Regularized Jacobi Iteration for Decentralized Convex Quadratic Optimization with Separable Constraints

Luca Deori, Kostas Margellos and Maria Prandini

***Abstract*—We consider multi-agent, convex quadratic optimization programs subject to separable constraints, where the constraint function of each agent involves only its local decision vector, while the decision vectors of all agents are coupled via a common objective function. We focus on a regularized variant of the so called Jacobi algorithm for decentralized computation in such problems. We provide a fixed-point theoretic analysis showing that the algorithm converges to a minimizer of the centralized problem under more relaxed conditions on the regularization coefficient from those available in the literature, and in particular with respect to scaled projected gradient algorithms. The efficacy of the proposed algorithm is illustrated by applying it to the problem of optimal charging of electric vehicles.**

***Index Terms*—Decentralized optimization, Jacobi algorithm, iterative methods, optimal charging control, electric vehicles.**

I. Introduction

OPTIMIZATION in multi-agent systems has attracted significant attention in the control and operations research communities, due to its applicability to different domains, e.g., energy [1], [2], mobility [3], [4], [5], robotic systems [6], etc. We focus on multi-agent optimization programs that are convex and are subject to constraints that are separable. The agents' decisions are, however, coupled by means of a common objective function, which is considered to be quadratic. The considered structure, although specific, captures a wide class of problems, like the electric vehicle charging problem studied in this paper, and is amenable to efficient numerical solvers tailored for quadratic optimization [7].

Solving such problems in a centralized fashion would require agents to share their local constraint functions, while even if this was possible it would unnecessarily increase the computational burden. To alleviate these issues we adopt an iterative, decentralized perspective, where agents perform local computations in parallel, and then exchange with each other their new solutions, or broadcast them to some central authority that sends an update to each agent. Admittedly, distributed optimization offers a more general communication setup, however, the fact that agents decision vectors are coupled via the objective function poses additional difficulties, preventing the use of distributed algorithms [8], [9]. Even upon an epigraphic reformulation, the resulting problem will not exhibit the structure typically encountered in distributed optimization, with the resulting coupling constraint not necessarily being of "budget" form [10]. A distributed gossip based gradient algorithm has been proposed in [11], but with reference to a noncooperative counterpart of the problem under study here. As such, it does not lead to a social welfare solution. Moreover, it requires an iteration varying step-size as opposed to the constant step-size considered in this paper.

Research was supported by the European Commission, H2020, under the project UnCoVerCPS, grant number 643921, and by EPSRC UK under the grant EP/P03277X/1. The authors would like to thank one anonymous reviewer for suggesting a refinement in the calculations of Section III-C.

L. Deori and M. Prandini are with the Dipartimento di Elettronica Informazione e Bioingegneria, Politecnico di Milano, Piazza Leonardo da Vinci 32, 20133 Milano, Italy, e-mail: {luca.deori, maria.prandini}@polimi.it

K. Margellos is with the Department of Engineering Science, University of Oxford, Parks Road, Oxford, OX1 3PJ, United Kingdom, e-mail: kostas.margellos@eng.ox.ac.uk

A. Related work

From a cooperative optimization point of view, algorithms for decentralized solutions to convex optimization problems with separable constraints can be found in [12], [13], and references therein. Two main algorithmic directions can be distinguished, both of them relying on an iterative process. The first one is based on each agent performing at every iteration a local gradient descent step, while keeping the decision variables of all other agents fixed to the values communicated at the previous iteration [14]–[16]. Under certain structural assumptions (differentiability of the objective function and Lipschitz continuity of its gradient), it is shown that this scheme converges to some minimizer of the centralized problem, for an appropriately chosen gradient step-size.

The second direction for decentralized optimization involves mainly the so called Jacobi algorithm, which serves as a proximal based alternative to gradient algorithms. The Gauss-Seidel algorithm exhibits similarities with the Jacobi one, but is not of parallelizable nature [17], unless a colouring scheme is adopted (see Section 1.2.4 in [12]). Under the Jacobi algorithmic setup, at every iteration, instead of performing a gradient step, each agent minimizes the common objective function subject to its local constraints in a best-response fashion, while keeping the decisions of all other agents fixed to their values at the previous iteration. In [12], it is shown that the Jacobi algorithm converges under certain contractiveness requirements, which are typically satisfied only under strong (or strict in case of quadratic objective functions) convexity assumptions that are, however, not imposed in the current work. In [18], [19], a regularized version of the Jacobi algorithm is proposed, however, an explicit condition on the regularization coefficient for convergence is not provided. A similar parallelizable, albeit different scheme, has been presented in [4], [20], [21], without employing regularization, while [22], [23] follow a randomized block coordinate descent approach and provide convergence results concerning the expected value of the objective functions. The results most closely related to our work appear in [24], [25]. In all aforementioned references, however, unlike our paper, convergence is limited to the optimal value and not in iterates.

From a non-cooperative perspective there has recently been a notable research activity using tools from mean-field and

aggregative game theory. Under a deterministic, discrete-time setting, [3], [5], [26] deal with the non-cooperative counterpart of our work, showing convergence not to a minimizer, but to an approximate Nash equilibrium of a related game, and to an exact Nash equilibrium in the limiting case where the number of agents tends to infinity. Using an approach similar to the regularized Jacobi algorithm it is shown in [27] that convergence to an exact Nash equilibrium for a finite number of agents can be achieved. A similar result, using a gradient based variant is recently provided in [28].

B. Contributions of this work and organization of the paper

We adopt a cooperative point of view, and consider a regularized Jacobi algorithm similar to the one in [18], [19], [24]. Our contributions extend these results as follows:
1. Focusing on the case where the objective function is quadratic, we show that the iterates generated by the regularized Jacobi algorithm converge to an optimal solution of the centralized problem counterpart, as opposed to the weaker statement that the iterates sequence achieves the optimal value allowing, however, an oscillatory behaviour (i.e., all limit points are optimal solutions) [24]. To achieve this, we follow a fundamentally different analysis from [24], relying on an operator theoretic approach. Our result serves as the Jacobi counterpart of gradient methods, thus complementing the work of [12], [29], [30]. The recent paper [31] shows convergence to an optimal solution of the centralized problem counterpart as well. However, the converge proof in [31, Theorem 1] strongly depends on results of this paper and relies on the agents' constraint sets to be convex polyhedra while our result requires these sets to be only compact and convex.
2. As opposed to [18], [19], we provide an explicit calculation of the regularization coefficient that ensures convergence, and show that the condition of Theorem 1 constitutes a relaxed version of that of [24] (see Theorem 3 and discussion on constant step-sizes therein), as well as of that of unscaled projected gradient methods (see Proposition 3.3 in Chapter 3 of [12] for convergence in value, and Theorem 4.1 in [29] or Theorem 2 in [30] for convergence in iterates) that ends up being the same with that of [24]. We also show that the main Jacobi iteration can be written as a scaled projected gradient step and derive an improved convergence condition (however, concerning convergence in optimal value not in iterates) under a particular choice of the scaling matrix and projection norm. Notably, the condition of Theorem 1 is less conservative. This improvement can affect significantly how well-behaved numerically the underlying optimization programs are.
3. From an application point of view, we extend the results of [4] on electric vehicle charging, achieving convergence to an optimal charging solution as opposed to convergence in value.

The results obtained here extend significantly our earlier work in [32], where no formal comparison with the gradient methods and [24] was provided. It should be noted that [12], [24], provide algorithms that are limited to convergence in optimal value under more restrictive choices on the step-size, however, are applicable to convex function and not necessarily quadratic as the focus of this paper. Our results can be extended to the non-quadratic case (the proof is similar to [24]), one can show convergence as far as the optimal value is concerned using a less restrictive step-size condition. We refer to this condition in Remark 3, while the reader is referred to the technical memorandum [33] for more details and proofs.

Section II introduces the problem under study and states the proposed algorithm. In Section III we provide the main convergence result and a comparison with scaled projected gradient methods and the algorithm of [24]. Section IV provides an extensive simulation study for the electric vehicle charging control case study, while Section V concludes the paper and outlines some directions for future research.

II. Decentralized problem formulation

A. Motivating example: Optimal charging of electric vehicles

We consider the problem of optimizing the charging strategy for a fleet of m plug-in electric vehicles (PEVs) over a finite horizon T. Following [3], [5], [26], the PEV charging problem is given by the following optimization problem.

$$\min_{\left\{\{x^i(t)\}_{i=1}^m\right\}_{t=0}^T} \frac{1}{m}\sum_{t=0}^{T} p(t)\Big(d(t)+\sum_{i=1}^{m} x^i(t)\Big)^2 \tag{1}$$
$$\text{subject to}$$
$$\sum_{t=0}^{T} x^i(t)=\gamma^i, \text{ for all } i=1,\dots,m$$
$$\underline{x}^i(t)\le x^i(t)\le \overline{x}^i(t), \text{ for all } i=1,\dots,m,\ t=0,\dots,T,$$

where $p(t)\in\mathbb{R}$ is an electricity price coefficient at time t, $d(t)\in\mathbb{R}$ represents the non-PEV demand at time t, $x^i(t)\in\mathbb{R}$ is the charging rate of vehicle i at time t, $\gamma^i\in\mathbb{R}$ represents a prescribed charging level to be reached by each vehicle i at the end of the considered time horizon, and $\underline{x}^i(t),\overline{x}^i(t)\in\mathbb{R}$ are bounds on the minimum and maximum value of $x^i(t)$, respectively. The objective function in (1) encodes the total electricity cost given by the demand (both PEVs and non-PEVs) multiplied by the price of electricity, which in turn depends linearly on the total demand through $p(t)$, thus giving rise to the quadratic function in (1). This linear dependency of price with respect to the total demand models the fact that agents/vehicles are price anticipating authorities, anticipating their consumption to have an effect on the electricity price (see introduction in [2] for further elaboration on price anticipating agents). Problem (1) can be rewritten as

$$\min_{x\in\mathbb{R}^{m(T+1)}} (d+Ax)^\top P(d+Ax) \tag{2}$$
$$\text{subject to: } x^i\in X^i, \text{ for all } i=1,\dots,m,$$

where $P=(1/m)\mathrm{diag}(p)\in\mathbb{R}^{(T+1)\times(T+1)}$, and $\mathrm{diag}(p)$ is a matrix with $p=(p(0),\dots,p(T))\in\mathbb{R}^{T+1}$ on its diagonal. $A=\mathbf{1}_{1\times m}\otimes I\in\mathbb{R}^{(T+1)\times m(T+1)}$, where $\otimes$ denotes the Kronecker product, and $I\in\mathbb{R}^{(T+1)\times(T+1)}$ the identity matrix. Moreover, $d=(d(0),\dots,d(T))\in\mathbb{R}^{T+1}$, $x=(x^1,\dots,x^m)\in\mathbb{R}^{m(T+1)}$, $x^i=(x^i(0),\dots,x^i(T))\in\mathbb{R}^{T+1}$, and X^i is the constraint set of vehicle i, $i=1,\dots,m$, in (1).

Algorithm 1 Decentralized algorithm

1: **Initialization**
2: $k = 0$.
3: Consider $x_0^i \in X^i$, for all $i = 1, \ldots, m$.
4: **For** $i = 1, \ldots, m$ **repeat until convergence**
5: Agent i receives x_k^{-i} from central authority.
6: $x_{k+1}^i = \arg\min_{z^i \in X^i} \{ f(z^i, x_k^{-i}) + c\|z^i - x_k^i\|^2 \}$.
7: $k \leftarrow k + 1$.

B. Problem statement

Motivated by the electric vehicle charging control problem in (2), we consider the following class of programs:

$$\mathcal{P} : \min_{\{x^i \in \mathbb{R}^{n_i}\}_{i=1}^m} f(x^1, \ldots, x^m) \tag{3}$$

$$\text{subject to: } x^i \in X^i, \text{ for all } i = 1, \ldots, m, \tag{4}$$

where each agent i, $i = 1, 2, \ldots, m$, has a local decision vector $x^i \in \mathbb{R}^{n_i}$ and a local constraint set $X^i \subseteq \mathbb{R}^{n_i}$, and cooperates to determine a minimizer of $f : \mathbb{R}^{n_1} \times \ldots \times \mathbb{R}^{n_m} \to \mathbb{R}$, which couples its decision vector with those of the other agents.

Assumption 1. *The objective function $f : \mathbb{R}^{n_1} \times \ldots \times \mathbb{R}^{n_m} \to \mathbb{R}$ is given by $f(x^1, \ldots, x^m) = x^\top Q x + q^\top x$, where $x = [(x^1)^\top, \ldots, (x^m)^\top]^\top \in \mathbb{R}^n$ with $n = \sum_{i=1}^m n_i$, $Q \in \mathbb{R}^{n \times n}$ is symmetric and positive semi-definite ($Q = Q^\top \succeq 0$) and $q \in \mathbb{R}^n$. Moreover, the sets $X^i \subseteq \mathbb{R}^{n_i}$, $i = 1, \ldots, m$, are non-empty, compact and convex.*

Note that Q is assumed to be symmetric without loss of generality; in the opposite case it could be split in a symmetric and an antisymmetric part, with the latter giving rise to terms that simplify each other.

Remark 1 (Problem generalization)**.** *We also allow for objective functions of the form $f(x^1, \ldots, x^m) = x^\top Q x + q^\top x + \sum_{i=0}^m g^i(x^i)$, where the $g^i(x^i)$,$i = 1, \ldots, m$, are convex functions that could encode a utility function for each agent. In this case an epigraphic reformulation can be exploited to bring the cost back to be quadratic. Letting $y^i = [x^{i,\top}\ h^i]^\top$ be the decision vector of agent i, the local constraint set can be then defined as $Y^i = X^i \cap \{g^i(x^i) \leq h^i\}$, while the objective function can be rewritten as $x^\top Q x + q^\top x + \sum_{i=0}^m h^i$, which is quadratic in $y = [y^{1,\top}\ \ldots\ y^{m,\top}]^\top$.*

Under Assumption 1, the function f is convex and hence continuous, while the constraint set $X = X^1 \times \cdots \times X^m$ is non-empty and compact, as result of Weierstrass' theorem [12, Proposition A8, p. 625], $\mathcal{P}$ admits at least one optimal solution. However, $\mathcal{P}$ does not necessarily admit a unique minimizer.

With a slight abuse of notation, for each i, $i = 1, \ldots, m$, let $f(\cdot, x^{-i}) : \mathbb{R}^{n_i} \to \mathbb{R}$ be the objective function in (3) as a function of the decision vector x^i of agent i, when the decision vectors of all other agents are fixed to $x^{-i} \in \mathbb{R}^{n-n_i}$. We will occasionally also write $f(x)$ instead of $f(x^1, \ldots, x^m)$.

C. Regularized Jacobi algorithm

Solving problem $\mathcal{P}$ in a centralized fashion is not always possible since agents may not be willing to share X^i, $i = 1, \ldots, m$. Moreover, even if this was the case, solving $\mathcal{P}$ in one shot might be computationally challenging. To overcome this and account for information sharing issues, motivated by the separable structure of $\mathcal{P}$ we follow a decentralized, iterative approach as described in Algorithm 1.

Initially, each agent i, $i = 1, \ldots, m$, starts with some value $x_0^i \in X^i$, such that $\left(x_0^1, \ldots, x_0^m\right)$ is feasible (step 3, Algorithm 1). At iteration $k + 1$, each agent i receives x_k^{-i} (step 5, Algorithm 1) from the central authority, and updates its estimate for x^i by solving a local minimization problem (step 6, Algorithm 1). The performance criterion in this local problem is a linear combination of the objective $f(z^i, x_k^{-i})$, where the variables of all other agents apart from the i-th one are fixed to their values at iteration k, and a quadratic regularization term, penalizing the difference between z^i and the value of agent's i own variable at iteration k, i.e., x_k^i. The relative importance of these two terms is dictated by the regularization coefficient $c \in \mathbb{R}_+$, which plays a key role in determining the convergence properties of Algorithm 1. Note that under Assumption 1, and due to the presence of the quadratic penalty term, the resulting problem is strictly convex with respect to z^i, and hence admits a unique minimizer.

Remark 2 (Information exchange)**.** *To implement Algorithm 1, at iteration $k + 1$, it is needed that some central authority collects and broadcasts the current solution of each agent to all others, so that each of them can compute $f(\cdot, x_k^{-i})$. However, in the case where the coupling in the objective function is only through the average of some agents' variables as in the example of Section II-A, at every iteration k the central authority needs to broadcast only the average of the agents' decisions, or in other words the cumulative charging $d + Ax_k$ with reference to the electric vehicle case study. Each agent will then be able to compute $f(\cdot, x_k^{-i})$ by subtracting from the average the value its local decision vector x_k^i.*

III. Main convergence result

We start defining some matrices that will be used in the following: for all $i = 1, \ldots, m$, let $Q_{i,i}$ denote the i-th block of Q, with row and column indices corresponding to x^i, where $x = [x^{1,\top}\ \ldots\ x^{m,\top}]^\top$. Denote then by Q_d a block diagonal matrix whose i-th block is $Q_{i,i}$, and let $Q_z = Q - Q_d$ denote the off (block) diagonal part of Q. Since Q is assumed to be symmetric, Q_z is symmetric as well and its eigenvalues are all real. Since Q_z has zero trace, at least one of its eigenvalues will be non-negative. As a result, $\lambda_{Q_z}^{\max} \geq 0$, where $\lambda_{Q_z}^{\max}$ denotes the maximum eigenvalue of Q_z.

Theorem 1. *Under Assumption 1, if $c > \lambda_{Q_z}^{\max}$, then Algorithm 1 converges to a minimizer of $\mathcal{P}$.*

Theorem 1 provides an explicit bound on c that ensures convergence. Such a bound is derived by a fixed-point theoretic approach. Note that if the objective function f was strictly convex with respect to x, then the standard Jacobi iteration of [12] can be adopted instead of the regularized version. In that case, geometric convergence to some minimizer of $\mathcal{P}$ is guaranteed by means of Proposition 3.5 in [12].

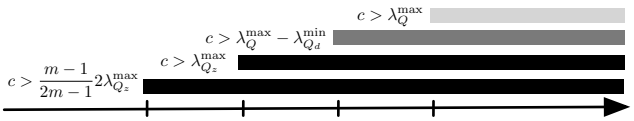

Fig. 1: Bar plot for the bound on c (from bottom to top): Algorithm of [33] (convergence in value); Theorem 1 (convergence in iterates); scaled projected gradient algorithm (convergence in value); Algorithm of [24] (convergence in value) and unscaled projected gradient algorithm (convergence in iterates).

Remark 3 (Connection with [24] and extensions)**.** *For the more general case of a convex objective function, by [24, Theorem 3] (see constant step-size condition), it can be shown that Algorithm 1 converges to the optimal value of $\mathcal{P}$ for c greater than one half of the Lipschitz constant of the objective function gradient, which for the case of quadratic objective functions is $2\lambda_Q^{\max}$, thus leading to $c > \lambda_Q^{\max}$. The sequence of iterates, however, may not converge and exhibit an oscillatory behaviour. Under the same condition on c it is shown unscaled projected gradient algorithms with step-size $1/c$ (i.e., two over the Lipschitz constant of the gradient) can converge not only in value, but also in iterates (see Proposition 3.3 in Chapter 3 of [12] for convergence in value, and Theorem 4.1 in [29] or Theorem 2 in [30] for convergence in iterates).*

In Section III-C we write the Jacobi iteration as a scaled projected gradient step and show that it converges in value but not in iterates if $c > \lambda_Q^{\max} - \lambda_{Q_d}^{\min}$, which is less restrictive than the aforementioned conditions. This result is strengthened even further if $c > \lambda_{Q_z}^{\max}$ according to Theorem 1.

By Theorem 3 of [33] it can be shown that, as far as the optimal value is concerned, Algorithm 1 converges for $c > \frac{m-1}{2m-1} 2\lambda_{Q_z}^{\max}$. The latter is a relaxed version for the condition $c > \lambda_{Q_z}^{\max}$ of Theorem 1, since $\frac{1}{2} > \frac{m-1}{2m-1}$, for all m. However, Theorem 1 ensures convergence to some minimizer and not just convergence in value. This result is shown in [33] using an analysis similar to the proof of Theorem 3 in [24], that is based on Proposition 1 and sequence convergence properties (see Exercise 1.19 in [34] (p. 18)). The relationship between the various conditions on c is pictorially shown in Figure 1.

A. Preliminary results

The results of this section hold under Assumption 2.

Assumption 2. *The function $f : \mathbb{R}^{n_1} \times \ldots \times \mathbb{R}^{n_m} \to \mathbb{R}$ is continuously differentiable, and jointly convex with respect to all arguments, i.e., convex with respect to x. The sets $X^i \subseteq \mathbb{R}^{n_i}$, $i = 1, \ldots, m$, are non-empty, compact and convex.*

1) Minimizers and fixed-points definitions: By (3)-(4), the set of minimizers of $\mathcal{P}$ is given by

$$M = \arg \min_{\{z^i \in X^i\}_{i=1}^m} f(z^1, \ldots, z^m) \subseteq X. \tag{5}$$

Following the discussion below Assumption 1, M is non-empty. Note that M is not necessarily a singleton; this is the case if f is jointly strictly convex with respect to its arguments.

For each i, $i = 1, \ldots, m$, consider the mappings $T^i : X \to X^i$ and $\widetilde{T}^i : X \to X^i$, defined such that, for any $x = (x^1, \ldots, x^m) \in X$,

$$T^i(x) = \arg \min_{z^i \in X^i} \|z^i - x^i\|^2 \tag{6}$$
$$\text{subject to: } f(z^i, x^{-i}) \leq \min_{\zeta^i \in X^i} f(\zeta^i, x^{-i}),$$

$$\widetilde{T}^i(x) = \arg \min_{z^i \in X^i} \left\{ f(z^i, x^{-i}) + c\|z^i - x^i\|^2 \right\}. \tag{7}$$

The mapping in (6) serves as a tie-break rule to select, in case $f(\cdot, x^{-i})$ admits multiple minimizers over X^i, the one closer to x^i with respect to the Euclidean norm. Note that in (6) and (7) we use equality instead of inclusion since the corresponding minimizers $T^i(x)$ and $\widetilde{T}^i(x)$, respectively, are unique. Note also that with x_k in place of x, (7) implies that the update step 6 in Algorithm 1 can be equivalently represented by $x_{k+1}^i = \widetilde{T}^i(x_k)$.

Define also the mappings $T : X \to X$ and $\widetilde{T} : X \to X$, such that their components are given by T^i and $\widetilde{T}^i$,

$$T(x) = \arg \min_{z \in X} \sum_{i=1}^m \|z^i - x^i\|^2 \tag{8}$$
$$\text{subject to: } f(z^i, x^{-i}) \leq \min_{\zeta^i \in X^i} f(\zeta^i, x^{-i}), \ \forall i = 1, \ldots, m,$$

$$\widetilde{T}(x) = \arg \min_{z \in X} \sum_{i=1}^m \left\{ f(z^i, x^{-i}) + c\|z^i - x^i\|^2 \right\}, \tag{9}$$

where the terms inside the summation in (8) and (9) are decoupled. The set of fixed-points of T and $\widetilde{T}$ is, respectively, are given by

$$F_T = \left\{ x \in X : \ x^i = T^i(x), \text{ for all } i = 1, \ldots, m \right\}, \tag{10}$$
$$F_{\widetilde{T}} = \left\{ x \in X : \ x^i = \widetilde{T}^i(x), \text{ for all } i = 1, \ldots, m \right\}. \tag{11}$$

2) Connections between minimizers and fixed-points: We report here a fundamental optimality result.

Proposition 1 ([12, Proposition 3.1])**.** *Assume that f is a continuously differentiable function and X is a non-empty, closed and convex set. We then have that,*
1) if $x \in X$ minimizes f over X, then $(z - x)^\top \nabla f(x) \geq 0$, for all $z \in X$.
2) if f is also convex on X, then the condition of the previous part is also sufficient for $x \in \arg\min_{z \in X} f(z)$.

We show that the set of minimizers M of $\mathcal{P}$ in (5) and the set of fixed-points F_T of the mapping T in (8) coincide.

Proposition 2. *Under Assumption 2, $M = F_T$.*

Proof. 1) $M \subseteq F_T$: Fix any $x \in M$. For $i = 1, \ldots, m$, denote x by (x^i, x^{-i}). The fact that $x \in M$ implies that $f(x^i, x^{-i})$ will be no greater than $f(\zeta^i, x^{-i})$, for all $\zeta^i \in X^i$, i.e., $f(x^i, x^{-i}) \leq \min_{\zeta^i \in X^i} f(\zeta^i, x^{-i})$, which means that x satisfies the inequality in (8). Moreover x is also optimal for the objective function in (8), since it results in zero cost. Hence, by (8), x is a fixed-point of T, i.e., $x \in F_T$.
2) $F_T \subseteq M$: Fix any $x \in F_T$. By the definition of F_T we have $f(x^i, x^{-i}) \leq \min_{\zeta^i \in X^i} f(\zeta^i, x^{-i})$, for all $i = 1, \ldots, m$. The last statement implies that x^i is the minimizer of $f(\cdot, x^{-i})$

over X^i. For all $i = 1, \ldots, m$, by the first part of Proposition 1 (with $f(\cdot, x^{-i})$ in place of f) we then have that

$$(z^i - x^i)^\top \nabla^i f(x^i, x^{-i}) \geq 0, \text{ for all } z^i \in X^i, \tag{12}$$

where $\nabla^i f(x^i, x^{-i})$ is the i-th component of the gradient $\nabla f(\cdot, x^{-i})$ of $f(\cdot, x^{-i})$, evaluated at x^i. By (12), we then have that $\sum_{i=1}^m (z^i - x^i)^\top \nabla^i f(x^i, x^{-i}) \geq 0$ for all $z^i \in X^i$, $i = 1, \ldots, m$, which, by setting $x = (x^1, \ldots, x^m)$, $z = (z^1, \ldots, z^m)$, can be written as $(z - x)^\top \nabla f(x) \geq 0$, for all $z \in X$. By the second part of Proposition 1, and since f is jointly convex with respect to all elements of x, the last statement implies that x minimizes f over X, i.e., $x \in M$. □

The connection between minimizers, fixed-points and variational inequalities similar to (12), has been also investigated in [35], in the context of non-cooperative games.

Proposition 3. *Under Assumption 2, $F_T = F_{\widetilde{T}}$.*

Proof. 1) $F_T \subseteq F_{\widetilde{T}}$: Fix any $x \in F_T$. By (10), this is equivalent to the fact that $x^i = T^i(x)$, for all $i = 1, \ldots, m$, which, due to the definition of T implies that, for all $i = 1, \ldots, m$, $f(x^i, x^{-i}) \leq \min_{\zeta^i \in X^i} f(\zeta^i, x^{-i})$. This implies that x^i minimizes $f(\cdot, x^{-i})$ over X^i, hence, by the first part of Proposition 1 (with $f(\cdot, x^{-i})$ in place of f) we have that $(z^i - x^i)^\top \nabla^i f(x^i, x^{-i}) \geq 0$, for all $z^i \in X^i$. Let $f_c(z^i, x) = f(z^i, x^{-i}) + c\|z^i - x^i\|^2$, for all z^i, $i = 1, \ldots, m$, and notice that $\nabla f_c(x^i, x) = \nabla f(x^i, x^{-i})$, since the gradient of the quadratic penalty term vanishes at x^i. We then have that, for all $i = 1, \ldots, m$,

$$(z^i - x^i)^\top \nabla^i f_c(x^i, x) \geq 0, \text{ for all } z^i \in X^i. \tag{13}$$

Since $f_c(\cdot, x)$ is strictly convex with respect to its first argument, by the second part of Proposition 1 (with $f_c(\cdot, x)$ in place of f), (13) implies that, for all $i = 1, \ldots, m$, x^i is the unique minimizer of $f_c(\cdot, x)$ over X^i, i.e.,

$$x^i = \arg\min_{z^i \in X^i} f(z^i, x^{-i}) + c\|z^i - x^i\|^2. \tag{14}$$

By (7), (14) is equivalent to $x^i = \widetilde{T}^i(x)$, for all $i = 1, \ldots, m$. 2) $F_{\widetilde{T}} \subseteq F_T$: Fix any $x \in F_{\widetilde{T}}$. By (11) this is equivalent to the fact that $x^i = \widetilde{T}^i(x)$, for all $i = 1, \ldots, m$, which, by the definition of $\widetilde{T}^i$ in (7), implies that, for all $i = 1, \ldots, m$,

$$x^i = \arg\min_{z^i \in X^i} f(z^i, x^{-i}) + c\|z^i - x^i\|^2. \tag{15}$$

Let again $f_c(z^i, x) = f(z^i, x^{-i}) + c\|z^i - x^i\|^2$. Equation (15) implies then that, for all $i = 1, \ldots, m$, x^i minimizes $f_c(\cdot, x)$ over X^i, and by the first part of Proposition 1 (with $f_c(\cdot, x)$ in place of f) leads to $(z^i - x^i)^\top \nabla^i f_c(x^i, x) \geq 0$, for all $z^i \in X^i$. Notice that $\nabla f_c(x^i, x) = \nabla f(x^i, x^{-i})$, since the gradient of $c\|z^i - x^i\|^2$ with respect to z^i vanishes at x^i. Therefore, for all $i = 1, \ldots, m$, we have that

$$(z^i - x^i)^\top \nabla^i f(x^i, x^{-i}) \geq 0, \text{ for all } z^i \in X^i. \tag{16}$$

Since $f(\cdot, x^{-i})$ is convex with respect to its first argument, by the second part of Proposition 1, (16) implies that x^i minimizes $f(\cdot, x^{-i})$ over X^i. In other words, $x^i \in \arg\min_{z^i \in X^i} f(z^i, x^{-i})$, for all $i = 1, \ldots, m$. This in turn implies that, for all $i = 1, \ldots, m$, $f(x^i, x^{-i}) \leq f(z^i, x^{-i})$, for all $z^i \in X^i$, i.e., $f(x^i, x^{-i}) \leq \min_{z^i \in X^i} f(z^i, x^{-i})$. The last inequality shows that x satisfies the inequality in (8). Moreover, it minimizes the objective function in (8), since it results in zero cost, so $x = T(x)$. □

By Propositions 2 and 3 we have that the set of minimizers M of $\mathcal{P}$ coincides with the fixed-points of the mapping $\widetilde{T}$.

Corollary 1. *Under Assumption 2, $M = F_{\widetilde{T}}$.*

B. Proof of Theorem 1

Step 6 of Algorithm 1 can be equivalently written as $x^i_{k+1} = \widetilde{T}^i(x_k)$, which entails that $x_{k+1} = \widetilde{T}(x_k)$, i.e., a Picard-Banach iteration of $\widetilde{T}$ (see [36] (Chapter 1.2) for a definition). Since $\widetilde{T}$ is non-empty (it coincides with M by Corollary 1), we only need to prove that $\widetilde{T}$ is firmly non-expansive (see [37] (Section 1) for a definition in general Hilbert spaces). If that is the case, then, by [37], [38], we have that the Picard-Banach iteration converges to a fixed-point of $\widetilde{T}$, for any initial condition x_0. By Corollary 1 this fixed-point will also be a minimizer of $\mathcal{P}$. We next show that if $c > \lambda^{\max}_{Q_z}$, then $\widetilde{T}(\cdot)$ is indeed firmly non-expansive with respect to $\|\cdot\|_{Q_d + I_c - Q}$ (I_c is the identity matrix I of appropriate dimensions weighted by c), i.e.,

$$\begin{aligned}&\|\widetilde{T}(x) - \widetilde{T}(y)\|^2_{Q_d + I_c - Q} \\ &\qquad \leq (x - y)^\top (Q_d + I_c - Q)(\widetilde{T}(x) - \widetilde{T}(y)),\end{aligned} \tag{17}$$

thus establishing Theorem 1. To this end, by Assumption 1,

$$\begin{aligned}\widetilde{T}(x) &= \arg\min_{z \in X} \sum_{i=1}^m f(z^i, x^{-i}) + c\|z^i - x^i\|^2 \\ &= \arg\min_{z \in X} \sum_{i=1}^m (z^i)^\top (Q_{i,i} + I_c) z^i \\ &\qquad + (2(x^{-i})^\top Q_{-i,i} - 2(x^i)^\top I_c + q_i^\top) z^i \\ &= \arg\min_{z \in X} z^\top (Q_d + I_c) z + (2x^\top Q_z - 2x^\top I_c + q^\top) z.\end{aligned} \tag{18}$$

Notice the slight abuse of notation in (18), where I_c in the second and the third equality are not of the same dimension. Let $\xi(x) = (Q_d + I_c)^{-1}(I_c x - Q_z x - q/2)$ denote the unconstrained minimizer of (18). We then have that

$$\begin{aligned}\widetilde{T}(x) &= \arg\min_{z \in X} (z - \xi(x))^\top (Q_d + I_c)(z - \xi(x)) \\ &= [\xi(x)]^X_{Q_d + I_c},\end{aligned} \tag{19}$$

where $[\xi(x)]^X_{Q_d + I_c}$ denotes the projection, with respect to $\|\cdot\|_{Q_d + I_c}$, of $\xi(x)$ on X. Note that $Q_d + I_c$ is positive definite for $c \in \mathbb{R}_+$, so its inverse exists and the projection is well defined. We then have that

$$\begin{aligned}&\|\widetilde{T}(x) - \widetilde{T}(y)\|^2_{Q_d + I_c} \\ &= \|\,[\xi(x)]^X_{Q_d + I_c} - [\xi(y)]^X_{Q_d + I_c}\,\|^2_{Q_d + I_c} \\ &\leq (\xi(x) - \xi(y))^\top (Q_d + I_c)([\xi(x)]^X_{Q_d + I_c} - [\xi(y)]^X_{Q_d + I_c}) \\ &= (x - y)^\top (I - Q(Q_d + I_c)^{-1})(Q_d + I_c) \\ &\qquad \times ([\xi(x)]^X_{Q_d + I_c} - [\xi(y)]^X_{Q_d + I_c}) \\ &= (x - y)^\top (Q_d + I_c - Q)([\xi(x)]^X_{Q_d + I_c} - [\xi(y)]^X_{Q_d + I_c}),\end{aligned} \tag{20}$$

where the first inequality follows from the definition of a firmly non-expansive mapping and the fact that any projection mapping is firmly non-expansive (see Proposition 4.8 in [39]). The second equality is due to the definition ξ, and the last one follows after performing the matrix multiplication. Since $Q \succeq 0$, then $\|\widetilde{T}(x)-\widetilde{T}(y)\|^2_{Q_d+I_c-Q} \leq \|\widetilde{T}(x)-\widetilde{T}(y)\|^2_{Q_d+I_c}$. This, together with (20), implies that

$$\|\widetilde{T}(x)-\widetilde{T}(y)\|^2_{Q_d+I_c-Q} \leq (x-y)^\top(Q_d+I_c-Q)(\widetilde{T}(x)-\widetilde{T}(y)). \quad (21)$$

By the definition of a firmly non-expansive mapping [39], (21) implies that, if $Q_d+I_c-Q \succ 0$, $\widetilde{T}$ is firmly non-expansive with respect to $||\cdot||_{Q_d+I_c-Q}$. The condition $Q_d+I_c-Q \succ 0$ can be satisfied by choosing $c > \lambda^{\max}_{Q_z}$.

C. Connection with gradient algorithms

Recalling the formulation in (18) and (19), $x^i_{k+1} = \widetilde{T}^i(x_k)$, $i = 1,\dots,m$, in step 6 of Algorithm 1 can be equivalently written as a scaled projected gradient step as follows:

$$\begin{aligned} x_{k+1} &= [\xi(x_k)]^X_{Q_d+I_c} \\ &= \Big[(Q_d+I_c)^{-1}(Q_d+I_c-Q)x_k - (Q_d+I_c)^{-1}\frac{q}{2}\Big]^X_{Q_d+I_c} \\ &= \Big[x_k - (Q_d+I_c)^{-1}(Qx_k+\frac{q}{2})\Big]^X_{Q_d+I_c} \\ &= \Big[x_k - \frac{1}{2c}(\frac{Q_d}{c}+I)^{-1}(2Qx_k+q)\Big]^X_{\frac{Q_d}{c}+I}, \end{aligned} \quad (22)$$

where the first equality follows by the definition of $\xi(x)$ and of Q_z, and the last equality is obtained scaling by c. The gradient $2Qx+q$ of the original cost appears from the definition of $\xi(x)$, $1/(2c)$ plays the role of the gradient step-size, and $(\frac{Q_d}{c}+I)$ is the scaling matrix (see [12, Section 3.3.3]). Notice that $Q_d \succeq 0$, as a result of $Q \succeq 0$. Therefore, for any $c > 0$,

$$\begin{aligned} (x-y)^\top(\frac{Q_d}{c}+I)(x-y) &\geq (1+\lambda^{\min}_{\frac{Q_d}{c}})\|x-y\|^2_2, \\ &= (1+\frac{1}{c}\lambda^{\min}_{Q_d})\|x-y\|^2_2, \text{ for all } x,y \in X, \end{aligned} \quad (23)$$

where $\lambda^{\min}_{Q_d}$ denotes the minimum eigenvalue of Q_d, and is non-negative as a result of $Q_d \succeq 0$.

We can write x_{k+1} in (22) as the unique solution of

$$\begin{aligned} x_{k+1} = \arg\min_{z\in X} c(z-x_k)^\top(\frac{Q_d}{c}+I)(z-x_k) \\ + (z-x_k)^\top(2Qx_k+q). \end{aligned} \quad (24)$$

By optimality of x_{k+1}, and since $z = x_k \in X$ results in zero objective value, we have that $c(x_{k+1}-x_k)^\top(\frac{Q_d}{c}+I)(x_{k+1}-x_k) + (x_{k+1}-x_k)^\top(2Qx_k+q) \leq 0$, which implies that

$$\begin{aligned} (x_{k+1}-x_k)^\top(2Qx_k+q) \\ \leq -c(x_{k+1}-x_k)^\top(\frac{Q_d}{c}+I)(x_{k+1}-x_k). \end{aligned} \quad (25)$$

By (23) and (25) (notice that $x_{k+1}, x_k \in X$), we have that

$$\begin{aligned} (x_{k+1}-x_k)^\top(2Qx_k+q) \\ \leq -c(1+\frac{1}{c}\lambda^{\min}_{Q_d})\|x_{k+1}-x_k\|^2_2. \end{aligned} \quad (26)$$

By the Descent Lemma (Lemma 2.1 in [12]), for the quadratic objective function of Assumption 1 we obtain that

$$\begin{aligned} f(x_{k+1}) \leq f(x_k) + (x_{k+1}-x_k)^\top(2Qx_k+q) \\ + \lambda^{\max}_Q\|x_{k+1}-x_k\|^2_2, \end{aligned} \quad (27)$$

where $\lambda^{\max}_Q$ denotes the maximum eigenvalue of Q, which equals half of the Lipschitz constant of the gradient of f. By (27) and (26) we then have that

$$f(x_{k+1}) \leq f(x_k) - (c-(\lambda^{\max}_Q - \lambda^{\min}_{Q_d}))\|x_{k+1}-x_k\|^2_2,$$

which implies that $f(x_{k+1}) \leq f(x_k)$ if if $c > \lambda^{\max}_Q - \lambda^{\min}_{Q_d}$. This monotonicity condition and (23) ensure that the scaled projected gradient iteration (22) converges to the minimum value of $\mathcal{P}$ via Proposition 3.7, p. 217 of [12], i.e., all iterates limit points are optimal. However, convergence in iterates cannot be shown; see [40] for further elaboration. Note that the particular step-size for which convergence is ensured was not quantified in Proposition 3.7, p. 217 of [12].

1) Comparison with the scaled projected gradient algorithm and the result of [24]: In the scaled projected gradient algorithm it was shown that the step-size should be chosen so that $c > \lambda^{\max}_Q - \lambda^{\min}_{Q_d}$. Instead, Theorem 1 requires $c > \lambda^{\max}_{Q_z}$. The latter is a less restrictive condition since $\lambda^{\max}_Q - \lambda^{\min}_{Q_d} \geq \lambda^{\max}_{Q_z}$. Indeed, let v be the eigenvector corresponding to the eigenvalue $\lambda^{\max}_{Q_z}$. Then, we have that $\lambda^{\max}_{Q_z}v^\top v = v^\top Q_z v = v^\top(Q-Q_d)v \leq v^\top Qv - \lambda^{\min}_{Q_d}v^\top v$, where the first equality follows from the fact that v is the eigenvector corresponding to $\lambda^{\max}_{Q_z}$ and the inequality follows from $Q_d \succeq \lambda^{\min}_{Q_d}I$. This implies that

$$\begin{aligned} \lambda^{\max}_{Q_z} \leq \frac{v^\top Qv}{v^\top v} - \lambda^{\min}_{Q_d} \leq \max_{z\neq 0}\frac{z^\top Qz}{z^\top z} - \lambda^{\min}_{Q_d} \\ = \lambda^{\max}_Q - \lambda^{\min}_{Q_d}, \end{aligned} \quad (28)$$

where the last equality follows recalling the definition of the induced 2-norm of a symmetric square matrix.

By (28), it follows that $\lambda^{\max}_{Q_z} \leq \lambda^{\max}_Q$, thus rendering $c > \lambda^{\max}_{Q_z}$ less restrictive than the condition of Theorem 3 in [24] (which is in turn the same with unscaled projected gradient methods) as well (see also Remark 3). For a qualitative comparison we refer to Figure 2, which is a result of the fact that $\lambda^{\max}_{Q_z} \leq \lambda^{\max}_Q - \lambda^{\min}_{Q_d} \leq \lambda^{\max}_Q$.

2) Dependence of the regularization coefficient c on the number of agents: The condition on c imposed by Theorem 1 can be no more restrictive compared to the one derived in Section III-C for an appropriate scaled projected gradient algorithm, as well as the one of [24]. In particular, the higher c the more ill-conditioned from a numerical point of view the resulting optimization program tends to be. In this section, we provide a numerical investigation to quantify the relative gain $\frac{(\lambda^{\max}_Q-\lambda^{\min}_{Q_d})-\lambda^{\max}_{Q_z}}{(\lambda^{\max}_Q-\lambda^{\min}_{Q_d})}$. To achieve this, we considered matrices $Q \succeq 0$ with blocks of dimension 2, 4 and 8, and elements randomly chosen from a uniform distribution with support $[0,1]$. For different dimensions of Q we averaged the relative gain across 50 different random extractions of Q for each block structure. By inspection of Figure 2 the difference is always positive and, despite decreasing with the number of

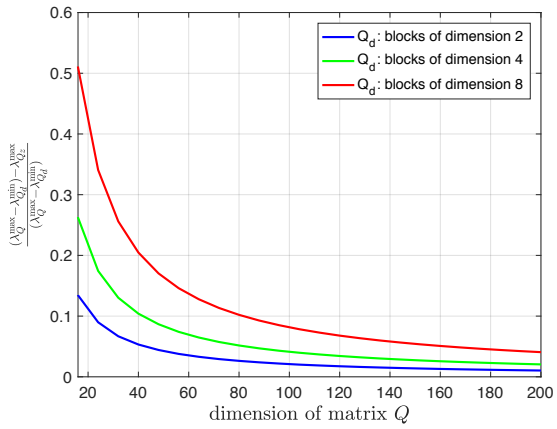

Fig. 2: Relative difference $\frac{(\lambda_Q^{\max}-\lambda_{Q_d}^{\min})-\lambda_{Q_z}^{\max}}{(\lambda_Q^{\max}-\lambda_{Q_d}^{\min})}$ for random matrices $Q \succeq 0$ of different dimension and block structure, averaged across 50 different extractions for each case.

agents, it can be quite significant in particular for problems with blocks of high dimension.

Comparing with the condition $c > \lambda_Q^{\max}$ of [24], this difference would have been more pronounced. In that case, the more the diagonal part of Q is dominant with respect to the off diagonal part, the more the difference between $\lambda_{Q_z}^{\max}$ and $\lambda_Q^{\max}$ becomes significant. For example, for $Q = \mathbf{1}_{m\times m} + mI_{m\times m}$, $\lambda_{Q_z}^{\max} = m-1$, $\lambda_Q^{\max} - \lambda_{Q_d}^{\min} = m$, while $\lambda_Q^{\max} = 2m$.

IV. Optimal Charging of Electric Vehicles Revisited

We revisit problem (2) which can be solved via Algorithm 1. Note that the objective function in (2) is not strictly convex as $A^\top PA = \mathbf{1}_{m\times m} \otimes P$, and it exhibits a structure that allows for reduced information exchange as described in Remark 1. Indeed, at iteration $k+1$ of Algorithm 1, the central authority needs to collect the solution of each agent but it only has to broadcast $\bar{x}_k = d + Ax_k$. Each agent i can then compute its objective as $f(z^i, x_k^{-i}) = (\bar{x}_k - x_k^i + z^i)^\top P(\bar{x}_k - x_k^i + z^i)$. Step 6 in Algorithm 1 for problem (2) reduces then to

$$x_{k+1}^i = \widetilde{T}^i(x_k) = \arg\min_{z^i \in X^i} \left\{(\bar{x}_k - x_k^i + z^i)^\top P(\bar{x}_k - x_k^i + z^i) + c\|z^i - x_k^i\|^2\right\}.$$

We consider $m = 100$, $T = 24$, and $\gamma^i \in [0.1, 0.3]$, $i = 1,\dots,m$. The bounds on $x^i(t)$ are taken to be $\underline{x}^i(t) = 0$ and $\overline{x}^i(t) = 0.02$, for all $i = 1,\dots,m$, $t = 0,\dots,T$. The non-PEV demand profile is retrieved from [3], whereas the price coefficient is $p(t) = 0.15$, $t = 0,\dots,T$. Note that, as in [5], $x^i(t)$ corresponds to normalized charging rate, which is then rescaled to be turned into reasonable power values. All optimization problems are solved using CPLEX [41].

Problem (2) is solved first in a centralized fashion, achieving an optimal value $f^\star = 2.67$. It is then solved by means of Algorithm 1, setting $c = 0.1485$ (as required by Theorem 1, since $\lambda_{Q_z}^{\max} = 0.1485$). After 30 iterations $f(x_{30}) - f^\star = 1.95 \cdot 10^{-6}$, thus achieving numerical convergence. The left panel of Figure 3 shows that PEV demand is optimized so that the over-night valley of the non-PEV demand is nearly

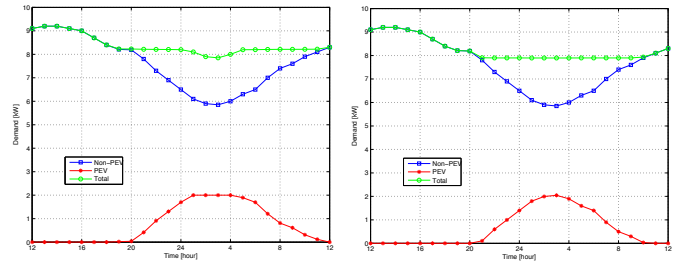

Fig. 3: Demand along a one day time horizon: non-PEV demand (blue squares), PEV demand computed via Algorithm 1 (red stars), and total demand (green circles). Left panel: $m = 100$ agents; Right panel: $m = 1000$ agents.

c	0	0.05	0.075	0.1	0.1478	0.2	0.4
k	-	-	10	16	27	37	77

TABLE I: Number of iterations k for $\frac{f(x_k)-f^\star}{f^\star} < 10^{-6}$, for different values of c.

filled-up. Note that due to the constraints in (2), there is still a small "dip" over hours 1 and 4.

Considering the same setting, we perform a parametric analysis, running Algorithm 1 for different values of c. In Table I the number of iterations needed to achieve a relative error $\frac{f(x_k)-f^\star}{f^\star} < 10^{-6}$ is reported. It can be observed that, as c increases, numerical convergence requires more iterations. Choosing $c > 0.1485$ the condition of Theorem 1 is satisfied, and Algorithm 1 always converges. This is also the case for some lower values of c, i.e., whenever $c > 0.1478$, as justified by Remark 3. However, for even lower values of c convergence is not guaranteed by our analysis (see table entries for $c = 0$ and $c = 0.05$), while for some of them numerical convergence is achieved (see table entries for $c = 0.075$ and $c = 0.1$).

We consider now $m = 1000$, a charging level such that $\gamma^i \in [0.005, 0.025]$, for all $i = 1,\dots,m$, and bounds on $x^i(t)$ to $\underline{x}^i(t) = 0$ and $\overline{x}^i(t) = 0.0025$, for all $i = 1,\dots,m$, $t = 0,\dots,T$. All the other parameters are left unchanged with respect to the previous set-up. After 30 iterations $f(x_{30}) - f^\star = 8.18 \cdot 10^{-7}$, thus achieving numerical convergence. However, as shown in the right panel of Figure 3, the peak of the PEV demand now perfectly fills the over-night valley of the non-PEV demand, preventing the small "dip" at the upper panel of Figure 3. As shown in Figure 4, our algorithm converges to a minimizer of (1), as opposed to the algorithm in [4] that ensures convergence only to the optimal value.

V. Concluding Remarks

In this paper, we investigated convergence of a decentralized, regularized Jacobi algorithm for multi-agent, convex quadratic optimization programs with a common objective and subject to separable constraints. We extended the convergence results of [24] to show that iterates converge to some optimizer of the centralized problem counterpart instead of convergence in optimal value. We show that this occurs under milder conditions than [24], and scaled projected gradient algorithms for which we derive a convergence condition. The efficacy of the proposed algorithm was illustrated by applying it to the problem of optimal charging of electric vehicles, achieving convergence with a finite number of vehicles.

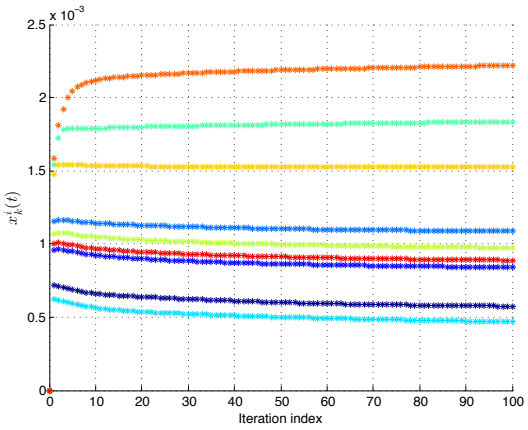

Fig. 4: Evolution of the iterates $x_k^i(t)$ generated by Algorithm 1 at $t = 12$ as a function of the iteration index k, for $i = 1, \ldots, 10$, i.e., the first 10 vehicles of the 1000-vehicle fleet.

Current work concentrates on exploiting the results of [24] and the recent investigations of [31] to provide convergence rate conditions for our setting and quantify the effect of the regularization coefficient on the convergence rate. At the same time, we aim at exploiting the results of [42] to account for affine coupling constraints.

References

[1] F. Bagagiolo and D. Bauso, "Mean-field games and dynamic demand management in power grids," *Dyn. Games & Appl.*, vol. 4, no. 2, pp. 155–176, 2014.

[2] B. Gharesifard, T. Basar, and A. Dominguez-Garcia, "Price-based coordinated aggregation of networked distributed energy resources," *IEEE Trans. Aut. Con.*, vol. 61, no. 10, pp. 2936–2946, Oct 2016.

[3] Z. Ma, D. Callaway, and I. Hiskens, "Decentralized charging control of large populations of plug-in electric vehicles," *IEEE Trans. Con. Sys. Tech.*, vol. 21, no. 1, pp. 67–78, 2013.

[4] L. Gan, U. Topcu, and S. Low, "Optimal Decentralized Protocol for Electric Vehicle Charging," *IEEE Trans. Power Sys.*, vol. 28, no. 2, pp. 940 – 951, 2013.

[5] F. Parise, M. Colombino, S. Grammatico, and J. Lygeros, "Mean field constrained charging policy for large populations of plug-in electric vehicles," *IEEE Conf. on Dec. Con*, pp. 5101–5106, 2014.

[6] M. Pavone, E. Frazzoli, and F. Bullo, "Decentralized Algorithms for Stochastic and Dynamic Vehicle Routing with General Demand Distribution," *IEEE Conf. Dec. Con.*, pp. 4869 – 4874, 2007.

[7] G. Banjac, B. Stellato, N. Moehle, P. Goulart, A. Bemporad, and S. Boyd, "Embedded code generation using the OSQP solver," *IEEE Conf. Dec. Con., to appear*, pp. 1–6, 2017.

[8] A. Nedic, A. Ozdaglar, and P. Parrilo, "Constrained consensus and optimization in multi-agent networks," *IEEE Trans. Aut. Con.*, vol. 55, no. 4, pp. 922–938, 2010.

[9] M. Zhu and S. Martinez, "On distributed convex optimization under inequality and equality constraints," *IEEE Trans. Aut. Con.*, vol. 57, no. 1, pp. 151–164, 2012.

[10] J. Koshal, A. Nedich, and U. Shanbhag, "Multiuser optimization: distributed algorithms and error analysis," *SIAM J. Opt.*, vol. 21, no. 3, pp. 1046–1081, 2011.

[11] ——, "A gossip algorithm for aggregative games on graphs," *IEEE Conf. Dec. Con.*, pp. 4840 – 4845, 2012.

[12] D. Bertsekas and J. Tsitsiklis, *Parallel and distributed computation: Numerical methods*. Athena Scientific, 1989.

[13] N. Parikh and S. Boyd, "Proximal algorithms," *Foundations and Trends in Optimization*, vol. 1, no. 3, pp. 123 – 231, 2013.

[14] A. Goldstein, "Convex programming in Hilbert space," *Bull. Amer. Math. Soc.*, vol. 70, pp. 709–710, 1964.

[15] E. Levitin and B. Poljak, "Costrained minimization methods," *USSR Comp. Math. & Phys. (English translation)*, vol. 1, pp. 1–50, 1965.

[16] D. Bertsekas, "On the Goldstein-Levitin-Polyak gradient projection method," *IEEE Trans. Aut. Con.*, vol. 2, pp. 174–184, 1976.

[17] H. Attouch, J. Bolte, and B. Svaiter, "Convergence of descent methods for semi-algebraic and tame problems: proximal algorithms, forward-backward splitting, and regularized GaussSeidel methods," *Mathematical Programming*, vol. 127, no. 1, p. 91129, 2013.

[18] G. Cohen, "Optimization by decomposition and coordination: A unified approach," *IEEE Trans. Aut. Con.*, vol. 23, no. 2, pp. 222–232, 1978.

[19] ——, "Auxiliary problem principle and decomposition of optimization problems," *J. Opt. Theory & Appl.*, vol. 32, no. 3, pp. 277–305, 1980.

[20] M. Patriksson, "Decomposition methods for differentiable optimization problems over cartesian product sets," *Comp. Opt. & Appl.*, vol. 9, no. 1, pp. 5–42, 1998.

[21] D. Zhu and P. Marcotte, "Coupling the auxiliary problem principle with descent methods of pseudoconvex programming," *Eur. J. of Oper. Res.*, vol. 83, no. 3, pp. 670–685, 1995.

[22] P. Richtárik and M. Takáč, "Parallel coordinate descent methods for big data optimization," *Math. Prog.*, vol. 156, no. 1-2, pp. 433–484, 2016.

[23] I. Necoara and D. Clipici, "Parallel random coordinate descent method for composite minimization: Convergence analysis and error bounds," *SIAM J. Opt.*, vol. 26, no. 1, pp. 197–226, 2016.

[24] G. Scutari, F. Facchinei, P. Song, D. Palomar, and J. Pang, "Decomposition by partial linearization: Parallel optimization of multi-agent systems," *IEEE Trans. Sig. Proc.*, vol. 62, no. 3, pp. 641–656, 2014.

[25] F. Facchinei, G. Scutari, and S. Sagratella, "Parallel selective algorithms for nonconvex big data optimization," *IEEE Trans. Sig. Proc.*, vol. 63, no. 7, pp. 1874–1889, 2015.

[26] S. Grammatico, F. Parise, M. Colombino, and J. Lygeros, "Decentralized convergence to nash equilibria in constrained deterministic mean field control," *IEEE Trans. Aut. Con.*, vol. 61, no. 11, pp. 3315–3329, Nov 2016.

[27] G. Scutari, F. Facchinei, J. Pang, and D. Palomar, "Real and complex monotone communication games," *IEEE Trans. Inf. Theory*, vol. 60, no. 7, pp. 4197–4231, 2014.

[28] D. Paccagnan, M. Kamgarpour, and J. Lygeros, "On Aggregative and Mean Field Games with Applications to Electricity Markets," *Eur. Con. Conf.*, pp. 196 – 201, 2016.

[29] Z. Luo and P. Tseng, "On the linear convergence of descent methods for convex essentially smooth minimization," *SIAM J. Con. & Opt.*, vol. 30, no. 2, pp. 408–425, 1992.

[30] A. Nedich, "Gradient Methods for Constrained Optimization," *Lecture Notes, UIUC*, 2008.

[31] G. Banjac, K. Margellos, and P. J. Goulart, "On the convergence of a regularized jacobi algorithm for convex optimization," *IEEE Trans. Aut. Con., to appear*, pp. 1–7, 2017.

[32] L. Deori, K. Margellos, and M. Prandini, "On decentralized convex optimization in a multi-agent setting with separable constraints and its application to optimal charging of electric vehicles," in *IEEE Conf. Dec. Con.*, Dec 2016, pp. 6044–6049.

[33] ——, "Regularized Jacobi iteration for decentralized convex optimization with separable constraints," *Technical Report*, pp. 1–11, 2017. [Online]. Available: https://arxiv.org/pdf/1604.07814.pdf

[34] T. Rockafellar and R. Wets, *Variational Analysis*. Springer-Verlag Berlin Heidelberg, 1998.

[35] F. Facchinei, A. Fischer, and V. Piccialli, "On generalized Nash games and variational inequalities," *Oper. Res. Let.*, vol. 35, no. 2, pp. 159–164, 2007.

[36] V. Berinde, *Iterative Approximation of Fixed Points*. Springer, 2007.

[37] P. Combettes and T. Pennanen, "Generalized Mann iterates for constructing fixed points in Hilbert spaces," *J. Math. Anal. & Appl.*, vol. 275, no. 2, p. 521536, 2002.

[38] Z. Opial, "Weak convergence of the sequence of successive approximations for nonexpansive mappings," *Bull. Amer. Math. Soc.*, vol. 73, no. 4, pp. 591–597, 1967.

[39] H. Bauschke and P. Combettes, *Convex analysis and monotone operator theory in Hilbert spaces*. Springer, 2010.

[40] S. Bonettini and M. Prato, "New convergence results for the scaled gradient projection method," *Inverse Problems*, vol. 31, no. 9, pp. 408–425, 2015.

[41] "IBM ILOG CPLEX Optimizer," 2015. [Online]. Available: http://www-01.ibm.com/software/commerce/optimization/cplex-optimizer/index.html

[42] D. Paccagnan, B. Gentile, F. Parise, M. Kamgarpour, and J. Lygeros, "Distributed computation of generalized Nash equilibria in quadratic aggregative games with affine coupling constraints," *IEEE Conf. Dec. Con.*, pp. 6123–6128, 2016.