Fast–Lipschitz Optimization with Wireless Sensor Networks Applications

Carlo Fischione
Automatic Control Lab
KTH Royal Institute of Technology
100-44, Stockholm, Sweden
carlofi@ee.kth.se

Ulf Jönsson
Optimization and Systems Theory
KTH Royal Institute of Technology
100-44, Stockholm, Sweden
ulfj@math.kth.se

ABSTRACT

Motivated by the need for fast computations demanded by wireless sensor networks, the new F-Lipschitz optimization theory is introduced for a novel class of optimization problems. These problems are defined by simple qualifying properties specified in terms of increasing objective function and contractive constraints. It is shown that feasible F-Lipschitz problems have always a unique optimal solution that satisfies the constraints at equality. The solution is obtained quickly by asynchronous algorithms of certified convergence. F-Lipschitz optimization can be applied to both centralized and distributed optimization. Compared to traditional Lagrangian methods, which often converge linearly, the convergence time of centralized F-Lipschitz problems is at least superlinear. Distributed F-Lipschitz algorithms converge fast, as opposed to traditional Lagrangian decomposition and parallelization methods, which generally converge slowly and at the price of many message passings. In both cases, the computational complexity is much lower than traditional Lagrangian methods. Examples of application of the new optimization method are given for distributed detection and radio power control in wireless sensor networks. The drawback of the F-Lipschitz optimization is that it might be difficult to check the qualifying properties. For more general optimization problems, it is suggested that it is convenient to have conditions ensuring that the solution satisfies the constraints at equality.

Categories and Subject Descriptors

G.1.6 [**Mathematics of Computing**]: Optimization—*Constrained optimization*

General Terms

Algorithms, Design, Performance, Standardization, Theory.

Keywords

Wireless Sensor Networks; Convex and Non-convex Optimization; Distributed Optimization; Parallel and Distributed Computation; Interference Function Theory.

IPSN'11, April 12–14, 2011, Chicago, Illinois.

1. INTRODUCTION

Numerous engineering applications have pushed a huge development of mathematical optimization theory in recent years [1]. From large scale systems, to smaller systems, optimization is a pervasive tool to take decisions, improve efficiency, and reduce the development costs. Wireless sensor networks, with their applications to smart grids, water distribution, vehicular networks, are networked systems in which decision variables must be quickly optimized by algorithms of light computational cost. Unfortunately, many traditional optimization methods are difficult to use in wireless sensor networks due to the complex operations or the high number of messages that have to be exchanged among nodes to compute the optimal solution.

Wireless sensor networks are characterized by small and cheap hardware platforms. Thus they have limited computational capabilities. It follows that there is the need of fast, simple, and robust to errors and noises computations to solve optimization problems, both in a centralized and in a distributed set-up [2] – [5]. In networked systems, the computation of the solution to these problems must be distributed when the dynamics of the communication channels or the topology may quickly change and the network is large. There may be lack of central coordination, it could be difficult to have such a coordination, or it could be simply impossible to have a coordination point where solutions to problems that involve information from the overall network can be achieved. See the IPSN paper [2] for an interesting discussion on optimization for wireless sensor networks. In all these cases, the solution must be computed by distributed algorithms, where the computation is split into sub tasks that are distributed to the local nodes of the networks that cooperate in parallel [6, 7].

Convex optimization has played a dominant role in many engineering problems [8], both centralized and distributed. Problems are often approximated as convex ones, given the availability of interior point methods based on Lagrangian duality to compute the optimal solution [1]. However, not all problems are convex and it is well known that some convex problems can be solved by faster methods, such as the iterative contraction mappings of the interference function optimization theory used in wireless communication [9]. The decomposition methods developed in the parallel and distributed computation theory are the fundamental reference mathematical tools to cope with distributed optimization [6]. However, these methods often converge slowly due to many message exchanges among the nodes, and do not consider that the communication among nodes may be extremely expensive. This is the typical case of wireless sensor networks, where transmitting information demands about one hundred times the energy needed to perform computations [3]. In this case, it is important that the number of messages exchanged among nodes to compute the optimal solu-

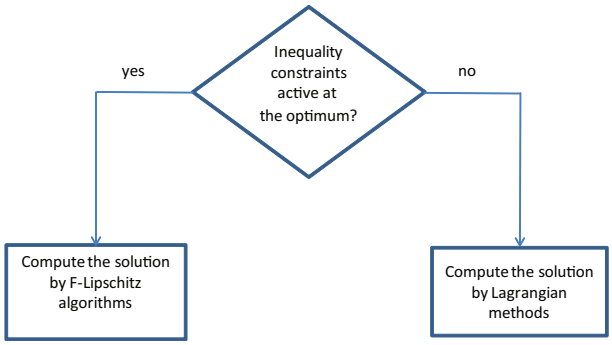

Figure 1: F-Lipschitz optimization defines a class of optimization problems for which all the constraints are satisfied at equality when computed at the optimal solution, including the inequality constraints. The solution to the set of equations given by the projected constraints is the optimal solution, which avoids using Lagrangian methods. These methods are computationally much more expensive, particularly for distributed optimization over wireless sensor networks. The challenging part of F-Lipschitz optimization is the availability of conditions ensuring that the constraints are active at the optimum without knowing what is the optimal solution in advance. For this reason, we restrict ourselves to problems in the form (3.1) **presented in Section 3. However, the method can be used for problems in the more general form** (3.3)**, as we see in Section 3.2.**

tion be as limited as possible.

In this paper, we propose a new optimization approach particularly suitable for wireless sensor networks. We define a new class of problems that is characterized by the availability of fast and simple algorithms for the computation of the optimal solution. We denote it F-Lipschitz to evidence that it is fast, hence "F", and based on a Lipschitz property of the constraints. In particular, it is assumed that the objective function is increasing, whereas the constraints be transformed into contractive Lipschitz functions. However, we show that in a number of cases problems in canonical form [8] can be cast to F-Lipschitz ones. To compute the optimal solution to centralized optimization problems, F-Lipschitz algorithms do not require Lagrangian methods, but superlinear iterations based on a solution of a system of equations given by the constraints, as illustrated in Fig. 1. In distributed optimization, we propose an algorithm that do not require Lagrangian decomposition and parallelization methods, but simple asynchronous iterative methods. We show that the computation of the optimal solution of an F-Lipschitz problem is robust to quantization errors and not sensitive to perturbation of the constraints, which is quite important for wireless sensor networks with nodes having low computational precision. We show that F-Lipschitz optimization solves much more efficiently problems traditionally solved by Lagrangian methods. The approach presented in this paper covers several problems in the general interference function theory [9]–[11]. Fig. 2 illustrates the intersection between the F-Lipschitz optimization and other classic optimization areas.

The remainder of the paper is organized as follows: In Section 2, two wireless sensor networks motivating examples are presented. The definition of F-Lipschitz optimization is given in Section 3, along with properties and algorithms to compute the optimal solution. In Section 4 we show that the motivating examples of Section 2 are F-Lipschitz and we illustrate how they are solved. Finally, conclusions and future perspectives are given in Section 5.

1.1 Notation

We use the notation $\mathbb{R}^n_+$ to denote the set of strictly positive valued real vectors. By $\cdot^T$ we denote the transpose of a vector or of a matrix. By $|\cdot|$ we denote the absolute value of a real number. For $\boldsymbol{x} \in \mathbb{R}^n$ we let

$$\|\boldsymbol{x}\|_1 = \sum_{k=1}^{n} |x_k| \quad \text{and} \quad \|\boldsymbol{x}\|_\infty = \max_{k=1}^{n} |x_k|,$$

and we will use the duality relation $|\boldsymbol{x}^T\mathbf{y}| \leq \|\boldsymbol{x}\|_1 \|\mathbf{y}\|_\infty$. For a matrix $\mathbf{A} \in \mathbb{R}^{n\times n}$ we use the induced $\|\cdot\|_\infty$ and induced $\|\cdot\|_1$ norms defined as

$$\|\mathbf{A}\|_1 = \max_{i=1}^{n} \sum_{j=1}^{n} |a_{ij}| \quad \text{and} \quad \|\mathbf{A}\|_\infty = \max_{j=1}^{n} \sum_{i=1}^{n} |a_{ij}|.$$

The spectral radius of a matrix is defined as $\rho(\mathbf{A}) = \max\{|\lambda| : \lambda \in \text{eig}(\mathbf{A})\}$.

By $\mathbf{a} \preceq \mathbf{b}$ and $\mathbf{a} \succeq \mathbf{b}$ we denote the element-wise inequalities between the vectors $\mathbf{a}$ and $\mathbf{b}$. A matrix is called positive, $\mathbf{A} \succeq 0$, if $a_{ij} \geq 0$, $\forall i, j$. By $\mathbf{I}$ and $\mathbf{1}$ we denote the identity matrix and the vector $(1, \ldots, 1)^T$, respectively, whose dimensions are clear from the context. Given the set $\mathscr{D} = [x_{1,\min}, x_{1,\max}] \times [x_{2,\min}, x_{2,\max}] \ldots [x_{n,\min}, x_{n,\max}] \in \mathbb{R}^n$, with $-\infty < x_{i,\min} < x_{i,\max} < \infty$, for $i = 1, \ldots, n$, and the vector $\boldsymbol{x} \in \mathscr{D}$, we use the notation $[x_i]^{\mathscr{D}}$ to denote the orthogonal projection with respect to the Euclidean norm of the i-th component of the vector $\boldsymbol{x}$ onto the i-th component of the closed set $\mathscr{D}$, namely $[x_i]^{\mathscr{D}} = x_i$ if $x_i \in [x_{i,\min}, x_{i,\max}]$, $[x_i]^{\mathscr{D}} = x_{i,\min}$ if $x_i < x_{i,\min}$, or $[x_i]^{\mathscr{D}} = x_{i,\max}$ if $x_i > x_{i,\max}$.

∇ denotes the gradient operator. Given a scalar function $f(\boldsymbol{x}) : \mathbb{R}^n \to \mathbb{R}$,

$$\nabla f(\boldsymbol{x}) = \left[\frac{df_1(\boldsymbol{x})}{dx_1}, \ldots, \frac{df_n(\boldsymbol{x})}{dx_n}\right]^T.$$

Given a vector function $\mathbf{F}(\boldsymbol{x}) : \mathbb{R}^n \to \mathbb{R}^n$, we use the gradient matrix definition

$$\nabla \mathbf{F}(\boldsymbol{x}) = \begin{bmatrix} \nabla F_1(\boldsymbol{x}) & \ldots & \nabla F_n(\boldsymbol{x}) \end{bmatrix},$$

which is the transpose of the Jacobian matrix.

When we study vector optimization problems, we consider always Pareto optimal solutions. Therefore, we use the notation "optimal solution" to mean "Pareto optimal solution".

2. MOTIVATING EXAMPLES

In this section we describe some motivating examples where there is a need of fast optimization for wireless sensor networks. We argue that these problems cannot be solved efficiently with traditional approaches.

2.1 Distributed Detection

A classical problem in wireless sensor networks is the binary hypothesis testing or detection [8, 13]. In distributed detection theory, it is assumed that every node i of a network of n nodes wants to decide if an event out of two happened. The event detection is usually modelled by a Gaussian random variable that is associated to the first event H_0 or to the second event H_1:

$$\Gamma_i(s) = w_i(s) \quad \text{if} \quad H_0 \tag{2.1}$$

$$\Gamma_i(s) = E + w_i(s) \quad \text{if} \quad H_1 \tag{2.2}$$

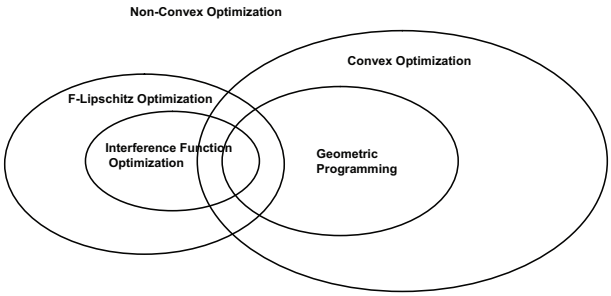

Figure 2: F-Lipschitz optimization theory includes the interference function optimization of type-I [9] for smooth functions, and, for example, part of convex optimization and geometric programming [12]. It follows that optimization problems previously solved by, e.g., convex solvers or geometric programming solvers, that fall in the area of F-Lipschitz optimization can now be solved much more efficiently by the methods proposed in this paper.

where s denotes that $\Gamma_i(s)$ is the outcome of a random variable at sample s, and E is a signal level that appears when the event H_1 occurs. The term $w_i(s)$ is a Gaussian random variable of zero mean and variance σ^2 that models the detection uncertainty. To overcome such an uncertainty, node i takes S samples and builds the so-called Likelihood ratio test as

$$T_i = \frac{1}{S}\sum_{s=1}^{S}\Gamma_i(s) \gtreqless x_i\,.$$

If $T_i \leq x_i$, where x_i is a detection threshold, then the node may decide for hypothesis H_0. If $T_i > x_i$, then the node may decide for hypothesis H_1. The ratio test gives a probability of false alarm, namely the probability that H_0 happened but H_1 was instead detected. The ratio test gives also a probability of misdetection, namely the probability that H_1 happened but H_0 was instead detected. It is easy to show that these probabilities have the following expressions [13]:

$$P_{\rm fa}^{(i)}(x_i) = \Pr[T_i > x_i|H_0] = Q\left(\frac{x_i}{\sqrt{\frac{\sigma^2}{S}}}\right)\,,$$

$$P_{\rm md}^{(i)}(x_i) = \Pr[T_i \leq x_i|H_1] = Q\left(\frac{E-x_i}{\sqrt{\frac{\sigma^2}{S}}}\right)\,,$$

where

$$Q(x) = \frac{1}{\sqrt{2\pi}}\int_x^\infty e^{-\frac{t^2}{2}}\,dt$$

is the complementary standard Gaussian distribution. Usually, the threshold x_i must be optimized so that to minimize the false alarm probability and while keeping under control the probability of misdetection. However, a node i may improve the detection performance by taking advantage of the probabilities communicated by neighboring nodes. The node makes an optimal decision on which event occurred by taking into account the opinions (the probabilities) transmitted by neighbors. This is done by combination with weighting factors, so that the final probability of misdetection at node i is expressed as

$$\sum_{j=1}^{n} b_{i,j}P_{\rm md}^{(j)}(x_j)\,.$$

Here, $b_{i,j} \geq 0, \forall i,j, \sum_{j=1}^{n} b_{i,j} = 1$, where $b_{i,i} \neq 0$ and $b_{i,j} = 0$ if node i does not communicate its probability with j. The total probability of false-alarm, which must be minimized in the network, is expressed as

$$\sum_{i=1}^{n} P_{\rm fa}^{(i)}(x_i)\,.$$

A global optimization problem can be posed where the objective is the minimization of the probabilities of false alarm while the probabilities of misdetection are kept under control [14]:

$$\begin{aligned}\min_{\boldsymbol{x}} \quad & \sum_{i=1}^{n} P_{\rm fa}^{(i)}(x_i) \qquad (2.3)\\ \text{s.t.} \quad & \sum_{j=1}^{n} b_{i,j}P_{\rm md}^{(j)}(x_j) \leq c_i, \quad i=1,\ldots,n\,,\\ & 0 \preceq \boldsymbol{x} \preceq E\mathbf{1}\,.\end{aligned}$$

The second set of constraints is because an unconstrained minimization of the probabilities of false alarm would increase dramatically the probability of misdetection. The third set of constraints is introduced for physical reasons [14]. Due to the property of the $Q(\cdot)$ function, it is easy to show that the problem is convex. In a distributed set-up, this would allow the computation of the optimal solution by the usual decomposition methods and the Lagrangian message passing. However, this is prohibitive for wireless sensor networks. We show in this paper that this problem can be transformed into an F-Lipschitz one and can be solved much more efficiently.

2.2 Radio Power Allocation with Intermodulation Powers

In wireless systems, the problem of allocating the transmit radio powers is cast as an optimization problem. The radio power used to transmit signals must be minimized to reduce the nodes energy consumption and the interference caused to other wireless transmissions. At the same time, the radio powers should be high enough to allow the receivers detecting successfully the transmitted signals. For illustrative purposes, we consider the basic problem of radio power transmission that is of interest for sensor networks. However, more advanced radio problems can be investigated, as those listed in [9]. We consider a high data rate wireless sensor network of n transmitter nodes, where node i transmits at a radio power p_i, $i=1,\ldots,n$. Since the data rate is high, it is meaningful to make a radio power control to save energy and prolong the lifetime of the sensor nodes. Let $\mathbf{p} \in \mathbb{R}^n$ the vector that contains the radio powers. In the network, there are n receiver nodes, where node i receives the power $G_{ii}p_i$ of the signal from transmitter i. G_{ii} is the channel gain from transmitter node i to receiver node i. Receiver node i is subject also to an interference from the signals from other transmitters, which is $\sum_{k\neq i} G_{ik}p_k$ and from the thermal noise σ_i. The signal to interference plus noise ratio of the i-th transmitter-receiver pair is defined as

$$\text{SINR}_i = \frac{G_{ii}p_i}{\sigma_i + \sum_{k\neq i} G_{ik}p_k + \sum_{k\neq i} M_{ik}p_i^2p_k^2}\,,$$

where G_{ik} models the wireless channel gain between the transmitter i and the receiver j, and M_{ik} is a intermodulation terms introduced by the amplifier of the receiver. Typically, these intermodulation terms are present when the amplifiers are built out of cheap and unreliable components, which may be typical for wire-

less sensor nodes. M_{ik}, $k \neq i$, assumes values smaller then G_{ik}, and can take on both positive and negative values.

The power optimization problem is usually written in the following form

$$\begin{aligned} \min_{\mathbf{p}} \quad & \mathbf{p} \\ \text{s.t.} \quad & \text{SINR}_i \geq S_{\min}, \quad i = 1, \ldots, n\,, \\ & p_{\min}\mathbf{1} \preceq \mathbf{p} \preceq p_{\max}\mathbf{1}\,, \end{aligned} \tag{2.4}$$

where $S_{\min}$ is the minimum required SINR and ensures that the signal of transmitter i can be received with the desired quality. The box constraints on the radio powers is naturally due to that transmitters have a minimum and maximum power level that can be used. Notice that the solution to this problem must be achieved in a distributed fashion, namely every node must be able to compute its own radio power because there is no time for node i to send the wireless channel coefficients G_{ij}, which is measured at node i, to a central solver and wait back for the optimal power to use: in the meantime G_{ij} would have changed due to the dynamic of the wireless channel and the optimal solution would be outdated [15]. Problem (2.4) is an unsolved optimization problem in communication theory. The interference function theory [9, 10], which is based on the monotonic and scalable property of the interference, is the fundamental reference to solve these problems. However, it cannot be used here because the intermodulation coefficients can be negative and this makes the interference term not monotonic and scalable. By the same argument, the geometric programming theory, which has also been widely employed to solve power allocation problems [16], cannot be used. Moreover, the problem is not convex. The classical approach would be to use parallelization and decomposition methods [6], provided that strong duality holds. Then, one would use iterative computation of the primal decision variables and dual variables until convergence is achieved. This makes it hard computing the solution of the problem (2.4), particularly when such a solution has to be computed by nodes of reduced computational capability. An alternative theory is needed. This theory is developed in the following.

3. F-LIPSCHITZ OPTIMIZATION

In this section we give the definition of an F-Lipschitz optimization problem, we characterize the existence and uniqueness of the optimal solution, we give algorithms to compute such a solution both in a closed form, when possible, and numerically. We characterize several features of the new optimization, including sensitivity analysis and robustness to quantization.

Definition 3.1 (F-Lipschitz optimization). *An F-Lipschitz optimization problem is defined as*

$$\begin{aligned} \max_{\boldsymbol{x}} \quad & f_0(\boldsymbol{x}) && \text{(3.1a)} \\ \text{s.t.} \quad & x_i \leq f_i(\boldsymbol{x})\,, \quad i = 1, \ldots, l && \text{(3.1b)} \\ & x_i = h_i(\boldsymbol{x})\,, \quad i = l+1, \ldots, n && \text{(3.1c)} \\ & \boldsymbol{x} \in \mathscr{D}\,, && \text{(3.1d)} \end{aligned}$$

where $\mathscr{D} \subset \mathbb{R}^n$ is a non empty, convex, and compact set, $l \leq n$, with objective and constraints being continuous differentiable functions such that

$$\begin{aligned} & f_0(\boldsymbol{x}) : \mathscr{D} \to \mathbb{R}^m\,, \qquad m \geq 1 \\ & f_i(\boldsymbol{x}) : \mathscr{D} \to \mathbb{R}\,, \qquad i = 1, \ldots, l \\ & h_i(\boldsymbol{x}) : \mathscr{D} \to \mathbb{R}\,, \qquad i = l+1, \ldots, n \end{aligned}$$

Let $\mathbf{f}(\boldsymbol{x}) = [f_1(\boldsymbol{x}), f_2(\boldsymbol{x}), \ldots, f_l(\boldsymbol{x})]^T$, $\mathbf{h}(\boldsymbol{x}) = [h_{l+1}(\boldsymbol{x}), h_{l+2}(\boldsymbol{x}), \ldots, h_n(\boldsymbol{x})]^T$, and $\mathbf{F}(\boldsymbol{x}) = [F_i(\boldsymbol{x})] = [\mathbf{f}(\boldsymbol{x})^T\ \mathbf{h}(\boldsymbol{x})^T]^T$. The following properties must be verified:

$$\begin{aligned} & 1.a \quad \nabla f_0(\boldsymbol{x}) \succ 0\,, \textit{ i.e., } f_0(\boldsymbol{x}) \textit{ is strictly increasing,} && \text{(3.2a)} \\ & 1.b \quad |\nabla \mathbf{F}(\boldsymbol{x})|_\infty < 1\,, && \text{(3.2b)} \\ & \textit{and either} \\ & 2.a \quad \nabla_j F_i(\boldsymbol{x}) \geq 0 \quad \forall i, j\,, && \text{(3.2c)} \\ & \textit{or} \\ & 3.a \quad \nabla_i f_0(\boldsymbol{x}) = \nabla_j f_0(\boldsymbol{x})\,, && \text{(3.2d)} \\ & 3.b \quad \nabla_j F_i(\boldsymbol{x}) \leq 0 \quad \forall i, j\,, && \text{(3.2e)} \\ & 3.c \quad |\nabla \mathbf{F}(\boldsymbol{x})|_1 < 1\,, && \text{(3.2f)} \\ & \textit{or} \\ & 4.a \quad f_0(\boldsymbol{x}) \in \mathbb{R}\,, && \text{(3.2g)} \\ & 4.b \quad |\nabla \mathbf{F}(\boldsymbol{x})|_1 \leq \frac{\delta}{\delta + \Delta}\,, && \text{(3.2h)} \\ & \qquad \delta = \min_{i, \boldsymbol{x} \in \mathscr{D}} \nabla_i f_0(\boldsymbol{x})\,, && \text{(3.2i)} \\ & \qquad \Delta = \max_{i, \boldsymbol{x} \in \mathscr{D}} \nabla_i f_0(\boldsymbol{x})\,. && \text{(3.2j)} \end{aligned}$$

Note the Lipschitz contractivity conditions in $1.b$, $3.c$ and $4.b$ imply that $\forall \boldsymbol{x}, \mathbf{y} \in \mathscr{D}$,

$$|F_i(\boldsymbol{x}) - F_i(\mathbf{y})| \leq \alpha_i ||\boldsymbol{x} - \mathbf{y}||, \quad i = 1, \ldots, n,$$

with $\alpha_i = \max_{\boldsymbol{x}} \|\nabla F_i(\boldsymbol{x})\| < 1 \quad \forall i = 1, \ldots n$.

We call properties (3.2a) – (3.2h) the *qualifying properties* of an F-Lipschitz optimization problem.

Note that it is possible that $l = n$, in which case there are no equality constraints. It is also possible that $l = 0$, in which case there are no inequality constraints. The objective function and the constraints are allowed to be linear or non linear functions, as for instance concave, convex, monomial, posynomial, etc. For example, the functions $f_i(\boldsymbol{x})$ and $h_i(\boldsymbol{x})$ can be convex. This makes the constraints $\boldsymbol{x} - \mathbf{F}(\boldsymbol{x}) \preceq 0$ non-convex in general, and therefore difficult to solve. The objective function (3.1a) is allowed to be both a decomposable or a non-decomposable function of the decision variables. Note that the objective function is allowed to be a vector in $\mathbb{R}^m$. When $m = 1$ we have a scalar optimization. In general, the problem is a multi-objective optimization one with m criteria. Examples of F-Lipschitz objective functions are

$$\begin{aligned} & f_0(\boldsymbol{x}) = \boldsymbol{x} \in \mathbb{R}^n \\ & f_0(\boldsymbol{x}) = \mathbf{c}^T \boldsymbol{x}, \quad \mathbf{c} \in \mathbb{R}^n, \quad \mathbf{c} \succ 0\,. \end{aligned}$$

Given that an F-Lipschitz problem can be a multi-objective problem, we recall the concept of Pareto optimal solutions.

Consider the following set

$$\begin{aligned} \mathscr{A} = \{ & \boldsymbol{x} \in \mathscr{D} : x_i \leq f_i(\boldsymbol{x}), i = 1, \ldots, l, \quad x_i = h_i(\boldsymbol{x}), \\ & i = l+1, \ldots, n \}\,, \end{aligned}$$

and let $\mathscr{B} \in \mathbb{R}^m$ be the image set of $f_0(\boldsymbol{x})$,i.e., $f_0(\boldsymbol{x}) : \mathscr{A} \to \mathscr{B}$. Then, we consider that the set $\mathscr{B}$ is partially ordered in a natural way, namely if $\boldsymbol{x}, \mathbf{y} \in \mathscr{B}$ then $\boldsymbol{x} \succeq \mathbf{y}$ if $x_i \geq y_i\ \forall i$ (e.g., $\mathbb{R}^m_+$ is the ordering cone).

Definition 3.2 (Pareto Optimal). *A vector $\boldsymbol{x}^*$ is called a Pareto optimal (or an Edgeworth-Pareto optimal) point if there is no $\boldsymbol{x} \in \mathscr{A}$ such that $f_0(\boldsymbol{x}) \succeq f_0(\boldsymbol{x}^*)$, i.e., if $f_0(\boldsymbol{x}^*)$ is a maximal element of the set $\mathscr{B}$ with respect to the natural partial ordering defined by the cone $\mathbb{R}^m_+$ [17].*

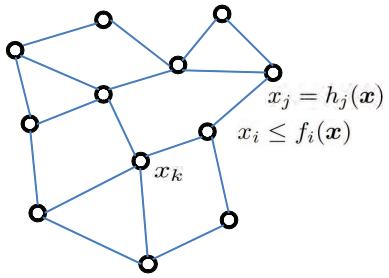

Figure 3: An optimization problem in the form (3.1) associated to a wireless sensor network. Every node has associated a decision variable and a constraint, or a group of decision variables and associated constraints. The line between nodes i and j denotes that the two nodes are able to communicate directly, without the need of a routing protocol.

In practice, a Pareto optimal solution is a vector for which is impossible to improve one component without decreasing another component. The Pareto optimal solutions are derived by converting a vector optimization problem into a scalar one via scalarization of the objective function [17].

Problem (3.1) can be used in centralized setting and in distributed settings. In this last case, a network of nodes needs to compute the optimal solution. The decision variable x_i is associated to a node i of the network for which we need to solve problem (3.1). Analogously, constraint i is associated to node i. In this paper, we are interested in the case in which every node needs primarily its own decision variable x_i, as it is common in wireless sensor network applications. An illustrative example of such a network is given in Fig. 3.

In the following subsection, we establish the existence of solutions to F-Lipschitz problems.

3.1 Existence and Uniqueness of Solutions

Here we give one of the core contributions of this paper. We show that for F-Lipschitz problems there is a unique optimal solution that is given by the (projected) system of constraints at equality. We have the following result:

Theorem 3.3. *Let the F-Lipschitz optimization problem in* (3.1) *be feasible. Then, the problem admits a unique Pareto optimum* $\boldsymbol{x}^* \in \mathscr{D}$ *given by the solutions of the following set of equations:*

$$\begin{aligned} x_i^* &= [f_i(\boldsymbol{x}^*)]^{\mathscr{D}} \quad i = 1, \dots, l \\ x_i^* &= h_i(\boldsymbol{x}^*) \quad i = l+1, \dots, n\,. \end{aligned}$$

PROOF. *A proof is presented in Appendix A.1.* □

Remark 3.4. *It follows from the proof that assumption* (3.2f) *could be replaced by the assumption that* $\nabla \mathbf{F}(\boldsymbol{x})^2 \succeq 0$.

The next example shows that the condition (3.2f) in general cannot be relaxed.

Example 3.5. *The following problem is not F-Lipschitz:*

$$\begin{aligned} \max \quad & \mathbf{1}^T \boldsymbol{x} \\ \text{s.t.} \quad & x_1 \leq 1 \\ & x_2 \leq -0.9 x_1 \\ & x_3 \leq -0.9 x_2 \\ & \boldsymbol{x} \in \mathscr{D} = \{\boldsymbol{x} : -10 \leq x_k \leq 10;\ k = 1, 2, 3\}\,. \end{aligned}$$

It has contractive constraints but does not satisfy (3.2f). *If the inequality constraints are active we get* $\boldsymbol{x}^* = (1, -0.9, -0.9)$ *with objective value* $\mathbf{1}^T \boldsymbol{x}^* = -0.8$. *However, this is not an optimal solution since the feasible point* $\boldsymbol{x} = 0$ *has a larger objective value* $\mathbf{1}^T \boldsymbol{x} = 0$. *Hence the conclusion of Theorem 3.3 does not hold.*

Example 3.6. *The following problem is not F-Lipschitz:*

$$\begin{aligned} \max \quad & \mathbf{1}^T \boldsymbol{x} \\ \text{s.t.} \quad & x_1 \leq \epsilon x_3 + 1 \\ & x_2 \leq -a x_1 \\ & x_3 \leq -a x_1 - a x_2 \\ & \boldsymbol{x} \in \mathscr{D} = \{\boldsymbol{x} : -10 \leq x_k \leq 10;\ k = 1, 2, 3\} \end{aligned}$$

where $0 < \epsilon < 1$ *and* $0 < a < 0.5$ *satisfies* (3.2f) *but not our sign conventions on the coefficients in* (3.2e). *If the inequality constraints are active we get* $\boldsymbol{x}^* = (1 - \epsilon a/\Delta, -a/\Delta, -a/\Delta)$, *where* $\Delta = 1/(1 + \epsilon a(1-a))$. *This gives an objective value* $\mathbf{1}^T \boldsymbol{x}^* = (1 - \epsilon a^2 - 2a)/\Delta$ *which is negative if* $a \approx 0.5$ *and* $\epsilon \approx 1$. *This is not the maximal solution since the feasible point* $\boldsymbol{x} = 0$ *has a larger objective value* $\mathbf{1}^T \boldsymbol{x} = 0$. *Hence the conclusion of Theorem 3.3 does not hold.*

In the next subsection we establish an alternative form that will be used in some of our examples.

3.2 Problems in Canonical Form

We have defined F-Lipschitz optimization problems by the special form in problem (3.1). It is common in the optimization literature to have problems stated in the following form, which is often referred to as canonical form [8]:

$$\begin{aligned} \min_{\boldsymbol{x}} \quad & g_0(\boldsymbol{x}) && && \text{(3.3a)} \\ \text{s.t.} \quad & g_i(\boldsymbol{x}) \leq 0\,, && i = 1, \dots, l && \text{(3.3b)} \\ & p_i(\boldsymbol{x}) = 0\,, && i = l+1, \dots, n && \text{(3.3c)} \\ & \boldsymbol{x} \in \mathscr{D}\,, \end{aligned}$$

where

$$\begin{aligned} g_0(\boldsymbol{x}) &: \mathscr{D} \to \mathbb{R}^m\,, && m \leq n \\ g_i(\boldsymbol{x}) &: \mathscr{D} \to \mathbb{R}\,, && i = 1, \dots, l \\ p_i(\boldsymbol{x}) &: \mathscr{D} \to \mathbb{R}\,, && i = l+1, \dots, n \end{aligned}$$

Problem (3.3) can be converted into a F-Lipschitz-like problem (3.1) by the following transformations

$$\begin{aligned} \max_{\boldsymbol{x}} \quad & f_0(\boldsymbol{x}) && && \text{(3.4a)} \\ \text{s.t.} \quad & x_i \leq f_i(\boldsymbol{x})\,, && i = 1, \dots, l\,, && \text{(3.4b)} \\ & x_i = h_i(\boldsymbol{x}) && i = l+1, \dots, n\,, && \text{(3.4c)} \\ & \boldsymbol{x} \in \mathscr{D}\,, \end{aligned}$$

where

$$\begin{aligned} f_0(\boldsymbol{x}) &= -g_0(\boldsymbol{x})\,, && && \text{(3.5)} \\ f_i(\boldsymbol{x}) &= x_i - \gamma_i g_i(\boldsymbol{x})\,, && i = 1, \dots, l\,, && \text{(3.6)} \\ h_i(\boldsymbol{x}) &= x_i - \mu_i p_i(\boldsymbol{x})\,, && i = l+1, \dots, n\,, && \text{(3.7)} \end{aligned}$$

with $\gamma_i > 0$, $i = 1, \dots, l$ and $\mu_i \in \mathbb{R}$, $i = l+1, \dots, n$. We let

$$\mathbf{G}(\boldsymbol{x}) = [g_1(\boldsymbol{x}), \dots, g_l(\boldsymbol{x}), p_{l+1}(\boldsymbol{x}), \dots, p_n(\boldsymbol{x})]^T\,.$$

Problem (3.3) and problem (3.4) have the same optimal solution because the constraints of problem (3.4) hold if and only if the constraints of problem (3.3) hold, since $\gamma_i > 0$ and $\mu_i \neq 0\ \forall i$.

Clearly, problem (3.4) is in general not an F-Lipschitz one. It is interesting to establish when problem (3.4) is F-Lipschitz, namely when qualifying properties (3.2a) – (3.2h) are satisfied for (3.4). We have the following result:

Theorem 3.7. *Consider the optimization problems* (3.3) *and* (3.4). *Suppose that* $\forall\, \boldsymbol{x} \in \mathscr{D}$

$$1.a \quad \nabla g_0(\boldsymbol{x}) \prec 0\,, \tag{3.8a}$$
$$1.b \quad \nabla_i G_i(\boldsymbol{x}) > 0 \quad \forall i\,, \tag{3.8b}$$
$$1.c \quad \nabla_i G_i(\boldsymbol{x}) > \sum_{j\neq i} |\nabla_j G_i(\boldsymbol{x})| \quad \forall i\,,\,, \tag{3.8c}$$

and either

$$2.a \quad \nabla_j G_i(\boldsymbol{x}) \leq 0 \quad \forall j \neq i\,, \tag{3.8d}$$

or

$$3.a \quad \nabla_i g_0(\boldsymbol{x}) = \nabla_j g_0(\boldsymbol{x})\,, \tag{3.8e}$$
$$3.b \quad \nabla_j G_i(\boldsymbol{x}) \geq 0 \quad \forall j \neq i\,,$$
$$3.c \quad \nabla_i G_i(\boldsymbol{x}) > \sum_{j\neq i} |\nabla_i G_j(\boldsymbol{x})| \quad \forall i\,, \tag{3.8f}$$

or

$$4.a \quad g_0(\boldsymbol{x}) \in \mathbb{R}\,, \tag{3.8g}$$
$$4.b \quad \frac{\delta}{\delta+\Delta}\nabla_i G_i(\boldsymbol{x}) > \sum_{j\neq i} |\nabla_i G_j(\boldsymbol{x})| \quad \forall i\,, \tag{3.8h}$$

where δ *and* Δ *are defined in Eqs.* (3.2i) *and* (3.2j). *Then, problem* (3.4) *is F-Lipschitz.*

PROOF. A proof is provided in Appendix A.2. □

Note that (3.8c) is a condition on the rows of the gradient of $\mathbf{G}(x)$, whereas (3.8f) and (3.8h) are per-column conditions on the gradient.

This theorem can be used to show that some times convex optimization problems can be cast to F-Lipschitz and thus solved efficiently. By the same theorem, it is also possible to show that, for example, a class of geometric and signomial programming problems [12] is solved by an F-Lipschitz approach.

3.3 Computation of the Optimal Solution

In subsection 3.1 we have proved that there is a unique optimal solution to F-Lipschitz optimization problems, which is achieved by solving the system of equations given by the projected constraints at equality. In this section, we show that the complexity to solve F-Lipschitz optimization problems is quite low compared to Lagrangian methods, which are the ones traditionally employed.

If the set of equations given by the projected constraints at equality can be solved in a closed form, then we have the optimal solution in a closed form, otherwise we need numerical algorithms. These algorithms to solve systems of equations are well known, see [1, 18] as fundamental references. In the following, we summarize a low-computational complexity technique for centralized optimization problems, namely for problems that can be solved by a central node. We then present a simple algorithm for distributed optimization problems, where there is the lack of a central computational unit and every node cooperate to compute the solution in a distributed fashion. For simplicity we focus on first order techniques, though the convergence speed of numerical algorithms can be increased by heavy ball methods [18].

3.3.1 Centralized Computation

Since the functions $\mathbf{f}(x)$ and $\mathbf{h}(\boldsymbol{x})$ of problem (3.1) are differentiable within $\mathscr{D}$, then the Newton's method can be applied, which consists of the iterations

$$\boldsymbol{x}(k+1) = \left[\boldsymbol{x}(k) - \beta\left(I - \nabla\mathbf{F}(\boldsymbol{x}(k))\right)^{-1}\left(\boldsymbol{x}(k) - \mathbf{F}(\boldsymbol{x}(k))\right)\right]^{\mathscr{D}}\,, \tag{3.9}$$

where β is a positive scalar that can be chosen so that the modulus of the previous mapping is contractive and as small as possible so to have the fastest convergence. The iterations can be initialized by any $x(0) \in \mathscr{D}$, and convergence is certified if the gradient matrix $I - \nabla\mathbf{F}(\boldsymbol{x})$ is invertible:

Lemma 3.8. *Consider the function* $\mathbf{F}(\boldsymbol{x}) = [\mathbf{f}(\boldsymbol{x})^T\mathbf{h}(\boldsymbol{x})^T]^T$ *of problem* (3.1) *and let qualifying property* (3.2b) *hold. Then the matrix* $I - \nabla\mathbf{F}(\boldsymbol{x})$ *has full rank.*

PROOF. The simple proof is based on that the matrix $\nabla\mathbf{F}(\boldsymbol{x})$ is contractive. Its spectral radius is $\rho(\nabla\mathbf{F}(\boldsymbol{x})) \leq \|\nabla\mathbf{F}(\boldsymbol{x})\| < 1$ [19], thus the eigenvalues of $I - \nabla\mathbf{F}(\boldsymbol{x})$ are strictly positive and the matrix is invertible. □

It is well known that the convergence speed of the Newton's algorithm is quite fast, namely it is superlinear. As Bertsekas writes, it is the most complex and also the fastest among the gradient methods [1]. It has however the drawback of requiring the computation of the inverse of a matrix at each step. To avoid this computational burden, several other algorithms have been developed, including gradient methods, conjugate direction methods, quasi-Newton methods, and non-derivative methods [1]. These methods are characterized by different convergence speeds and computational complexities. One can chose the most suitable according to the nature of the F-Lipschitz constraints. For example, if these constraints are quadratic, then gradient methods are good candidates. See [1] for rules and recommendations on what method for different kind of functions.

3.3.2 Distributed Computation

There are many algorithms to compute the solution to a system of equations in a distributed set-up. See [6] as a fundamental reference. Here, we present a simple iterative and asynchronous distributed algorithm. Recall that in the distributed set-up we assume that there is a global optimization problem that needs to be solved by the wireless sensor network, where each decision variable is associated to a node and every node cooperatively communicates with other nodes to compute the optimal solution without central coordination. Such a communication could be subject to delays and losses due to the underlying communication channel. We have the following result:

Proposition 3.9. *Let* $\boldsymbol{x}(0) \in$ *be an initial guess of the optimal solution to a feasible F-Lipschitz problem* (3.1). *Let* $\boldsymbol{x}^i(k) = [x_1(\tau_1^i(k)), x_2(\tau_2^i(k)), \ldots, x_n^i(\tau_n(k))]$ *the vector of decision variables available at node* i *at time* $k \in \mathbb{N}_+$, *where* $\tau_j^i(k)$ *is the delay with which the decision variable of node* j *reaches node* i. *Then, the following iterative algorithm converges to the optimal solution:*

$$\begin{aligned} x_i(k+1) &= [f_i(\boldsymbol{x}^i(k))]^{\mathscr{D}} \quad i = 1, \ldots, l \\ x_i(k+1) &= [h_i(\boldsymbol{x}^i(k))]^{\mathscr{D}} \quad i = l+1, \ldots, n \end{aligned} \tag{3.10}$$

where $k \in \mathbb{N}_+$ *is an integer associated to the iterations.*

PROOF. From Theorem 3.3 the unique optimal solution to a feasible F-Lipschitz problem (3.1) is given by the constraints at equality. Since the right hand side of the constraints is contractive for qualifying properties (3.2), the convergence of algorithm (3.10) to the optimal solution is guaranteed as $k \to \infty$ by applying the asynchronous converge theorem in [6], which concludes the proof. □

According to previous proposition, every node i of the network updates its decision variable by the iterative algorithm (3.10) and by using the decision variables of other nodes that are available at that time. Notice that when $f_i(\boldsymbol{x})$ depends only on the decision variables of the neighboring nodes, the communications of these variables is fast and practical. This is the case of the example we discuss in Section 2.1. In other situations, $f_i(\boldsymbol{x})$ can be given by an oracle locally at node i without any direct communication of decision variables from the other nodes, as in the radio power optimization problems we illustrate in the example in Section 2.2 (see also, e.g., [11]).

Proposition 3.10. *Consider the distributed algorithm for the synchronous computation of the optimal solution, i.e., $\tau_j^i(k)$ is constant $\forall i, j$. Let*

$$d = \max_{\boldsymbol{x},\mathbf{y}\in\mathscr{D}} \|\boldsymbol{x} - \mathbf{y}\|_\infty \,.$$

Let ε be the desired precision with which the optimal solution $\boldsymbol{x}^*$ must be known. Suppose $\varepsilon < d$. Then, an upper bound to the number of iterations needed for the algorithm to converge is $O(\bar{k})$, where

$$\bar{k} = \frac{|\ln \varepsilon| - |\ln d|}{|\ln \alpha|} \,,$$

where α is the Lipshitz constant of $\mathbf{F}(\boldsymbol{x})$ for $\boldsymbol{x} \in \mathscr{D}$.

PROOF. From qualifying properties (3.2), the iterations of Algorithm (3.10) are contractive. It follows that $\|\boldsymbol{x}(k) - \boldsymbol{x}^*\| \leq \alpha^k \|\boldsymbol{x}(0) - \boldsymbol{x}^*\| \leq \alpha^k d$, whereby $\bar{k}$ follows immediately by recalling that $\alpha < 1$ and $\varepsilon < d$. □

The previous proposition is useful in that it allows us to upper bound the number of iterations to compute the optimal solution with a desired accuracy. Notice that the upper bound does not depend on the number of variables, but on the Lipscitz constant of the contraction mapping. This result is remarkable, because for F-Lipschitz optimization the convergence speed of Algorithm (3.10) may be arbitrarily high regardless of the number of decision variables or nodes of the network. However, note that this is possible under the assumption that every nodes works in parallel at the same pace and that the delay with which variables are communicated is fixed.

We now compare the convergence speed of F-Lipschitz algorithm to the one achieved by using the traditional decomposition methods based on Lagrange multipliers. To do that, we need to show that strong duality applies, as we see in the following subsection. Strong duality will be also very useful to study the sensitivity and stability of the optimal solution to perturbations of the constraints.

3.4 Computational complexity

The results of this section are useful to compare the convergence speed of the F-Lipschitz algorithms presented in subsection 3.3 to the traditional Lagrangian algorithms that one would use for nonlinear optimization problems. In this section, we show that it is possible to solve F-Lipschitz optimization problems by Lagrangian iterative methods because strong duality applies. Moreover, strong duality is useful to characterize the sensitivity to the optimal solution to perturbations of the constraints.

For analytical simplicity, we show that strong duality applies to the optimization problem (3.1) in the case when there are only inequality constraints. We assume for simplicity that the box constraints, $\boldsymbol{x} \in \mathscr{D}$, can be held implicit. Note, however, that the upper bound constraints $x_i \leq x_{i,\max}$ poses no problem since they satisfy the invexity inequality used in the proof.

Theorem 3.11. *Consider the optimization problem* (3.1) *in the case when there are no equality constraints and suppose the problem feasible. Then strong duality applies and the KKT conditions are necessary and sufficient to compute the optimal solution of problem* (3.1)*.*

PROOF. See Appendix A.3. □

Given that we have strong duality, we can solve problem (3.1) by the KKT conditions:

$$\begin{aligned}
&x_i - f_i(\boldsymbol{x}^*) \leq 0 \quad i = 1, \ldots, n \\
&\lambda_i^* \geq 0 \qquad\quad i = 1, \ldots, n \\
&\lambda_i^*(x_i^* - f_i(\boldsymbol{x}^*)) = 0 \quad i = 1, \ldots, n \\
&\nabla_{\boldsymbol{x}} L(\boldsymbol{x}^*, \boldsymbol{\lambda}^*) = 0 \,,
\end{aligned}$$

where $\boldsymbol{\lambda} = [\lambda_i] \in \mathbb{R}^n$ is the vector of Lagrange multipliers and $L(\boldsymbol{x}, \boldsymbol{\lambda})$ is the Lagrange function associated to the scalarized version [8] of problem (3.1):

$$L(\boldsymbol{x}, \boldsymbol{\lambda}) = -\boldsymbol{\mu}^T f_0(\boldsymbol{x}) + \sum_{i=1}^n \lambda_i (x_i - f_i(\boldsymbol{x}))$$

where $\boldsymbol{\mu} \succ 0$. Finding $\boldsymbol{\lambda}^*$ and $\boldsymbol{x}^*$ that solve these conditions is much more expensive than solving the system of equations given by the projected constraints at the equality as proposed in Subsection 3.3.

When the Lagrangian function gives closed form multipliers, one would have to solve two systems of n equations in n variables and recover the primal variables. Namely, 1) one would have to take the derivative of the lagrangian with respect to $\boldsymbol{x}$ and solve for $\boldsymbol{x}$ as function of $\boldsymbol{\lambda}$ the system of n equations $\nabla_{\boldsymbol{x}} L(\boldsymbol{x}, \boldsymbol{\lambda}) = 0$, thus achieving $\boldsymbol{x}^*(\boldsymbol{\lambda})$. Then, 2) one would have to plug in such a solution in $L(\boldsymbol{x}, \boldsymbol{\lambda})$, and solve a further system of n equations $\nabla_{\boldsymbol{\lambda}} L(\boldsymbol{x}^*(\boldsymbol{\lambda}), \boldsymbol{\lambda}) = 0$, thus achieving $\boldsymbol{\lambda}^*$. Finally, 3) one would have to recover the optimal solution by computing the function $\boldsymbol{x}^*(\boldsymbol{\lambda}^*)$.

When the Lagrangian function does not give closed form multipliers, one would have to resort to numerical iterative Lagrangian methods. There are several such methods to compute the optimal solution to problem (3.1), such as barrier and interior point methods, penalty and augmented lagrangian methods, primal-dual interior point methods, etc., that we can use to solve an F-Lipschitz optimization problem. These methods differ for the convergence speed and computational complexity. Typically, at each iteration of these methods some further optimization problem and/or matrix inversions are required. To give an idea of convergence speed comparison with the F-Lipschitz algorithms described in Subsection 3.3, we consider as reference a first order Lagrangian method for optimization problems with equality constraints. By using this method, the optimal solution to problem (3.1) is given by the iterations

$$\boldsymbol{x}(k+1) = \boldsymbol{x}(k) - \beta \nabla_{\boldsymbol{x}} L(\boldsymbol{x}(k), \boldsymbol{\lambda}(k)) \tag{3.11}$$

$$\boldsymbol{\lambda}(k+1) = \boldsymbol{\lambda}(k) - \beta \nabla_{\boldsymbol{\lambda}} L(\boldsymbol{x}(\boldsymbol{\lambda}(k)), \boldsymbol{\lambda}(k)) \,, \tag{3.12}$$

where β is a positive scalar. It is clear that iterations (3.11) and (3.12) are computationally much more expensive than the algorithms in Subsection 3.3 because these iterations involve the computation of the gradient of the Lagrangian function and the update of the decision variables plus the Lagrangian multipliers. In a distributed set-up, the situation is even worse, because every node needs to send to all other nodes the Lagrange multipliers either directly or by incremental techniques [2, 7].

3.5 Sensitivity Analysis

It is interesting to establish the sensitivity of the optimal solution to perturbations of the constraints. When an F-Lipschitz optimization problem is implemented on hardware platforms having limited computational capabilities, such as in wireless sensor networks, it may happen that the constraints are altered by errors, noises, and quantization. However, the optimal solution to F-Lipschitz problems is not sensitive to perturbations. We prove the result under the simplifying assumptions that 1) there are no equality constraints, 2) the box constraints are not active at the optimal solution. We have the following result:

Claim 3.12. *Consider the optimization problem* (3.1) *in the case when there are no equality constraints and suppose the problem is feasible. Then the unique global optimum is not sensitive to perturbations to the constraints.*

PROOF. See Appendix A.4 □

3.6 Robustness to Quantization Errors

When the computation of the optimal solution is performed by resource constrained nodes that introduce quantization errors, such as wireless sensor networks, it is important to study the robustness of the computation of the optimal solution to these errors.

The quantization error is modelled as a random variable $q_i(k)$, $i = 1, \ldots, n$, having some distribution within the interval $[-q_{\max}, q_{\max}]$, where $q_{\max}$ is the maximum quantization error. We do not assume any specific distribution, and we model the quantization process for the synchronous distributed algorithm as

$$\tilde{x}_i(k+1) = f_i(\tilde{\boldsymbol{x}}(k)) + q_i(k) \quad i = 1, \ldots, l$$
$$\tilde{x}_i(k+1) = h_i(\tilde{\boldsymbol{x}}(k)) + q_i(k) \quad i = l+1, \ldots, n$$

It follows that the decision variables at time k $\tilde{x}_i(k)$'s are affected by two quantization errors: the errors coming from the variables that concur in the computation of $\tilde{x}_i(k+1)$ in $f_i(\tilde{\boldsymbol{x}}(k))$, and the error affecting the computation of $f_i(\tilde{\boldsymbol{x}}(k))$, which is $q_i(k)$. These errors may hinder the convergence of the algorithm. We would like to study the stability of iterations when there are these quantization errors. We have the following result:

Proposition 3.13. *Let $\boldsymbol{x}^*$ be the optimal solution to* (3.1)*. Let $\alpha_{\max} = \max_i \alpha_i$. Let $q_{\max}$ the maximum quantization error. Then the optimal solution computed by the synchronous Algorithm* (3.10) *with quantization errors satisfies*

$$\lim_{k\to\infty} \|\tilde{\boldsymbol{x}}(k) - \boldsymbol{x}^*\|_\infty \leq \frac{1}{1-\alpha_{\max}} q_{\max}\,.$$

PROOF. See Appendix A.5. □

From this proposition we see that if we would like to minimize the effect of the quantization error, than we should have $\alpha_{\max}$ as small as possible. This gives also a fast convergence speed of the iterations. Therefore, the faster the convergence speed, the lower the quantization error that will affect the computation of the optimal solution. This is natural since at each iteration there is a quantization error that keeps accumulating. Reducing these iterations, reduces the entity of such an accumulation. We would like to mention that the analysis developed in this section applies also to traditional Lagrangian methods in (3.11) and (3.12).

4. EXAMPLES OF APPLICATIONS

In this section, we show that the wireless sensor network motivating examples of Section 2 are F-Lipschitz.

4.1 Distributed Detection

The distributed detection optimization problem (2.3) can be solved by F-Lipschitz methods. We observe that the cost function is strictly decreasing in the decision variables since the complementary Gaussian distribution $Q(\cdot)$ is increasing in its argument. Analogously, the constraints have strictly positive derivatives since they are given by the positive combination of Q functions with negative argument, thus condition (3.8b) of Theorem 3.7 applies. If (3.8c) and (3.8f) also hold, the the problem is F-Lipschitz. This can be easily achieved by imposing that the weighting coefficients are such that

$$b_{ii} e^{-\frac{E^2}{2\frac{\sigma^2}{S}}} \geq \sum_{j\neq i} b_{ij} \quad \forall i\,, \tag{4.1}$$

and

$$\frac{\delta}{\delta+\Delta} b_{ii} e^{-\frac{E^2}{2\frac{\sigma^2}{S}}} \geq \sum_{j\neq i} b_{ji} \quad \forall i\,, \tag{4.2}$$

where

$$\delta = \min_{i,\boldsymbol{x}\in\mathscr{D}} \nabla_i f_0(\boldsymbol{x}) = \min_{i,\boldsymbol{x}\in\mathscr{D}} \frac{1}{\sqrt{2\pi\frac{\sigma^2}{S}}} e^{-\frac{x_i^2}{2\frac{\sigma^2}{S}}} = \frac{1}{\sqrt{2\pi\frac{\sigma^2}{S}}}\,,$$

$$\Delta = \max_{i,\boldsymbol{x}\in\mathscr{D}} \nabla_i f_0(\boldsymbol{x}) = \max_{i,\boldsymbol{x}\in\mathscr{D}} \frac{1}{\sqrt{2\pi\frac{\sigma^2}{S}}} e^{-\frac{x_i^2}{2\frac{\sigma^2}{S}}} = \frac{1}{\sqrt{2\pi\frac{\sigma^2}{S}}}\,.$$

The assumptions (4.1) and (4.2) have a natural interpretation: the weighting coefficients that node i uses for its own opinion should be high enough so to trust more the own opinion than other node's opinions. By these assumptions, problem (2.3) has the optimal solution satisfying all the constraints at equality. The solution may be computed by algorithm (3.9) in a centralized set-up, or by algorithm (3.10) in a distributed set-up.

We can transform the problem in the form (3.1) by positive scalars γ_i, as shown in Section 3.2. In particular, we define

$$f_0(\boldsymbol{x}) = -\sum_{i=1}^n P_{\text{fa}}^{(i)}(x_i)\,,$$

and $f_i(\boldsymbol{x}) = x_i - \gamma_i g_i(\boldsymbol{x})$ with

$$g_i(\boldsymbol{x}) = \sum_{j=1}^n b_{i,j} P_{\text{md}}^{(j)}(x_j) - c_i \quad \forall i\,,$$

where

$$\begin{aligned}\gamma_i &\leq \min_{0\preceq\boldsymbol{x}\preceq E\mathbf{1}} \frac{1}{\nabla_i g_i(\boldsymbol{x})} \\ &= \min_{0\preceq\boldsymbol{x}\preceq E\mathbf{1}} \frac{1}{\frac{1}{\sqrt{2\pi\frac{\sigma^2}{S}}} \sum_{j=1}^n b_{i,j} e^{-\frac{(E-x_j)^2}{\sqrt{\frac{\sigma^2}{S}}}}} \\ &= \sqrt{2\pi\frac{\sigma^2}{S}} \quad \forall i\,,\end{aligned}$$

ensures contractivity of the mapping.

We performed numerical simulations by Matlab. We considered a network of $n = 5$ nodes, where each node detected an event as a Gaussian random variable of average 0 if the hypothesis H_0 happened and of average $E = 3$ if the hypothesis H_1 happened. We set $\sigma = 1$ and $S = 1$. We chose that every node needs to have at most a probability $c_i = 1/15$ $\forall i$ as a total probability of miss detection. Moreover, we generated randomly $b_{i,j}$ $\forall i, j$ with a

uniform distribution in $[0,1]$ and then we normalized these coefficients so that (4.1) and (4.2) hold. Since optimization problem is convex, we used the interior point method by the Matlab function `fmincon` with the option "interior-point", whereas the F-Lipschitz solution was provided by solving the system of equations given by the constraints at the equality by the Matlab function `fsolve`. We observed a convergence of `fsolve` in about 5 iterations and 36 function evaluations, whereas `fmincon` converges in about 31 iterations and 231 function evaluations. Similar results hold for other choices of the detection parameters. The reduction of number of iterations and computational complexity ensured by our method is remarkable.

4.2 Radio Power Allocation

The power allocation problem of Section 2.2 is F-Lipschitz. We can see this by making the variable substitution $x_i = -p_i$. Then, problem (2.4) can be rewritten as

$$\begin{aligned} \max_{\boldsymbol{x}} \quad & \boldsymbol{x} \qquad (4.3)\\ \text{s.t.} \quad & G_{ii}x_i + S_{\min}(\sigma_i - \sum_{k\neq i} G_{ik}x_k + \sum_{k\neq i} M_{ik}x_i^2x_k^2) \leq 0,\\ & \quad i = 1,\ldots,n\,,\\ & -p_{\max}\mathbf{1} \preceq \boldsymbol{x} \preceq -p_{\min}\mathbf{1}\,. \end{aligned}$$

Since M_{ik} is smaller than G_{ii} and G_{ik} $\forall i,k$, it is reasonable to assume that for $-\mathbf{1}p_{\max} \preceq \boldsymbol{x} \preceq -\mathbf{1}p_{\min}$ the following conditions hold: $G_{ii} > 2S_{\min}\sum_{k\neq i} M_{ik}x_ix_k^2$, $G_{ik} > 2M_{ik}x_i^2x_k$ and

$$G_{ii} + 2S_{\min}\sum_{k\neq i} M_{ik}p_{\min}^3 \geq S_{\min}\sum_{k\neq i} G_{ik}\,.$$

This assumption gives

$$\nabla_i g_i(\boldsymbol{x}) > \sum_{i\neq j} |\nabla_j g_i(\boldsymbol{x})|\,,$$

for $-\mathbf{1}p_{\max} \preceq \boldsymbol{x} \preceq -\mathbf{1}p_{\min}$. Therefore, Theorem 3.7 applies (see conditions Eq. (3.8d) and Eq. (3.8c)) and problem (4.3) has the optimal solution satisfying all the constraints at equality. The solution could be centrally computed by algorithm (3.9), or by algorithm (3.10) in a distributed set-up. In this last case, we have to transform the problem in the form (3.1) by positive scalars γ_i, as shown in Section 3.2, thus achieving the equivalent constraints $f_i(\boldsymbol{x}) = x_i - \gamma_i g_i(\boldsymbol{x})$, where

$$g_i(\boldsymbol{x}) = G_{ii}x_i + S_{\min}(\sigma_i - \sum_{k\neq i} G_{ik}x_k + \sum_{k\neq i} M_{ik}x_i^2x_k^2) \quad \forall i\,.$$

This function is contractive Lipschitz provided that one chooses

$$\begin{aligned} \gamma_i &\leq \min_{-\mathbf{1}p_{\max}\preceq\boldsymbol{x}\preceq-\mathbf{1}p_{\min}} \frac{1}{\nabla_i g_i(\boldsymbol{x})}\\ &= \min_{-\mathbf{1}p_{\max}\preceq\boldsymbol{x}\preceq-\mathbf{1}p_{\min}} \frac{1}{G_{ii} + 2S_{\min}\sum_{k\neq i} M_{ik}x_ix_k^2}\\ &= \frac{1}{G_{ii} + 2S_{\min}p_{\max}^3\sum_{k\neq i} M_{ik}} \end{aligned}$$

which ensures contractivity with respect to the infinity norm.

We performed numerical simulations by Matlab with 10 transmitter–receiver pairs of nodes. We considered a wireless sensor network where the figures for the wireless channel and noises are taken coherently with the Tmote Sky sensor nodes [20], which features the CC2420 radio transceiver module by Chipcon [21]. The noise is set to $\sigma_i = -130$ dBm $\forall i$, $S_{\min} = 1$, $\forall i$, $G_{ij} = -90$ dBm $\forall i \neq j$ and $M_{ij} = -120$ dBm $\forall i \neq j$. Moreover, $p_{\min} = -25$ dBm and $p_{\max} = 0$ dBm. See [22] for details. We observed a convergence of algorithm (3.10) in less than 10 iterations, whereas a traditional method based on the Lagrangian dual function converges in about 40 iterations. Once again, the reduction of number of iterations and computational complexity ensured by our method is remarkable.

5. CONCLUSIONS AND FUTURE WORK

In this paper we presented the F-Lipschitz optimization, which enables fast computations of the solution of a class of convex and non-convex problems. The central idea was to show that the optimal solution is achieved when all the constraints hold at equality. We showed that this optimization method can solve problems much more efficiently than traditional Lagrangian methods, including convex problems. If an F-Lipschitz optimization problem must be solved by distributed operations, then our optimization method uses simple and fast distributed asynchronous algorithms, which are very appealing for wireless sensor networks. We showed that some typical optimization problems that arise in wireless sensor networks are F-Lipschitz, such as distributed detection and radio power control.

We believe that F-Lipschitz optimization may have many developments. There can be more qualifying properties for which this optimization applies, such as when constraints have the Jacobian always positive definite (which is sufficiently ensured by the qualifying F-Lipschitz conditions). In many situations, it could be interesting to approximate problems to F-Lipschitz ones, given the numerous useful properties of this optimization. For example, if the Jacobian of the constraints does not have all the off-diagonal elements of the same sign, then one could set to zero these outliers and see how suboptimal is the solution by following a similar approach proposed in [23]. Another interesting line of research is about the extension of our method to more general objective functions, e.g., non necessarily increasing functions. The introduction of slack variables that make the constraints hold at equality might be useful, even though it appears challenging to tie these variables to the objective function and show that the technique is more convenient than Lagrangian methods. For problems with general objective functions and constraints, it could be computationally useful to check the existence of conditions for which the constraints hold at equality. If we know that strong duality holds, then strictly positive Lagrangian multipliers associated to inequality constraints are a sufficient (but not necessary) condition for the solution to satisfy the constraints at equality. We are currently investigating when these conditions hold. Finally, we are currently investigating whether knowing that some of the inequality constraints hold at equality at the optimum is of help to reduce the computations for the optimal solution. This might be done by not including these constraints in the Lagrangian.

6. ACKNOWLEDGEMENTS

C. Fischione is grateful to A. Speranzon of the United Technology Research Center, East Hartford, CT, USA, for the numerous and fruitful discussions on distributed estimation topics, which have led to the development of the F-Lipschitz optimization.

The authors acknowledge the support of the Swedish Research Council and of the EU projects FeedNetBack and Hycon2.

7. REFERENCES

[1] D. P. Bertsekas, *Non Linear Programming*, 2nd ed. Athena Scientific, 2004.

[2] M. Rabbat and R. Nowak, "Distributed optimization in sensor networks," in *ACM/IEEE IPSN*, 2004.
[3] H. Gharavi and P. R. Kumar, "Toward a theory of in-network computation in wireless sensor networks," *IEEE Communication Magazine*, pp. 97–107, 2006.
[4] P. Wan and M. D. Lemmon, "Event-triggered distributed optimization in sensor networks," in *ACM/IEEE IPSN*, 2009.
[5] H. Wang, N. Agoulmine, M. Ma, and J. Yanliang, "Network lifetime optimization in wireless sensor networks," *IEEE Journal on Selected Areas on Communication*, vol. 28, no. 7, pp. 1127–1137, September 2010.
[6] D. P. Bertsekas and J. N. Tsitsiklis, *Parallel and Distributed Computation: Numerical Methods*. Athena Scientific, 1997.
[7] B. Johansson, "On distributed optimization in networked systems," Ph.D. dissertation, KTH, 2009.
[8] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.
[9] R. D. Yates, "A framework for uplink power contorl in cellular radio system," *IEEE Journal on Selected Areas in Communications*, vol. 13, pp. 1341–1347, September 1995.
[10] J. Herdtner and E. K. P. Chong, "Analysis of a class of distributed asynchronous power control algorithm for cellular wireles system," *IEEE Journal on Selected Areas in Communications*, vol. 18, no. 3, pp. 436–446, March 2000.
[11] M. Schubert and H. Boche, *QoS-Based Resource Allocation and Transceiver Optimization, Foundations and Trends in Communications and Information Theory*. NOW, 2005.
[12] S. Boyd, S. J. Kim, L. Vandenberghe, A. Hassibi, "A tutorial on geometric programming," *Optimization and Engineering*, vol. 1, no. 1, p. 1, 1 2006.
[13] S. M. Kay, *Fundamentals of statistical signal processing: detection theory*. Prentice Hall, Inc., 1998.
[14] P. Paysarvi-Hoseini and N. C. Beaulieu, "Optimal wideband spectrum sensing framework for cognitive radio systems," *IEEE Transaction on Signal Processing*, vol. 42, 2011, to appear.
[15] G. L. Stüber, *Priciples of Mobile Communication*. Kluwer Academic Publishers, 1996.
[16] S. Kandukuri, S. Boyd, "Optimal power control in interference-limited fading wireless channels with outage-probability specifications," *IEEE Transaction on Wireless Communications, Vol. 1, pp 46–55*, January 2002.
[17] J. Jhan, *Vector Optimization*. Springer-Verlag, 2004.
[18] B. Polyak, *Introduction to Optimization*. Optimization Software, Inc., 1987.
[19] R. A. Horn and C. A. Johnson, *Matrix Analysis*. Cambridge University press, 1985.
[20] *Tmote Sky Data Sheet*, Moteiv, San Francisco, CA, 2006. [Online]. Available: http://www.moteiv.com/products/docs/tmote-sky-datasheet.pdf
[21] Chipcon Products, "2.4 GHz IEEE 802.15.4 ZigBee-ready RF Transceiver," Texas Instruments, Tech. Rep., 2006.
[22] C. Fischione, K. H. Johansson, A. Sangiovanni-Vincentelli, and B. Zurita Ares, "Energy consumption of minimum energy coding in cdma wireless sensor networks," *IEEE Transactions on Wireless Communications*, vol. 8, pp. 985–994, February 2009.
[23] A. Rantzer, "On prize mechanisms in linear quadratic team theory," in *IEEE Conference on Decision and Control*, December 2007.
[24] M. A. Hanson, "On sufficiency of the Kuhn-Tucker conditions," in *Journal of Mathematical Analysis and Applications*, vol. 80, 1981, pp. 545–Ű550.
[25] R. Osuna-Gomez, A. Rufian-Lizana, and P. Ruiz-Canales, "Invex functions and generalized convexity in multiobjective programming," in *Journal of optimization theory and applications*, September 1998, pp. 665 –661.
[26] R. Horst, P. M. Pardalos, and N. V. Thoai, *Introduction to Global Optimization, Nonconvex Optimization and its Applications*. Kluwer Academic Publisher, 1995.

APPENDIX

A.1 Proof of Theorem 3.3

Let

$$\mathbf{F}(\boldsymbol{x}) = \begin{bmatrix} \mathbf{f}(\boldsymbol{x}) \\ \mathbf{h}(\boldsymbol{x}) \end{bmatrix}^{\mathscr{D}}.$$

It follows from that Banach fixed point theorem that there exists a unique fixed point to $\boldsymbol{x} = \mathbf{F}(\boldsymbol{x})$ in the compact set $\mathscr{D}$. By our assumption, the fixed point must be feasible to the F-Lipschitz optimization problem.

It remains to show that the fixed point, $\boldsymbol{x}^*$, is optimal to problem (3.1). Take an arbitrary feasible point such that

$$\Psi(\boldsymbol{x}) := \boldsymbol{x} - \mathbf{F}(\boldsymbol{x}) = -\mathbf{b}$$

for some nonzero $\mathbf{b} \geq 0$. Let us compare the solutions

$$\Psi(\boldsymbol{x}) = -\mathbf{b} \qquad \Psi(\boldsymbol{x}^*) = 0.$$

Assuming that $\boldsymbol{x} + t(\boldsymbol{x}^* - \boldsymbol{x})$, $t \in [0,1]$ intersects the boundary of $\mathscr{D}$ only at a finite number of points, we have

$$\begin{aligned} \mathbf{b} = \Psi(\boldsymbol{x}^*) - \Psi(\boldsymbol{x}) &= \int_0^1 J\Psi(\boldsymbol{x} + t(\boldsymbol{x}^* - \boldsymbol{x}))(\boldsymbol{x}^* - \boldsymbol{x})dt \\ &= \left(I - \int_0^1 J\mathbf{F}(\boldsymbol{x} + t(\boldsymbol{x}^* - \boldsymbol{x}))dt\right)(\boldsymbol{x}^* - \boldsymbol{x}) \end{aligned} \tag{A.1}$$

where $J\mathbf{F}(\cdot) = \nabla\mathbf{F}(\cdot)^T$ denotes the Jacobian. Let us define

$$\mathbf{A} = [a_{ij}]_{i,j=1}^n, \quad a_{ij} = \int_0^1 \nabla_j F_i(\boldsymbol{x} + t(\boldsymbol{x}^* - \boldsymbol{x}))dt.$$

Then (A.1) can be rewritten as $\mathbf{y} = \mathbf{A}\mathbf{y} + \mathbf{b}$.

For any scalarization $\boldsymbol{\mu}^T f_0(\boldsymbol{x})$, where $\boldsymbol{\mu} \succ 0$, the change in the cost function from $\boldsymbol{x}$ to $\boldsymbol{x}^*$ is

$$\begin{aligned} \boldsymbol{\mu}^T f_0(\boldsymbol{x}^*) - \boldsymbol{\mu}^T f_0(\boldsymbol{x}) &= \int_0^1 \boldsymbol{\mu}^T J\mathbf{f}_0(\boldsymbol{x} + t(\boldsymbol{x}^* - \boldsymbol{x}))dt(\boldsymbol{x}^* - \boldsymbol{x}) \\ &=: \mathbf{c}^T(\boldsymbol{x}^* - \boldsymbol{x}), \end{aligned}$$

where $\mathbf{c} = [c_i]$, $c_i > 0$ and if in addition (3.2e) holds have $\mathbf{c} = c\mathbf{1}$ for some $c > 0$. We want to show that $\mathbf{c}^T\mathbf{y} > 0$ when $\mathbf{A}\mathbf{y} = \mathbf{b}$.

We consider the two cases. In the case when (3.2c) and (3.2b) hold we have $a_{ij} \geq 0$ and $\sum_{j=1}^n a_{ij} < 1$ (or $|A|_1 < 1$). Then $(I - \mathbf{A})^{-1} \succeq 0$ (positive matrix) with strictly positive diagonal elements. It is easy to see that this implies $\mathbf{c}^T(I - \mathbf{A})^{-1} \succ 0$ and hence $\mathbf{c}^T\mathbf{y} = \mathbf{c}^T(I - \mathbf{A})^{-1}\mathbf{b} > 0$.

In case when (3.2e) and (3.2f) hold we have $a_{ij} \leq 0$, $\sum_{i=1}^n a_{ij} < 1$ (or $|A|_\infty < 1$) and $\mathbf{c} = c\mathbf{1}$, for some $c > 0$. We need to show that $\mathbf{c}^T\mathbf{y} > 0$, where $\mathbf{y} = \mathbf{A}\mathbf{y} + \mathbf{b}$. Since $|A|_\infty < 1$ it follows that $\rho(A) < 1$ and $I - \mathbf{A}$ is invertible so our desired condition is $\mathbf{1}^T(I - \mathbf{A})^{-1}\mathbf{b} = \mathbf{b}^T(I - \mathbf{A}^T)^{-1}\mathbf{1} > 0$. It is thus equivalent to show that $\mathbf{b}^T\mathbf{z} > 0$ for $\mathbf{z} = \mathbf{A}^T\mathbf{z} + \mathbf{1} = \mathbf{A}^{2T}\mathbf{z} + (\mathbf{A}^T + I)\mathbf{1}$. Since

by assumption $\mathbf{A}^{2T} \succeq 0$ and $\rho(\mathbf{A}^{2T}) \leq |\mathbf{A}|_\infty^2 < 1$ it follows that $(I - \mathbf{A}^{2T})^{-1} \succeq 0$ with strictly positive diagonal elements. Hence,

$$\mathbf{b}^T \mathbf{z} = \mathbf{b}^T (I - \mathbf{A}^{2T})^{-1} (\mathbf{A}^T + I)\mathbf{1} > 0,$$

since $(\mathbf{A}^T + I)\mathbf{1} > 0$.

The case when the qualifying conditions (3.2h) hold can be treated as in the proof of Claim 3.12 below.

A.2 Proof of Theorem 3.7

We show that by the conditions of the theorem, the F-Lipschitz properties are satisfied for the optimization problem (3.4).

Conditions (3.8) imply that conditions (3.2a), (3.2c) and (3.2e) of an F-Lipschitz optimization problem are verified. Therefore, we need only to show contractivity of the right-hand side of the constraints of problem (3.4), namely that condition (3.8d) and (3.8c) imply (3.2c) and (3.2b). We focus on the inequality functions $f_i(\boldsymbol{x})$, $i = 1 \ldots, l$. The arguments used to show that these functions are contractive apply also to the equality functions $h_i(\boldsymbol{x})$, for $i = l+1, \ldots, n$.

Consider the constraints of problem (3.4). The function $f_i(\boldsymbol{x})$ is contractive with respect to the norm 1 if $\forall \boldsymbol{x} \in \mathscr{D}$, then $||\nabla f_i(\boldsymbol{x})||_\infty < 1$, namely

$$|1 - \gamma_i \nabla_i g_i(\boldsymbol{x})| + \gamma_i \sum_{i \neq j} |\nabla_j g_i(\boldsymbol{x})| < 1\,. \tag{A.2}$$

We show next that this holds if condition in (3.8c) holds. Let us choose γ_i such that

$$1 - \gamma_i \nabla_i g_i(\boldsymbol{x}) \geq 0\,.$$

Inequality (A.2) holds if

$$\gamma_i(-\nabla_i g_i(\boldsymbol{x}) + \sum_{i \neq j} |\nabla_j g_i(\boldsymbol{x})|) < 0\,,$$

which is certainly verified since $\gamma_i > 0$. The other cases follow in the same fashion.

Therefore, by the conditions of the theorem, problem (3.4) satisfies the F-Lipschitz qualifying properties. This concludes the proof.

A.3 Proof of Theorem 3.11

To give the proof, we need first a definition and two intermediate technical results. We use some results in Hanson's invex analysis [24]:

Definition A.1. *[24, 25] A continuously differentiable function $f(\boldsymbol{x}) : \mathscr{D} \to \mathbb{R}$ is invex if $\forall\ \boldsymbol{x} \in \mathscr{D}$ and $\mathbf{y} \in \mathscr{D}$ there exists a function $\boldsymbol{\eta}(\boldsymbol{x}, \mathbf{y}) : \mathscr{D} \times \mathscr{D} \to \mathbb{R}^n$ such that*

$$f(\boldsymbol{x}) - f(\mathbf{y}) \geq \boldsymbol{\eta}^T(\boldsymbol{x}, \mathbf{y}) \nabla f(\boldsymbol{x})\,.$$

Lemma A.2. *Let $q_i(\boldsymbol{x}) = x_i - f_i(\boldsymbol{x})$, $i = 1, \ldots, n$, where $f_i(\boldsymbol{x}) : \mathscr{D} \to \mathbb{R}$ is a Lipschitz function with modulus $\alpha_i < 1$. Then, $q_i(\boldsymbol{x})$ is invex with respect to the function*

$$\boldsymbol{\eta}_i(\boldsymbol{x}, \mathbf{y}) = -\frac{(1 - \alpha_i)\mathbf{1}}{\mathbf{1}^T \nabla q_i(\boldsymbol{x})} \|\boldsymbol{x} - \mathbf{y}\|\,.$$

PROOF. Since the function $f_i(\boldsymbol{x})$ is Lipschitz contractive, $q_i(\boldsymbol{x})$ is Lipschitz with constant $1 - \alpha_i$ and we have

$$\begin{aligned} |q_i(\boldsymbol{x}) - q_i(\mathbf{y})| &\leq (1 - \alpha_i)\|\boldsymbol{x} - \mathbf{y}\| \\ &= \mathbf{1}^T \nabla q_i(\boldsymbol{x}) \frac{(1 - \alpha_i)}{\mathbf{1}^T \nabla q_i(\boldsymbol{x})} \|\boldsymbol{x} - \mathbf{y}\|\,, \end{aligned}$$

where the last inequality is allowed by that $\mathbf{1}^T \nabla q_i(\boldsymbol{x}) = \mathbf{1}^T(\mathbf{e}_i - \nabla f(\boldsymbol{x})) > 0$, where $\mathbf{e}_i$ is the all zero vector with the exception of the i-th element, which is 1. It follows that

$$q_i(\boldsymbol{x}) - q_i(\mathbf{y}) \geq -\mathbf{1}^T \frac{(1 - \alpha_i)\nabla q_i(\boldsymbol{x})}{\mathbf{1}^T \nabla q_i(\boldsymbol{x})} \|\boldsymbol{x} - \mathbf{y}\|\,,$$

whereby the proposition follows by observing that

$$\boldsymbol{\eta}_i(\boldsymbol{x}, \mathbf{y}) = -\frac{(1 - \alpha_i)\mathbf{1}}{\mathbf{1}^T \nabla q_i(\boldsymbol{x})} \|\boldsymbol{x} - \mathbf{y}\|\,. \tag{A.3}$$

□

Lemma A.3. *Let $f(\boldsymbol{x}) = \sum_j \mu_j f_{0j}(\boldsymbol{x})$ the scalar objective function associated to the vector optimization problem* (3.1)*, where $0 < \mu_j \leq 1\ \forall\ j$. Let L_j be the modulus of $f_{0j}(\boldsymbol{x})\ \forall\ j$. Then, $f(\boldsymbol{x})$ is invex with respect to the function $\boldsymbol{\eta}_0(\boldsymbol{x}, \mathbf{y}) : \mathscr{D} \times \mathscr{D} \to \mathbb{R}^n$, with*

$$\boldsymbol{\eta}_0(\boldsymbol{x}, \mathbf{y}) = -\frac{\sum_j \mu_j L_j \mathbf{1}}{\mathbf{1}^T \nabla f(\boldsymbol{x})} \|\boldsymbol{x} - \mathbf{y}\|\,. \tag{A.4}$$

PROOF. The proof follows the same steps of Lemma A.2. □

We are now in the position to prove Theorem 3.11.

PROOF. Strong duality follows by showing that the KKT conditions are necessary and sufficient. The necessary condition is given by the Mangasarian-Fromowitz constraint qualification [26, pag. 25]. For this constraint qualification to hold, it is necessary that the gradient of the inequality constraints is a full rank matrix, which we know from Lemma 3.8.

Finally, the sufficiency of the KKT conditions is ensured by showing that the objective function and the constraints have a common invex function, see [24]: From Lemma A.2 and Lemma A.3, we define the common invex function $\boldsymbol{\eta}(\boldsymbol{x}, \mathbf{y}) : \mathscr{D} \times \mathscr{D} \to \mathbb{R}^n$ such that

$$\boldsymbol{\eta}(\boldsymbol{x}, \mathbf{y}) = \min_i \boldsymbol{\eta}_i(\boldsymbol{x}, \mathbf{y}) \qquad i = 0, \ldots, n\,,$$

where $\boldsymbol{\eta}_0(\boldsymbol{x}, \mathbf{y})$ is given in (A.4) and $\boldsymbol{\eta}_i(\boldsymbol{x}, \mathbf{y})$ is given in (A.3). Note that $-\infty < \boldsymbol{\eta}_i(\boldsymbol{x}, \mathbf{y}) < 0$ for $i = 1, \ldots, n$, and that $-\infty < \boldsymbol{\eta}_0(\boldsymbol{x}, \mathbf{y}) < 0$ because $1^T \nabla f(\boldsymbol{x}) > 0\ \forall \boldsymbol{x} \in \mathscr{D}$. □

A.4 Proof of Claim 3.12

Since strong duality holds from Theorem 3.11, we use the Lagrange dual problem to establish the sensitivity. Consider the scalarized problem associated to (3.1), where the vector objective function is converted into a scalar function by the Pareto coefficients $\boldsymbol{\mu} = [\mu_i]$, with $\boldsymbol{\mu} \succ 0$, and with $\mathbf{1}^T \boldsymbol{\mu} = 1$ [8]. The scalarized objective function of problem (3.1) is $f_0(\boldsymbol{x}) = [f_{0i}(\boldsymbol{x})] \in \mathbb{R}^m$, with $f_{0i}(\boldsymbol{x}) : \mathscr{D} \to \mathbb{R}$, $i = 1, \ldots, n$, we have:

$$\max_{\boldsymbol{x}} \quad \frac{\theta}{\max_{\boldsymbol{x} \in \mathscr{D}} \nabla \boldsymbol{\mu}^T f_0(\boldsymbol{x})} \boldsymbol{\mu}^T f_0(\boldsymbol{x}) \tag{A.5a}$$

$$\text{s.t.} \quad x_i \leq F_i(\boldsymbol{x}) \quad i = 1, \ldots, n \tag{A.5b}$$

$$\boldsymbol{x} \in \mathscr{D}\,.$$

Notice that we have normalized the scalarized objective function of problem (3.1) by an arbitrary constant given by $\theta > 0$ divided by the maximum of the derivative. Clearly, this does not affect the optimal solution as long as $\theta > 0$. Let $\boldsymbol{\lambda} = [\lambda_i] \in \mathbb{R}^n$ be the vector of Lagrangian multipliers. We show next that $0 < \boldsymbol{\lambda} < \mathbf{1}$ provided that θ is small enough, which ensures a low sensitivity of the optimization problem to perturbations of the constraints [1].

The Lagrange multipliers associated to this problem must satisfy the following condition coming from the derivative of the Lagrange

dual function:

$$(I - \nabla \mathbf{F}(\boldsymbol{x}))\boldsymbol{\lambda} = \frac{\theta \nabla \boldsymbol{\mu}^T f_0(\boldsymbol{x})}{\max_{\boldsymbol{x}\in\mathscr{D}} \nabla \boldsymbol{\mu}^T f_0(\boldsymbol{x})}\,, \qquad \text{(A.6)}$$

where $\mathbf{F}(\boldsymbol{x}) = [F_1(\boldsymbol{x}),\, F_2(\boldsymbol{x}), \ldots, F_n(\boldsymbol{x})]^T$. For simplicity of notation, let

$$\mathbf{b}(\boldsymbol{x}) = \frac{\theta}{\max_{\boldsymbol{x}\in\mathscr{D}} \nabla \boldsymbol{\mu}^T f_0(\boldsymbol{x})} \nabla \boldsymbol{\mu}^T f_0(\boldsymbol{x})\,.$$

Clearly, $0 < b_i < \theta, \forall i$. Then, it trivially follows that (A.6) gives

$$\boldsymbol{\lambda} = \nabla \mathbf{F}(\boldsymbol{x}^*)\boldsymbol{\lambda} + \mathbf{b}(\boldsymbol{x}^*)\,, \qquad \text{(A.7)}$$

at the optimal point $\boldsymbol{x}^*$. Now, suppose that the first condition of the second qualifying property (3.2c) holds, namely $\nabla_j f_i(\boldsymbol{x}) \geq 0\ \forall i \neq j$. Then since $\rho(\nabla \mathbf{F}(\boldsymbol{x})) < \|\nabla \mathbf{F}(\boldsymbol{x})\|_\infty < 1$ it follows that $(\mathbf{I} - \nabla \mathbf{F}(\boldsymbol{x}))^{-1} \succeq 0$ with strictly positive diagonal elements. Hence at the optimal point $\boldsymbol{x}^*$ we get $0 \prec \boldsymbol{\lambda} = (\mathbf{I} - \nabla \mathbf{F}(\boldsymbol{x}^*))^{-1}\mathbf{b}(\boldsymbol{x}^*) \prec \mathbf{1}$ given that θ is small enough.

Suppose that the second condition of the second qualifying property holds, namely $\nabla_j f_i(\boldsymbol{x}) \leq 0, \forall i, j$. We may rewrite (A.7) as

$$\boldsymbol{\lambda} = \nabla \mathbf{F}(x^*)\boldsymbol{\lambda} + \mathbf{b}(x^*) = \nabla \mathbf{F}(x^*)^2\boldsymbol{\lambda} + (\nabla \mathbf{F}(x^*) + \mathbf{I})\mathbf{b}(\boldsymbol{x}^*)$$

Due to (3.2e) we have $\nabla \mathbf{F}(\boldsymbol{x}^*)^2 \succeq 0$ and $\mathbf{b}(x) = c \cdot \theta \cdot \mathbf{1}$, for some positive constant c. Condition (3.2f) implies that $\rho(\nabla \mathbf{F}(\boldsymbol{x}^*)^2) \leq \|\nabla \mathbf{F}(\boldsymbol{x}^*)\|_1^2 < 1$ and thus that $(\mathbf{I} - \nabla \mathbf{F}(\boldsymbol{x}^*)^2)^{-1} \succeq 0$ with positive diagonal elements and moreover that $(\nabla \mathbf{F}(x^*) + \mathbf{I})\mathbf{b}(\boldsymbol{x}^*) \succ 0$. Hence, we have that $0 \prec \boldsymbol{\lambda} = (\mathbf{I} - \nabla \mathbf{F}(\boldsymbol{x}^*)^2)^{-1}(\nabla \mathbf{F}(x^*) + \mathbf{I})\mathbf{b}(\boldsymbol{x}^*) \prec \mathbf{1}$ given that θ is small enough.

Suppose that the qualifying condition (3.2h) holds. We have

$$\lambda_i^* = -\sum_{j=1}^n \nabla_i F_j(\boldsymbol{x}^*)\lambda_j^* + b_i(\boldsymbol{x}^*)\,, \quad \forall i\,. \qquad \text{(A.8)}$$

where $\mathbf{F}(\boldsymbol{x}) = [F_1(\boldsymbol{x}),\, F_2(\boldsymbol{x}), \ldots, F_n(\boldsymbol{x})]^T$. It follows that

$$\lambda_{\max} \leq \sum_{j=1}^n |\nabla_i F_j(\boldsymbol{x}^*)|\lambda_{\max} + b_i(\boldsymbol{x}^*)\,, \quad \forall i\,,$$

where $\lambda_{\max} = \max_i \lambda_i^*$. Thus

$$\lambda_{\max} \leq \sum_{j=1}^n \max_i |\nabla_i F_j(\boldsymbol{x}^*)|\lambda_{\max} + \max_i b_i(\boldsymbol{x}^*)\,,$$

whereby

$$\lambda_{\max} \leq \frac{\max_i b_i(\boldsymbol{x}^*)}{1 - \sum_{j=1}^n \max_i |\nabla_i F_j(\boldsymbol{x}^*)|} \quad \forall i\,.$$

The denominator of the previous inequality is always positive by the assumption that $\|\nabla F(\boldsymbol{x})\|_1 < 1$. We can compute a lower bound for the lagrangians. Let $\lambda_{\min} = \min_i \lambda_i^*$, then

$$\lambda_{\min} \geq -\sum_{j=1}^n |\nabla_i F_j(\boldsymbol{x}^*)|\lambda_{\max} + b_i(\boldsymbol{x}^*)\,, \quad \forall i\,,$$

Thus

$$\begin{aligned}
\lambda_{\min} &\geq -\sum_{j=1}^n \max_i |\nabla_i F_j(\boldsymbol{x}^*)|\lambda_{\max} + \min_i b_i(\boldsymbol{x}^*) \\
&\geq -\frac{-\max_i b_i(\boldsymbol{x}^*)}{1 - \sum_{j=1}^n \max_i |\nabla_i F_j(\boldsymbol{x}^*)|} \sum_{j=1}^n \max_i |\nabla_i F_j(\boldsymbol{x}^*)| \\
&\quad + \min_i b_i(\boldsymbol{x}^*) \\
&= \frac{\min_i b_i(\boldsymbol{x}^*)}{1 - \sum_{j=1}^n \max_i |\nabla_i F_j(\boldsymbol{x}^*)|} \\
&\quad - \frac{(\min_i b_i(\boldsymbol{x}^*) + \max_i b_i(\boldsymbol{x}^*)) \sum_{j=1}^n \max_i |\nabla_i F_j(\boldsymbol{x}^*)|}{1 - \sum_{j=1}^n \max_i |\nabla_i F_j(\boldsymbol{x}^*)|}\,.
\end{aligned}$$

We have that $\lambda_{\min}$ is positive if the numerator and the denominator of the previous inequality are positive, which happens if

$$\sum_{j=1}^n \max_i |\nabla_i F_j(\boldsymbol{x}^*)| < 1\,,$$

and

$$\sum_{j=1}^n \max_i |\nabla_i F_j(\boldsymbol{x}^*)| < \frac{\min_i b_i(\boldsymbol{x}^*)}{\min_i b_i(\boldsymbol{x}^*) + \max_i b_i(\boldsymbol{x}^*)}\,. \qquad \text{(A.9)}$$

Clearly, between the last two inequalities, (A.9) wins, which holds true by the assumption of the theorem. But then, the qualifying condition (3.2h) implies that the lagrangian are strictly positive and less than 1 by tuning θ. This concludes the proof.

A.5 Proof of Proposition 3.13

Throughout the proof, we let the norm $\|\cdot\|$ be the max norm. Let $\boldsymbol{x}(0) = \tilde{\boldsymbol{x}}(0)$ the initial feasible vector used to compute the optimal solution. It holds that $\lim_{k\to\infty} \boldsymbol{x}(k) = \lim_{k\to\infty} \mathbf{F}(\boldsymbol{x}(k-1)) = \boldsymbol{x}^*$ and $\boldsymbol{x}^* = \mathbf{F}(\boldsymbol{x}^*)$. We have

$$\tilde{\boldsymbol{x}}(1) = \mathbf{F}(\tilde{\boldsymbol{x}}(0)) + \mathbf{q}(0)\,.$$

By using the Lipschitz condition we get

$$\begin{aligned}
\|\tilde{\boldsymbol{x}}(1) - \boldsymbol{x}^*\| &= \|\mathbf{F}(\tilde{\boldsymbol{x}}(0)) - \mathbf{F}(\boldsymbol{x}^*) + \mathbf{q}(0)\| \\
&\leq \alpha_{\max}\|\tilde{\boldsymbol{x}}(0) - \boldsymbol{x}^*\| + \alpha_{\max}\|\mathbf{q}(0)\|,
\end{aligned}$$

where we used that $\mathbf{F}(\boldsymbol{x})$ is Lipschitz. Similarly,

$$\begin{aligned}
\|\tilde{\boldsymbol{x}}(2) - \boldsymbol{x}^*\| &= \|\mathbf{F}(\tilde{\boldsymbol{x}}(1)) - \mathbf{F}(\boldsymbol{x}^*) + \mathbf{q}(1)\| \\
&\leq \alpha_{\max}\|\tilde{\boldsymbol{x}}(1) - \boldsymbol{x}^*\| + \|\mathbf{q}(1)\| \\
&= \alpha_{\max}\|\mathbf{F}(\tilde{\boldsymbol{x}}(0)) - \mathbf{F}(x^*) + \mathbf{q}(0)\| + \|\mathbf{q}(1)\| \\
&\leq \alpha_{\max}^2\|\tilde{\boldsymbol{x}}(0) - \boldsymbol{x}^*\| + \alpha_{\max}\|\mathbf{q}(0)\| + \|\mathbf{q}(1)\|
\end{aligned}$$

and it is straightforward to generalize the iterations to

$$\begin{aligned}
\|\tilde{\boldsymbol{x}}(k) - \boldsymbol{x}^*\| &\leq \alpha_{\max}^k\|\tilde{\boldsymbol{x}}(0) - \boldsymbol{x}^*\| + \sum_{j=0}^{k-1} \alpha_{\max}^{k-1-j}\|\mathbf{q}(j)\| \\
&\leq \alpha_{\max}^k\|\tilde{\boldsymbol{x}}(0) - \boldsymbol{x}^*\| + \sum_{j=0}^{k-1} \alpha_{\max}^{k-1-j} q_{\max} \\
&\leq \alpha_{\max}^k\|\tilde{\boldsymbol{x}}(0) - \boldsymbol{x}^*\| + \frac{1 - \alpha_{\max}^{k-2}}{1 - \alpha_{\max}} q_{\max}\,,
\end{aligned}$$

where we used that $\|q(j)\| \leq q_{\max}$, and that $\|\boldsymbol{\alpha}\| < 1$. The result follows by taking the limit as $k \to \infty$.