} \mathcal{L}(x_i, \mu_i) - \mathcal{L}(\hat{x}_k, \hat{\mu}_k) \leq \frac{\|x_0 - \hat{x}_k\|^2}{2\alpha k} + \frac{\alpha L^2}{2},$$

$$-\frac{\|\mu_0 - \hat{\mu}_k\|^2}{2\alpha k} - \frac{\alpha L^2}{2} \leq \frac{1}{k}\sum_{i=0}^{k-1} \mathcal{L}(x_i, \mu_i) - \mathcal{L}(\hat{x}_k, \hat{\mu}_k).$$

By multiplying with -1 the preceding relations and combining them, we see that

$$-\frac{\|x_0 - \hat{x}_k\|^2}{2\alpha k} - \frac{\alpha L^2}{2} \leq \mathcal{L}(\hat{x}_k, \hat{\mu}_k) - \frac{1}{k}\sum_{i=0}^{k-1} \mathcal{L}(x_i, \mu_i) \leq \frac{\|\mu_0 - \hat{\mu}_k\|^2}{2\alpha k} + \frac{\alpha L^2}{2}. \tag{10}$$

The result follows by summing relations (9) and (10). ∎

The result in part (a) of the preceding proposition provides bounds on the averaged function values $\frac{1}{k}\sum_{i=0}^{k-1} \mathcal{L}(x_i, \mu_i)$ in terms of the distances of the initial iterates μ_0 and x_0 from the vectors x^* and μ^* that constitute a saddle point of $\mathcal{L}$. In particular, the averaged function values $\frac{1}{k}\sum_{i=0}^{k-1} \mathcal{L}(x_i, \mu_i)$ converge to the saddle point value $\mathcal{L}(x^*, \mu^*)$ within error level $\alpha L^2/2$ with rate $1/k$. The result in part (b) gives bounds on the function value $\mathcal{L}(\hat{x}_k, \hat{\mu}_k)$ of the averaged iterates $\hat{x}_k$ and $\hat{\mu}_k$ in terms of the distances of the averaged iterates from the initial iterates and saddle point vectors. Under the assumption that the iterates generated by the subgradient algorithm (5)-(6) are bounded (which holds when the sets X and M are compact), this result shows that the function values of the averaged iterates $\mathcal{L}(\hat{x}_k, \hat{\mu}_k)$ converge to the saddle-point value $\mathcal{L}(x^*, \mu^*)$ within error level αL^2 with rate $1/k$. The error level is due to our use of a constant stepsize and can be controlled by choosing a smaller stepsize value. The estimate of the preceding proposition provides explicit tradeoffs between accuracy and computational complexity in choosing the stepsize value.

4 Lagrangian Duality

In this section, we consider a major application of the subgradient method developed so far, which is to the Lagrangian function of an optimization problem. In particular, we consider a constrained optimization problem, which we refer to as primal problem, and the corresponding Lagrangian dual problem. Motivated by the standard characterization of the primal-dual optimal solutions as the saddle points of the Lagrangian function, we use the subgradient method of Eqs. (5)-(6) to construct *approximate primal and dual optimal solutions* with per-iteration performance bounds. We start by introducing primal and dual problems.

4.1 Primal and Dual Problems

We consider the following constrained optimization problem:

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & g(x) \leq 0 \\ & x \in X, \end{array} \tag{11}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ is a convex function, $g = (g_1, \ldots, g_m)'$ and each $g_j : \mathbb{R}^n \to \mathbb{R}$ is a convex function, and $X \subset \mathbb{R}^n$ is a nonempty closed convex set. We refer to this problem as the *primal problem*. We denote the primal optimal value by f^*.

The dual problem of (11) is defined through Lagrangian relaxation of the inequality constraints $g(x) \leq 0$, and is given by

$$\begin{array}{ll} \text{maximize} & q(\mu) \\ \text{subject to} & \mu \geq 0 \\ & \mu \in \mathbb{R}^m. \end{array}$$

The dual objective function $q : \mathbb{R}^m_+ \to \mathbb{R}$ is defined by

$$q(\mu) = \inf_{x \in X} \mathcal{L}(x, \mu), \tag{12}$$

where $\mathcal{L}(x, \mu) : X \times \mathbb{R}^m_+ \to \mathbb{R}$ is the Lagrangian function defined by

$$\mathcal{L}(x, \mu) = f(x) + \mu' g(x). \tag{13}$$

We denote the dual optimal value by q^* and the dual optimal set by M^*.

It is well-known that for any $\mu \geq 0$, the dual function value $q(\mu)$ is a lower bound on the primal optimal value f^*, i.e., *weak duality* holds. Moreover, under standard Constraint Qualifications (such as the Slater condition which will be discussed in the following section), the optimal values of the primal and the dual problems are equal, i.e., there is *zero duality gap*, and there exists a dual optimal solution, i.e., the set M^* is nonempty (see for example Bertsekas [5] or Bertsekas, Nedić, and Ozdaglar [6]). Therefore, in view of the favorable structure of the dual problem, it is a common approach to solve the dual problem using subgradient methods, hence providing approximate dual optimal solutions and bounds on the primal optimal value in finite number of iterations.

More recent literature also considered exploiting the subgradient information generated in the dual space directly to produce approximate primal solutions (see Sherali and Choi [28], Larsson, Patriksson, and Strömberg [15, 16, 17], Nesterov [25], and Nedić and Ozdaglar [22]).

The use of subgradient methods for solving the dual problem relies on the assumption that the subgradients of the dual function can be evaluated efficiently at each iteration. Due to the structure of the dual function q, the subgradients of q at a vector μ are related to the vectors x_μ attaining the minimium in Eq. (12). Specifically, we have the following relation for the subdifferential set $\partial q(\mu)$ of the dual function q at a given $\mu \geq 0$ (see Bertsekas, Nedić, and Ozdaglar [6], Proposition 4.5.1):

$$\operatorname{conv}(\{g(x_\mu) \mid x_\mu \in X_\mu\}) \subseteq \partial q(\mu), \qquad X_\mu = \{x_\mu \in X \mid q(\mu) = f(x_\mu) + \mu' g(x_\mu)\}.$$

Hence, effective use of dual subgradient methods requires efficient computation of the minimizer x_μ in Eq. (12) at each iteration (which is possible when the Lagrangian function has a favorable structure). In many applications however, the Lagrangian function lacks any special structure that allows efficient computation of the minimizer x_μ, and therefore efficient computation of a subgradient of the dual function.

In the rest of the paper, we focus on this case. In particular, we want to investigate auxiliary procedures for generating directions that may be good approximations of the subgradients. Our development is motivated by the following standard result that characterizes the primal-dual optimal solutions as the saddle points of the Lagrangian function (see Bertsekas [5], Prop. 5.1.6).

Theorem 1 *(Saddle-Point Theorem)* The pair (x^*, μ^*) with $x^* \in X$ and $\mu^* \geq 0$ is a primal-dual optimal solution pair if and only if $x^* \in X$, $\mu^* \geq 0$, and the following relation holds:

$$\mathcal{L}(x^*, \mu) \leq \mathcal{L}(x^*, \mu^*) \leq \mathcal{L}(x, \mu^*) \qquad \text{for all } x \in X,\ \mu \geq 0,$$

i.e., (x^*, μ^*) is a saddle point of the Lagrangian function $\mathcal{L}(x, \mu)$.

Based on this characterization, we will use the subgradient method of the previous section for finding the saddle points of the Lagrangian function. The averaging scheme will allow us to construct approximate primal and dual optimal solutions with per-iteration error bounds. We then provide stronger error estimates under the Slater constraint qualification. The key feature that allows us to obtain stronger bounds is the boundedness of the dual optimal solution set under Slater condition. This feature is elaborated further in the following section.

4.2 Boundedness of the Optimal Dual Set

In this section, we show that any set of the form $\{\mu \geq 0 \mid q(\mu) \geq c\}$ for a fixed c (which includes the dual optimal set as a special case for $c = q^*$) is bounded under the standard Slater constraint qualification, formally given in the following.

Assumption 2 *(Slater Condition)* There exists a vector $\bar{x} \in \mathbb{R}^n$ such that

$$g_j(\bar{x}) < 0 \qquad \text{for all } j = 1, \dots, m.$$

We refer to a vector $\bar{x}$ satisfying the Slater condition as *a Slater vector*.

In addition to guaranteeing zero duality gap and the existence of a dual optimal solution, Slater condition also implies that the dual optimal set is bounded (see for example Hiriart-Urruty and Lemaréchal [11]). This property of the dual optimal set under the Slater condition, has been observed and used as early as in Uzawa's analysis of Arrow-Hurwicz gradient method in [31]. Nevertheless this fact has not been utilized in most analysis of subgradient methods in the literature[2].

The following proposition extends the result on the optimal dual set boundedness under the Slater condition. In particular, it shows that the Slater condition guarantees the boundedness of the (level) sets $\{\mu \geq 0 \mid q(\mu) \geq c\}$. This result was shown in Nedić and Ozdaglar [22]; therefore the proof is omitted here.

Lemma 3 Let $\bar{\mu} \geq 0$ be a vector, and consider the set $Q_{\bar{\mu}} = \{\mu \geq 0 \mid q(\mu) \geq q(\bar{\mu})\}$. Let Slater condition hold (cf. Assumption 2). Then, the set $Q_{\bar{\mu}}$ is bounded and, in particular, we have

$$\|\mu\|_1 \leq \frac{1}{\gamma}\ (f(\bar{x}) - q(\bar{\mu})) \qquad \text{for all } \mu \in Q_{\bar{\mu}},$$

where $\gamma = \min_{1 \leq j \leq m}\{-g_j(\bar{x})\}$ and $\bar{x}$ is a Slater vector.

An immediate implication of this result is that for any dual optimal solution μ^*, we have

$$\|\mu^*\|_1 \leq \frac{1}{\gamma}\ (f(\bar{x}) - q^*)\,. \tag{14}$$

This fact will be used in constructing a slight variation of the subgradient method of (5)-(6) in the following section.

5 Primal-Dual Subgradient Method

In this section, we apply the subgradient method of Eqs. (5)-(6) to the Lagrangian function of Eq. (13). Since the Lagrangian function $\mathcal{L}(x, \mu)$ is defined over $X \times \mathbb{R}^m_+$, an application of the method involves choosing the set M to be the nonnegative orthant $\mathbb{R}^m_+$. In this case the dual iterates μ_k generated by the algorithm need not be bounded. Note that the subgradients with respect to x and μ of the Lagrangian function $\mathcal{L}(x, \mu)$ at the vector (x_k, μ_k) are given by

$$\mathcal{L}_x(x_k, \mu_k) = s_f(x_k) + \sum_{j=1}^{m} \mu_k^j s_{g_j}(x_k), \qquad \mathcal{L}_\mu(x_k, \mu_k) = g(x_k), \tag{15}$$

where μ_k^j is the j-th component of the vector μ_k, while $s_f(x_k)$ and $s_{g_j}(x_k)$ are respectively subgradients of f and g_j at x_k. If the dual iterates μ_k are not bounded, then the subgradients $(\mathcal{L}_x(x_k, \mu_k), \mathcal{L}_\mu(x_k, \mu_k))$ need not be bounded. This violates the Subgradient

[2]See also our recent work [22] for the analysis of dual subgradient methods under Slater assumption.

Boundedness assumption (cf. Assumption 1). In the existing literature on saddle-point subgradient methods, such boundedness assumptions have been often assumed without any explicit guarantees on the boundedness of the dual iterates (see Gol'shtein [10] and Korpelevich [14]).

One of the contributions of this paper is the following variation of the subgradient method of (5)-(6) using the Slater condition, which allows us to dispense with the boundedness assumption on the dual iterates μ_k generated by the algorithm. In particular we consider a *primal-dual subgradient method* in which the iterates are generated by the following:

$$x_{k+1} = \mathcal{P}_X\left[x_k - \alpha \mathcal{L}_x(x_k, \mu_k)\right] \qquad \text{for } k = 0, 1, \ldots. \tag{16}$$

$$\mu_{k+1} = \mathcal{P}_D[\mu_k + \alpha \mathcal{L}_\mu(x_k, \mu_k)] \qquad \text{for } k = 0, 1, \ldots, \tag{17}$$

where the set D is a compact convex set that contains the set of dual optimal solutions (to be discussed shortly) and $\mathcal{P}_X$ and $\mathcal{P}_D$ denote the projection on the sets X and D respectively. The vectors $x_0 \in X$ and $\mu_0 \geq 0$ are initial iterates, and the scalar $\alpha > 0$ is a constant stepsize. Here $\mathcal{L}_x(x_k, \mu)$ denotes a subgradient with respect to x of the Lagrangian function $\mathcal{L}(x, \mu)$ at the vector x_k. Similarly, $\mathcal{L}_\mu(x, \mu_k)$ denotes a subgradient with respect to μ of the Lagrangian function $\mathcal{L}(x, \mu)$ at the vector μ_k [cf. Eq. (15)].

Under the Slater condition, the dual optimal set M^* is nonempty and bounded, and a bound on the norms of the dual optimal solutions is given by

$$\|\mu^*\|_1 \leq \frac{1}{\gamma}\left(f(\bar{x}) - q^*\right) \qquad \text{for all } \mu^* \in M^*,$$

with $\gamma = \min_{1 \leq j \leq m}\{-g_j(\bar{x})\}$ and $\bar{x}$ a Slater vector [cf. Eq. (14)]. Thus, having the dual value $\tilde{q} = q(\tilde{\mu})$ for some $\tilde{\mu} \geq 0$, since $q^* \geq \tilde{q}$, we obtain

$$\|\mu^*\|_1 \leq \frac{1}{\gamma}\left(f(\bar{x}) - \tilde{q}\right) \qquad \text{for all } \mu^* \in M^*. \tag{18}$$

This motivates the following choice for the set D:

$$D = \left\{\mu \geq 0 \;\Big|\; \|\mu\| \leq \frac{f(\bar{x}) - \tilde{q}}{\gamma} + r\right\}, \tag{19}$$

with a scalar $r > 0$. Clearly, the set D is compact and convex, and it contains the set of dual optimal solutions in view of relation (18) and the fact $\|y\| \leq \|y\|_1$ for any vector y. Under this choice of set D, the dual sequence $\{\mu_k\}$ is bounded. This allows us to make the following assumption:

Assumption 3 *(Bounded Subgradients)* Let the sequences $\{x_k\}$ and $\{\mu_k\}$ be generated by the subgradient algorithm (16)–(17). The subgradients $\mathcal{L}_x(x_k, \mu_k)$ and $\mathcal{L}_\mu(x_k, \mu_k)$ are uniformly bounded for all k, i.e., there exists some $L > 0$ such that

$$\max_{k \geq 0}\, \max\left\{\|\mathcal{L}_x(x_k, \mu_k)\|, \|\mathcal{L}_\mu(x_k, \mu_k)\|\right\} \leq L.$$

This assumption is satisfied for example when f and all g_j's are affine functions, or when the set X is compact. In the latter case, since the functions f and g_j's are convex over $\mathbb{R}^n$, they are also continuous over $\mathbb{R}^n$, and therefore the sets $\cup_{x\in X}\partial f(x)$ and $\cup_{x\in X}\partial g_j(x)$ are bounded (see Bertsekas, Nedić, and Ozdaglar [6], Proposition 4.2.3). Moreover, $\max_{x\in X}\|g(x)\|$ is finite, thus we can provide a uniform upper bound on the norm of the subgradient sequence.

5.1 Approximate Primal Solutions under Slater

In this section, we focus on constructing primal solutions by using the iterates generated by the primal-dual subgradient algorithm (5)-(6). Our goal is to generate *approximate* primal solutions and provide performance guarantees in terms of bounds on the amount of feasibility violation and primal cost values at each iteration. We generate primal solutions by averaging the vectors from the sequence $\{x_k\}$. In particular, we define $\hat{x}_k$ as the average of the vectors $x_0, \ldots, x_{k-1}$, i.e.,

$$\hat{x}_k = \frac{1}{k}\sum_{i=0}^{k-1} x_i \qquad \text{for all } k \geq 1. \tag{20}$$

The average vectors $\hat{x}_k$ lie in the set X because X is convex and $x_i \in X$ for all i. However, these vectors need not satisfy the primal inequality constraints $g_j(x) \leq 0$, $j = 0, \ldots, m$, and therefore, they can be primal infeasible.

In the next proposition, we provide per-iterate error estimates on the feasibility violation and primal cost values of the average vectors $\hat{x}_k$.

Proposition 2 Let the Slater condition and Bounded Subgradient assumption hold (cf. Assumptions 2 and 3). Let the sequences $\{x_k\}$ and $\{\mu_k\}$ be generated by the subgradient algorithm of Eqs. (16)-(17). Let $\{\hat{x}_k\}$ be the sequence of the primal averages as defined in Eq. (20), and let x^* be a primal optimal solution. Then, for all $k \geq 1$, we have:

(a) An upper bound on the amount of constraint violation of the vector $\hat{x}_k$ is given by

$$\|g(\hat{x}_k)^+\| \leq \frac{2}{k\alpha r}\left(\frac{f(\bar{x}) - \tilde{q}}{\gamma} + r\right)^2 + \frac{\|x_0 - x^*\|^2}{2k\alpha r} + \frac{\alpha L^2}{2r}.$$

(b) An upper bound on the primal cost of the vector $\hat{x}_k$ is given by

$$f(\hat{x}_k) \leq f^* + \frac{\|\mu_0\|^2}{2k\alpha} + \frac{\|x_0 - x^*\|^2}{2k\alpha} + \alpha L^2.$$

(c) A lower bound on the primal cost of the vector $\hat{x}_k$ is given by

$$f(\hat{x}_k) \geq f^* - \left(\frac{f(\bar{x}) - \tilde{q}}{\gamma}\right)\|g(\hat{x}_k)^+\|.$$

Here, the scalars $r > 0$ and $\tilde{q}$ with $\tilde{q} \leq q^*$ are those from the definition of the set D in Eq. (19), $\gamma = \min_{1\leq j\leq m}\{-g_j(\bar{x})\}$, $\bar{x}$ is the Slater vector of Assumption 2, and L is the subgradient norm bound of Assumption 3.

Proof.
(a) Using the definition of the iterate μ_{k+1} in Eq. (17) and the nonexpansive property of projection on a closed convex set, we obtain for all $\mu \in D$ and all $i \geq 0$,

$$\begin{aligned}
\|\mu_{i+1} - \mu\|^2 &= \|\mathcal{P}_D\left[\mu_i + \alpha \mathcal{L}_\mu(x_i, \mu_i)\right] - \mu\|^2 \\
&\leq \|\mu_i + \alpha \mathcal{L}_\mu(x_i, \mu_i) - \mu\|^2 \\
&\leq \|\mu_i - \mu\|^2 + 2\alpha(\mu_i - \mu)' \mathcal{L}_\mu(x_i, \mu_i) + \alpha^2 \|\mathcal{L}_\mu(x_i, \mu_i)\|^2 \\
&\leq \|\mu_i - \mu\|^2 + 2\alpha(\mu_i - \mu)' \mathcal{L}_\mu(x_i, \mu_i) + \alpha^2 L^2,
\end{aligned}$$

where the last inequality follows from the boundedness of the subgradients [cf. Assumption 3]. Therefore, for any $\mu \in D$,

$$(\mu - \mu_i)' \mathcal{L}_\mu(x_i, \mu_i) \leq \frac{\|\mu_i - \mu\|^2 - \|\mu_{i+1} - \mu\|^2}{2\alpha} + \frac{\alpha L^2}{2} \qquad \text{for all } i \geq 0. \tag{21}$$

Since $\mathcal{L}_\mu(x_i, \mu_i)$ is a subgradient of the function $\mathcal{L}(x_i, \mu)$ at $\mu = \mu_i$, using the subgradient inequality [cf. Eq. (2)], we obtain for any dual optimal solution μ^*,

$$(\mu_i - \mu^*)' \mathcal{L}_\mu(x_i, \mu_i) \leq \mathcal{L}(x_i, \mu_i) - \mathcal{L}(x_i, \mu^*) \qquad \text{for all } i \geq 0.$$

Since (x^*, μ^*) is a primal-dual optimal solution pair and $x_i \in X$, it follows from Theorem 1 that

$$\mathcal{L}(x_i, \mu^*) \geq \mathcal{L}(x^*, \mu^*) = f^*.$$

Combining the preceding two relations, we obtain

$$(\mu_i - \mu^*)' \mathcal{L}_\mu(x_i, \mu_i) \leq \mathcal{L}(x_i, \mu_i) - f^*.$$

We then have for all $\mu \in D$ and all $i \geq 0$,

$$\begin{aligned}
(\mu - \mu^*)' \mathcal{L}_\mu(x_i, \mu_i) &= (\mu - \mu^* - \mu_i + \mu_i)' \mathcal{L}_\mu(x_i, \mu_i) \\
&= (\mu - \mu_i)' \mathcal{L}_\mu(x_i, \mu_i) + (\mu_i - \mu^*) \mathcal{L}_\mu(x_i, \mu_i) \\
&\leq (\mu - \mu_i)' \mathcal{L}_\mu(x_i, \mu_i) + \mathcal{L}(x_i, \mu_i) - f^*.
\end{aligned}$$

From the preceding relation and Eq. (21), we obtain for any $\mu \in D$,

$$(\mu - \mu^*)' \mathcal{L}_\mu(x_i, \mu_i) \leq \frac{\|\mu_i - \mu\|^2 - \|\mu_{i+1} - \mu\|^2}{2\alpha} + \frac{\alpha L^2}{2} + \mathcal{L}(x_i, \mu_i) - f^* \qquad \text{for all } i \geq 0.$$

Summing over $i = 0, \ldots, k-1$ for $k \geq 1$, we obtain for any $\mu \in D$ and $k \geq 1$,

$$\begin{aligned}
\sum_{i=0}^{k-1} (\mu - \mu^*)' \mathcal{L}_\mu(x_i, \mu_i) &\leq \frac{\|\mu_0 - \mu\|^2 - \|\mu_k - \mu\|^2}{2\alpha} + \frac{\alpha k L^2}{2} + \sum_{i=0}^{k-1} \mathcal{L}(x_i, \mu_i) - k f^* \\
&\leq \frac{\|\mu_0 - \mu\|^2}{2\alpha} + \frac{\alpha k L^2}{2} + \sum_{i=0}^{k-1} \mathcal{L}(x_i, \mu_i) - k f^*.
\end{aligned}$$

Therefore, for any $k \geq 1$,

$$\max_{\mu \in D} \left\{ \sum_{i=0}^{k-1} (\mu - \mu^*)' \mathcal{L}_\mu(x_i, \mu_i) \right\} \leq \frac{1}{2\alpha} \max_{\mu \in D} \|\mu_0 - \mu\|^2 + \frac{\alpha k L^2}{2} + \sum_{i=0}^{k-1} \mathcal{L}(x_i, \mu_i) - kf^*. \quad (22)$$

We now provide a lower estimate on the left-hand side of the preceding relation. Let $k \geq 1$ be arbitrary and, for simplicity, we suppress the explicit dependence on k by letting

$$s = \sum_{i=0}^{k-1} \mathcal{L}_\mu(x_i, \mu_i). \quad (23)$$

In view of the fact $\mathcal{L}_\mu(x_i, \mu_i) = g(x_i)$, it follows that

$$s = \sum_{i=0}^{k-1} g(x_i). \quad (24)$$

By convexity of the functions g_j, it further follows that $s \geq kg(\hat{x}_k)$. Hence, if $s^+ = 0$, then the bound in part (a) of this proposition trivially holds. Therefore, assume that $s^+ \neq 0$ and define a vector $\bar{\mu}$ as follows:

$$\bar{\mu} = \mu^* + r \frac{s^+}{\|s^+\|}.$$

Note that $\bar{\mu} \geq 0$ since $\mu^* \geq 0$, $s^+ \geq 0$ and $r > 0$. By Lemma 3, the dual optimal solution set is bounded and, in particular, $\|\mu^*\| \leq \frac{f(\bar{x}) - q^*}{\gamma}$. Furthermore, since $\tilde{q} \leq q$, it follows that $\|\mu^*\| \leq \frac{f(\bar{x}) - \tilde{q}}{\gamma}$ for any dual solution μ^*. Therefore, by the definition of the vector $\bar{\mu}$, we have

$$\|\bar{\mu}\| \leq \|\mu^*\| + r \leq \frac{f(\bar{x}) - \tilde{q}}{\gamma} + r, \quad (25)$$

implying that $\bar{\mu} \in D$. Using the definition of the vector s in Eq. (23) and relation (22), we obtain

$$\begin{aligned} (\bar{\mu} - \mu^*)' s &= \sum_{i=0}^{k-1} (\bar{\mu} - \mu^*)' \mathcal{L}_\mu(x_i, \mu_i) \\ &\leq \max_{\mu \in D} \left\{ \sum_{i=0}^{k-1} (\mu - \mu^*)' \mathcal{L}_\mu(x_i, \mu_i) \right\} \\ &\leq \frac{1}{2\alpha} \max_{\mu \in D} \|\mu_0 - \mu\|^2 + \frac{\alpha k L^2}{2} + \sum_{i=0}^{k-1} \mathcal{L}(x_i, \mu_i) - kf^*. \end{aligned} \quad (26)$$

Since $\bar{\mu} - \mu^* = r \frac{s^+}{\|s^+\|}$, we have $(\bar{\mu} - \mu^*)' s = r\|s^+\|$. By using the expression for s given in Eq. (24), we obtain

$$(\bar{\mu} - \mu^*)' s = r \left\| \left[\sum_{i=0}^{k-1} g(x_i) \right]^+ \right\|.$$

Substituting the preceding equality in Eq. (26) and dividing both sides by r, we obtain

$$\Big\|\Big[\sum_{i=0}^{k-1} g(x_i)\Big]^+\Big\| \le \frac{1}{2\alpha r}\ \max_{\mu\in D}\ \|\mu_0-\mu\|^2 + \frac{\alpha k L^2}{2r} + \frac{1}{r}\left(\sum_{i=0}^{k-1}\mathcal{L}(x_i,\mu_i) - kf^*\right). \tag{27}$$

Dividing both sides of this relation by k, and using the convexity of the functions g_j in $g = (g_1, \ldots, g_m)$ together with the definition of the average primal vector $\hat{x}_k$ yields

$$\|g(\hat{x}_k)^+\| \le \frac{1}{k}\Big\|\Big[\sum_{i=0}^{k-1} g(x_i)\Big]^+\Big\| \le \frac{1}{2k\alpha r}\ \max_{\mu\in D}\ \|\mu_0-\mu\|^2 + \frac{\alpha L^2}{2r} + \frac{1}{r}\left(\frac{1}{k}\sum_{i=0}^{k-1}\mathcal{L}(x_i,\mu_i) - f^*\right). \tag{28}$$

Since $\mu_0 \in D$, it follows that

$$\max_{\mu\in D}\|\mu_0-\mu\|^2 \le \max_{\mu\in D}(\|\mu_0\|+\|\mu\|)^2 \le 4\max_{\mu\in D}\|\mu\|^2.$$

By using the definition of the set D [cf. Eq. (19)], we have

$$\max_{\mu\in D}\|\mu\| \le \frac{f(\bar{x})-\tilde{q}}{\gamma} + r.$$

Using Proposition 1(a) with the substitution $\mathcal{L}(x^*,\mu^*) = f^*$ and $M = D$, we can also provide an upper bound on the last term in Eq. (27):

$$\frac{1}{k}\sum_{i=0}^{k-1}\mathcal{L}(x_i,\mu_i) - f^* \le \frac{\|x_0-x^*\|^2}{2\alpha k} + \frac{\alpha L^2}{2}. \tag{29}$$

Substituting the preceding three estimates in the relation of Eq. (28) we obtain the desired upper bound on the amount of constraint violation of the average vector $\hat{x}_k$, i.e.,

$$\|g(\hat{x}_k)^+\| \le \frac{2}{k\alpha r}\ \left(\frac{f(\bar{x})-\tilde{q}}{\gamma} + r\right)^2 + \frac{\|x_0-x^*\|^2}{2k\alpha r} + \frac{\alpha L^2}{r}.$$

(b) By the convexity of the cost function f, we have

$$f(\hat{x}_k) \le \frac{1}{k}\sum_{i=0}^{k-1} f(x_i) = \frac{1}{k}\sum_{i=0}^{k-1}\mathcal{L}(x_i,\mu_i) - \frac{1}{k}\sum_{i=0}^{k-1}\mu_i' g(x_i).$$

Thus, it follows by Eq. (29) that

$$f(\hat{x}_k) - f^* \le \frac{1}{k}\sum_{i=0}^{k-1}\mathcal{L}(x_i,\mu_i) - f^* - \frac{1}{k}\sum_{i=0}^{k-1}\mu_i' g(x_i) \le \frac{\|x_0-x^*\|^2}{2\alpha k} + \frac{\alpha L^2}{2} - \frac{1}{k}\sum_{i=0}^{k-1}\mu_i' g(x_i). \tag{30}$$

We next provide a lower bound on $-\frac{1}{k}\sum_{i=0}^{k-1}\mu_i' g(x_i)$. Note that $0 \in D$, so that from Lemma 1(b) with $\mu = 0$, we have

$$\|\mu_{k+1}\|^2 \le \|\mu_k\|^2 + 2\alpha\mu_k' g(x_k) + \alpha^2 L^2.$$

Thus, by Eq. (7) of Lemma 2, it follows that

$$-2\alpha\mu_k' g(x_k) \le \|\mu_k\|^2 - \|\mu_{k+1}\|^2 + \alpha^2 L^2,$$

implying that

$$-\mu_k' g(x_k) \le \frac{1}{2\alpha}\left(\|\mu_k\|^2 - \|\mu_{k+1}\|^2\right) + \frac{\alpha L^2}{2}.$$

By summing these relations, we obtain

$$-\sum_{i=0}^{k-1} \mu_i' g(x_i) \le \frac{1}{2\alpha}\left(\|\mu_0\|^2 - \|\mu_k\|^2\right) + \frac{k\alpha L^2}{2}.$$

Hence,

$$-\frac{1}{k}\sum_{i=0}^{k-1} \mu_i' g(x_i) \le \frac{\|\mu_0\|^2}{2\alpha k} + \frac{\alpha L^2}{2}.$$

The estimate follows by substituting the preceding relation in Eq. (30).
(c) For any dual optimal solution μ^*, we have

$$f(\hat{x}_k) = f(\hat{x}_k) + (\mu^*)' g(\hat{x}_k) - (\mu^*)' g(\hat{x}_k) = \mathcal{L}(\hat{x}_k, \mu^*) - (\mu^*)' g(\hat{x}_k).$$

By the Saddle-Point Theorem [cf. Theorem 1], it follows that

$$\mathcal{L}(\hat{x}_k, \mu^*) \ge \mathcal{L}(x^*, \mu^*) = f^*.$$

Hence,

$$f(\hat{x}_k) \ge f^* - (\mu^*)' g(\hat{x}_k). \tag{31}$$

Furthermore, since $\mu^* \ge 0$ and $g(\hat{x}_k) \le g^+(\hat{x}_k)$, it follows that

$$-(\mu^*)' g(\hat{x}_k) \ge -(\mu^*)' g^+(\hat{x}_k) \ge -\|\mu^*\| \|g^+(\hat{x}_k)\|.$$

Since $\|\mu^*\| \le \frac{f(\bar{x}) - q^*}{\gamma}$ and $q^* \ge \tilde{q}$, it follows that $\|\mu^*\| \le \frac{f(\bar{x}) - \tilde{q}}{\gamma}$, which when substituted in Eq. (31) yields the desired estimate. ■

We note here that the primal-dual method of Eqs. (5)-(6) with the set D as given in Eq. (19) couples the computation of multipliers through the projection operation. In some applications, it might be desirable to accommodate distributed computation models whereby the multiplier components μ^j are processed in a distributed manner among a set of processors or agents. To accommodate such computations, one may modify the subgradient method of Eqs. (16)-(17) by replacing the set D of Eq. (19) with the following set

$$D_\infty = \left\{\mu \ge 0 \;\middle|\; \|\mu\|_\infty \le \frac{f(\bar{x}) - \tilde{q}}{\gamma} + r\right\}.$$

It can be seen that the results of Proposition 2 also hold for this choice of the projection set. In particular, this can be seen by following the same line of argument as in the proof of Proposition 2 [by using the fact $\|y\|_\infty \le \|y\|$ in Eq. (25)].

5.2 Optimal Choice for the set D

The bound provided in part (a) of Proposition 2 is a function of the paramater r, which is used in the definition of the set D [cf. (19)]. We next consider selecting the parameter r such that the right-hand side of the bound in part (a) of Proposition 2 is minimized at each iteration k. Given some $k \geq 1$, we choose r as the optimal solution of the problem

$$\min_{r>0} \left\{ \frac{2}{k\alpha r} \left(\frac{f(\bar{x}) - \tilde{q}}{\gamma} + r \right)^2 + \frac{\|x_0 - x^*\|^2}{2k\alpha r} + \frac{\alpha L^2}{2r} \right\}.$$

It can be seen that the optimal solution of the preceding problem, denoted by $r^*(k)$, is given by

$$r^*(k) = \sqrt{\left(\frac{f(\bar{x}) - \tilde{q}}{\gamma} \right)^2 + \frac{\|x_0 - x^*\|^2}{4} + \frac{k\alpha^2 L^2}{4}} \qquad \text{for } k \geq 1. \tag{32}$$

Consider now an algorithm where the dual iterates are obtained by

$$\mu_{i+1} = \mathcal{P}_{D_k} \left[\mu_i + \alpha \mathcal{L}_\mu(x_i, \mu_i) \right] \qquad \text{for each } i \geq 0,$$

with $\mu_0 \in D_0$ and the set D_k given by

$$D_k = \left\{ \mu \geq 0 \mid \|\mu\| \leq \frac{f(\bar{x}) - \tilde{q}}{\gamma} + r^*(k) \right\},$$

where $r^*(k)$ is given by Eq. (32). Hence, at each iteration i, the algorithm projects onto the set D_k, which contains the set of dual optimal solutions M^*.

Substituting $r^*(k)$ in the bound of Proposition 2(a), we can see that

$$\begin{aligned} \|g(\hat{x}_k)^+\| &\leq \frac{4}{k\alpha} \left(\frac{f(\bar{x}) - \tilde{q}}{\gamma} + \sqrt{\left(\frac{f(\bar{x}) - \tilde{q}}{\gamma} \right)^2 + \frac{\|x_0 - x^*\|^2}{4} + \frac{k\alpha^2 L^2}{4}} \right) \\ &\leq \frac{4}{k\alpha} \left[\frac{2(f(\bar{x}) - \tilde{q})}{\gamma} + \frac{\|x_0 - x^*\|}{2} + \frac{\alpha L \sqrt{k}}{2} \right] \\ &= \frac{8}{k\alpha} \left(\frac{f(\bar{x}) - \tilde{q}}{\gamma} \right) + \frac{2\|x_0 - x^*\|}{k\alpha} + \frac{2L}{\sqrt{k}}. \end{aligned}$$

The preceding discussion combined with Proposition 2(a) immediately yields the following result:

Proposition 3 Let the Slater condition and Bounded Subgradient assumption hold [cf. Assumptions 2 and 3]. Fix a positive integer $k \geq 1$, and define the set D_k as

$$D_k = \left\{ \mu \geq 0 \;\middle|\; \|\mu\|_2 \leq \frac{f(\bar{x}) - \tilde{q}}{\gamma} + \sqrt{\left(\frac{f(\bar{x}) - \tilde{q}}{\gamma} \right)^2 + \frac{\|x_0 - x^*\|^2}{4} + \frac{k\alpha^2 L^2}{4}} \right\}, \tag{33}$$

where $\bar{x}$ is the Slater vector of Assumption 2, $\gamma = \min_{1 \leq j \leq m}\{-g_j(\bar{x})\}$, and L is the bound on the subgradient norm of Assumption 3. Let the primal-dual sequence $\{x_i, \mu_i\}$

be generated by the following modified subgradient method: Let $x_0 \in X$ and $\mu_0 \in D_k$. For each $i \geq 0$, the iterates x_i and μ_i are obtained by

$$x_{i+1} = \mathcal{P}_X \left[x_i - \alpha \mathcal{L}_x(x_i, \mu_i) \right],$$

$$\mu_{i+1} = \mathcal{P}_{D_k} \left[\mu_i + \alpha \mathcal{L}_\mu(x_i, \mu_i) \right],$$

i.e., at each iteration, the dual solutions are projected onto the set D_k. Then, an upper bound on the amount of feasibility violation of the vector $\hat{x}_k$ is given by

$$\|g(\hat{x}_k)^+\| \leq \frac{8}{k\alpha} \left(\frac{f(\bar{x}) - \tilde{q}}{\gamma} \right) + \frac{2\|x_0 - x^*\|}{k\alpha} + \frac{2L}{\sqrt{k}}. \tag{34}$$

This result shows that for a given k, the error estimate provided in Eq. (34) can be achieved if we use a primal-dual subgradient method where each dual iterate is projected on the set D_k defined in Eq. (33). Given a pre-specified accuracy for the amount of feasibility violation, this bound can be used to select the stepsize value and the set D_k for the algorithm.

6 Conclusions

We presented a subgradient method for generating approximate saddle points for a convex-concave function $\mathcal{L}(x, \mu)$ defined over $X \times M$. This algorithm takes steps along directions defined by the subgradients of the function $\mathcal{L}(x, \mu)$ with respect to x and μ, and uses an averaging scheme to produce approximate saddle points. We showed that under the assumption that the iterates generated by the algorithm are bounded (which will hold when the sets X and M are compact), the function values $\mathcal{L}(\hat{x}_k, \hat{\mu}_k)$, at the iterate averages $\hat{x}_k$ and $\hat{\mu}_k$, converge to the function value at a saddle point with rate $1/k$ and with an error level which is a function of the stepsize value.

We then focused on Lagrangian duality, where we consider a convex primal optimization problem and its Lagrangian dual. We proposed using our subgradient algorithm for generating saddle points of the Lagrangian function of the primal problem, which provide approximate primal-dual optimal solutions. Under Slater condition, we studied a variation of our algorithm which allowed us to generate convergence rate estimates without imposing boundedness assumptions on the generated iterates.

An important application of our work is in resource allocation problems for networked-systems. In this context, Lagrangian duality and subgradient methods for solving the dual problem have been used with great success in developing decentralized network algorithms (see Low and Lapsley [18], Srikant [30], and Chiang *et al.* [7]). This approach relies on the assumption that the subgradient of the dual function can be evaluated efficiently. This implicitly assumes that the Lagrangian function is decomposable, which will hold when the primal problem has a separable structure. Our results in this paper can be used to generalize this approach to more general non-separable resource allocation problems.

References

[1] K.J. Arrow, L. Hurwicz, and H. Uzawa, *Studies in linear and non-linear programming*, Stanford University Press, Stanford, CA, 1958.

[2] A. Auslender and M. Teboulle, *Interior gradient and proximal methods for convex and conic optimization*, SIAM Journal on Optimization **16** (2006), 697–725.

[3] ______, *Projected subgradient methods with non-Euclidean distances for nondifferentiable convex minimization and variational inequalities*, Mathematical Programming (2007).

[4] A. Beck and M. Teboulle, *Mirror descent and nonlinear projected subgradient methods for convex optimization*, Operations Research Letters **31** (2003), 167–175.

[5] D.P. Bertsekas, *Nonlinear programming*, Athena Scientific, Cambridge, Massachusetts, 1999.

[6] D.P. Bertsekas, A. Nedić, and A.E. Ozdaglar, *Convex analysis and optimization*, Athena Scientific, Cambridge, Massachusetts, 2003.

[7] M. Chiang, S.H. Low, A.R. Calderbank, and J.C. Doyle, *Layering as optimization decomposition: A mathematical theory of network architectures*, Proceedings of the IEEE **95** (2007), no. 1, 255–312.

[8] R. Correa and C. Lemaréchal, *Convergence of some algorithms for convex minimization*, Mathematical Programming **62** (1993), 261–271.

[9] V.F. Demyanov and L.V. Vasilyev, *Nondifferentiable optimization*, Optimization Software, Inc., New York, 1985.

[10] E.G. Gol'shtein, *A generalized gradient method for finding saddle points*, Matekon **10** (1974), 36–52.

[11] J.-B. Hiriart-Urruty and C. Lemaréchal, *Convex analysis and minimization algorithms*, Springer-Verlag, Berlin, 1996.

[12] M. Kallio and C.H. Rosa, *Large-scale convex optimization via saddle-point computation*, Operations Research **47** (1999), 93–101.

[13] M. Kallio and A. Ruszczyński, *Perturbation methods for saddle point computation*, Tech. Report WP-94-38, International Institute for Applied Systems Analysis, 1994.

[14] G.M. Korpelevich, *The extragradient method for finding saddle points and other problems*, Matekon **13** (1977), 35–49.

[15] T. Larsson, M. Patriksson, and A. Strömberg, *Ergodic results and bounds on the optimal value in subgradient optimization*, Operations Research Proceedings (P. Kelinschmidt et al., ed.), Springer, 1995, pp. 30–35.

[16] T. Larsson, M. Patriksson, and A. Strömberg, *Ergodic convergence in subgradient optimization*, Optimization Methods and Software **9** (1998), 93–120.

[17] ______, *Ergodic primal convergence in dual subgradient schemes for convex programming*, Mathematical Programming **86** (1999), 283–312.

[18] S. Low and D.E. Lapsley, *Optimization flow control, I: Basic algorithm and convergence*, IEEE/ACM Transactions on Networking **7** (1999), no. 6, 861–874.

[19] D. Maistroskii, *Gradient methods for finding saddle points*, Matekon **13** (1977), 3–22.

[20] A. Nedić and D.P. Bertsekas, *Convergence rate of incremental subgradient algorithms*, Stochastic Optimization: Algorithms and Applications (S. Uryasev and P. Pardalos, eds.), Kluwer Academic Publishers, 2000.

[21] ______, *Incremental subgradient methods for nondifferentiable optimization*, SIAM Journal on Optimization **56** (2001), no. 1, 109–138.

[22] A. Nedić and A. Ozdaglar, *Approximate primal solutions and rate analysis for dual subgradient methods*, Tech. Report 2753, LIDS, 2007.

[23] 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 Journal on Optimization **15** (2004), no. 1, 229–251.

[24] A.S. Nemirovski and D.B. Judin, *Cezare convergence of gradient method approximation of saddle points for convex-concave functions*, Doklady Akademii Nauk SSSR **239** (1978), 1056–1059.

[25] Y. Nesterov, *Primal-dual subgradient methods for convex problems*, Tech. Report 67, Center for Operations Research and Econometrics (CORE), Catholic University of Louvain (UCL), 2005.

[26] B.T. Polyak, *Introduction to optimisation*, Optimization Software, Inc., New York, 1987.

[27] P.A. Samuelson, *Market mechanisms and maximizations*, Tech. Report P-69, RAND Corporation, 1949.

[28] H.D. Sherali and G. Choi, *Recovery of primal solutions when using subgradient optimization methods to solve Lagrangian duals of linear programs*, Operations Research Letters **19** (1996), 105–113.

[29] N.Z. Shor, *Minimization methods for nondifferentiable functions*, Translated from Russian by K.C. Kiwiel and A. Ruszczynski, Springer, Berlin, 1985.

[30] R. Srikant, *Mathematics of Internet congestion control*, Birkhauser, 2004.

[31] H. Uzawa, *Iterative methods in concave programming*, Studies in Linear and Nonlinear Programming (K. Arrow, L. Hurwicz, and H. Uzawa, eds.), Stanford University Press, 1958, pp. 154–165.

[32] I.Y. Zabotin, *A subgradient method for finding a saddle point of a convex-concave function*, Issledovania po prikladnoi matematike **15** (1988), 6–12.