Extensions of Fast-Lipschitz Optimization

Martin Jakobsson, Sindri Magnússon, Carlo Fischione, *Member, IEEE*, and
Pradeep Chathuranga Weeraddana, *Member, IEEE*

Abstract—The need of fast distributed solvers for optimization problems in networked systems has motivated the recent development of the Fast-Lipschitz optimization framework. In such an optimization, problems satisfying certain qualifying conditions, such as monotonicity of the objective function and contractivity of the constraints, have a unique optimal solution obtained via fast distributed algorithms that compute the fixed point of the constraints. This paper extends the set of problems for which the Fast-Lipschitz framework applies. Existing assumptions on the problem form are relaxed and new and generalized qualifying conditions are established by novel results based on Lagrangian duality. It is shown for which cases of more constraints than decision variables, and less constraints than decision variables Fast-Lipschitz optimization applies. New results are obtained by imposing non strict monotonicity of the objective functions. The extended Fast-Lipschitz framework is illustrated by a number of examples, including network optimization and optimal control problems.

Index Terms—Fast-Lipschitz optimization, Lagrangian duality.

I. INTRODUCTION

FAST-Lipschitz optimization is a recently proposed class or problems where the special structure ensures that the optimal point is the fixed point of the constraints.[1] These problems can therefore be solved through simple and decentralized algorithms, making the framework an interesting alternative for distributed applications over networks. The special structure of Fast-Lipschitz problems makes them useful also in centralized applications, where it is guaranteed that the solutions can be obtained through simple system of equations that do not use the traditional Lagrangian approach.

In a general networked optimization problem, the network nodes or agents must coordinate their actions to optimize a network-wide objective function. When information such as nodes' objectives, constraints and decision variables are distributed among the nodes and physically scattered across the network, or if the amount of information is too large to collect centrally, it can be impractical or even impossible to centrally compute the solution. For example, collecting information in one place might be too expensive if the network has limited communication resources, or it may be too slow if the solution is needed at the local nodes in real time. In these situations, fast distributed solution algorithms must be used.

Distributed optimization has a long history, and much of the recent developments build upon the work of Tsitsiklis [2], [3]. Several approaches exist for solving these problems, such as primal and dual decomposition methods. In these methods, the primal or the dual problem is decomposed into local subproblems solved at the nodes. The subproblems are coordinated through a centralized master problem, which is usually solved by gradient or subgradient methods [4]. These methods have found many applications in network utility maximization, e.g., [5]–[7]. Due to the slow convergence of subgradient algorithms, recent works have explored higher order methods. For example, [8] replaces the subgradient step with a Newton-like step. Another decomposition approach which is faster and more robust than the standard decomposition methods is the alternating direction method of multipliers (ADMM). The method has recently attracted a substantial interest, especially for problems with large data sets [9]–[13].

Although the decomposition methods mentioned above distribute the computational workload over the network nodes, the subproblems must still be coordinated among the nodes. For example, dual decomposition based methods must update dual variables in a central master problem. This requires the network to iteratively 1) transfer information form all nodes to some central "master" node; 2) centrally update the dual variables; 3) broadcast the updated dual variables to all nodes of the network and back to 1) until convergence.

To avoid the centralized master node, peer-to-peer or multi-agent methods have recently been proposed to coordinate the subproblem through local neighbor interactions based on consensus algorithms, e.g., [14]–[18]. In these algorithms, nodes update their decision variables as convex combinations of the decision variables of their neighbors without a centralized master. In [19], the consensus algorithm has been combined with gradient descent to solve an unconstrained optimization problem where the objective is a sum of local, convex functions. The work is extended in [20] and [21], which investigate constraints and randomness. While the previous papers solve the primal problem, [22]–[24] use consensus-based algorithms for the dual problem. Higher order methods are considered also for peer-to-peer optimization, e.g., [25] solves an unconstrained primal problem, whereas [26] solves a linearly constrained dual problem, both by approximating Newton's algorithm through consensus.

Consensus based methods have many benefits, such as resilience to node failures and changing network topology. However, since every round of consensus requires message passings,

[1]We give a brief technical summary of Fast-Lipschitz optimization in Section II.

also these methods may suffer from communication overheads. A recent study of the tradeoff between communication and local computation can be found in [27]. The communication overhead is a problem especially in large scale distributed networks or wireless sensor networks, where the energy expenditure for communication can be orders of magnitude larger than the energy for computation [28].

The methods discussed thus far assume convex problems. There are other classes of algorithms that do not rely necessarily on convexity, but on other structural properties. Three such classes are abstract optimization [29], which generalizes linear optimization, monotonic optimization [30]–[32], where the monotonicity of the objective function is used to iteratively refine a solution within bounds of the feasible region, and Interference Function optimization [33]–[37], which is the fundamental framework to solve radio power control problems over wireless networks. Given the importance of Interference Function optimization as a precursor of Fast Lipschitz optimization, and considering that we will give some application examples later on in this paper, we give below some technical detail on such an optimization framework.

In *Interference Function optimization*, typical radio power control problems are solved in a simple distributed (as in decentralized) way. In such an optimization approach, it is assumed that the nodes of a wireless network transmit signals of a certain level of radio power, say node i transmits with power p_i. The signal is received at the intended receiver corrupted by multiplicative wireless channel attenuations and additive interference by other transmitter nodes. The level of power that the transmit power of node i has to overcome at the receiver in order to get the signal decoded is usually denoted by $\mathcal{I}_i(\mathbf{p})$, the *interference function* of transmitter i [36], where $\mathbf{p} = [p_1, p_2, \ldots, p_n]^T$ is the vector of all transmit radio powers of the n transmitters in the wireless network. The goal of the basic power control problem is to minimize the radio powers p_i, while overcoming the interference at each receiver, i.e.,

$$\min_{\mathbf{p}} \quad \mathbf{p}$$
$$\text{s.t.} \quad p_i \geq \mathcal{I}_i(\mathbf{p}) \quad \forall i. \tag{1}$$

Note that the problem above is a vector optimization, where the minimization is carried out with respect to the non-negative orthant (see Section II-B for details). Roughly speaking, the minimization of problem (1) makes all the components of $\mathbf{p}$ become small simultaneously. The solution of problem (1) is a particularly successful instance of distributed optimization. Affine versions of $\mathcal{I}_i(\mathbf{p})$ are the simplest and best studied type of interference function, but in theory one can consider functions $\mathcal{I}(\mathbf{p})$ of any form. The first distributed algorithm to solve such a problem was proposed in [33], and improved in [34] and [35]. The algorithm was later generalized to the Interference Function framework by Yates [36]. In this framework, a function $\mathcal{I}(\mathbf{p})$ is called *standard* if, for all $\mathbf{p} \geq 0$, it fulfils

- *Monotonicity*: If $\mathbf{p} \geq \mathbf{p}'$, then $\mathcal{I}(\mathbf{p}) \geq \mathcal{I}(\mathbf{p}')$,
- *Scalability*: For all $\alpha > 1, \alpha\mathcal{I}(\mathbf{p}) > \mathcal{I}(\alpha\mathbf{p})$.

When problem (1) above is feasible, and the functions $\mathcal{I}_i(\mathbf{p})$ are standard, the unique optimal solution is given by the fixed point of the iteration

$$p_i^{k+1} := \mathcal{I}_i(\mathbf{p}^k) \tag{2}$$

or $\mathbf{p}^{k+1} := \mathcal{I}(\mathbf{p}^k)$ in vector form. Here, p_i^k is the power of transmitter i at time k and $\mathbf{p}^k = [p_1^k, p_2^k, \ldots, p_n^k]^T$. The computation of the optimal solution by these iterations is much simpler than using the classical parallelization and decomposition methods. This is because there is no longer a need to centrally collect, compute and redistribute the coupling variables of the problem since $\mathcal{I}_i(\mathbf{p}^k)$ can be know locally at node i [36]. Even in a centralized setting, iteration (2) is simpler than traditional Lagrangian methods, since no dual variables need to be stored and manipulated. The iterations require only that every node successively updates its transmit power by using local knowledge of other nodes' current decision variables (radio powers). Another advantage is that the algorithm converges even though such a knowledge is delayed, i.e., when the decision variables p_j^k of other nodes come with some delay at node i [38].

Extensions of the Interference Function framework have been proposed in [39], where optimization problems whose interference functions are not standard by Yates' definition are investigated. Instead, they introduce Type-II standard functions, which for all $\mathbf{p} \geq \mathbf{0}$ fulfill

- *Type-II monotonicity*: If $\mathbf{p} \leq \mathbf{p}'$, then $\mathcal{I}(\mathbf{p}) \geq \mathcal{I}(\mathbf{p}')$.
- *Type-II scalability*: $\forall \alpha > 1, \mathcal{I}(\alpha\mathbf{p}) > (1/\alpha)\mathcal{I}(\mathbf{p})$.

These functions are shown to have the same fixed point properties as Yates' standard functions, i.e., problem (1) with Type-II standard constraints can be solved through repeated iterations of the constraints (2).

Fast-Lipschitz optimization is a natural generalization of the Interference Function approach on how to solve distributed optimization problems over networks by using fixed point iterations similar to (2), but when the constraints are neither standard nor Type-II standard [28]. It also considers more general objective functions $\mathbf{g}_0(\mathbf{p})$ for problem (1). The framework considers a class of possibly non-convex and multi-objective problems with monotonic objective functions. These problems have unique Pareto optimal solutions, well defined by a contractive system of equations formed by the problem constraints. Therefore, Fast-Lipschitz problems are solved without having to introduce Lagrangian functions and dual variables or consensus based iterations. This makes the framework particularly well suited when highly decentralized solutions, with few coordination messages, are required. This is important in typical areas such as wireless sensor networks and in multi-agent systems.

In this paper, we substantially extend the class of problems that are currently solvable with the Fast-Lipschitz framework. In particular, we 1) introduce a new qualifying condition that unifies and extends the existing conditions of [28], 2) consider problems with more, or less, constraints than variables, which was not considered in [28], and 3) study objective functions that are not strictly monotonic, which was not considered in [28].

The remainder of this paper is organized as follows. In Section II we clarify our notation and give a brief overview of the current state of Fast-Lipschitz optimization. Section III presents a new, general qualifying condition for Fast-Lipschitz optimization. The new qualifying condition is proved (using an approach different from [28]) in Section III-C. In Section IV we give special cases of the general qualifying condition that have much less analytical and computational complexity, and highlight the connection of these special cases to existing qualifying conditions and related work. Furthermore, Section V

relaxes some of the requirements of Fast-Lipschitz form by considering problems with more constraints than variables in Section V-A, and problems with less constraints than variables in Section V-B, while Section V-C refines some of the requirements on the objective function. Section VI-A features an example that illustrates some of the new results of this paper. In Section VI-C the new results developed in this paper are applied to a family of optimal control problems, where the problem structure is utilized to determine the optimal solution without computations. Finally, the paper is concluded in Section VII.

II. PRELIMINARIES

This section clarifies notation and recalls Fast-Lipschitz optimization to provide the essential background definitions for the core contribution of this paper.

A. Notation

Vectors and matrices are denoted by bold lower and upper case letters, respectively. The components of a vector $\mathbf{x}$ are denoted x_i or $[\mathbf{x}]_i$. Similarly, the elements of the matrix $\mathbf{A}$ are denoted A_{ij} or $[\mathbf{A}]_{ij}$. The transpose of a vector or matrix is denoted $\cdot^T$. $\mathbf{I}$ and $\mathbf{1}$ denote the identity matrix and the vector of ones. A vector or matrix where all elements are zero is denoted by $\mathbf{0}$.

The gradient $\nabla \mathbf{f}(\mathbf{x})$ is the transpose of the Jacobian matrix, i.e., $[\nabla \mathbf{f}(\mathbf{x})]_{ij} = \partial f_j(\mathbf{x})/\partial x_i$, whereas $\nabla_i \mathbf{f}(\mathbf{x})$ denotes the ith row of $\nabla \mathbf{f}(\mathbf{x})$. Note that $\nabla \mathbf{f}(\mathbf{x})^k = (\nabla \mathbf{f}(\mathbf{x}))^k$, which is not to be confused with the kth derivative. The spectral radius of $\mathbf{A}$ is denoted $\rho(\mathbf{A})$. Vector norms are denoted $||| \cdot |||$ and matrix norms are denoted $\| \cdot \|$. Unless specified $||| \cdot |||$ and $\| \cdot \|$ denote arbitrary norms. $\||\mathbf{A}\|| = \max_i \sum_j |A_{ij}|$ is the norm induced by the ℓ_∞ vector norm.

All inequalities are intended *element-wise*, i.e., they have nothing to do with positive definiteness.

B. Vector Optimization and Pareto Optimality

In this paper we are concerned with maximization of vector valued objective functions $\mathbf{f}_0(\mathbf{x}) \in \mathbb{R}^m$. A vector optimization problem involves a proper cone $\mathcal{K}$ (see, e.g., [40, Section 4.7]). In this paper we focus on the case when $\mathcal{K}$ is non-negative orthant

$$\mathcal{K} = \mathbb{R}_+^m \triangleq \{\mathbf{y} : y_i \geq 0, \quad \forall i = 1, \ldots, m\}.$$

This maximization of a vector is formally expressed as

$$\text{maximize } (\text{with respect to } \mathbb{R}_+^m) \ \mathbf{f}_0(\mathbf{x}).^2 \quad (3)$$

The goal of the maximization (3) is to find the decision vector $\mathbf{x}$ such that the components of $\mathbf{f}_0(\mathbf{x})$ are as big as possible with respect to the cone $\mathbb{R}_+^m$. In particular, when comparing two vectors $\mathbf{x}$ and $\mathbf{y}$ with respect to the cone $\mathbb{R}_+^m$, we say $\mathbf{x} \succeq_{\mathbb{R}_+^m} \mathbf{y}$ if $\mathbf{x} - \mathbf{y} \in \mathbb{R}_+^m$ and $\mathbf{x} \preceq_{\mathbb{R}_+^m} \mathbf{y}$ if $\mathbf{y} - \mathbf{x} \in \mathbb{R}_+^m$. Note that this corresponds exactly to the component-wise inequalities ($\mathbf{y} \geq \mathbf{x}$

and $\mathbf{y} \leq \mathbf{x}$), and for this reason we will use $\geq$ rather than $\succeq_{\mathbb{R}_+^m}$ throughout the rest of the paper. Unlike scalars, where it must hold that either $a \geq b$, or $a < b$, two vectors might not be comparable. For example, for $\mathbf{x} = [1 \ 2]^T$ and $\mathbf{y} = [3 \ 1]^T$, we have $\mathbf{x} \not\geq \mathbf{y}$ and $\mathbf{y} \not\geq \mathbf{x}$.

Based on the discussion above, $\mathbf{x}_1$ is preferable to $\mathbf{x}_2$ in problem (3) if $\mathbf{f}_0(\mathbf{x}_1) \geq \mathbf{f}_0(\mathbf{x}_2)$. A feasible decision variable $\hat{\mathbf{x}}$ is said to be *Pareto optimal* if there is no other feasible vector $\mathbf{x}$ such that $\mathbf{f}_0(\mathbf{x}) \geq \mathbf{f}_0(\hat{\mathbf{x}})$. A problem can have several Pareto optimal points—in this case each Pareto optimal point is incomparable to any other Pareto optimal points, but preferable to any point that is not Pareto optimal. If a problem only has one Pareto optimal point $\mathbf{x}^\star$, then $\mathbf{x}^\star$ is the unambiguously best choice among the decision vectors.

Scalarization is a useful technique for generating Pareto optimal points. Scalarization is performed by picking a weight vector $\boldsymbol{\mu}$ in the interior of $\mathbb{R}_+^m$, i.e., $\boldsymbol{\mu} > \mathbf{0}$, and solving the scalar optimization problem

$$\max \boldsymbol{\mu}^T \mathbf{f}_0(\mathbf{x}) = \sum_{i=1}^m \mu_i [\mathbf{f}_0(\mathbf{x})]_i. \quad (4)$$

Any point $\hat{\mathbf{x}}$ that is optimal in problem (4) is Pareto optimal in (3). Although scalarization does not necessarily generate all Pareto optimal points, one can show that if a point $\mathbf{x}^\star$ is optimal for *all* scalarization vectors $\boldsymbol{\mu} > \mathbf{0}$, then $\mathbf{x}^\star$ is the unique Pareto optimal point.

C. Fast-Lipschitz Optimization

We will now give a formal definition of Fast-Lipschitz problems. For a thorough discussion of Fast-Lipschitz properties we refer the reader to the above mentioned paper.

Definition 1: A problem is said to be on *Fast-Lipschitz form* if it can be written

$$\begin{aligned} \max \quad & \mathbf{f}_0(\mathbf{x}) \\ \text{s.t.} \quad & x_i \leq f_i(\mathbf{x}) \quad \forall i \in \mathbb{A} \\ & x_i = f_i(\mathbf{x}) \quad \forall i \in \mathbb{B} \end{aligned} \quad (5)$$

where

- $\mathbf{f}_0 : \mathbb{R}^n \to \mathbb{R}^m$ is a differentiable scalar ($m = 1$) or vector valued ($m \geq 2$) function,
- $\mathbb{A}$ and $\mathbb{B}$ are complementary subsets of $\{1, \ldots, n\}$,
- $f_i : \mathbb{R}^n \to \mathbb{R}$ are differentiable functions.

We will refer to problem (5) as our main problem. From the individual constraint functions we form the vector valued function $\mathbf{f} : \mathbb{R}^n \to \mathbb{R}^n$ as $\mathbf{f}(\mathbf{x}) = [f_1(\mathbf{x}) \cdots f_n(\mathbf{x})]^T$.

Remark 2: The characteristic feature of Fast-Lipschitz form is a pairing such that each variable x_i is associated to one constraint $f_i(\mathbf{x})$. The form $x \leq f(\mathbf{x})$ is general, since any constraint on canonical form, $g(\mathbf{x}) \leq 0$, can be written $x \leq x - \gamma g(\mathbf{x})$ for some positive constant γ.

Definition 3: For the rest of the paper, we will restrict our attention to a *bounding box* $\mathcal{D} = \{\mathbf{x} \in \mathbb{R}^n | \mathbf{a} \leq \mathbf{x} \leq \mathbf{b}\}$. We assume $\mathcal{D}$ contains all candidates for optimality and that $\mathbf{f}$ maps $\mathcal{D}$ into $\mathcal{D}$, $\mathbf{f} : \mathcal{D} \to \mathcal{D}$. This box arise naturally in practice, since any real-world quantity or decision must be bounded.

Definition 4: A problem is said to be *Fast-Lipschitz* when it can be written in Fast-Lipschitz form and admits a unique

---

²In this paper we only consider optimization with respect to the cone $\mathbb{R}_+^m$. Therefore, for notational simplicity we simply write $\max \mathbf{f}_0(\mathbf{x})$, and the vector optimization (3) should be understood whenever $\mathbf{f}_0(\mathbf{x})$ is vector valued.

Pareto optimal solution $\mathbf{x}^\star$, defined as the unique solution to the system of equations

$$\mathbf{x}^\star = \mathbf{f}(\mathbf{x}^\star). \qquad (6)$$

A problem written in Fast-Lipschitz form is not necessarily Fast-Lipschitz. The following *qualifying conditions* guarantee that a problem in Fast-Lipschitz form is Fast-Lipschitz.

Old Qualifying Conditions: For all $\mathbf{x}$ in $\mathcal{D}$, $\mathbf{f}_0(\mathbf{x})$ and $\mathbf{f}(\mathbf{x})$ should be everywhere differentiable and fulfill at least one of the following cases (e.g., (0) and (i), or (0) and (ii)):

$$
\begin{array}{rll}
(\mathbf{0}) & \nabla\mathbf{f}_0(\mathbf{x}) > \mathbf{0} & (7a) \\
\text{AND } (\mathbf{i.a}) & \nabla\mathbf{f}(\mathbf{x}) \geq \mathbf{0} & (7b) \\
(\mathbf{i.b}) & \||\nabla\mathbf{f}(\mathbf{x})\|| < 1 & (7c) \\
\text{OR } (\mathbf{ii.a}) & \mathbf{f}_0(\mathbf{x}) = c\mathbf{1}^T\mathbf{x}, \text{ with } c > 0 & (7d) \\
(\mathbf{ii.b}) & \nabla\mathbf{f}(\mathbf{x}) \leq \mathbf{0} & (7e) \\
& (\text{or more generally, } \nabla\mathbf{f}(\mathbf{x})^2 \geq \mathbf{0}) & \\
(\mathbf{ii.c}) & \||\nabla\mathbf{f}(\mathbf{x})\||_\infty < 1 & (7f) \\
\text{OR } (\mathbf{iii.a}) & \mathbf{f}_0(\mathbf{x}) \in \mathbb{R} & (7g) \\
\\
(\mathbf{iii.b}) & \||\nabla\mathbf{f}(\mathbf{x})\||_\infty < \dfrac{\bar{\delta}}{\bar{\delta}+\bar{\Delta}} & \\
& \text{where } \begin{array}{l} \bar{\delta} \triangleq \min_i \min_{\mathbf{x}\in\mathcal{D}} \nabla_i\mathbf{f}_0(\mathbf{x}) \\ \bar{\Delta} \triangleq \max_i \max_{\mathbf{x}\in\mathcal{D}} \nabla_i\mathbf{f}_0(\mathbf{x}). \end{array} & (7h)
\end{array}
$$

Theorem 5 ([28, Theorem 3.3]): A problem in Fast-Lipschitz form that fulfills any pair of the Old Qualifying Conditions is Fast-Lipschitz, i.e., it has a unique Pareto optimal point given by $\mathbf{x}^\star = \mathbf{f}(\mathbf{x}^\star)$.

Once it is known that a problem is Fast-Lipschitz, computing the solution becomes a matter of solving the system of (6), which in general is much easier than solving an optimization problem using Lagrangian multipliers. This is particularly evident when $\mathbf{f}(\mathbf{x})$ is contractive on $\mathcal{D}$, a property assured by the qualifying conditions. In this case, the iterations $\mathbf{x}^{k+1} := \mathbf{f}(\mathbf{x}^k)$ converge geometrically to the optimal point $\mathbf{x}^\star$, starting from any initial point $\mathbf{x}^0 \in \mathcal{D}$.

Fast-Lipschitz optimization problems need not be convex, but convex problems that fulfill the qualifying conditions can be rewritten and solved with the Fast-Lipschitz framework. For example, this is true for any problem where the constraints are standard:

Proposition 6 ([41, Theorem 4.2]): If problem (1) is feasible, and the constraints standard, then the problem is Fast-Lipschitz.

This concludes the preliminary part of the paper. We are now ready to present the core contributions of this paper.

III. A GENERAL QUALIFYING CONDITION

This section presents a new qualifying condition, that generalizes and unifies the conditions of Section II-C. The new condition is introduced in Section III-A, together Theorem 7, which formally states the role of the condition. Section III-B shows that problems fulfilling the general qualifying condition of Section III-A have contractive constraints. These results enable finding the optimal point through fixed point iterations, and they also function as a preliminary result to the proof of Theorem 7 in Section III-C.

From this point and onwards we will do small change of terminology. We will call each set of related conditions a "Qualifying Condition," rather than a "case" of the qualifying conditions. We will still use the notion of "case" when referring

to the Old Qualifying Conditions (7). For example, "case (i)" will refer to the groups (7a)–(7c) of the old qualifying conditions, while "case (ii)" refers to (7a) and (7d)–(7f).

A. New Qualifying Conditions

Consider once again the optimization problem (5) on Fast-Lipschitz form, surrounded by the bounding box $\mathcal{D}$. Just as in Section II-C, the qualifying conditions presented in this section are used to ensure that a problem in Fast-Lipschitz form also is Fast-Lipschitz. In the upcoming qualifying conditions, we will use the ratio

$$\mathfrak{q}(\mathbf{x}) \triangleq \min_j \frac{\min_i [\nabla\mathbf{f}_0(\mathbf{x})]_{ij}}{\max_i [\nabla\mathbf{f}_0(\mathbf{x})]_{ij}}. \qquad (8)$$

The value of $\mathfrak{q}(\mathbf{x})$ is the smallest ratio of any two elements from the same column of $\nabla\mathbf{f}_0(\mathbf{x})$. When $\nabla\mathbf{f}_0(\mathbf{x}) \geq \mathbf{0}$, $\mathfrak{q}(\mathbf{x})$ is always non-negative. Moreover, $\mathfrak{q}(\mathbf{x}) \leq 1$ by construction, with equality if and only if all rows of $\nabla\mathbf{f}_0(\mathbf{x})$ are identical. In fact, $\mathfrak{q}(\mathbf{x})$ can be seen as a penalty for when the objective function gradient points in a different direction than the vector $\mathbf{1}$.

We are now ready to state the new qualifying conditions. We will begin with the most general form of the qualifying conditions. Special cases of this condition will be discussed in Section IV.

General Qualifying Condition	
GQC	**(GQC.a)** $\nabla\mathbf{f}_0(\mathbf{x}) \geq \mathbf{0}$ with non-zero rows **(GQC.b)** $\||\nabla\mathbf{f}(\mathbf{x})\|| < 1$ There exists a $k \in \{1, 2, \dots\} \cup \infty$ such that **(GQC.c)** When $k < \infty$, then $\qquad \nabla\mathbf{f}(\mathbf{x})^k \geq \mathbf{0}$ **(GQC.d)** When $k > 1$, then $\qquad \left\|\sum_{l=1}^{k-1} \nabla\mathbf{f}(\mathbf{x})^l\right\|_\infty < \mathfrak{q}(\mathbf{x})$

In the condition above, we allow the parameter k to be any positive integer, or infinity. Each value of k corresponds to a different case in the proof of Theorem 7 below. In the extremes $k = \infty$ and $k = 1$, there is no need to fulfill conditions (GQC.c) and (GQC.d) respectively. Both these cases will be discussed further in Section IV.

We now give Theorem 7, which plays a role analogous to that of Theorem 5 in Section II-C.

Theorem 7: Assume problem (5) is feasible, and that GQC holds for every $\mathbf{x} \in \mathcal{D}$. Then, the problem is Fast-Lipschitz, i.e., the unique Pareto optimal solution is given by $\mathbf{x}^\star = \mathbf{f}(\mathbf{x}^\star)$.

Proof: The theorem is proved in Section III-C. ∎

Remark 8: It should be emphasized that all qualifying conditions of this paper are only sufficient, i.e., a problem that does not fulfil the qualifying conditions can still be Fast-Lipschitz. For example, by considering certain transformations of the constraint functions, it is possible to relax condition the condition (GQC.d) that contains the norm $\|| \cdot \||_\infty$. One can also relax certain requirements on the Fast-Lipschitz form of problem (5), (e.g., by considering problems with more variables than constraints or problem where the objective function only depends on a subset of the variables (see Section V).

B. Contraction Properties of Fast-Lipschitz Problems

In this subsection we briefly discuss the contractiveness of $\mathbf{f}(\mathbf{x})$. This is important since it allows a Fast-Lipschitz optimization problem to be solved through fixed point methods. These results will also be used in the proof of Theorem 7 in the next subsection.

Once optimization problem (5) is shown to be Fast-Lipschitz, solving it becomes a matter of finding the point $\mathbf{x}^\star = \mathbf{f}(\mathbf{x}^\star)$. In a centralized setting, one can use any suitable method for solving such a system of equations, e.g., Newton like methods. However, if $\mathbf{f}(\mathbf{x})$ is contractive, the simplest way to solve $\mathbf{x}^\star = \mathbf{f}(\mathbf{x}^\star)$ is to repeatedly evaluate $\mathbf{f}(\mathbf{x})$. This method works both in a centralized and a distributed setting. Compared to other distributed methods, it has the benefit of being totally decentralized, i.e., there is no master problem or coordinating node. Furthermore, the iterations converge even when some of the nodes, due to dropped or delayed packets, only have access to outdated information of the neighbors' decisions, see [28, Proposition 3.6] or [38] for details. We show below that if the general qualifying condition GQC holds, then $\mathbf{f}(\mathbf{x})$ is contractive.

Proposition 9 ([38, Sec. 3.1]): Let $\mathbf{f}$ be a mapping from a closed subset of $\mathbb{R}^n$ into itself, $\mathbf{f}: \mathcal{X} \to \mathcal{X}$. If there is a norm $\|.\|$ and a scalar $\alpha < 1$ such that $\|\mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{y})\| \leq \alpha \|\mathbf{x} - \mathbf{y}\|$ for all $\mathbf{x}, \mathbf{y} \in \mathcal{X}$, then $\mathbf{f}(\mathbf{x})$ is a contraction mapping. As a result:

- $\mathbf{x}^\star = \mathbf{f}(\mathbf{x}^\star)$ is the unique fixed point of $\mathbf{f}$ in $\mathcal{X}$.
- For every initial point $\mathbf{x}^0 \in \mathcal{X}$, the sequence $\mathbf{x}^{k+1} := \mathbf{f}(\mathbf{x}^k)$ converges linearly to $\mathbf{x}^\star$.

Since we know $\mathbf{f}: \mathcal{D} \to \mathcal{D}$, $\mathbf{f}$ is a contraction mapping if we can find a vector norm such that $\|\mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{y})\| \leq \alpha \|\mathbf{x} - \mathbf{y}\|$ for all $\mathbf{x}, \mathbf{y} \in \mathcal{D}$.

Lemma 10: If the general qualifying condition GQC holds, $\mathbf{f}(\mathbf{x})$ is contractive.

Proof: Parameterize the line between $\mathbf{x}, \mathbf{y} \in \mathcal{D}$ by $\mathbf{u}(t) = t\mathbf{x} + (1-t)\mathbf{y}, 0 \leq t \leq 1$, and let $\mathbf{g}(t) = \mathbf{f}(\mathbf{u}(t))$. Then, $\mathrm{d}\mathbf{g}(t)/\mathrm{d}t = \nabla\mathbf{f}(\mathbf{u}(t))^T(\mathbf{x} - \mathbf{y})$, wherefore

$$\mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{y}) = \mathbf{g}(1) - \mathbf{g}(0) = \int_0^1 \frac{\mathrm{d}\mathbf{g}(t)}{\mathrm{d}t}\mathrm{d}t0$$

$$= \int_0^1 \nabla\mathbf{f}(\mathbf{u}(t))^T \, \mathrm{d}t \, (\mathbf{x} - \mathbf{y}) \triangleq \mathbf{A}(\mathbf{x} - \mathbf{y})$$

where each element in $\mathbf{A}$ is the integral of the corresponding element in $\nabla\mathbf{f}(\mathbf{u}(t))$. Let $\|\cdot\|_a$ be the matrix norm that satisfies condition (GQC.b), and define $\|\cdot\|_b$ such that $\||\mathbf{A}\||_b \triangleq \||\mathbf{A}^T\||_a$. It is straight-forward to show that $\|\cdot\|_b$ inherits the matrix norm properties of $\|\cdot\|_a$. We can now bound

$$\||\mathbf{A}\||_b = \left\|\left\|\left\|\int_0^1 \nabla\mathbf{f}(\mathbf{u}(t))^T \, \mathrm{d}t\right\|\right\|\right\|_b \leq \int_0^1 \left\|\left\|\left\|\nabla\mathbf{f}(\mathbf{u}(t))^T\right\|\right\|\right\|_b \mathrm{d}t$$

$$\leq \int_0^1 \max_{\mathbf{z} \in \mathcal{D}} \||\nabla\mathbf{f}(\mathbf{z})^T\||_b \, \mathrm{d}t = \max_{\mathbf{z} \in \mathcal{D}} \||\nabla\mathbf{f}(\mathbf{z})^T\||_b$$

$$= \max_{\mathbf{z} \in \mathcal{D}} \||\nabla\mathbf{f}(\mathbf{z})\||_a \triangleq \alpha.$$

The first inequality above is the triangle inequality. The second inequality holds since $\mathcal{D}$ is convex whereby $\mathbf{u}(t) \in \mathcal{D}$ for all t.

The maximum exists since $\mathcal{D}$ is compact, and $\alpha < 1$ by (GQC.b).

From [42, Theorem 5.6.26], we know that there exists an *induced* matrix norm $\||.\||_V$ such that $\||\mathbf{M}\||_V \leq \||\mathbf{M}\||_a$ for every matrix $\mathbf{M} \in \mathbb{R}^{n \times n}$. Let $\|.\|_v$ be the vector norm that induces $\||.\||_V$. By the properties of induced matrix norms we get

$$\|\mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{y})\|_v = \|\mathbf{A}(\mathbf{x} - \mathbf{y})\|_v \leq \||\mathbf{A}\||_V \|(\mathbf{x} - \mathbf{y})\|_v$$
$$\leq \||\mathbf{A}\||_a \|(\mathbf{x} - \mathbf{y})\|_v \leq \alpha \|(\mathbf{x} - \mathbf{y})\|_v$$

as desired. Since $\mathbf{f}(\mathbf{x}): \mathcal{D} \to \mathcal{D}$, $\mathbf{f}(\mathbf{x})$ is a contraction mapping. This concludes the proof. ∎

We are now ready for the main proof of the paper.

C. Proof of Theorem 7

In this section we prove Theorem 7, which is one of the main contributions of this paper. The proof will be given as a series of lemmas as outlined in the following.

Proof of Theorem 7: Consider optimization problem (5). When the general qualifying condition GQC holds for all $\mathbf{x} \in \mathcal{D}$, then the following steps ensure that $\mathbf{x}^\star = \mathbf{f}(\mathbf{x}^\star)$ is the unique optimal solution.

1) First, we restrict ourselves to optimization problems on Fast-Lipschitz form with only inequality constraints, without loss of generality by Lemma 11 below.
2) The inequality-only constrained optimization problem allows us to show that all feasible points of the optimization problem are regular [43], wherefore any optimal point $\hat{\mathbf{x}}$ must fulfill the KKT-conditions, see Lemma 13 below.
3) Any point that fulfills the KKT-conditions must be a fixed point of $\mathbf{f}(\mathbf{x})$, see Lemma 15.
4) Finally, we show that there exists a unique fixed point $\mathbf{x}^\star = \mathbf{f}(\mathbf{x}^\star)$ by Proposition 9 and Lemma 10. Therefore, $\mathbf{x}^\star$ is the only point fulfilling the KKT-conditions and $\mathbf{x}^\star$ must be the optimum. ∎

The first lemma allows us to focus on problems consisting only of inequality constraints. This is an important step that permits us to use the KKT conditions (see, e.g., [38, 3.1.1]) to establish the existence an uniqueness of optimal solutions.

Lemma 11: If the inequality-only constrained optimization problem

$$\begin{aligned} \max \quad & \mathbf{f}_0(\mathbf{x}) \\ \text{s.t.} \quad & x_i \leq f_i(\mathbf{x}) \quad \forall i \in \{1, \ldots, n\} \end{aligned} \tag{9}$$

is Fast-Lipschitz, then so is any problem

$$\begin{aligned} \max \quad & \mathbf{f}_0(\mathbf{x}) \\ \text{s.t.} \quad & x_i \leq f_i(\mathbf{x}) \quad \forall i \in \mathbb{A} \quad \mathbb{A} \cup \mathbb{B} = \{1, \ldots, n\} \\ & x_i = f_i(\mathbf{x}) \quad \forall i \in \mathbb{B}, \quad \mathbb{A} \cap \mathbb{B} = \varnothing \end{aligned} \tag{10}$$

obtained by switching any number of the inequalities for equalities.

Proof: Let $\mathcal{F}_9$ and $\mathcal{F}_{10}$ be the feasible regions of problem (9) and problem (10) respectively. The point $\mathbf{x}^\star = \mathbf{f}(\mathbf{x}^\star)$ is feasible in both problems and by definition uniquely optimal for problem (9) since it is Fast-Lipschitz. Suppose, contrary to the lemma, that problem (10) is not Fast-Lipschitz. Then there exists some feasible point $\hat{\mathbf{x}} \in \mathcal{F}_{10} \subset \mathcal{F}_9$ such that $\mathbf{f}_0(\hat{\mathbf{x}}) \geq \mathbf{f}_0(\mathbf{x}^\star)$ which contradicts the unique optimality of $\mathbf{x}^\star$ in problem (9). ∎

With Lemma 11 in mind, we develop the rest of the proof of Theorem 7 by focusing on the inequality-only constrained problem (9) instead of main problem (5). The inequality-only problem on canonical form is

$$\min \quad -\mathbf{f}_0(\mathbf{x})$$
$$\text{s.t.} \quad g_i(\mathbf{x}) = x_i - f_i(\mathbf{x}) \leq 0 \quad \forall i. \tag{11}$$

Definition 12: In problem (11), a point $\mathbf{x}$ is *regular* if the gradients of all active constraints at $\mathbf{x}$ form a linearly independent set (see [43]).

Lemma 13: If problem (9) fulfills the general qualifying condition GQC, then every point $\mathbf{x} \in \mathcal{D}$ in problem (11) is regular.

Proof: The gradients of the individual constraints $g_i(\mathbf{x})$ are the columns of $\nabla \mathbf{g}(\mathbf{x}) = \mathbf{I} - \nabla \mathbf{f}(\mathbf{x})$. Since condition (GQC.b) implies $\rho(\nabla \mathbf{f}(\mathbf{x})) \leq \|\|\nabla \mathbf{f}(\mathbf{x})\|\| < 1$, the eigenvalues of $\nabla \mathbf{g}(\mathbf{x})$ lie in a ball of radius $\rho(\nabla \mathbf{f}(\mathbf{x})) < 1$, centered at 1. Hence, no eigenvalue of $\nabla \mathbf{g}(\mathbf{x})$ is zero and $\nabla \mathbf{g}(\mathbf{x})$ is invertible, wherefore the constraint gradients $\nabla \mathbf{g}_i(\mathbf{x})$ (the columns of $\nabla \mathbf{g}(\mathbf{x})$) form a linearly independent set. ■

We now scalarize problem (11), by considering

$$\min \quad -\boldsymbol{\mu}^T \mathbf{f}_0(\mathbf{x})$$
$$\text{s.t.} \quad g_i(\mathbf{x}) = x_i - f_i(\mathbf{x}) \leq 0 \quad \forall i. \tag{12}$$

for a positive vector $\boldsymbol{\mu} \in \mathbb{R}^m$ (see Section II-B). Since an arbitrary scaling of $\boldsymbol{\mu}$ does not change the solution of the problem, we restrict ourselves to $\boldsymbol{\mu}$ fulfilling $\sum_k \mu_k = 1$.

Introduce dual variables $\boldsymbol{\lambda} \in \mathbb{R}^n$, and form the Lagrangian function $L(\mathbf{x}, \boldsymbol{\lambda}) = -\boldsymbol{\mu}^T \mathbf{f}_0(\mathbf{x}) + \boldsymbol{\lambda}^T (\mathbf{x} - \mathbf{f}(\mathbf{x}))$. Since every $\mathbf{x} \in \mathcal{D}$ in problem (12) is regular (Lemma 13), any pair $(\hat{\mathbf{x}}, \hat{\boldsymbol{\lambda}})$ of locally optimal variables must satisfy the KKT-conditions (see e.g., [38, 3.1.1]). In particular, $\hat{\mathbf{x}}$ must be a minimizer of $L(\mathbf{x}, \hat{\boldsymbol{\lambda}})$, which requires

$$\nabla_{\mathbf{x}} L(\hat{\mathbf{x}}, \hat{\boldsymbol{\lambda}}) = -\nabla \mathbf{f}_0(\hat{\mathbf{x}})\boldsymbol{\mu} + \hat{\boldsymbol{\lambda}} - \nabla \mathbf{f}(\hat{\mathbf{x}})\hat{\boldsymbol{\lambda}} = \mathbf{0} \tag{13}$$

and complementarity must hold, i.e.,

$$\hat{\lambda}_i (\hat{x}_i - f_i(\hat{\mathbf{x}})) = 0 \quad \forall i. \tag{14}$$

We will soon show that GQC implies $\hat{\boldsymbol{\lambda}} > \mathbf{0}$. To this end, the following remark is useful.

Remark 14: Let $\boldsymbol{\lambda} = \mathbf{A}\mathbf{c}$. If $\mathbf{c} > \mathbf{0}$ and $\mathbf{A} \geq \mathbf{0}$ with non-zero rows, then $\boldsymbol{\lambda} > \mathbf{0}$.

The statement above is trivial, but we give it as Remark 14 since we will refer to it several times throughout the paper. Note that $\mathbf{A} \geq \mathbf{0}$ and $\mathbf{c} > \mathbf{0}$ is not sufficient for $\mathbf{A}\mathbf{c} > \mathbf{0}$, this is the reason that condition (GQC.a) requires non-zero rows.

The following lemma is the main part of the proof of Theorem 7, and establishes that the optimal dual variable is strictly positive.

Lemma 15: Whenever GQC holds, any pair $(\hat{\mathbf{x}}, \hat{\boldsymbol{\lambda}})$ with $\hat{\mathbf{x}} \in \mathcal{D}$ satisfying (13) must have $\hat{\boldsymbol{\lambda}} > \mathbf{0}$.

Proof: For notational convenience, let us fix one $\mathbf{x} \in \mathcal{D}$, which we denote $\hat{\mathbf{x}}$, and introduce

$$\mathbf{A} \stackrel{\triangle}{=} \nabla \mathbf{f}(\hat{\mathbf{x}}) \in \mathbb{R}^{n \times n} \quad \text{and} \quad \mathbf{c} \stackrel{\triangle}{=} \nabla \mathbf{f}_0(\hat{\mathbf{x}})\boldsymbol{\mu} \in \mathbb{R}^n. \tag{15}$$

Note that condition (GQC.a) and Remark 14 give $\mathbf{c} > \mathbf{0}$ for every $\boldsymbol{\mu} > \mathbf{0}$.

With the new notation, (13) can be written as

$$-\mathbf{c} + \hat{\boldsymbol{\lambda}} - \mathbf{A}\hat{\boldsymbol{\lambda}} = \mathbf{0}, \text{ or } \hat{\boldsymbol{\lambda}} = (\mathbf{I} - \mathbf{A})^{-1}\mathbf{c} \tag{16}$$

whenever $\mathbf{I} - \mathbf{A} = \mathbf{I} - \nabla \mathbf{f}(\hat{\mathbf{x}})$ is invertible, which is true for all $\mathbf{x} \in \mathcal{D}$ (proof of Lemma 13). The expansion of this inverse gives

$$\hat{\boldsymbol{\lambda}} = (\mathbf{I} + \mathbf{A} + \mathbf{A}^2 + \cdots)\mathbf{c} \tag{17a}$$
$$= (\mathbf{I} + \mathbf{A}^k + \mathbf{A}^{2k} + \cdots)(\mathbf{I} + \mathbf{A} + \cdots + \mathbf{A}^{k-1})\mathbf{c}$$
$$= \underbrace{\sum_{l=0}^{\infty} (\mathbf{A}^k)^l}_{\text{"left matrix"}} \underbrace{(\mathbf{I} + \mathbf{A} + \cdots + \mathbf{A}^{k-1})\mathbf{c}}_{\text{"right vector"}}. \tag{17b}$$

The first step is showing that the "left matrix" above is non-negative with non-zero rows. If $k \to \infty$, the "right vector" becomes identical to the right hand side of (17a). The "left matrix" must therefore equals identity,[3] which is non-negative with non-zero rows. For all other $k < \infty$, the "left matrix" is ensured non-negative by condition (GQC.c) (with non-zero rows guaranteed by the first term $\mathbf{A}^0 = \mathbf{I}$).

When the "left matrix" is non-negative with non-zero rows, a sufficient condition for $\hat{\boldsymbol{\lambda}} > \mathbf{0}$ is that the right vector is positive (Remark 14), i.e.,

$$\left(\mathbf{I} + \underbrace{\mathbf{A} + \cdots + \mathbf{A}^{k-1}}_{\stackrel{\triangle}{=}\mathbf{B}}\right)\mathbf{c} > \mathbf{0} \quad \Leftrightarrow \quad -\mathbf{B}\mathbf{c} < \mathbf{c}. \tag{18}$$

When $k = 1$, then $\mathbf{B} = \mathbf{0}$ and (18) holds trivially, since $\mathbf{c} > \mathbf{0}$. For $k > 1$, let $c_{\min}$ and $c_{\max}$ be the minimum and maximum elements of $\mathbf{c}$ and consider row i of inequality (18). We can now bound the right side by $c_{\min} \leq c_i$ and the left side by

$$|[-\mathbf{B}\mathbf{c}]_i| = \left|\sum_{j=1}^{n} B_{ij}c_j\right| \leq \sum_{j=1}^{n} |B_{ij}||c_j|$$
$$\leq \max_i \sum_{j=1}^{n} |B_{ij}|c_{\max} = \|\|\mathbf{B}\|\|_\infty c_{\max}. \tag{19}$$

Therefore, (18) holds if $\|\|\mathbf{B}\|\|_\infty c_{\max} < c_{\min}$, or

$$\|\|\mathbf{B}\|\|_\infty < \frac{c_{\min}}{c_{\max}}. \tag{20}$$

Let $\nabla_i \mathbf{f}_0(\hat{\mathbf{x}})$ be the ith row of $\nabla \mathbf{f}_0(\hat{\mathbf{x}})$ and define

$$\mathbf{a}(\boldsymbol{\mu}) = \underset{\nabla_i \mathbf{f}_0(\hat{\mathbf{x}})}{\operatorname{argmin}} \nabla_i \mathbf{f}_0(\hat{\mathbf{x}})\boldsymbol{\mu}, \text{ and } \mathbf{b}(\boldsymbol{\mu}) = \underset{\nabla_i \mathbf{f}_0(\hat{\mathbf{x}})}{\operatorname{argmax}} \nabla_i \mathbf{f}_0(\hat{\mathbf{x}})\boldsymbol{\mu}.$$

Let $d_k(\boldsymbol{\mu}) = a_k(\boldsymbol{\mu})/b_k(\boldsymbol{\mu})$ and express the components of $\mathbf{a}$ as $a_k(\boldsymbol{\mu}) = d_k(\boldsymbol{\mu})b_k(\boldsymbol{\mu})$. Since $\mathbf{c} = \nabla \mathbf{f}_0(\hat{\mathbf{x}})\boldsymbol{\mu}$, we have

$$c_{\min} = \mathbf{a}(\boldsymbol{\mu})\nabla \mathbf{f}_0(\hat{\mathbf{x}}) = \sum_k a_k(\boldsymbol{\mu})\mu_k = \sum_k d_k(\boldsymbol{\mu})b_k(\boldsymbol{\mu})\mu_k$$

and

$$c_{\max} = \mathbf{b}(\boldsymbol{\mu})\nabla \mathbf{f}_0(\hat{\mathbf{x}}) = \sum_k b_k(\boldsymbol{\mu})\mu_k.$$

---

[3]A different way to see this is by noting that $\lim_{k \to \infty} \mathbf{A}^k = 0$ since $\rho(|A|)\mathbf{1}$ by condition (GQC.b). The "left matrix" therefore evaluates to $\sum_{l=1}^{\infty} \mathbf{0}^l$, where only the term $\mathbf{0}^0 = \mathbf{I}$ gives a contribution.

The fraction in (20) can therefore be bounded by

$$\frac{c_{\min}}{c_{\max}} = \frac{\sum_k a_k(\boldsymbol{\mu}) \cdot \mu_k}{\sum_k b_k(\boldsymbol{\mu}) \cdot \mu_k} = \frac{\sum_k d_k(\boldsymbol{\mu}) \cdot b_k(\boldsymbol{\mu}) \cdot \mu_k}{\sum_k b_k(\boldsymbol{\mu}) \cdot \mu_k}$$
$$\geq \frac{\sum_k d_{\min}(\boldsymbol{\mu}) \cdot b_k(\boldsymbol{\mu}) \cdot \mu_k}{\sum_k b_k(\boldsymbol{\mu}) \cdot \mu_k} = d_{\min}(\boldsymbol{\mu}) \qquad (21)$$

where

$$d_{\min}(\boldsymbol{\mu}) = \min_k d_k(\boldsymbol{\mu}) = \min_k \frac{a_k(\boldsymbol{\mu})}{b_k(\boldsymbol{\mu})}$$
$$\geq \min_k \frac{\min_i [\nabla \mathbf{f}_0(\hat{\mathbf{x}})]_{ik}}{\max_i [\nabla \mathbf{f}_0(\hat{\mathbf{x}})]_{ik}} = \mathfrak{q}(\hat{\mathbf{x}}) \qquad (22)$$

with $\mathfrak{q}(\mathbf{x})$ defined in (8). By the definition of $\mathbf{B}$ and condition (GQC.d) we get

$$\||\mathbf{B}|\|_\infty = \left\|\left\| \sum_{l=1}^{k-1} \nabla \mathbf{f}(\hat{\mathbf{x}})^l \right\|\right\|_\infty < \mathfrak{q}(\hat{\mathbf{x}})$$

which together with inequalities (20)–(22) ensures

$$\||\mathbf{B}|\|_\infty = \left\|\left\| \sum_{l=1}^{k-1} \nabla \mathbf{f}(\hat{\mathbf{x}}) \right\|\right\|_\infty < \mathfrak{q}(\hat{\mathbf{x}}) \leq d_{\min} \leq \frac{c_{\min}}{c_{\max}}$$

wherefore $\hat{\boldsymbol{\lambda}} > \mathbf{0}$ by inequalities (18)–(20) and Remark 14. ∎

We now know that every pair $(\hat{\mathbf{x}}, \hat{\boldsymbol{\lambda}})$ satisfying the KKT conditions must have $\hat{\boldsymbol{\lambda}} > \mathbf{0}$, provided that GQC holds. Furthermore, strict complementarity ((14)) must hold. Since $\hat{\lambda}_i > 0$, we have $\hat{\lambda}_i(\hat{x}_i - f_i(\hat{\mathbf{x}})) = 0$ only if $\hat{x}_i - f_i(\hat{\mathbf{x}}) = 0$, i.e., $\hat{\mathbf{x}} = \mathbf{f}(\hat{\mathbf{x}})$. In other words, any candidate for primal optimality must be a fixed point of $\mathbf{f}(\mathbf{x})$. It remains to show that there always exists a fixed point, and that this fixed point is unique. But this is already done, since GQC and Lemma 10 ensure $\mathbf{f}(\mathbf{x})$ is contractive, and the result follow from Proposition 9.

By now, we have shown that the unique point $\mathbf{x}^\star = \mathbf{f}(\mathbf{x}^\star)$ is the only possible optimum of problem (12). Since this is true for *any* scalarization vector $\boldsymbol{\mu} > \mathbf{0}$, $\mathbf{x}^\star$ is the unique Pareto optimal point of problem (9). Finally, Lemma 11 extends this to the originally considered problem (5). By this, we have taken all the steps to prove Theorem 7. ∎

In the next section we revisit the GQC, through a number of special cases.

IV. SPECIAL CASES OF GQC

In this section, we discuss the general qualifying condition (GQC) in more detail. Moreover, we present several new qualifying conditions, each of which implies GQC.

GQC has the benefit of giving a unified view of the qualifying conditions. This is convenient for proving properties of Fast-Lipschitz problems, and also for giving an overall understanding for what the qualifying conditions ensure. However, GQC may not always be suitable when determining whether or not a given problem (or class of problems) is Fast-Lipschitz. This is because the generality of GQC comes at the price of analytical and computational complexity and cumbersome notation. For example, conditions (GQC.c) and (GQC.d) become increasingly tedious to verify as the integer k grows. However, as we will see, the special cases can yield clean and easily verifiable conditions, which are much easier to use in practice than GQC.

Furthermore, the specialized cases provide easy comparison to the old qualifying conditions, and other related work such as the standard and type-II standard function of [36] and [39]. We start with the simplest case Q_1, given below.

Qualifying Condition 1	
Q_1	**(Q_1.a)** $\nabla \mathbf{f}_0(\mathbf{x}) \geq \mathbf{0}$ with non-zero rows
	(Q_1.b) $\||\nabla \mathbf{f}(\mathbf{x})|\| < 1$
	(Q_1.c) $\nabla \mathbf{f}(\mathbf{x}) \geq \mathbf{0}$

Qualifying condition Q_1 is the special case of GQC when $k = 1$. It is the simplest case of Fast-Lipschitz optimization, and only requires a monotonic objective function $\mathbf{f}_0(\mathbf{x})$ and a monotonic, contractive constraint function $\mathbf{f}(\mathbf{x})$. Q_1 is highly related to standard interference functions [36]. In fact, any problem (5) with monotonic objective function and standard constraints is Fast-Lipschitz [41, Theorem 4.2], [44]. The difference between Q_1 and case (i) of the Old Qualifying Conditions (7) lies in condition (Q_1.a), where we now allow $\nabla \mathbf{f}_0(\mathbf{x}) \geq \mathbf{0}$ as long as no row consists only of zeros.

Qualifying Condition 2 is a simplified version of GQC with $k = 2$.

Qualifying Condition 2	
Q_2	**(Q_2.a)** $\nabla \mathbf{f}_0(\mathbf{x}) > \mathbf{0}$
	(Q_2.b) $\nabla \mathbf{f}(\mathbf{x})^2 \geq \mathbf{0}, \quad$ (e.g., $\nabla \mathbf{f}(\mathbf{x}) \leq \mathbf{0}$)
	(Q_2.c) $\||\nabla \mathbf{f}(\mathbf{x})|\|_\infty < \mathfrak{q}(\mathbf{x})$
	$\quad$ where $\mathfrak{q}(\mathbf{x}) \triangleq \min_j \frac{\min_i [\nabla \mathbf{f}_0(\mathbf{x})]_{ij}}{\max_i [\nabla \mathbf{f}_0(\mathbf{x})]_{ij}}$

Proposition 16: Qualifying condition Q_2 implies GQC.

Proof: Condition (GQC.a) is implied by (Q_2.a). Note that if any element of $\nabla \mathbf{f}_0(\mathbf{x})$ is zero, then $\mathfrak{q}(\mathbf{x}) = 0$ and condition (Q_2.c) cannot be fulfilled. We can therefore, without loss of generality, use a strict inequality in (Q_2.a). Condition (GQC.b) is fulfilled since $\||\nabla \mathbf{f}(\mathbf{x})|\|_\infty < \mathfrak{q}(\mathbf{x})$ by condition (Q_2.c), and $\mathfrak{q}(\mathbf{x}) \leq 1$. Finally, conditions (GQC.c) and (GQC.d) are (for $k = 2$) given exactly by (Q_2.b) and (Q_2.c), respectively. ∎

Condition (Q_2.b) requires that the square of the gradient is non-negative, but is particularly easy to verify when $\nabla \mathbf{f}(\mathbf{x}) \leq \mathbf{0}$. Note that also $\nabla \mathbf{f}(\mathbf{x}) \geq \mathbf{0}$ fulfills (Q_2.b), i.e., (Q_2.b) is more general than (Q_1.c). However, this generalization comes at a cost since (Q_2.c) is more restrictive than (Q_1.b) in the sense that it requires the specific norm $\||\cdot|\|$ and that $\mathfrak{q}(\mathbf{x})$ in general is less than one.

The formulation where $\nabla \mathbf{f}(\mathbf{x}) \leq \mathbf{0}$ corresponds to a non-increasing objective function, and the norm $\||\nabla \mathbf{f}(\mathbf{x})|\|_\infty$ is small enough, Q_2 is closely related to type-II standard functions [44].

Qualifying Condition 3	
Q_3	**(Q_3.a)** $\nabla \mathbf{f}_0(\mathbf{x}) > \mathbf{0}$
	(Q_3.b) $\||\nabla \mathbf{f}(\mathbf{x})|\|_\infty < \frac{\mathfrak{q}(\mathbf{x})}{1 + \mathfrak{q}(\mathbf{x})}$

Qualifying condition Q_3 can be seen as the special case of GQC when $k = \infty$, as we see in Proposition 17. In contrast to the other qualifying conditions, Q_3 does not require non-negativity of $\nabla \mathbf{f}(\mathbf{x})$ (or $(\nabla \mathbf{f}(\mathbf{x}))^k$). It is only required that $\mathbf{f}_0$

is monotonic and the infinity norm of $\nabla \mathbf{f}(\mathbf{x})$ is small enough. However, (Q_3.b) is significantly stricter than (Q_2.c) because $0 \leq \mathfrak{q}(\mathbf{x}) \leq 1$ implies $|||\nabla \mathbf{f}(\mathbf{x})|||_\infty < \mathfrak{q}(\mathbf{x})/(1+\mathfrak{q}(\mathbf{x})) \leq 1/2$ in (Q_3.b).

Proposition 17: Qualifying condition Q_3 implies GQC.

Proof: Condition (Q_3.a) implies (GQC.a) and condition (Q_3.b) implies $|||\nabla \mathbf{f}(\mathbf{x})|||_\infty < 1/2$, whereby condition (GQC.b) is fulfilled. Condition (GQC.c) is irrelevant when $k = \infty$.[4] Finally, condition (GQC.d) is implied by (Q_3.b) because

$$|||\nabla \mathbf{f}(\mathbf{x})|||_\infty < \frac{\mathfrak{q}(\mathbf{x})}{1+\mathfrak{q}(\mathbf{x})} \implies \frac{|||\nabla \mathbf{f}(\mathbf{x})|||_\infty}{1-|||\nabla \mathbf{f}(\mathbf{x})|||_\infty} < \mathfrak{q}(\mathbf{x}) \quad (23)$$

where the implication above follows from that:

$$\left|\left|\left|\sum_{l=1}^\infty \nabla \mathbf{f}(\mathbf{x})^l\right|\right|\right|_\infty \leq \sum_{l=1}^\infty |||\nabla \mathbf{f}(\mathbf{x})^l|||$$
$$\leq \sum_{l=1}^\infty |||\nabla \mathbf{f}(\mathbf{x})|||_\infty^l = \frac{|||\nabla \mathbf{f}(\mathbf{x})|||_\infty}{1-|||\nabla \mathbf{f}(\mathbf{x})|||_\infty}$$
$$< \mathfrak{q}(\mathbf{x}).$$

The first two inequalities follow from sub-additive and sub-multiplicative properties of matrix norms. The equality follows from the geometric series because $|||\nabla \mathbf{f}(\mathbf{x})|||_\infty < 1$, and the last inequality is expression (23). This concludes the proof. ∎

The qualifying conditions can be further simplified by introducing $\delta(\mathbf{x}) \triangleq \min_{i,j} [\nabla \mathbf{f}_0(\mathbf{x})]_{ij}$ and $\Delta(\mathbf{x}) \triangleq \max_{i,j} [\nabla \mathbf{f}_0(\mathbf{x})]_{ij}$. These are the smallest and largest elements of $\nabla \mathbf{f}_0(\mathbf{x})$, regardless of column. The difference compared to $\bar{\delta}$ and $\bar{\Delta}$ in (7h) is that we now evaluate δ and Δ for each $\mathbf{x}$, instead of taking the extremes over all $\mathbf{x}$. We can now bound

$$\mathfrak{q}(\mathbf{x}) = \min_j \frac{\min_i [\nabla \mathbf{f}_0]_{ij}}{\max_i [\nabla \mathbf{f}_0]_{ij}} \geq \frac{\min_{ij} [\nabla \mathbf{f}_0]_{ij}}{\max_{ij} [\nabla \mathbf{f}_0]_{ij}} = \frac{\delta(\mathbf{x})}{\Delta(\mathbf{x})}. \quad (24)$$

Since both $\mathfrak{q}(\mathbf{x})$ and $\mathfrak{q}(\mathbf{x})/(1+\mathfrak{q}(\mathbf{x}))$ are increasing in $\mathfrak{q}$ (recall $\mathfrak{q}(\mathbf{x}) \geq 0$), we can lower bound $\mathfrak{q}(\mathbf{x})$ by $\delta(\mathbf{x})/\Delta(\mathbf{x})$ in any one of the qualifying conditions above. This gives the remaining qualifying conditions of this section. They are all special cases of previous conditions—easier to verify and analyze, at the expense of being more conservative.

Qualifying Condition 4	
Q_4	(Q_4.a) $\nabla \mathbf{f}_0(\mathbf{x}) > \mathbf{0}$ (Q_4.b) $\nabla \mathbf{f}(\mathbf{x})^2 \geq \mathbf{0}$, (e.g., $\nabla \mathbf{f}(\mathbf{x}) \leq \mathbf{0}$) ($Q_4$.c) $\|\|\|\nabla \mathbf{f}(\mathbf{x})\|\|\|_\infty < \dfrac{\delta(\mathbf{x})}{\Delta(\mathbf{x})}$
Qualifying Condition 5	
Q_5	(Q_5.a) $\nabla \mathbf{f}_0(\mathbf{x}) > \mathbf{0}$ (Q_5.b) $\|\|\|\nabla \mathbf{f}(\mathbf{x})\|\|\|_\infty < \dfrac{\delta(\mathbf{x})}{\delta(\mathbf{x})+\Delta(\mathbf{x})}$

Qualifying conditions Q_4 and Q_5 are obtained when inserting inequality (24) in (Q_2.c) and (Q_3.b) respectively. They imply Q_2 and Q_3 which in turn imply GQC by construction. Note that qualifying conditions Q_4 and Q_5 have previously appeared as [1, case (ii)–(iii)].

[4]This is discussed in the proof of Theorem 7, after (17).

We will now end this section with the observation that the old qualifying conditions are a special case of GQC.

Proposition 18: The Old Qualifying Conditions (7) imply GQC.

Proof: We will show how each case of the old qualifying conditions imply one of the qualifying conditions 1–5 above, which in turn implies GQC.

case (i) ⇒ Q_1: Condition (Q_1.a) is implied by (7a), (Q_1.b) and (Q_1.c) are the same as (7b) and (7c).

case (ii) ⇒ Q_4: Conditions (Q_4.a) and (Q_4.b) are the same as (7a) and (7e). Condition (7d) implies $\nabla \mathbf{f}_0(\mathbf{x}) = c\mathbf{1}$, whereby $\delta(\mathbf{x}) = \Delta(\mathbf{x}) = c$. Condition ($Q_4$.c) therefore requires that $|||\nabla \mathbf{f}(\mathbf{x})|||_\infty < 1$, which is ensured by (7f).

case (iii) ⇒ Q_5: Condition (7g) requires $\mathbf{f}_0(\mathbf{x})$ to be scalar valued, whereby $\nabla \mathbf{f}_0(\mathbf{x})$ only has one column. The deltas of condition (7h) can therefore be written as $\bar{\delta} \triangleq \min_{\mathbf{x}\in\mathcal{D}} \delta(\mathbf{x}) \leq \delta(\mathbf{x})$ and $\bar{\Delta} \triangleq \max_{\mathbf{x}\in\mathcal{D}} \Delta(\mathbf{x}) \geq \Delta(\mathbf{x})$. This gives

$$\frac{\delta_{\min}}{\Delta_{\max}} \leq \frac{\delta(\mathbf{x})}{\Delta(\mathbf{x})} \quad (25)$$

whereby

$$|||\nabla \mathbf{f}(\mathbf{x})|||_\infty < \frac{\bar{\delta}}{\bar{\delta}+\bar{\Delta}} = \frac{\bar{\delta}/\bar{\Delta}}{1+\bar{\delta}/\bar{\Delta}} \leq \frac{\delta(\mathbf{x})/\Delta(\mathbf{x})}{1+\delta(\mathbf{x})/\Delta(\mathbf{x})}$$
$$= \frac{\delta(\mathbf{x})}{\delta(\mathbf{x})+\Delta(\mathbf{x})}$$

and (Q_5.b) is ensured. The first inequality is condition (7h), the second inequality follows from (25) because $h(a) = a/(1+a)$ is an increasing function of a.

We have now showed how each case of the Old Qualifying Conditions implies GQC, through the implication chain

$$\text{GQC} \begin{array}{c} \nLeftarrow \\ \Longleftarrow \\ \Longleftarrow \end{array} \begin{array}{ccccc} Q_1 & \Longleftarrow & (i) & & \\ Q_2 & \Longleftarrow & Q_4 & \Longleftarrow & (ii) \\ Q_3 & \Longleftarrow & Q_5 & \Longleftarrow & (iii). \end{array}$$

This concludes the proof. ∎

The next section will loosen some of the assumptions of the optimization problem structure, i.e., investigate problems not entirely in Fast-Lipschitz form.

V. RELAXATIONS OF THE FAST-LIPSCHITZ FORM

This section considers relaxations of the Fast-Lipschitz form that, e.g., require the same number of constraints as variables. Section V-A shows a technique for handling more constraints than variables, and Section V-B shows a situation with fewer constraints than variables. Finally, Section V-C discuss the case when the objective function does not depend on all variables.

A. Additional Constraints

In this section we complement the Fast-Lipschitz form (5) with an additional set $\mathcal{X}$. Hence, we consider the problem

$$\begin{aligned} \max \quad & \mathbf{f}_0(\mathbf{x}) \\ \text{s.t.} \quad & x_i \leq f_i(\mathbf{x}) \quad \forall i \in \mathbb{A} \\ & x_i = f_i(\mathbf{x}) \quad \forall i \in \mathbb{B} \\ & \mathbf{x} \in \mathcal{X}. \end{aligned} \quad (26)$$

Corollary 19: If GQC holds, and $\mathcal{X}$ contains a point $\mathbf{x}^\star = \mathbf{f}(\mathbf{x}^\star)$, then problem (26) is Fast-Lipschitz.

Proof: Relax the problem by removing the new constraint $\mathbf{x} \in \mathcal{X}$. The relaxed problem is our main problem (5), whereby the qualifying conditions and Theorem 7 ensure $\mathbf{x}^\star = \mathbf{f}(\mathbf{x}^\star)$ is the unique optimum. Since $\mathbf{x}^\star \in \mathcal{X}$, this is also the unique optimum of problem (26). ∎

In theory, we can handle any set $\mathcal{X}$ provided we can show $\mathbf{x}^\star \in \mathcal{X}$. For example, $\mathcal{X}$ does not need to be bounded, convex, or even connected (i.e., $\mathcal{X}$ can consist of mutually disconnected subsets). In practice however, the most common form of $\mathcal{X}$ is a box constraint, for example a requirement of non-negativity, $\mathcal{X} = \{\mathbf{x} : \mathbf{0} \leq \mathbf{x}\}$. In these cases, $\mathcal{X}$ becomes the natural choice for the imagined bounding box $\mathcal{D}$.

B. Fewer Constraints Than Variables—Constant Constraints

The Fast-Lipschitz form in problem (5) requires one (and only one) constraint f_i for each variable x_i. In this section we will look at the case when the number of constraints (f_i) are fewer than the number of variables. We will assume that the individual variables are upper and lower bounded, which is always the case for problems of engineering interest, such as wireless networks. This means we get an extra constraint set $\mathcal{X}$ as discussed in Section V-A. We investigate the case when all constraints are inequalities. It follows from Lemma 11 that the following results are true also for problems with equality constraints.

Consider a partitioned variable $\mathbf{x} = [\mathbf{y}^T \ \mathbf{z}^T]^T \in \mathcal{X} \subset \mathbb{R}^n$

$$
\begin{aligned}
\max \quad & \mathbf{f}_0(\mathbf{x}) \\
\text{s.t.} \quad & \mathbf{y} \leq \mathbf{f_y}(\mathbf{x}) \\
& \mathbf{x} \in \mathcal{X}, \\
\text{with } \mathcal{X} = & \left\{ \mathbf{x} : \begin{array}{l} \mathbf{y} \in \mathcal{X_y} = \{\mathbf{y} : \mathbf{a_y} \leq \mathbf{y} \leq \mathbf{b_y}\} \\ \mathbf{z} \in \mathcal{X_z} = \{\mathbf{z} : \mathbf{a_z} \leq \mathbf{z} \leq \mathbf{b_z}\} \end{array} \right\}.
\end{aligned} \quad (27)
$$

In the formulation above, there are no constraints f_i for the variables z_i. However, by enforcing $\mathbf{z} \in \mathcal{X_z}$ twice we get the equivalent problem

$$
\begin{aligned}
\max \quad & \mathbf{f}_0(\mathbf{x}) \\
\text{s.t.} \quad & \mathbf{y} \leq \mathbf{f_y}(\mathbf{x}) \\
& \mathbf{z} \leq \mathbf{f_z}(\mathbf{x}) = \mathbf{b_z} \\
& \mathbf{x} \in \mathcal{X}.
\end{aligned} \quad (28)
$$

This problem has the right form (26) and is Fast-Lipschitz if

$$
\nabla \mathbf{f} = \begin{bmatrix} \nabla_\mathbf{y} \mathbf{f_x}(\mathbf{x}) & \nabla_\mathbf{y} \mathbf{f_z}(\mathbf{x}) \\ \nabla_\mathbf{z} \mathbf{f_y}(\mathbf{x}) & \nabla_\mathbf{z} \mathbf{f_z}(\mathbf{x}) \end{bmatrix} \quad \text{and} \quad \nabla \mathbf{f}_0 = \begin{bmatrix} \nabla_\mathbf{y} \mathbf{f}_0(\mathbf{x}) \\ \nabla_\mathbf{z} \mathbf{f}_0(\mathbf{x}) \end{bmatrix}
$$

fulfills GQC. Since $\mathbf{f_z}(\mathbf{x}) = \mathbf{b_z}$ is constant, $\nabla \mathbf{f}(\mathbf{x})$ simplifies to

$$
\nabla \mathbf{f} = \begin{bmatrix} \nabla_\mathbf{y} \mathbf{f_y}(\mathbf{x}) & \mathbf{0} \\ \nabla_\mathbf{z} \mathbf{f_y}(\mathbf{x}) & \mathbf{0} \end{bmatrix}.
$$

The special structure of $\nabla \mathbf{f}_0(\mathbf{x})$ can be exploited to construct less restrictive qualifying conditions.

First, consider problem (27) with a fixed $\mathbf{z} \in \mathcal{X_z}$. The problem can be written

$$
\begin{aligned}
\max \quad & \mathbf{f}_{0|\mathbf{z}}(\mathbf{y}) \\
\text{s.t.} \quad & \mathbf{y} \leq \mathbf{f}_{\mathbf{y}|\mathbf{z}}(\mathbf{y}) \\
& \mathbf{y} \in \mathcal{X_y}.
\end{aligned} \quad (29)
$$

We will refer to problem (29) as the *subproblem* $(\mathbf{f}_{0|\mathbf{z}}, \mathbf{f}_{\mathbf{y}|\mathbf{z}})$.

Proposition 20: Consider problem (28). Suppose that (a) the subproblem $(\mathbf{f}_{0|\mathbf{z}}, \mathbf{f}_{\mathbf{y}|\mathbf{z}})$ fulfills GQC for all $\mathbf{z} \in \mathcal{X_z}$, and it holds for all $\mathbf{x} \in \mathcal{X}$, that either.

(b.i) $\nabla_\mathbf{z} \mathbf{f}_0(\mathbf{x}) \geq \mathbf{0}$, and $\nabla_\mathbf{z} \mathbf{f_y}(\mathbf{x}) \geq \mathbf{0}$ with non-zero rows, or

(b.ii) $\nabla_\mathbf{z} \mathbf{f}_0(\mathbf{x}) \geq \mathbf{0}$ with non-zero rows, and $\nabla_\mathbf{z} \mathbf{f_y}(\mathbf{x}) \geq \mathbf{0}$, or

(b.iii) $\||\nabla_\mathbf{z} \mathbf{f_y}(\mathbf{x})|\|_\infty / (1 - \||\nabla_\mathbf{y} \mathbf{f_y}(\mathbf{x})|\|_\infty) < \delta_\mathbf{z}(\mathbf{x}) / \Delta_\mathbf{y}(\mathbf{x})$, where $\delta_\mathbf{z}(\mathbf{x}) = \min_{ij} [\nabla_\mathbf{z} \mathbf{f}_0(\mathbf{x})]_{ij}$ and $\Delta_\mathbf{y}(\mathbf{x}) = \max_{ij} [\nabla_\mathbf{y} \mathbf{f}_0(\mathbf{x})]_{ij}$.

Then, problem (27) is Fast-Lipschitz.

Proof: We refer to the arguments of Section III-C where we modify Lemma 15 as follows.

The particular form and partitioning remains in the definitions (15), giving

$$
\mathbf{A} = \begin{bmatrix} \nabla_\mathbf{y} \mathbf{f_y}(\hat{\mathbf{x}}) & \mathbf{0} \\ \nabla_\mathbf{z} \mathbf{f_y}(\hat{\mathbf{x}}) & \mathbf{0} \end{bmatrix} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{0} \\ \mathbf{A}_{21} & \mathbf{0} \end{bmatrix}
$$

$$
\mathbf{c} = \begin{bmatrix} \nabla_\mathbf{y} \mathbf{f}_0(\hat{\mathbf{x}}) \\ \nabla_\mathbf{z} \mathbf{f}_0(\hat{\mathbf{x}}) \end{bmatrix} \boldsymbol{\mu} = \begin{bmatrix} \mathbf{c}_1 \\ \mathbf{c}_2 \end{bmatrix}.
$$

Consider again (16) and denote $\mathbf{E} \triangleq (\mathbf{I} - \mathbf{A})^{-1}$, i.e., $\boldsymbol{\lambda} = \mathbf{E}\mathbf{c}$. As in the proof of Lemma 15, $\mathbf{E}$ is well defined if $\rho(\mathbf{A}) < 1$. This is the case, since the eigenvalues of a block triangular matrix are the union of the eigenvalues of the diagonal blocks, wherefore $\rho(\mathbf{A}) = \rho(\mathbf{A}_{11}) < 1$ by assumption (a).

As in the proof of Lemma 15, we must show $\boldsymbol{\lambda} > \mathbf{0}$. This time we will make use of the block structure of $\mathbf{A}$ and $\mathbf{c}$.[5] From the block matrix inverse formula we get

$$
\boldsymbol{\lambda} = (\mathbf{I} - \mathbf{A})^{-1} \mathbf{c} = \begin{bmatrix} \mathbf{E}_{11} & \mathbf{E}_{12} \\ \mathbf{E}_{21} & \mathbf{E}_{22} \end{bmatrix} \begin{bmatrix} \mathbf{c}_1 \\ \mathbf{c}_2 \end{bmatrix}
$$

where $\mathbf{E}_{11} = (\mathbf{I} - \mathbf{A}_{11})^{-1}$, $\mathbf{E}_{12} = \mathbf{0}$, $\mathbf{E}_{21} = \mathbf{A}_{21} \mathbf{E}_{11}$ and $\mathbf{E}_{22} = \mathbf{I}$, i.e.,

$$
\begin{bmatrix} \boldsymbol{\lambda}_1 \\ \boldsymbol{\lambda}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{E}_{11} \mathbf{c}_1 \\ \mathbf{E}_{21} \mathbf{c}_1 + \mathbf{c}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{E}_{11} \mathbf{c}_1 \\ \mathbf{A}_{21} \mathbf{E}_{11} \mathbf{c}_1 + \mathbf{c}_2 \end{bmatrix}.
$$

Note that $\boldsymbol{\lambda}_1 = \mathbf{E}_{11} \mathbf{c}_1 = (\mathbf{I} - \mathbf{A}_{11})^{-1} \mathbf{c}_1$, so $\boldsymbol{\lambda}_1 > \mathbf{0}$ since the subproblem $(\mathbf{f}_{0|\mathbf{z}}, \mathbf{f}_{y|\mathbf{z}})$ fulfills Lemma 15 by assumption (a). The second block component is

$$
\boldsymbol{\lambda}_2 = \mathbf{A}_{21} \mathbf{E}_{11} \mathbf{c}_1 + \mathbf{c}_2 = \mathbf{A}_{21} \boldsymbol{\lambda}_1 + \mathbf{c}_2. \quad (31)
$$

Given that $\boldsymbol{\lambda}_1 > \mathbf{0}$, we need to show $\boldsymbol{\lambda}_2 > \mathbf{0}$ if either of assumptions (b.i)–(b.iii) hold.

We start with assumption (b.i), which ensures $\mathbf{c}_2 \geq \mathbf{0}$ and $\mathbf{A}_{21} \geq \mathbf{0}$ with non-zero rows (so $\mathbf{A}_{21} \boldsymbol{\lambda}_1 > \mathbf{0}$ by Remark 14), wherefore $\boldsymbol{\lambda}_2 > \mathbf{0}$ by (31).

Assumption (b.ii) assures $\mathbf{A}_{21} \geq \mathbf{0}$ and $\mathbf{c}_2 > \mathbf{0}$, wherefore $\boldsymbol{\lambda}_2 > \mathbf{0}$ by (31).

Finally, assuming (b.iii) is fulfilled, we see that $\boldsymbol{\lambda}_2 > \mathbf{0}$ if $\mathbf{c}_2 > -\mathbf{A}_{21} \mathbf{E}_{11} \mathbf{c}_1$. Analogous to (19) and (20), this holds if $\min_i [\mathbf{c}_2]_i > \||\mathbf{A}_{21} \mathbf{E}_{11}|\| \max_i [\mathbf{c}_1]_i$, or since $\mathbf{c}_1 > \mathbf{0}$ and $\mathbf{E}_{11} = (\mathbf{I} - \mathbf{A}_{11})^{-1} = \sum_{k=0}^\infty \mathbf{A}_{11}^k$, when

$$
\frac{\min_i [\mathbf{c}_2]_i}{\max_i [\mathbf{c}_1]_i} > \left\| \mathbf{A}_{21} \sum_{k=0}^\infty \mathbf{A}_{11}^k \right\|_\infty. \quad (32)
$$

---

[5]Formulas for the inverse of a block matrix, as well as the products of two block matrices can be found in [45].

By the triangle inequality, the sub-multiplicative property of matrix norms and the geometric series, the right side of (32) can be upper bounded by

$$\frac{\||\mathbf{A}_{21}\||_\infty}{1 - \||\mathbf{A}_{11}\||_\infty} \geq \left\|\left\| \mathbf{A}_{21} \sum_{k=0}^\infty \mathbf{A}_{11}^k \right\|\right\|_\infty. \quad (33)$$

Finally the definitions of $\mathbf{c}_i$ and $\sum_j \mu_j = 1$ gives

$$\begin{aligned}
\frac{\min_i [\mathbf{c}_2]_i}{\max_i [\mathbf{c}_1]_i} &= \frac{\min_i \sum_j [\nabla_\mathbf{z} \mathbf{f}_0(\mathbf{x})]_{ij}\, \mu_j}{\max_i \sum_j [\nabla_\mathbf{y} \mathbf{f}_0(\mathbf{x})]_{ij}\, \mu_j} \\
&\geq \frac{\min_{ij} [\nabla_\mathbf{z} \mathbf{f}_0(\mathbf{x})]_{ij} \sum_j \mu_j}{\max_{ij} [\nabla_\mathbf{y} \mathbf{f}_0(\mathbf{x})]_{ij} \sum_j \mu_j} \\
&= \frac{\delta_\mathbf{z}(\mathbf{x})}{\Delta_\mathbf{y}(\mathbf{x})}. \quad (34)
\end{aligned}$$

By combining inequalities (32)–(34), a sufficient condition ensuring $\lambda_2 > 0$ is

$$\frac{\||\mathbf{A}_{21}\||_\infty}{1 - \||\mathbf{A}_{11}\||_\infty} < \frac{\delta_\mathbf{z}(\mathbf{x})}{\Delta_\mathbf{y}(\mathbf{x})}$$

which is guaranteed by assumption (d). This concludes the proof. ∎

C. Non-Strictly Monotonic Objective Function—Variables Missing in Objective Function

Sometimes it is practical or necessary to formulate problems where not all variables appear in the objective function.

For example, the problem

$$\begin{aligned}
\max\quad &\mathbf{f}_0(\mathbf{x}) \\
\text{s.t.}\quad &\mathbf{x} \leq \mathbf{f}_x(\mathbf{x}, \mathbf{z}) \\
&\mathbf{z} \leq \mathbf{f}_z(\mathbf{x}, \mathbf{z})
\end{aligned}$$

has some variables not affecting the objective function and is not in Fast-Lipschitz form. Redefining $\mathbf{f}_0 = \mathbf{f}_0(\mathbf{x}, \mathbf{z})$ gives a problem of the right form (5), but $\nabla_\mathbf{z} \mathbf{f}_0(\mathbf{x}, \mathbf{z}) = \mathbf{0}$ everywhere. Therefore, condition (GQC.a) and Theorem 7 can not be used to classify the problem as Fast-Lipschitz.

The situation above is a special case of the following problem. Consider a partitioned optimization variable $(\mathbf{x}, \mathbf{z})$ and the problem

$$\begin{aligned}
\max\quad &\mathbf{f}_0(\mathbf{x}, \mathbf{z}) \\
\text{s.t.}\quad &\mathbf{x} \leq \mathbf{f}_\mathbf{x}(\mathbf{x}, \mathbf{z}) \\
&\mathbf{z} \leq \mathbf{f}_\mathbf{z}(\mathbf{x}, \mathbf{z}). \quad (35)
\end{aligned}$$

Suppose $\mathbf{f}_0$ is monotonic, i.e., $\nabla \mathbf{f}_0 \geq \mathbf{0}$, but partitioned such that

$$\nabla \mathbf{f}_0(\mathbf{x}, \mathbf{z}) = \begin{bmatrix} \nabla_\mathbf{x} \mathbf{f}_0(\mathbf{x}, \mathbf{z}) \\ \nabla_\mathbf{z} \mathbf{f}_0(\mathbf{x}, \mathbf{z}) \end{bmatrix}$$

where $\nabla_\mathbf{x} \mathbf{f}_0(\mathbf{x}, \mathbf{z})$ has non-zero rows for all $(\mathbf{x}, \mathbf{z}) \in \mathcal{D}$, while $\nabla_\mathbf{z} \mathbf{f}_0(\mathbf{x}, \mathbf{z})$ can have zero rows. By partitioning the objective function gradient, one can find situations when problem (35) is actually Fast-Lipschitz.

Proposition 21: Consider problem (35). If it holds, for all $\mathbf{x}$ in $\mathcal{D}$, that

(a) $\nabla_\mathbf{x} \mathbf{f}_0(\mathbf{x}, \mathbf{z}) > \mathbf{0}$ and $\nabla_\mathbf{z} \mathbf{f}_0(\mathbf{x}, \mathbf{z}) \geq \mathbf{0}$

(b) $\nabla \mathbf{f}(\mathbf{x}, \mathbf{z}) = \begin{bmatrix} \nabla_\mathbf{x} \mathbf{f}_\mathbf{x}(\mathbf{x}, \mathbf{z}) & \nabla_\mathbf{x} \mathbf{f}_\mathbf{z}(\mathbf{x}, \mathbf{z}) \\ \nabla_\mathbf{z} \mathbf{f}_\mathbf{x}(\mathbf{x}, \mathbf{z}) & \nabla_\mathbf{z} \mathbf{f}_\mathbf{z}(\mathbf{x}, \mathbf{z}) \end{bmatrix} \geq \mathbf{0},$

(c) $\||\nabla \mathbf{f}(\mathbf{x})\|| < 1$ for some matrix norm, and

(d) $\nabla_\mathbf{z} \mathbf{f}_\mathbf{x}(\mathbf{x}, \mathbf{z})$ has non-zero rows.

Then, problem (35) is Fast-Lipschitz.

Remark 22: The condition that the ith row of $\nabla_\mathbf{z} \mathbf{f}_\mathbf{x}(\mathbf{x}, \mathbf{z})$ is non-zero means that an increase in the variable z_i will allow an increase of some variable x_j, which in turn will influence the objective.

Proof: The proof of Theorem 7 can be reused, with some alterations to Lemma 15.

The partitioning of $\nabla \mathbf{f}$ and $\nabla \mathbf{f}_0$ remains in $\mathbf{A}$ and $\mathbf{c}$, i.e.,

$$\mathbf{A} = \begin{bmatrix} \nabla_\mathbf{x} \mathbf{f}_\mathbf{x}(\mathbf{x}, \mathbf{z}) & \nabla_\mathbf{x} \mathbf{f}_\mathbf{z}(\mathbf{x}, \mathbf{z}) \\ \nabla_\mathbf{z} \mathbf{f}_\mathbf{x}(\mathbf{x}, \mathbf{z}) & \nabla_\mathbf{z} \mathbf{f}_\mathbf{z}(\mathbf{x}, \mathbf{z}) \end{bmatrix} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix}$$

$$\mathbf{c} = \begin{bmatrix} \nabla_\mathbf{x} \mathbf{f}_0(\mathbf{x}, \mathbf{z}) \\ \nabla_\mathbf{z} \mathbf{f}_0(\mathbf{x}, \mathbf{z}) \end{bmatrix} \quad \boldsymbol{\mu} = \begin{bmatrix} \mathbf{c}_1 \\ \mathbf{c}_2 \end{bmatrix}.$$

Assumption (a) and $\boldsymbol{\mu} > \mathbf{0}$ gives $\mathbf{c}_1 > \mathbf{0}$ and $\mathbf{c}_2 \geq \mathbf{0}$. Just as in Lemma 15, assumptions (b) and (c) guarantee the existence and non-negativity of $\mathbf{E} = (\mathbf{I} - \mathbf{A})^{-1}$, so $\boldsymbol{\lambda} = (\mathbf{I} - \mathbf{A})^{-1} \mathbf{c} \geq \mathbf{0}$ is well defined and non-negative.[6] Thus, it remains to show $\boldsymbol{\lambda} = \mathbf{E} \mathbf{c} > \mathbf{0}$.

Expressing the inverse $\mathbf{E} = (\mathbf{I} - \mathbf{A})^{-1}$ block-wise, we have

$$(\mathbf{I} - \mathbf{A})^{-1} = \begin{bmatrix} \mathbf{I} - \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{I} - \mathbf{A}_{22} \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{E}_{11} & \mathbf{E}_{12} \\ \mathbf{E}_{21} & \mathbf{E}_{22} \end{bmatrix}$$

where $\mathbf{E}_{11} = (\mathbf{I} - (\mathbf{A}_{11} + \mathbf{A}_{12}(\mathbf{I} - \mathbf{A}_{22})^{-1} \mathbf{A}_{21}))^{-1}$ and $\mathbf{E}_{21} = (\mathbf{I} - \mathbf{A}_{22})^{-1} \mathbf{A}_{21} \mathbf{E}_{11}$. We now have

$$\begin{bmatrix} \boldsymbol{\lambda}_1 \\ \boldsymbol{\lambda}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{E}_{11} & \mathbf{E}_{12} \\ \mathbf{E}_{21} & \mathbf{E}_{22} \end{bmatrix} \begin{bmatrix} \mathbf{c}_1 \\ \mathbf{c}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{E}_{11} \mathbf{c}_1 \\ \mathbf{E}_{21} \mathbf{c}_1 \end{bmatrix} + \begin{bmatrix} \mathbf{E}_{12} \\ \mathbf{E}_{22} \end{bmatrix} \mathbf{c}_2.$$

Since $\mathbf{E} \geq \mathbf{0}$, all blocks $\mathbf{E}_{ij}$ are non-negative. The second term is always non-negative and can be ignored, it is enough to show that the first term is strictly positive.

Since $\mathbf{c}_1 > \mathbf{0}$, and $\mathbf{E}_{11} \geq \mathbf{0}$, we have that $\boldsymbol{\lambda}_1 > \mathbf{0}$ if $\mathbf{E}_{11}$ has non-zero rows (Remark 14). This is always the case since $\mathbf{E}_{11}$ (defined as an inverse) is invertible. The second component can be expressed in terms of the first component: $\boldsymbol{\lambda}_2 = (\mathbf{I} - \mathbf{A}_{22})^{-1} \mathbf{A}_{21} \boldsymbol{\lambda}_1$. When $\boldsymbol{\lambda}_1 > \mathbf{0}$, $\mathbf{A}_{21} \boldsymbol{\lambda}_1 > \mathbf{0}$ if $\mathbf{A}_{21} = \nabla_\mathbf{z} \mathbf{f}_\mathbf{x}(\mathbf{x}, \mathbf{z})$ has non-zero rows (Remark 14). This is true by assumption (d). Since $(\mathbf{I} - \mathbf{A}_{22})^{-1}$ is invertible it has non-zero rows and $\boldsymbol{\lambda}_2 = (\mathbf{I} - \mathbf{A}_{22})^{-1}(\mathbf{A}_{21} \boldsymbol{\lambda}_1) > \mathbf{0}$, which concludes the proof. ∎

We end this section by a general example.

Example: Start with problem (5). For a lighter notation, we assume all constraints are inequalities. Transforming the problem to an equivalent problem on epigraph form gives

$$\begin{aligned}
\max\quad &\mathbf{t} \\
\text{s.t.}\quad &\mathbf{t} \leq \mathbf{f}_0(\mathbf{x}) \\
&\mathbf{x} \leq \mathbf{f}(\mathbf{x}).
\end{aligned}$$

---

[6] In contrast to Lemma 15, we only have the weak inequality $\mathbf{c} \geq \mathbf{0}$.

This problem has a (non-strictly) monotonic objective, regardless of $\mathbf{f}_0(\mathbf{x})$. By writing this as

$$\begin{aligned} \max \quad & \mathbf{g}_0(\mathbf{t}, \mathbf{x}) = \mathbf{t} \\ \text{s.t.} \quad & \mathbf{t} \leq \mathbf{g}_t(\mathbf{t}, \mathbf{x}) = \mathbf{f}_0(\mathbf{x}) \\ & \mathbf{x} \leq \mathbf{g}_x(\mathbf{t}, \mathbf{x}) = \mathbf{f}(\mathbf{x}) \end{aligned}$$

we obtain

$$\nabla \mathbf{g}_0(\mathbf{t}, \mathbf{x}) = \begin{bmatrix} \nabla_t \mathbf{g}_0(\mathbf{t}, \mathbf{x}) \\ \nabla_x \mathbf{g}_0(\mathbf{t}, \mathbf{x}) \end{bmatrix} = \begin{bmatrix} \mathbf{I} \\ \mathbf{0} \end{bmatrix}$$

$$\begin{aligned} \nabla \mathbf{g}(\mathbf{t}, \mathbf{x}) &= \begin{bmatrix} \nabla_t \mathbf{g}_t(\mathbf{t}, \mathbf{x}) & \nabla_t \mathbf{g}_x(\mathbf{t}, \mathbf{x}) \\ \nabla_x \mathbf{g}_t(\mathbf{t}, \mathbf{x}) & \nabla_x \mathbf{g}_x(\mathbf{t}, \mathbf{x}) \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \nabla \mathbf{f}_0(\mathbf{x}) & \nabla \mathbf{f}(\mathbf{x}) \end{bmatrix}. \end{aligned}$$

Proposition 21 can now be applied, and the problem is Fast-Lipschitz if $\nabla \mathbf{g} \geq \mathbf{0}$, $\rho(\nabla \mathbf{g}) < 1$ and $\nabla \mathbf{f}_0$ has non-zero rows. Since the eigenvalues of a block triangular matrix are the union of the eigenvalues of the diagonal blocks, we have $\rho(\nabla \mathbf{g}) = \rho(\nabla \mathbf{f})$, wherefore the problem is Fast-Lipschitz if

- $\nabla \mathbf{f}(\mathbf{x}) \geq \mathbf{0}$, $\||\nabla \mathbf{f}(\mathbf{x})|\| < 1$, and
- $\nabla \mathbf{f}_0(\mathbf{x}) \geq \mathbf{0}$ with non-zero rows.

This is precisely qualifying condition Q_1 applied to the original problem (5), i.e., no generalization was achieved by considering the epigraph form of the problem.

This concludes the main part of the paper. In the following section we will illustrate the new theory with two examples.

VI. EXAMPLES

We begin by illustrating Fast-Lipschitz optimization on a non-convex optimization example in Section VI-A. In Section VI-B and C, we apply the novel results established in this paper to state conditions for when two types of optimal control problems are easily solvable. Finally, in Section VI-D we compare an interior point method with finding the fixed point of the constraints on a problem which is both convex and Fast-Lipschitz.

A. Simple Non-Convex Example

Consider the problem

$$\begin{aligned} \max \quad & \mathbf{f}_0(\mathbf{x}) \\ \text{s.t.} \quad & \mathbf{x} \leq \mathbf{f}(\mathbf{x}) \\ & \mathbf{x} \in \mathcal{X} = \{\mathbf{x} : \mathbf{0} \leq \mathbf{x} \leq \mathbf{1}\} \end{aligned} \tag{36}$$

where $\mathbf{x} \in \mathbb{R}^2$

$$\mathbf{f}_0(\mathbf{x}) = \begin{bmatrix} 2x_1 + x_2 \\ x_1 + 2x_2 \end{bmatrix}, \quad \text{and} \quad \mathbf{f}(\mathbf{x}) = 0.5 \begin{bmatrix} 1 + ax_2^2 \\ 1 + bx_1^2 \end{bmatrix}.$$

If either a or b are positive, the problem is not convex (the canonical constraint functions $x_i - f_i(\mathbf{x})$ become concave).

At this point we do not know if $\mathbf{x}^\star = \mathbf{f}(\mathbf{x}^\star) \in \mathcal{X}$. However, following the results of Section V-A, we can assume this is the case and examine whether:

$$\begin{aligned} \max \quad & \mathbf{f}_0(\mathbf{x}) \\ \text{s.t.} \quad & \mathbf{x} \leq \mathbf{f}(\mathbf{x}) \end{aligned} \tag{37}$$

fulfills GQC (or $Q_1 - Q_5$). The qualifying conditions must apply in the box $\mathcal{D}$, which we select equal to $\mathcal{X}$ (no point outside of $\mathcal{D} = \mathcal{X}$ is feasible, hence all feasible points lie in $\mathcal{D}$).

The gradients of the objective and constraint functions are

$$\nabla \mathbf{f}_0 = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \quad \text{and} \quad \nabla \mathbf{f} = \begin{bmatrix} 0 & bx_1 \\ ax_2 & 0 \end{bmatrix}.$$

Since $\nabla \mathbf{f}_0 > \mathbf{0}$ for all $\mathbf{x}$, the assumptions (GQC.a) on the objective function is always fulfilled. We will now check all sign combinations of a and b.

$a, b \geq 0$: If both a and b are non-negative, then $\nabla \mathbf{f}(\mathbf{x}) \geq \mathbf{0}$ for all $\mathbf{x}$ in $\mathcal{D}$ and condition $(Q_1.c)$ holds. To verify condition $(Q_1.b)$, one must find a norm $\|\cdot\|$ such that $\||\nabla \mathbf{f}(\mathbf{x})|\| < 1$ for all $\mathbf{x}$ in $\mathcal{D}$. When the ∞-norm is used, we get

$$\max_{\mathbf{x} \in \mathcal{D}} \||\nabla \mathbf{f}(\mathbf{x})|\|_\infty = \max_{\mathbf{x} \in \mathcal{D}} \max \{|bx_1|, |ax_2|\} \leq \max\{b, a\} < 1$$

if $a, b < 1$. Thus, when $0 \leq a, b < 1$, the problem is Fast-Lipschitz by Q_1.

$a, b \leq 0$: If a and b are instead non-positive, we have $\nabla \mathbf{f}(\mathbf{x}) \leq \mathbf{0}$ for all $\mathbf{x}$ in $\mathcal{D}$ and condition $(Q_4.b)$ is fulfilled. In order to verify $(Q_4.c)$, we need $\delta(\mathbf{x})$ and $\Delta(\mathbf{x})$. These are defined pointwise in $\mathbf{x}$, as the smallest and largest (in absolute value) element of $\nabla \mathbf{f}_0$, i.e., $\delta(\mathbf{x}) = 1$ and $\Delta(\mathbf{x}) = 2$ $\forall \mathbf{x}$. Condition $(Q_4.c)$ now requires, for all $\mathbf{x} \in \mathcal{D}$, that $\||\nabla \mathbf{f}(\mathbf{x})|\|_\infty < \delta(\mathbf{x})/\Delta(\mathbf{x}) = 1/2$. From the previous case we know that $\||\nabla \mathbf{f}(\mathbf{x})|\|_\infty \leq \max\{|a|, |b|\}$ for all $\mathbf{x} \in \mathcal{D}$, so the problem is guaranteed Fast-Lipschitz by qualifying condition Q_4, provided that $-1/2 < a, b \leq 0$. Note that this would not have met the old case (ii) in (7d), since $\mathbf{f}_0(\mathbf{x}) \neq c\mathbf{1}^T \mathbf{x}$.

$ab < 0$: When a and b have different signs, neither $(Q_1.c)$, nor $(Q_2.b)$ holds. Instead, one can try qualifying condition Q_5, which does not place any sign restrictions on $\nabla \mathbf{f}(\mathbf{x})$. It is only required, in $(Q_3.b)$, that

$$\||\nabla \mathbf{f}(\mathbf{x})|\|_\infty < \frac{\delta(\mathbf{x})}{\delta(\mathbf{x}) + \Delta(\mathbf{x})}.$$

These quantities are unchanged from the previous cases, so $\delta(\mathbf{x}) = 1, \Delta(\mathbf{x}) = 2$ and $\||\nabla \mathbf{f}(\mathbf{x})|\|_\infty \leq \max\{|a|, |b|\}$, so the problem is Fast-Lipschitz by qualifying condition Q_5 if both $|a|$ and $|b|$ are less than $1/3$. Also in this case, the Old Qualifying Conditions would not have worked since case (iii) in (7g) requires a scalar objective function.

Solution of the Problem: If problem (37) is Fast-Lipschitz by any of the cases above, the optimal point $\mathbf{x}^\star$ is found by solving $\mathbf{x}^\star = \mathbf{f}(\mathbf{x}^\star)$. We now solve the problem when $a = -0.3$ and $b = 0.3$, by iterating

$$\mathbf{x}^{k+1} := \mathbf{f}(\mathbf{x}^k). \tag{38}$$

This sequence will converge to $\mathbf{x}^\star = \mathbf{f}(\mathbf{x}^\star)$ since the qualifying conditions imply that $\mathbf{f}$ is contractive (Lemma 10). The iterates $\mathbf{x}^k$ of (38), together with the feasible region of the problem is shown in Fig. 1(a). Clearly, $\mathbf{x}^\star = \mathbf{f}(\mathbf{x}^\star) \in \mathcal{X}$, so Proposition 19
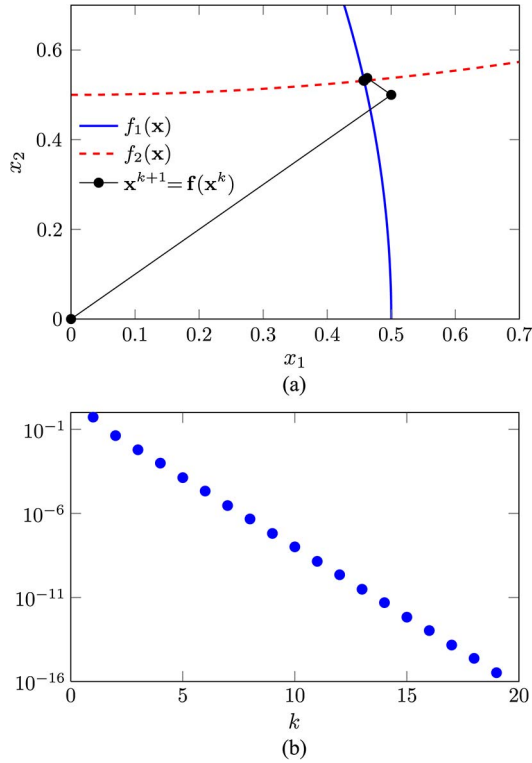
Fig. 1. Plots from the simple example in Section VI-A. (a) Illustration of the feasible region of problem (37) with $a = -0.3$ and $b = 0.3$ (making the feasible region non-convex). The iterates (38) of the solution quickly converges to $\mathbf{x}^\star = \mathbf{f}(\mathbf{x}^\star)$, where all constraints are active. (b) Convergence of the iterates (38), measured in the ∞-norm. The convergence is geometric (linear in the log domain) and the optimal solution is found within an accuracy of 10^{-6} after 8 iterations.

applies and $\mathbf{x}^\star$ is optimal also for the original problem (36). The convergence the iterations (38) is shown in Fig. 1(b).

B. Inventory Control Example

We now demonstrate how Fast-Lipschitz optimization can be used for solving Inventory control problems, a class of optimal control problems with numerous practical applications, e.g., see Example 1.1.1., pp. 3–4 in [46]. Note that, without loss of generality, the inventory control dynamics (see (39) below) can be considered as flow constraints over a network with tree topology. Therefore, our result are easily generalized to network flow optimization problems over trees.

Consider the dynamics of an inventory

$$s^{i+1} = cs^i + u^i - d^i \tag{39}$$

where at time i, $s^i \geq 0$ is the number of stored items, $u^i > 0$ is the number of ordered (or produced) items, and $d^i \geq 0$ is the demand. The parameter $c > 0$ represents the rate change in the number of stored items between time periods, i.e., storage losses if $c < 1$ or storage gains if $c > 1$. Such variations are natural in many inventories, for example $c < 1$ can represent the percentage of crops that gets ruined in storage between time periods, and $c > 1$ can represent interest rate when (39) describes a bank account where u and d are the deposits and withdrawals, respectively. The goal of inventory planning is to

find an ordering policy that minimizes the costs of ordering and storing. Formally the problem is given as follows:

$$\min_{\{u^i\}_{i=1}^N} \quad \sum_{i=1}^N a_i s^i + b_i u^i$$
$$\text{s.t.} \quad s^1 = 0$$
$$s^{i+1} = cs^i + u^i - d^i, \ i = 1, \dots, N-1$$
$$0 \leq s^i, \quad i = 2, \dots, N+1$$
$$0 \leq u^i, \quad i = 1, \dots, N \tag{40}$$

where $a_i, b_i > 0$ represent the storage costs and ordering costs, respectively, at each time i. Using the Fast-Lipschitz qualifying conditions, we can find conditions on a_i, b_i, and c, for which the optimal policy of (40) can be easily computed.

Result 23: Consider Problem (40) and set $a_{\min} = \min\{a_i\}_{i=2}^{N+1}$ and $b_{\max} = \max\{b_i\}_{i=1}^N$. If $b_{\max}(c+1) < a_{\min}$, then the optimal policy is $u^i = d^i$, for $i = 1, \dots, N$.

Before deriving Result 23, consider a relaxed version of Problem (40) denoted by (R) where the constraint $0 \leq u^i$ is relaxed. It is easily verified that if the optimal choice of $\{s^i\}_{i=2}^{N+1}$ in (R) is $s^i = 0$ for $i = 2, \dots, N+1$, then $u^i = d^i$ is the optimal policy of (40). Therefore, we next show that Result 23 holds by equivalently formulate (R) in a form which reveals that (R) actually satisfies Fast-Lipschitz qualifying conditions when $b_{\max}(c+1) < a_{\min}$, implying that $s^i = 0$.

To reveal the potential Fast-Lipschitz structure of (R), we use the change of variables $z^i = -s^i$ and $y^i = -u^i/\alpha$ and reorder the terms to obtain the equivalent formulation

$$\max_{\{y^i\}_{i=1}^N} \quad \sum_{i=1}^N a_i z^i + \alpha b_i y^i$$
$$\text{s.t.} \quad \mathbf{y} = f_{\mathbf{y}}(\mathbf{y}, \mathbf{z})$$
$$\mathbf{z} \leq 0 \tag{41}$$

where $\mathbf{y} = [y^1, \dots, y^n]^T$, $\mathbf{z} = [z^2, \dots, z^{n+1}]^T$, $[f_{\mathbf{y}}(\mathbf{y}, \mathbf{z})]_i = (z^{i+1} - cz^i)/\alpha$, and $z^1 = 0$. Problem (41) is on the form of Problem (28) and hence we can use Proposition 20. Condition (a) of Proposition 20 clearly holds since $f_{\mathbf{y}}(\mathbf{y}, \mathbf{z})$ is constant when $\mathbf{z}$ are fixed, and condition (b.iii) holds when

$$\frac{1+c}{\alpha} < \frac{a_{\min}}{\alpha b_{\max}} \tag{42}$$

where we have used that $\||\nabla_{\mathbf{z}} f_{\mathbf{y}}(\mathbf{y}, \mathbf{z})\||_\infty = (1+c)/\alpha$, $\||\nabla_{\mathbf{y}} f_{\mathbf{y}}(\mathbf{y}, \mathbf{z})\||_\infty = 0$, $\delta_{\mathbf{z}}(\mathbf{z}, \mathbf{y}) = a_{\min}$, and $\Delta_{\mathbf{y}}(\mathbf{z}, \mathbf{y}) = b_{\max}$. By multiplying (42) by α and rearranging we see that (41) and (R) are Fast-Lipschitz when $b_{\max}(c+1) < a_{\min}$.

C. Optimal Control Example

Consider a discrete time system with the state $\mathbf{s} \in \mathbb{R}^n$, the bounded and positive control $\mathbf{u} \in \mathbb{R}^p$, and the dynamics

$$\mathbf{s}^{i+1} = \mathbf{f}(\mathbf{s}^i, \mathbf{u}^i) + \mathbf{w}^i \tag{43}$$

where $\mathbf{f} : \mathbb{R}^n \times \mathbb{R}^p \to \mathbb{R}^n$ and $\mathbf{w}^i \in \mathbb{R}^n$ is an additive, bounded disturbance. At each time instance, the system has a cost $g(\mathbf{s}^i, \mathbf{u}^i)$ that is strictly increasing in all variables. The design objective is to choose the control inputs $\{\mathbf{u}^i\}_{i=1}^N$ that minimizes the accumulated cost over N periods. The optimal control

problem, given the disturbances $\{\mathbf{w}^i\}_{i=1}^N$ and the initial state $\mathbf{s}_{\text{init}}$, becomes

$$\min_{\{u^i\}_{i=1}^N} \quad \sum_{i=1}^N g(\mathbf{s}_i, \mathbf{u}_i)$$
$$\text{s.t.} \quad \mathbf{s}^1 = \mathbf{s}_{\text{init}}$$
$$\mathbf{s}^{i+1} = \mathbf{f}(\mathbf{s}^i, \mathbf{u}^i) + \mathbf{w}^i, \; i = 1, \dots, N-1$$
$$0 \le \mathbf{u}^i \le \mathbf{u}_{\max}, \quad i = 1, \dots, N. \tag{44}$$

It can be verified that if $\mathbf{f}(\mathbf{s}, \mathbf{u})$ is increasing in $\mathbf{u}$, then the optimal control is $\mathbf{u}^{i\star} = \mathbf{0}$. However, when $\mathbf{f}(\mathbf{s}, \mathbf{u})$ is decreasing in $\mathbf{u}$, then there is a tradeoff between choosing a small $\mathbf{u}^i$ to make $g(\mathbf{s}^i, \mathbf{u}^i)$ small, or a large $\mathbf{u}^i$ to make $g(\mathbf{s}^{i+1}, \mathbf{u}^{i+1})$ small. Therefore, we consider only the non-trivial case when $\mathbf{f}(\mathbf{s}, \mathbf{u})$ is decreasing in $\mathbf{u}$. With Fast-Lipschitz optimization it is possible to determine conditions on the costs and dynamics of problem (44), under which the optimal control is simply given by $\mathbf{u}^{i\star} = \mathbf{0}$ for all i.

Result 24: Consider problem (44) and assume that $g(\mathbf{s}, \mathbf{u})$ is increasing in $\mathbf{s}$ and $\mathbf{u}$ and $\mathbf{f}(\mathbf{s}, \mathbf{u})$ is decreasing in $\mathbf{u}$. Assume further that the pair $\nabla_{\mathbf{s}}\mathbf{f}(\mathbf{s}, \mathbf{u})$ and $\nabla_{\mathbf{s}}g(\mathbf{s}, \mathbf{u})$ fulfills the GQC (in place of $\nabla\mathbf{f}(\mathbf{x})$ and $\nabla f_0(\mathbf{x})$ respectively) and

$$\frac{\max_{\mathbf{s}, \mathbf{u}} \|\|\nabla_{\mathbf{u}}\mathbf{f}(\mathbf{s}, \mathbf{u})\|\|_\infty}{1 - \max_{\mathbf{s}, \mathbf{u}} \|\|\nabla_{\mathbf{s}}\mathbf{f}(\mathbf{s}, \mathbf{u})(\mathbf{s}, \mathbf{u})\|\|} < \frac{\min_{\mathbf{s}, \mathbf{u}} \min_i [\nabla_{\mathbf{u}}g(\mathbf{s}, \mathbf{u})]_i}{\max_{\mathbf{s}, \mathbf{u}} \max_i [\nabla_{\mathbf{s}}g(\mathbf{s}, \mathbf{u})]_i} \tag{45}$$

for all allowed $\mathbf{u}$ and all reachable $\mathbf{s}$. Then, the optimal solution $\{\mathbf{u}^{i\star}\}_{i=1}^N$ is given by $\mathbf{u}^{i\star} = \mathbf{0}$ for all i, regardless of the problem horizon N, the initial state $\mathbf{s}_{\text{init}}$, and the disturbances $\{\mathbf{w}^i\}_{i=1}^N$.

The result can be derived be considering a Fast-Lipschitz problem equivalent to (44) as follows. Introduce the vectors

$$\mathbf{y} = \begin{bmatrix} -\mathbf{s}^1 \\ \vdots \\ -\mathbf{s}^N \end{bmatrix} \in \mathbb{R}^{nN}, \; \mathbf{z} = \begin{bmatrix} -\mathbf{u}^1 \\ \vdots \\ -\mathbf{u}^N \end{bmatrix} \in \mathbb{R}^{pN}$$

$$\bar{\mathbf{w}} = \begin{bmatrix} \mathbf{w}^1 \\ \vdots \\ \mathbf{w}^N \end{bmatrix} \in \mathbb{R}^{nN}$$

and let $\mathbf{y}^i = -\mathbf{s}^i$, $\mathbf{z}^i = -\mathbf{u}^i$, and $\mathbf{x} = [\mathbf{y}^T \; \mathbf{z}^T]^T$. Problem (44) can then be transformed to the equivalent problem

$$\max_{\mathbf{x}} \quad f_0(\mathbf{x})$$
$$\text{s.t.} \quad \mathbf{y} = \mathbf{f}_{\mathbf{y}}(\mathbf{x})$$
$$\mathbf{z}_{\min} \le \mathbf{z} \le \mathbf{0} \tag{46}$$

where the ith $(n \times 1)$-block of $\mathbf{f}_{\mathbf{y}}(\mathbf{x})$ is given by

$$\mathbf{f}_{\mathbf{y}[i]}(\mathbf{x}) = \begin{cases} -\mathbf{s}_{\text{init}} & i = 1 \\ -\mathbf{f}(-\mathbf{y}^i, -\mathbf{z}^i) - \mathbf{w}^i & i = 2, \dots, N \end{cases}$$

and $f_0(\mathbf{x}) = -\sum_{i=1}^N g(-\mathbf{y}_i, -\mathbf{z}_i)$. By Corollary 19 and Lemma 11, (46) is Fast-Lipschitz if the relaxed problem

$$\max_{\mathbf{x}} \quad f_0(\mathbf{x})$$
$$\text{s.t.} \quad \mathbf{y} \le \mathbf{f}_{\mathbf{y}}(\mathbf{x})$$
$$\mathbf{z} \le \mathbf{0} \tag{47}$$

is Fast-Lipschitz. But (47) is on the form of (28) so Proposition 20 can be used to determine when it is Fast-Lipschitz.

To use Proposition 20 we need to determine the form of $\nabla_{\mathbf{y}}\mathbf{f}_{\mathbf{y}}(\mathbf{x})$ and $\nabla_{\mathbf{y}}\mathbf{f}_{\mathbf{y}}(\mathbf{x})$. Simple calculations show that $\nabla_{\mathbf{y}}\mathbf{f}_{\mathbf{y}}(\mathbf{x})$ and $\nabla_{\mathbf{y}}\mathbf{f}_{\mathbf{y}}(\mathbf{x})$ are $N \times N$ block matrices of $n \times n$ and $n \times p$ blocks, respectively. Moreover, all blocks are zero except the block sub-diagonal, the blocks below the block diagonal, which are $(\nabla_{\mathbf{s}}\mathbf{f}(-\mathbf{y}^1, -\mathbf{z}^1), \dots, \nabla_{\mathbf{s}}\mathbf{f}(-\mathbf{y}^{N-1}, -\mathbf{z}^{N-1}))$ and $(\nabla_{\mathbf{u}}\mathbf{f}(-\mathbf{y}^1, -\mathbf{z}^1), \dots, \nabla_{\mathbf{u}}\mathbf{f}(-\mathbf{y}^{N-1}, -\mathbf{z}^{N-1}))$ for $\nabla_{\mathbf{y}}\mathbf{f}_{\mathbf{y}}(\mathbf{x})$ and $\nabla_{\mathbf{y}}\mathbf{f}_{\mathbf{y}}(\mathbf{x})$, respectively.

Now we can verify that condition (a) of Proposition 20 holds, since $\nabla_{\mathbf{y}}\mathbf{f}_{\mathbf{y}}(\mathbf{x})$ and $\nabla_{\mathbf{y}}\mathbf{f}_{\mathbf{y}}(\mathbf{x})$ are sub-diagonal with $\nabla_{\mathbf{s}}\mathbf{f}$ and $\nabla_{\mathbf{u}}\mathbf{f}$, respectively, on the sub-diagonal and by the assumptions of Remark (24) $\nabla_{\mathbf{s}}\mathbf{f}$ and $\nabla_{\mathbf{u}}\mathbf{f}$ fulfill the GQC. Next we determine under what condition (b.iii) of Proposition 20 holds. By using again the structure of $\nabla_{\mathbf{y}}\mathbf{f}_{\mathbf{y}}(\mathbf{x})$ and $\nabla_{\mathbf{y}}\mathbf{f}_{\mathbf{y}}(\mathbf{x})$ we get

$$\|\|\nabla_{\mathbf{y}}\mathbf{f}_{\mathbf{y}}(\mathbf{x})\|\|_\infty \le \max_{\mathbf{s}, \mathbf{u}} \|\|\nabla_{\mathbf{s}}\mathbf{f}(\mathbf{s}, \mathbf{u})\|\|_\infty \tag{48}$$

$$\|\|\nabla_{\mathbf{z}}\mathbf{f}_{\mathbf{y}}(\mathbf{x})\|\|_\infty \le \max_{\mathbf{s}, \mathbf{u}} \|\|\nabla_{\mathbf{u}}\mathbf{f}(\mathbf{s}, \mathbf{u})\|\|_\infty \tag{49}$$

$$\Delta_{\mathbf{y}}(\mathbf{x}) = \max_i [\nabla_{\mathbf{y}}f_0(\mathbf{x})]_i \le \max_{\mathbf{s}, \mathbf{u}} \max_i [\nabla_{\mathbf{s}}g(\mathbf{s}, \mathbf{u})]_i \tag{50}$$

$$\delta_{\mathbf{z}}(\mathbf{x}) = \min_i [\nabla_{\mathbf{z}}f_0(\mathbf{x})]_i \le \min_{\mathbf{s}, \mathbf{u}} \min_i [\nabla_{\mathbf{u}}g(\mathbf{s}, \mathbf{u})]_i \tag{51}$$

where the optimizations is over the permissible controls $\mathbf{u}$ and the reachable states $\mathbf{s}$. Condition (b.iii) is now fulfilled if

$$\frac{\|\|\nabla_{\mathbf{z}}\mathbf{f}_{\mathbf{y}}(\mathbf{x})\|\|_\infty}{1 - \|\|\nabla_{\mathbf{y}}\mathbf{f}_{\mathbf{y}}(\mathbf{x})\|\|_\infty} < \frac{\delta_{\mathbf{z}}(\mathbf{x})}{\Delta_{\mathbf{y}}(\mathbf{x})}$$

i.e., if

$$\frac{\max_{\mathbf{s}, \mathbf{u}} \|\|\nabla_{\mathbf{u}}\mathbf{f}(\mathbf{s}, \mathbf{u})\|\|_\infty}{1 - \max_{\mathbf{s}, \mathbf{u}} \|\|\nabla_{\mathbf{s}}\mathbf{f}(\mathbf{s}, \mathbf{u})\|\|_\infty} < \frac{\min_{\mathbf{s}, \mathbf{u}} \min_i [\nabla_{\mathbf{u}}g(\mathbf{s}, \mathbf{u})]_i}{\max_{\mathbf{s}, \mathbf{u}} \max_i [\nabla_{\mathbf{s}}g(\mathbf{s}, \mathbf{u})]_i}.$$

When the inequality above holds, problem (47) fulfills Proposition 20, whereby problems (47) and (46) are Fast-Lipschitz. This implies that the optimal solution is given by $\mathbf{y}^\star = \mathbf{f}_{\mathbf{y}}(\mathbf{y}^\star, \mathbf{z}^\star)$ and $\mathbf{z}^\star = \mathbf{0}$ and hence the optimal solution of (44) is given by $\mathbf{u}^{i\star} = -\mathbf{z}^{i\star} = \mathbf{0}$ for all i.

We now apply Result 24 on two optimal control problems.

a) Linear first order system: Consider for illustrative purposes a linear first order system with a linear cost function. The optimal control problem (44) becomes

$$\min_{\{u^i\}_{i=1}^N} \quad \sum_{i=1}^N g(s_i, u_i)$$
$$\text{s.t.} \quad s^1 = s_{\text{init}}$$
$$s^{i+1} = f(s^i, u^i) + w^i, \; i = 1, \dots, N-1$$
$$0 \le u^i \le u_{\max}, \quad i = 1, \dots, N \tag{52}$$

where $f(s, u) = as - bu$ and $g(s, u) = c_s s + c_u u$ with $a, b, c_s, c_u > 0$. In this case, the gradients are given by $\nabla_s f(s, u) = a$, $\nabla_u f(s, u) = -b$, $\nabla_s g(s, u) = c_s$ and $\nabla_u g(s, u) = c_u$. Since the gradients are scalar and constant, Result 24 applies to problem (52) provided that $a < 1$ and $a/(1-a) < c_u/c_s$. If this is the case, the optimal solution must be $u^{i\star} = 0$ for all i. We remark that this is an important and non obvious result.

b) Non-linear first order system: Consider again problem (52), but with the dynamics given by

$$f(s, u) = \frac{as}{1+s} s - bu. \tag{53}$$

We continue to assume $a, b > 0$, and the cost function $g(s, u)$ is left unchanged from the linear case. The factor $as/(s+1)$ can be seen as a variable decay-rate, under which large states decay with a factor close to a while small states decay almost instantly. Assume that the disturbances w^i are non-negative, and let the upper limit of u^i depend on the current state s^i, $u \leq u_{\max}(s^i)$ where $u_{\max}(s) = as^2/b(1+s)$. Note that Corollary 19 still applies when moving from problem (46) to problem (47) because $u^{i\star} = 0 \leq u_{\max}(s^i)$, so the optimal values $u^{i\star} = 0$ of the relaxed problem (47) are always feasible in problem (46). The modified constraints $u_{\max}(s^i)$ ensure $f(s^i, u^i) \geq 0$, whereby a non-negative $s_{\text{init}} \geq 0$ implies that all future states s^i are non-negative. All gradients from the linear example are left unchanged, except

$$\nabla_s f(s, u) = a \frac{s^2 + 2s}{s^2 + 2s + 1}.$$

This gradient can be bounded by $0 \leq \nabla_s f(s, u) \leq a$, since the modified constraint $u_{\max}(s^i)$ ensures that only non-negative states s can be reached. Therefore, it holds that $\max_{s,u} \| \| \nabla_s f(s, u) \| \|_\infty \leq a$ whereby the conditions

$$|a| < 1 \quad \text{and} \quad \frac{b}{1-a} < \frac{c_u}{c_s} \qquad (54)$$

remain unchanged from the linear example. When these conditions hold, we know that $u^i = 0$ for all i is the optimal solution. Because of the non-affine equality constraints in (53), problem (54) is a non-convex problem in N variables subject to $2N$ non-trivial constraints when $u_{\max}$ is a function of s^i. However, by using the Fast-Lipschitz properties of the problem we have solved it without performing any calculations, except for those involved in verifying the assumptions.

We now conclude this example by numerically solving two instances of problem (52).

c) Numerical example: Consider two instances of problem (52) with the non-linear dynamics (53) denoted by (P1) and (P2). (P1) has $b = 0.3$ and (P2) has $b = 0.5$, otherwise the problems share all parameters where $N = 20$, $s_{\text{init}} = 1$, $a = 0.5$, $c_s = 3$, $c_u = 2$, and w^i are uniform random on [0,1]. Both (P1) and (P2) have $0 \leq a < 1$, but only (P1) fulfills (54). We can therefore guarantee that $u^i = 0$ for all i solves (P1), but we cannot say anything about (P2), since the conditions in this paper are only sufficient. The two problems are solved numerically with Matlab's built-in solver fmincon, and the results are shown in Fig. 2. As expected by Result 24, the optimal actions for (P1) is to always keep the control variables at the lower bound and (P2) illustrates a case where Result 24 does not apply.

D. Fast-Lipschitz Versus an Interior Point Method

Interior point methods (IPM) are usually the main candidate for solving constraint convex optimization problems. However, when a problem is also Fast-Lipschitz then the optimization problem can also be solved by finding the fixed point of the constraints. In this section, we compare the two approaches
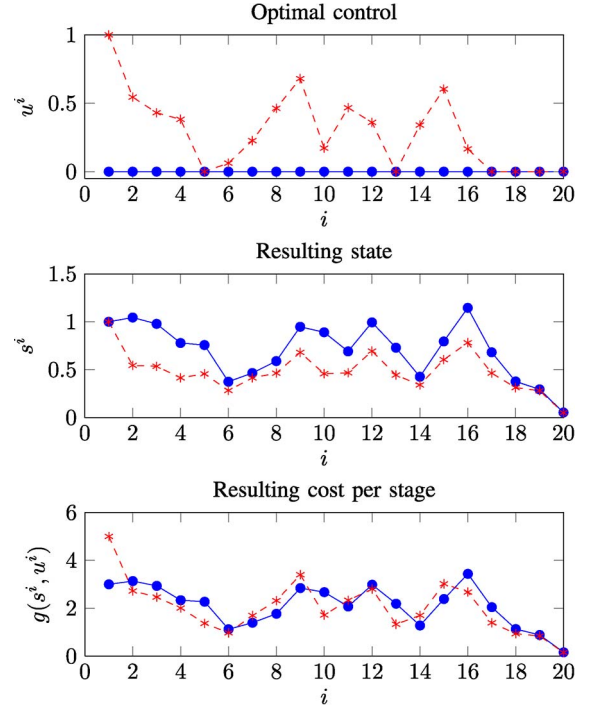

Fig. 2.   The optimal control and the resulting states and cost for Problems (P1) (blue circles) and (P2) (red stars), described in Section VI-C c).
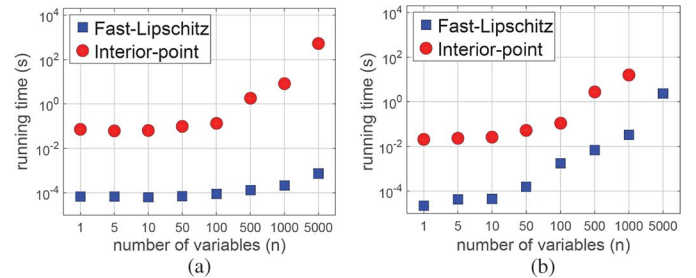

Fig. 3.   The running times for solving Sparse and Full by exploiting their Fast-Lipschitz structure and finding the fixed point of the constraint and by using an interior point method. (a) Sparse. (b) Full.

on problems of varying size that are both convex and Fast-Lipschitz. We consider problems on the form

$$\min_{\mathbf{x}} \quad \sum_{i=1}^{n} e^{-x_i}$$
$$\text{s.t.} \quad \mathbf{x} \leq \mathbf{A}\mathbf{x} + \mathbf{r}, \quad \forall i \in \{1, \dots, n\} \qquad (55)$$

where $\mathbf{x} = \{x_i\}_{i=1}^n \in \mathbb{R}^n$ and $\mathbf{r} \in \mathbb{R}^n$. We consider two instances of problem (55), Sparse: where $[A]_{i,i+1} = [A]_{i+1,i} = 1/4$ for $i = 1, \dots, n-1$ and $[A]_{i+1,i} = 0$ otherwise and the elements of $\mathbf{r}$ are uniform random on [0, 4], and Full: where $[A]_{i,j} = 1/(2n)$ for $i \neq j$ and $[A]_{i,j} = 0$ otherwise and the elements of $\mathbf{r}$ are uniform random on [0, 1]. Both problem instances are convex and Fast-Lipschitz.

We compare the running times needed to solve Sparse and Full using an IPM and solving $(\mathbf{I} - \mathbf{A})\mathbf{x} = \mathbf{r}$, i.e., exploiting the Fast-Lipschitz structure. We use Matlab's built-in solvers, the IPM in fmincon and backslash to solve $(\mathbf{I} - \mathbf{A})\mathbf{x} = \mathbf{r}$. The problem size n varies from 1 to 5000.

Fig. 3 depicts the results, where each data point is an average of 10 runs. The worst case relative standard deviation (std), i.e., std relative to the mean, of a data point is 0.26. The

data point for $n = 5000$ in `Full` is missing since the IPM always failed due to memory error. In all cases, solving the linear system is significantly faster than using the IPM. The differences are particularly apparent in `Sparse` (Fig. 3(a)) where Fast-Lipschitz goes from being 10^3 times faster than the IPM when $n = 1$ to being $7 \cdot 10^5$ times faster when $n = 5000$. This superiority of solving a system of equations over solving general constraint optimization problems is intuitive because most optimization techniques, including IPM's, are based on sequentially solving systems of equations with the same number of variables as the original problem (or more variables, e.g., primal-dual methods). Therefore, when a convex problem is known to be Fast-Lipschitz it can significantly accelerate the convergence speed by exploiting the Fast-Lipschitz structure.

VII. Conclusion

In this paper we significantly extended the previous Fast-Lipschitz framework proposed in [28]. A new set of qualifying conditions, that unify and generalize previous conditions, was presented. Furthermore, we investigated problems that deviate from the required Fast-Lipschitz form, either by not having the same number of constraints as variables or by having variables that do not affect the objective function. For these cases we established conditions for which they are still Fast-Lipschitz. Based on these new results, a larger set of convex and non-convex optimization problems can be solved by the Fast-Lipschitz method of this paper, both in a centralized and in a distributed set-up. This avoids using Lagrangian methods, which are inefficient in terms of computations and communication complexity especially when used over networks.

Several possible extensions remain to consider, for example:

- We believe that the Fast-Lipschitz optimization framework can be extended to cover also non-smooth problems. A potential benefit of such an extension would be to form expressions such as $\mathbf{x} \leq \mathbf{f}(\mathbf{x}) = \min\{\mathbf{f}_1(\mathbf{x}), \mathbf{f}_2(\mathbf{x})\}$, whereby problems with more constraints than variables ($(\mathbf{f}(\mathbf{x}) : \mathbb{R}^n \to \mathbb{R}^p$ where $p > n$) could be considered.

- So far, only problems with $\nabla \mathbf{f}_0(\mathbf{x}) \geq \mathbf{0}$ have been considered. The standard inequality we have used is a partial ordering induced by the non-negative orthant $\mathbb{R}_+^m$, i.e., $\nabla \mathbf{f}_0(\mathbf{x}) \succeq_{\mathbb{R}_+^m} \mathbf{0}$. A possible extension is to allow for problems where $\nabla \mathbf{f}_0(\mathbf{x}) \succeq_{\mathcal{K}} \mathbf{0}$ for a more general cone $\mathcal{K}$. This would allow a tradeoff between conditions for $\nabla \mathbf{f}_0(\mathbf{x})$ and conditions on $\nabla \mathbf{f}(\mathbf{x})$, which may give more flexibility in the qualifying conditions.

  We have already seen a similar example of this, when case (ii) of the old conditions is generalized to Q_2. In this case, condition (7d) required $\mathbf{f}_0(\mathbf{x}) = c\mathbf{1}^T\mathbf{x}$ for some $c > 0$, which can also be stated

$$\mathbf{f}_0(\mathbf{x}) \in \mathbb{R}, \quad \text{and} \quad \nabla \mathbf{f}_0(\mathbf{x}) \succeq_{\mathcal{K}_1} \mathbf{0}$$

where $\mathcal{K}_1$ is the cone (ray) generated by the single vector $\mathbf{1}$. In Q_2 it is instead required that $\nabla \mathbf{f}_0(\mathbf{x}) > \mathbf{0}$, which is more general. This, however, comes at the price of a less general requirement on $\mathbf{f}(\mathbf{x})$, since $(Q_2.c)$ requires

$$\|\|\nabla \mathbf{f}(\mathbf{x})\|\|_\infty < \mathfrak{q}(\mathbf{x}) \leq 1$$

while (7f) only requires $\|\|\nabla \mathbf{f}(\mathbf{x})\|\|_\infty < 1$.

References

[1] M. Jakobsson and C. Fischione, "Extensions of Fast-Lipschitz optimization for convex and non-convex problems," in *Proc. 3rd IFAC Workshop Distrib. Estim. Control Netw. Syst. (NecSys)*, 2012. [Online]. Available: http://www.researchgate.net/publication/269259498_Extensions_of_Fast-Lipschitz_Optimization_for_Convex_and_Non-Convex_Problems

[2] J. Tsitsiklis, "Problems in Decentralized Decision Making and Computation," Ph.D. dissertation, MIT, Cambridge, MA, USA, 1984.

[3] J. Tsitsiklis, D. Bertsekas, and M. Athans, "Distributed asynchronous deterministic and stochastic gradient optimization algorithms," *IEEE Trans. Autom. Control*, vol. 31, pp. 803–812, 1986.

[4] S. Boyd, L. Xiao, A. Mutapcic, and J. Mattingley, Notes on Decomposition Methods, 2007.

[5] F. P. Kelly, A. K. Maulloo, and D. K. H. Tan, "Rate control for communication networks: Shadow prices, proportional fairness and stability," *J. Oper. Res. Soc.*, vol. 49, no. 3, 1998. [Online]. Available: http://www.jstor.org/stable/3010473?seq=1#page_scan_tab_contents

[6] S. Low and D. Lapsley, "Optimization flow control—I: Basic algorithm and convergence," *IEEE/ACM Trans. Netw.*, vol. 7, pp. 861–874, 1999.

[7] D. Palomar and M. Chiang, "Alternative distributed algorithms for network utility maximization: Framework and applications," *IEEE Trans. Autom. Control*, vol. 52, no. 12, pp. 2254–2269, 2007.

[8] E. Wei, A. Ozdaglar, and A. Jadbabaie, "A Distributed Newton Method for Network Utility Maximization, I: Algorithm," LIDS, Tech. Rep. 2832, 2011.

[9] D. Gabay and B. Mercier, "A dual algorithm for the solution of nonlinear variational problems via finite element approximation," *Comp. Math. Applicat.*, vol. 2, no. 1, pp. 17–40, 1976.

[10] J. Eckstein and D. P. Bertsekas, "On the douglas-rachford splitting method and the proximal point algorithm for maximal monotone operators," *Math. Programming*, vol. 55, pp. 293–318, 1992.

[11] I. Schizas, A. Ribeiro, and G. Giannakis, "Consensus in ad hoc WSNs with noisy links—Part I: Distributed estimation of deterministic signals," *IEEE Trans. Signal Process.*, vol. 56, no. 1, pp. 350–364, 2008.

[12] I. Schizas, G. Giannakis, S. Roumeliotis, and A. Ribeiro, "Consensus in ad hoc WSNs with noisy links—Part II: Distributed estimation and smoothing of random signals," *IEEE Trans. Signal Process.*, vol. 56, no. 4, pp. 1650–1666, 2008.

[13] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends Mach. Learning*, vol. 3, no. 1, pp. 1–122, 2011.

[14] A. Jadbabaie, J. Lin, and A. S. Morse, "Coordination of groups of mobile autonomous agents using nearest neighbor rules," *IEEE Trans. Autom. Control*, vol. 48, pp. 988–1001, Jun. 2003.

[15] S. Boyd, A. Ghosh, B. Prabhakar, and D. Shah, "Randomized gossip algorithms," *IEEE Trans. Inform. Theory*, vol. 52, pp. 2508–2530, 2006.

[16] A. Dimakis, A. Sarwate, and M. Wainwright, "Geographic gossip: Efficient averaging for sensor networks," *IEEE Trans. Signal Process.*, vol. 56, no. 3, pp. 1205–1216, 2008.

[17] A. Olshevsky and J. Tsitsiklis, "Convergence speed in distributed consensus and averaging," *SIAM J. Control Optm.*, vol. 48, no. 1, pp. 33–55, 2009.

[18] A. Chiuso, F. Fagnani, L. Schenato, and S. Zampieri, "Gossip algorithms for distributed ranking," in *Proc. Amer. Control Conf.*, 2011, pp. 5468–5473.

[19] A. Nedić and A. Ozdaglar, "Distributed subgradient methods for multi-agent optimization," *IEEE Trans. Autom. Control*, vol. 54, pp. 48–61, 2009.

[20] A. Nedić, A. Ozdaglar, and P. Parrilo, "Constrained consensus and optimization in multi-agent networks," *IEEE Trans. Autom. Control*, vol. 55, pp. 922–938, Apr. 2010.

[21] S. Sundhar Ram, A. Nedić, and V. Veeravalli, "Distributed stochastic subgradient projection algorithms for convex optimization," *J. Optim. Theory Applicat.*, vol. 147, pp. 516–545, 2010.

[22] J. Duchi, A. Agarwal, and M. Wainwright, "Dual averaging for distributed optimization: Convergence analysis and network scaling," *IEEE Trans. Autom. Control*, vol. 57, pp. 592–606, Mar. 2012.

[23] K. I. Tsianos, S. Lawlor, and M. G. Rabbat, "Push-sum distributed dual averaging for convex optimization," in *Proc. IEEE 51st Conf. Decision Control (CDC)*, Dec. 2012, pp. 5453–5458.

[24] M. Zhu and S. Martinez, "On distributed convex optimization under inequality and equality constraints," *IEEE Trans. Autom. Control*, vol. 57, pp. 151–164, 2012.

[25] F. Zanella, D. Varagnolo, A. Cenedese, G. Pillonetto, and L. Schenato, "Newton-Raphson consensus for distributed convex optimization," in *Proc. IEEE 50th Conf. Decision Control*, 2011, pp. 5917–5922.

[26] M. Zargham, A. Ribeiro, and A. Jadbabaie, "Network optimization under uncertainty," in *Proc. IEEE 51st Conf. Decision Control*, Dec. 2012, pp. 7470–7475.

[27] K. I. Tsianos, S. Lawlor, and M. G. Rabbat, "Communication/computation tradeoffs in consensus-based distributed optimization," *Adv. Neural Inform. Process. Syst. 25*, pp. 1952–1960, 2012.

[28] C. Fischione, "Fast-Lipschitz optimization with wireless sensor networks applications," *IEEE Trans. Autom. Control*, vol. 56, pp. 2319–2331, 2011.

[29] G. Notarstefano and F. Bullo, "Distributed abstract optimization via constraints consensus: Theory and applications," *IEEE Trans. Autom. Control*, vol. 56, pp. 2247–2261, 2011.

[30] H. Tuy, "Monotonic optimization: Problems and solution approaches," *SIAM J. Optim.*, vol. 11, no. 2, pp. 464–494, 2000.

[31] E. Jorswieck and E. Larsson, "Monotonic optimization framework for the two-user MISO interference channel," *IEEE Trans. Commun.*, vol. 58, no. 7, pp. 2159–2168, 2010.

[32] E. Björnson, G. Zheng, M. Bengtsson, and B. Ottersten, "Robust monotonic optimization framework for multicell MISO systems," *IEEE Trans. Signal Process.*, vol. 60, no. 5, pp. 2508–2523, 2012.

[33] J. Zander, "Distributed cochannel interference control in cellular radio systems," *IEEE Trans. Veh. Technol.*, vol. 41, pp. 305–311, Aug. 1992.

[34] S. Grandhi, R. Vijayan, and D. Goodman, "Distributed power control in cellular radio systems," *IEEE Trans. Commun.*, vol. 42, pp. 226–228, 1994.

[35] G. Foschini and Z. Miljanic, "A simple distributed autonomous power control algorithm and its convergence," *IEEE Trans. Veh. Technol.*, vol. 42, pp. 641–646, 1993.

[36] R. Yates, "A framework for uplink power control in cellular radio systems," *IEEE J. Sel. Areas Commun.*, vol. 13, pp. 1341–1347, 1995.

[37] J. Herdtner and E. Chong, "Analysis of a class of distributed asynchronous power control algorithms for cellular wireless systems," *IEEE J. Sel. Areas Commun.*, vol. 18, no. 3, pp. 436–446, 2000.

[38] D. P. Bertsekas and J. Tsitsiklis, *Parallel and Distributed Computation: Numerical Methods*. Belmont, MA, USA: Athena Scientific, 1997.

[39] C. Sung and K. Leung, "A generalized framework for distributed power control in wireless networks," *IEEE Trans. Inf. Theory*, vol. 51, pp. 2625–2635, 2005.

[40] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.

[41] C. Fischione, F-Lipschitz Optimization, Tech. Rep. IR-EE-RT 2010:023. [Online]. Available: http://www.ee.kth.se/php/modules/publications/reports/2010/IR-EE-RT_2010_023.pdf

[42] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 1985.

[43] D. P. Bertsekas, *Nonlinear Programming*, 2nd ed. Belmont, MA, USA: Athena Scientific, Sep. 1999.

[44] M. Jakobsson and C. Fischione, "Optimality of radio power control via fast-Lipschitz optimization," *Available on ArXiv*, 2014.

[45] K. B. Petersen and M. S. Pedersen, *The Matrix Cookbook*, Nov. 14, 2008. [Online]. Available: http://www.math.uwaterloo.ca/~hwolkowi/matrixcookbook.pdf

[46] D. P. Bertsekas, *Dynamic Programming and Optimal Control Volume I*, 3rd ed. Belmont, MA, USA: Athena Scientific, 2005.

Sindri Magnússon received the B.Sc. degree in mathematics from the University of Iceland, Reykjavik, Iceland, in 2011, and the M.S. degree in Mathematics from KTH Royal Institute of Technology, Stockholm, Sweden, in 2013, where he is currently pursuing the Ph.D. degree.

He has held a six-months visiting research position at Harvard University, Cambridge, MA, sponsored by the Engblom Foundation. His research interests include distributed optimization, both theory and applications.

Carlo Fischione (M'05) received the the Dr.Eng. degree in electronic engineering (with highest honors) and the Ph.D. degree in electrical and information engineering from the University of L'Aquila, Italy, in 2001 and 2005, respectively.

He is a tenured Associate Professor at KTH Royal Institute of Technology, Electrical Engineering and ACCESS Linnaeus Center, Automatic Control Lab, Stockholm, Sweden. He has held research positions at the Massachusetts Institute of Technology (MIT), Cambridge, 2015, the University of California at Berkeley, Berkeley (2004–2005, Visiting Scholar, and 2007–2008, Research Associate), and Royal Institute of Technology, Stockholm, Sweden (2005–2007, Research Associate). He has co-authored over 100 publications, including book, book chapters, international journals and conferences, and international patents. His current research interests include optimization and parallel computation with applications to wireless sensor networks, cyberphysical systems, and wireless networks.

Dr. Fischione received the best paper award from the IEEE TRANSACTIONS ON INDUSTRIAL INFORMATICS in 2007, the Best Paper Award at the IEEE International Conference on Mobile Ad-hoc and Sensor System 2005 and 2009, the Best Paper Award of the IEEE Sweden VT-COM-IT Chapter of 2014, the Best Business Idea award from VentureCup East Sweden, 2010, the Ferdinando Filauro award from University of LAquila, Italy, 2003, the Higher Education award from Abruzzo Region Government, Italy, 2004, and the Junior Research award from Swedish Research Council, 2007, the Silver Ear of Wheat award in history from the Municipality of Tornimparte, Italy, 2012. He has chaired or served as a technical member of program committees of several international conferences and is serving as referee for technical journals. Meanwhile, he also has offered his advice as a consultant to numerous technology companies such as Berkeley Wireless Sensor Network Lab, Ericsson Research, Synopsys, and United Technology Research Center. He is co-funder and CTO of the sensor networks start-up companies Lokkupp (indoor navigation) and Modern Ancient Instruments Networked (MIND). He is an Ordinary Member of the academy of history Deputazione Abruzzese di Storia Patria (DASP).

Martin Jakobsson received the M.Eng. degree in systems and control and the Licentiate degree in electrical engineering from KTH Royal Institute of Technology, Stockholm Sweden, in 2011 and 2014, respectively.

He is currently an Engineer at Scania AB, Södertälje, Sweden.

Pradeep Chathuranga Weeraddana (S'08–M'11) received the M.Eng degree in telecommunication from the School of Engineering and Technology, Asian Institute of Technology, Thailand, in 2007 and the Ph.D. degree from University of Oulu, Finland, in 2011.

He was a Postdoctoral Researcher in the Department of Automatic Control, School of Electrical Engineering and ACCESS Linnaeus Center, KTH Royal Institute of Technology, Stockholm, Sweden. He is currently a Senior Lecturer at the Sri Lankan Institute of Information Technology, Malabe, Sri lanka. His research interests include application of optimization techniques in various application domains, such as signal processing, wireless communications, smart grids, privacy, and security.