Mathematical Programming
https://doi.org/10.1007/s10107-023-02051-2

FULL LENGTH PAPER

Series A

Semiglobal exponential stability of the discrete-time Arrow-Hurwicz-Uzawa primal-dual algorithm for constrained optimization

Michelangelo Bin[1,2] · **Ivano Notarnicola**[1] · **Thomas Parisini**[2,3,4]

Received: 10 January 2023 / Accepted: 16 December 2023

Abstract

We consider the discrete-time Arrow-Hurwicz-Uzawa primal-dual algorithm, also known as the first-order Lagrangian method, for constrained optimization problems involving a smooth strongly convex cost and smooth convex constraints. We prove that, for every given compact set of initial conditions, there always exists a sufficiently small stepsize guaranteeing exponential stability of the optimal primal-dual solution of the problem with a domain of attraction including the initialization set. Inspired by the analysis of nonlinear oscillators, the stability proof is based on a non-quadratic Lyapunov function including a nonlinear cross term.

Keywords Constraint programming · Convex programming · Primal-dual algorithm · Discrete-time control/observation systems

Mathematics Subject Classification 68Q25 · 37N40 · 90C25 · 93C55

T.P. and M.B. acknowledge funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no. 739551 (KIOS CoE) and by the Italian Ministry for Research in the framework of the 2017 Program for Research Projects of National Interest (PRIN), Grant no. 2017YKXYXJ.

✉ Michelangelo Bin
michelangelo.bin@unibo.it

Ivano Notarnicola
ivano.notarnicola@unibo.it

Thomas Parisini
t.parisini@imperial.ac.uk

1 Department of Electrical, Electronic, and Information Engineering, University of Bologna, Bologna, Italy

2 Department of Electrical and Electronic Engineering, Imperial College London, London, UK

3 KIOS Research and Innovation Center of Excellence, University of Cyprus, Nicosia, Cyprus

4 Department of Engineering and Architecture, University of Trieste, Trieste, Italy

Published online: 12 February 2024

1 Introduction

1.1 Problem overview, contribution, and literature review

In this article, we study the convergence and Lyapunov stability properties of the following discrete-time first-order primal-dual algorithm

$$x^{t+1} = x^t - \gamma \left(\nabla f(x^t) + \sum_{i=1}^{r} \lambda_i^t \nabla g_i(x^t) \right), \qquad x^0 \in \mathbb{R}^n, \tag{1a}$$

$$\lambda_i^{t+1} = \max\left\{0,\ \lambda_i^t + \gamma g_i(x^t)\right\}, \qquad \lambda_i^0 \geq 0, \qquad i = 1, \ldots, r, \tag{1b}$$

in which $t \in \mathbb{N}$ is the iteration counter, $n, r \in \mathbb{N}$ are arbitrary, $x^t \in \mathbb{R}^n$ is the primal variable, $\lambda_i^t \in \mathbb{R}_{\geq 0}$ are the dual variables, $f : \mathbb{R}^n \to \mathbb{R}$ and $g = (g_1, \ldots, g_r) : \mathbb{R}^n \to \mathbb{R}^r$ are convex functions, and $\gamma > 0$ is a parameter called the *stepsize*. Algorithm (1) gives an iterative procedure to compute a solution of the constrained optimization problem

$$\begin{aligned} &\min_{x \in \mathbb{R}^n} \ f(x) \\ &\text{subj. to } \ g_i(x) \leq 0, \quad i = 1, \ldots, r, \end{aligned} \tag{2}$$

and it is a slight variation of Uzawa's original method [29]. In particular, when $\lambda_i^t + \gamma g_i(x^t) \geq 0$ for all $i = 1, \ldots, r$, Eqs. (1) take the form of a discrete-time version of the Arrow-Hurwicz saddle-point dynamics [1] (see also [18]) applied to the Lagrangian function of (2), which reads as

$$L(x, \lambda) := f(x) + \sum_{i=1}^{r} \lambda_i g_i(x).$$

Indeed, (1) can be rewritten in compact form as

$$\begin{aligned} x^{t+1} &= x^t - \gamma \frac{\partial L}{\partial x}(x^t, \lambda^t), & x^0 &\in \mathbb{R}^n, \\ \lambda^{t+1} &= \max\left\{0,\ \lambda^t + \gamma \frac{\partial L}{\partial \lambda}(x^t, \lambda^t)\right\}, & \lambda^0 &\in (\mathbb{R}_{\geq 0})^r, \end{aligned}$$

in which the max operator acts component-wise. Hence, (1) is a first-order Lagrangian method.

In his original paper [29], Uzawa provided a proof of non-local stability and convergence of (1). However, his arguments were later found wrong (see, e.g., [25, Sec. 1]). Other existing proofs, which can be found, for instance, in [25] and the classical textbook [2], only provide *local* convergence guarantees to a saddle point of the Lagrangian function L. These results are based on the linear approximation of the algorithm around the optimal point (see, e.g., [2, Sec. 4.4]) and, hence, can only guarantee convergence from a (possibly very small) neighborhood of the optimum, whose size is not guaranteed to increase as the stepsize γ decreases. Nonlocal convergence results have been obtained in [15, 16] at the cost, however, of adopting a diminishing stepsize. An extension of the latter results to a stochastic setting is studied in [30] in the context of distributed network utility maximization. Other nonlocal, yet approximate, convergence bounds have been given in [23] under the assumption of gradient boundedness. Finally, it is worth noticing that versions of (1) tailored for quadratic programs have been widely studied in the context of imaging; see, e.g., [3]. Despite the numerous results and the long history of Algorithm (1), to the best of the authors' knowledge, a purely discrete-time analysis providing nonlocal convergence and stability guarantees is still missing.

Interestingly, guarantees of such kind do exist for the continuous-time version of Algorithm (1), which is also known as *saddle-point* or *saddle-flow dynamics*. See, for instance, references [4–7, 9, 11, 12, 26] and extensions covering the augmented Lagrangian version [28] of (1) and its proximal regularization [10]. In particular, in continuous time one can achieve global [5–7, 9, 11] and even exponential [4, 26] convergence in some cases, resulting in a sharp distinction between the continuous- and discrete-time domains. However, the results attained for continuous-time algorithms are typically not preserved under (Euler) discretization and, therefore, cannot be used to assess equivalent properties on their (first-order) discretization. In particular, the continuous-time algorithms cited above, for which global and/or exponential stability guarantees do exist, typically consist in differential equations defined by a vector field that is discontinuous at some relevant points. Yet, continuity is generally required to apply the basic discretization theorems (see, e.g., [27, Sects. 2.1.1 and 2.3]). Among the continuous-time algorithms employing continuous vector fields and ensuring global convergence it is worth mentioning [13, Eq. (5)] (see also [8]). However, a simple counterexample, similar to that reported later in Sect. 1.2, can be used to show that the Euler discretization of such an algorithm cannot be globally convergent. Hence, also in these cases, the continuous-time results are not directly extendable to cover the algorithms' discretization.

In view of the above discussion, to the best of the authors' knowledge, a nonlocal stability and convergence proof for the discrete-time algorithm (1) is still an open, long-standing problem, even under strong convexity of f and convexity of g (which we shall assume later on). In this article, we aim to fill this gap by providing a purely discrete-time *semiglobal* asymptotic stability analysis for (1). As shown in Sect. 1.2, global convergence is generically not possible for Algorithm (1); hence, a semiglobal result is the best one can achieve in the general case. Specific contributions are highlighted in the next section.

1.2 Contributions

Under widely adopted regularity and convexity assumptions on f and g detailed later in Sect. 2.1, we prove that the minimizer of (2) (which is unique under the assumptions of the article) is *semiglobally*[1] Lyapunov stable and exponentially attractive for Algorithm (1). More specifically, we show that there exists an equilibrium $(x^\star, \lambda^\star) \in \mathbb{R}^n \times \mathbb{R}^r$ for (1), with $x^\star$ being the optimal solution of (2) and $\lambda^\star$ the corresponding optimal Lagrange multiplier, and that, for every arbitrary compact set $\Xi_0 \subseteq \mathbb{R}^n \times \mathbb{R}^r$ of initial conditions for (1), there exists $\gamma^\star > 0$, such that, for all $\gamma \in (0, \gamma^\star)$, the following properties hold:

1. *Lyapunov stability:* for every $\varepsilon > 0$, there exists $\delta > 0$, such that $|x^0 - x^\star| < \delta$ and $|\lambda^0 - \lambda^\star| < \delta$ imply $|x^t - x^\star| < \varepsilon$ and $|\lambda^t - \lambda^\star| < \varepsilon$ for all $t \in \mathbb{N}$.
2. *Attractiveness:* every solution [2] of (1) with $(x^0, \lambda^0) \in \Xi_0$ satisfies $\lim_{t\to\infty} |x^t - x^\star| = 0$ and $\lim_{t\to\infty} |\lambda^t - \lambda^\star| = 0$.
3. *Exponential (or linear) convergence:* there exist $\sigma > 0$ and $\mu \in (0, 1)$ (depending on γ and Ξ_0) such that every solution of (1) with $(x^0, \lambda^0) \in \Xi_0$ satisfies

$$\forall t \in \mathbb{N}, \quad |(x^t, \lambda^t) - (x^\star, \lambda^\star)| \leq \sigma \mu^t |(x^0, \lambda^0) - (x^\star, \lambda^\star)|.$$

The previously-defined stability notion, known as *Lyapunov stability* [19], [17, Ch. 4], is a continuity property of the algorithm's trajectories with respect to variations of the initial conditions. It guarantees that small deviations of the initial conditions from the optimal point $(x^\star, \lambda^\star)$ do not lead to large deviations from it along the algorithm's trajectories. We underline that Lyapunov stability, attractiveness, and exponential convergence are guaranteed from an arbitrarily large compact initialization set Ξ_0, provided that γ is chosen sufficiently small. This semiglobal result is strictly stronger than its *local* counterpart that, indeed, would only guarantee the existence of a (possibly very small) neighborhood Ξ_0 of $(x^\star, \lambda^\star)$ from which the previous properties hold. However, it is also weaker than a *global* result, for which a single γ would work for all possible initialization sets Ξ_0. Nevertheless, we observe that the lack of global convergence is not a shortcoming of our analysis; indeed, global convergence is, in general, not possible for (1). This can be seen by means of a simple counterexample. Take $n = r = 1$, $f(x) = x^2$ and $g(x) = x^2 - 1$. Fix $\gamma > 0$ arbitrarily. Then, every solution with initial conditions satisfying

$$x^0 \geq 2, \qquad \lambda^0 \geq \frac{1+\sqrt{2}}{2\gamma} \tag{3}$$

diverges. In fact from (1), one obtains that, for all $t \in \mathbb{N}$,

$$\begin{cases} x^t \geq 2, \\ \lambda^t \geq \dfrac{1+\sqrt{2}}{2\gamma} \end{cases} \implies \begin{cases} |x^{t+1}|^2 = \left(1 - 2\gamma - 2\gamma\lambda^t\right)^2 |x^t|^2 \geq 2|x^t|^2 \geq 2, \\ \lambda^{t+1} = \lambda^t + \gamma\left(|x^t|^2 - 1\right) \geq \lambda^t \geq \dfrac{1+\sqrt{2}}{2\gamma}. \end{cases}$$

[1] We borrow the terminology from control theory (see, e.g., [14, Ch. 9]) and we say that a property P on the solutions of (1) holds semiglobally if, for every arbitrary compact subset $\Xi_0 \subseteq \mathbb{R}^n \times \mathbb{R}^r$ of initial conditions for (1), there exists $\bar{\gamma} > 0$, such that, for all $\gamma \in (0, \bar{\gamma})$, P holds for all the trajectories generated by (1) that originate in Ξ_0.

[2] Here and throughout the article, a *solution* of (5) is meant as any function $t \mapsto (x(t), \lambda(t))$ solving such recursive equations.

By induction, one thus obtains that (3) implies $x^t \geq 2$ and $\lambda^t \geq \frac{1+\sqrt{2}}{2\gamma}$ for all $t \in \mathbb{N}$ and, moreover, that $|x^{t+1}|^2 \geq 2|x^t|^2$ holds for all $t \in \mathbb{N}$. Hence, the trajectory x diverges exponentially.

1.3 A systems-theoretic approach

Local stability and convergence results based on the linear approximation of the algorithm's equations cannot be easily extended to nonlocal results where the nonlinear terms dominate. Instead, the analysis approach pursued in this article is based on the theory of Lyapunov functions [17, Chapter 4], which is better suited to handle purely nonlinear problems like the one considered in the paper. Finding a suitable Lyapunov function is in general difficult, and a counterexample can be used to show that the simple choice $(x - x^\star)^2 + (\lambda - \lambda^\star)^2$, used by Uzawa in the aforementioned article [29], would not work. In this direction, it helps to look at (1) from a different perspective. Namely, by ignoring the "max" in the equation of λ, we can look at (1) as the Euler discretization (with sampling time γ) of the following continuous-time system (consider $r = 1$ for simplicity)

$$\begin{aligned} \dot{x} &= -\nabla g(x)\lambda - \nabla f(x), \\ \dot{\lambda} &= g(x). \end{aligned}$$

This is the equation of a nonlinear oscillator with $\nabla g(x)$ playing the role of the natural frequency, and $-\nabla f(x)$ that of a nonlinear damping term. It is well-known [17, Example 4.4], that Lyapunov functions for nonlinear oscillators must have a cross-term. This is what ultimately motivated the specific choice for the Lyapunov function used in this article, formally defined in (25). In turn, the introduction of a suitable cross-term, which can be seen as a modification of Uzawa's Lyapunov candidate function, turned out to be key for proving stability and convergence.

1.4 Organization and notation

Organization. In Sect. 2, we detail the basic assumptions and link the equilibria of (1) to the optimal solution of (2). In Sect. 3, we state the main result of the paper proving semiglobal exponential stability of the optimal equilibrium. Finally, the proof of the main result is presented in Sect. 4.

Notation. Set inclusion (either strict or not) is denoted by $\subseteq$. If S is a set and $\sim$ a binary relation on it, for $s \in S$ we let $S_{\sim s} := \{z \in S : z \sim s\}$. The closed ball of radius r centered at $\bar{x} \in \mathbb{R}^n$ is denoted by $\overline{\mathbb{B}}_r(\bar{x}) := \{x \in \mathbb{R}^n : |x - \bar{x}| \leq r\}$. We identify linear operators $\mathbb{R}^m \to \mathbb{R}^n$ with their matrix representation with respect to the standard bases of $\mathbb{R}^m$ and $\mathbb{R}^n$. If $A, B \in \mathbb{R}^{n\times n}$, $A \geq B$ means that $A - B$ is positive semidefinite. Given a scalar function $f : \mathbb{R}^n \to \mathbb{R}$, we define its gradient as $\nabla f(\cdot) = (\partial f(\cdot)/\partial x_1, \dots, \partial f(\cdot)/\partial x_n) \in \mathbb{R}^n$. Given a vector field $g : \mathbb{R}^n \to \mathbb{R}^r$, we let $\nabla g(\cdot) := \left[\nabla g_1(\cdot) \cdots \nabla g_r(\cdot)\right] \in \mathbb{R}^{n\times r}$. We equip $\mathbb{R}^n$ with the standard inner product $\langle x \mid y \rangle := \sum_{i=1}^n x_i y_i$, and we denote the induced Euclidean norm by

$|x|:=\sqrt{\langle x \mid x \rangle}$. If $x \in \mathbb{R}^n$ and $\sim$ is a binary relation on $\mathbb{R}$, $x \sim 0$ means $x_i \sim 0$ for all $i = 1, \dots, n$. Similarly, $\max\{0, x\}:=(\max\{0, x_1\}, \dots, \max\{0, x_n\})$ and, for $y \in \mathbb{R}^n$, $\max\{y, x\}:=(\max\{y_1, x_1\}, \dots, \max\{y_n, x_n\})$. For notation convenience, we write (x, y) in place of $((x, y))$. For instance, $\overline{\mathbb{B}}_r(1, 2)$ is the closed ball of radius r in $\mathbb{R}^2$ centered at $(1, 2) \in \mathbb{R}^2$. We denote by $(\cdot)^+$ the "shift" operator such that, for a discrete-time signal $x : \mathbb{N} \to \mathbb{R}^n$, $x^+(t) = x(t + 1)$. For brevity, in dealing with time signals, we use the notation x^t in place of $x(t)$.

2 The framework

2.1 Standing assumptions and optimality conditions

We consider Algorithm (1) and Problem (2) under the following assumptions.

Assumption 1 The functions f and g_i satisfy the following properties:

A. f is strongly convex and twice continuously differentiable;
B. for all $i = 1, \dots, r$, g_i is convex and twice continuously differentiable;
C. there exists $\bar{x} \in \mathbb{R}^n$ such that $g_i(\bar{x}) \le 0$ for all $i = 1, \dots, r$.

The conditions asked by Assumption 1 are widely adopted [2]. In particular, they imply that the optimization problem (2) has a unique solution, as established by the lemma below.

Lemma 1 *Suppose that Assumption 1 holds. Then, there exists a unique $x^\star \in \mathbb{R}^n$ solving (2).*

The proof of Lemma 1 is given in the Appendix. Throughout the article, we denote by $x^\star$ the unique optimal solution of (2). Moreover, we let

$$A(x^\star):=\{i \in \{1, \dots, r\} \ : \ g_i(x^\star) = 0\}$$

denote the set of indices of the *active* constraints at $x^\star$. Then, in addition to Assumption 1, we assume that $x^\star$ is a *regular* point in the following sense.

Assumption 2 The vectors $\{\nabla g_i(x^\star) \ : \ i \in A(x^\star)\}$ are linearly independent.

Like Assumption 1, also Assumption 2 is customary [2]. In particular, Lemma 1 (hence, Assumption 1) and Assumption 2 imply that there necessarily exists a unique $\lambda^\star \in \mathbb{R}^r$ such that the so-called *KKT conditions* hold (see, e.g., [2, Prop. 3.3.1])

$$\nabla f(x^\star) + \nabla g(x^\star)\lambda^\star = 0, \tag{4a}$$
$$g_i(x^\star) \le 0, \quad \lambda_i^\star \ge 0, \quad \lambda_i^\star g_i(x^\star) = 0, \quad \forall i = 1, \dots, r. \tag{4b}$$

Notice that Conditions (4) are also sufficient. Namely, if some $(x^\star, \lambda^\star)$ satisfies (4), then $x = x^\star$ is the optimal solution of (2) (see, e.g., [2, Prop. 3.3.4]).

2.2 Optimality and equilibria

Algorithm (1) can be rewritten in compact form as[3]

$$x^{+} = x - \gamma \nabla L(x, \lambda), \qquad x^{0} \in \mathbb{R}^{n}, \tag{5a}$$

$$\lambda^{+} = \max\{0,\ \lambda + \gamma g(x)\}, \qquad \lambda^{0} \geq 0, \tag{5b}$$

where $L(x, \lambda) := f(x) + \langle \lambda \mid g(x) \rangle$ denotes the Lagrangian function associated with (2). We remark that the initialization $\lambda^0 \geq 0$ is only assumed to simplify the analysis and it is not necessary. Indeed, (5b) trivially implies $\lambda^t \geq 0$ for all $t \geq 1$ even if $\lambda^0 < 0$.

The following lemma characterizes the equilibria of (5) in terms of the optimality conditions (4).

Lemma 2 $(x, \lambda) \in \mathbb{R}^n \times (\mathbb{R}_{\geq 0})^r$ *is an equilibrium of (5) if and only if it satisfies (4).*

Proof The proof simply follows by noticing that $x^+ = x$ if and only if $\nabla L(x, \lambda) = 0$, which is (4a), and $\lambda^+ = \lambda$ if and only if $0 = \max\{-\lambda, \gamma g(x)\}$, which is equivalent to (4b) since $\lambda \geq 0$. □

The discussion of Sect. 2.1 and Lemma 2 ultimately imply that (5) has a unique equilibrium $(x^\star, \lambda^\star) \in \mathbb{R}^n \times (\mathbb{R}_{\geq 0})^r$ satisfying (4) and such that $x^\star$ solves (2). In the remainder of the article, we study the stability and exponential attractiveness properties of such an equilibrium.

3 Main result

In this section, we state and discuss the main result of the article establishing semiglobal exponential stability of the optimal equilibrium $(x^\star, \lambda^\star)$ for Algorithm (5).

Theorem 1 *Suppose that Assumptions 1 and 2 hold. Then, for every compact subset $\Xi_0 \subseteq \mathbb{R}^n \times (\mathbb{R}_{\geq 0})^r$ of initial conditions for (5), there exists $\bar{\gamma} > 0$, and for every $\gamma \in (0, \bar{\gamma})$, there exist $\mu = \mu(\gamma) \in (0, 1)$ and $\sigma = \sigma(\gamma) > 0$, such that every solution (x, λ) of (5) with $(x^0, \lambda^0) \in \Xi_0$ satisfies*

$$\forall t \in \mathbb{N}, \qquad |(x^t - x^\star, \lambda^t - \lambda^\star)| \leq \sigma \mu^t |(x^0 - x^\star, \lambda^0 - \lambda^\star)|. \tag{6}$$

The proof of Theorem 1 is presented in Sect. 4. Clearly, (6) implies that the optimal equilibrium $(x^\star, \lambda^\star)$ is Lyapunov stable for (5) and semiglobally exponentially attractive with the convergence rate μ and the constant σ depending on γ. As shown in the proof of the theorem (see, in particular, Sect. 4.7) for a fixed $\gamma > 0$, the constants μ and σ are estimated as

$$\mu := \sqrt{1 - \frac{1}{6}\min\{2,\ \gamma c_0,\ \gamma^2 k_2^2\}}, \qquad \sigma := \sqrt{3}\mu^{-T},$$

[3] With a slight abuse of notation, for ease of presentation we denote by $\nabla L(x, \lambda) := \partial L(x, \lambda)/\partial x$ the gradient with respect to the x variable only.

in which

$$T:=\frac{6(3K_0^2-\min\{\gamma^2h^2,\varepsilon^2\})}{\min\left\{2,\ \gamma c_0,\ \gamma^2k_2^2\right\}\min\left\{\gamma^2h^2,\varepsilon^2\right\}},$$

for suitable positive constants $c_0, K_0, h, k_2, \varepsilon$ defined in the proof of Theorem 1 (see Sects. 4.1 and 4.2). In particular, c_0 is the convexity parameter of f such that (8) holds, $K_0>0$ is any scalar such that $\Xi_0 \subseteq \overline{\mathbb{B}}_{K_0}(x^\star,\lambda^\star)$ (see (7)), $h:=\min_{i\notin A(x^\star)}|g_i(x^\star)|$, $k_2>0$ is the Lipschitz constant of $(x,\lambda)\mapsto\nabla L(x,\lambda)$ on a suitably-defined compact superset of $\overline{\mathbb{B}}_{K_0}(x^\star,\lambda^\star)$ (see (9)), and $\varepsilon>0$ is a possibly "small" scalar fixed in (18) so that, for all $x\in\mathbb{R}^n$ satisfying $|x-x^\star|\le\varepsilon$, $g_i(x)<0$ for all $i\notin A(x^\star)$, and $\nabla g_{\mathrm{A}}(x)^\top\nabla g_{\mathrm{A}}(x)$ is uniformly positive definite.

The above estimates of μ and σ highlight the worst-case dependency of the convergence properties of the algorithm from the stepsize (γ), the convexity properties of the cost function (c_0), the "size" of the domain of attraction (K_0), the smoothness of the cost function and the constraint functions (k_2), and the "regularity" (or independence) of the active constraints (ε). In this respect, we observe that (6) only gives a worst-case estimate of the error decrease, and it does not characterize exactly the actual algorithm's convergence rate.

4 Proof of Theorem 1

In this section, we prove Theorem 1. We organize the proof in seven subsections. In Sects. 4.1 and 4.2, we first present some preliminary definitions and technical lemmas. In Sect. 4.3, we construct a Lyapunov candidate function that, unlike the one used by Uzawa in [29], includes a cross-term proportional to $\langle x-x^\star\,|\,\nabla g(x)(\lambda-\lambda^\star)\rangle$. In Sects. 4.4 and 4.5, we study the descent properties of such a Lyapunov candidate. The analysis is divided into two cases, depending on how x is far from $x^\star$. It turns out that the aforementioned cross-term is key to prove that the Lyapunov function decreases when x is close to $x^\star$, as it produces a negative term proportional to $|\tilde\lambda_{\mathrm{A}}|^2$ in the evolution equation of the Lyapunov candidate (see (46)). In turn, this term was missing from Uzawa's analysis in [29]. In Sect. 4.6, we use the Lyapunov candidate to establish equiboundedness of the solutions and convergence to the optimum $(x^\star,\lambda^\star)$. Finally, in Sect. 4.7, we prove the exponential bound (6).

We fix now, once and for all, an arbitrary compact set $\Xi_0\subseteq\mathbb{R}^n\times(\mathbb{R}_{\ge0})^r$ for the initial conditions of (5). We stress that Ξ_0 can be any, arbitrarily large, compact set.

4.1 Preliminary definitions

Consider a $K_0>0$ be such that

$$\Xi_0\subseteq\overline{\mathbb{B}}_{K_0}(x^\star,\lambda^\star). \tag{7}$$

Let us fix once and for all

$$K \geq 2K_0 + 1,$$

and define

$$\mathscr{K} := \overline{\mathbb{B}}_K(x^\star, \lambda^\star), \qquad \mathscr{K}_x := \left\{x \in \mathbb{R}^n \ : \ \exists \lambda \in \mathbb{R}^r,\ (x, \lambda) \in \mathscr{K}\right\}.$$

Since f is strongly convex (Assumption 1-A), there exists $c_0 > 0$ such that

$$\forall x \in \mathbb{R}, \quad \langle x - x^\star \mid \nabla f(x) - \nabla f(x^\star) \rangle \geq c_0 |x - x^\star|^2. \tag{8}$$

Since $\mathscr{K}$ is compact, the smoothness assumptions 1-A and 1-B together with the optimality conditions (4) imply the existence of $k_1, k_2, k_3, k_4 > 0$ such that

$$\begin{aligned}
&\forall x, \xi \in \mathscr{K}_x, && |g(x) - g(\xi)| \leq k_1 |x - \xi|,\\
&\forall (x, \lambda) \in \mathscr{K}, && |\nabla L(x, \lambda)| \leq k_2(|x - x^\star| + |\lambda - \lambda^\star|),\\
&\forall x, \xi \in \mathscr{K}_x, && |\nabla f(x) - \nabla f(\xi)| \leq k_3 |x - \xi|,\\
&\forall x, \xi \in \mathscr{K}_x, && |\nabla g(x) - \nabla g(\xi)| \leq k_4 |x - \xi|.
\end{aligned} \tag{9}$$

Moreover, we can define the following constants

$$k_5 := \sup_{(x,\lambda)\in\mathscr{K}} |\nabla L(x, \lambda)|, \quad k_6 := \sup_{x\in\mathscr{K}_x} |\nabla g(x)|, \quad k_7 := \sup_{x\in\mathscr{K}_x} |g(x)|. \tag{10}$$

Let $r_a \leq r$ denote the number of active constraints at $x^\star$. Without loss of generality, we assume that these active constraints are associated with the indices $i \in \mathrm{A} := \{1, \ldots, r_a\}$. Thus, we have $g_i(x^\star) = 0$ for all $i \in \mathrm{A}$, and $g_i(x^\star) < 0$ for all $i \in \mathrm{I} := \{r_a + 1, \ldots, r\}$. Let $\lambda_{\mathrm{A}} := (\lambda_1, \ldots, \lambda_{r_a})$ collect all multipliers associated with active constraints, and $\lambda_{\mathrm{I}} := (\lambda_{r_a+1}, \ldots, \lambda_r)$ those associated with inactive constraints. Let g_{A} and g_{I} be defined accordingly. Then

$$g_{\mathrm{A}}(x^\star) = 0, \qquad \lambda^\star_{\mathrm{A}} \geq 0, \qquad g_{\mathrm{I}}(x^\star) < 0, \qquad \lambda^\star_{\mathrm{I}} = 0. \tag{11}$$

and, for all $(x, \lambda) \in \mathbb{R}^n \times (\mathbb{R}_{\geq 0})^r$,

$$|\lambda|^2 = |\lambda_{\mathrm{A}}|^2 + |\lambda_{\mathrm{I}}|^2, \qquad \nabla g(x)\lambda = \nabla g_{\mathrm{A}}(x)\lambda_{\mathrm{A}} + \nabla g_{\mathrm{I}}(x)\lambda_{\mathrm{I}}. \tag{12}$$

Moreover, Assumption 2 implies

$$\nabla g_{\mathrm{A}}(x^\star)^\top \nabla g_{\mathrm{A}}(x^\star) > 0. \tag{13}$$

Since ∇g is continuous (Assumption 1-B), there exist $q > 0$ and $\bar{\varepsilon}_1 > 0$ such that the following conditions hold

$$\forall x \in \mathbb{R}^n, \quad |x - x^\star| \leq \bar{\varepsilon}_1 \implies \nabla g_{\mathrm{A}}(x)^\top \nabla g_{\mathrm{A}}(x) \geq qI, \tag{14a}$$

$$\forall x \in \mathbb{R}^n, \quad |x - x^\star| \le \bar{\varepsilon}_1 \implies g_{\mathrm{I}}(x) < 0. \tag{14b}$$

For ease of notation, define

$$\begin{aligned}
&\alpha_1 := \frac{q}{k_4k_5 + k_6(k_3 + k_4|\lambda^\star|)}, \qquad \alpha_2 := \frac{k_4k_5 + k_6(k_3 + k_4|\lambda^\star|)}{2\alpha_1} + k_1k_6,\\
&\alpha_3 := \frac{3k_1k_2k_6 + k_2k_4k_5}{2}, \qquad \alpha_4 := \frac{k_1k_2k_6 + 3k_2k_4k_5}{2},\\
&\alpha_7 := k_2k_6 + + \frac{\beta}{2}\big(Kk_4k_6 + k_4k_5 + k_2k_6\big), \qquad \alpha_8 := \frac{\beta}{2}\left(k_1^2k_6^2 + Kk_4k_6k_2^2\right),\\
&\alpha_9 := \beta\left(\frac{k_2k_6}{2} + k_6^2\right) + k_2k_6, \qquad \alpha_{10} := \beta\frac{Kk_4k_6k_2^2}{2},\\
&\alpha_{11} := \frac{\beta}{2}\left(k_4k_5 + k_2k_6\frac{1+\delta_1+\delta_1^2}{\delta_1}\right) + k_2k_6\frac{1+\delta_1}{\delta_1} + k_6^2 + \beta k_6^2\frac{1+2\delta_1}{4\delta_1} + \beta Kk_4k_6,
\end{aligned} \tag{15}$$

in which we denote

$$\beta := 6\frac{k_2^2}{q}, \qquad \delta_1 := \frac{k_2^2}{8\alpha_9}, \qquad \delta_2 := \begin{cases} 0 & \text{if } r = r_a \\ \frac{c_0}{2(r-r_a)(1+\beta)k_1} & \text{if } r < r_a. \end{cases} \tag{16}$$

Next, define

$$h := \min_{i\in \mathrm{I}} |g_i(x^\star)|, \qquad \bar{\varepsilon} := \min\left\{\bar{\varepsilon}_1,\ \frac{h}{4k_1(1+2\beta)}\right\}, \tag{17}$$

(notice that $h > 0$ in view of (11)) and we fix once and for all (and arbitrarily)

$$\varepsilon \in (0, \bar{\varepsilon}). \tag{18}$$

Finally, with

$$\bar{\gamma}_1 := \frac{1}{16K_0k_5}, \qquad \bar{\gamma}_2 := \frac{1}{2k_5}, \qquad \bar{\gamma}_3 := \frac{1}{16K_0k_7}, \qquad \bar{\gamma}_4 := \frac{1}{2k_7}, \tag{19a}$$

$$\bar{\gamma}_5 := \frac{1}{\beta k_6}, \qquad \bar{\gamma}_6 := \frac{c_0\varepsilon^2}{\beta\big(4K_0^2k_4k_5 + 2K_0k_6k_7 + k_5k_6K\big) + k_7^2 + k_5^2}, \tag{19b}$$

$$\bar{\gamma}_7 := \min\left\{\frac{c_0}{2(k_1^2 + 2k_2^2 + \beta\alpha_2)},\ \sqrt{\frac{c_0}{2\beta\alpha_3}}\right\}, \qquad \bar{\gamma}_8 := \frac{k_2^2}{2\beta\alpha_4}, \tag{19c}$$

$$\bar{\gamma}_9 := \min\left\{\frac{1}{2\sqrt{\alpha_{11}}},\ \frac{\delta_2}{4(1+\beta)k_1}\right\}, \quad \bar{\gamma}_{10} := \frac{h}{4\alpha_{11}K}, \tag{19d}$$

$$\bar{\gamma}_{11} := \min\left\{\frac{c_0}{8\alpha_7},\ \frac{1}{2}\sqrt[3]{\frac{c_0}{\alpha_{10}}},\ \frac{k_2}{2\sqrt{2\alpha_{10}}}\right\}, \quad \bar{\gamma}_{12} := \frac{2h}{Kk_2^2}, \tag{19e}$$

we fix arbitrarily, once and for all, the value of γ as

$$\gamma \in (0, \bar{\gamma}), \qquad \bar{\gamma} := \min_{i=1,\dots,12} \bar{\gamma}_i. \tag{20}$$

The specific value of each of the above-defined constants is motivated by the derivations carried out in the following subsections. We stated all the definitions here to highlight that no circular dependencies arise. Specifically, one can readily verify that: *(i)* the constants K_0 and K only depend on the optimal point $(x^\star, \lambda^\star)$ and the initialization set Ξ_0; *(ii)* the constants $k_1, \dots, k_7$, defined in (9)–(10), only depend on the functions f and g, on $(x^\star, \lambda^\star)$, and on the previously-defined constant K; *(iii)* q and $\bar{\varepsilon}_1$ in (14) only depend on g; *(iv)* β only depends on k_2 and q; *(v)* δ_1 and δ_2 only depend on β and k_1, k_2, k_6; *(vi)* the constants $\alpha_1, \dots, \alpha_{11}$ only depend on the previously-defined quantities; *(vii)* h, $\bar{\varepsilon}$ and, hence, ε, only depend on g, $x^\star$, $\bar{\varepsilon}_1$, k_1 and β; *(viii)* the remaining constants $\bar{\gamma}_1, \dots, \bar{\gamma}_{12}$ only depend on the previously-defined constants.

4.2 Preparatory lemmas

In this subsection, we prove some preliminary technical lemmas that will be used in the forthcoming analysis. For notational convenience, we let

$$\tilde{x} = x - x^\star, \qquad \tilde{\lambda} = \lambda - \lambda^\star.$$

In view of (5), these variables satisfy the recursion

$$\tilde{x}^+ = \tilde{x} - \gamma \nabla L(x, \lambda), \tag{21a}$$

$$\tilde{\lambda}^+ = \max\big\{ -\lambda^\star,\ \tilde{\lambda} + \gamma g(x) \big\}. \tag{21b}$$

Since $\lambda^+ = \max\{0, \lambda + \gamma g(x)\} \geq \lambda + \gamma g(x)$ and since $\lambda^\star \geq 0$ in view of (4), we have $-2\langle \lambda^+ \mid \lambda^\star \rangle \leq -2\langle \lambda \mid \lambda^\star \rangle - 2\gamma \langle \lambda^\star \mid g(x) \rangle$. Therefore, we can write

$$|\tilde{x}^+|^2 = |\tilde{x} - \gamma \nabla L(x, \lambda)|^2 = |\tilde{x}|^2 - 2\gamma \langle \tilde{x} \mid \nabla L(x, \lambda) \rangle + \gamma^2 |\nabla L(x, \lambda)|^2 \tag{22a}$$

and

$$\begin{aligned} |\tilde{\lambda}^+|^2 &= \big|\lambda^+ - \lambda^\star\big|^2 = |\lambda^+|^2 - 2\langle \lambda^+ \mid \lambda^\star \rangle + |\lambda^\star|^2 \\ &\leq |\lambda + \gamma g(x)|^2 - 2\langle \lambda \mid \lambda^\star \rangle - 2\gamma \langle \lambda^\star \mid g(x) \rangle + |\lambda^\star|^2 \\ &\leq |\lambda|^2 + |\lambda^\star|^2 - 2\langle \lambda \mid \lambda^\star \rangle + 2\gamma \langle \tilde{\lambda} \mid g(x) \rangle + \gamma^2 |g(x)|^2 \\ &= |\tilde{\lambda}|^2 + 2\gamma \langle \tilde{\lambda} \mid g(x) \rangle + \gamma^2 |g(x)|^2. \end{aligned} \tag{22b}$$

Lemma 3 *Suppose that Assumption 1 holds and let $c_0 > 0$ be given by (8). Then,*

$$\forall (x, \lambda) \in \mathbb{R}^n \times (\mathbb{R}_{\geq 0})^r, \quad \langle \tilde{x} \mid \nabla f(x) + \nabla g(x)\lambda \rangle - \langle \tilde{\lambda} \mid g(x) \rangle \geq c_0 |\tilde{x}|^2.$$

Proof Since $\lambda^\star \geq 0$ and $\langle \tilde{x} \mid \nabla f(x)\rangle \geq \langle \tilde{x} \mid \nabla f(x^\star)\rangle + c_0|\tilde{x}|^2$ (see the strong convexity condition in (8)), we can write

$$\begin{aligned}
&\langle \tilde{x} \mid \nabla f(x) + \nabla g(x)\lambda\rangle - \langle \tilde{\lambda} \mid g(x)\rangle \\
&\quad = \langle \tilde{x} \mid \nabla f(x)\rangle + \langle \nabla g(x)^\top \tilde{x} - g(x) \mid \lambda\rangle + \langle g(x) \mid \lambda^\star\rangle \\
&\quad \geq \langle \tilde{x} \mid \nabla f(x^\star)\rangle + c_0|\tilde{x}|^2 + \langle \nabla g(x)^\top \tilde{x} - g(x) \mid \lambda\rangle + \langle g(x^\star) + \nabla g(x^\star)^\top \tilde{x} \mid \lambda^\star\rangle \\
&\quad = c_0|\tilde{x}|^2 + \langle \tilde{x} \mid \underbrace{\nabla f(x^\star) + \nabla g(x^\star)\lambda^\star}_{=0}\rangle + \underbrace{\langle g(x^\star) \mid \lambda^\star\rangle}_{=0 \text{ by } (4)} + \underbrace{\langle \nabla g(x)^\top \tilde{x} - g(x) \mid \lambda\rangle}_{\substack{\geq -\langle g(x^\star) \mid \lambda\rangle \geq 0 \\ \text{by } (4) \text{ and } (23) \text{ with} (w,y)=(x^\star,x)}} \\
&\quad \geq c_0|\tilde{x}|^2,
\end{aligned}$$

where, in the first inequality, we also used convexity[4] of the g_i (cf. condition (23) with the identification $(w, y) = (x, x^\star)$). □

Lemma 4 *Suppose that Assumption 1 holds, and let γ satisfy (20). Then, system (5) satisfies*

$$(x, \lambda) \in \overline{\mathbb{B}}_{2K_0}(x^\star, \lambda^\star) \implies (x^+, \lambda^+) \in \mathscr{K}.$$

Proof With reference to the constants introduced in (9)–(10), notice that, since $\gamma < \min\{\bar{\gamma}_1, \bar{\gamma}_2, \bar{\gamma}_3, \bar{\gamma}_4\}$ (see 20) and $\overline{\mathbb{B}}_{2K_0}(x^\star, \lambda^\star) \subseteq \mathscr{K}$, then (19a), (21), and (22) imply

$$\begin{aligned}
|\tilde{x}^+|^2 &= |\tilde{x}|^2 - 2\gamma\langle \tilde{x} \mid \nabla L(x, \lambda)\rangle + \gamma^2|\nabla L(x, \lambda)|^2 \\
&\leq |\tilde{x}|^2 + \gamma 4K_0k_5 + \gamma^2k_5^2 \leq |\tilde{x}|^2 + \frac{1}{2} \\
|\tilde{\lambda}^+|^2 &\leq |\tilde{\lambda}|^2 + 2\gamma\langle \tilde{\lambda} \mid g(x)\rangle + \gamma^2|g(x)|^2 \leq |\tilde{\lambda}|^2 + \gamma 4K_0k_7 + \gamma^2k_7^2 \leq |\tilde{\lambda}|^2 + \frac{1}{2}
\end{aligned}$$

for all $(x, \lambda) \in \overline{\mathbb{B}}_{2K_0}(x^\star, \lambda^\star)$. In the previous inequalities, we have used the fact that $\gamma < \min\{\bar{\gamma}_1, \bar{\gamma}_2, \bar{\gamma}_3, \bar{\gamma}_4\}$ implies

$$\begin{aligned}
\gamma 4K_0k_5 + \gamma^2k_5^2 &< \bar{\gamma}_1 4K_0k_5 + \bar{\gamma}_2^2k_5^2 = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}, \\
\gamma 4K_0k_7 + \gamma^2k_7^2 &< \bar{\gamma}_3 4K_0k_7 + \bar{\gamma}_4^2k_7^2 = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}.
\end{aligned}$$

Combining the previous inequalities, we then get

$$|\tilde{x}^+|^2 + |\tilde{\lambda}^+|^2 \leq |\tilde{x}|^2 + |\tilde{\lambda}|^2 + 1 \implies |(\tilde{x}^+, \tilde{\lambda}^+)| \leq |(\tilde{x}, \tilde{\lambda})| + 1 \leq 2K_0 + 1,$$

which implies $(x^+, \lambda^+) \in \mathscr{K}$. □

---

[4] Being each g_i convex and $\mathscr{C}^1$ (Assumption 1-B), it satisfies

$$\forall w, y \in \mathbb{R}^n, \qquad g_i(w) \geq g_i(y) + \nabla g_i(y)^\top (w - y). \tag{23}$$

Lemma 5 *Every solution of (5) satisfies* $\lambda^t \geq 0$ *and* $|\tilde{\lambda}^{t+1} - \tilde{\lambda}^t| \leq \gamma |g(x^t)|$ *for all* $t \in \mathbb{N}$.

Proof The fact that $\lambda^t \geq 0$ for all $t \geq 0$ is obvious. Regarding the second claim, pick $i \in \{1, \ldots, r\}$ and $t \in \mathbb{N}$ arbitrarily. From (21b), we obtain

$$\tilde{\lambda}_i^{t+1} = \max\left\{0, \lambda_i^t + \gamma g_i(x^t)\right\} - \lambda_i^\star. \tag{24}$$

First, assume that $\lambda_i^t + \gamma g_i(x^t) \geq 0$. Then (24) yields $\tilde{\lambda}_i^{t+1} - \tilde{\lambda}_i^t = \gamma g_i(x^t)$, hence $|\tilde{\lambda}_i^{t+1} - \tilde{\lambda}_i^t| = \gamma |g_i(x^t)|$. On the other hand, suppose that $\lambda_i^t + \gamma g_i(x^t) < 0$. Since $\lambda_i^t \geq 0$, then $g_i(x^t) < 0$, and (24) implies

$$|\tilde{\lambda}_i^{t+1} - \tilde{\lambda}_i^t| = | - \lambda_i^\star - \tilde{\lambda}_i^t| = | - \lambda_i^t| = \lambda_i^t \leq -\gamma g_i(x^t) = \gamma |g_i(x^t)|.$$

Hence, in both cases, $|\tilde{\lambda}_i^{t+1} - \tilde{\lambda}_i^t| \leq \gamma |g_i(x^t)|$. As i was arbitrary, we obtain

$$|\tilde{\lambda}^{t+1} - \tilde{\lambda}^t|^2 = \sum_{i=1,\ldots,r} |\tilde{\lambda}_i^{t+1} - \tilde{\lambda}_i^t|^2 \leq \gamma^2 \sum_{i=1,\ldots,r} |g_i(x^t)|^2 = \gamma^2 |g(x^t)|^2,$$

which concludes the proof. □

4.3 The Lyapunov candidate

Next, we propose the Lyapunov candidate used later to establish stability and convergence. In this part, we prove some of its basic properties. Specifically, with β defined in (16), we define the Lyapunov candidate

$$V(x, \lambda) := |\tilde{x}|^2 + |\tilde{\lambda}|^2 + \gamma\beta \langle \tilde{x} \mid \nabla g(x)\tilde{\lambda} \rangle. \tag{25}$$

The following lemma shows that V is positive definite with respect to $(x^\star, \lambda^\star)$.

Lemma 6 *Suppose that Assumption 1 holds, and let* γ *satisfy* (20)*. Then,*

$$\forall (x, \lambda) \in \mathscr{K}, \qquad \frac{1}{2}|(\tilde{x}, \tilde{\lambda})|^2 \leq V(x, \lambda) \leq \frac{3}{2}|(\tilde{x}, \tilde{\lambda})|^2. \tag{26}$$

Proof As for the upper bound in (26), notice that $(x, \lambda) \in \mathscr{K}$ and $\gamma < \bar{\gamma}_5$ (see (19b)) imply

$$\begin{aligned} V(x, \lambda) &\leq |\tilde{x}|^2 + |\tilde{\lambda}|^2 + \gamma\beta k_6 |\tilde{x}||\tilde{\lambda}| \leq |\tilde{x}|^2 + |\tilde{\lambda}|^2 + \gamma \frac{\beta k_6}{2}(|\tilde{x}|^2 + |\tilde{\lambda}|^2) \\ &\leq \frac{3}{2}(|\tilde{x}|^2 + |\tilde{\lambda}|^2), \end{aligned}$$

in which, in the second inequality, we used the Young's inequality $|\tilde{x}||\tilde{\lambda}| \leq \frac{1}{2}(|\tilde{x}|^2 + |\tilde{\lambda}|^2)$. Similarly, we obtain

$$\begin{aligned} |\tilde{x}|^2 + |\tilde{\lambda}|^2 &= V(x,\lambda) - \gamma\beta\langle \tilde{x} \mid \nabla g(x)\tilde{\lambda}\rangle \leq V(x,\lambda) + \gamma\beta k_6 |\tilde{x}||\tilde{\lambda}| \\ &\leq V(x,\lambda) + \frac{1}{2}(|\tilde{x}|^2 + |\tilde{\lambda}|^2), \end{aligned}$$

which gives the lower bound in (26). □

Next, we define

$$\Omega_\rho := \{(x,\lambda) \,:\, V(x,\lambda) \leq \rho\}, \tag{27}$$

with

$$\rho := \frac{3}{2} K_0^2. \tag{28}$$

Then, the following lemma shows that the level set Ω_ρ lies in between $\overline{\mathbb{B}}_{K_0}(x^\star, \lambda^\star)$ and $\overline{\mathbb{B}}_{2K_0}(x^\star, \lambda^\star)$.

Lemma 7 *Suppose that Assumption 1 holds, and let γ satisfy (20). Then,*

$$\overline{\mathbb{B}}_{K_0}(x^\star, \lambda^\star) \subseteq \Omega_\rho \subseteq \overline{\mathbb{B}}_{2K_0}(x^\star, \lambda^\star).$$

Proof In view of Lemma 6, we have

$$(x,\lambda) \in \overline{\mathbb{B}}_{K_0}(x^\star, \lambda^\star) \implies V(x,\lambda) \leq \frac{3}{2} K_0^2 = \rho \implies (x,\lambda) \in \Omega_\rho,$$

which proves the first inclusion. As for the second inclusion, we have

$$\begin{aligned} (x,\lambda) \in \Omega_\rho &\implies |(\tilde{x},\tilde{\lambda})|^2 \leq 2V(x,\lambda) \leq 2\rho = 3K_0^2 \\ &\implies |(\tilde{x},\tilde{\lambda})| \leq \sqrt{3} K_0 \leq 2K_0, \end{aligned}$$

which implies $(x,\lambda) \in \overline{\mathbb{B}}_{2K_0}(x^\star, \lambda^\star)$. □

In the next two subsections, we show that the Lyapunov candidate V in (25) is strictly decreasing on Ω_ρ along the solutions of (5). We subdivide the proof in two cases, corresponding to the partition of Ω_ρ in the following two sets

$$\Omega_\rho^{>\varepsilon} := \big\{(x,\lambda) \in \Omega_\rho \,:\, |x - x^\star| > \varepsilon\big\}, \qquad \Omega_\rho^{\leq\varepsilon} := \big\{(x,\lambda) \in \Omega_\rho \,:\, |x - x^\star| \leq \varepsilon\big\}, \tag{29}$$

where, we recall, ε has been fixed to satisfy (18).

As a preliminary step, common to both cases, we combine the inequalities (22) to obtain

$$V(x,\lambda)^+ \leq |\tilde{x}|^2 + |\tilde{\lambda}|^2 + \gamma^2\Big(|g(x)|^2 + |\nabla L(x,\lambda)|^2\Big) \\ -2\gamma\Big(\langle \tilde{x} \mid \nabla L(x,\lambda)\rangle - \langle \tilde{\lambda} \mid g(x)\rangle\Big) + \gamma\beta\langle \tilde{x}^+ \mid \nabla g(x^+)\tilde{\lambda}^+\rangle. \quad (30)$$

4.4 Descent on $\Omega_\rho^{>\varepsilon}$

We first focus on the last term of (30). In view of (21), and by adding and subtracting proper cross terms, we obtain

$$\begin{aligned}\gamma\beta\langle \tilde{x}^+ \mid \nabla g(x^+)\tilde{\lambda}^+\rangle &= \gamma\beta\langle \tilde{x} \mid \nabla g(x^+)\tilde{\lambda}^+\rangle - \gamma^2\beta\langle \nabla L(x,\lambda) \mid \nabla g(x^+)\tilde{\lambda}^+\rangle \\ &= \gamma\beta\langle \tilde{x} \mid \nabla g(x^+)\tilde{\lambda}\rangle + \gamma\beta\langle \tilde{x} \mid \nabla g(x^+)(\tilde{\lambda}^+ - \tilde{\lambda})\rangle \\ &\quad - \gamma^2\beta\langle \nabla L(x,\lambda) \mid \nabla g(x^+)\tilde{\lambda}^+\rangle \\ &= \gamma\beta\langle \tilde{x} \mid \nabla g(x)\tilde{\lambda}\rangle + \gamma\beta\langle \tilde{x} \mid (\nabla g(x^+) - \nabla g(x))\tilde{\lambda}\rangle \\ &\quad + \gamma\beta\langle \tilde{x} \mid \nabla g(x^+)(\tilde{\lambda}^+ - \tilde{\lambda})\rangle - \gamma^2\beta\langle \nabla L(x,\lambda) \mid \nabla g(x^+)\tilde{\lambda}^+\rangle.\end{aligned} \quad (31)$$

We recall that from Lemmas 4 and 7 it follows that

$$(x,\lambda) \in \Omega_\rho \implies (x,\lambda) \in \overline{\mathbb{B}}_{2K_0}(x^\star,\lambda^\star) \implies (x^+,\lambda^+) \in \mathscr{K}. \quad (32)$$

Therefore, if $(x,\lambda) \in \Omega_\rho$, the bounds (9) and (10) apply to both (x,λ) and (x^+,λ^+). In particular, we have

$$|\nabla g(x^+) - \nabla g(x)| \leq \gamma k_4 |\nabla L(x,\lambda)|. \quad (33)$$

Hence, as long as $(x,\lambda) \in \Omega_\rho$, we can further manipulate (31) by using (33), (9), (10) and Lemma 5 to obtain

$$\begin{aligned}\gamma\beta\langle \tilde{x}^+ \mid \nabla g(x^+)\tilde{\lambda}^+\rangle &\leq \gamma\beta\langle \tilde{x} \mid \nabla g(x)\tilde{\lambda}\rangle + \gamma^2\beta k_4 |\nabla L(x,\lambda)||\tilde{x}||\tilde{\lambda}| \\ &\quad + \gamma^2\beta k_6 |\tilde{x}||g(x)| + \gamma^2\beta k_5 k_6 |\tilde{\lambda}^+| \\ &\leq \gamma\beta\langle \tilde{x} \mid \nabla g(x)\tilde{\lambda}\rangle + \gamma^2\beta\big(4K_0^2 k_4 k_5 + 2K_0 k_6 k_7 + k_5 k_6 K\big),\end{aligned}$$

in which we also used the fact that, since $(x,\lambda) \in \Omega_\rho$, then $|\tilde{\lambda}^+| \leq K$ as implied by Lemma 4. Hence, by using Lemma 3 and $\gamma < \bar{\gamma}_6$ (see (19b)), from (30) we obtain

$$\begin{aligned}&V(x,\lambda)^+ - V(x,\lambda) \\ &\quad \leq -2\gamma c_0 |\tilde{x}|^2 + \gamma^2\Big(g(x)^2 + |\nabla L(x,\lambda)|^2\Big) \\ &\qquad + \gamma^2\beta\big(4K_0^2 k_4 k_5 + 2K_0 k_6 k_7 + k_5 k_6 K\big)\end{aligned}$$

$$\leq -2\gamma c_0|\tilde{x}|^2 + \gamma^2\Big(\beta\big(4K_0^2k_4k_5 + 2K_0k_6k_7 + k_5k_6K\big) + k_7^2 + k_5^2\Big)$$
$$\leq -2\gamma c_0|\tilde{x}|^2 + \gamma c_0\varepsilon^2.$$

Since $(x,\lambda) \in \Omega_\rho^{>\varepsilon} \implies |\tilde{x}|^2 \geq \varepsilon^2$, we finally conclude that

$$\forall (x,\lambda) \in \Omega_\rho^{>\varepsilon}, \qquad V(x,\lambda)^+ - V(x,\lambda) \leq -\gamma c_0 \varepsilon^2 < 0. \tag{34}$$

4.5 Descent on $\Omega_\rho^{\leq\varepsilon}$

Recall the decomposition of λ in λ_A and λ_I (Sect. 4.1), in which $A = \{1,\ldots,r_a\}$ is the set of indices i associated with active constraints (i.e., satisfying $g_i(x^\star) = 0$) and $I = \{r_a+1,\ldots,r\}$ that of indices i associated with inactive constraints (i.e., satisfying $g_i(x^\star) < 0$). Notice that (11) implies $\tilde{\lambda}_I = \lambda_I$. Moreover, since $\nabla g(x)\tilde{\lambda} = \nabla g_A(x)\tilde{\lambda}_A + \nabla g_I(x)\tilde{\lambda}_I = \nabla g_A(x)\tilde{\lambda}_A + \nabla g_I(x)\lambda_I$, we can rewrite V as

$$V(x,\lambda) = V_A(x,\lambda) + V_I(x,\lambda), \tag{35}$$

in which

$$V_A(x,\lambda) := |\tilde{x}|^2 + |\tilde{\lambda}_A|^2 + \gamma\beta\langle \tilde{x} \mid \nabla g_A(x)\tilde{\lambda}_A\rangle, \tag{36a}$$
$$V_I(x,\lambda) := |\lambda_I|^2 + \gamma\beta\langle \tilde{x} \mid \nabla g_I(x)\lambda_I\rangle. \tag{36b}$$

Notice that $V_A(x,\lambda)$ only depends on λ_A, and not on λ_I. In the next sections we analyze the behavior of V_A and V_I on $\Omega_\rho^{\leq\varepsilon}$.

4.5.1 Bounding $V_A(x,\lambda)^+$ on $\Omega_\rho^{\leq\varepsilon}$

With slight abuse of notation, define

$$\nabla L_A(x,\lambda_A) := \nabla f(x) + \nabla g_A(x)\lambda_A, \quad x_A^+ := x - \gamma\nabla L_A(x,\lambda_A), \quad \tilde{x}_A^+ := x_A^+ - x^\star. \tag{37}$$

We notice that, in view of (37), if $\lambda_I = 0$, then $x^+ = x_A^+$.

With the previous definitions in mind, notice that (12) implies

$$\nabla L(x,\lambda) = \nabla L_A(x,\lambda_A) + \nabla g_I(x)\lambda_I,$$
$$\tilde{x}_A^+ = \tilde{x} - \gamma\nabla L_A(x,\lambda_A) = \tilde{x}^+ + \gamma\nabla g_I(x)\tilde{\lambda}_I.$$

In addition, bounds analogous to (9) and (22b) hold for L_A and λ_A. Hence, using (12), (22a), and proceeding as in (22b), we obtain

$$V_A(x,\lambda)^+ \leq U(x,\lambda) + W(x,\lambda), \tag{38}$$

in which

$$\begin{aligned} U(x,\lambda) &:= |\tilde{x}|^2 + |\tilde{\lambda}_{\mathrm{A}}|^2 + \gamma^2\Big(|g_{\mathrm{A}}(x)|^2 + |\nabla L_{\mathrm{A}}(x,\lambda_{\mathrm{A}})|^2\Big) \\ &\quad - 2\gamma\Big(\langle \tilde{x} \mid \nabla L_{\mathrm{A}}(x,\lambda_{\mathrm{A}})\rangle - \langle \tilde{\lambda}_{\mathrm{A}} \mid g_{\mathrm{A}}(x)\rangle\Big) \\ &\quad + \gamma\beta\langle \tilde{x}_{\mathrm{A}}^{+} \mid \nabla g_{\mathrm{A}}(x_{\mathrm{A}}^{+})\tilde{\lambda}_{\mathrm{A}}^{+}\rangle \end{aligned} \tag{39}$$

and

$$\begin{aligned} W(x,\lambda) &:= \mathscr{E}_1 + \mathscr{E}_2 + \mathscr{E}_3 + \mathscr{E}_4, \\ \mathscr{E}_1 &:= \gamma^2\left(2\langle \nabla L_{\mathrm{A}}(x,\lambda_{\mathrm{A}}) \mid \nabla g_{\mathrm{I}}(x)\lambda_{\mathrm{I}}\rangle + |\nabla g_{\mathrm{I}}(x)\lambda_{\mathrm{I}}|^2\right), \\ \mathscr{E}_2 &:= -2\gamma\langle \tilde{x} \mid \nabla g_{\mathrm{I}}(x)\lambda_{\mathrm{I}}\rangle, \\ \mathscr{E}_3 &:= -\gamma^2\beta\langle \nabla g_{\mathrm{I}}(x)\lambda_{\mathrm{I}} \mid \nabla g_{\mathrm{A}}(x^{+})\tilde{\lambda}_{\mathrm{A}}^{+}\rangle, \\ \mathscr{E}_4 &:= \gamma\beta\langle \tilde{x}_{\mathrm{A}}^{+} \mid (\nabla g_{\mathrm{A}}(x^{+}) - \nabla g_{\mathrm{A}}(x_{\mathrm{A}}^{+}))\tilde{\lambda}_{\mathrm{A}}^{+}\rangle. \end{aligned} \tag{40}$$

We notice that $U(x,\lambda)$ only depends on λ_{A}, and not on λ_{I}.

In the following, we bound the two terms in (38) separately. As for $U(x,\lambda)$, we start by noticing that (4a), (8), (9), and (11) imply

$$|g_{\mathrm{A}}(x)| = |g_{\mathrm{A}}(x) - g_{\mathrm{A}}(x^{\star})| \le k_1|\tilde{x}|, \tag{41a}$$

$$\nabla L_{\mathrm{A}}(x^{\star},\lambda_{\mathrm{A}}^{\star}) = \nabla L(x^{\star},\lambda^{\star}) = 0, \tag{41b}$$

$$|\nabla L_{\mathrm{A}}(x,\lambda_{\mathrm{A}})| = |\nabla L_{\mathrm{A}}(x,\lambda_{\mathrm{A}}) - \nabla L_{\mathrm{A}}(x^{\star},\lambda_{\mathrm{A}}^{\star})| \le k_2\big(|\tilde{x}| + |\tilde{\lambda}_{\mathrm{A}}|\big), \tag{41c}$$

$$\langle \tilde{x} \mid \nabla L_{\mathrm{A}}(x,\lambda_{\mathrm{A}})\rangle - \langle \tilde{\lambda}_{\mathrm{A}} \mid g_{\mathrm{A}}(x)\rangle \ge c_0|\tilde{x}|^2, \tag{41d}$$

for all $(x,\lambda) \in \Omega_\rho$. In particular, (41d) can be derived by means of the same arguments used to prove Lemma 3 in view of (41b). Moreover, we observe that

$$(x,\lambda) \in \Omega_\rho \implies (x_{\mathrm{A}}^{+}, (\lambda_{\mathrm{A}},0)^{+}) \in \mathscr{K} \implies |\tilde{x}_{\mathrm{A}}^{+}|^2 + |\tilde{\lambda}_{\mathrm{A}}^{+}|^2 \le K^2. \tag{42}$$

The implications (42) can be proved as follows. By Lemma 7, $(x,\lambda) \in \Omega_\rho \implies (x,\lambda) \in \overline{\mathbb{B}}_{2K_0}(x^{\star},\lambda^{\star})$. Thus, $|(\tilde{x},(\tilde{\lambda}_{\mathrm{A}},0))| \le |(\tilde{x},(\tilde{\lambda}_{\mathrm{A}},\lambda_{\mathrm{I}}))| = |(\tilde{x},\tilde{\lambda})| \le 2K_0$ where, in the last equality, we have used $\lambda_{\mathrm{I}}^{\star} = 0$. This implies $(x,(\lambda_{\mathrm{A}},0)) \in \overline{\mathbb{B}}_{2K_0}(x^{\star},\lambda^{\star})$. Moreover, by (37), $\lambda_{\mathrm{I}} = 0 \implies x^{+} = x_{\mathrm{A}}^{+}$. Therefore, from Lemma 4 we obtain $(x_{\mathrm{A}}^{+},(\lambda_{\mathrm{A}},0)^{+}) = (x^{+},(\lambda_{\mathrm{A}},0)^{+}) \in \mathscr{K}$, which proves (42).

Conditions (9), (37) and (42) also imply

$$\forall (x,\lambda) \in \Omega_\rho, \quad |\nabla g_{\mathrm{A}}(x_{\mathrm{A}}^{+}) - \nabla g_{\mathrm{A}}(x)| \le k_4|x_{\mathrm{A}}^{+} - x| = \gamma k_4|\nabla L_{\mathrm{A}}(x,\lambda_{\mathrm{A}})|, \tag{43}$$

which will be useful later in the forthcoming computations.

Next, by using (9), (41), and Lemma 3, we obtain

$$\begin{aligned} U(x,\lambda) - V_{\mathrm{A}}(x,\lambda) &= \gamma^2\Big(|g_{\mathrm{A}}(x)|^2 + |\nabla L_{\mathrm{A}}(x,\lambda_{\mathrm{A}})|^2\Big) \\ &\quad - 2\gamma\Big(\langle \tilde{x} \mid \nabla L_{\mathrm{A}}(x,\lambda_{\mathrm{A}})\rangle - \langle \tilde{\lambda}_{\mathrm{A}} \mid g_{\mathrm{A}}(x)\rangle\Big) \\ &\quad + \gamma\beta\langle \tilde{x}_{\mathrm{A}}^+ \mid \nabla g_{\mathrm{A}}(x_{\mathrm{A}}^+)\tilde{\lambda}_{\mathrm{A}}^+\rangle - \gamma\beta\langle \tilde{x} \mid \nabla g_{\mathrm{A}}(x)\tilde{\lambda}_{\mathrm{A}}\rangle \\ &\le \big(\gamma^2 k_1^2 + \gamma^2 2k_2^2 - 2\gamma c_0\big)|\tilde{x}|^2 + \gamma^2 2k_2^2|\tilde{\lambda}_{\mathrm{A}}|^2 \\ &\quad + \gamma\beta\left(\langle \tilde{x}_{\mathrm{A}}^+ \mid \nabla g_{\mathrm{A}}(x_{\mathrm{A}}^+)\tilde{\lambda}_{\mathrm{A}}^+\rangle - \langle \tilde{x} \mid \nabla g_{\mathrm{A}}(x)\tilde{\lambda}_{\mathrm{A}}\rangle\right). \end{aligned} \tag{44}$$

The last term in (44) can be expressed as

$$\langle \tilde{x}_{\mathrm{A}}^+ \mid \nabla g_{\mathrm{A}}(x_{\mathrm{A}}^+)\tilde{\lambda}_{\mathrm{A}}^+\rangle - \langle \tilde{x} \mid \nabla g_{\mathrm{A}}(x)\tilde{\lambda}_{\mathrm{A}}\rangle = \mathscr{T}_1 + \mathscr{T}_2 + \mathscr{T}_3 + \mathscr{T}_4 + \mathscr{T}_5, \tag{45}$$

in which

$$\begin{aligned} &\mathscr{T}_1 := \langle \tilde{x} \mid (\nabla g_{\mathrm{A}}(x_{\mathrm{A}}^+) - \nabla g_{\mathrm{A}}(x))\tilde{\lambda}_{\mathrm{A}}\rangle, \\ &\mathscr{T}_2 := \langle \tilde{x} \mid \nabla g_{\mathrm{A}}(x_{\mathrm{A}}^+)(\tilde{\lambda}_{\mathrm{A}}^+ - \tilde{\lambda}_{\mathrm{A}})\rangle, \\ &\mathscr{T}_3 := -\gamma\langle \nabla L_{\mathrm{A}}(x,\lambda_{\mathrm{A}}) \mid \nabla g_{\mathrm{A}}(x_{\mathrm{A}}^+)(\tilde{\lambda}_{\mathrm{A}}^+ - \tilde{\lambda}_{\mathrm{A}})\rangle, \\ &\mathscr{T}_4 := \gamma\langle \nabla L_{\mathrm{A}}(x,\lambda_{\mathrm{A}}) \mid (\nabla g_{\mathrm{A}}(x) - \nabla g_{\mathrm{A}}(x_{\mathrm{A}}^+))\tilde{\lambda}_{\mathrm{A}}\rangle, \\ &\mathscr{T}_5 := -\gamma\langle \nabla L_{\mathrm{A}}(x,\lambda_{\mathrm{A}}) \mid \nabla g_{\mathrm{A}}(x)\tilde{\lambda}_{\mathrm{A}}\rangle. \end{aligned}$$

We now proceed in bounding all terms $\mathscr{T}_j$, $j = 1, \ldots, 5$, one-by-one. With α_1 defined in (15), by using (43) and the Young's inequality we obtain

$$\mathscr{T}_1 \le \gamma k_4|\nabla L_{\mathrm{A}}(x,\lambda_{\mathrm{A}})||\tilde{x}||\tilde{\lambda}_{\mathrm{A}}| \le \gamma k_4 k_5|\tilde{x}||\tilde{\lambda}_{\mathrm{A}}| \le \gamma k_4 k_5\left(\frac{1}{2\alpha_1}|\tilde{x}|^2 + \frac{\alpha_1}{2}|\tilde{\lambda}_{\mathrm{A}}|^2\right),$$

for all $(x,\lambda) \in \Omega_\rho$. Conditions (41), (42) and Lemma 5 also imply

$$\begin{aligned} \mathscr{T}_2 &\le |\nabla g_{\mathrm{A}}(x_{\mathrm{A}}^+)||\tilde{x}||\tilde{\lambda}_{\mathrm{A}}^+ - \tilde{\lambda}_{\mathrm{A}}| \le \gamma k_6|\tilde{x}||g_{\mathrm{A}}(x)| \le \gamma k_1 k_6|\tilde{x}|^2, \\ \mathscr{T}_3 &\le \gamma^2|\nabla g_{\mathrm{A}}(x_{\mathrm{A}}^+)||\nabla L_{\mathrm{A}}(x,\lambda_{\mathrm{A}})||g_{\mathrm{A}}(x)| \le \gamma^2 k_6 k_2 k_1(|\tilde{x}|^2 + |\tilde{x}||\tilde{\lambda}_{\mathrm{A}}|) \\ &\le \gamma^2 k_1 k_2 k_6\left(\frac{3}{2}|\tilde{x}|^2 + \frac{1}{2}|\tilde{\lambda}_{\mathrm{A}}|^2\right), \\ \mathscr{T}_4 &\le \gamma|\nabla L_{\mathrm{A}}(x,\lambda_{\mathrm{A}})||\nabla g_{\mathrm{A}}(x) - \nabla g_{\mathrm{A}}(x_{\mathrm{A}}^+)||\tilde{\lambda}_{\mathrm{A}}| \\ &\le \gamma^2 k_4|\nabla L_{\mathrm{A}}(x,\lambda_{\mathrm{A}})|^2|\tilde{\lambda}_{\mathrm{A}}| \le \gamma^2 k_2 k_4 k_5(|\tilde{x}||\tilde{\lambda}_{\mathrm{A}}| + |\tilde{\lambda}_{\mathrm{A}}|^2) \\ &\le \gamma^2 k_2 k_4 k_5\left(\frac{1}{2}|\tilde{x}|^2 + \frac{3}{2}|\tilde{\lambda}_{\mathrm{A}}|^2\right), \end{aligned}$$

for all $(x,\lambda) \in \Omega_\rho$.

Finally, by using (41b) and $\nabla g_{\mathrm{A}}(x)^\top \nabla g_{\mathrm{A}}(x) \geq qI$ for all $(x, \lambda) \in \Omega_\rho^{\leq\varepsilon}$ (see (14) and (18)), we obtain

$$\begin{aligned}
\mathscr{T}_5 &= -\gamma\langle \nabla L_{\mathrm{A}}(x, \lambda_{\mathrm{A}}) - \nabla L_{\mathrm{A}}(x^\star, \lambda_{\mathrm{A}}^\star) \,|\, \nabla g_{\mathrm{A}}(x)\tilde{\lambda}_{\mathrm{A}}\rangle \\
&= -\gamma\langle \nabla f(x) - \nabla f(x^\star) \,|\, \nabla g_{\mathrm{A}}(x)\tilde{\lambda}_{\mathrm{A}}\rangle - \gamma|\nabla g_{\mathrm{A}}(x)\tilde{\lambda}_{\mathrm{A}}|^2 \\
&\quad - \gamma\langle(\nabla g_{\mathrm{A}}(x) - \nabla g_{\mathrm{A}}(x^\star))\lambda_{\mathrm{A}}^\star \,|\, \nabla g_{\mathrm{A}}(x)\tilde{\lambda}_{\mathrm{A}}\rangle \\
&\leq \gamma k_6(k_3 + k_4|\lambda_{\mathrm{A}}^\star|)|\tilde{x}||\tilde{\lambda}_{\mathrm{A}}| - \gamma q|\tilde{\lambda}_{\mathrm{A}}|^2 \\
&\leq \gamma k_6(k_3 + k_4|\lambda^\star|)\left(\frac{1}{2\alpha_1}|\tilde{x}|^2 + \frac{\alpha_1}{2}|\tilde{\lambda}_{\mathrm{A}}|^2\right) - \gamma q|\tilde{\lambda}_{\mathrm{A}}|^2,
\end{aligned} \tag{46}$$

for all $(x, \lambda) \in \Omega_\rho^{\leq\varepsilon}$.

In view of the previous bounds, and by using (15) and (16), we can further manipulate (45) to obtain

$$\begin{aligned}
&\langle \tilde{x}_{\mathrm{A}}^+ \,|\, \nabla g_{\mathrm{A}}(x_{\mathrm{A}}^+)\tilde{\lambda}_{\mathrm{A}}^+\rangle - \langle \tilde{x} \,|\, \nabla g_{\mathrm{A}}(x)\tilde{\lambda}_{\mathrm{A}}\rangle \\
&\quad\leq \gamma\left(\frac{k_4k_5 + k_6(k_3 + k_4|\lambda^\star|)}{2\alpha_1} + k_1k_6 + \gamma\frac{3k_1k_2k_6 + k_2k_4k_5}{2}\right)|\tilde{x}|^2 \\
&\qquad + \gamma\left(-q + \frac{(k_4k_5 + k_6(k_3 + k_4|\lambda^\star|))\alpha_1}{2} + \gamma\frac{k_1k_2k_6 + 3k_2k_4k_5}{2}\right)|\tilde{\lambda}_{\mathrm{A}}|^2 \\
&\quad= \gamma\big(\alpha_2 + \gamma\alpha_3\big)|\tilde{x}|^2 + \gamma\left(-\frac{1}{2}q + \gamma\alpha_4\right)|\tilde{\lambda}_{\mathrm{A}}|^2
\end{aligned}$$

for all $(x, \lambda) \in \Omega_\rho^{\leq\varepsilon}$.

Then, going back to (44), we get

$$\begin{aligned}
U(x, \lambda) - V_{\mathrm{A}}(x, \lambda) \leq \gamma\Big(\gamma\big(k_1^2 + 2k_2^2 + \beta\alpha_2\big) + \gamma^2\beta\alpha_3 - 2c_0\Big)|\tilde{x}|^2& \\
+ \gamma^2\left(2k_2^2 + \gamma\beta\alpha_4 - \frac{1}{2}\beta q\right)|\tilde{\lambda}_{\mathrm{A}}|^2&
\end{aligned}$$

for all $(x, \lambda) \in \Omega_\rho^{\leq\varepsilon}$. By using the definition of β given in (16) and $\gamma < \bar{\gamma}_7$ (see (19c)), we obtain

$$\forall (x, \lambda) \in \Omega_\rho^{\leq\varepsilon}, \quad U(x, \lambda) - V_{\mathrm{A}}(x, \lambda) \leq -\gamma c_0|\tilde{x}|^2 + \gamma^2\big(\gamma\beta\alpha_4 - k_2^2\big)|\tilde{\lambda}_{\mathrm{A}}|^2.$$

By using $\gamma < \bar{\gamma}_8$ (see 19c), we can finally write

$$\forall (x, \lambda) \in \Omega_\rho^{\leq\varepsilon}, \quad U(x, \lambda) \leq V_{\mathrm{A}}(x, \lambda) - \gamma c_0|\tilde{x}|^2 - \frac{\gamma^2k_2^2}{2}|\tilde{\lambda}_{\mathrm{A}}|^2.$$

Summarizing the bounds derived so far, from (38) we obtain

$$\forall(x,\lambda)\in\Omega_\rho^{\le\varepsilon},\quad V_{\mathrm{A}}(x,\lambda)^+\le V_{\mathrm{A}}(x,\lambda)-\gamma c_0|\tilde{x}|^2-\frac{\gamma^2k_2^2}{2}|\tilde{\lambda}_{\mathrm{A}}|^2+W(x,\lambda), \tag{47}$$

and we can now proceed in bounding $W(x,\lambda)$.

With reference to the definition of $W(x,\lambda)$ in (40), we bound the terms $\mathscr{E}_1,\dots,\mathscr{E}_4$ one-by-one. Consider term $\mathscr{E}_1$. By using (9), (10), and (41c), we obtain

$$\begin{aligned}2\langle\nabla L_{\mathrm{A}}(x,\lambda_{\mathrm{A}})\mid\nabla g_{\mathrm{I}}(x)\lambda_{\mathrm{I}}\rangle&\le 2|\nabla L_{\mathrm{A}}(x,\lambda_{\mathrm{A}})||\nabla g_{\mathrm{I}}(x)||\lambda_{\mathrm{I}}|\\&\le k_2k_6|\tilde{x}|^2+k_2k_6\delta_1|\tilde{\lambda}_{\mathrm{A}}|^2+k_2k_6\frac{1+\delta_1}{\delta_1}|\lambda_{\mathrm{I}}|^2\end{aligned}$$

for all $(x,\lambda)\in\Omega_\rho^{\le\varepsilon}$, in which δ_1 is defined in (16). As a consequence, we obtain

$$\forall(x,\lambda)\in\Omega_\rho^{\le\varepsilon},\quad \mathscr{E}_1\le\gamma^2\left(k_2k_6|\tilde{x}|^2+k_2k_6\delta_1|\tilde{\lambda}_{\mathrm{A}}|^2+k_2k_6\frac{1+\delta_1}{\delta_1}|\lambda_{\mathrm{I}}|^2+k_6^2|\lambda_{\mathrm{I}}|^2\right), \tag{48}$$

Next, as for $\mathscr{E}_2$, we notice that $\lambda_{\mathrm{I}}\ge 0$ and convexity of each g_i (see (23)) imply

$$\begin{aligned}\langle\tilde{x}\mid\nabla g_{\mathrm{I}}(x)\lambda_{\mathrm{I}}\rangle&=\sum_{i=r_a+1}^{r}\langle\tilde{x}\mid\nabla g_i(x)\lambda_i\rangle=\sum_{i=r_a+1}^{r}\lambda_i\tilde{x}^\top\nabla g_i(x)\\&\ge\sum_{i=r_a+1}^{r}\lambda_i(g_i(x)-g_i(x^\star))=\langle g_{\mathrm{I}}(x)-g_{\mathrm{I}}(x^\star)\mid\lambda_{\mathrm{I}}\rangle.\end{aligned} \tag{49}$$

Hence,

$$\forall(x,\lambda)\in\Omega_\rho^{\le\varepsilon},\qquad \mathscr{E}_2\le-2\gamma\langle g_{\mathrm{I}}(x)-g_{\mathrm{I}}(x^\star)\mid\lambda_{\mathrm{I}}\rangle. \tag{50}$$

Furthermore, regarding $\mathscr{E}_3$, by means of the same arguments of Lemma 5, one can show that $|\tilde{\lambda}_{\mathrm{A}}^+-\tilde{\lambda}_{\mathrm{A}}|\le\gamma|g_{\mathrm{A}}(x)|\le\gamma k_1|\tilde{x}|$ for all $(x,\lambda)\in\Omega_\rho^{\le\varepsilon}$ (in which we also used (41a). Thus, in view of Lemma 4,

$$\begin{aligned}\langle\nabla g_{\mathrm{I}}(x)\lambda_{\mathrm{I}}\mid\nabla g_{\mathrm{A}}(x^+)\tilde{\lambda}_{\mathrm{A}}^+\rangle&\le|\nabla g_{\mathrm{I}}(x)||\nabla g_{\mathrm{A}}(x^+)||\lambda_{\mathrm{I}}|(|\tilde{\lambda}_{\mathrm{A}}|+|\tilde{\lambda}_{\mathrm{A}}^+-\tilde{\lambda}_{\mathrm{A}}|)\\&\le k_6^2|\lambda_{\mathrm{I}}|\Big(|\tilde{\lambda}_{\mathrm{A}}|+\gamma k_1|\tilde{x}|\Big)\\&\le k_6^2\delta_1|\tilde{\lambda}_{\mathrm{A}}|^2+k_6^2\frac{1+2\delta_1}{4\delta_1}|\lambda_{\mathrm{I}}|^2+\frac{\gamma^2k_1^2k_6^2}{2}|\tilde{x}|^2,\end{aligned}$$

which implies

$$\forall(x,\lambda)\in\Omega_\rho^{\le\varepsilon},\quad \mathscr{E}_3\le\beta\gamma^2\left(k_6^2\delta_1|\tilde{\lambda}_{\mathrm{A}}|^2+k_6^2\frac{1+2\delta_1}{4\delta_1}|\lambda_{\mathrm{I}}|^2+\frac{\gamma^2k_1^2k_6^2}{2}|\tilde{x}|^2\right). \tag{51}$$

Lastly, for what concerns $\mathscr{E}_4$, we use (42) to obtain $|\tilde{\lambda}_{\mathrm{A}}^{+}| \leq K$ and $|\nabla g_{\mathrm{A}}(x^{+}) - \nabla g_{\mathrm{A}}(x_{\mathrm{A}}^{+})| \leq k_4|x^{+} - x_{\mathrm{A}}^{+}| \leq \gamma k_4 |\nabla g_{\mathrm{I}}(x)\lambda_{\mathrm{I}}| \leq \gamma k_4 k_6 |\lambda_{\mathrm{I}}|$ for all $(x, \lambda) \in \Omega_{\rho}^{\leq \varepsilon}$. These inequalities and (41c) imply

$$\begin{aligned}\langle \tilde{x}_{\mathrm{A}}^{+} \mid (\nabla g_{\mathrm{A}}(x^{+}) - \nabla g_{\mathrm{A}}(x_{\mathrm{A}}^{+}))\tilde{\lambda}_{\mathrm{A}}^{+}\rangle &\leq |\tilde{x}_{\mathrm{A}}^{+}||\nabla g_{\mathrm{A}}(x^{+}) - \nabla g_{\mathrm{A}}(x_{\mathrm{A}}^{+})||\tilde{\lambda}_{\mathrm{A}}^{+}| \\ &\leq \gamma K k_4 k_6 |\tilde{x}_{\mathrm{A}}^{+}||\lambda_{\mathrm{I}}| \\ &\leq \gamma K k_4 k_6 \left(|\tilde{x}| + \gamma |\nabla L_{\mathrm{A}}(x, \lambda_{\mathrm{A}})|\right) |\lambda_{\mathrm{I}}| \\ &\leq \gamma K k_4 k_6 ((1 + \gamma k_2)|\tilde{x}| + \gamma k_2 |\tilde{\lambda}_{\mathrm{A}}|)|\lambda_{\mathrm{I}}| \\ &\leq \gamma K k_4 k_6 \left(\frac{(1+\gamma k_2)^2}{2}|\tilde{x}|^2 + \frac{\gamma^2 k_2^2}{2}|\tilde{\lambda}_{\mathrm{A}}|^2 + |\lambda_{\mathrm{I}}|^2\right),\end{aligned}$$

for all $(x, \lambda) \in \Omega_{\rho}^{\leq \varepsilon}$. Hence, we obtain

$$\forall (x, \lambda) \in \Omega_{\rho}^{\leq \varepsilon}, \quad \mathscr{E}_4 \leq \gamma^2 \beta K k_4 k_6 \left(\frac{1 + \gamma^2 k_2^2}{2}|\tilde{x}|^2 + \frac{\gamma^2 k_2^2}{2}|\tilde{\lambda}_{\mathrm{A}}|^2 + |\lambda_{\mathrm{I}}|^2\right). \tag{52}$$

Using (40), (48), (50), (51), and (52), we obtain

$$\begin{aligned}W(x, \lambda) \leq &-2\gamma \langle g_{\mathrm{I}}(x) - g_{\mathrm{I}}(x^{\star}) \mid \lambda_{\mathrm{I}}\rangle \\ &+ \gamma^2 \left(k_2 k_6 + \beta \gamma^2 \frac{k_1^2 k_6^2}{2} + \beta \frac{K k_4 k_6 \left(1 + \gamma^2 k_2^2\right)}{2}\right)|\tilde{x}|^2 \\ &+ \gamma^2 \left(k_2 k_6 \frac{1 + \delta_1}{\delta_1} + k_6^2 + \beta k_6^2 \frac{1 + 2\delta_1}{4\delta_1} + \beta K k_4 k_6\right)|\lambda_{\mathrm{I}}|^2 \\ &+ \gamma^2 \left(k_2 k_6 \delta_1 + \beta k_6^2 \delta_1 + \gamma^2 \beta \frac{K k_4 k_6 k_2^2}{2}\right)|\tilde{\lambda}_{\mathrm{A}}|^2\end{aligned} \tag{53}$$

for all $(x, \lambda) \in \Omega_{\rho}^{\leq \varepsilon}$.

Finally, we can further bound (47) using (53) as

$$\begin{aligned}V_{\mathrm{A}}(x, \lambda)^{+} \leq\ &V_{\mathrm{A}}(x, \lambda) - 2\gamma \langle g_{\mathrm{I}}(x) - g_{\mathrm{I}}(x^{\star}) \mid \lambda_{\mathrm{I}}\rangle \\ &+ \gamma \left(\gamma \left(k_2 k_6 + \beta \gamma^2 \frac{k_1^2 k_6^2}{2} + \beta \frac{K k_4 k_6 \left(1 + \gamma^2 k_2^2\right)}{2}\right) - c_0\right)|\tilde{x}|^2 \\ &+ \gamma^2 \left(k_2 k_6 \frac{1 + \delta_1}{\delta_1} + k_6^2 + \beta k_6^2 \frac{1 + 2\delta_1}{4\delta_1} + \beta K k_4 k_6\right)|\lambda_{\mathrm{I}}|^2 \\ &+ \gamma^2 \left(k_2 k_6 \delta_1 + \beta k_6^2 \delta_1 + \gamma^2 \beta \frac{K k_4 k_6 k_2^2}{2} - \frac{k_2^2}{2}\right)|\tilde{\lambda}_{\mathrm{A}}|^2,\end{aligned} \tag{54}$$

for all $(x, \lambda) \in \Omega_{\rho}^{\leq \varepsilon}$.

4.5.2 Bounding $V_{\mathrm{I}}(x,\lambda)^+$ on $\Omega_\rho^{\le\varepsilon}$

Consider now the function V_{I}, defined in (36b). We start noticing that, since $\tilde\lambda_{\mathrm{I}} = \lambda_{\mathrm{I}}$, then (1b) implies

$$\forall i \in \mathrm{I}, \qquad |\tilde\lambda_i^+|^2 - |\tilde\lambda_i|^2 = \begin{cases} -|\lambda_i|^2 & \lambda_i + \gamma g_i(x) \le 0 \\ \gamma\big(2\lambda_i + \gamma g_i(x)\big)g_i(x) & \lambda_i + \gamma g_i(x) > 0. \end{cases}$$

In view of (14b) and (18), $g_i(x) < 0$ holds for all $i \in \mathrm{I}$ and all $(x,\lambda) \in \Omega_\rho^{\le\varepsilon}$. Thus, $\lambda_i + \gamma g_i(x) > 0$ implies

$$\gamma\big(2\lambda_i + \gamma g_i(x)\big)g_i(x) = \gamma g_i(x)\lambda_i + \gamma\big(\lambda_i + \gamma g_i(x)\big)g_i(x) < \gamma g_i(x)\lambda_i.$$

Therefore, for every $(x,\lambda) \in \Omega_\rho^{\le\varepsilon}$, one has

$$|\tilde\lambda_{\mathrm{I}}^+|^2 - |\tilde\lambda_{\mathrm{I}}|^2 = \sum_{i=r_a+1}^{r} \left(|\tilde\lambda_i^+|^2 - |\tilde\lambda_i|^2\right) \le \sum_{i=r_a+1}^{r} \max\{-|\lambda_i|^2, \gamma g_i(x)\lambda_i\}. \tag{55}$$

We now consider the increment of the cross term in V_{I}, which satisfies

$$\begin{aligned}\langle \tilde x^+ \mid \nabla g_{\mathrm{I}}(x^+)\lambda_{\mathrm{I}}^+\rangle - \langle \tilde x \mid \nabla g_{\mathrm{I}}(x)\lambda_{\mathrm{I}}\rangle &= \langle \tilde x \mid (\nabla g_{\mathrm{I}}(x^+) - \nabla g_{\mathrm{I}}(x))\lambda_{\mathrm{I}}^+\rangle \\ &\quad - \gamma\langle \nabla L(x,\lambda) \mid \nabla g_{\mathrm{I}}(x^+)\lambda_{\mathrm{I}}^+\rangle + \langle \tilde x \mid \nabla g_{\mathrm{I}}(x)(\lambda_{\mathrm{I}}^+ - \lambda_{\mathrm{I}})\rangle.\end{aligned}$$

We bound the three terms one by one. First, notice that (14b) and (55) imply $|\lambda_{\mathrm{I}}^+| \le |\lambda_{\mathrm{I}}|$ for all $(x,\lambda) \in \Omega_\rho^{\le\varepsilon}$. Hence, proceeding as in previous section, we obtain

$$\langle \tilde x \mid (\nabla g_{\mathrm{I}}(x^+) - \nabla g_{\mathrm{I}}(x))\lambda_{\mathrm{I}}^+\rangle \le \gamma k_4 |\tilde x||\nabla L(x,\lambda)||\lambda_{\mathrm{I}}| \le \gamma\frac{k_4k_5}{2}\left(|\tilde x|^2 + |\lambda_{\mathrm{I}}|^2\right)$$

and

$$\begin{aligned}-\gamma\langle \nabla L(x,\lambda) \mid \nabla g_{\mathrm{I}}(x^+)\lambda_{\mathrm{I}}^+\rangle &\le \gamma|\nabla L(x,\lambda)||\nabla g_{\mathrm{I}}(x^+)||\lambda_{\mathrm{I}}| \\ &\le \gamma k_2k_6(|\tilde x| + |\tilde\lambda|)|\lambda_{\mathrm{I}}| \\ &\le \gamma\frac{k_2k_6}{2}\left(|\tilde x|^2 + \delta_1|\tilde\lambda_{\mathrm{A}}|^2 + \frac{1+\delta_1+\delta_1^2}{\delta_1}|\lambda_{\mathrm{I}}|^2\right)\end{aligned}$$

for all $(x,\lambda) \in \Omega_\rho^{\le\varepsilon}$. Lastly, since (14b) implies $\lambda_i^+ - \lambda_i \le 0$ for all $i \in \mathrm{I}$ and all $(x,\lambda) \in \Omega_\rho^{\le\varepsilon}$, then using convexity of each g_i (see (23)) as in (49), we obtain

$$\langle \tilde x \mid \nabla g_{\mathrm{I}}(x)(\lambda_{\mathrm{I}}^+ - \lambda_{\mathrm{I}})\rangle = \sum_{i=r_a+1}^{r} (\lambda_i^+ - \lambda_i)\tilde x^\top \nabla g_i(x) \le \langle g_{\mathrm{I}}(x) - g_{\mathrm{I}}(x^\star) \mid \lambda_{\mathrm{I}}^+ - \lambda_{\mathrm{I}}\rangle$$

for all $(x, \lambda) \in \Omega_\rho^{\leq\varepsilon}$.

Combining the previous bounds and (55), we can write

$$\begin{aligned} V_{\mathrm{I}}(x,\lambda)^+ - V_{\mathrm{I}}(x,\lambda) &= |\tilde{\lambda}_{\mathrm{I}}^+|^2 - |\tilde{\lambda}_{\mathrm{I}}|^2 + \gamma\beta\langle \tilde{x}^+ \mid \nabla g_{\mathrm{I}}(x^+)\tilde{\lambda}_{\mathrm{I}}^+\rangle - \gamma\beta\langle \tilde{x} \mid \nabla g_{\mathrm{I}}(x)\tilde{\lambda}_{\mathrm{I}}\rangle \\ &\leq \sum_{i=r_a+1}^{r} \max\{-|\lambda_i|^2, \gamma g_i(x)\lambda_i\} + \gamma\beta\langle g_{\mathrm{I}}(x) - g_{\mathrm{I}}(x^\star) \mid \lambda_{\mathrm{I}}^+ - \lambda_{\mathrm{I}}\rangle \\ &\quad + \gamma^2\frac{\beta}{2}(k_4k_5 + k_2k_6)\,|\tilde{x}|^2 + \gamma^2\beta\delta_1\frac{k_2k_6}{2}|\tilde{\lambda}_{\mathrm{A}}|^2 \\ &\quad + \gamma^2\frac{\beta}{2}\left(k_4k_5 + k_2k_6\frac{1+\delta_1+\delta_1^2}{\delta_1}\right)|\lambda_{\mathrm{I}}|^2, \end{aligned} \tag{56}$$

for all $(x, \lambda) \in \Omega_\rho^{\leq\varepsilon}$.

4.5.3 Bounding $V(x, \lambda)^+$ on $\Omega_\rho^{\leq\varepsilon}$

Finally, we can merge the bounds (54) and (56) derived in previous Sects. 4.5.1 and 4.5.2 to obtain from (35) the following bound for V

$$\begin{aligned} V(x,\lambda)^+ \leq V(x,\lambda) &+ \sum_{i=r_a+1}^{r} \max\{-|\lambda_i|^2, \gamma g_i(x)\lambda_i\} \\ &- 2\gamma\langle g_{\mathrm{I}}(x) - g_{\mathrm{I}}(x^\star) \mid \lambda_{\mathrm{I}}\rangle + \gamma\beta\langle g_{\mathrm{I}}(x) - g_{\mathrm{I}}(x^\star) \mid \lambda_{\mathrm{I}}^+ - \lambda_{\mathrm{I}}\rangle \\ &+ \left(\gamma^2\alpha_7 + \gamma^4\alpha_8 - \gamma c_0\right)|\tilde{x}|^2 + \gamma^2\alpha_{11}|\lambda_{\mathrm{I}}|^2 \\ &+ \left(\gamma^2\left(\delta_1\alpha_9 - \frac{k_2^2}{2}\right) + \gamma^4\alpha_{10}\right)|\tilde{\lambda}_{\mathrm{A}}|^2, \end{aligned} \tag{57}$$

for all $(x, \lambda) \in \Omega_\rho^{\leq\varepsilon}$, in which $\alpha_7, \alpha_8, \alpha_9, \alpha_{10}, \alpha_{11}$ are defined in (15).

Grouping all terms involving λ_i for $i \in \mathrm{I}$ (recall that $\mathrm{I} = \{r_a + 1, \ldots, r\}$), we can rewrite (57) as

$$\begin{aligned} V(x^+,\lambda^+) \leq V(x,\lambda) + \sum_{i=r_a+1}^{r} \Delta_i \;+\; \left(\gamma^2\alpha_7 + \gamma^4\alpha_8 - \gamma c_0\right)|\tilde{x}|^2 \\ + \left(\gamma^2\left(\delta_1\alpha_9 - \frac{k_2^2}{2}\right) + \gamma^4\alpha_{10}\right)|\tilde{\lambda}_{\mathrm{A}}|^2, \end{aligned} \tag{58}$$

in which

$$\begin{aligned} \Delta_i := \max\{-|\lambda_i|^2, \gamma g_i(x)\lambda_i\} + \gamma^2\alpha_{11}|\lambda_i|^2 \\ - 2\gamma(g_i(x) - g_i(x^\star))\lambda_i + \gamma\beta(g_i(x) - g_i(x^\star))(\lambda_i^+ - \lambda_i). \end{aligned}$$

Next, we derive a bound for Δ_i. For each $i \in \mathrm{I}$, we two cases may occur:

C1. $-|\lambda_i|^2 \geq \gamma g_i(x)\lambda_i$, which is true if and only if $|\lambda_i| \leq \gamma |g_i(x)|$;
C2. $-|\lambda_i|^2 < \gamma g_i(x)\lambda_i$, which is true if and only if $|\lambda_i| > \gamma |g_i(x)|$.

In the first case C1, we have $\max\{-|\lambda_i|^2, \gamma g_i(x)\lambda_i\} = -|\lambda_i|^2$. Since (14b) and (55) imply $|\lambda_i^+| \leq |\lambda_i|$, and hence $|\lambda_i^+ - \lambda_i| \leq 2|\lambda_i|$, we can write

$$
\begin{aligned}
-2\gamma(g_i(x) - g_i(x^\star))\lambda_i &+ \gamma\beta(g_i(x) - g_i(x^\star))(\lambda_i^+ - \lambda_i) \\
&\leq 2(1+\beta)\gamma|g_i(x) - g_i(x^\star)||\lambda_i| \\
&\leq 2(1+\beta)k_1\gamma|\tilde{x}||\lambda_i| \leq (1+\beta)k_1\gamma\left(\delta_2|\tilde{x}|^2 + \frac{1}{\delta_2}|\lambda_i|^2\right),
\end{aligned}
$$

for all $(x, \lambda) \in \Omega_\rho^{\leq\varepsilon}$, in which δ_2 is defined in (16). The above inequality and $\gamma < \bar{\gamma}_9$ (see (19d)) lead to

$$
\begin{aligned}
C1 \implies \Delta_i &\leq \left(\gamma^2\alpha_{11} + \gamma\frac{(1+\beta)k_1}{\delta_2} - 1\right)|\lambda_i|^2 + \gamma(1+\beta)k_1\delta_2|\tilde{x}|^2 \\
&\leq -\frac{1}{2}|\lambda_i|^2 + \gamma(1+\beta)k_1\delta_2|\tilde{x}|^2.
\end{aligned} \tag{59}
$$

In the second case C2, we have

$$
\max\{-|\lambda_i|^2, \gamma g_i(x)\lambda_i\} = \gamma g_i(x)\lambda_i = \gamma g_i(x^\star)\lambda_i + \gamma(g_i(x) - g_i(x^\star))\lambda_i.
$$

Moreover, in view of (11), (17), and (18), $g_i(x^\star) = -|g_i(x^\star)| \leq -h$ for all $i \in \mathrm{I}$. Hence, using again $|\lambda_i^+ - \lambda_i| \leq 2|\lambda_i|$, and $|\lambda_i| \leq K$, and since $(x, \lambda) \in \Omega_\rho^{\leq\varepsilon} \implies |\tilde{x}| \leq \varepsilon$, we obtain

$$
\begin{aligned}
\Delta_i &= \gamma g_i(x^\star)\lambda_i - \gamma(g_i(x) - g_i(x^\star))\lambda_i + \gamma\beta(g_i(x) - g_i(x^\star))(\lambda_i^+ - \lambda_i) + \gamma^2\alpha_{11}|\lambda_i|^2 \\
&\leq \gamma\big(\gamma\alpha_{11}K + (1+2\beta)|g_i(x) - g_i(x^\star)| - h\big)|\lambda_i| \\
&\leq \gamma\big(\gamma\alpha_{11}K + (1+2\beta)k_1\varepsilon - h\big)|\lambda_i|
\end{aligned}
$$

for all $(x, \lambda) \in \Omega_\rho^{\leq\varepsilon}$. Using $\gamma < \bar{\gamma}_{10}$ (see (19d), (17) and (18)) thus yields

$$
C2 \implies \Delta_i \leq -\gamma\frac{h}{2}|\lambda_i| \leq -\gamma\frac{h}{2}|\lambda_i| + \gamma(1+\beta)k_1\delta_2|\tilde{x}|^2 \tag{60}
$$

for all $(x, \lambda) \in \Omega_\rho^{\leq\varepsilon}$.

By joining (59) and (60), we thus obtain

$$
\forall(x, \lambda) \in \Omega_\rho^{\leq\varepsilon}, \quad \Delta_i \leq \frac{1}{2}\max\left\{-|\lambda_i|^2,\ -\gamma h|\lambda_i|\right\} + \gamma(1+\beta)k_1\delta_2|\tilde{x}|^2,
$$

for all $i \in \mathrm{I}$. Finally, including the latter inequality in (58), and using $\gamma \leq \bar{\gamma}_{11}$ (see (19e)) and the definition of δ_1 and δ_2 (see (16)), we obtain

$$
V(x^+, \lambda^+) \leq V(x, \lambda) - \frac{1}{2}\sum_{i=r_a+1}^{r} \min\left\{|\lambda_i|^2,\ \gamma h|\lambda_i|\right\}
$$

$$
\begin{aligned}
&+ \Big(\gamma(r-r_a)(1+\beta)k_1\delta_2 + \gamma^2\alpha_7 + \gamma^4\alpha_8 - \gamma c_0\Big)|\tilde{x}|^2 \\
&+ \left(\gamma^2\left(\delta_1\alpha_9 - \frac{k_2^2}{2}\right) + \gamma^4\alpha_{10}\right)|\tilde{\lambda}_{\mathrm{A}}|^2 \\
&= V(x,\lambda) - \frac{1}{2}\sum_{i=r_a+1}^{r} \min\left\{|\lambda_i|^2,\ \gamma h|\lambda_i|\right\} \\
&+ \gamma\left(\gamma\alpha_7 + \gamma^3\alpha_8 - \frac{c_0}{2}\right)|\tilde{x}|^2 + \gamma^2\left(\gamma^2\alpha_{10} - \frac{3k_2^2}{8}\right)|\tilde{\lambda}_{\mathrm{A}}|^2 \\
&\le V(x,\lambda) - \frac{1}{2}\sum_{i=r_a+1}^{r} \min\left\{|\lambda_i|^2,\ \gamma h|\lambda_i|\right\} - \frac{1}{4}\gamma c_0|\tilde{x}|^2 - \frac{1}{4}\gamma^2 k_2^2|\tilde{\lambda}_{\mathrm{A}}|^2
\end{aligned}
\tag{61}
$$

for all $(x,\lambda) \in \Omega_\rho^{\le\varepsilon}$.

4.6 Equiboundedness and convergence

The lemma below summarizes the results of the previous subsections.

Lemma 8 *Suppose that Assumptions 1 and 2 hold, and let γ satisfy (20). Then*

$$
\begin{aligned}
&V(x,\lambda)^+ - V(x,\lambda) \\
&\le -\min\left\{\gamma c_0\varepsilon^2,\ \frac{1}{2}\sum_{i=r_a+1}^{r}\min\left\{|\lambda_i|^2,\ \gamma h|\lambda_i|\right\} + \frac{1}{4}\gamma c_0|\tilde{x}|^2 + \frac{1}{4}\gamma^2 k_2^2|\tilde{\lambda}_A|^2\right\},
\end{aligned}
$$

for all $(x,\lambda) \in \Omega_\rho$.

Proof The proof directly follows from (34) and (61) since $\Omega_\rho = \Omega_\rho^{>\varepsilon} \cup \Omega_\rho^{\le\varepsilon}$. □

Lemma 8 ultimately enables us to conclude that the following implications hold for all $t \ge 0$

$$
(x^t,\lambda^t) \in \Omega_\rho \implies V(x^{t+1},\lambda^{t+1}) \le V(x^t,\lambda^t) \le \rho \implies (x^{t+1},\lambda^{t+1}) \in \Omega_\rho.
$$

Since, in view of (7) and Lemma 7, we have

$$
(x_0,\lambda_0) \in \Xi_0 \subseteq \overline{\mathbb{B}}_{K_0}(x^\star,\lambda^\star) \subseteq \Omega_\rho,
$$

then we claim by induction on t that every solution of (5) originating in Ξ_0 satisfies

$$
\forall t \in \mathbb{N}, \qquad (x^t,\lambda^t) \in \Omega_\rho. \tag{62}
$$

Relation (62) implies that all the trajectories of (5) originating in Ξ_0 are equibounded. Moreover, (62), Lemma 6, and Lemma 8 imply that every solution of (5) originating in Ξ_0 satisfies $V(x^t,\lambda^t) \to 0$ and $(x^t,\lambda^t) \to (x^\star,\lambda^\star)$.

4.7 Convergence rate and exponential bound

We now conclude the proof of the theorem by establishing the claimed exponential bound. As a first step, we prove the following lemma showing that, for every $\eta > 0$, all solutions of (5) originating in Ξ_0 enter into an invariant set where $V(x, \lambda) \leq \frac{1}{2}\min\left\{\eta^2, \varepsilon^2\right\}$ in the same common time.

Lemma 9 *Suppose that Assumptions 1 and 2 hold, and let γ satisfy (20). For every $\eta > 0$, let*

$$T{:=}\frac{6(2\rho - \min\{\eta^2, \varepsilon^2\})}{\min\left\{2,\ \gamma c_0,\ \gamma^2 k_2^2\right\}\min\left\{\eta^2, \varepsilon^2\right\}}. \tag{63}$$

Then, every solution of (5) originating in Ξ_0 satisfies

$$\forall t \geq T, \qquad V(x^t, \lambda^t) \leq \frac{1}{2}\min\left\{\eta^2, \varepsilon^2\right\}.$$

Proof Fix $\eta > 0$ arbitrarily and define

$$\upsilon{:=}\frac{1}{2}\min\left\{\eta^2, \varepsilon^2\right\}.$$

Pick a solution (x, λ) of (5) originating in Ξ_0, and let $\tau \in \mathbb{N}$ be such that $V(x^t, \lambda^t) > \upsilon$ for all $t \in \mathbb{N}_{<\tau}$ and $V(x_\tau, \lambda_\tau) \leq \upsilon$. The existence of such τ is implied by the convergence of (x, λ) to $(x^\star, \lambda^\star)$ established in previous Sect. 4.6. In view of (62), Lemma 8 implies $V(x^t, \lambda^t) \leq \upsilon$ for all $t \geq \tau$. Therefore, to prove the lemma, it suffices to show that $\tau \leq T$, with T defined in (63). By contradiction, suppose $\tau > T$. Then, $V(x^t, \lambda^t) > \upsilon$ for all $t \in \mathbb{N}_{\leq T}$.

For each $t \in \mathbb{N}$, let $\mathrm{I}_1^t \subseteq \mathrm{I}$ be the set of $i \in \mathrm{I}$ such that $\lambda_i^t \leq \gamma h$. Let $I_2^t = \mathrm{I}^t \setminus \mathrm{I}_1^t$. Then, we obtain (we omit the time dependency for readability)

$$\sum_{i=r_a+1}^{r} \min\left\{|\lambda_i|^2,\ \gamma h|\lambda_i|\right\} = |\lambda_{\mathrm{I}_1}|^2 + \gamma h \sum_{i\in\mathrm{I}_2}|\lambda_i|.$$

Moreover, Lemma 8 and Lemma 6 (see (26)) imply

$$\begin{aligned}
&V(x,\lambda)^+ - V(x,\lambda)\\
&\quad\leq \max\left\{-\gamma c_0\varepsilon^2,\ -\frac{1}{2}\gamma h\sum_{i\in\mathrm{I}_2}|\lambda_i| - \frac{1}{4}\min\{2,\gamma c_0,\gamma^2 k_2^2\}\left(|\tilde{x}|^2 + |\tilde{\lambda}_{\mathrm{A}}|^2 + |\lambda_{\mathrm{I}_1}|^2\right)\right\}\\
&\quad= \max\left\{-\gamma c_0\varepsilon^2,\ -\frac{1}{2}\gamma h\sum_{i\in\mathrm{I}_2}|\lambda_i| - \frac{1}{4}\min\{2,\gamma c_0,\gamma^2 k_2^2\}\left(|(\tilde{x},\tilde{\lambda})|^2 - |\lambda_{\mathrm{I}_2}|^2\right)\right\}.
\end{aligned} \tag{64}$$

Since $(x, \lambda) \in \Omega_\rho$ implies $|\lambda_i| \leq K$ for all $i \in \mathrm{I}$, then $\gamma < \bar{\gamma}_{12}$ (see (19e)) yields

$$\frac{1}{4}\min\{2,\ \gamma c_0,\ \gamma^2 k_2^2\}|\lambda_{\mathrm{I}_2}|^2 - \frac{1}{2}\sum_{i\in\mathrm{I}_2}\gamma h|\lambda_i| \leq \sum_{i\in\mathrm{I}_2}\left(\frac{\gamma^2 k_2^2}{4}|\lambda_i|^2 - \frac{\gamma h}{2}|\lambda_i|\right)$$
$$\leq \frac{\gamma}{2}\sum_{i\in\mathrm{I}_2}\left(\gamma\frac{k_2^2 K}{2} - h\right)|\lambda_i| \leq 0.$$

Hence, from (64) we obtain

$$V(x,\lambda)^+ - V(x,\lambda) \leq \max\left\{-\gamma c_0\varepsilon^2,\ -\frac{1}{6}\min\{2, \gamma c_0, \gamma^2 k_2^2\}V(x,\lambda)\right\}.$$

Using $t \leq T \implies V(x^t, \lambda^t) > \upsilon$, we then obtain

$$\begin{aligned}
&V(x^{t+1}, \lambda^{t+1})\\
&\leq V(x^t,\lambda^t) - \min\left\{\gamma c_0\varepsilon^2, \frac{1}{12}\min\{2, \gamma c_0, \gamma^2 k_2^2\}\min\{\eta^2, \varepsilon^2\}\right\}\\
&= V(x^t,\lambda^t) - \min\left\{\gamma c_0\varepsilon^2, \frac{1}{12}\gamma c_0\varepsilon^2, \frac{1}{12}\gamma c_0\eta^2, \frac{1}{12}\min\{2, \gamma^2 k_2^2\}\min\{\eta^2, \varepsilon^2\}\right\}\\
&= V(x^t,\lambda^t) - \min\left\{\frac{1}{12}\gamma c_0\varepsilon^2, \frac{1}{12}\gamma c_0\eta^2, \frac{1}{12}\min\{2, \gamma^2 k_2^2\}\min\{\eta^2, \varepsilon^2\}\right\}\\
&= V - \frac{1}{12}\min\{2, \gamma c_0, \gamma^2 k_2^2\}\min\{\eta^2, \varepsilon^2\}.
\end{aligned}$$

Namely,

$$\forall t \in \mathbb{N}_{\leq T}, \quad V(x^{t+1}, \lambda^{t+1}) \leq V(x^t, \lambda^t) - \chi(\gamma)$$

in which

$$\begin{aligned}
\chi(\gamma) &:= \frac{1}{6}\min\left\{2,\ \gamma c_0,\ \gamma^2 k_2^2\right\},\\
\upsilon &= \frac{1}{12}\min\left\{2,\ \gamma c_0,\ \gamma^2 k_2^2\right\}\min\left\{\eta^2, \varepsilon^2\right\}.
\end{aligned}$$

As $V(x^0, \lambda^0) \leq \rho$ (by Lemma 7) we thus obtain

$$V(x^T, \lambda^T) \leq \rho - \chi(\gamma)T = \upsilon,$$

which contradicts $V(x^T, \lambda^T) > \upsilon$, so that the proof follows. □

The following lemma, instead, provides conditions for local exponential convergence.

Lemma 10 *Suppose that Assumptions 1 and 2 hold, and let γ satisfy (20). Let $\mu \in [0, 1)$ and $a \in (0, \rho)$ be such that*

$$V(x, \lambda) \le a \implies V(x^+, \lambda^+) \le \mu^2 V(x, \lambda).$$

Finally, let $T \in \mathbb{N}$ be such that every solution of (5) originating in Ξ_0 satisfies $V(x^T, \lambda^T) \le a$. Then, every solution of (5) originating in Ξ_0 also satisfies

$$\forall t \in \mathbb{N}, \quad |(\tilde{x}^t, \tilde{\lambda}^t)| \le \sqrt{3}\mu^{t-T}|(\tilde{x}^0, \tilde{\lambda}^0)|.$$

Proof Pick a solution of (5) originating in Ξ_0. Lemma 8 and (62) and imply $V(x^T, \lambda^T) \le V(x^0, \lambda^0)$. Moreover, as $a < \rho$, Lemma 8 implies $V(x^t, \lambda^t) \le V(x^T, \lambda^T) \le a$ for all $t \ge T$.

Hence, in view of Lemma 6, we obtain

$$\begin{aligned} &\forall t \ge T, \quad V(x^t, \lambda^t) \le \mu^{2(t-T)} V(x^T, \lambda^T) \le \mu^{2(t-T)} V(x^0, \lambda^0) \le \mu^{2(t-T)} \frac{3}{2} |(\tilde{x}^0, \tilde{\lambda}^0)|^2 \\ &\implies \forall t \ge T, \quad |(\tilde{x}^t, \tilde{\lambda}^t)|^2 \le 2V(x^t, \lambda^t) \le 3\mu^{2(t-T)} |(\tilde{x}^0, \tilde{\lambda}^0)|^2. \end{aligned}$$

Instead, for $t \le T$, one has

$$|(\tilde{x}^t, \tilde{\lambda}^t)|^2 \le 2V(x^t, \lambda^t) \le 2V(x^0, \lambda^0) \le 3V(x^0, \lambda^0) \le 3\mu^{2(t-T)} V(x^0, \lambda^0),$$

where we used the fact that, since $\mu \in [0, 1)$, then $\mu^{2(t-T)} \ge 1$ for all $t \le T$. □

With Lemmas 9 and 10 at hand, we can now prove the claimed exponential bound. First, assume that

$$V(x, \lambda) \le a := \frac{1}{2} \min\left\{\varepsilon^2, \ \gamma^2 h^2, \ 2\rho\right\}. \tag{65}$$

Using $|\tilde{x}|^2 \le |(\tilde{x}, \tilde{\lambda})|^2 \le 2V(x, \lambda)$ and $|\lambda_i|^2 \le |(\tilde{x}, \tilde{\lambda})|^2 \le 2V(x, \lambda)$ for all $i \in \mathrm{I}$ (in view of Lemma 6), we get that (65) implies

$$\begin{aligned} &(x, \lambda) \in \Omega_\rho^{\le \varepsilon}, \\ &\forall i \in \mathrm{I}, \quad |\lambda_i| \le \gamma h, \end{aligned}$$

and, hence,

$$\forall i \in \mathrm{I}, \quad \min\left\{|\lambda_i|^2, \ \gamma h |\lambda_i|\right\} = |\lambda_i|^2.$$

Then, we can manipulate (61) exploiting Lemma 6 to assert that, if (65) holds, then

$$\begin{aligned} V(x^+, \lambda^+) &\le V(x, \lambda) - \frac{1}{2}|\lambda_{\mathrm{I}}|^2 - \frac{1}{4}\gamma c_0 |\tilde{x}|^2 - \frac{1}{4}\gamma^2 k_2^2 |\tilde{\lambda}_{\mathrm{A}}|^2 \\ &\le V(x, \lambda) - \frac{1}{4} \min\left\{2, \ \gamma c_0, \ \gamma^2 k_2^2\right\} \left(|\tilde{x}|^2 + |\tilde{\lambda}|^2\right) \end{aligned}$$

$$\le \left(1 - \frac{1}{6}\min\left\{2,\ \gamma c_0,\ \gamma^2 k_2^2\right\}\right) V(x,\lambda)$$
$$= \mu^2 V(x,\lambda) \tag{66}$$

with

$$\mu{:=}\sqrt{1 - \frac{1}{6}\min\left\{2,\ \gamma c_0,\ \gamma^2 k_2^2\right\}} \in [0, 1).$$

Thus, we have established the implication

$$V(x,\lambda) \le a \implies V(x^+,\lambda^+) \le \mu^2 V(x,\lambda) \tag{67}$$

$a \in (0, \rho)$ defined in (65).

Next, we apply Lemma 9 with

$$\eta{:=}\gamma h,$$

obtaining that every solution of (5) originating in Ξ_0 satisfies

$$\forall t \ge T, \quad V(x^t, \lambda^t) \le a \tag{68}$$

in which T has the expression (63) with $\eta = \gamma h$.

The claim of the theorem finally follows from Lemma 10 in view of (65), (66), and (68).

5 Conclusions

This article considered the long-standing open problem of nonlocal asymptotic stability of the popular discrete-time primal-dual algorithm (1) for convex, constrained optimization. In particular, under due convexity and regularity assumptions, it is proved that an optimal equilibrium exists, it is unique, and it is semiglobally asymptotically stable. Namely, for every compact set of initial conditions, there exists a sufficiently small stepsize, such that the sequences generated by the algorithm converge to the optimal solution of the optimization problem and to the optimal Lagrange multipliers. Moreover, convergence is exponential, and the optimal point is Lyapunov stable. As shown in Sect. 1.2, global asymptotic stability cannot be established for the considered algorithm, so as semiglobal guarantees are the best achievable in the general case.

The key idea inspiring the stability analysis pursued in the article was to look at Algorithm (1) as a discrete-time dynamical system sharing many similarities with a nonlinear oscillator. This motivated the usage of a non-trivial Lyapunov function with a suitably-defined cross-term, unlike Uzawa's previous attempt in [29].

Finally, it is worth remarking that the impossibility of global convergence of the algorithm complicates the development of robustness corollaries, which cannot be global in the size of the uncertainty. As a consequence, such shortfall poses new challenges in the design of distributed algorithms based on (1) and targeting, e.g., consensus optimization problems over networks [20–22, 24]. Future research will mainly focus on this latter extension.

Funding Open access funding provided by Alma Mater Studiorum - Università di Bologna within the CRUI-CARE Agreement.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

A Proof of Lemma 1

First, notice that, since by 1-B each g_i is continuous, then $G_i := \{x \in \mathbb{R}^n \ : \ g_i(x) \leq 0\}$ is closed for all $i = 1, \dots, r$. Hence, $G := \cap_{i=1,\dots,r} G_i$ is closed. Condition 1-C further implies that G is nonempty.

Since, by 1-A, f is continuously differentiable and strongly convex, it has a global minimum $z \in \mathbb{R}^n$ at which $\nabla f(z) = 0$, and there exists $m > 0$ such that

$$\forall w, y \in \mathbb{R}^n, \qquad f(y) \geq f(w) + \nabla f(w)(y - w) + m|y - w|^2. \tag{69}$$

In particular, (69) implies that, for all $c > f(z)$, $f^{-1}([f(z), c])$ is compact.

Let $(c_n)_{n\in\mathbb{N}}$ be an increasing sequence satisfying $c_n \to \infty$. Then, (69) implies that

$$\bigcup_{n\in\mathbb{N}} f^{-1}([f(z), c_n]) = \mathbb{R}^n. \tag{70}$$

Indeed, pick $x \in \mathbb{R}^n$. Since $c_n \to \infty$, we can find $n_x \in \mathbb{N}$ sufficiently large so that $c_{n_x} \geq f(z) + |\nabla f(x)||x - z|$. Then, (69) applied with $w = x$ and $y = z$ implies $f(x) \leq c_{n_x}$, i.e., $x \in f^{-1}([f(z), c_{n_x}])$. In view of this, we can assume without loss of generality that, for all $n \in \mathbb{N}$, $\varnothing \neq A_n := f^{-1}([f(z), c_n]) \cap G$.

Since f is continuous and A_n is compact and nonempty, for each $n \in \mathbb{N}$ there exists $x_n \in A_n$ satisfying $x_n = \operatorname{argmin}_{x\in A_n} f(x)$.

Since $c_{n+1} \geq c_n$, then $A_n \subseteq A_{n+1}$ and, hence, $f(x_{n+1}) \leq f(x_n)$. The sequence $(f(x_n))_{n\in\mathbb{N}}$ is therefore decreasing and lower-bounded by $f(z)$. Hence, it has a limit $f_\infty := \lim_{n\to\infty} f(x_n) = \inf_{n\in\mathbb{N}} f(x_n)$.

By using (69) with $(w, y) = (z, x_n)$, we obtain

$$\forall n \in \mathbb{N}, \quad m|x_n - z|^2 \leq f(x_n) - f(z) \leq f(x_0) - f(z).$$

Hence, the sequence x_n has a converging subsequence that converges to some $x^\star \in \mathbb{R}^n$ satisfying $f(x^\star) = f_\infty$.

Since G is closed and $x_n \in A_n \subseteq G$ for all $n \in \mathbb{N}$, then $x^\star \in G$. Thus, $x^\star$ it is feasible for (2).

In view of (70), for every $x \in G$, there exists $\bar{n} \in \mathbb{N}$ such that $x \in A_{\bar{n}}$. Hence, $f(x) \geq f(x_{\bar{n}}) \geq f_\infty = f(x^\star)$. Since $x \in G$ was arbitrary, this shows that $x^\star$ is a minimum of f in G.

Finally, the uniqueness of $x^\star$ follows by the fact that f is strongly convex and G is a convex set since g_i is convex for all $i = 1, \ldots, r$. Indeed, suppose $\bar{x} \in G$ is another point satisfying $f(\bar{x}) = f(x^\star) = f_\infty$. Then, $|\bar{x} - x^\star| > 0$ and, since G is convex, for every $t \in (0, 1)$, $t\bar{x} + (1-t)x^\star \in G$. Since f is strongly convex then this implies that, for some $m > 0$,

$$f(t\bar{x} + (1-t)x^\star) \leq \underbrace{tf(\bar{x}) + (1-t)f(x^\star)}_{=f_\infty} - mt(1-t)|\bar{x} - x^\star|^2 < f_\infty,$$

which violates the fact shown earlier that $f(x) \geq f_\infty$ for all $x \in G$.

References

1. Arrow, K.J., Hurwicz, L., Uzawa, H.: Studies in Linear and Non-linear Programming. Stanford University Press, Redwood City (1958)
2. Bertsekas, D.P.: Nonlinear Programming. Athena Scientific, Nashua (1999)
3. Chambolle, A., Pock, T.: A first-order primal-dual algorithm for convex problems with applications to imaging. J. Math. Imaging Vis. **40**(1), 120–145 (2011)
4. Chen, J., Lau, V.K.: Convergence analysis of saddle point problems in time varying wireless systems-control theoretical approach. IEEE Trans. Signal Process. **60**(1), 443–452 (2011)
5. Cherukuri, A., Gharesifard, B., Cortes, J.: Saddle-point dynamics: conditions for asymptotic stability of saddle points. SIAM J. Control Optim. **55**(1), 486–511 (2017)
6. Cherukuri, A., Mallada, E., Cortés, J.: Asymptotic convergence of constrained primal-dual dynamics. Syst. Control Lett. **87**, 10–15 (2016)
7. Cherukuri, A., Mallada, E., Low, S., Cortés, J.: The role of convexity in saddle-point dynamics: Lyapunov function and robustness. IEEE Trans. Autom. Control **63**(8), 2449–2464 (2017)
8. Dürr, H.B., Zeng, C., Ebenbauer, C.: Saddle point seeking for convex optimization problems. In: 9th IFAC Symposium on Nonlinear Control Systems, vol. 46, pp. 540–545 (2013)
9. Feijer, D., Paganini, F.: Stability of primal-dual gradient dynamics and applications to network optimization. Automatica **46**(12), 1974–1981 (2010)
10. Goldsztajn, D., Paganini, F.: Proximal regularization for the saddle point gradient dynamics. IEEE Trans. Autom. Control **66**(9), 4385–4392 (2020)
11. Holding, T., Lestas, I.: Stability and instability in saddle point dynamics-part i. IEEE Trans. Autom. Control **66**(7), 2933–2944 (2020)
12. Holding, T., Lestas, I.: Stability and instability in saddle point dynamics-part ii: the subgradient method. IEEE Trans. Autom. Control **66**(7), 2945–2960 (2020)
13. Hu, G., Pang, Y., Sun, C., Hong, Y.: Distributed Nash equilibrium seeking: continuous-time control-theoretic approaches. IEEE Control Syst. Mag. **42**(4), 68–86 (2022)
14. Isidori, A.: Nonlinear Control Systems, 3rd edn. Springer-Verlag, London (1995)

15. Kallio, M., Rosa, C.H.: Large-scale convex optimization via saddle point computation. Oper. Res. **47**(1), 93–101 (1999)
16. Kallio, M., Ruszczynski, A.: Perturbation methods for saddle point computation. Tech. Rep, IIASA (1994)
17. Khalil, H.K.: Nonlinear Systems, 3rd edn. Pearson, London (2002)
18. Kose, T.: Solutions of saddle value problems by differential equations. Econometrica J. Econom. Soc. **24**, 59–70 (1956)
19. Lyapunov, A.M.: The general problem of the stability of motion. Ph.D. thesis, University of Kharkov (1892)
20. Nedić, A.: Convergence rate of distributed averaging dynamics and optimization in networks. Found. Trends® Syst. Control **2**(1), 1–100 (2015)
21. Nedić, A., Liu, J.: Distributed optimization for control. Annu. Rev. Control Robot. Auton. Syst. **1**, 77–103 (2018)
22. Nedić, A., Olshevsky, A., Rabbat, M.G.: Network topology and communication-computation tradeoffs in decentralized optimization. Proc. IEEE **106**(5), 953–976 (2018)
23. Nedić, A., Ozdaglar, A.: Subgradient methods for saddle-point problems. J. Optim. Theory Appl. **142**(1), 205–228 (2009)
24. Notarstefano, G., Notarnicola, I., Camisa, A.: Distributed optimization for smart cyber-physical networks. Found. Trends® Syst. Control **7**(3), 253–383 (2019)
25. Polyak, B.: Iterative methods using Lagrange multipliers for solving extremal problems with constraints of the equation type. USSR Comput. Math. Math. Phys. **10**(5), 42–52 (1970)
26. Qu, G., Li, N.: On the exponential stability of primal-dual gradient dynamics. IEEE Control Syst. Lett. **3**(1), 43–48 (2018)
27. Stetter, H.J.: Analysis of Discretization Methods for Ordinary Differential Equations. Springer, Berlin (1973)
28. Tang, Y., Qu, G., Li, N.: Semi-global exponential stability of augmented primal-dual gradient dynamics for constrained convex optimization. Syst. Control Lett. **144**, 104754 (2020)
29. Uzawa, H.: Iterative methods for concave programming. Stud. Linear Non Linear Program. **6**, 154–165 (1958)
30. Zhang, J., Zheng, D., Chiang, M.: The impact of stochastic noisy feedback on distributed network utility maximization. IEEE Trans. Inf. Theory **54**(2), 645–665 (2008)