Quantitative Sensitivity Bounds for Nonlinear Programming and Time-varying Optimization

Irina Subotić, Student Member, IEEE, Adrian Hauswirth, Student Member, IEEE, and Florian Dörfler, Member, IEEE

Abstract—Inspired by classical sensitivity results for nonlinear optimization, we derive and discuss new quantitative bounds to characterize the solution map and dual variables of a parametrized nonlinear program. In particular, we derive explicit expressions for the local and global Lipschitz constants of the solution map of non-convex or convex optimization problems, respectively. Our results are geared towards the study of time-varying optimization problems which are commonplace in various applications of online optimization, including power systems, robotics, signal processing and more. In this context, our results can be used to bound the rate of change of the optimizer. To illustrate the use of our sensitivity bounds we generalize existing results for time-varying optimization problems which are commonplace in various applications of online optimization, including power systems, robotics, signal processing and more. Roughly speaking, for these problems we can distinguish between running algorithms [4], [18], that do not anticipate the change in an optimizer, as opposed to predictive schemes that exploit some information about the change of the optimization problem [3], [19]–[21].

Depending on the application, the optimization problem might be unconstrained with a time-varying cost function [18], [22], constrained to a stationary set but with time-varying cost function [4], [23], or have time-varying constraints as well [24]. If constraints are time-varying it is possible to eliminate equality constraints or apply barrier functions [20], [25] to the objective to arrive at an unconstrained problem. All of these setups require the solution of the time-varying problem to be a well-behaved function of time. In other words, an optimizer should be isolated, if not unique, and vary slowly so that an online algorithm is able to approximately track it. We refer to the asymptotic distance between the true optimizer and the algorithm state (or iterate) as tracking performance.

To analyze the tracking performance, most works have focused on convex problems, with strongly convex objectives that are guaranteed to admit a unique global optimizer. Moreover, many works establish tracking performance under the assumption that the rate of change of the optimizer is bounded. However, apart from special cases, it is unclear if and how this assumption is satisfied, and which rates are achievable.

Therefore, in this paper, we study how the rate of change of optimizers can be bounded for a general class of nonlinear optimization problems. More precisely, using the classical sensitivity results for parametrized, nonlinear optimization problems [26]–[29] we establish sufficient conditions under which the instantaneous optimizer and its dual variables are continuous functions of a perturbation (not necessarily time), with a quantifiable sensitivity. Moreover, we provide and discuss special cases of constrained parametrized optimization problems for which the sensitivities of optimizer and dual variables take a simple and easy-to-interpret form.

As a second contribution, in view of the renewed interest in continuous-time optimization [30]–[32], we show the usefulness of our sensitivity results by quantifying the tracking performance of continuous-time running algorithms for time-varying optimization. In particular, we generalize and formalize a monotonicity-like property that can be used to provide asymptotic bounds on the tracking error of, almost arbitrary, well-defined continuous-time schemes.

Finally, as a third contribution, we illustrate our results by introducing a novel continuous-time algorithm, termed sweeping gradient flow and modeled as a perturbed sweeping process [33], [34], to track the solution of a convex optimization problem with time-varying objective and constraints. The proposed scheme is inherently discontinuous and defined by a differential inclusion. However, the scheme is of conceptual importance since it can be interpreted as the continuous-time limit of an online projected gradient descent [4], [23]. Moreover, sweeping processes of this type, which essentially model a continuous projection onto a time-varying set, have been considered in feedback optimization [5]–[7] where constraint enforcement may rely on input saturation in physical plants.

The paper is structured as follows: In Section II we fix the notation and recall standard optimality conditions for nonlinear programming, as well as results from sensitivity analysis. Section III deals with quantifying the local differentiability and Lipschitz continuity of the solution maps for nonlinear programs. Then, in Section IV we introduce and discuss...
assumptions that allow for global statements about solution maps and their sensitivities. These findings are discussed in Section V using illustrative special cases. Section VI provides a general result to quantify the tracking performance of continuous-time algorithms under a monotonicity-like property. Next, Section VII introduces a “sweeping gradient flow” as a continuous-time scheme for the time-varying optimization and discusses its tracking guarantees for a special class of convex problems. Finally, Section VIII provides illustrative examples validating our theoretical results. Finally, Section IX summarizes our findings and discusses open problems.

II. PRELIMINARY

We consider $\mathbb{R}^n$ with the usual inner product $(\cdot, \cdot)$ and 2-norm $\|\cdot\|$. The non-negative orthant is denoted by $\mathbb{R}^n_0$. The identity and zero matrices of appropriate dimension are written as $I$ and $0$, respectively. For a vector $v \in \mathbb{R}^n$ and an index set $I \subset \{1, \ldots, n\}$, $v_I \in \mathbb{R}^{|I|}$ denotes the vector corresponding to the $I$-th components of $v$. If $A \in \mathbb{R}^{m \times n}$ is a matrix, $A_I$ denotes the submatrix consisting of the $I$-th rows of $A$. Furthermore, $A^T$ denotes its transpose of $A$ and $\|A\| := \sup_{\|x\|=1} \|Ax\|/\|x\|$ its operator norm. If $A \in \mathbb{R}^{m \times n}$ is symmetric, $\lambda_{A}^{\max}$ and $\lambda_{A}^{\min}$ denote the maximum and minimum eigenvalues of $A$, respectively. We write $A > 0$ ($A \succeq 0$) if $A$ is positive (semi) definite. For a matrix $B \in \mathbb{R}^{m \times m}$, $\sigma_B^{\max}$ and $\sigma_B^{\min}$ are the maximum and minimum singular values of $B$, respectively. Namely, we have $\sigma_B^{\max} = \|B\|$.

A map $F : \mathcal{X} \to \mathbb{R}^n$ with $\mathcal{X} \subset \mathbb{R}^n$ is $\ell$-Lipschitz if

$$\|F(y) - F(x)\| \leq \ell \|y - x\| \quad (1)$$

holds for all $y, x \in \mathcal{X}$, and $F$ is locally $\ell'$-Lipschitz at $x' \in \mathcal{X}$ if for every $\ell' > 0$ there exists a (relative) neighborhood $N_{x'}$ of $x'$ such that $\{1\}$ holds for all $x, y \in N_{x'}$. In other words, $\ell'$ is the largest lower bound on $\ell$ such that $\{1\}$ is satisfied, i.e., $\ell' = \sup_{x' \to x} \|F(y) - F(x)\|/\|y - x\|$.

Derivatives are understood in the sense of Fréchet, and we denote the Jacobian matrix of $F$ at $x$ by $\nabla F(x) \in \mathbb{R}^{n \times n}$. If $F$ is differentiable at $x$ then it is locally $\ell$-Lipschitz at $x$ with $\ell = \|\nabla F(x)\|$. The Jacobian at $x$, with respect to a subset of variables $x_1 \in \mathcal{X}^{n_1}$ is denoted by $\nabla_{x_1} F(x) \in \mathbb{R}^{m \times n_1}$.

If $m = 1$, we call the column vector $\nabla F(x) := \nabla F(x)^T$ the gradient of $F$, and $\nabla^2_{xx} F(x) \in \mathbb{R}^{n \times n}$ denotes the Hessian of $F$ at $x$. However, when differentiating $F$ twice, with respect to different subsets of variables $x_1 \in \mathcal{X}^{n_1}$ and $x_2 \in \mathcal{X}^{n_2}$, we define $\nabla^2_{x_1 x_2} F(x) := \nabla^2_{x_1} (\nabla_{x_2} F(x)) \in \mathbb{R}^{n_1 \times n_2}$, in particular, we have $\nabla^2_{x_1 x_2} F(x) = (\nabla^2_{x_2 x_1} F(x))^T$.

A. Sensitivity Analysis for Nonlinear Optimization

Throughout the paper, we consider the following parametrized nonlinear optimization problem and special cases thereof:

$$\begin{aligned}
& \text{minimize} \\
& \quad \quad f(x, \xi) \\
& \text{subject to} \\
& \quad \quad h(x, \xi) = 0 \quad (2) \\
& \quad \quad g(x, \xi) \leq 0,
\end{aligned}$$

where $x \in \mathbb{R}^n$ is the decision variable, $\xi \in \Xi \subset \mathbb{R}^r$ is a parameter. Further, the objective $f : \mathbb{R}^n \times \Xi \to \mathbb{R}$ and the constraint functions $h : \mathbb{R}^n \times \Xi \to \mathbb{R}$ and $g : \mathbb{R}^n \times \Xi \to \mathbb{R}^m$ are twice continuously differentiable in $(x, \xi)$. Further, we define the parameterized feasible set $\mathcal{X}(\xi) := \{x \in \mathbb{R}^n | h(x, \xi) = 0, g(x, \xi) \leq 0\}$. Given $x \in \mathcal{X}(\xi)$, let $I^f_1 := \{i | g_i(x, \xi) = 0\}$ denote the set of active inequality constraints at $x$ for $\xi \in \Xi$ and $I^f_2 := \{i | g_i(x, \xi) < 0\}$ the set of inactive inequality constraints at $x \in \Xi$.

The following standard definitions and results closely follow [26], [29], [35], but they can also be found in any good textbook on nonlinear optimization, e.g., [36], [37].

Definition 1 (LICQ). Given $\xi \in \Xi$, and $x \in \mathcal{X}(\xi)$, the linear independence constraint qualification (LICQ) holds at $x$ if the matrix $[\nabla_x h(x, \xi)^T \nabla_x g_{I^f_1}(x, \xi)^T]$ has full rank.

The (parametrized) Lagrangian of (2) is defined as $L(x, \lambda, \mu, \xi) := f(x, \xi) + \lambda^T h(x, \xi) + \mu^T g(x, \xi)$ for all $x \in \mathbb{R}^n$, all Lagrange multipliers $\mu \in \mathbb{R}^{m_0}$, and all $\xi \in \Xi$.

Definition 2 (KKT). Given $\xi$ for some $\xi \in \Xi$, the Karush-Kuhn-Tucker (KKT; first-order optimality) conditions hold at $(x^*, \lambda^*, \mu^*) \in \mathcal{X}(\xi) \times \mathbb{R}^m \times \mathbb{R}^{m_0}_0$ if $\nabla_x L(x^*, \lambda^*, \mu^*, \xi) = 0$ and $\mu^*_i = 0$ holds for all $i \in I^f_2$.

Under LICQ (but also under weaker constraint qualifications, such as the Mangasarian-Fromowitz CQ), every optimizer is a KKT point [36 Th. 4.2.13]. However, the uniqueness of $\mu^*$ holds only under LICQ.

Theorem 1. Given $\xi \in \Xi$, if $x^*$ is a (local) solution of (2) and LICQ holds at $x^*$, then there exists a unique $(\lambda^*, \mu^*) \in \mathbb{R}^m \times \mathbb{R}^{m_0}_0$ such that the KKT conditions are satisfied at $(x^*, \lambda^*, \mu^*)$.

Theorem 2. Given $\xi \in \Xi$, if $(x^*, \lambda^*, \mu^*)$ satisfies LICQ at $x^*$, KKT, and SSOSC, then $x^*$ is a strict local minimizer of (2).

Given $\xi \in \Xi$, we say that $(x^*, \lambda^*, \mu^*) \in \mathcal{X}(\xi) \times \mathbb{R}^m \times \mathbb{R}^{m_0}_0$ is a regular minimizer of (2) if LICQ is satisfied at $x^*$ and $(x^*, \lambda^*, \mu^*)$ satisfies KKT and SSOSC. For regular optimizers, the following sensitivity result holds [35 Thm. 2.3.3]:

Theorem 3. Consider (2) and let $f, g, h$ be twice continuously differentiable in $(x, \xi)$. Further, let $\tilde{x}$ be a regular minimizer of (2) for $\xi$ with multiplier $(\lambda^*, \mu^*)$. Then, on a neighborhood $N \subset \Xi$ of $\xi$, there exist continuous maps $x^* : N \to \mathbb{R}^n$, $\lambda^* : N \to \mathbb{R}^m$, and $\mu^* : N \to \mathbb{R}^{m_0}_0$ such that

$(i) \; x^*(\xi) = \tilde{x}, \; \lambda^*(\xi) = \tilde{\lambda}$, and $\mu^*(\xi) = \tilde{\mu}$. 

1As a slightly less stringent condition for (i, (ii), and (iii)) to hold, we require only that $f$ and $g$ are twice continuously differentiable in $x$ and that $f$, $g$, $h$, $\nabla_x f$, $\nabla_g h$, $\nabla_{xx} f$, $\nabla_{xx} g$ and $\nabla_x^2 h$ are continuous in $\xi$ [35 Thm.2.3.2].
(ii) for all $\xi \in \mathcal{N}$, LICQ is satisfied for $x^*(\xi)$, and KKT and SSOSC hold for $(x^*(\xi), \lambda^*(\xi), \mu^*(\xi))$.

(iii) for all $\xi \in \mathcal{N}$, $x^*(\xi)$ is a local minimizer for $\{, \mu^*(\xi)\}$ and $(\lambda^*(\xi), \mu^*(\xi))$ is the corresponding Lagrange multiplier.

(iv) $x^*$, $\lambda^*$, and $\mu^*$ are locally Lipschitz at $\xi$.

In particular, note that $x^*$ and $\mu^*$ are feasible on $\mathcal{N}$, i.e., $x^*(\xi) \in \mathcal{X}(\xi)$ and $\mu^*(\xi) \in \mathbb{R}^M_{\geq 0}$ hold for all $\xi \in \mathcal{N}$.

### III. Quantifying the Lipschitz Continuity of Solution Maps

Theorem 3 and similar results in [20, 23, 29] guarantee the Lipschitz continuity of the solution map $x^*$, however, they do not quantify the Lipschitz constants of $x^*$ and $\mu^*$.

Hence, in this section, we refine these results and give quantitative bounds on the rate of change of $x^*$. The key insight for this purpose, also used in [29], is that a KKT point $(x^*, \mu^*) \in \mathcal{X}(\xi) \times \mathbb{R}^m$ solves the system

$$ F(x^*, \mu^*, \xi) := \begin{bmatrix} \nabla_x L(x^*, \mu^*, \xi) \\ h(x^*, \xi) \\ \text{diag}(\mu^*)g(x^*, \xi) \end{bmatrix} = 0 \in \mathbb{R}^{n+p+m}. \tag{3} $$

Under additional assumptions, the implicit function theorem can be applied to (3) to express $(x^*, \mu^*)$ as a function of $\xi$.

Throughout the remainder of this section, we consider the same setup as in Theorem 3. Hence, existence and local Lipschitz continuity of $x^*: \mathcal{N} \to \mathbb{R}^n$ and $\mu^*: \mathcal{N} \to \mathbb{R}^m$ for some neighborhood $\mathcal{N} \subset \Xi$ around $\xi$, for which $(\bar{x}, \bar{\mu})$ is a regular optimizer, are guaranteed. Further, we use the shorthand notations $\mathbf{1}(\xi) := \mathbf{1}_\mathcal{X}(\xi)$, $\mathbf{1}(\xi) := \mathbf{1}_\mathcal{X}(\xi)$, and, for all $\xi \in \mathcal{N}$, we define

$$ A(\xi) := \nabla^2_{xx} L(x^*(\xi), \lambda^*(\xi), \mu^*(\xi), \xi) \in \mathbb{R}^{n \times n} , $$

$$ B(\xi) := \begin{bmatrix} \nabla_x h(x^*(\xi), \xi) \\ \nabla_\theta \Theta_1(x^*(\xi), \xi) \end{bmatrix} \in \mathbb{R}^{n+|\mathcal{I}^*| \times n} , $$

$$ L^*(\xi) := \nabla^2_{xx} L(x^*(\xi), \lambda^*(\xi), \mu^*(\xi), \xi) \in \mathbb{R}^{n \times r} , $$

$$ G^*(\xi) := \begin{bmatrix} \nabla_\theta h(x^*(\xi), \xi) \\ \nabla_\theta \Theta_1(x^*(\xi), \xi) \end{bmatrix} \in \mathbb{R}^{n \times |\mathcal{I}^*| \times r} . $$

If (2) does not have any equality constraints and $\mathbf{1}(\xi) = 0$, then we follow the convention that $B(\xi) = 0$ and $G^*(\xi) = 0$.

In the following, unless there is any ambiguity, we drop the argument from any map whose sole argument is $\xi$.

Finally, we make one simplifying assumption that can, presumably, be slightly relaxed in future work:

**Assumption 1.** In the setup of Theorem 3 the matrix $A(\xi)$ is positive definite for all $\xi \in \Xi$.

In particular, Assumption 1 is satisfied for convex optimization problems with strongly convex objective – though one can perceive also scenarios with weaker regularity assumptions. Furthermore, the assumption is common in multi-parametric programming, when the the KKT system needs to be solved explicitly (e.g., [38]). This assumption will be made in forthcoming sections to provide global bounds on the Lipschitz constants of $x^*$ and $(\lambda^*, \mu^*)$.

To quantify the Lipschitz constant of the solution map we have to distinguish between two cases related to constraints becoming active or inactive when varying $\xi$. Namely, if every active constraint is associated with a positive multiplier (which we refer to as strict complementarity below), these constraints remain active for small variations in $\xi$. In this case, the solution maps $x^*$ and $\mu^*$ are differentiable. However, when a constraint is active but its multiplier is zero, then differentiability is in general not guaranteed. In this case, we need to make a careful case distinction considering the constraint as being either active or inactive.

#### A. Sensitivity Analysis under Strict Complementarity

Given a KKT-point $(x^*, \lambda^*, \mu^*)$ of (2) for a given $\xi \in \Xi$, we define the sets of strongly active inequality constraints as

$$ \mathbf{1}(x^*, \lambda^*, \mu^*) := \{ i \in \mathcal{I}, | \mu_i^* > 0 \} $$

and we say that $(x^*, \lambda^*, \mu^*)$ satisfies strict complementary slackness (SCS) if $\mathbf{1}(x^*, \lambda^*, \mu^*) = \mathcal{I}^*$, i.e., if all active inequality constraints are strongly active.

Under strict complementary slackness, the solution maps $x^*$ and $(\lambda^*, \mu^*)$ are continuously differentiable in a neighborhood of $\xi$. The following result gives an explicit expression for the respective derivatives.

**Proposition 4.** Consider the same setup as in Theorem 3 and let SCS and Assumption 1 hold at $(x^*, \lambda^*, \mu^*)$. Then, in a neighborhood $\mathcal{N} \subset \Xi$ of $\xi$, the maps $x^*: \mathcal{N} \to \mathbb{R}^n$ and $(\lambda^*, \mu^*): \mathcal{N} \to \mathbb{R}^p \times \mathcal{M} \to \mathbb{R}^{M_{\geq 0}}$ are continuously differentiable with

$$ \nabla_\xi x^* = -\Pi A^{-1} L^* - S \Sigma G^* \in \mathbb{R}^{n \times r} , $$

$$ \nabla_\xi (\lambda^*) = B^\top \Sigma (A^{-1} L^* - B^\top G^*) \in \mathbb{R}^{n \times |\mathcal{I}^*| \times r} \tag{4} $$

and $\nabla_\xi \mu_i^* = 0$, where $B^\top := B^\top (B A^{-1} B^\top)^{-1} \in \mathbb{R}^{n \times (|\mathcal{I}^*|)}$ is the pseudoinverse of $B$, and $\Sigma, \Pi \in \mathbb{R}^{n \times n}$ are given by

$$ A := A^{-1} B^\top (B A^{-1} B^\top)^{-1} \text{ and } \Pi := \Pi - \Sigma. $$

If (2) has only equality constraints and none of them are active at $\xi$, then we have $B^\top G^* = 0$.

**Proof.** We follow the previous procedure as in [26, 35]. However, Assumption 1 the matrix inversion Lemma 5 in the appendix, and some attention to details allow us to derive the explicit expressions [4] for the derivative of the solution map.

For the moment let $z^* = (x^*, \lambda^*, \mu^*)$. The implicit function theorem [39, Ch. 9] applied to (3) states that $\nabla_\xi z^* = - (\nabla_\xi F)^{-1} \nabla_\xi F$. Hence, without loss of generality, assume that the first $|\mathcal{I}^*|$ constraints are (strongly) active and, by SCS, the remaining constraints are inactive. Then, $\nabla F$ can be written as

$$ \nabla_\xi F := \begin{bmatrix} A & B^\top & B^\top \\ D_\mu B & 0 & 0 \\ 0 & 0 & D_\mu G^* \end{bmatrix} , \quad \nabla_\xi F := \begin{bmatrix} L^* \\ 0 \\ 0 \end{bmatrix} , $$

where we partitioned and regrouped the matrices according to $(x^*, [\lambda^*, \mu_i^*], [\mu_i^*])$, and used $B(\xi) := \nabla_\theta \Theta_1(x^*(\xi), \xi)$. Further, we have used $D_\mu := \text{diag}(\mu_i^*)$ and $D_\mu := \text{diag}(g_\xi)$. Denote the upper left part of $\nabla_\xi F$ by

$$ M := \begin{bmatrix} A & B^\top \\ D_\mu B & 0 \end{bmatrix}. \tag{5} $$
Then, using Lemma 5 (and assuming for the moment that $M$ is invertible), we have that
\[
(\nabla_{(x,\mu)} F)^{-1} = \begin{bmatrix} M^{-1} & \ast \\ 0 & 0 \end{bmatrix},
\]
where $\ast$ denotes non-zero, but irrelevant components.

Hence, we can already conclude from the derivatives for \( \nabla \xi^z \) that \( \nabla \xi \mu^*_z \) is invertible.

Note that $A$ is invertible by Assumption 1 and the Schur complement of $M$ given by $-D_\mu B^T A^{-1} B^T$ is invertible because $D_\mu$ in invertible by SCS. Hence, $M$ is nonsingular, and its inverse can be be computed according to Lemma 5 as
\[
M^{-1} = \begin{bmatrix} M_1 & M_2 \\ M_3 & M_4 \end{bmatrix},
\]
where
\[
\begin{align*}
M_1 &= (I - A^{-1} B^T (D_\mu B A^{-1} B^T)^{-1} D_\mu B) A^{-1} \\
M_2 &= A^{-1} B^T (D_\mu B A^{-1} B^T)^{-1} = A^{-1} B^T (BA^{-1} B^T)^{-1} D_\mu^{-1} = A^{-1} B^T (BA^{-1} B^T)^{-1} D_\mu^{-1} D_\mu B B^T D_\mu^{-1} = \Sigma B^T D_\mu^{-1}, \\
M_3 &= (D_\mu B A^{-1} B^T)^{-1} D_\mu B A^{-1} = B^T A^{-1} B (BA^{-1} B^T)^{-1} A^{-1} A = B^T A^{-1} A, \\
M_4 &= -(D_\mu B A^{-1} B^T)^{-1} = -(BA^{-1} B^T)^{-1} D_\mu^{-1} = -B^T A^{-1} A^{-1} B (BA^{-1} B^T)^{-1} D_\mu^{-1} D_\mu B B^T D_\mu^{-1} = -B^T A^{-1} A^{-1} A = B^T A^{-1} A^{-1} A.
\end{align*}
\]
\[
\begin{align*}
\nabla \xi^z &= -(\nabla \xi^{-1} (\xi^z))^\top \nabla \xi^z \\
\nabla \xi \mu^*_z &= \nabla \xi \mu^*_z \\
\nabla \xi \mu^*_z &= \begin{bmatrix} M_1 & M_2 \\ M_3 & M_4 \end{bmatrix} \nabla \xi^z
\end{align*}
\]
which yields the desired expressions.

We defer an in-depth discussion of (4) to Section V where we explore special cases. One special feature of (4) that we require though is the fact that $\Pi$ and $\Sigma$ can be interpreted as (oblique) projections onto the kernel of $B$ and its orthogonal complement, respectively. Consequently, exploiting non-expansivity, their norm can be bounded independently of $B$:

**Lemma 1.** Consider the setup of Proposition 4. For all $\xi \in \mathcal{N}$, it holds that $\max\{\|\Pi(\xi)\|,\|\Sigma(\xi)\|\} \leq \sqrt{\lambda_{\max} A(\xi)/(\lambda_{\min} A(\xi))}$.

**Proof.** As before, we drop the argument $\xi$. Consider
\[
\begin{align*}
\text{minimize} \quad & \frac{1}{2} (v - w) \top A(v - w) \quad \text{subject to} \quad Bw = 0. \\
\text{Since} \quad & A \text{ is positive definite for all} \quad \xi \in \mathcal{E} \quad \text{by Assumption 1}, \quad \text{we can rewrite as} \\
\text{minimize} \quad & \frac{1}{2} \|Qv - \hat{w}\|^2 \quad \text{subject to} \quad \hat{B} \hat{w} = 0
\end{align*}
\]
where $Q := A^{1/2}$ is the unique symmetric positive definite matrix such that $A = Q^2$, $\hat{w} := Qw$, and $\hat{B} := BQ^{-1}$.

Since (7) is a Euclidean projection onto the kernel of $\hat{B}$, its solution can be stated explicitly as [40, eq. 5.13.3]
\[
\hat{w}^* = Qv - \hat{B}^T (\hat{B} \hat{B}^T)^{-1} \hat{B} Qv = Q(I - A^{-1} B^T (BA^{-1} B^T)^{-1} B)v = Q \Pi v.
\]
By non-expansivity of an orthogonal projection onto a subspace, we have $\|\hat{w}^*\| \leq \|Qv\|$, and therefore we can write
\[
\lambda^\min_Q \|\Pi v\| \leq \|Q v\| \leq \|\hat{w}^*\| \leq \|Qv\| \leq \lambda^\max_Q \|v\|.
\]
Rearranging the leftmost and rightmost terms together with the fact that $\lambda^\max_A = (\lambda^\max_B)^2$ and $\lambda^\max_A = (\lambda^\min_B)^2$ yields the desired bound for $\|\Pi\|$.

The projection on the orthogonal complement of $\ker \hat{B}$ is given by $\Sigma := I - \Pi$ [40, eq. 5.13.6]. The same bound as for $\|\Pi\|$ holds for $\|\Sigma\|$ since $\Sigma$ is also non-expansive.

**Lemma 2** allows us to bound (4) and thereby establish bounds for the local Lipschitz constants of $x^*$ and $\mu^*$ at $\xi$.

**Corollary 5.** Consider the setup of Proposition 4 Then,
\[
\|\nabla \xi x^*(\xi)\| \leq \ell_{x^*} \quad \text{and} \quad \|\nabla \xi (\lambda^*_{\mu^*})\| \leq \ell_{(\lambda^*,\mu^*)}
\]
holds for all $\xi \in \mathcal{N}$ where
\[
\ell_{x^*} := \frac{\lambda_{\max} A(\xi)}{\lambda_{\min} A(\xi)} \left( \frac{\|L^*(\xi)\|}{\lambda_{\min} A(\xi)} + \frac{\|G^*(\xi)\|}{\sigma_{\min} B(\xi)} \right),
\]
\[
\ell_{(\lambda^*,\mu^*)} := \frac{\lambda_{\max} A(\xi)}{\lambda_{\min} A(\xi)^{3/2}} \left( \left( \frac{\|L^*(\xi)\|}{\lambda_{\min} A(\xi)} \right)^{1/2} + \frac{\|G^*(\xi)\|}{\sigma_{\min} B(\xi)} \right).
\]

**Proof.** Using singular value decomposition $B^\top = V^\top \Lambda^\top U$, where $B^\top$ (i.e., $\Lambda^\top$) has a full rank since the columns of $B$ are linearly independent. Next we note that $B^\top = V^\top A^\top U$ also has a full rank and corresponds to the right pseudoinverse of $B$, where $\Lambda^\top$ is the right pseudoinverse of $\Lambda$. Comparing the expressions for $B^\top$ and $B^\top$ we have that for the minimal and maximal singular values $\sigma^\max_B = 1/\sigma^\min_B$ and $\sigma^\min_B = 1/\sigma^\max_B$ hold. Consequently, $\|B\| = 1/\sigma^\max_B(\xi)$ holds. Finally, the bounds in (8) follows from applying the triangle inequality, Cauchy-Schwarz, and Lemma 1 to (4).

**B. Sensitivity Analysis for General Regular Optimizers**

If the SCS assumption in Proposition 4 does not hold, then at least one constraint $i$ is weakly active, i.e., $gi(x^*(\xi)) = 0$ but $\mu^*_i(\xi) = 0$. Weakly active constraints occur, for example, when the unconstrained minimizer of $f$ lies exactly on a constraint surface. In this case, the constraint is active, but is not “pushing” the optimizer inwards.

Whenever we vary $\xi$ and a constraint changes from being strongly active to inactive (or vice versa), the constraint is momentarily weakly active for some $\xi$. This is a consequence of the continuity of $(x^*,\mu^*)$.

For values of $\xi$, for which some constraints are weakly active, $(x^*,\lambda^*,\mu^*)$ might not be differentiable, i.e., (3) might not apply. Consequently, the best we can hope for in the
general case where SCS does not hold, is a bound on the Lipschitz constant of \((x^*, \lambda^*, \mu^*)\).

To deal with weakly active constraints, we follow the approach in [35]. For this purpose, consider the same setup as in Proposition 4 and let \((\tilde{x}^*, \tilde{\lambda}^*, \tilde{\mu}^*)\) be a regular optimizer for (2) for \(\xi \in \Xi\), satisfying LICQ at \(x^*\), KKT, and SSOSC but not necessarily SCS. Further, let \(R\) be any index set satisfying
\[
\mathbf{I}^*(\tilde{x}^*, \tilde{\mu}^*, \tilde{\xi}) \subset R \subset \mathbf{I}^*(\xi),
\]
that is, \(R\) includes all strongly active constraints and an arbitrary subset of weakly active constraints at \((\tilde{x}^*, \tilde{\lambda}^*, \tilde{\mu}^*, \tilde{\xi})\). Next, consider the equality-constrained problem

\[
\text{minimize } f(x) \\
\text{subject to } h(x, \xi) = 0, \quad \ell_R(x, \xi) = 0. \tag{10}
\]

By considering (10) we make a choice for every constraint that is weakly active at \((\tilde{x}^*, \tilde{\lambda}^*, \tilde{\mu}^*)\) to consider it like a strongly active constraint (by including it in \(R\)) or to ignore it.

Note that Proposition 4 and Corollary 3 can be applied to (10). In particular, the SCS condition is vacuous because, let \(x_R^* : N_R \rightarrow \mathbb{R}^n\) and \((\lambda_R^*, \mu_R^*) : N_R \rightarrow \mathbb{R}^{n+1}\) denote the primal and dual solution map of (10) in a neighborhood \(N_R\) of \(\xi\).

In the proof of [35] Thm 2.3.2] it was established\(^3\) that for every \(\xi\) in a neighborhood \(N \subset \bigcap_R N_R\) of \(\xi\) the following holds: If \((x^*(\xi), \lambda^*(\xi), \mu^*(\xi))\) is a regular optimizer of (2), then there exists an index set \(R\) satisfying (9) such that \((x^*(\xi), \lambda^*(\xi), \mu^*(\xi))\) is a regular optimizer of (10). In particular, we have \(x^*(\xi) = x_R^*(\xi), \lambda^*(\xi) = \lambda_R^*(\xi), \) and \(\mu^*(\xi) = 0\) for all \(i \notin R\).

Note that the set \(R\) for which this equivalence of solutions between (2) and (10) holds, depends on \(\xi\) and there does in general not exist a single set \(R\) that works for the entire neighborhood \(N\) of \(\xi\).

This key insight can be used to establish bounds on the Lipschitz constants of \(x^*\) and \(\mu^*\) at \(\xi\) in the absence of SCS:

**Theorem 6.** Consider the same setup as in Theorem 3 and let Assumption 4 hold at \((\tilde{x}^*, \tilde{\lambda}^*, \tilde{\mu}^*)\). Then, \(x^* : N \rightarrow \mathbb{R}^n\) and \((\lambda^*, \mu^*) : N \rightarrow \mathbb{R}^{n+1}\) are locally Lipschitz at \(\xi\) with bounds for the Lipschitz constants given by \(\ell_{x}^{*} \quad \text{and} \quad \ell_{x}(\lambda^*, \mu^*)\) as in (8).

The difference in (8) for Corollary 5 and Theorem 6 lies in the fact that for Corollary 5 the set \(\mathbf{I}^*\) (required for \(A, \beta, B, L^*, \) and \(G^*\)) contains only strictly active constraints, whereas for Theorem 6 it may also contain weakly active constraints.

**Proof.** Given \(R\) satisfying (9), let \(x_R^* : N \rightarrow \mathbb{R}^n, \lambda_R^* : N \rightarrow \mathbb{R}^m, \) and \(\mu_R^* : N \rightarrow \mathbb{R}^{n+1}\) the solution maps of (10) in a neighborhood \(N\) of \(\xi\). Consequently, we apply Corollary 3 to (10) for every \(R\) satisfying (9), and we let \(\ell_{x_R}^{*}\) and \(\ell_{x}(\lambda_R^*, \mu_R^*)\) denote the respective bounds on the Lipschitz constant of \(x_R^*\) and \((\lambda_R^*, \mu_R^*)\) at \(\xi\) according to (8). Similarly, let \(A_R^*(\xi), B_R^*(\xi), L_R^*(\xi)\) and \(G_R^*(\xi)\) denote the corresponding quantities for (10), evaluated at \(\xi\).

Since, for every \(\xi \in \Xi\) the solution \((x^*(\xi), \mu^*(\xi))\) is equal to \((x_R^*(\xi), \mu^*(\xi))\) for some \(R\) satisfying (9), we can upper bound the local Lipschitz constants of \(x^*\) and \((\lambda^*, \mu^*)\) at \(\xi\) by maximizing over all (finite) possibilities of \(R\) satisfying (9). Namely, for \(x^*\) we have
\[
\sup_{\xi} \frac{\|x_R^*(\xi) - x^{*}(\xi)\|}{\|\xi - \xi\|} \leq \max R \sup_{\xi} \frac{\|x_R^*(\xi) - x^{*}(\xi)\|}{\|\xi - \xi\|} \leq \max R \ell_{x},
\]
and the case for \((\lambda^*, \mu^*)\) follows analogously.

In particular, we claim that this maximum is achieved exactly for the choice \(R = I^*(\xi)\) where all strongly and weakly active constraints are considered.

To see this, note the following: Since lifting a vector to higher dimensions by adding components can only increase its 2-norm, we have \(\|G_R^*(\xi)\| \leq \|G_{I^*(\xi)}^*(\xi)\|\) for any \(R\) satisfying (9). Analogously, we have \(\sigma_{\min} B_R^*(\xi) \geq \sigma_{\min} B_{I^*(\xi)}\) since adding columns to \(B^T\) can only reduce its minimum singular value.

Next, at \(\xi\) we that \(\mu^*(\xi) = \mu_{I^*(\xi)}^*(\xi)\) for any \(R\) satisfying (9). Further, since \(\mu_{I^*(\xi)}^*(\xi) = 0\), for all weakly active constraints with index \(i\), we have \(L^*(\xi) = L_{I^*(\xi)}^*(\xi)\) and \(A(\xi) = A_{I^*(\xi)}(\xi)\). In other words, at \(\xi\) the weakly active constraints do not affect the quantities \(A\) and \(L^*\) which are derivatives of the Lagrangian.

Applying these facts to (8), it follows that \(\ell_{x}^{*} \leq \ell_{x}^{I^*(\xi)}\) and \(\ell_{x}(\lambda^*, \mu^*) \leq \ell_{x}(\lambda_{I^*(\xi)}^*, \mu_{I^*(\xi)}^*)\). for all \(R\) satisfying (9) and thus we have \(\ell_{x}^{*} = \max_{R} \ell_{x}^{R}\) and \(\ell_{x}(\lambda^*, \mu^*) = \max_{R} \ell_{x}(\lambda_{I^*(\xi)}^*, \mu_{I^*(\xi)}^*)\), which completes the proof.

**IV. ASSUMPTIONS FOR GLOBAL SOLUTION MAPS**

**Theorem 6** provides bounds on the local Lipschitz constants for \(x^*\) and \(\mu^*\) at \(\xi\). Next, we provide a set of assumptions to give global bounds on the sensitivity. Naturally, some of these assumptions are restrictive, but they provide important intuition and use cases for the general result in Theorem 6.

For simplicity, instead of (2), we focus on problems with only inequality constraints which require a more careful investigation than equality constraints as shown in the previous analysis, i.e., we consider

\[
\text{minimize } f(x, \xi) \\
\text{subject to } g(x, \xi) \leq 0, \tag{11}
\]

for the remainder of this section. All of the following assumptions and conclusions can be generalized.

First, it is necessary that (11) admits a solution for every \(\xi\):

**Assumption 2.** The problem (11) is feasible for all \(\xi \in \Xi\).

Next, in order to guarantee that the solution map \(x^* : \Xi \rightarrow \mathbb{R}^n\) is single-valued for all \(\xi \in \Xi\), it is convenient, if not necessary, to assume (strong) convexity:

**Assumption 3.** For (11), let \(f\) be \(\alpha\)-strongly convex and have a \(\beta\)-Lipschitz gradient \(\nabla_x f\) in \(x\) for all \(\xi \in \Xi\). Further, for all \(i = 1, \ldots, m\), let \(g_i\) be convex and \(\nabla g_i\) be \(\ell_i\)-Lipschitz in \(x\) for all \(\xi \in \Xi\).
**Assumption 3** not only guarantees that (2) admits a unique optimizer for every $\xi \in \Xi$, it also implies that any KKT point of (11) (trivially) satisfies the SSOSC since $\nabla^2_{xx} L$ is positive definite for all $(x, \mu, \xi) \in \mathbb{R}^n \times \mathbb{R}_0^m \times \Xi$. Assumption 3 also implies the lower bound $\lambda^\min(A(\xi)) \geq \alpha$ for all $\xi \in \Xi$ and thus replaces Assumption 1. Lipschitz continuity of $\nabla_x f$ and $\nabla_x g$ is required to upper bound $\lambda^\max(A)$ as discussed below. Moreover, we generally require the active constraints to have uniformly full rank:

**Assumption 4.** Given (11), there exists $\omega > 0$ such that, for all $\xi \in \Xi$ and all $x \in \mathcal{X}(\xi)$ such that $\xi^\star \not= 0$, we have

$$\omega^2 \Xi \leq \nabla_x g_{\xi^\star}(x, \xi) \nabla_x g_{\xi^\star}(x, \xi)^T .$$

In particular, Assumption 4 guarantees that LICQ is satisfied at all $x \in \mathcal{X}(\xi)$, for all $\xi \in \Xi$, and that $\sigma^\min_{B^T}(\xi) \geq \omega$ in (8) is lower bounded away from zero whenever at least one constraint is active. Note that Assumption 4 is independent of the choice of the cost function.

Further note that, under Assumptions 3 and 4, (11) has a unique regular optimizer for every $\xi \in \Xi$. Hence, by Theorem 3 continuous solution (and multiplier) maps exist around every $\xi \in \Xi$. This implies that the maps $x^\star : \Xi \to \mathbb{R}^n$ and $\mu^\star : \Xi \to \mathbb{R}^m$ exist globally on all of $\Xi$ and are continuous.

Finally, because $A = \nabla^2_{xx} f + \sum_i \mu_i \nabla_{xx} g$ and $L^\star = \nabla_{\xi^\star} f + \sum_i \mu_i \nabla_{\xi^\star} g$ are weighted sums over $\mu$, we generally require an upper bound on the dual multiplier $\mu^\star$.

**Assumption 5.** There exist $\zeta_i > 0$ such that, for all $\xi \in \Xi$ and every KKT point $(x^\star(\xi), \mu^\star(\xi))$ of (11), we have $\mu^\star_i(\xi) \leq \zeta_i$ for all $i = 1, \ldots, m$.

Finding (tight) upper bounds on the dual multipliers is tricky and depends on the problem structure.

If $\|\nabla_x f(x, \xi)\| \leq B_f$ and $\|\nabla_x g_i(x, \xi)\| \leq B_{g_i}$ are uniformly bounded for all $\xi \in \Xi$, all $x \in \mathcal{X}(\xi)$ and all $i = 1, \ldots, m$, then it is easy to see that $\mu^\star_i \leq \zeta_i := B_f / B_{g_i}$.

Another possibility, documented in [41], applies specifically to convex optimization problems (i.e., under Assumption 3). Namely, if for all $\xi \in \Xi$ a strict lower bound $f_{\xi}^\star$ (i.e., $f_{\xi}^\star < f(x, \xi)$ for all $x \in \mathcal{X}(\xi)$) and a strictly feasible point $\bar{x}_i$ (i.e., $g(\bar{x}_i, \xi) = 0$) are known, then we can choose

$$\zeta_i = \sup_{\xi \in \Xi} \frac{f(\bar{x}_i, \xi) - f_{\xi}^\star(\xi)}{-g_i(\bar{x}_i)} .$$

Combining Assumptions 3 and 5 we can guarantee that

$$A = \nabla^2_{xx} f + \sum_{i=1}^m \mu_i^\star \nabla^2_{xx} g \succeq \left( \beta + \sum_i \zeta_i \ell_i \right) I .$$

In other words, we have $\lambda^\max(A(\xi)) \leq \beta + \sum_i \zeta_i \ell_i$.

The assumptions made so far can be summarized in the following statement:

**Corollary 7.** Under Assumptions 2 to 5 the primal and dual solution maps $x^\star : \Xi \to \mathbb{R}^n$ and $\mu^\star : \Xi \to \mathbb{R}^m$ of (11) are single-valued and Lipschitz continuous (on $\Xi$) with respective Lipschitz constants

$$\ell_{x^\star} := \sqrt{\frac{\beta + \sum_i \zeta_i \ell_i}{\alpha} \left( \frac{L^\star}{\alpha} + \frac{\bar{G}^\star}{\omega} \right)} ,$$

$$\ell_{\mu^\star} := \frac{\beta + \sum_i \zeta_i \ell_i}{\alpha^{1/2} \omega} \left( \frac{L^\star}{\alpha} + \frac{\bar{G}^\star}{\omega} \right) ,$$

where $L^\star := \sup_{\xi \in \Xi} \|L^\star(\xi)\|$ and $\bar{G}^\star := \sup_{\xi \in \Xi} \|G^\star(\xi)\|$ are upper bounds on $L^\star$ and $G^\star$ over $\Xi$, respectively.

It remains to establish bounds on $L^\star$ and $G^\star$. Such bounds, however, are highly problem dependent, and therefore we consider only special cases in the next section.

**V. General Discussion of Sensitivity Bounds**

We now discuss special cases for the established sensitivity bounds and how to relax the differentiability of $f$ and $g$.

**A. Special Cases**

**Stationary Constraints:** We first assume that the constraints are stationary and only the cost function is perturbed, i.e., we consider the problem

$$\text{minimize } f(x, \xi) \quad \text{subject to } g(x) \leq 0$$

and assume that Assumptions 2 to 5 are satisfied. Let $(\bar{x}^\star, \bar{\mu}^\star)$ be a regular optimizer for $\xi$ for which SCS holds, and therefore Proposition 4 is applicable.

In this case, $G^\star(\xi)$ is vacuous and we have $L^\star = \nabla^2_{\xi^\star} L = \nabla^2_{\xi^\star} f$. Hence, the evolution of $x^\star$ and $\mu^\star$ around $\xi$ is governed by the first terms in (11). In particular, we have $\nabla_{\xi^\star} x^\star = -\Pi A^{-1} L^\star$, which is independent of $B^\star$ (and hence of $\sigma^\min_{B^\star}$). In other words, the conditioning of the constraints (as quantified by Assumption 4) does not affect the sensitivity of $x^\star$ (however, it does affect the sensitivity of $\mu^\star$).

In this context, we can also make sense of the operator $\Pi$ which projects the quantity $A^{-1} L^\star$ onto ker $B^\top$ which is the tangent space of the surface of (strongly) active constraints. In other words, $\Pi$ guarantees that $x^\star$ can only change along the surface spanned by the (strongly) active constraints, which is plausible since the constraints are stationary. Figure 1a illustrates this effect.
If we further assume that no constraints are active at $\bar{x}$, we find ourselves in a situation of unconstrained optimization. In this case, (4) reduces to $\nabla_x x^* = -\left(\nabla^2_{xx} f(\xi)\right)^{-1} \nabla^2_{xx} f$. This results can also be derived by a direct application of the implicit function theorem, for which $\ell_x^*$ reduces to

$$
\ell_x^* := \frac{\ell_f}{\alpha}, \quad \text{where} \quad \ell_f := \sup_{\xi \in \Xi} \|\nabla^2_{xx} f(x^*(\xi'), \xi')\|.
$$

(14)

In particular, for unconstrained time-varying optimization, the assumption that $\nabla_x f$ is $\beta$-Lipschitz is not required.

Translational Perturbation of Objective: A particular class of objective functions is when the perturbation is in the form of a translation, i.e., we consider

$$
\text{minimize} \quad \hat{f}(x - c(\xi)) \quad \text{subject to} \quad g(x) \leq 0, \quad (15)
$$

where $\hat{f} : \mathbb{R}^n \to \mathbb{R}$ is $\alpha$-strongly convex and $\beta$-smooth and $c : \mathbb{R}^n \to \mathbb{R}^n$ is continuously differentiable and $\ell_c$-Lipschitz. Otherwise, let Assumptions 2 to 5 hold. Considering $f(x, \xi) = f(x - c(\xi))$, the structure of (15) implies that $\nabla_x f(x, \xi) = \nabla \hat{f}(x - c(\xi))$, and therefore $\nabla^2_{xx} f(x, \xi) = -\nabla^2_{xx} \hat{f}(x - c(\xi)) \nabla c(\xi)$, and therefore $\|L^*\| \leq \beta \ell_c$ (since $\nabla_x g = 0$). In fact, $c$ does not need to be continuously differentiable, but only Lipschitz as discussed in the forthcoming Section V-B.

Right-hand Constraint Perturbations: Next, we consider a stationary cost function and a constraint parametization that takes the form of a right-hand-side perturbation, i.e.,

$$
\text{minimize} \quad f(x) \quad \text{subject to} \quad u(x) \leq v(\xi), \quad (16)
$$

where $u : \mathbb{R}^n \to \mathbb{R}^m$ and $v : \Xi \to \mathbb{R}^m$ are twice continuously differentiable, and Assumptions 2 to 5 hold.

Let $(\bar{x}^*, \bar{\mu}^*)$ be a regular optimizer for $\xi$ and let SC0 hold, such that $x^*$ and $\mu^*$ are differentiable at $\xi$.

Exploiting the structure of (16), we have $\nabla^2_{xx} g = 0$, and $\nabla^2_{xx} f = 0$, and thus $L^* = 0$. Consequently, according to Proposition 4, it holds that $\nabla_{xx} x^* = -\nabla B^T G^*$. Recall from Lemma 1 and its proof, that $\nabla$ denotes an orthogonal projection onto the space spanned by $Q^{-1} \nabla g_i(\xi)$ for all $i \in \Gamma_i(\xi)$ where $Q := A^T \Lambda^{1/2}$. Roughly speaking, as the constraints vary according to (16), the optimizer can only move $Q^{-1}$-orthogonally to the surface spanned by the active constraints. This behavior is illustrated in Figure 1b.

Linear Constraints: Yet, a more specific case is to assume that $\mathcal{X}(\xi)$ is a polyhedron, that is, to consider

$$
\text{minimize} \quad f(x) \quad \text{subject to} \quad Ux \leq v(\xi), \quad (17)
$$

where $U \in \mathbb{R}^{m \times n}$ is assumed to have full row rank (thus satisfying Assumption 4). The fact that in this case we have that $\nabla^2_{xx} g = 0$ (and that $L^* = \nabla^2_{xx} f$) obviates the need for Assumption 5, i.e., there is no need to estimate the upper bound of the dual multipliers. Hence, the bounds $\ell_x^*$ and $\ell_{\mu^*}$ in Corollary 7 simplify to

$$
\ell_x^* := \sqrt{\frac{\beta}{\alpha}} \frac{\ell_v}{\omega} \quad \text{and} \quad \ell_{\mu^*} := \frac{\beta^{3/2}}{\alpha} \frac{\ell_v}{\omega^2} \quad \text{and} \quad \ell_{v} := \sup_{\xi \in \Xi} \|\nabla v(\xi)\| \quad \text{is the Lipschitz constant of} \quad v.
$$

(18)

B. Relaxation of Differentiability Assumption

All results thus far require $f$ and $g$, from (11), to be twice continuously differentiable in $(x, \xi)$. Whether it is possible to directly relax this (possibly restrictive) assumption without losing Lipschitz continuity is still an open question.

However, this issue sometimes can be circumvented by a perturbation mapping that isolates the non-smoothness. Namely, assume that there exist a map $\tilde{\xi} : \mathbb{R}^p \to \mathbb{R}^r$ that is $\xi$-Lipschitz. Then, based on (11), we consider the problem

$$
\text{minimize} \quad f(x, \tilde{\xi}(\psi)),
$$

(19)

subject to $g(x, \tilde{\xi}(\psi)) \leq 0$ parametrized in $\psi$. If (11) admits a $\ell_{x^*}$-Lipschitz solution map with respect to $\xi$, then (19) admits a $\ell_{x^*}$-Lipschitz solution map $\bar{x}^*$, in $\psi$. This holds since the composition of Lipschitz maps is also Lipschitz.

Based on the special problem classes discussed in the previous subsection, we can also derive the following result which incorporates non-differentiable perturbations with tighter bound on the Lipschitz constant than derived from (19):

Corollary 8. Consider the problem

$$
\text{minimize} \quad \hat{f}(x - c(\xi)) \quad \text{subject to} \quad Ux \leq v(\xi)
$$

(20)

where $\xi \in \Xi \subset \mathbb{R}^r$. Define $\tilde{\mathcal{X}}(\xi) := \{x \mid Ux \leq v(\xi)\}$ and let

(i) $\hat{f} : \mathbb{R}^n \to \mathbb{R}$ be twice continuously differentiable, $\alpha$-strongly convex, with $\beta$-Lipschitz $\nabla \hat{f}$,

(ii) $c : \Xi \to \mathbb{R}^n$ and $v : \Xi \to \mathbb{R}^m$ be $\ell_c$- and $\ell_v$-Lipschitz, respectively, and

(iii) $U \in \mathbb{R}^{m \times n}$ and $\omega > 0$ such that for every $\xi \in \Xi$ and every $x \in \tilde{\mathcal{X}}(\xi)$ one has $\omega^2 \|U^T\| \leq \frac{\ell_c}{\ell_v} \frac{U^T \Lambda^{1/2}}{U^T \Lambda^{1/2}}$.

If $\tilde{\mathcal{X}}(\xi) \neq \emptyset$ for every $\xi \in \Xi$, then the primal and dual solution maps $x^* : \Xi \to \mathbb{R}^n$ and $\mu^* : \Xi \to \mathbb{R}^m$ of (20) are Lipschitz with respective bounds on Lipschitz constants

$$
\ell_{x^*} = \sqrt{\frac{\beta}{\alpha}} \left(\frac{\beta \ell_c}{\alpha} + \frac{\ell_v}{\omega}\right) \quad \text{and} \quad \ell_{\mu^*} = \beta^{3/2} \left(\frac{\beta \ell_c}{\alpha} + \frac{\ell_v}{\omega}\right).
$$

Proof. The key point about Corollary 8 is the fact that $c$ and $v$ are Lipschitz continuous, but not twice differentiable. To address this issue, consider, instead of (20), the problem

$$
\text{minimize} \quad \hat{f}(x - c) \quad \text{subject to} \quad Ux \leq v \quad (21)
$$

parametrized in $(c, v)$. Under the stated assumptions, the parametrized objective $(x, c) \mapsto \hat{f}(x - c)$ and constraint function $(x, v) \mapsto Ux - v$ are twice continuously differentiable in $(x, c)$ and $(x, v)$, respectively. Therefore, the solution map $(c, v) \mapsto \bar{x}^*(c, v)$ of (21) is Lipschitz by Corollary 7. In fact, by considering the reasoning for translational perturbations of the objective and right-hand perturbations of polyhedra in the previous subsection, we know that $\bar{x}^*$ is Lipschitz in $c$ and $v$ with Lipschitz constants $(\frac{\ell_c}{\alpha})^{3/2}$ and $\sqrt{\frac{2}{\alpha}} \frac{\ell_c}{\omega}$, respectively.

Moreover, replacing $c$ with $c(\xi)$ and using $\ell_c$-Lipschitz continuity of $c$ we have, for all $c, c \in \mathbb{R}^n$ and all $v \in \mathbb{R}^m$,

$$
\|\bar{x}^*(c(\xi), v) - \bar{x}^*(c(\xi), v)\| \leq \sqrt{\frac{2}{\alpha}} \frac{\ell_c}{\omega} \|\xi' - \xi\| \quad (22)
$$
Similarly, replacing $v$ with $v(\xi)$ we have
\[
\| \dot{x}^*(c, v(\xi')) - \dot{x}^*(c, v(\xi)) \| \leq \sqrt{2 \sum_{j=1}^{n} \ell_j |\xi'_j - \xi_j|} \tag{23}
\]
for all $c \in \mathbb{R}^n$ and all $v', v \in \mathbb{R}^m$.

The mutual uniformity of the these Lipschitz constants, i.e., the fact that (22) and (23) are independent of $v$ and $c$, respectively, allows us to write
\[
\|x^*(\xi') - x^*(\xi)\| \leq \sqrt{2 \sum_{j=1}^{n} \left( \frac{\beta_j}{\alpha_j} + \frac{\lambda_j}{\alpha_j} \right) |\xi'_j - \xi_j|},
\]
where $x^*(\xi) = \dot{x}^*(c(\xi), v(\xi))$ and thus the solution map $x^*: \Xi \rightarrow \mathbb{R}^n$ of (20) is Lipschitz with the desired Lipschitz constant. The Lipschitz constant for $\mu^*$ follows analogously. \hfill \Box

VI. TRACKING PERFORMANCE OF CONTINUOUS-TIME ALGORITHMS FOR TIME-VARYING OPTIMIZATION

In this section, we derive a tracking result that applies to any continuous-time algorithm with absolutely continuous trajectories that satisfy a common monotonicity-like property. Although the same property has previously been exploited for specific algorithms, our tracking result is more general since it does not require any additional assumptions on the (running) algorithm. For instance, our performance bound can be applied to algorithms whose trajectories are Krasovskii [31] Ch. 5 or Filippov [12] solutions of differential inclusions. In the forthcoming Section VII we require the generality of this result for our novel discontinuous sweeping gradient flow.

Generally, in this section, we consider a time-varying optimization problem with non-smooth variations in time whose instantaneous optimizer the running algorithm tracks. Hence, instead of periodically referring back to the sensitivity results of the previous sections, we make the following assumption:

**Assumption 6.** The problem (2) with $\xi = \dot{\xi}(t)$ and $\dot{\xi} : \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}^r$ admits a unique (global) optimizer for every $t \in \mathbb{R}_{\geq 0}$ and the solution map $\dot{x}^* : \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}^n$ is $\ell$-Lipschitz.

The following proposition guarantees the bounded asymptotic behavior of the running algorithm, i.e., the distance $\|x(t) - \dot{x}^*(t)\|$ between the trajectories of the running algorithm $x(t)$ and instantaneous optimizer $\dot{x}^*(t)$ is asymptotically bounded.

**Proposition 9.** Consider (2) with $\xi = \dot{\xi}(t)$ and $\dot{\xi} : \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}^r$ and let Assumption 6 be satisfied. Furthermore, let $x : \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}^n$ be absolutely continuous and assume that
\[
\langle \dot{x}(t), x(t) - \dot{x}^*(t) \rangle \leq -a \|x(t) - \dot{x}^*(t)\|^2, \tag{24}
\]
holds for almost every $t \in \mathbb{R}_{\geq 0}$ where $a \in \mathbb{R}_{\geq 0}$.

Then, \[ \lim_{t \to \infty} \sup_{t \geq 0} |x(t) - \dot{x}^*(t)| \leq \frac{c}{a} \] holds. Moreover, if $\|x(0) - \dot{x}^*(0)\| \leq \frac{c}{a}$, then $\|x(t) - \dot{x}^*(t)\| \leq \frac{c}{a}$ holds for all $t \geq 0$.

In the next subsection we provide the necessary technical results and the proof of Proposition 9.
To prove the lemma, we apply Lemma 2 to \( W \). Hence, we need to show that (i) the integral \( \lim_{t \to \infty} W(\tau) d\tau \) exists and is finite, and that (ii) \( W \) is uniformly continuous.

To show (i), we exploit the fact that the growth of \( V \) is bounded by \(-W\) to establish the bound

\[
\int_0^t W(\tau) d\tau \leq - \int_0^t \dot{V}(\tau) d\tau = V(0) - V(t) \leq V(0),
\]

where the last inequality follows from the fact that \( V(t) \) is non-negative for all \( t \in \mathbb{R}_{\geq 0} \), by definition. Further, the left-hand side integral is non-decreasing in \( t \) (since \( W(t) \geq 0 \) for all \( t \)). Hence, as \( t \to \infty \), the limit exists and is finite.

For (ii) it suffices to show that \( W \) is non-increasing itself. Then, since \( W \) is continuous and lower bounded, it follows from Lemma 3 that \( W \) is uniformly continuous.

Thus, we have shown that \( W \) is uniformly continuous and \( \lim_{t \to \infty} \int_0^t W(\tau) d\tau \) exists. Therefore, from Lemma 2, it follows that \( W(t) \to 0 \) for \( t \to \infty \) and consequently, from the definition of \( W \), we have \( \lim sup_{t \to \infty} \|y(t)\| \leq b/a \).

We show the last part of the lemma by contradiction. Assume that there exists \( T \geq 0 \) such that \( \|y(T)\| > b/a \) and that there exists \( t_1 \geq T \) such that \( \|y(t_1)\| > b/a \). That means \( V(t_1) > V(T) = 0 \), but \( V(T) = 0 \), hence \( V(T + \delta) > 0 \), for all \( \delta > 0 \). Consequently \( V(t_1) = 0 \) which is contradiction to \( V(t_1) > V(T) \), hence \( \|y(t)\| \leq b/a \), for all \( t \geq T \).

Note that the second part of Theorem 10 that guarantees the invariance of \( \mathcal{B} \), is essentially a consequence of Nagumo’s theorem [45, Thm. 1.2.1]. Likewise, Theorem 10 and its proof relate to Lyapunov functions, barrier functions, etc.

Lastly, in order to show that Proposition 9 follows from Theorem 10, we first note \( x \) and \( \tilde{x}^* \) are both (individually) differentiable for almost all \( t \geq 0 \) and thus they are (jointly) differentiable for almost all \( t \geq 0 \) (since the union of two zero-measure sets has measure zero). Hence, the instantaneous optimizer trajectory \( \tilde{x}^* \) satisfies Assumption 6.

A. Perturbed Sweeping Processes

Given a non-empty convex set \( \mathcal{X} \), recall that the normal cone of \( \mathcal{X} \) at \( x \in \mathcal{X} \) is given by

\[
N_x \mathcal{X} := \{y \mid \forall y \in \mathcal{X} : \langle y, y - x \rangle \leq 0\}.
\]

Sweeping processes [34], [46], [47] have originally been formulated to describe a sweeping effect of a moving impenetrable boundary on a mobile object. When the object itself is subject to a perturbation (e.g. a force), we use a perturbed sweeping process [33], [48] which is formally defined as

\[
\dot{x}(t) \in \Phi(x, t) - N_x \mathcal{X}(t), \quad x(0) \in \mathcal{X}(0).
\]

B. Sweeping Gradient Flows

As a new running algorithm for constrained time-varying optimization we consider a perturbed sweeping process of the form [27] where \( \Phi(x, t) := -\nabla_x f(x, t) \) is the negative gradient of a time-varying cost function \( f : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R} \), thus the resulting sweeping gradient flow is given by

\[
\dot{x}(t) \in -\nabla_x f(x, t) - N_x \mathcal{X}(t), \quad x(0) \in \mathcal{X}(0).
\]

VII. CONTINUOUS-TIME CONSTRAINED OPTIMIZATION

Before presenting a novel continuous-time optimization scheme, we review some results regarding sweeping processes.

4The original result applies to prox-regular sets in a Hilbert space instead of convex sets in \( \mathbb{R}^n \). Furthermore, in the original theorem (ii) is more general and requires \( \mathcal{X} \) to vary in a absolutely continuous way instead of a Lipschitz continuous way. Finally, in the original statement (iii) is split into a local Lipschitz and linear growth condition. For simplicity, we assume global Lipschitz continuity to meet both requirements.
In the following, we consider sweeping gradient flows for the special problem structure from Corollary 8 for which we can give easy-to-interpret results. However, (29) is well-defined for much more general setups\footnote{For instance, existence of trajectories for sweeping gradient flows is guaranteed even if \( X \) is not convex, but \textit{prox-regular}, as long as \( \nabla x f \) is measurable in \( t,X(t) \) is non-empty for all \( t \geq 0 \) and, roughly-speaking, \( X \) varies in an absolutely continuous way \cite{33}. To guarantee a global asymptotic bound on the tracking error, however, we generally require Assumptions \ref{ass:1} to \ref{ass:4} to hold such that Corollary \ref{cor:8} and Proposition \ref{prop:9} can be applied.}

The key insight from this section is that our sensitivity bounds and the generalized tracking results of the previous sections can be applied to non-trivial discontinuous optimization dynamics involving time-varying constraints. Namely, we establish the following result:

**Theorem 12.** Consider the problem

\[
\begin{array}{ll}
\text{minimize} & \hat{f}(x - c(t)) \\
\text{subject to} & Ux \leq v(t),
\end{array}
\]

where \( t \geq 0 \). Define \( \hat{X}(t) := \{x \mid Ux \leq v(t)\} \) and let

(i) \( \hat{f} : \mathbb{R}^n \to \mathbb{R} \) be twice continuously differentiable, \( \alpha \)-strongly convex, and \( \nabla \hat{f} \) is \( \beta \)-Lipschitz,

(ii) \( c : \mathbb{R}_0^+ \to \mathbb{R}^n \) and \( v : \mathbb{R}_0^+ \to \mathbb{R}^m \) be \( \ell_c \)- and \( \ell_v \)-Lipschitz continuous,

(iii) \( U \in \mathbb{R}^{n \times m} \) and \( \omega > 0 \) such that for every \( \xi \in \Xi \) and every \( x \in \hat{X}(\xi) \) one has \( \omega \|\xi\| \leq \|U_x\| \|\xi\|_V \).

If \( \hat{X}(t) := \{x \mid Ux \leq v(t)\} \neq \emptyset \) for all \( t \geq 0 \), then the sweeping gradient flow

\[
\dot{x} \in -\nabla \hat{f}(x - c(t)) - N_x \hat{X}(t)
\]

admits a complete solution \( x : \mathbb{R}_0^+ \to \mathbb{R}^n \) for every initial condition \( x(0) \in \hat{X}(0) \) and it holds that

\[
\limsup_{t \to \infty}\|x(t) - x^*(t)\| \leq \frac{\beta \|c(t)\|}{\alpha \varepsilon_0} \left( \frac{\beta L_c}{\alpha} + \frac{\varepsilon_0}{\omega} \right),
\]

where \( x^*(t) \) is the unique optimizer of (30) at time \( t \). Furthermore, if \( \|x(0) - x^*(0)\| \leq \frac{\beta \|c(0)\|}{\alpha \varepsilon_0} \), then \( \|x(t) - x^*(t)\| \leq \frac{\beta \|c(t)\|}{\alpha \varepsilon_0} \) holds for all \( t \geq 0 \).

To show Theorem 12 we first need to prove the existence of complete solutions for (29).

**Lemma 4.** Given the setup of Theorem 12 (31) admits a unique complete solution for any initial point \( x(0) \in \hat{X}(0) \).

Even though Lemma 4 is an existence result, we can prove it using the sensitivity results of the previous section.

**Proof.** The proof is an application of Theorem 11. In order to apply Theorem 11 to (31), note that for the setup in Theorem 12 the requirements (i) in Theorem 11 (non-empty closed convex \( \hat{X}(t) \)) and (iii) (Lipschitz vector field) hold by assumption on any compact interval \([0,T] \subset \mathbb{R}_0^+\). Also, \( \nabla \hat{f}(x,t) = \nabla \hat{f}(x - c(t)) \) is measurable in \( t \) since it is continuous in \( t \). It remains to show (ii) by establishing (28).

For this purpose, we consider \( d(z,\hat{X}(t)) \) as the solution of the parametrized optimization problem

\[
\begin{array}{ll}
\text{minimize} & \frac{1}{2}\|y - z\|^2_2 \\
\text{subject to} & Uy \leq v(t),
\end{array}
\]

where \( t \) varies but \( z \) is fixed. As a problem parametrized in \( t \), (33) falls into the class of problems to which Corollary 8 applies. Namely, we have \( \alpha = \beta = 1 \) (since \( \hat{f}(x) := \frac{1}{2}\|x\|^2 \)), \( \ell_c = 0 \) (since \( c \equiv 0 \)), and \( \omega \) as defined in Theorem 12. Further, by the assumption in Theorem 12 the feasible set of (33) (which is equivalent to \( \hat{X}(t) \)) is non-empty for all \( t \geq 0 \).

Therefore, by Corollary 8 the solution map \( t \mapsto y^*_t(t) \) of (33) is Lipschitz continuous with a bound on the Lipschitz constant that is independent of \( z \) and given by \( \ell_v = \ell_v/\omega \). Therefore, we have \( |y^*_t(t') - y^*_t(t)| \leq \ell_v |t' - t| \) for any \( z \in \mathbb{R}^n \) and \( t,t' \geq 0 \), and (ii) in Theorem 11 holds on any compact interval \([0,T] \).

Hence, Theorem 11 guarantees the existence of a unique solution of (31) for every initial condition \( x(0) \in \hat{X}(0) \), and for every compact interval \([0,T] \) and hence, by definition, a complete solution on \( \mathbb{R}_{0}^+\).

**Proof of Theorem 12** By Corollary 8 (30) admits a unique primal solution \( x^*(t) \) for every \( t \geq 0 \). In particular, the solution map \( t \mapsto x^*(t) \) of (30) is \( \ell_v \)-Lipschitz with \( \ell_v := \sqrt{\frac{\beta}{\alpha}} \left( \frac{\beta L_c}{\alpha} + \frac{\varepsilon_0}{\omega} \right) \). Consequently, Assumption 6 is satisfied. Furthermore, Lemma 4 guarantees the existence of a unique complete solution \( x : \mathbb{R}_0^+ \to \mathbb{R}^n \) for every initial condition \( x(0) \in \hat{X}(0) \). Next, we verify that (24) holds, i.e., we show that \( \langle \dot{x}(t), x(t) - x^*(t) \rangle \leq -\alpha \|x(t) - x^*(t)^2 \| \).

In the following let \( f(x,t) := \hat{f}(x - c(t)) \). Recall that, for almost all \( t \geq 0 \), we have \( \langle \dot{x}(t), x(t) - x^*(t) \rangle = -\nabla_x f(x(t),t) - \eta(t) \) for some \( \eta(t) \in N_x \hat{X}(t) \). Further, because \( x^*(t) \) solves (30), it satisfies the KKT conditions which is equivalent to saying that \( -\nabla_x f(x^*(t),t) - \eta^*(t) = 0 \) for some \( \eta^*(t) \in N_x \hat{X}(t) \).

Putting these two insights together, we have that

\[
\langle \dot{x}(t), x(t) - x^*(t) \rangle = -\nabla_x f(x(t),t) - \eta(t) + \nabla_x f(x^*(t),t) + \eta^*(t).
\]

Next, we omit the argument \( t \) for \( x, \dot{x}, \eta, \eta^* \) and \( X \). By definition (26) of the normal vectors \( \eta \in N_x \hat{X} \) and \( \eta^* \in N_x \hat{X} \), for \( \hat{X} \), we have \( \langle \eta, x - x^* \rangle \geq 0 \) and \( \langle \eta^*, x^* - x \rangle \geq 0 \), respectively. Since \( f \) is \( \alpha \)-strongly convex we have

\[
\langle \nabla f(x,t) - \nabla f(x^*(t),x - x^* \rangle \geq \alpha \|x - x^* \|^2_2.
\]

Combining these facts, we get

\[
\langle \dot{x}, x - x^* \rangle = -\langle \nabla f(x,t) - \nabla f(x^*(t),t) - \eta^*(t), x - x^* \rangle
\]

\[
= -\langle \eta, x - x^* \rangle - \langle \eta^*, x^* - x \rangle
\]

\[
= \langle -\nabla f(x,t) - \nabla f(x^*(t),t) - x - x^* \rangle
\]

\[
\leq -\alpha \|x - x^* \|^2_2.
\]

Consequently, by taking \( a = \alpha \), (24) holds. Thus, Proposition 9 is applicable and yields the desired asymptotic tracking bound and completes the proof.

**VIII. ILLUSTRATIVE, NUMERICAL EXAMPLES**

In this section, we provide two simple numerical examples to illustrate our results for simple time-varying optimization setups. First, we consider an unconstrained optimization problem in one dimension, with non-smooth change in time to demonstrate the tightness of our bound in that case. Second, we consider a constrained time-varying optimization problem to illustrate the behavior of sweeping gradient flows.
A. Time-Varying Unconstrained Optimization

Consider the problem of minimizing

\[ f(x, \tilde{\xi}(t)) = \|x - \tilde{\xi}(t)\|^2, \]

where \(\tilde{\xi}\) is a triangular wave (hence non-smooth) with period \(\tau = 4s\). That is, for all \(n \in \mathbb{N}\), \(\tilde{\xi}\) is defined as

\[ \tilde{\xi}(t) = \begin{cases} (t - \tau(n-1)) - 1, & t \in [\tau(n-1), \tau(n-0.5)) \\ (t - \tau(n-1)) + 3, & t \in [\tau(n-0.5), \tau n) \end{cases}. \]

In particular, in reference to [19], we have \(\psi = t\) and \(f(x, \xi) := \|x - \xi\|^2\) is \(\alpha\)-strongly convex with \(\alpha = 2\). Furthermore, the solution map given by \(\tilde{x}^*(t) = \tilde{\xi}(t)\) is \(\ell_{\tilde{\xi}}\)-Lipschitz with \(\ell_{\tilde{\xi}} = 1\). From [14], we have \(\sup_{\xi \in [-1,1]} \|\nabla_{\xi} \tilde{f}\| = 2\), the solution trajectory \(x^*\) is \(\ell_{\tilde{x}^*}\)-Lipschitz with respect to \(\xi\), where \(\ell_{\tilde{x}^*} = 1\). Therefore, the solution trajectory \(\tilde{x}^*\) is \(\ell_{\tilde{x}^*}\)-Lipschitz continuous in \(t\) with \(\ell_{\tilde{x}^*} := \ell_{\tilde{x}} \cdot \ell_{\tilde{\xi}} = 1\).

For all \(t \geq 0\), the unconstrained gradient flow is given by

\[ \dot{x}(t) = -\nabla_x f(x(t), \tilde{\xi}(t)) = -2(x(t) - \tilde{x}(t)). \]

Finally, combining the zero cost function gradient, at the unconstrained optimizer (i.e., \(\nabla_x f(\tilde{x}^*(t), \tilde{\xi}(t)) = 0\)) with the \(\alpha\)-strongly convex property [34], for almost all \(t \geq 0\) we have

\[ \langle \dot{x}, x(t) - \tilde{x}^*(t) \rangle \leq -\alpha \|x(t) - \tilde{x}^*(t)\|^2. \]

Hence, from Proposition 9, \(\limsup_{t \to \infty} \|x(t) - \tilde{x}^*(t)\| \leq 0.5\).

\[ \text{Figure 2} \] shows the trajectory \(x(t)\) (black, solid) obtained from the gradient descent algorithm with initial condition \(x(0) = 5\) and the instantaneous optimizer trajectory \(\tilde{x}^*(t)\) (vermilion, dashed).

\[ \text{Figure 3} \] shows the distance between the trajectories \(\|x(t) - \tilde{x}^*(t)\|\) (black, solid), and the analytical sensitivity bound \(\frac{\ell_x}{\alpha} = 0.5\) (blue, dashed). Easily, it can be noticed that as \(t \to \infty\), the distance always stays within the analytical sensitivity bound. Additionally, in this example, the analytical sensitivity bound is indeed tight. Moreover, this example perfectly illustrates the invariance property of a ball of radius \(\frac{\ell_x}{\alpha}\) around the time-varying unconstrained optimizer \(\tilde{x}^*(t)\). Namely, the algorithm trajectory never perfectly tracks the unconstrained optimizer trajectory, yet always stays within \(\frac{\ell_x}{\alpha}\) from the optimizer.

B. Time-Varying Constrained Optimization

In this example, we consider the same problem setup as in [Theorem 12] with smooth changes in time. Namely, let

\[ Q := \begin{bmatrix} 12 & -8 \\ -8 & 10 \end{bmatrix} \succ 0, \quad P(t) := -\begin{bmatrix} 3 \sin(t + 3) + 1.3t \\ 2 \tanh(t - 3) + 0.71t \end{bmatrix} \]

let \(c(t) := -\frac{1}{2}Q^{-1}P(t)\), and the cost function is defined as

\[ \hat{f}(x(t) - c(t)) = (x(t) + \frac{1}{2}Q^{-1}P(t))^T Q(x(t) + \frac{1}{2}Q^{-1}P(t)). \]

In particular, \(\hat{f}\) is \(\alpha\)-strongly convex with \(\alpha = 2\sigma_{\max}(Q)\), and gradient \(\nabla \hat{f}\) is \(\beta\)-Lipschitz with \(\beta = 2\sigma_{\max}(Q)\). Furthermore, we can compute \(\ell_c = \sup_{t \geq 0} \|\hat{P}(t)\|\).

Next, the constraint set is given by

\[ \tilde{X}(t) = \{x(t) \in \mathbb{R}^n \mid Ux(t) \leq v(t)\}, \]

where \(v(t) := (V_1 t + V_2)\) and

\[ U := \begin{bmatrix} -2 & 1 \\ 1 & -1 \\ 0.5 & 1 \\ -3 & -1 \end{bmatrix}, \quad V_1 := \begin{bmatrix} -0.05 \\ -0.3 \\ 0.25 \\ -0.5 \end{bmatrix}, \quad V_2 := \begin{bmatrix} -2 \\ 5 \\ 4 \\ 3 \end{bmatrix}. \]

In particular, we have \(\ell_v = \sigma_{\max}(V_1)\) and in this special case, we can determine \(\omega\) by inspection. Hence, we can compute the bound on the Lipschitz constant of the constrained optimizer \(\tilde{x}^*\) as \(\ell_{\tilde{x}} = \sqrt{\frac{2\sigma}{\ell_{\tilde{x}^*} + \ell_c \omega}}\), and we can compute the analytical tracking bound as \(\frac{\ell_{\tilde{x}^*}}{\alpha} \approx 0.5302\). Figure 4 presents the analytical tracking bound as \(\frac{\ell_{\tilde{x}^*}}{\alpha} \approx 0.5302\) (blue, solid) and the tracking error upper bound \(\frac{\ell_{\tilde{x}^*}}{\alpha} \approx 0.5302\) (blue, dashed). It can be seen that the tracking error is always within the bound, and that it is decreasing within the considered time interval.

In contrast to the unconstrained time-varying optimization, for the constrained time-varying optimization, the movement of both the cost function and the constraint set has to be considered. Figure 5 illustrates the sweeping gradient flow trajectory \(x(t)\) (black, solid), instantaneous optimizer trajectory \(x^*(t)\) (vermilion, dashed) and the time-varying constraint set \(\tilde{X}(t)\) (blue, solid), captured in a different time steps, i.e., \(t = 0s, t = 3.95s, t = 7.95s, t = 11.95s,\) and \(t = 16s\). Note, however, that contrary to the theoretical requirement, the problem becomes eventually infeasible with \(\tilde{X}(t)\) collapsing to an empty set.

IX. Conclusion

In this work, we have revisited classical sensitivity results to characterize and quantify the Lipschitz continuity primal and dual solutions of nonlinear perturbed optimization problems. Since time-varying optimization is a natural application of our results, we have established a general tracking result for any continuous-time schemes satisfying a monotonicity-like...
property. In particular, this result can be immediately applied to discontinuous dynamical systems. To illustrate this fact, we have introduced sweeping gradient flows, a novel discontinuous gradient descent scheme for convex, time-varying constrained optimization problems. This scheme is based on well-established perturbed sweeping processes and allows us to apply our generalized tracking result in a technically sophisticated context. Lastly, our numerical examples confirm our theoretical tracking guarantees.

It remains open whether some of our assumptions can be weakened: For instance, by considering the difference between cost function values, instead of the distance to the optimizer, it is presumably possible to establish tracking guarantees in terms of the objective value without requiring strong convexity. Similarly, there might be room to relax the LICQ assumption terms of the objective value without requiring strong convexity.

Finally, although our focus is on time-varying optimization, we envision our quantified sensitivity bounds to apply to other robustness questions about online optimization algorithms.

**APPENDIX**

**Lemma 5.** [49] Ch. 2.17 Let $A \in \mathbb{R}^{n \times n}, B, C^T \in \mathbb{R}^{n \times m}$ and $D \in \mathbb{R}^{m \times m}$. If $A$ is non-singular, then the inverse exists if and only if $D - BA^{-1}C$ is invertible. Then, we have

$$
M_1 = A^{-1} + A^{-1}C(D - BA^{-1}C)^{-1}BA^{-1},
$$

$$
M_2 = -A^{-1}C(D - BA^{-1}C)^{-1},
$$

$$
M_3 = -(D - BA^{-1}C)^{-1}BA^{-1},
$$

$$
M_4 = (D - BA^{-1}C)^{-1}.
$$

**REFERENCES**

[1] A. Simonetto, “Time-Varying Convex Optimization via Time-Varying Averaged Operators,” ArXiv170407338 Math, Nov. 2017.

[2] Y. Tang and S. Low, “Distributed algorithm for time-varying optimal power flow,” in 2017 IEEE 56th Annual Conference on Decision and Control (CDC). Melbourne, VIC: IEEE, Dec. 2017, pp. 3264–3270.

[3] A. Lesage-Landry, I. Shames, and J. A. Taylor, “Predictive online convex optimization,” Automatica, vol. 113, p. 108771, Mar. 2020.

[4] A. Mokhtari, S. Shahrampour, A. Jadbabaie, and A. Ribeiro, “Online optimization in dynamic environments: Improved regret rates for strongly convex problems,” in 2016 IEEE 55th Conference on Decision and Control (CDC). Las Vegas, NV, USA: IEEE, Dec. 2016, pp. 7195–7201.

[5] A. Hausrwirth, F. Dörfler, and A. R. Teel, “Anti-Windup Approximations of Oblique Projected Dynamical Systems for Feedback-based Optimization,” ArXiv200300478 MathOC, 2020.

[6] Y. Tang, E. D’All’Anese, A. Bernstein, and S. Low, “Running Primal-Dual Gradient Method for Time-Varying Nonconvex Problems,” ArXiv181200613 Math, Dec. 2018.

[7] A. Hauswirth, I. Subotić, S. Bolognani, G. Hug, and F. Dörfler, “Time-varying Projected Dynamical Systems with Applications to Feedback Optimization of Power Systems,” in 2018 IEEE Conference on Decision and Control (CDC), Miami Beach, FL, Dec. 2018, pp. 3258–3263.

[8] M. Colombini, E. D’All’Anese, and A. Bernstein, “Online Optimization as a Feedback Controller: Stability and Tracking,” IEEE Trans. Control Netw. Syst., 2019.

[9] A. Bernstein, E. D’All’Anese, and A. Simonetto, “Online Optimization with Feedback,” ArXiv180405159 Math, Apr. 2018.

[10] J. Liu, J. Marecek, A. Simonetto, and M. Takač, “A Coordinate-Descent Algorithm for Tracking Solutions in Time-Varying Optimal Power Flows,” in Power Systems Computation Conference (PSCC), Jun. 2018.

[11] E. D’All’Anese and A. Simonetto, “Optimal Power Flow Pursuit,” IEEE Trans. Smart Grid, vol. 9, no. 2, pp. 942–952, Mar. 2018.

[12] A. Hauswirth, A. Zanardi, S. Bolognani, F. Dörfler, and G. Hug, “Online optimization in closed loop on the power flow manifold,” in 2017 IEEE PowerTech, Manchester, UK, Jun. 2017.

[13] V. Zavala and M. Anitescu, “Real-Time Nonlinear Optimization as a Generalized Equation,” SIAM J. Control Optim., vol. 48, no. 8, pp. 5444–5467, Jan. 2010.

[14] M. Diehl, H. G. Bock, J. P. Schlöder, R. Findeisen, Z. Nagy, and F. Allgöwer, “Real-time optimization and nonlinear model predictive control of processes governed by differential-algebraic equations,” Journal of Process Control, vol. 12, no. 4, pp. 577–585, Jun. 2002.

[15] C. Feller and C. Ebenbauer, “A stabilizing iteration scheme for model predictive control based on relaxed barrier functions,” Automatica, vol. 80, pp. 328–339, Jun. 2017.

[16] S. Rahili, W. Ren, and S. Ghapani, “Distributed Convex Optimization of Time-Varying Cost Functions with Swarm Tracking Behavior for Continuous-time Dynamics,” ArXiv150704773 Math, Jul. 2015.

[17] C. Sun, Z. Feng, and G. Hu, “Distributed Time-Varying Formation and Optimization for Uncertain Euler-Lagrange Systems,” in 2018 IEEE 8th Annual International Conference on CYBER Technology in Automation, Control, and Intelligent Systems (CYBER). Tianjin, China: IEEE, Jul. 2018, pp. 889–894.

[18] A. Y. Popkov, “Gradient Methods for Nonstationary Unconstrained Optimization Problems,” Autom Remote Control, vol. 66, no. 6, pp. 883–891, Jun. 2005.

[19] A. Simonetto and E. D’All’Anese, “Prediction-Correction Algorithms for Time-Varying Constrained Optimization,” IEEE Trans. Signal Process., vol. 65, no. 20, pp. 5481–5494, Oct. 2017.

[20] Y. Fazlyab, S. Paternain, V. M. Preciado, and A. Ribeiro, “Prediction-Correction Interior-Point Method for Time-Varying Convex Optimization,” IEEE Trans. Automat. Contr., vol. 63, no. 7, pp. 1973–1986, Jul. 2018.
