Decentralized non-convex optimization via bi-level SQP and ADMM

Gösta Stomberg, Alexander Engelmann and Timm Faulwasser

Abstract— **Decentralized non-convex optimization is important in many problems of practical relevance. Existing decentralized methods, however, typically either lack convergence guarantees for general non-convex problems, or they suffer from a high subproblem complexity. We present a novel bi-level SQP method, where the inner quadratic problems are solved via ADMM. A decentralized stopping criterion from inexact Newton methods allows the early termination of ADMM as an inner algorithm to improve computational efficiency. The method has local convergence guarantees for non-convex problems. Moreover, it only solves sequences of Quadratic Programs, whereas many existing algorithms solve sequences of Nonlinear Programs. The method shows competitive numerical performance for an optimal power flow problem.**

I. Introduction

Decentralized optimization methods, i.e. methods which solve optimization problems purely based on neighbor-to-neighbor communication, are of interest in many applications such as Optimal Power Flow (OPF) and distributed Model Predictive Control (MPC) [1, 2]. In many cases, these applications require to solve optimization problems with non-convex objectives and non-convex constraints. At large, existing algorithms can be categorized as follows: a) they are decentralized, but lack guarantees for problems with non-convex constraints; b) they are not decentralized, i.e. they require central communication/coordination; or c) they solve non-convex Nonlinear Programs (NLP)s at a subsystem level, which increases complexity and impedes convergence guarantees.

The Alternating Direction Method of Multipliers (ADMM) is a decentralized method, which shows promising performance for a large variety of problems. ADMM is guaranteed to converge only for special classes of non-convex problems [3]. Decentralized schemes with global convergence guarantees based on augmented Lagrangian methods are presented in [4, 5]. Whereas [4] assumes polynomial objectives and equality constraints, [5] allows for more general constraints. A distributed method with local convergence guarantees for non-convex probelms is the Augmented Lagrangian Alternating Direction Inexact Newton (ALADIN) method [6]. Bi-level ALADIN variants employ inner algorithms to decompose the coordination step of ALADIN [7]. All the above methods require solving constrained non-convex NLPs to optimality, which can often not be guaranteed by numerical solvers for an arbitrary initialization [8].[1]

The authors are with the Institute for Energy Systems, Energy Efficiency and Energy Economics, TU Dortmund University, Dortmund, Germany. goesta.stomberg@tu-dortmund.de, {alexander.engelmann, timm.faulwasser}@ieee.org

[1]Most solvers are guaranteed to converge to a stationary point of a merit function, which is not necessarily a minimizer, cf. e.g. [8, Thm. 19.2].

An essentially decentralized interior point method with local convergence guarantees for general non-convex problems is proposed in [9]. Therein, the solution of NLPs by subsystems is avoided. However, the step-size selection requires scalar global communication. Tailored algorithms for distributed NMPC can be found e.g. in [2, 10]. These algorithms also suffer from either of the above drawbacks.

For centralized non-convex optimization, Sequential Quadratic Programming (SQP) has local or global convergence guarantees [8]. Bi-level algorithms with SQP on the outer level and ADMM as an inner algorithm are presented in [11, 12]. The scheme of [11] has global convergence guarantees for problems with non-convex objective and linearly coupled subproblems. However, it does not allow for local constraints or inequality constraints. In [12], non-convex objectives and non-convex local constraints are considered, but no convergence guarantees are established. Both approaches assume that ADMM solves the subproblem Quadratic Programs (QP)s to optimality in each SQP step, which is unrealistic in many applications. A further bi-level approach with ADMM as an inner method is presented in [13] for a Sequential Convex Programming-based distributed MPC scheme.

The focus of this work is a bi-level decentralized SQP method (d-SQP) with ADMM as inner algorithm. As the proposed scheme combines SQP with inner decentralized ADMM we obtain a fully decentralized algorithm. This paper presents two contributions. First, in contrast to the bi-level SQP method in [11], we establish convergence for nonlinear programs with non-convex objective *and* non-convex constraints. Second, we derive a novel stopping criterion for the inner ADMM iterations that guarantees local convergence of d-SQP despite inexact QP solutions. This is crucial as it may not be computationally feasible to obtain an exact subproblem solution via ADMM in each SQP step.

The paper is organized as follows: Section II introduces the problem formulation, SQP schemes, and ADMM. Section III derives d-SQP and establishes local convergence. Section IV analyses numerical results from OPF.

Notation: Given a matrix A and an integer j, $[A]_j$ denotes the jth row of A. For an index set $\mathcal{A}$, $[A]_{\mathcal{A}}$ denotes the matrix consisting of rows $[A]_j$ for all $j \in \mathcal{A}$. Likewise, $[a]_j$ is the jth component of vector a and $a_{\mathcal{A}}$ is the vector of components $[a]_j$ for all $j \in \mathcal{A}$. The concatenation of vectors x and y into a column vector is $\mathrm{col}(x, y)$. The ε-neighborhood around a point x is denoted by $\mathcal{B}_\varepsilon(x)$, i.e., $\mathcal{B}_\varepsilon(x) \doteq \{y \in \mathbb{R}^{n_x} \mid \|y - x\| \leq \varepsilon\}$, where $\|\cdot\|$ is some norm on $\mathbb{R}^n$. The condition number of a matrix $A \in \mathbb{R}^{n \times n}$ is defined as $\mathrm{cond}(A) \doteq \|A\| \|A^{-1}\|$. I is the identity matrix.

II. PROBLEM STATEMENT AND PRELIMINARIES

We consider Nonlinear Programs of the form

$$\min_{x_1,\ldots,x_S} \sum_{i\in\mathcal{S}} f_i(x_i) \tag{1a}$$

$$\text{subject to} \quad g_i(x_i) = 0 \mid \nu_i, \qquad \forall i \in \mathcal{S}, \tag{1b}$$

$$h_i(x_i) \leq 0 \mid \mu_i, \qquad \forall i \in \mathcal{S}, \tag{1c}$$

$$\sum_{i\in\mathcal{S}} E_i x_i = c \mid \lambda, \tag{1d}$$

where $\mathcal{S} = \{1,\ldots,S\}$ is a set of subsystems, each of which is equipped with $x_i \in \mathbb{R}^{n_i}$ and twice continuously differentiable functions $f_i : \mathbb{R}^{n_i} \to \mathbb{R}$, $g_i : \mathbb{R}^{n_i} \to \mathbb{R}^{n_{gi}}$, and $h_i : \mathbb{R}^{n_i} \to \mathbb{R}^{n_{hi}}$. The matrices $E_i \in \mathbb{R}^{n_c \times n_i}$ and the vector $c \in \mathbb{R}^{n_c}$ couple the subsystems. The notation in (1) highlights that $\nu_i \in \mathbb{R}^{n_{gi}}$, $\mu_i \in \mathbb{R}^{n_{hi}}$, and $\lambda \in \mathbb{R}^{n_c}$ are Lagrange multipliers associated with the respective constraints. The centralized variables are $x \doteq \mathrm{col}(x_i)_{i\in\mathcal{S}}$, $\nu \doteq \mathrm{col}(\nu_i)_{i\in\mathcal{S}}$, and $\mu \doteq \mathrm{col}(\mu_i)_{i\in\mathcal{S}}$. Define the Lagrangian of (1),

$$L(x,\nu,\mu,\lambda) = \left(\sum_{i\in\mathcal{S}} L_i(x_i,\nu_i,\mu_i,\lambda) \right) - \lambda^\top c,$$

where $L_i(\cdot) = f_i(x_i) + \nu_i^\top g_i(x_i) + \mu_i^\top h_i(x_i) + \lambda^\top E_i x_i$. Then, the Karush-Kuhn-Tucker (KKT) conditions of (1) read

$$\nabla f_i(x_i) + \nabla g_i(x_i)\nu_i + \nabla h_i(x_i)\mu_i + E_i^\top \lambda = 0, \quad i \in \mathcal{S}, \tag{2a}$$

$$g_i(x_i) = 0, \quad i \in \mathcal{S}, \tag{2b}$$

$$h_i(x_i) \leq 0, \ \mu_i \geq 0, \ \mu_i^\top h_i(x_i) = 0, \quad i \in \mathcal{S}, \tag{2c}$$

$$\sum_{i\in\mathcal{S}} E_i x_i = c. \tag{2d}$$

A. Sequential quadratic programming

SQP methods repeatedly solve quadratic approximations of (1), cf. [8, Ch. 18]. A quadratic approximation of (1) at a primal-dual iterate $(x^k, \nu^k, \mu^k, \lambda^k)$ is

$$\min_{s_1,\ldots,s_S} \sum_{i\in\mathcal{S}} \frac{1}{2} s_i^\top H_i^k s_i + \nabla f_i(x_i^k)^\top s_i \tag{3a}$$

subject to

$$g_i(x_i^k) + \nabla g_i(x_i^k)^\top s_i = 0 \mid \nu_i^{QP}, \quad \forall i \in \mathcal{S}, \tag{3b}$$

$$h_i(x_i^k) + \nabla h_i(x_i^k)^\top s_i \leq 0 \mid \mu_i^{QP}, \quad \forall i \in \mathcal{S}, \tag{3c}$$

$$\sum_{i\in\mathcal{S}} E_i(x_i^k + s_i) = c \mid \lambda^{QP}, \tag{3d}$$

where $H_i^k \doteq \nabla^2_{x_i x_i} L_i(x_i^k, \nu_i^k, \mu_i^k, \lambda^k)$. Observe that the KKT system (2) can be written as $F(x,\nu,\mu,\lambda) = 0$, where

$$F(x,\nu,\mu,\lambda) \doteq \begin{bmatrix} \nabla_{x_1} L_1(x_1,\nu_1,\mu_1,\lambda) \\ g_1(x_1) \\ \min(-h_1(x_1),\mu_1) \\ \vdots \\ \nabla_{x_S} L_S(x_S,\nu_S,\mu_S,\lambda) \\ g_S(x_S) \\ \min(-h_S(x_S),\mu_S) \\ \left(\sum_{i\in\mathcal{S}} E_i x_i\right) - c \end{bmatrix}. \tag{4}$$

The block rows $\min(-h_i(x_i),\mu_i)$ represent (2c). Algorithm 1 summarizes an SQP method for solving (1).

Algorithm 1 Inequality-constrained SQP for solving (1)

1: Initialization: x_i^0, ν_i^0, μ_i^0 for all $i \in \mathcal{S}$, λ^0, ϵ
2: **while** $\|F^k\| \not\leq \epsilon$ **do**
3:      compute $\nabla f_i(x_i^k), \nabla g_i(x_i^k), \nabla h_i(x_i^k), H_i^k, \ i \in \mathcal{S}$
4:      solve QP (3)
5:      $x_i^{k+1} = x_i^k + s_i^k, \ \nu_i^{k+1} = \nu_i^{QP}, \ \mu_i^{k+1} = \mu_i^{QP}, \lambda^{k+1} = \lambda^{QP}$
6: **end while**
7: **return** $x^\star$

Define $p \doteq \mathrm{col}(x,\nu,\mu,\lambda)$, let $p^\star \doteq \mathrm{col}(x^\star,\nu^\star,\mu^\star,\lambda^\star)$ be a KKT point of (1), and let $\mathcal{A}_i$ and $\mathcal{I}_i$ denote the sets of active and inactive inequality constraints at $x_i^\star$ respectively, i.e.,

$$\mathcal{A}_i \doteq \{j \in \{1,\ldots,n_{hi}\} | [h_i(x_i^\star)]_j = 0\},$$
$$\mathcal{I}_i \doteq \{j \in \{1,\ldots,n_{hi}\} | [h_i(x_i^\star)]_j < 0\}.$$

Assumption 1: The point $p^\star$ is a KKT point of (1) which, for all $i \in \mathcal{S}$, satisfies

i) $h_i(x_i^\star) + \mu_i^\star \neq 0$ (strict complementarity),
ii) $s_i^\top \nabla^2_{x_i x_i} L_i(x_i^\star, \nu_i^\star, \mu_i^\star, \lambda^\star) s_i > 0$ for all $s_i \neq 0$ with $\nabla g_i(x_i^\star)^\top s_i = 0.^2$

Furthermore, the matrix

$$\begin{bmatrix} \nabla_{x_1} g_1(x_1^\star)^\top & \cdots & \nabla_{x_S} g_S(x_S^\star)^\top \\ \nabla_{x_1}[h_1(x_1^\star)]_{\mathcal{A}_1}^\top & \cdots & \nabla_{x_S}[h_S(x_S^\star)]_{\mathcal{A}_S}^\top \\ E_1 & \cdots & E_S \end{bmatrix}$$

has full row rank, i.e., it satisfies the Linear Independence Constraint Qualification (LICQ). $\qquad\square$

Definition 1 (Convergence rates): We say that the sequence $\{p^k\} \subset \mathbb{R}^{n_p}$ converges to $p^\star \in \mathbb{R}^{n_p}$

i)   q-linearly, if $\|p^{k+1} - p^\star\| \leq c\|p^k - p^\star\| \quad \forall k \geq k_0$ for some $0 < c < 1$ and $k_0 \geq 0$.
ii)   q-{superlinearly, quadratically}, if $p^k \to p^\star$ and $\|p^{k+1} - p^\star\| = \{o(\|p^k - p^\star\|), O(\|p^k - p^\star\|^2)\}$ for $k \to \infty$.

Theorem 1 (Local convergence of SQP, [14, Thm. 5.31]): Let Assumption 1 hold. Then, there exists $\varepsilon_1 > 0$ such that for all $p^0 \in \mathcal{B}_{\varepsilon_1}(p^\star)$ the following holds.

i)   The sequence $\{p^k\}$ generated by Algorithm 1 converges q-superlinearly to $p^\star$.
ii)   If $\nabla^2_{x_i x_i} f_i, \nabla^2_{x_i x_i}[g_i]_j, j \in \{1,\ldots,n_{gi}\}$, and $\nabla^2_{x_i x_i}[h_i]_j, j \in \{1,\ldots,n_{hi}\}$ are locally Lipschitz continuous for all $i \in \mathcal{S}$, then the convergence rate is q-quadratic. $\qquad\square$

The proof of Theorem 1 given in [14] shows that inside $\mathcal{B}_{\varepsilon_1}(p^\star)$, the iterates generated by Algorithm 1 correspond to the iterates generated by Newton's method applied to the nonlinear system of equations $F(p) = 0$. In particular, for all $p^k \in \mathcal{B}_{\varepsilon_1}(p^\star)$, $F^k \doteq F(p^k)$ is continuously differentiable and $\nabla F^k \doteq \nabla F(p^k)$ is regular. Hence the Newton iteration $p^{k+1} = p^k + d$ with $F^k + \nabla F^k d = 0$, where

$$d = (\mathrm{col}(d_i)_{i\in\mathcal{S}}, \Delta\lambda), \tag{5}$$

$d_i \doteq \mathrm{col}(s_i, \Delta\nu_i, \Delta\mu_i)$, $\Delta\nu_i \doteq \nu_i^{QP} - \nu_i^k$, $\Delta\mu_i \doteq \mu_i^{QP} - \mu_i^k$, and $\Delta\lambda \doteq \lambda^{QP} - \lambda^k$ is well-defined inside $\mathcal{B}_{\varepsilon_1}$.

---

[2] This is a slightly stronger assumption than the Second-Order Sufficient Condition as we exclude the conditions $[\nabla h_i(x_i^\star)]_{\mathcal{A}_i}^\top s_i = 0$ and $E_i s_i = 0$.

Algorithm 2 ADMM for solving (6)

1: Initialization: $\bar{s}_i^0, \gamma_i^0$ for all $i \in \mathcal{S}$
2: **while** not converged **do**
3: $\quad s_i^{l+1} = \arg\min_{s_i \in \mathbb{S}_i^k} L_{\rho,i}^k(s_i, \bar{s}_i^l, \gamma_i^l)$ for all $i \in \mathcal{S}$
4: $\quad \bar{s}^{l+1} = \arg\min_{\bar{s} \in \mathbb{E}} \sum_{i \in \mathcal{S}} L_{\rho,i}^k(s_i^{l+1}, \bar{s}_i, \gamma_i^l)$
5: $\quad \gamma_i^{l+1} = \gamma_i^l + \rho(s_i^{l+1} - \bar{s}_i^{l+1})$ for all $i \in \mathcal{S}$
6: **end while**
7: **return** $(\bar{s}_i^l, \nu_i^{l,QP}, \mu_i^{l,QP}, \gamma_i^l)$ **for all** $i \in \mathcal{S}$

B. Alternating direction method of multipliers

We next recall how to solve (3) via ADMM. First, we reforumlate (3) as

$$\min_{s_1 \in \mathbb{S}_1^k, \bar{s}_1, \ldots, s_S \in \mathbb{S}_s^k, \bar{s}_S} \sum_{i \in \mathcal{S}} \phi_i^k(s_i) \tag{6a}$$

$$\text{subject to} \quad s_i - \bar{s}_i = 0 \mid \gamma_i \quad, \quad \forall i \in \mathcal{S}, \tag{6b}$$

$$\sum_{i \in \mathcal{S}} E_i(x_i^k + \bar{s}_i) = c \mid \lambda^{QP}, \tag{6c}$$

where $\bar{s}_i \in \mathbb{R}^{n_i}$ is an auxiliary decision variable for each subsystem $i \in \mathcal{S}$, $\phi_i^k(s_i) \doteq 1/2 s_i^\top H_i^k s_i + \nabla f_i(x_i^k)^\top s_i$ and

$$\mathbb{S}_i^k \doteq \left\{ s_i \in \mathbb{R}^{n_i} \,\middle|\, \begin{array}{l} g_i(x_i^k) + \nabla g_i(x_i^k)^\top s_i = 0 \\ h_i(x_i^k) + \nabla h_i(x_i^k)^\top s_i \leq 0 \end{array} \right\}.$$

Let $\bar{s} \doteq \text{col}(\bar{s}_i)_{i \in \mathcal{S}}$, $s \doteq \text{col}(s_i)_{i \in \mathcal{S}}$, and $\gamma \doteq \text{col}(\gamma_i)_{i \in \mathcal{S}}$. The augmented Lagrangian of (6) w.r.t. (6b) reads

$$L_\rho^k(s,\bar{s},\gamma) = \sum_{i \in \mathcal{S}} L_{\rho,i}^k(s_i,\bar{s}_i,\gamma_i) = \sum_{i \in \mathcal{S}} \phi_i^k(s_i) + \gamma_i^\top(s_i - \bar{s}_i) + \frac{\rho}{2}\|s_i - \bar{s}_i\|_2^2.$$

Denote the feasible set of the coupling constraints by $\mathbb{E} = \{\bar{s} \in \mathbb{R}^n \mid \sum_{i \in \mathcal{S}} E_i(x_i^k + \bar{s}_i) = c\}$. Algorithm 2 summarizes ADMM, where the dual variables $\nu_i^{l,QP}$ and $\mu_i^{l,QP}$ are obtained in Step 3. Step 3 can be executed by each subsystem in parallel by solving the local QP

$$\min_{s_i} \frac{1}{2} s_i^\top (H_i^k + \rho I)s_i + (\nabla f_i(x_i^k) + \gamma_i^l - \rho\bar{s}_i^l)^\top s_i \tag{7a}$$

subject to

$$g_i(x_i^k) + \nabla g_i(x_i^k)^\top s_i = 0 \mid \nu_i^{l,QP}, \tag{7b}$$

$$h_i(x_i^k) + \nabla h_i(x_i^k)^\top s_i \leq 0 \mid \mu_i^{l,QP}, \tag{7c}$$

where the iterates $\bar{s}_i^l$ and γ_i^l serve as parameters for the local QP of subsystem $i \in \mathcal{S}$.

Remark 1 (Decentralized ADMM): Step 4 in Algorithm 2 solves a QP subject to the coupling constraints and in general requires centralized computation. However, if (1) is given in so-called consensus form, and if γ_i^0 is chosen appropriately, then Step 4 is equivalent to a decentralized averaging step, i.e., it only requires neighbor-to-neighbor communication and local computation [15, Ch. 7]. A common setup where Problem (1) is given in consensus form occurs if the constraints (1d) couple original and copied variables between neighboring subsystems, cf. [16]. Then, if the SQP initialization satisfies $\sum_{i \in \mathcal{S}} E_i x_i^0 = 0$, we may replace Step 4 by a decentralized averaging step. $\quad\square$

III. DECENTRALIZED SEQUENTIAL QUADRATIC PROGRAMMING

The key idea of our approach is to solve QP (3) with ADMM. It may not be tractable to solve (3) to high precision in every SQP step. Therefore we next derive convergence results of the inexact outer SQP steps and then provide technical results for the subproblems based on well-known ADMM convergence properties.

A. Outer convergence

Theorem 1 establishes convergence if (3) is solved exactly in Step 4 of Algorithm 1. Instead, we allow inexact solutions of (3) and rely on the inexact Newton stopping criterion [17]

$$\|F^k + \nabla F^k d\| \leq \eta^k \|F^k\| \tag{8}$$

with $0 < \eta^k < 1$.

Lemma 1 (Local convergence with inexact SQP steps): Let Assumption 1 hold. Solve (3) inexactly in Step 4 of Algorithm 1, form d according to (5), and let d satisfy (8). Then there exist $\varepsilon_2 > 0$ and $\eta > 0$ such that for all $p^0 \in \mathcal{B}_{\varepsilon_2}(p^\star)$ the following holds.

i) If $\eta^k \leq \eta$, then the sequence $\{p^k\}$ generated by Algorithm 1 converges q-linearly to $p^\star$.
ii) If $\eta^k \to 0$, then the convergence rate is q-superlinear.
iii) If $\eta^k = O(\|F^k\|)$ and if $\nabla_{x_i x_i}^2 f_i, \nabla_{x_i x_i}^2 [g_i]_j, j \in \{1, \ldots, n_{gi}\}$, and $\nabla_{x_i x_i}^2 [h_i]_j, j \in \{1, \ldots, n_{hi}\}$ are Lipschitz continuous inside $\mathcal{B}_{\varepsilon_2}$, then the convergence rate is q-quadratic. $\quad\square$

Proof: We choose $\varepsilon_2 \leq \varepsilon_1$ with ε_1 from Theorem 1. Then, ∇F is regular inside $\mathcal{B}_{\varepsilon_2}(p^\star)$. To obtain statement i), we modify Theorem 2.3 from [17], which proves linear convergence in the sense that $\|\nabla F(p^\star)(p^{k+1} - p^\star)\| \leq t\|\nabla F(p^\star)(p^k - p^\star)\|$ with some t such that $\eta < t < 1$. From the second to last inequality follows $\|p^{k+1} - p^\star\| \leq \text{cond}(\nabla F(p^\star))t\|p^k - p^\star\|$ [18, Eq. 7]. Since $\nabla F(p^\star)$ is regular, $\text{cond}(\nabla F(p^\star))$ is finite and choosing t sufficiently small yields q-linear convergence with $c = \text{cond}(\nabla F(p^\star))t < 1$. Statements ii) and iii) follow directly from [17, Thm. 3.3]. $\quad\blacksquare$

The stopping criterion (8) guarantees local convergence of Algorithm 1 for iterates starting in $\mathcal{B}_{\varepsilon_2}$. However, in practice it is often not known whether p^0 lies inside $\mathcal{B}_{\varepsilon_2}$. For an initialization outside $\mathcal{B}_{\varepsilon_2}$, it may not be possible to evaluate (8) as the derivatives of the block rows $\min(-h_i(x_i^k), \mu_i^k), i \in \mathcal{S}$ in (4) are ill-defined. Hence, we develop a modified stopping criterion that is equivalent to (8) inside $\mathcal{B}_{\varepsilon_2}$, but which can also be evaluated outside $\mathcal{B}_{\varepsilon_2}$. We propose the modified stopping criterion

$$\|\tilde{F}^k + \nabla \tilde{F}^k d\| \leq \eta^k \|\tilde{F}^k\|, \tag{9}$$

with

$$\tilde{F} \doteq \begin{bmatrix} \nabla_{x_1} L_1(x_1, \nu_1, \mu_1, \lambda) \\ g_1(x_1) \\ \vdots \\ \nabla_{x_S} L_S(x_S, \nu_S, \mu_S, \lambda) \\ g_S(x_S) \\ \left(\sum_{i \in \mathcal{S}} E_i x_i\right) - c \end{bmatrix}, \nabla \tilde{F} = \begin{bmatrix} \nabla \tilde{F}_1 \cdots & 0 & \bar{E}_1^\top \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots \nabla \tilde{F}_S & \bar{E}_S^\top \\ \tilde{E}_1 & \cdots & \tilde{E}_S & 0 \end{bmatrix},$$

$$\nabla \tilde{F}_i \doteq \begin{bmatrix} \nabla_{x_i x_i}^2 L_i & \nabla g_i(x_i) & \nabla h_i(x_i) \\ \nabla g_i(x_i)^\top & 0 & 0 \end{bmatrix},$$

with $\tilde{E}_i \doteq \begin{bmatrix} E_i & 0 & 0 \end{bmatrix}$ and $\bar{E}_i \doteq \begin{bmatrix} E_i & 0 \end{bmatrix}$. Observe that $\tilde{F}$ does not include the block rows $\min(-h_i(x_i^k), \mu_i^k), i \in \mathcal{S}$ to avoid differentiability issues outside $\mathcal{B}_{\varepsilon_1}$. The subsystems can evaluate (9) locally and only communicate convergence flags, if $\sum_i E_i(x_i^k + s_i^k) = c$ and if $\|\cdot\|_\infty$ is chosen.

Lemma 2 (Modified stopping criterion): Let Assumption 1 hold, let $p^k \in \mathcal{B}_{\varepsilon_1}(p^\star)$ with ε_1 from Theorem 1 and let $(s, \nu^{QP}, \mu^{QP}, \lambda^{QP})$ be an inexact solution to (3) which has the same active set as $p^\star$, i.e.

$$[h_i(x_i^k)]_j + [\nabla h_i(x_i^k)^\top s_i]_j = 0, \ \forall j \in \mathcal{A}_i, \forall i \in \mathcal{S} \quad (10a)$$
$$[\mu_i^{QP}]_j = 0, \ \forall j \in \mathcal{I}_i, \forall i \in \mathcal{S}. \quad (10b)$$

Then, $\|\tilde{F}^k + \nabla \tilde{F}^k d\| \le \eta^k \|\tilde{F}^k\| \Longrightarrow \|F^k + \nabla F^k d\| \le \eta^k \|F^k\|.\ \square$

Proof: From (10) follows $\|\tilde{F}^k + \nabla \tilde{F}^k d\| = \|F^k + \nabla F^k d\|$. Since $\|\tilde{F}^k\| \le \|F^k\|$, we obtain the implication. ∎

B. Inner convergence

To apply the stopping criterion (9), we show that ADMM stops at the correct active set in a neighborhood of $p^\star$. We first show that the optimal active sets of (1), (6), and (7) are equivalent in a neighborhood of $p^\star$, if ADMM is initialized appropriately. Let $q \doteq \mathrm{col}(\bar{s}, \gamma)$ and $q_i \doteq \mathrm{col}(\bar{s}_i, \gamma_i)$.

Lemma 3 (Local active set equivalence of (1),(6),(7)): Let Assumption 1 hold, let $p^k \in \mathcal{B}_{\varepsilon_1}(p^\star)$, and let $(s^{k\star}, \bar{s}^{k\star}, \nu^{k\star,QP}, \mu^{k\star,QP}, \gamma^{k\star}, \lambda^{k\star,QP})$ be a KKT point of (6) formulated at p^k. Then there exists $\varepsilon_3 > 0$ such that for all $i \in \mathcal{S}$ the following holds. If $q \in \mathcal{B}_{\varepsilon_3}(q^{k\star})$, then the solution $(s_i^\star, \mu_i^{\star,QP})$ to (7) with parameters q_i has the same active set as $p^\star$, i.e.

$$[h_i(x_i^k)]_j + [\nabla h_i^j(x_i^k)^\top \tilde{s}_i^\star]_j = 0, \ \forall j \in \mathcal{A}_i, \forall i \in \mathcal{S},$$
$$[\tilde{\mu}_i^{\star,QP}]_j = 0, \ \forall j \in \mathcal{I}_i, \ \forall i \in \mathcal{S}. \ \square$$

Proof: We first show that, (a), for $p^k \in \mathcal{B}_{\varepsilon_1}(p^\star)$ problems (1), (6), and (7) have the same active set at their respective solution, if (7) is parameterized with $q^{k\star}$. We then show that, (b), this also holds for small variations in q.

(a) Assumption 1 implies that QPs (3) and (7) each have a unique KKT point. Furthermore, the KKT point of (3) has the same active set as $x^\star$ and satisfies strict complementarity for all $p^k \in \mathcal{B}_{\varepsilon_1}(p^\star)$ [14, Thm. 5.31]. Comparing the KKT conditions of (3) and (6), we see that if $(s^{k\star}, \nu^{k\star,QP}, \mu^{k\star,QP}, \lambda^{k\star,QP})$ is the KKT point of (3), then $(s^{k\star}, \bar{s}^{k\star}, \nu^{k\star,QP}, \mu^{k\star,QP}, \gamma^{k\star}, \lambda^{k\star,QP})$ with $\bar{s}^{k\star} = s^{k\star}$ and $\gamma^{k\star} = E^\top \lambda^{k\star,QP}$ with $E \doteq \begin{bmatrix} E_1 & \dots & E_S \end{bmatrix}$ is the KKT point of (6). Comparing the KKT conditions of (6) and (7), we see that if $(s^{k\star}, \bar{s}^{k\star}, \nu^{k\star,QP}, \mu^{k\star,QP}, \gamma^{k\star,QP}, \lambda^{k\star,QP})$ is the KKT point of (6), then $(s_i^{k\star}, \nu_i^{k\star,QP}, \mu_i^{k\star,QP})$ is a KKT point of (7) of subsystem i with parameters $q_i^{k\star}$.

(b) Since $(s_i^{k\star}, \nu_i^{k\star}, \mu_i^{k\star})$ is the only KKT point of (7), it satisfies strict complementarity for all $i \in \mathcal{S}$:

$$-[h_i(x_i^k)]_j - [\nabla h_i(x_i^k)^\top s_i^{k\star}]_j < [\mu_i^{k\star,QP}]_j, \ \forall j \in \mathcal{A}_i,$$
$$-[h_i(x_i^k)]_j - [\nabla h_i(x_i^k)^\top s_i^k]_j > [\mu_i^{k\star,QP}]_j, \ \forall j \in \mathcal{I}_i.$$

Therefore, there exists $\varepsilon_4 > 0$ such that the active set of (7) is constant for all $\mathrm{col}(\tilde{s}_i^\star, \tilde{\mu}_i^{\star,QP}) \in \mathcal{B}_{\varepsilon_4}(\mathrm{col}(s_i^{k\star}, \mu_i^{k\star,QP}))$. Assumption 1 implies that the solution $(s_i^\star, \mu_i^{\star,QP})$

to (7) is a continuous function of γ_i and $\bar{s}_i$ [19, Thm. 5.2]. Therefore, there exists $\varepsilon_3 > 0$ such that $\mathrm{col}(s_i^\star, \mu_i^{\star,QP}) \in \mathcal{B}_{\varepsilon_4}(\mathrm{col}(s_i^{k\star}, \mu_i^{k\star,QP}))$ for all $q \in \mathcal{B}_{\varepsilon_3}(q^{k\star})$. Hence, $(s_i^\star, \mu_i^{\star,QP})$ has the same active set as $(x_i^\star, \mu_i^\star)$ for all $q \in \mathcal{B}_{\varepsilon_3}(q^{k\star})$. ∎

We next show that ADMM can be initialized such that the active set stays constant.

Lemma 4 (ADMM active set): Let Assumption 1 hold and let $p^k \in \mathcal{B}_{\varepsilon_1}(p^\star)$. Then there exists $\varepsilon_5 > 0$ such that if $q^0 \in \mathcal{B}_{\varepsilon_5}(q^{k\star})$, then the iterates $(\bar{s}^l, \mu^{l,QP})$ have the same active set as $p^\star$ for all $l \ge 0$, i.e.

$$[h_i(x_i^k)]_j + [\nabla h_i(x_i^k)^\top \bar{s}_i^l]_j = 0, \forall j \in \mathcal{A}_i, \forall i \in \mathcal{S}, \forall l \ge 0,$$
$$[\mu_i^{l,QP}]_j = 0, \forall j \in \mathcal{I}_i, \ \forall i \in \mathcal{S}, \forall l \ge 0. \ \square$$

Proof: We first show that, (a), inside $\mathcal{B}_{\varepsilon_1}(p^\star)$ problem (6) is convex on the constraints. Then, (b), standard ADMM convergence results imply that the iterates will stay in a bounded neighborhood of $q^{k\star}$. Finally, (c), we show that this neighborhood can be chosen such that the ADMM iterates have the correct active set.

(a) Let $G_i^k \doteq \nabla g_i(x_i^k)^\top$. Recall that for $A \in \mathbb{R}^{m \times n}$ the null space of A and the range space of $A^\top$ together form $\mathbb{R}^n$ [8, p. 603]. We can therefore write $s_i = u_i + v_i$, where u_i lies in the null space of G_i^k and v_i lies in the range space of $G_i^{k\top}$. From (7b) and LICQ follows $v_i = -G_i^{k\top}(G_i^k G_i^{k\top})^{-1} g_i(x_i^k)$. Since v_i is uniquely determined and since $u_i^\top H_i^k u_i > 0$ because of Assumption 1, the local objective $\phi_i^k(s_i)$ is a convex function for all feasible s_i.

(b) ADMM is therefore guaranteed to converge to $(s^{k\star}, \bar{s}^{k\star}, \nu^{k\star,QP}, \mu^{k\star,QP}, \gamma^{k\star}, \lambda^{k\star,QP})$ and the ADMM iterates satisfy [20, Eq. (3.5)]

$$\rho \|\bar{s}^{l+1} - \bar{s}^{k\star}\|_2^2 + \frac{1}{\rho}\|\gamma^{l+1} - \gamma^{k\star}\|_2^2 \le \rho\|\bar{s}^l - \bar{s}^{k\star}\|_2^2 + \frac{1}{\rho}\|\gamma^l - \gamma^{k\star}\|_2^2$$

for all $l \ge 0$. If $0 < \rho < \infty$, then the above inequality implies that for all $\varepsilon > 0$, there exists $\varepsilon_5 > 0$ such that if $q^0 \in \mathcal{B}_{\varepsilon_5}(q^{k\star})$, then $q^l \in \mathcal{B}_\varepsilon(q^{k\star})$ for all $l \ge 0$.

(c) Strict complementarity at $\bar{s}^{k\star}$ implies that there exists $\varepsilon_6 > 0$ such that the active set is constant for all $\bar{s} \in \mathcal{B}_{\varepsilon_6}(\bar{s}^{k\star})$. Lemma 3 implies that $[\mu_i^{l,QP}]_{\mathcal{I}_i}$ is constant, if $q^l \in \mathcal{B}_{\varepsilon_3}(q^{k\star})$ for all $l \ge 0$. Because of (b), we can choose ε_5 such that $\bar{s}^l \in \mathcal{B}_{\varepsilon_6}(\bar{s}^{k\star})$ and $q^l \in \mathcal{B}_{\varepsilon_3}(q^{k\star})$ for all $l \ge 0$ and hence all iterates have the same active set as $p^\star$. ∎

C. Local convergence of decentralized SQP

Algorithm 3 summarizes the decentralized SQP method (d-SQP). The algorithm follows a bi-level structure and we denote the SQP (outer) iterations by k and the ADMM (inner) iterations by l. We are now ready to state our main result.

Theorem 2 (Local convergence of d-SQP): Let Assumption 1 hold and initialize ADMM in each SQP step with $\bar{s}_i^0 = 0$ and $\gamma_i^0 = E_i^\top \lambda^k$ for all $i \in \mathcal{S}$. Then, there exist $\varepsilon > 0$ and $\eta > 0$ such that for all $p^0 \in \mathcal{B}_\varepsilon(p^\star)$ the following holds.

i) If $\eta^k \le \eta$, then the sequence $\{p^k\}$ generated by Algorithm 3 converges to $p^\star$ and the convergence rate is q-linear in the outer iterations.

Algorithm 3 d-SQP for solving (1)

1: SQP initialization: $k = 0, x_i^0, \nu_i^0, \mu_i^0$ for all $i \in \mathcal{S}$, λ^0, η^0, ϵ
2: **while** $\|F^k\| \not\leq \epsilon$ **do**
3:     compute $\nabla f_i(x_i^k), \nabla g_i(x_i^k), \nabla h_i(x_i^k), H_i^k$ locally
4:     compute $\tilde{F}_i^k, \nabla \tilde{F}_i^k$ locally
5:     ADMM initialization: $l = 0, \bar{s}_i^0, \gamma_i^0$ for all $i \in \mathcal{S}$
6:     **while** $\|\tilde{F}^k + \nabla \tilde{F}^k d\| \not\leq \eta^k \|\tilde{F}^k\|$ **do**
7:         $s_i^{l+1} = \mathrm{argmin}_{s_i \in \mathbb{S}_i^k} L_{\rho,i}^k(s_i, \bar{s}_i^l, \gamma_i^l)$ for all $i \in \mathcal{S}$
8:         $\bar{s}^{l+1} = \mathrm{argmin}_{\bar{s} \in \mathbb{E}} \sum_{i \in \mathcal{S}} L_{\rho,i}^k(s_i^{l+1}, \bar{s}_i, \gamma_i^l)$
9:         $\gamma_i^{l+1} = \gamma_i^l + \rho(s_i^{l+1} - \bar{s}_i^{l+1})$ for all $i \in \mathcal{S}$
10:        $l = l + 1$
11:     **end while**
12:     $x_i^{k+1} = x_i^k + \bar{s}_i^l, \nu_i^{k+1} = \nu_i^{l,QP}, \mu_i^{k+1} = \mu_i^{l,QP}, \gamma_i^{k+1} = \gamma_i^l$
13:     update η^{k+1}
14:     $k = k + 1$
15: **end while**
16: **return** $x^\star$

---

  ii) If $\eta^k \to 0$, then the convergence rate is q-superlinear in the outer iterations.

  iii) If $\eta^k = O(\|\tilde{F}^k\|)$ and if $\nabla_{x_i x_i}^2 f_i, \nabla_{x_i x_i}^2 [g_i]_j, j \in \{1, \ldots, n_{gi}\}$, and $\nabla_{x_i x_i}^2 [h_i]_j, j \in \{1, \ldots, n_{hi}\}$ are Lipschitz continuous inside $\mathcal{B}_\varepsilon$, then the convergence rate is q-quadratic in the outer iterations. $\qquad\square$

    Proof: We first show that, (a), the ADMM initialization $\mathrm{col}(0, E^\top \lambda^k)$ lies within a neighborhood of the subproblem solution $q^{k\star}$ for all SQP iterations. We then show that, (b), this neighborhood can be chosen such that the ADMM iterations are at the correct active set at all iterations. Finally, (c), we invoke Lemmas 1–2 to prove convergence.

    (a) We first choose $\varepsilon \leq \varepsilon_2$. The convergence of the Newton and inexact Newton methods implies that if $p^k \in \mathcal{B}_\varepsilon(p^\star)$, then $p^{k+1} \in \mathcal{B}_\varepsilon(p^\star)$ holds for the Newton method as well as for the inexact Newton method [17, Thm. 2.3]. Therefore $\|p^{k+1} - p^k\| \leq 2\varepsilon$. Recall that $\|\mathrm{col}(x, y)\| \geq \|x\|$ for any vectors x and y. Hence, $\|\mathrm{col}(x^{k+1} - x^k, \lambda^{k+1} - \lambda^k)\| \leq 2\varepsilon$. Recall that $\gamma = E^\top \lambda$. Hence, $\|\mathrm{col}(x^{k+1} - x^k, \gamma^{k+1} - \gamma^k)\| \leq 2\varepsilon_7$ with $\varepsilon_7 = \max(\|E^\top\|, 1)\varepsilon$. That is, $\mathrm{col}(0, \gamma^k) \in \mathcal{B}_{2\varepsilon_7}(\mathrm{col}(x^{k+1} - x^k, \gamma^{k+1}))$. The iterations of the Newton method read $x^{k+1} = x^k + \bar{s}^{k\star}$ and $\lambda^{k+1} = \lambda^{k\star,QP}$ and we hence get $\mathrm{col}(0, E^\top \lambda^k) \in \mathcal{B}_{2\varepsilon_7}(q^{k\star})$.

    (b) Lemma 4 shows that the ADMM iterates $(\bar{s}^l, \mu^{l,QP})$ are at the correct active set, if $q^0 \in \mathcal{B}_{\varepsilon_5}(q^{k\star})$. We therefore choose $\varepsilon \leq \min(\varepsilon_2, \varepsilon_5/(2\max(1, \|E^\top\|)))$. The ADMM initialization then satisfies $q^0 = \mathrm{col}(0, E^\top \lambda) \in \mathcal{B}_{\varepsilon_5}(q^{k\star})$. By Lemma 4, the iterates $(\bar{s}^l, \mu^{l,QP})$ therefore are at the correct active set for all ADMM iterations $l \geq 0$ and for all SQP iterations $k \geq 0$.

    (c) The ADMM convergence invoked in the proof of Lemma 4 ensures that ADMM can satisfy the stopping criterion (9). By Lemma 2, satisfaction of (9) together with the correct active set implies satisfaction of the inexact Newton stopping criterion (8). Lemma 1 then implies local q-convergence in the outer iterations. $\qquad\blacksquare$

D. Communication requirements and discussion

    Four steps of Algorithm 3 require communication between subsystems: Step 1 to initialize λ^0, Steps 2 and 6 to evaluate the stopping criteria, and Step 8 for the $\bar{s}$ update in ADMM.

    It is well-known that Step 8 may be computed efficiently as a decentralized averaging step, if (1) is a consensus-type problem, cf. [15, 16]. That is, the update of $\bar{s}$ in Step 8 then requires only the communication of vectors between neighboring subsystems if we choose $\gamma_i^0 = E_i^\top \lambda^0$.

    If $\|\cdot\|_\infty$ is chosen, then the stopping criteria can be evaluated locally and in Steps 2 and 6 only convergence flags must be communicated, because the update $x^{k+1} = x^k + \bar{s}^k$ ensures $\sum_{i \in \mathcal{S}} E_i x_i^k = c$ for all $k \geq 1$.

    Remark 2 (Dual iterates of the coupling constraints): Theorem 2 proves convergence with the dual variable λ^k. However, instead of computing λ^k explicitly, Algorithm 3 exploits that $E_i^\top \lambda = \gamma_i$ to evaluate Steps 2, 4, and 6. $\qquad\square$

    Remark 3 (Hessian regularization): ADMM is guaranteed to converge if the cost functions $\phi_i^k(s_i)$ are convex on the constraints $\mathbb{S}_i^k$. That is, we do not require H_i^k to be positive definite, but only positive definite on the equality constraint null space. However, this may not be the case if d-SQP is initialized far away from $p^\star$. To this end, we regularize the reduced Hessian via [8, Eq. 3.43] with $\delta = 10^{-4}$, where δ is in the notation of [8]. $\qquad\square$

    Remark 4 (Inner method and application to QPs): Here, we present d-SQP with ADMM as an inner method for solving (3). Instead of ADMM, other decentralized methods with guaranteed convergence to a KKT point for convex QPs can also be used. If (1) itself is a QP, then d-SQP is equivalent to applying the inner method to a possibly regularized version of (1). $\qquad\square$

IV. NUMERICAL RESULTS

    We compare the performance of d-SQP to four other methods for an OPF problem for the IEEE 118-bus system. The first method for comparison is standalone ADMM, which solves (1) directly and which we denote as ADMM in the following. The second and third methods are bi-level ALADIN variants, where the ALADIN coordination QP is solved via essentially decentralized Conjugate Gradients (d-CG) or via an ADMM variant (d-ADMM). The fourth method is an essentially decentralized Interior Point method (d-IP) [9].

    The problem consists of four subsystems with a total of $n = 576$ variables, $n_g = 470$ non-linear equality constraints, $n_h = 792$ linear inequality constraints, and $n_c = 52$ coupling constraints. We initialize voltage magnitudes as 1 and remaining variables as 0.

    For ADMM, we tune ρ for fastest convergence in the set $\{10^2, 10^3, 10^4\}$ and obtain $\rho = 10^3$. The respective subproblems in ADMM, and ALADIN are solved with IPOPT [21] and the QPs in d-SQP are solved with qpOASES [22]. For d-SQP, we choose $\eta^0 = 0.8$, $\eta^{k+1} = 0.9\eta^k$, use $\rho = 10^3$, and initialize the inner iterations with $(\bar{s}_i^0, \gamma_i^0) = (0, E_i^\top \lambda^k)$. For bi-level ALADIN, we run either 50 d-CG iterations or 100 d-ADMM iterations per ALADIN iteration. For d-IP, we use
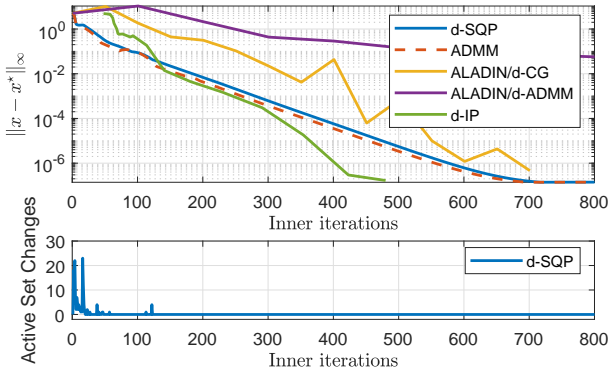
Fig. 1. d-SQP, ADMM, bi-level ALADIN, and d-IP for 118-bus OPF.

the parameters from [9]. We use CasADi [23] to compute sensitivities via algorithmic differentiation in all methods.

The top part of Figure 1 shows the convergence to the minimizer found by IPOPT.[3] For d-SQP, bi-level ALADIN, and d-IP we count the inner iterations. Figure 1 shows that d-SQP requires more iterations than ADMM to achieve an equivalent accuracy in $\|x - x^\star\|_\infty$. However, at the level of each subsystem, d-SQP only solves QPs whereas ADMM solves NLPs. As a result, in our prototypical implementation ADMM takes $67\,\mathrm{s}$ to achieve $\|x - x^\star\|_\infty < 10^{-6}$, whereas d-SQP only takes $30\,\mathrm{s}$. Moreover, d-SQP is guaranteed to converge locally for this problem, whereas ADMM might diverge [1]. The bottom plot of Figure 1 shows the number of inequality constraints which toggle between being active and inactive. The active set stays constant after 127 iterations, which indicates that the area of local convergence is reached.

All methods require local neighbor-to-neighbor communication of the same complexity: $2n_c$ floats for all subsystems combined per inner iteration. ALADIN/d-CG and d-IP further communicate two floats globally per d-CG iteration. The faster convergence of d-IP in Figure 1 therefore comes at the cost of global communication with low complexity. Hence, the bi-level SQP scheme shows competitive *decentralized* performance *and* it admits local convergence guarantees.

V. Conclusion and Outlook

This paper has established new convergence guarantees for decentralized SQP schemes under non-convex constraints. The proposed method solves quadratic subproblems with ADMM. In contrast to existing decentralized SQP methods, we allow for inexact ADMM solutions via an appropriate decentralized stopping criterion for the inner iterations. Numerical results show competitive performance to further optimization schemes for an example from power systems. Future work will consider the globalization of the method and its application to distributed model predictive control.

---

[3]The convergence results for ADMM shown in Figure 1 differ from the results shown in [9]. Here we use a different formulation of the augmented Lagrangian to facilitate the averaging step, cf. [16].

References

[1] K. Christakou, D.-C. Tomozei, J.-Y. Le Boudec, and M. Paolone, "AC OPF in radial distribution networks – Part I: On the limits of the branch flow convexification and the alternating direction method of multipliers," *Electr. Power Syst. Res.*, vol. 143, pp. 438–450, 2017.

[2] B. T. Stewart, S. J. Wright, and J. B. Rawlings, "Cooperative distributed model predictive control for nonlinear systems," *J. Process Control*, vol. 21, no. 5, pp. 698–704, 2011.

[3] Y. Wang, W. Yin, and J. Zeng, "Global convergence of ADMM in nonconvex nonsmooth optimization," *J. Sci. Comput.*, vol. 78, no. 1, pp. 29–63, 2019.

[4] J.-H. Hours and C. N. Jones, "A parametric nonconvex decomposition algorithm for real-time and distributed NMPC," *IEEE Trans. Autom Control*, vol. 61, no. 2, pp. 287–302, 2016.

[5] K. Sun and X. A. Sun, "A two-level distributed algorithm for general constrained non-convex optimization with global convergence," *arXiv preprint arXiv:1902.07654*, 2019.

[6] B. Houska, J. Frasch, and M. Diehl, "An augmented lagrangian based algorithm for distributed nonconvex optimization," *SIAM J. Optim.*, vol. 26, no. 2, pp. 1101–1127, 2016.

[7] A. Engelmann, Y. Jiang, B. Houska, and T. Faulwasser, "Decomposition of nonconvex optimization via bi-level distributed ALADIN," *IEEE Trans. Control Network Syst.*, vol. 7, no. 4, pp. 1848–1858, 2020.

[8] J. Nocedal and S. Wright, *Numerical Optimization*. Springer Science & Business Media, New York, 2006.

[9] A. Engelmann, G. Stomberg, and T. Faulwasser, "Toward decentralized interior point methods for control," in *Proc. of the 60th IEEE Conference on Decision and Control*, 2021.

[10] I. Necoara, C. Savorgnan, D. Q. Tran, J. Suykens, and M. Diehl, "Distributed nonlinear optimal control using sequential convex programming and smoothing techniques," in *Proc. of the 48th IEEE Conference on Decision and Control*, 2009, pp. 543–548.

[11] M. Liu and J. Jian, "An ADMM-based SQP method for separably smooth nonconvex optimization," *J. Inequal. Appl.*, vol. 2020, no. 1, pp. 1–17, 2020.

[12] L. Lu, "Separable nonlinear model predictive control via sequential quadratic programming for large-scale systems," *IFAC-PapersOnLine*, vol. 48, no. 23, pp. 495–500, 2015.

[13] V.-A. Le and T. X. Nghiem, "Gaussian process based distributed model predictive control for multi-agent systems using sequential convex programming and ADMM," in *Proc. of the IEEE Conference on Control Technology and Applications*, IEEE, 2020, pp. 31–36.

[14] C. Geiger and C. Kanzow, *Theorie und Numerik restringierter Optimierungsaufgaben*. Springer, 2002.

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

[16] G. Stomberg, A. Engelmann, and T. Faulwasser, "A compendium of optimization algorithms for distributed linear-quadratic MPC," *at - Automatisierungstechnik*, vol. 70, no. 4, pp. 317–330, 2022.

[17] R. S. Dembo, S. C. Eisenstat, and T. Steihaug, "Inexact newton methods," *SIAM J. Numer. Anal.*, vol. 19, no. 2, pp. 400–408, 1982.

[18] B. Morini, "Convergence behaviour of inexact newton methods," *Math. Comput.*, vol. 68, no. 228, pp. 1604–1613, 1999.

[19] A. V. Fiacco and Y. Ishizuka, "Sensitivity and stability analysis for nonlinear programming," *Ann. Oper. Res.*, vol. 27, no. 1, pp. 215–235, 1990.

[20] B. He and X. Yuan, "On non-ergodic convergence rate of Douglas–Rachford alternating direction method of multipliers," *Numer. Math.*, vol. 130, no. 3, pp. 567–577, 2015.

[21] L. T. Biegler and V. M. Zavala, "Large-scale nonlinear programming using ipopt: An integrating framework for enterprise-wide dynamic optimization," *Comput Chem Eng*, vol. 33, no. 3, pp. 575–582, 2009.

[22] H. J. Ferreau, C. Kirches, A. Potschka, H. G. Bock, and M. Diehl, "qpOASES: A parametric active-set algorithm for quadratic programming," *Mathematical Programming Computation*, vol. 6, no. 4, pp. 327–363, 2014.

[23] J. A. E. Andersson, J. Gillis, G. Horn, J. B. Rawlings, and M. Diehl, "CasADi: A software framework for nonlinear optimization and optimal control," *Math. Program.*, vol. 11, no. 1, pp. 1–36, 2019.