Overlapping Schwarz Decomposition for Constrained Quadratic Programs

Sungho Shin, Mihai Anitescu, Victor M. Zavala

Abstract—We present an overlapping Schwarz decomposition algorithm for constrained quadratic programs (QPs). Schwarz algorithms have been traditionally used to solve linear algebra systems arising from partial differential equations, but we have recently shown that they are also effective at solving structured optimization problems. In the proposed scheme, we consider QPs whose algebraic structure can be represented by graphs. The graph domain is partitioned into overlapping subdomains, yielding a set of coupled subproblems. The algorithm computes the solution of the subproblems in parallel and enforces convergence by updating primal-dual information in the coupled regions. We show that convergence is guaranteed if the overlap is sufficiently large and that the convergence rate improves exponentially with the size of the overlap. Convergence results rely on a key property of graph-structured problems that is known as exponential decay of sensitivity. Here, we establish conditions under which this property holds for constrained QPs, thus extending existing work addressing unconstrained QPs. The numerical behavior of the Schwarz scheme is demonstrated by using a DC optimal power flow problem defined over a network with 9,241 nodes.

I. INTRODUCTION

Structured quadratic programs (QPs) arise in a number of applications such as optimal power flow (OPF), optimization with embedded partial differential equations (PDEs), model predictive control, and multistage stochastic programming. A wide range of decomposition schemes have been proposed to tackle such problems; these include Lagrangian dual decomposition [1], the alternating direction method of multipliers (ADMM) [2], and Jacobi/Gauss-Seidel methods [3]. The basic tenet behind such algorithms is to decompose the original problem into subproblems and to coordinate subproblem solutions by using primal-dual information at the boundary of the subdomains. A disadvantage of these schemes is that convergence can be slow [4].

Overlapping Schwarz algorithms have been used recently to solve large structured optimization problems, and they have been demonstrated to outperform popular schemes such as ADMM and Jacobi/Gauss-Seidel [5]. Schwarz algorithms were originally developed for the parallel solution of linear algebra systems arising from partial differential equations, but such schemes can also be used to handle general linear systems and optimization problems by exploiting their underlying algebraic topology [6]. As the name suggests, overlapping Schwarz algorithms decompose the full problem (the underlying graph) into subproblems that are defined over overlapping subdomains. In the context of QPs that are unconstrained and convex, we have shown that the convergence rate of Schwarz algorithms improves exponentially with the size of the overlapping region [5]. Overlapping Schwarz schemes provide a bridge between fully distributed Jacobi/Gauss-Seidel algorithms (no overlap) and centralized algorithms (where the overlap is the entire domain).

This paper presents a Schwarz algorithm for constrained QPs. We analyze the convergence of the algorithm and derive an explicit relationship between its convergence rate and the size of the overlap. In particular, we show that the algorithm converges with sufficiently large overlap and that the convergence rate exponentially improves with the size of overlap. This convergence result relies on a property called exponential decay of sensitivity. The property states that the sensitivity of the primal-dual solution at a given node decays exponentially with respect to the distance from the perturbation. Such a property has been established for optimal control problems (the graph is a line) [7], [8] and for unconstrained QPs (general graph) [5]. This paper establishes the property for constrained QPs over general graphs.

The paper is organized as follows. In the remainder of this section we introduce basic notation and the problem under study. In Section II we introduce the overlapping Schwarz algorithm. In Section III we present the main theoretical results. We first analyze the sensitivity of the solution of structured QPs against parametric perturbations and then use the results to establish convergence conditions for the algorithm. Numerical results are given in Section IV.

Notation. The set of real numbers and the set of integers are denoted by $\mathbb{R}$ and $\mathbb{Z}$, respectively, and we define $\mathbb{I}_{a,b} := \mathbb{I} \cap [a,b]$, $\mathbb{I}_{>0} := \mathbb{I} \cap (0,\infty)$, $\mathbb{R}_{>0} := (0,\infty)$, and $\mathbb{R} := \mathbb{R} \cup \{\infty\}$. By default, vectors are assumed to be column vectors, and we use the syntax $(M_1,\ldots,M_n) := [M_1^\top \cdots M_n^\top]^\top$, $(M_i)_{i \in \mathcal{I}} := (M_{i_1},\ldots,M_{i_m})$, and $(M_{i,j})_{i,j \in \mathcal{J}} := \{M_{i,j}\}_{i,j \in \mathcal{J}}$. Vector 2-norms and induced 2-norms are denoted by $\|\cdot\|$ and $\|\cdot\|_2$, respectively. For matrices $A$ and $B$, $A \succeq B$ indicates that $A - B$ is positive semi-definite while $A \succeq B$ represents a componentwise inequality.

Setting. We consider a potentially infinite parent graph...
$\mathcal{G}(\mathcal{V}, \mathcal{E})$, where $\mathcal{V}$ is the set of nodes and $\mathcal{E}$ is the set of edges. We also consider the finite node subset $U \subseteq \mathcal{V}$ and the following QP:

$$
\begin{align*}
\min_{\{x_i\}_{i \in U}} & \sum_{i \in U} \sum_{j \in N_U(i)} \frac{1}{2} x_i^T Q_{i,j} x_j - \sum_{i \in U} f_i^T x_i \quad (1a) \\
\text{s.t.} & \sum_{j \in N_U(i)} A_{i,j}^x x_j = g_i^x, \quad (\lambda_i^x), \ i \in U \\
& \sum_{j \in N_U(i)} A_{i,j}^g x_j \geq g_i^g, \quad (\lambda_i^g), \ i \in U. \quad (1c)
\end{align*}
$$

Here $x_i \in \mathbb{R}^{r_i}$ are the decision variables; $\lambda_i^x \in \mathbb{R}^{m_i^x}$ and $\lambda_i^g \in \mathbb{R}^{m_i^g}$ are the dual variables; $Q_{i,j} \in \mathbb{R}^{r_i \times r_j}$, $A_{i,j}^x \in \mathbb{R}^{m_i^x \times r_j}$, $A_{i,j}^g \in \mathbb{R}^{r_i \times m_i^g}$, $f_i \in \mathbb{R}^{r_i}$, $g_i^x \in \mathbb{R}^{m_i^x}$, and $g_i^g \in \mathbb{R}^{m_i^g}$ are the data; and $N_U[X] := N_U(X) \cup X$, where $N_U(X) := \{ j \in U : \exists i \in X \text{ such that } \{i, j\} \in \mathcal{E} \}$ and the argument is considered as a singleton if $X$ is a single node. We define $A_{i,j} := (A_{i,j}^x, A_{i,j}^g)$, $g_i := (g_i^x, g_i^g)$, $\lambda_i := (\lambda_i^x, \lambda_i^g)$, $z_i := (x_i, \lambda_i)$, $d_i := (f_i, g_i)$, $m_i := m_i^x + m_i^g$, and $n_i := r_i + m_i$. We assume that $Q_{i,j} = Q_{j,i}$.

An equivalent problem can be written in a compact form:

$$
P_U(d_U) : \min_{x_U} \frac{1}{2} x_U^T Q_U x_U - f_U^T x_U \quad \text{s.t.} \quad A_U^x x_U = g_U^x, \quad (\lambda_U^x)$$

$$
A_U^g x_U \geq g_U^g, \quad (\lambda_U^g).
$$

Here, $x_U := \{x_i\}_{i \in U}$; $\lambda_U^x := \{\lambda_i^x\}_{i \in U}$; $\lambda_U^g := \{\lambda_i^g\}_{i \in U}$; $z_U := \{z_i\}_{i \in U}$; $d_U := \{d_i\}_{i \in U}$; $Q_U := \{Q_{i,j}\}_{i,j \in U}$; $A_U^x := \{A_{i,j}^x\}_{i,j \in U}$; $A_U^g := \{A_{i,j}^g\}_{i,j \in U}$; $A_U := \{A_{i,j}\}_{i,j \in U}$; $U := \{U_i\}_{i \in U}$; $m_U := \sum_{i \in U} m_i$; and $n_U := \sum_{i \in U} n_i$. The problem is denoted as the parametric form $P_U(d_U)$.

II. OVERLAPPING SCHWARZ ALGORITHM

This section introduces the Schwarz algorithm for the solution of $P_V(d_V)$ (referred to as the full problem) with $V \subseteq \mathcal{V}$. We consider a non-overlapping partition $\{V_k \subseteq \mathcal{V}\}_{k=1}^K$ of $\mathcal{V}$ and an overlapping partition $\{W_k \subseteq \mathcal{V}\}_{k=1}^K$ of $\mathcal{V}$ such that $V_k \subseteq W_k$ holds for $k \in 1:K$. We call $V_1, \ldots, V_K$ original subdomains and $W_1, \ldots, W_K$ expanded subdomains. The Schwarz algorithm is detailed below.

Algorithm 1 Overlapping Schwarz Algorithm

**Require**: $z^{(0)}_V$, $\{V_k\}_{k=1}^K$, $\{W_k\}_{k=1}^K$

1. for $\ell = 0, 1, \ldots$
2. for (in parallel) $k = 1$ to $K$
3. $z^{(\ell+1)}_V = T_{V_k \leftarrow W_k} z^{(\ell)}_W W_k - H_{W_k} z^{(\ell)}_W$
4. end for
5. end for

**Ensure**: $z^{(\ell)}_V$

Here, we use a syntax that can be applied to any $U \subseteq \mathcal{V}$:

$H_U := \{H_{i,j}\}_{i,j \in U}$; $H_U := \{H_{i,j}\}_{i \in U, j \in N_U(i)}$; $H_{i,j} := \{Q_{i,j} A_{i,j}^x A_{i,j}^g 0\}$; and $z_U := \{z_i\}_{i \in N_U(U)}$. Furthermore,

$z_U^T(\cdot)$ is the primal-dual solution mapping of the parametric optimization problem $P_U(\cdot)$; $T_{U \leftarrow U} := \{T_{i,j}\}_{i \in U, j \in U}$, where $U_1, U_2 \subseteq \mathcal{V}$ and $T_{i,j} = I_{n_i \times n_j}$ if $i = j$ and $0_{n_i \times n_j}$ otherwise. Note that $z_U$ is supposed to represent the solution information that is complementary to $U$. The full complementary solution information includes the solution on $V \setminus U$. However, the variables and constraints in $U$ are coupled only with $N_U(U)$, so it suffices to incorporate information only for the coupled complementary region $N_U(U)$. Therefore, we will abuse the term complementary solution to represent the coupled complementary solution $z_{-U}$.

The core part of the algorithm (line 3) consists of three steps: subproblem solution, solution restriction, and primal-dual exchange. In the first step, one formulates the subproblem for the $k$th subdomain as $P_{W_k}(d_{W_k} - H_{W_k} z^{(\ell)}_{W_k})$ (this formulation will be justified later in Lemma 6). The subproblem incorporates complementary solution information $H_{W_k} z^{(\ell)}_{W_k}$, which is obtained during the third inner step of the previous iteration step. The subproblem is solved to obtain its solution $z^{(\ell)}_{W_k}(d_{W_k} - H_{W_k} z^{(\ell)}_{W_k})$. Here, we observe that solution multiplicity exists at the overlapping region. To remove such multiplicity, we restrict the solution in the second step. In particular, we abandon the primal-dual solutions associated with $W_k \setminus V_k$ (subdomain region acquired by expansion) and take only those solutions associated with $V_k$ (the original subdomain region). This procedure is represented by the restriction operator $T_{V_k \leftarrow W_k}$. After restriction, the solutions are assembled over $k \in 1:K$ to make the next guess of the solution $z^{(\ell+1)}_V$. In the third step, the primal-dual solutions are exchanged across the subdomains to update the complementary information $H_{W_k} z^{(\ell)}_{W_k}$ for each subproblem. The schematic of the algorithm is shown in Fig. 1. The algorithm can be implemented in a fully distributed manner, and different updating schemes can be used (e.g., Gauss-Seidel or asynchronous); see [5] for details.

III. MAIN RESULTS

In this section we analyze the convergence of the Schwarz algorithm. We will see that parametric sensitivity plays a central role in convergence behavior because, intuitively, primal-dual solutions of the neighbors of a subdomain enter as parametric perturbations. We first analyze the parametric
Proof. We let $\mathbb{B}_U = \{B(1), \ldots, B(T)\}$ (note that $\mathbb{B}_U$ is finite).
We define $\pi(t) : [0,1] \rightarrow \mathbb{R}$ to be the mapping from $s \in [0,1]$ to the objective value of $z_{B(t)}^U(d_U^t)$ for $P_U(d_U^t)$ (the value is $+\infty$ if $B(t)$ is infeasible). Also, we define $\pi(s) := \min_{t \in I_{1:T}} \pi(t)(s)$. By Lemma 1, $\pi(t)$ is the objective value mapping of $P_U(d_U^t)$ from $s \in [0,1]$.

One can see that $z_{B(t)}^U(d_U^t)$ is affine in $s$; thus the feasibility conditions for $z_{B(t)}^U(d_U^t)$ can be expressed by a finite number of affine equalities and inequalities. This implies that the set of $s \in [0,1]$ on which $\pi(s) < \infty$ is obtained as a closed interval in $[0,1]$. Accordingly, $\pi(t)(\cdot)$ is a quadratic function on a closed interval support. Now, we collect the endpoints of such intervals over $t \in I_{1:T}$ to construct $\Pi := \{\hat{s}_0 = 0, \ldots, \hat{s}_{N_d} = 1\}$. For each $k \in I_{1:N_q}$, we collect $T_k := \{t \in I_{1:T} : \pi(t)(\cdot) < \infty \ (\hat{s}_{k-1}, \hat{s}_k)\}$. Observe that (i) $\pi(t)$ with $t \in T_k$ are quadratic on $(\hat{s}_{k-1}, \hat{s}_k)$ and (ii) $\pi(t) = \min_{t \in T_k} \pi(t)(\cdot)$ on $(\hat{s}_{k-1}, \hat{s}_k)$. By Lemma 3 (stated below, applicable due to (i)), we have that each $(\hat{s}_{k-1}, \hat{s}_k)$ can be further divided by using $\tilde{\Pi}_k := \tilde{\Pi}_{k,0} \cup \tilde{\Pi}_{k,k'}$ with $k' \in I_{1:N_d}$ and $k \in I_{1:N_q}$, there exists $t \in T_k$ such that $\pi(t) = \min_{t \in T_k} \pi(t)(\cdot)$ (recall observation (ii)).

We now let $\{\hat{s}_0, \ldots, \hat{s}_{N_d}\} = \bigcup_{k=1}^{N_q} \tilde{\Pi}_k$. One can observe that, for each $k \in I_{1:N_q}$, there exists $t \in I_{1:T}$ such that $\pi(t)(\cdot) = \pi(\cdot)$ on $(\hat{s}_{k-1}, \hat{s}_k)$. We choose such $B(t)$ as $B_k$; it is known that the objective value mapping of a QP is continuous on its support (e.g., see [10, Corollary 9] or [11, Theorem 5.53]); thus $\pi(t)$ is continuous on its support. By the continuity of $\pi(\cdot)$ and $\pi(t)(\cdot)$ on their supports, we have that $\pi(t)(\cdot) = \pi(\cdot)$ on $(\hat{s}_{k-1}, \hat{s}_k)$ implies that the same holds on $[\hat{s}_{k-1}, \hat{s}_k]$. Finally, one can check that $B_k$ is optimal for $P_U(d_U^t)$ with $s \in [\hat{s}_{k-1}, \hat{s}_k]$. The desired $\{\hat{s}_k\}_{k=0}^{N_q}$ and $\{B_k\}_{k=1}^{N_q}$ are thus obtained.

\textbf{Lemma 3.} Let $q_1, \ldots, q_{N_q} : (a,b) \rightarrow \mathbb{R}$ be quadratic functions. We let $q : (a,b) \rightarrow \mathbb{R}$ be $q(\cdot) := \min_{i \in I_{1:N_q}} q_i(\cdot)$. Then there exist $\{a_0 = a, a_1, \ldots, a_k = b\}$ such that for each $k \in I_{1:N_q}$, there exists $i \in I_{1:N_q}$ such that $q_i(\cdot) = q(\cdot)$ on $(a_{k-1}, a_k)$.

Proof. For each $(i,j) \in I_{1:N_q} \times I_{1:N_q}$, we let $I_{i,j} := \{x \in (a,b) : q_i(x) \leq q_j(x)\}$. Since $q_i(\cdot)$ and $q_j(\cdot)$ are quadratic, we have that $I_{i,j}$ is obtained as a union of intervals (not necessarily open or closed). Then we can define $I_i := \bigcup_{j \in I_{1:N_q}} I_{i,j}$. Since $I_{i,j}$ are unions of intervals, we have that $I_i$ is also a union of intervals. By collecting the end points of the intervals in $I_i$, $i \in I_{1:N_q}$, we can construct $\{a_0, \ldots, a_{N_K}\}$. In particular, $\{a_0, \ldots, a_{N_K}\} = \bigcup_{i \in I_{1:N_q}} \text{closure}(I_i) \setminus \text{interior}(I_i)$. We observe that $\bigcup_{i \in I_{1:N_q}} I_i = (a,b)$; thus, for any $k \in I_{1:N_q}$, $(a_{k-1}, a_k)$ is contained in $I_i$ for some $i \in I_{1:N_q}$. This means $q_i(\cdot) = q(\cdot)$ on $(a_{k-1}, a_k)$. The proof is complete.

An important implication of Lemma 2 is that the solution path obtained by the perturbation on $d_U$ can be divided into
multiple paths, each of which is a basic solution mapping. Thus, given Lemma 2, it suffices to study the sensitivities only of the basic solution mappings.

B. Exponential Decay of Sensitivity

We now establish our main sensitivity result for the constrained QP, known as exponential decay of sensitivity. 

**Theorem 1** (Exponential Decay of Sensitivity). Let Assumption 1 hold. The following holds for $d_U, d'_U \in \mathbb{R}^n$:

\[
\| T_{i-U} (z^{v}_{U}(d) - z^{v}_{U}(d')) \| = \sum_{j \in U} \Gamma_U \rho_U \left( \frac{\Delta_U(i,j) - 1}{2} \right) \| d_j - d'_j \|
\]

with $\Gamma_U := \sigma_{\max} / \sigma_U^2$; $\rho_U := (\sigma_{\max}^2 - \sigma_U^2) / (\sigma_{\max}^2 + \sigma_U^2)$; $\sigma_{\max} := \max_{B \in \mathcal{B}_U} \sigma_{\max} (H_U[B,B])$; $\sigma_{\min} := \min_{B \in \mathcal{B}_U} \sigma_{\min} (H_U[B,B])$. 

Here $\lceil \cdot \rceil$ denotes the ceiling operator, and $\Delta_U(i,j)$ denotes the geodesic distance between $i, j \in U$ on the subgraph of $G(\mathcal{V}, \mathcal{E})$ induced by $U$, or the number of elements in the shortest path $\{ e_q \in \mathcal{E} : e_q \subseteq \cup_{i \in 1} \}$. From the definition of bases and basic solutions, we have:

\[
\sum_{i \in U} (\overline{H}_U(i,i)) = 0.
\]

**Lemma 4.** If $\Delta_U(i,j) > q \in \mathbb{N}_0$, then $H_U[B,B] = 0$.

**Proof.** We use the notation $\overline{H}_U := H_U[B,B]$. We proceed by induction. From the fact that $H_{i,j} > 0$ if $\Delta_U(i,j) > 1$, one can observe that $H_{i,i} = 0$ if $\Delta_U(i,i) > 1$. Hence, the claim holds for $q = 1$. Now suppose the claim holds for $q = q'$. One can easily see that triangle inequality holds for distance $\Delta_U(\cdot, \cdot)$. From triangle inequality, if $\Delta_U(i,j) > q' + 1$, either $\Delta_U(i,l) > q'$ or $\Delta_U(l,j) > 1$ holds for any $l \in U$. Thus, if $\Delta_U(i,j) > q' + 1$, then $\overline{H}_U(i,j) = 0$.

\[
\| (H_U[B,B])_{i,j} \| \leq \Gamma_U \rho_U \left( \frac{\Delta_U(i,j) - 1}{2} \right) \| d_j - d'_j \|
\]

By definition of bases and basic solutions, we have:

\[
\| (z^B_{U}(d^B_{U}))_{i,j} \| \leq \Gamma_U \rho_U \left( \frac{\Delta_U(i,j) - 1}{2} \right) \| d_j - d'_j \|
\]

From (9)-(10), the triangle inequality holds for distance $\Delta_U(\cdot, \cdot)$, which is the condition for $\Delta_U(\cdot, \cdot)$. Therefore, the effect of the perturbation decays as one moves away from the perturbation.
C. Convergence Analysis

In this subsection, we formally establish the convergence of Algorithm 1. To facilitate the later discussion, we first introduce some notation: \(z_t^U := \{z_t^i\}_{i \in U}\), where \(z_t^i := T_{i-t}z_t^i(d_V)\) and \(U \subseteq V\). Furthermore, we define \(\|\cdot\|_{U,1} := \sum_{i \in U} \|T_{i-t}z_t^i(\cdot)\|\) and \(\|\cdot\|_{U,\infty} := \max_{i \in U} \|T_{i-t}z_t^i(\cdot)\|.\) Note that \(\|\cdot\|_{1,1}, \|\cdot\|_{1,\infty} : \mathbb{R}^{nu} \to \mathbb{R}_{\geq 0}\) are vector norms; accordingly, \(\|\cdot\|_{U,1}, \|\cdot\|_{U,\infty} : \mathbb{R}^{nu \times q} \to \mathbb{R}_{\geq 0}\) are induced matrix norms for any \(q \in \mathbb{N}_{\geq 0}\).

**Assumption 2.** Assumption 1 holds with \(D_U := \{d_U - H_{-U}z_{-U} \in \mathbb{R}^{nu}\}\) for any \(U \subseteq V\).

Here \(n_{-U} = \sum_{i \in N_U(U)} n_i\). While Assumption 2 is strong, we believe it can be relaxed; but doing so meaningfully is a technically extensive endeavor beyond the scope of this communication format. We note, however, that it is always satisfied for bound constraints and for augmented Lagrangian reformulations of the original problem by using slacks to obtain only bound inequality constraints and then penalizing all equality constraints, which can approximate the infinite set of the original problem arbitrarily well.

**Assumption 3.** \(\omega := \min_{k \in \mathbb{N}_{\leq K}} \Delta_V(V_k, N_V(W_k)) - 1 \geq 1\).

Here, we abuse the notation by letting \(\Delta_V(U, U') := \min_{\omega \in (U, U') \Delta_V(U, U') \Delta_U(U, U')}, \) where \(U, U' \subseteq V\). We call \(\omega\) the size of overlap. Note that an overlapping partition \(\{W_k\}_{k=1}^{K}\) with size \(\omega\) can be constructed from a non-overlapping partition \(\{V_k\}_{k=1}^{K}\) by expanding each original subdomain using \(W_k = \{i \in V : \Delta_V(i, V_k) \leq \omega\}\).

**Assumption 4.** \(\sigma := \inf_{U \subseteq V} \sigma_U > 0, \sigma := \sup_{U \subseteq V} \sigma_U < \infty.\)

Assumption 4 trivially holds if the parent graph \(G(V, E)\) is finite, but it may be violated if the parent graph is infinite. In particular, \(\sigma_V\) or \(\sigma_V\) may tend to zero or infinity as \(V\) grows. Checking the validity of Assumption 4 for the infinite parent graph case is beyond the scope of this paper; sufficient conditions for this to hold will be studied in the future.

We note, however, that Assumptions 2 and 4 hold for augmented Lagrangian reformulations with bounded data when the objective matrix has bounded entries and is strongly diagonal dominant, which is a way we can approximate most QPs with some regularization; in contrast, Assumption 3 can be satisfied by construction. Therefore our setup contains a large set of problems or arbitrarily close approximations of them.

**Lemma 6.** Let Assumption 2 hold. For any \(U \subseteq V\) we have that \(z_{t+1}^U = z_t^U(d_U - H_{-U}z_{-U}^U).\)

**Proof.** Since \(P_U(\cdot)\) is a convex quadratic program, the KKT conditions are necessary and sufficient for optimality. By Assumption 2, the primal-dual solution is unique; therefore, it suffices to prove that \(z_{t+1}^U\) satisfies the KKT conditions of \(P_U(d_U - H_{-U}z_{-U}^U)\). From the KKT conditions of \(P_U(d_V)\), we have

\[
Q_Ux_U^1 + A_U^T\lambda_U = f_U, \quad A_U^Tz_U^1 = g_U, \quad A_U^T x_U^1 \geq g_U^T \lambda_U^T \geq 0, \quad \text{diag}(A_U^T)(A_U^T x_U^1 - g_U^T) = 0.
\]  

(12)

By extracting the rows associated with \(U\) and rearranging equations and inequalities, we obtain

\[
Q_Ux_U^1 + A_U^T\lambda_U = f_U - Q_{-U}x_{-U} - A_{-U}^T\lambda_{-U} \quad A_U^Tz_U^1 = g_U - A_{-U}^T x_{-U} - A_{-U}^T x_{-U}^1, \quad \lambda_U^T \geq 0, \quad \text{diag}(A_U^T)(A_U^T x_U^1 - g_U^T + A_{-U}^T x_{-U}^1) = 0.
\]

(13)

Here, note that \(A_{-U} := \{A_{i,i} \in U, j \in N_V(U)\} \) and \(A_{-U} := \{A_{i,j} \in U, j \in N_V(U)\}\) (they are not transpose to each other). Conditions (13) imply that \(z_{t+1}^U\) satisfies the KKT conditions for \(P_U(d_U - H_{-U}z_{-U}^U)\). Thus, it is the solution. ⊓⊔

Lemma 6 provides a form of consistent subproblems whose solution recovers a piece of the full solution as long as the complementary solution information is accurate. Thus, it justifies the subproblem formulation in Algorithm 1.

**Remark 1.** Lemma 6 reveals that the algorithm applies an (overlapping) block-Jacobi scheme to the saddle point problem for the Lagrangian. In particular, Algorithm 1 is equivalent to performing, for \(t = 0, 1, \ldots\),

\[
z_{\bar{V}}^{(t+1)} = T_{\bar{V}} - z_{\bar{V}}^{(t)} \quad \text{argminmax}_{z_{\bar{V}}^{(t)}} \quad L(z_{\bar{V}}^{(t)}, z_{\bar{V}}^{(t)}), \quad k \in \mathbb{I}_{1:K},
\]

where \(L(\cdot)\) is the Lagrangian of \(P_U(d_V)\), \(z_U := \mathbb{R}^{nu} \times \mathbb{R}^{m_U^\ell} \times \mathbb{R}^{m_U^f}\), and argminmax(\cdot) denotes the saddle point where the given function is minimized over the primal directions and maximized over the dual directions.

We can now state our main result.

**Theorem 2.** (Convergence of Overlapping Schwarz) Let Assumptions 2–4 hold. The sequence \(\{z_{V}^{(t)}\}_{t=0}^{\infty}\) generated by Algorithm 1 satisfies

\[
\|z_{V}^{(t)} - z_{V}^{(t)}\|_{V, \infty} \leq (\alpha_V(\omega))^{t} \|z_{V}^{(0)} - z_{V}^{(0)}\|_{V, \infty}.
\]

(14)

Here, \(\alpha_V(\omega) := \rho \Gamma_0(1/(\omega-1/2)); \quad \rho := \max_{U \subseteq V} \|H_{-U}\|_{U, 1}; \quad \Gamma_0 := \sigma / \sigma^2; \quad \rho := (\sigma^2 - \sigma^2)/(\sigma^2 + \sigma^2).
\]

**Proof.** By Lemma 6, we have that, for any \(i \in V_k,\)

\[
\|z_{i}^{(t+1)} - z_{i}^{(t)}\| \leq \|T_{i-t} z_{i}^U (d_{W_k} - H_{-W_k} z_{l-W_k}) - T_{i-t} z_{i}^U (d_{W_k} - H_{-W_k} z_{l-W_k})\| \leq \sum_{j \in W_k} \Gamma_{W_k}^{\max_{[\Delta W_k(i,j)-1/2]} \times \|T_{j-t} z_{j}^U (H_{-W_k} (z_{l-W_k} - z_{l-W_k}))\|}.
\]

Here the inequality follows from Theorem 1. A key observation is that \(T_{j-t} z_{j}^U (H_{-W_k} (z_{l-W_k} - z_{l-W_k})) = 0\) only if \(\Delta_V(j, N_V(W_k)) = 1\).

Such a \(j\) satisfies \(\Delta_{W_k}(i, j) \geq \omega\), for any \(i \in V_k,\) by
the definition of $\omega$, the triangle inequality for $\Delta V(\cdot, \cdot)$, and $\Delta W_k(i, j) \geq \Delta V(i, j)$. Therefore, for any $i \in V$,
\[
\left\| z^{(l+1)}_i - z^*_i \right\| \leq \left( \max_{k \in 1:K} \| H_{-W_k} \| w_k, 1 \right) \| \omega - \alpha_k \| z^{(l)}_i - z^*_i \|_{V, \infty}.
\]

One can show that $\max_{k \in 1:K} \alpha_k \leq \alpha_V(\omega)$ with Assumptions 3–4. This implies $\| z^{(l+1)}_V - z^*_V \|_{V, \infty} \leq \alpha_V(\omega)\| z^{(l)}_V - z^*_V \|_{V, \infty}$, which yields (14).

The upper bound of convergence rate $\alpha_V(\omega)$ decays with an increase in the size of overlap $\omega$ (recall that $\rho \in (0, 1)$). In the case of a finite parent graph, Algorithm 1 converges if $\omega$ is sufficiently large, since either (i) $\alpha_V(\omega) < 1$ or (ii) each subproblem becomes the full problem (the solution is obtained in one iteration).

If the parent graph is infinite (such a setting is relevant for time grids, unbounded physical space domains, and scenario trees), we can make $V$ arbitrarily large, and thus the limiting behavior of $\alpha_V$ is of interest. Theorem 2 suggests that constant $\Gamma$ and the exponential decay factor $\rho$ do not deteriorate as $V$ grows, but $R_V$ may grow with $V$. Here, $R_V$ represents the maximum coupling between the subdomains $U \subseteq V$ with its surroundings $V \setminus U$. Accordingly, the growth of $R_V$ is determined by the topology (as long as $\| H_{-W_k} \|$ are uniformly bounded). In particular, if the parent graph $G(V, E)$ is finite-dimensional (e.g., grid points in $\mathbb{I}^d$ with $d < \infty$), $R_V$ grows in a polynomial manner with the diameter of $V$. In such a case, a sufficiently large $\omega$ can be found so that the decay of $\rho^{(\omega-1)/2}$ offsets the growth of $R_V$ (an exponentially decaying function times a polynomial converges). On the other hand, in the case of scenario trees, where $R_V$ may grow exponentially, the upper bound derived in Theorem 2 may not provide a tight upper bound, and $\alpha_V(\omega)$ may diverge as $V$ grows no matter how large $\omega$ is.

Note that there exists an inherent trade-off between the convergence rate and subproblem complexity. The convergence rate improves with $\omega$ but the subproblem solution times also increase with $\omega$. Therefore, to achieve maximum performance, one needs to tune $\omega$.

D. Monitoring Convergence

Convergence can be monitored by checking the residuals to the KKT conditions (12) of the full problem $P_V(d_V)$. However, a more convenient surrogate of the full KKT residuals can be derived as follows.

Proposition 1. Suppose that Assumptions 2–3 hold, and let $\{ z^{(l)}_V \}_{l=0}^\infty$ be a sequence generated by Algorithm 1 and $\tilde{z}^{(l)}_{W_k} := z^{(l)}_{W_k} - H_{-W_k} z^{(l-1)}_{W_k}$. Then $z^{(l)}_V \to z^*_V$ as $l \to \infty$ if the following holds for $k \in 1:K$:
\[
E^{(l)}_k := z^{(l)(k)}_{N^*_V(V_k)} - z^{(l)(k)}_{N^*_V(V_k)} \to 0, \text{ as } l \to \infty.
\]

Proof. We make two observations. (i) The KKT conditions of $P_{W_k}(d_{W_k} - H_{-W_k} z^{(l-1)}_{W_k})$ and the fact that $N^*_V(V_k) \subseteq W_k$ (from Assumption 3) imply that (13) holds when we replace $U \leftarrow V_k, z^{(l)}_U \to z^{(l)(k)}_{V_k}$, and $z^+_U \to z^{(l)(k)}_{N^*_V(V_k)}$ for any $k \in 1:K$. (ii) The residual of the KKT systems can be obtained from the residuals to (13) by replacing $U \leftarrow V_k, z^{(l)}_U \to z^{(l)(k)}_{V_k}$, and $z^+_U \to z^{(l)(k)}_{N^*_V(V_k)}$ and collecting the residuals for $k \in 1:K$. Note that the residuals of the KKT conditions (13) are continuous with respect to $z^{(l)}_V$. By using a continuity argument, (15) and observations (i)–(ii), the limit points of $\{ z^{(l)}_V \}_{l=1}^\infty$ satisfy (13) for $z^{(l)}_V$. By the uniqueness of the solution such a limit point is unique, and this implies $z^{(l)}_V \to z^*_V$ as $l \to \infty$.

We can define primal-dual errors as $\epsilon_p := \max_{k \in 1:K} \| pr(E^{(l)}_k) \|_\infty$ and $\epsilon_u := \max_{k \in 1:K} \| du(E^{(l)}_k) \|_\infty$, where $pr(\cdot)$ and $du(\cdot)$ extract the indices associated with primal variables and dual variables, respectively. Convergence criteria can thus be set to the following: Stop if $(\epsilon_p < \epsilon_{p0}) \land (\epsilon_u < \epsilon_{u0})$.

IV. Numerical Example

Consider the regularized DC OPF problem [13] over a network $G(V, E)$:
\[
\min_{\{\theta_i\}_{i \in V}, \{ P_{Q} \}_{Q \in E} } \sum_{q \in \Omega} c_{q,1} P_{q} + c_{q,2} P^{2}_{q} + \frac{1}{2} \sum_{i \in V} (\theta_i - \theta_j)^2 \quad (16a)
\]
\[\text{s.t. } \sum_{P_{q} \in \Omega} P_{q} - \sum_{j \in N^*_V[i]} B_{i,j}(\theta_i - \theta_j) = P^L_i, i \in V \quad (16b)\]
\[P_{q} \leq P_{q} \leq P_{q}^u, q \in \Omega, \theta_i = \theta^r_i, i \in V^r \quad (16c)\]
\[-B_{i,j} \leq \theta_i - \theta_j \leq B_{i,j}, \{ i, j \} \in E. \quad (16d)\]

Here, $\Omega_i$ is the set of generators in node $i$; $\Omega := \bigcup_{i \in V} \Omega_i$ is the set of all generators; $\gamma^r$ is the set of reference nodes; $\theta_i$ are the voltage angles; $P_{q}^u$ are the active power generations; $c_{q,1}$ and $c_{q,2}$ are the generation cost coefficients; $\gamma$ is the regularization coefficient; $B_{i,j}$ are the line susceptances; $P^L_i$ are the active power loads; $P_{q}^u$ and $P_{q}^l$ are the lower and upper bounds of active power generations, respectively; $\theta_i$ are the voltage angle separation limits; and $\theta^r$ are reference voltage angles. The problems are modeled in the algebraic modeling language JuMP [14] and solved with the nonlinear programming solver Ipopt [15]. The 9,241-bus test case obtained from pglib-opf (v19.05) is used [16]. We modified the data by placing artificial storage with infinite capacity and high charge/discharge cost in each node. The network node set is partitioned into 16 subdomains using the graph partitioning tool Metis [17], and each subdomain is expanded to obtain an overlapping partition. The Schwarz scheme is run on a multicore parallel computing server (shared memory and 32 CPUs of Intel Xeon CPU E5-2698 v3 running at 2.30 GHz) using the Julia package Distributed.jl. One master process and 16 worker processes are used (one process per one subproblem). We use $\gamma = 10^3$, $\epsilon_{p0} = 10^{-2}$, and $\epsilon_{u0} = 10^2$. The scripts to reproduce the results can be found here https://github.com/zavalab/JuliaBox/tree/master/SchwarzQPcons.

Convergence results are shown in Fig. 2: we vary the size of overlap $\omega$ and show the evolution of the objective.
value and primal error. The black dashed line represents the optimal objective value (obtained by solving the problem with IPOPT) and the error tolerances. The total computation times and iteration counts are compared in Table I. We can see that increasing \( \omega \) accelerates convergence (reduces the iteration count roughly proportionally with the size of the overlap) but does not always reduce the computation time. The reason is that the subproblem complexity increases with \( \omega \); thus, one can see that an optimal overlap exists.

V. Conclusions

We have presented an overlapping Schwarz algorithm for solving constrained quadratic programs. We show that convergence relies on an exponential decay of the sensitivity result, which we establish for the setting of interest. The algorithm was demonstrated by using a large DC optimal power flow problem. In future work we will study the convergence of the algorithm in a nonlinear setting, and we will conduct more extensive benchmark studies.

Acknowledgment

This material is based upon work supported by the U.S. Department of Energy, Office of Science, Office of Advanced Scientific Computing Research (ASCR) under Contract DE-AC02-06CH11347 and by NSF through award CNS-1545046.

References

[1] C. Lemaréchal, “Lagrangian relaxation,” in Computational combinatorial optimization. Springer, 2001, pp. 112–156.

Fig. 2. Profiles for objective and primal error for different overlap sizes.

| Table I: Effect of overlap on solution times and iteration counts |
|---------------------|-----------------|----------------|
| \( \omega = 1 \)     | \( \omega = 5 \) | \( \omega = 10 \) |
| Solution time        | 12.5 sec        | 17.7 sec        | 24.2 sec        |
| Iterations           | 190 iter        | 39 iter         | 25 iter         |

The reason is that the subproblem complexity increases with \( \omega \); thus, one can see that an optimal overlap exists.

Table I: Effect of overlap on solution times and iteration counts

value and primal error. The black dashed line represents the optimal objective value (obtained by solving the problem with IPOPT) and the error tolerances. The total computation times and iteration counts are compared in Table I. We can see that increasing \( \omega \) accelerates convergence (reduces the iteration count roughly proportionally with the size of the overlap) but does not always reduce the computation time. The reason is that the subproblem complexity increases with \( \omega \); thus, one can see that an optimal overlap exists.

V. Conclusions

We have presented an overlapping Schwarz algorithm for solving constrained quadratic programs. We show that convergence relies on an exponential decay of the sensitivity result, which we establish for the setting of interest. The algorithm was demonstrated by using a large DC optimal power flow problem. In future work we will study the convergence of the algorithm in a nonlinear setting, and we will conduct more extensive benchmark studies.

Acknowledgment

This material is based upon work supported by the U.S. Department of Energy, Office of Science, Office of Advanced Scientific Computing Research (ASCR) under Contract DE-AC02-06CH11347 and by NSF through award CNS-1545046.

References

[1] C. Lemaréchal, “Lagrangian relaxation,” in Computational combinatorial optimization. Springer, 2001, pp. 112–156.