EXPONENTIAL DECAY OF SENSITIVITY IN GRAPH-STRUCTURED NONLINEAR PROGRAMS

SUNGHO SHIN*, MIHAI ANITESCU†, AND VICTOR M. ZAVALA‡

Abstract. We study solution sensitivity for nonlinear programs (NLPs) whose structure is induced by a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$. These graph-structured NLPs arise in many applications such as dynamic optimization, stochastic optimization, optimization with partial differential equations, and network optimization. We show that the sensitivity of the primal-dual solution at node $i \in \mathcal{V}$ against a data perturbation at node $j \in \mathcal{V}$ is bounded by $\Upsilon \rho^{d_{\mathcal{G}}(i,j)}$ for constants $\Upsilon > 0$ and $\rho \in (0, 1)$ and where $d_{\mathcal{G}}(i, j)$ is the distance between i and j on $\mathcal{G}$. In other words, the sensitivity of the solution decays exponentially with the distance to the perturbation point. This result, which we call exponential decay of sensitivity (EDS), holds under fairly standard assumptions used in classical NLP sensitivity theory: the strong second-order sufficiency condition and the linear independence constraint qualification. We also present conditions under which the constants (Υ, ρ) remain uniformly bounded; this allows us to characterize behavior for NLPs defined over subgraphs of infinite graphs (e.g., as those arising in problems with unbounded domains). Our results provide new insights on how perturbations propagate through the NLP graph and on how the problem formulation influences such propagation. Specifically, we provide empirical evidence that positive objective curvature and constraint flexibility tend to dampen propagation. The developments are illustrated with numerical examples.

Key words. sensitivity, nonlinear programming, graphs.

AMS subject classifications. 49Q12, 90C30, 90C35

1. Introduction.

We study the *graph-structured* nonlinear program (NLP):

$$\min_{\{x_i \in \mathbb{R}^{r_i}\}_{i \in \mathcal{V}}} \sum_{i \in \mathcal{V}} f_i(\{x_j\}_{j \in N_{\mathcal{G}}[i]}; \{p_j\}_{j \in N_{\mathcal{G}}[i]}) \tag{1.1a}$$

$$\text{s.t. } c_i^E(\{x_j\}_{j \in N_{\mathcal{G}}[i]}; \{p_j\}_{j \in N_{\mathcal{G}}[i]}) = 0,\ i \in \mathcal{V},\ (y_i^E) \tag{1.1b}$$

$$c_i^I(\{x_j\}_{j \in N_{\mathcal{G}}[i]}; \{p_j\}_{j \in N_{\mathcal{G}}[i]}) \geq 0,\ i \in \mathcal{V},\ (y_i^I), \tag{1.1c}$$

where the undirected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ is an ordered pair of the nonempty, strictly ordered node set $\mathcal{V}$ and the edge set $\mathcal{E} \subseteq \{\{i, j\} \subseteq \mathcal{V} : i \neq j\}$; $N_{\mathcal{G}}[i] := \{j \in \mathcal{V} : \{i, j\} \in \mathcal{E}\} \cup \{i\}$ denotes the closed neighborhood of $i \in \mathcal{V}$ on $\mathcal{G}$. For each node $i \in \mathcal{V}$, $x_i \in \mathbb{R}^{r_i}$ is the primal variable vector; $p_i \in \mathbb{R}^{l_i}$ is the data vector; $f_i : \prod_{j \in N_{\mathcal{G}}[i]} \mathbb{R}^{r_j} \times \prod_{j \in N_{\mathcal{G}}[i]} \mathbb{R}^{l_j} \to \mathbb{R}$ is the objective function; $c_i^E : \prod_{j \in N_{\mathcal{G}}[i]} \mathbb{R}^{r_j} \times \prod_{j \in N_{\mathcal{G}}[i]} \mathbb{R}^{l_j} \to \mathbb{R}^{m_i^E}$ is the equality constraint vector function; $c_i^I : \prod_{j \in N_{\mathcal{G}}[i]} \mathbb{R}^{r_j} \times \prod_{j \in N_{\mathcal{G}}[i]} \mathbb{R}^{l_j} \to \mathbb{R}^{m_i^I}$ is the inequality constraint vector function; and $y_i^E \in \mathbb{R}^{m_i^E}$ and $y_i^I \in \mathbb{R}^{m_i^I}$ are the dual variable vectors associated with (1.1b) and (1.1c), respectively.

The graph-structured NLP is composed of a collection of nodes $i \in \mathcal{V}$ wherein each node i has its own variables x_i, data p_i, objective function $f_i(\cdot)$, and constraint functions $c_i^E(\cdot), c_i^I(\cdot)$. The objective and constraint functions of node i depend on the variables $\{x_j\}_{j \in N_{\mathcal{G}}[i]}$ and data $\{p_j\}_{j \in N_{\mathcal{G}}[i]}$ from neighboring nodes $N_{\mathcal{G}}[i]$. In other

*Department of Chemical and Biological Engineering, University of Wisconsin-Madison, Madison, WI (sungho.shin@wisc.edu),

†Mathematics and Computer Science Division, Argonne National Laboratory, Argonne, IL and Department of Statistics, University of Chicago, Chicago, IL(anitescu@mcs.anl.gov),

‡Department of Chemical and Biological Engineering, University of Wisconsin-Madison, Madison, WI and Mathematics and Computer Science Division, Argonne National Laboratory, Argonne, IL (victor.zavala@wisc.edu)

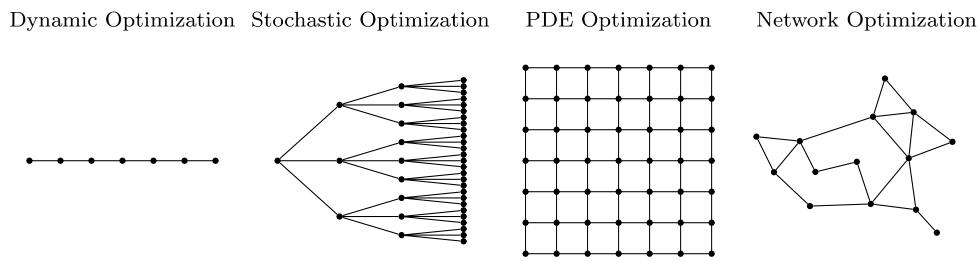

FIG. 1. *Illustration of graphs associated with various graph-structured NLPs.*

words, node i is coupled algebraically to its neighbors, and the topology of such connectivity is dictated by $\mathcal{G}$. To enable compact notation, we use the following definitions: $c_i(\cdot) := [c_i^E(\cdot); c_i^I(\cdot)]$; $y_i := [y_i^E; y_i^I]$; $z_i := [x_i; y_i]$; $m_i = m_i^E + m_i^I$; and $n_i = r_i + m_i$. Furthermore, we define $\boldsymbol{x} := \{x_i\}_{i\in\mathcal{V}}$; $\boldsymbol{y}^E := \{y_i^E\}_{i\in\mathcal{V}}$; $\boldsymbol{y}^I := \{y_i^I\}_{i\in\mathcal{V}}$; $\boldsymbol{y} := \{y_i\}_{i\in\mathcal{V}}$, $\boldsymbol{z} := \{z_i\}_{i\in\mathcal{V}}$; and $\boldsymbol{p} := \{p_i\}_{i\in\mathcal{V}}$.

A wide range of problems fall into our definition of graph-structured NLPs; examples include dynamic optimization (e.g., optimal control, long-term planning, and state estimation) [4, 5], multi-stage stochastic programs [26, 29], optimization with partial differential equations (PDEs) [7], and network optimization (e.g., energy networks, supply chains) [10, 11, 21, 36]. Typical graphs associated with such problems are illustrated in Figure 1. A few concrete examples will be discussed in Section 5.

Our work is motivated by the following question:

(Q1) *How does the primal-dual solution at node $i \in \mathcal{V}$ change when the data at node $j \in \mathcal{V}$ is perturbed?*

We provide an answer to this question by identifying conditions under which we can find *nodal* sensitivity coefficients $\{C_{ij} \in \mathbb{R}_{\geq 0}\}_{i,j\in\mathcal{V}}$ satisfying:

$$\|z_i^\dagger(\boldsymbol{p}) - z_i^\dagger(\boldsymbol{p}')\| \leq \sum_{j\in\mathcal{V}} C_{ij}\|p_j - p_j'\|, \quad i \in \mathcal{V}, \tag{1.2}$$

where $\boldsymbol{z}^\dagger(\boldsymbol{p})$ and $\boldsymbol{z}^\dagger(\boldsymbol{p}')$ are solutions of NLP (1.1) for data $\boldsymbol{p}$ and $\boldsymbol{p}'$, respectively. Furthermore, $\boldsymbol{p}$ and $\boldsymbol{p}'$ are perturbations of the base data $\boldsymbol{p}^\star$ (with associated base solution $\boldsymbol{z}^\star$).

Our main result (Theorem 3.5) shows that if: (i) the strong second order sufficiency condition (SSOSC) and the linear independence constraint qualification (LICQ) hold at $\boldsymbol{z}^\star$, and (ii) $\boldsymbol{p}$ and $\boldsymbol{p}'$ are sufficiently close to $\boldsymbol{p}^\star$, then (1.2) holds with $C_{ij} = \Upsilon\rho^{\lceil d_{\mathcal{G}}(i,j)/4-1\rceil_+}$. Here, $\Upsilon > 0$ and $\rho \in (0,1)$ are constants, $d_{\mathcal{G}}(i,j)$ is the graph distance between nodes i and j on $\mathcal{G}$, and $\lceil\cdot\rceil_+$ denotes the smallest non-negative integer that is greater than or equal to the argument. In other words, *solution sensitivity decays exponentially* with respect to the distance to the perturbation point. We call this property *exponential decay of sensitivity* (EDS). This result is a specialization of classical sensitivity results for general NLPs [9, 14, 15, 27, 28] to a graph-structured setting. Specifically, classical results establish an *overall* sensitivity coefficient C satisfying $\|\boldsymbol{z}^\dagger(\boldsymbol{p}) - \boldsymbol{z}^\dagger(\boldsymbol{p}')\| \leq C\|\boldsymbol{p} - \boldsymbol{p}'\|$, while here we establish *nodal* sensitivity coefficients $\{C_{ij}\}_{i,j\in\mathcal{V}}$ satisfying (1.2). Our results thus provide intuition into how perturbations propagate through the structure of the NLP.

The constants (Υ, ρ) play key roles in the magnitude and decay rate of sensitivity. We will see that these constants depend on the singular values of the Hessian of the Lagrangian; as such, we establish conditions under which the singular values remain uniformly bounded. This uniformity property will be particularly relevant when analyzing the sensitivity of NLPs that are defined over subgraphs of an infinite graph (a graph with an arbitrarily large domain). Such graphs can be used to analyze the limiting behavior of certain problem classes such as dynamic optimization problems over infinite horizons or of PDE optimization problems over unbounded domains. We show that (Υ, ρ) remain uniformly bounded under *uniform* boundedness conditions for graph degrees and second-order derivatives and under *uniform* SSOSC and LICQ. Unfortunately, these conditions are difficult to verify in practice (e.g., because the problem becomes arbitrarily large); accordingly, we establish sufficient conditions for uniform SSOSC and LICQ that can be verified in practice; in particular, we show that *block* SSOSC and LICQ conditions (assuming uniform SSOSC and LICQ hold over individual *blocks*) guarantee uniform SSOSC and LICQ for the entire NLP.

Question (Q1) has been recently addressed in specific settings such as nonlinear dynamic optimization [22,24,31,34] and graph-structured quadratic programs [30,33]. Our work generalizes such results. Addressing this question is crucial for understanding solution stability of a wide range of problem classes, for designing approximation schemes (often cast as parametric perturbations) [6, 13, 23, 32, 34], and for designing solution algorithms [24, 30, 31]. For instance, it has been recently shown that EDS plays a central role in assessing the impact of coarsening schemes [19,32] for dynamic optimization and in establishing convergence of overlapping Schwarz algorithms for graph-structured problems [24, 30, 31]. From an application stand-point, our results seek to provide new insights on how perturbations propagate through graphs and on how the problem formulation influences such propagation. Specifically, we provide empirical evidence that positive objective curvature and constraint flexibility tend to dampen propagation (promote sensitivity decay). Such insights can be used, for instance, to design systems that dampen (or magnify) perturbations or to identify system elements that are sensitive (or insensitive) to perturbations.

The paper is organized as follows: In Section 2 we present basic results for graph-induced matrix properties (these generalize the results in [12,33] and will serve as an analytical tool for sensitivity analysis). In Section 3 we present the main sensitivity results; specifically, we apply graph-induced matrix properties to classical NLP sensitivity theory to establish bounds on the nodal sensitivity coefficients in (1.2). In Section 4, we study uniform boundedness conditions for the sensitivity coefficients to analyze the NLPs with infinite graphs. Numerical results are provided in Section 5, followed by conclusions in Section 6.

Basic Notation: The set of real numbers and the set of integers are denoted by $\mathbb{R}$ and $\mathbb{I}$, respectively. We define $\mathbb{I}_A := \mathbb{I} \cap A$, where A is a set; $\mathbb{I}_{>0} := \mathbb{I} \cap (0, \infty)$; $\mathbb{I}_{\geq 0} := \mathbb{I} \cap [0, \infty)$; $\mathbb{R}_{>0} := (0, \infty)$; and $\mathbb{R}_{\geq 0} := [0, \infty)$. For $A \subseteq X$, $f : X \to Y$, and $x \in X$, where X and Y are linear spaces, $A + x := \{x' + x : x' \in A\}$ and $f(X) := \{f(x) \in Y : x \in X\}$. Vectors are treated as column vectors. We use the syntax: $[M_1; \cdots ; M_n] := [M_1^\top \cdots M_n^\top]^\top$; $\{M_i\}_{i \in U} := [M_{i_1}; \cdots ; M_{i_m}]$; $\{M_{i,j}\}_{i \in U, j \in V} := \{\{M_{i,j}^\top\}_{j \in V}^\top\}_{i \in U}$, where, $U = \{i_1 < \cdots < i_m\}$ and $V = \{j_1 < \cdots < j_n\}$ are strictly ordered sets. Furthermore, $v[i]$ is the i-th component of vector v; $M[i, j]$ is the (i, j)-th component of matrix M; $v[I] := \{v[i]\}_{i \in I}$; $M[I, J] := \{M[i, j]\}_{i \in I, j \in J}$. For a function $\phi : \mathbb{R}^n \to \mathbb{R}$ and variable vectors $y \in \mathbb{R}^p$, $z \in \mathbb{R}^q$, $\nabla^2_{yz}\phi(x) := \{\frac{\partial^2}{\partial y[i] \partial z[j]}\phi(x)\}_{i \in \mathbb{I}_{[1,p]}, j \in \mathbb{I}_{[1,q]}}$. For a vector function $\varphi : \mathbb{R}^n \to \mathbb{R}^m$ and a variable vector $w \in \mathbb{R}^s$, $\nabla_w \varphi(x) := \{\frac{\partial}{\partial w[j]}\varphi(x)[i]\}_{i\in\mathbb{I}_{[1,m]}, j\in\mathbb{I}_{[1,s]}}$. We use the shorthand notation $\nabla^2\phi(x) := \nabla^2_{xx}\phi(x)$, $\nabla\varphi(x) := \nabla_x\varphi(x)$. Vector 2-norms and induced 2-norms of matrices are denoted by $\|\cdot\|$. For matrices A and B, $A \succ (\succeq) B$ indicates that $A - B$ is positive (semi)-definite while $A > (\geq) B$ denotes a componentwise inequality. The identity matrix is denoted as $\boldsymbol{I}$ and the zero matrix or vector are denoted as $\mathbf{0}$. Specific notations will be introduced at the first appearance.

Remark 1.1. In some applications (e.g., energy networks), we might encounter variables, data, objectives, and constraints defined over edges (not explicitly expressed in Problem (1.1)). Such information can be captured within "super-nodes" that encapsulate edges (this is possible because the formulation allows for nodes with different numbers of variables and constraints). Alternatively, one may treat edges in the graph as nodes and rewrite the problem with a newly defined lifted graph $\widetilde{\mathcal{G}} = (\widetilde{\mathcal{V}}, \widetilde{\mathcal{E}})$, where $\widetilde{\mathcal{V}} := \mathcal{V} \cup \mathcal{E}$ and $\widetilde{\mathcal{E}} := \{\{i, e\} : i \in \mathcal{V},\ i \in e,\ e \in \mathcal{E}\}$ (an order needs to be assigned for $\widetilde{\mathcal{V}}$).

2. Graph-Induced Matrix Properties.

This section derives basic properties of *graph-induced matrices*. The results in this section will be crucial in deriving the sensitivity results of interest. Properties of graph-induced *positive definite* matrices are reported in [12, 33]; here, we establish properties for general (non-symmetric and indefinite) matrices. We begin by introducing the notion of distance on graphs and establish basic properties for such distance.

DEFINITION 2.1 (Graph Distance and Diameter). *The distance $d_{\mathcal{G}}(i,j)$ between nodes $i, j \in \mathcal{V}$ on graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ is the number of edges in the shortest path connecting them. Furthermore, the diameter $D_{\mathcal{G}}$ of $\mathcal{G}$ is the largest distance between any pair of nodes in V.*

PROPOSITION 2.2. *The distance $d_{\mathcal{G}} : V \times V \to \mathbb{I}_{\geq 0}$ is a metric on V; that is, (a) $d_{\mathcal{G}}(i,j) \geq 0$ for any $i, j \in \mathcal{V}$; (b) $i = j$ if and only if $d_{\mathcal{G}}(i,j) = 0$; (c) $d_{\mathcal{G}}(i,j) = d_{\mathcal{G}}(j,i)$ for any $i, j \in \mathcal{V}$; (d) $d_{\mathcal{G}}(i,j) \leq d_{\mathcal{G}}(i,k) + d_{\mathcal{G}}(k,j)$ for any $i, j, k \in \mathcal{V}$.*

The proof of this result is straightforward and is thus omitted. We now introduce the concept of graph-induced *matrix bandwidth*.

DEFINITION 2.3 (Graph-Induced Matrix Bandwidth). *Consider a matrix $X \in \mathbb{R}^{m\times n}$, a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, and index sets $\mathcal{I} = \{I_i\}_{i\in\mathcal{V}}$, $\mathcal{J} = \{J_i\}_{i\in\mathcal{V}}$ that partition*[1] *$\mathbb{I}_{[1,m]}$ and $\mathbb{I}_{[1,n]}$, respectively. Matrix X is said to have bandwidth B induced by an ordered triple $(\mathcal{G}, \mathcal{I}, \mathcal{J})$, if B is the smallest nonnegative integer such that $X_{[i][j]} = \mathbf{0}$ for any $i, j \in \mathcal{V}$ with $d_{\mathcal{G}}(i,j) > B$, where $X_{[i][j]} := X[I_i, J_j]$.*

We refer to $X_{[i][j]}$ as the $[i][j]$-block of matrix X. The bandwidth defined above is a generalization of the standard notion of matrix bandwidth [16, Section 1.2.1]. If the matrix $X \in \mathbb{R}^{n\times n}$ is square, $\mathcal{V} = \mathbb{I}_{[1,n]}$, $\mathcal{E} = \{\{i, i+1\}\}_{i=1}^{n-1}$, and $\mathcal{I} = \mathcal{J} = \{\{i\}\}_{i=1}^{n}$, then the graph-induced matrix bandwidth reduces to the standard definition of matrix bandwidth.

Definition 2.3 enables a formal definition of graph-induced matrices; specifically, a graph-induced matrix is a matrix X that has a triple $(\mathcal{G}, \mathcal{I}, \mathcal{J})$ such that the bandwidth B of X, induced by $(\mathcal{G}, \mathcal{I}, \mathcal{J})$, is much smaller than the diameter $D_{\mathcal{G}}$ of $\mathcal{G}$ (i.e., $B \ll$

[1] In this paper, we call a family $\{X_1, \cdots, X_k\}$ of subsets of X to be a partition if $\bigcup_{k=1}^K X_k = X$ and $X_1, \cdots, X_k$ are disjoint; here, *we allow X_k to be empty sets*. Note that this differs from the standard definition of a partition, where the nonempty nature of X_k is enforced. This modification allows us to handle nodes with empty variables, constraints, or data.

$D_{\mathcal{G}}$). This corresponds to the notion of a *block-banded matrix*. Block-diagonal matrices whose blocks are defined by $\mathcal{I}, \mathcal{J}$ (including identity matrices and zero matrices) always have bandwidth of zero (by Proposition 2.2(b)). We now state basic properties of the matrix bandwidth.

LEMMA 2.4. *Consider $X \in \mathbb{R}^{m\times n}$ with bandwidth no greater than B_X induced by $(\mathcal{G},\mathcal{I},\mathcal{J})$; we have that: (a) $X^\top$ has bandwidth not greater than B_X induced by $(\mathcal{G},\mathcal{J},\mathcal{I})$; (b) if $Y \in \mathbb{R}^{m\times n}$ has bandwidth not greater than B_Y induced by $(\mathcal{G},\mathcal{I},\mathcal{J})$, then $X+Y$ has bandwidth not greater than $\max(B_X, B_Y)$ induced by $(\mathcal{G},\mathcal{I},\mathcal{J})$; (c) if $W \in \mathbb{R}^{n\times k}$ has bandwidth not greater than B_W induced by $(\mathcal{G},\mathcal{J},\mathcal{K})$, then XW has bandwidth not greater than $B_X + B_W$ induced by $(\mathcal{G},\mathcal{I},\mathcal{K})$.*

Proof of (a). We have that $(X^\top)_{[i][j]} = (X^\top)[J_i][I_j] = (X[I_j][J_i])^\top = (X_{[j][i]})^\top$. From the assumption that X has bandwidth not greater than B_X and Proposition 2.2(c), $(X^\top)_{[i][j]} = \mathbf{0}$ if $d_{\mathcal{G}}(i,j) > B_X$; therefore, X has bandwidth not greater than B_X, and induced by $(\mathcal{G},\mathcal{J},\mathcal{I})$. ◻

Proof of (b). We have that $X_{[i][j]} = \mathbf{0}$ and $Y_{[i][j]} = \mathbf{0}$ if $d_{\mathcal{G}}(i,j) > \max(B_X, B_Y)$, which yields $(X+Y)_{[i][j]} = \mathbf{0}$ if $d_{\mathcal{G}}(i,j) > \max(B_X, B_Y)$. Thus, $X+Y$ has bandwidth not greater than $\max(B_X, B_Y)$, and induced by $(\mathcal{G},\mathcal{I},\mathcal{J})$. ◻

Proof of (c). If $d_{\mathcal{G}}(i,j) > B_X + B_W$, from Proposition 2.2(d) we have that for any $k \in \mathcal{V}$, $d_{\mathcal{G}}(i,k) > B_X$ or $d_{\mathcal{G}}(j,k) > B_W$. Thus, if $d_{\mathcal{G}}(i,j) > B_X + B_W$, we have $(XW)_{[i][j]} = \sum_{k\in\mathcal{V}} X_{[i][k]} W_{[k][j]} = \mathbf{0}$ (where the first equality comes from the block matrix multiplication law and the second equality comes from observing that $d_{\mathcal{G}}(i,k) > B_X$ or $d_{\mathcal{G}}(j,k) > B_W$). Therefore, XW has bandwidth not greater than $B_X + B_W$, and induced by $(\mathcal{G},\mathcal{I},\mathcal{K})$. ◻

Lemma 2.4 implies that graph-induced properties of a matrix are preserved under transposition, addition, and multiplication (as long as the associated index sets are compatible). Using Lemma 2.4, we can establish properties for the inverse of a graph-induced matrix (this is the main result of this section).

THEOREM 2.5. *Consider a nonsingular matrix $X \in \mathbb{R}^{n\times n}$ with bandwidth not greater than $B_X \geq 1$ induced by $(\mathcal{G},\mathcal{K},\mathcal{P})$, $Y \in \mathbb{R}^{n\times m}$ with bandwidth not greater than B_Y induced by $(\mathcal{G},\mathcal{K},\mathcal{J})$, and $W \in \mathbb{R}^{\ell\times n}$ with bandwidth not greater than B_W induced by $(\mathcal{G},\mathcal{I},\mathcal{P})$; for constants $\overline{\sigma}_X \geq \overline{\sigma}(X)$, $\overline{\sigma}_Y \geq \overline{\sigma}(Y)$, $\overline{\sigma}_W \geq \overline{\sigma}(W)$, and $0 < \underline{\sigma}_X \leq \underline{\sigma}(X)$ (where $\overline{\sigma}(\cdot)$ and $\underline{\sigma}(\cdot)$ denote the largest and smallest non-trivial singular values of the argument),[2] the following holds:*

$$\|(WX^{-1}Y)_{[i][j]}\| \leq \frac{\overline{\sigma}_X\overline{\sigma}_Y\overline{\sigma}_W}{\underline{\sigma}_X^2}\left(\frac{\overline{\sigma}_X^2 - \underline{\sigma}_X^2}{\overline{\sigma}_X^2 + \underline{\sigma}_X^2}\right)^{\left\lceil \frac{d_{\mathcal{G}}(i,j) - B_X - B_Y - B_W}{2B_X}\right\rceil_+}, \quad i,j \in \mathcal{V}, \tag{2.1}$$

where $(WX^{-1}Y)_{[i][j]} := (WX^{-1}Y)[I_i, J_j]$ and $\lceil\cdot\rceil_+$ is the smallest non-negative integer that is greater than or equal to the argument.

Proof. By definition of singular values and the assumptions on $\overline{\sigma}_X$, $\overline{\sigma}_Y$, $\overline{\sigma}_W$, and $\underline{\sigma}_X$, we have $\underline{\sigma}_X^2\boldsymbol{I} \preceq \underline{\sigma}(X)^2\boldsymbol{I} \preceq XX^\top \preceq \overline{\sigma}(X)^2\boldsymbol{I} \preceq \overline{\sigma}_X^2\boldsymbol{I}$. From this, one can obtain:

$$\frac{\underline{\sigma}_X^2 - \overline{\sigma}_X^2}{\underline{\sigma}_X^2 + \overline{\sigma}_X^2}\boldsymbol{I} \preceq \boldsymbol{I} - \frac{2}{\underline{\sigma}_X^2 + \overline{\sigma}_X^2}XX^\top \preceq \frac{-\underline{\sigma}_X^2 + \overline{\sigma}_X^2}{\underline{\sigma}_X^2 + \overline{\sigma}_X^2}\boldsymbol{I}. \tag{2.2}$$

[2]Non-trivial in the sense that we exclude the singular values that are trivially zero due to the nonsquare size of the matrix.

By Lemma 2.4(c), $XX^\top$ has bandwidth not greater than $2B_X$ induced by $(\mathcal{G},\mathcal{K},\mathcal{K})$. Furthermore, $XX^\top$ is nonsingular (from the nonsingularity of X). These observations imply that:

$$WX^{-1}Y = \frac{2}{\underline{\sigma}_X^2+\overline{\sigma}_X^2} WX^\top \left(\frac{2}{\underline{\sigma}_X^2+\overline{\sigma}_X^2} XX^\top \right)^{-1} Y \tag{2.3a}$$

$$= \frac{2}{\underline{\sigma}_X^2+\overline{\sigma}_X^2} WX^\top \left(\boldsymbol{I} - (\boldsymbol{I} - \frac{2}{\underline{\sigma}_X^2+\overline{\sigma}_X^2} XX^\top) \right)^{-1} Y \tag{2.3b}$$

$$= \frac{2}{\underline{\sigma}_X^2+\overline{\sigma}_X^2} WX^\top \left(\sum_{q=0}^{\infty} \left(\boldsymbol{I} - \frac{2}{\underline{\sigma}_X^2+\overline{\sigma}_X^2} XX^\top \right)^q \right) Y, \tag{2.3c}$$

$$= \frac{2}{\underline{\sigma}_X^2+\overline{\sigma}_X^2} \sum_{q=0}^{\infty} WX^\top \left(\boldsymbol{I} - \frac{2}{\underline{\sigma}_X^2+\overline{\sigma}_X^2} XX^\top \right)^q Y. \tag{2.3d}$$

The second equality follows from a simple algebraic manipulation, the third equality follows from [20, Corollary 5.6.16] and (2.2), and the last equality follows from the fact that the series in (2.3d) converges (due to (2.2)). Furthermore, from Lemma 2.4, we see that $WX^\top \left(\boldsymbol{I} - \frac{2}{\underline{\sigma}_X^2+\overline{\sigma}_X^2} XX^\top \right)^q Y$ has bandwidth not greater than $(2q+1)B_X + B_Y + B_W$ and induced by $(\mathcal{G},\mathcal{I},\mathcal{J})$. By extracting submatrices defined by the row index I_i and the column index J_j from (2.3), one can obtain:

$$(WX^{-1}Y)_{[i][j]} = \frac{2}{\underline{\sigma}_X^2+\overline{\sigma}_X^2} \sum_{q=q_0(i,j)}^{\infty} \left(WX^\top \left(\boldsymbol{I} - \frac{2XX^\top}{\underline{\sigma}_X^2+\overline{\sigma}_X^2} \right)^q Y \right)_{[i][j]}$$

where $q_0(i,j) := \left\lceil \frac{d_{\mathcal{G}}(i,j)-B_X-B_Y-B_W}{2B_X} \right\rceil_+$; the summation over $q = 0, \cdots, q_0(i,j)-1$ is zero; because such q satisfy $(2q+1)B_X + B_Y + B_W < d_{\mathcal{G}}(i,j)$, thus

$$\left(WX^\top \left(\boldsymbol{I} - \frac{2}{\underline{\sigma}_X^2+\overline{\sigma}_X^2} XX^\top \right)^q Y \right)_{[i][j]} = \boldsymbol{0}.$$

Using the triangle inequality and the fact that the matrix norm of a submatrix is smaller than that of the original matrix,

$$\begin{aligned} \|(WX^{-1}Y)_{[i][j]}\| &\le \frac{2}{\underline{\sigma}_X^2+\overline{\sigma}_X^2} \sum_{q=q_0(i,j)}^{\infty} \left\| WX^\top \left(\boldsymbol{I} - \frac{2XX^\top}{\underline{\sigma}_X^2+\overline{\sigma}_X^2} \right)^q Y \right\| \\ &\le \frac{2}{\underline{\sigma}_X^2+\overline{\sigma}_X^2} \sum_{q=q_0(i,j)}^{\infty} \overline{\sigma}_W \overline{\sigma}_X \left(\frac{\overline{\sigma}_X^2-\underline{\sigma}_X^2}{\overline{\sigma}_X^2+\underline{\sigma}_X^2} \right)^q \overline{\sigma}_Y \\ &\le \frac{\overline{\sigma}_X\overline{\sigma}_Y\overline{\sigma}_W}{\underline{\sigma}_X^2} \left(\frac{\overline{\sigma}_X^2-\underline{\sigma}_X^2}{\overline{\sigma}_X^2+\underline{\sigma}_X^2} \right)^{\left\lceil \frac{d_{\mathcal{G}}(i,j)-B_X-B_Y-B_W}{2B_X} \right\rceil_+}. \end{aligned} \tag{2.4}$$

The second inequality follows from the submultiplicativity of the matrix norm and (2.2); and the last inequality follows from the summation of geometric series. Therefore, (2.1) is obtained. ∎

The result indicates that the norm of the $[i][j]$-block of $WX^{-1}Y$ *decays exponentially* with respect to the distance between nodes i and j on $\mathcal{G}$; the decay rate becomes faster (smaller) as the condition number $\overline{\sigma}(X)/\underline{\sigma}(X)$ of X decreases; and the decay rate becomes faster as the bandwidths B_X, B_Y, and B_W decrease. This property will be key in establishing EDS for the graph-structured NLP (1.1) and hints at the fact that EDS arises from connectivity induced by the graph (at the linear algebra level).

Theorem 2.5 is a generalization of [33, Theorem 1] (which assumes positive definiteness of X and $Y = W = \boldsymbol{I}$). We also note that [12] has studied exponential decay of the components of the inverse of banded matrices. Specifically, in [12, Theorem 2.4], exponential decay for indefinite banded matrices (with the standard definition of bandwidth) is established. Furthermore, in [12, Proposition 5.1], a less general form of Theorem 2.5 is presented; however, graph-induced matrices are not formally introduced, and only positive definite matrices are considered. Theorem 2.5 generalizes these results by introducing the notion of graph-induced matrices and by considering non-symmetric and indefinite matrices.

3. Exponential Decay of Sensitivity.

This section aims to provide an answer to question (Q1). The sketch of our analysis is as follows: we invoke classical results of NLP sensitivity theory [14, 28] to obtain an explicit representation for the one-sided directional derivative of the primal-dual solution mapping w.r.t. to the data; the representation involves the inverse of a graph-induced matrix. The results from the previous section are then applied to this representation to establish bounds on the nodal sensitivity coefficients. Finally, the one-sided directional derivative is integrated over the line segment between a pair of data points that are within the neighborhood of the base data to obtain the result in the form of (1.2). This yields explicit expressions for (Υ, ρ).

To enable compact notation, we introduce the following definitions:

$$
\begin{aligned}
\boldsymbol{f}(\boldsymbol{x};\boldsymbol{p}) &:= \sum_{i\in\mathcal{V}} f_i(\{x_j\}_{j\in N_\mathcal{G}[i]};\{p_j\}_{j\in N_\mathcal{G}[i]});\\
\boldsymbol{c}^E(\boldsymbol{x};\boldsymbol{p}) &:= \{c_i^E(\{x_j\}_{j\in N_\mathcal{G}[i]};\{p_j\}_{j\in N_\mathcal{G}[i]})\}_{i\in\mathcal{V}};\\
\boldsymbol{c}^I(\boldsymbol{x};\boldsymbol{p}) &:= \{c_i^I(\{x_j\}_{j\in N_\mathcal{G}[i]};\{p_j\}_{j\in N_\mathcal{G}[i]})\}_{i\in\mathcal{V}};\\
\boldsymbol{c}(\boldsymbol{x};\boldsymbol{p}) &:= \{c_i(\{x_j\}_{j\in N_\mathcal{G}[i]};\{p_j\}_{j\in N_\mathcal{G}[i]})\}_{i\in\mathcal{V}};
\end{aligned}
$$

$\boldsymbol{r} = \sum_{i\in\mathcal{V}} r_i$; $\boldsymbol{m}^E = \sum_{i\in\mathcal{V}} m_i^E$; $\boldsymbol{m}^I = \sum_{i\in\mathcal{V}} m_i^I$; $\boldsymbol{m} = \sum_{i\in\mathcal{V}} m_i$; $\boldsymbol{n} = \sum_{i\in\mathcal{V}} n_i$; $\boldsymbol{l} = \sum_{i\in\mathcal{V}} l_i$. Boldface symbols are used whenever a variable or a function is associated with more than one node. Using these definitions, (1.1) can be expressed as a general parametric NLP of the form:

$$
\begin{aligned}
P(\boldsymbol{p}) :\ \min_{\boldsymbol{x}}\ & \boldsymbol{f}(\boldsymbol{x};\boldsymbol{p})\\
\text{s.t. } & \boldsymbol{c}^E(\boldsymbol{x};\boldsymbol{p}) = \boldsymbol{0},\ (\boldsymbol{y}^E)\\
& \boldsymbol{c}^I(\boldsymbol{x};\boldsymbol{p}) \geq \boldsymbol{0},\ \ (\boldsymbol{y}^I).
\end{aligned}
$$

We denote this problem as $P(\boldsymbol{p})$; its Lagrange function $\mathcal{L} : \mathbb{R}^{\boldsymbol{n}} \times \mathbb{R}^{\boldsymbol{l}} \to \mathbb{R}$ is given by $\mathcal{L}(\boldsymbol{z};\boldsymbol{p}) := \boldsymbol{f}(\boldsymbol{x};\boldsymbol{p}) - \boldsymbol{y}^\top \boldsymbol{c}(\boldsymbol{x};\boldsymbol{p})$.

3.1. Preliminaries.

We use $\boldsymbol{z}^\star \in \mathbb{R}^{\boldsymbol{n}}$ to denote the primal-dual base solution of $P(\boldsymbol{p}^\star)$. We denote the primal and dual components of $\boldsymbol{z}^\star = \{z_i^\star\}_{i\in\mathcal{V}} = \{[x_i^\star; y_i^\star]\}_{i\in\mathcal{V}}$ as $\boldsymbol{x}^\star := \{x_i^\star\}_{i\in\mathcal{V}}$ and $\boldsymbol{y}^\star = \{y_i^\star\}_{i\in\mathcal{V}}$, respectively. We say that $\boldsymbol{z}^\star$ is a primal-dual solution if $\boldsymbol{x}^\star$ satisfies the first-order optimality conditions with Lagrange multiplier $\boldsymbol{y}^\star$ (see [25]).

We now make key assumptions that are necessary to establish our main sensitivity result.

Assumption 3.1 (Twice Continuous Differentiability of Functions). *The functions $\boldsymbol{f} : \mathbb{R}^{\boldsymbol{r}} \times \mathbb{R}^{\boldsymbol{l}} \to \mathbb{R}$ and $\boldsymbol{c} : \mathbb{R}^{\boldsymbol{r}} \times \mathbb{R}^{\boldsymbol{l}} \to \mathbb{R}^{\boldsymbol{m}}$ are twice continuously differentiable in the neighborhood of $[\boldsymbol{x}^\star; \boldsymbol{p}^\star]$.*

Assumption 3.2 (Regularity of Solution). *The base solution $\boldsymbol{z}^\star$ satisfies SSOSC and LICQ.*

We recall that SSOSC requires positive definiteness of the reduced Hessian of the Lagrangian at $\boldsymbol{z}^\star$. The reduced Hessian is the Hessian projected on the null space defined by equality constraints and active inequality constraints with nonzero duals. LICQ requires that the constraint Jacobian defined by equality and active inequality constraints are linearly independent at $\boldsymbol{z}^\star$. These requirements are stated formally as:

$$\text{(SSOSC)} \qquad \mathrm{ReH}(\nabla^2_{\boldsymbol{xx}}\mathcal{L}(\boldsymbol{z}^\star;\boldsymbol{p}^\star), \nabla_{\boldsymbol{x}}\boldsymbol{c}(\boldsymbol{x}^\star;\boldsymbol{p}^\star)[\mathcal{A}^0(\boldsymbol{p}^\star), :]) \succ \boldsymbol{0}$$

$$\text{(LICQ)} \qquad \underline{\sigma}(\nabla_{\boldsymbol{x}}\boldsymbol{c}(\boldsymbol{x}^\star;\boldsymbol{p}^\star)[\mathcal{A}^1(\boldsymbol{p}^\star), :]) > 0.$$

Here, $\mathrm{ReH}(H, A) := Z^\top H Z$ is the reduced Hessian, where Z is a null-space matrix of A, and

$$\mathcal{A}^0(\boldsymbol{p}^\star) := \mathcal{A}^E \cup \{i \in \mathcal{A}^I : \boldsymbol{c}(\boldsymbol{x}^\star)[i] = 0,\ \boldsymbol{y}^\star[i] \neq 0\} \tag{3.1}$$

$$\mathcal{A}^1(\boldsymbol{p}^\star) := \mathcal{A}^E \cup \{i \in \mathcal{A}^I : \boldsymbol{c}(\boldsymbol{x}^\star)[i] = 0\}, \tag{3.2}$$

where $\mathcal{A}^E$ and $\mathcal{A}^I$ are the set of equality and inequality constraint indices within $\mathbb{I}_{[1,\boldsymbol{m}]}$, respectively.

SSOSC and LICQ are standard assumptions used in NLP sensitivity theory. For instance, Assumption 3.2 guarantees strong regularity of the generalized equation (GE) representation of the first-order optimality conditions of (1.1) at $z^\star$ [28]. Strong regularity is then used to establish properties for the solution mapping for the parametric NLP $P(\boldsymbol{p})$. In what follows, we refer to $\boldsymbol{p}^\star$ as the base data and $\boldsymbol{z}^\star$ as the base solution.

Lemma 3.3. *Under Assumptions 3.1, 3.2, there exist neighborhoods $\mathbb{P} \subseteq \mathbb{R}^{\boldsymbol{l}}$ of $\boldsymbol{p}^\star$ and $\mathbb{Z} \subseteq \mathbb{R}^{\boldsymbol{n}}$ of $\boldsymbol{z}^\star$ and a continuous function $\boldsymbol{z}^\dagger : \mathbb{P} \to \mathbb{Z}$ such that $\boldsymbol{z}^\dagger(\boldsymbol{p})$ is a primal-dual solution of $P(\boldsymbol{p})$ that satisfies SSOSC and LICQ. Furthermore, for any $\boldsymbol{q} := \{q_i\}_{i\in\mathcal{V}} \in \mathbb{R}^{\boldsymbol{l}}$, the one-sided directional derivative of $\boldsymbol{z}^\dagger(\cdot)$ is given by:*

$$D_{\boldsymbol{q}}\boldsymbol{z}^\dagger(\boldsymbol{p}) := \lim_{h\searrow 0} \frac{\boldsymbol{z}^\dagger(\boldsymbol{p}+\boldsymbol{q}h) - \boldsymbol{z}^\dagger(\boldsymbol{p})}{h};$$

with $D_{\boldsymbol{q}}\boldsymbol{x}^\dagger(\boldsymbol{p}) = \boldsymbol{\xi}_i^\dagger(\boldsymbol{p},\boldsymbol{q})$, $D_{\boldsymbol{q}}\boldsymbol{y}^\dagger(\boldsymbol{p},\boldsymbol{q}) = \boldsymbol{\eta}^\dagger(\boldsymbol{p},\boldsymbol{q})$. We also have that $\boldsymbol{\xi}^\dagger(\boldsymbol{p},\boldsymbol{q})$ and $\boldsymbol{\eta}^\dagger(\boldsymbol{p},\boldsymbol{q})[\mathcal{A}^1(\boldsymbol{p})]$ is a primal-dual solution of the quadratic program:

$$QP(\boldsymbol{p},\boldsymbol{q}): \quad \min_{\boldsymbol{\xi}} \frac{1}{2}\boldsymbol{\xi}^\top \boldsymbol{Q}(\boldsymbol{p})\boldsymbol{\xi} + \boldsymbol{\xi}^\top \boldsymbol{S}(\boldsymbol{p})\boldsymbol{q}, \tag{3.3a}$$

$$\text{s.t. } \boldsymbol{A}(\boldsymbol{p})[i,:]^\top \boldsymbol{\xi} + \boldsymbol{B}(\boldsymbol{p})[i,:]\boldsymbol{q} = 0,\ i \in \mathcal{A}^0(\boldsymbol{p}),\ (\boldsymbol{\eta}[i]) \tag{3.3b}$$

$$\boldsymbol{A}(\boldsymbol{p})[i,:]^\top \boldsymbol{\xi} + \boldsymbol{B}(\boldsymbol{p})[i,:]\boldsymbol{q} \geq 0,\ i \in \mathcal{A}^1(\boldsymbol{p}) \setminus \mathcal{A}^0(\boldsymbol{p}),\ (\boldsymbol{\eta}[i]) \tag{3.3c}$$

and $\boldsymbol{\eta}^\dagger(\boldsymbol{p},\boldsymbol{q})[\mathbb{I}_{[1,\boldsymbol{m}]} \setminus \mathcal{A}^1(\boldsymbol{p})] = \boldsymbol{0}$ (i.e., the free dual variables are fixed to zero), where:

$$\mathcal{A}^0(\boldsymbol{p}) := \mathcal{A}^E \cup \{i \in \mathcal{A}^I : \boldsymbol{c}(\boldsymbol{x}^\dagger(\boldsymbol{p}))[i] = 0,\ \boldsymbol{y}^\dagger(\boldsymbol{p})[i] \neq 0\} \tag{3.4a}$$

$$\text{(3.4b)} \qquad \mathcal{A}^1(\boldsymbol{p}) := \mathcal{A}^E \cup \{i \in \mathcal{A}^I : \boldsymbol{c}(\boldsymbol{x}^\dagger(\boldsymbol{p}))[i] = 0\}$$

$$\text{(3.4c)} \qquad \boldsymbol{Q}(\boldsymbol{p}) := \nabla_{\boldsymbol{xx}}\mathcal{L}(\boldsymbol{z}^\dagger(\boldsymbol{p});\boldsymbol{p}); \quad \boldsymbol{S}(\boldsymbol{p}) := \nabla_{\boldsymbol{xp}}\mathcal{L}(\boldsymbol{z}^\dagger(\boldsymbol{p});\boldsymbol{p})$$

$$\text{(3.4d)} \qquad \boldsymbol{A}(\boldsymbol{p}) := \nabla_{\boldsymbol{x}}\boldsymbol{c}(\boldsymbol{x}^\dagger(\boldsymbol{p});\boldsymbol{p}); \quad \boldsymbol{B}(\boldsymbol{p}) := \nabla_{\boldsymbol{p}}\boldsymbol{c}(\boldsymbol{x}^\dagger(\boldsymbol{p});\boldsymbol{p}).$$

Moreover, a unique primal-dual solution of $QP(\boldsymbol{p},\boldsymbol{q})$ exists for any $\boldsymbol{p} \in \mathbb{P}$ and $\boldsymbol{q} \in \mathbb{R}^{\boldsymbol{l}}$; thus, $\boldsymbol{\xi}^\dagger(\cdot,\cdot)$ and $\boldsymbol{\eta}^\dagger(\cdot,\cdot)$ are well-defined.

Proof. The results in [14, Theorem 2G.8] ensure semidifferentiability (which guarantees continuity) of the solution of the first-order optimality conditions for $P(\boldsymbol{p})$ over a certain neighborhood of $\boldsymbol{p}^\star$. This result is established by using the GE representation of the first-order conditions. Furthermore, [14, Theorem 2G.9] establishes that over a certain neighborhood $\mathbb{P}$, the solution mapping of the GE satisfies SSOSC and LICQ; that is, within $\mathbb{P}$, the solution mapping for the GE is the solution mapping for $P(\boldsymbol{p})$ at which SSOSC and LICQ are satisfied. Moreover, by [14, Theorem 2G.8], the one-sided directional derivative of the solution mapping for the GE (which exists for any direction $\boldsymbol{q} \in \mathbb{R}^{\boldsymbol{l}}$ by semidifferentiability [14, Theorem 2D.1]) can be evaluted by using the linearized GE. The linear GE are the first-order optimality conditions for $QP(\boldsymbol{p},\boldsymbol{q})$ (see [14, Equation (35)]); here, the first-order conditions are necessary and sufficienct conditions for the optimality due to the convexity $QP(\boldsymbol{p},\boldsymbol{q})$ (guaranteed by SSOSC and LICQ of $\boldsymbol{z}^\star$ for the original problem). As such, the one-sided directional derivative of the solution mapping for $P(\boldsymbol{p})$ can be evaluated by solving $QP(\boldsymbol{p},\boldsymbol{q})$. Finally, the strong regularity of the GE at $z^\star$ (obtained under SSOSC and LICQ) guarantees that there exists a unique solution of the linearized GE [28], which in turn guarantees the existence of a unique solution of $QP(\boldsymbol{p},\boldsymbol{q})$. ∎

Under Lemma 3.3, the rate of change $D_{\boldsymbol{q}}\boldsymbol{z}^\dagger(\boldsymbol{p})$ of the primal-dual solution of $P(\boldsymbol{p})$ (for a given direction $\boldsymbol{q}$) can be quantified by using the solution of $QP(\boldsymbol{p},\boldsymbol{q})$. For given $\boldsymbol{p}$ and $\boldsymbol{q}$, the parameters in $QP(\boldsymbol{p},\boldsymbol{q})$ can be evaluated explicitly and thus its solution can be calculated; as such, Lemma 3.3 provides a computational procedure to evaluate primal-dual solution sensitivity.

The quadratic program $QP(\boldsymbol{p},\boldsymbol{q})$ plays a central role in our analysis and we thus examine its properties in more detail. The first-order conditions of this problem are:

$$\text{(3.5a)} \qquad \boldsymbol{Q}(\boldsymbol{p})\boldsymbol{\xi} + \boldsymbol{S}(\boldsymbol{p})\boldsymbol{q} + \boldsymbol{A}^\top\boldsymbol{\eta} = \boldsymbol{0}$$

$$\text{(3.5b)} \qquad \boldsymbol{A}(\boldsymbol{p})[i,:]\boldsymbol{\xi} + \boldsymbol{B}(\boldsymbol{p})[i,:]\boldsymbol{q} = 0,\ i \in \mathcal{A}^0(\boldsymbol{p})$$

$$\text{(3.5c)} \qquad \boldsymbol{A}(\boldsymbol{p})[i,:]\boldsymbol{\xi} + \boldsymbol{B}(\boldsymbol{p})[i,:]\boldsymbol{q} \geq 0,\ i \in \mathcal{A}^1(\boldsymbol{p}) \setminus \mathcal{A}^0(\boldsymbol{p})$$

$$\text{(3.5d)} \qquad \boldsymbol{\eta}[i] \geq 0,\ i \in \mathcal{A}^1(\boldsymbol{p}) \setminus \mathcal{A}^0(\boldsymbol{p})$$

$$\text{(3.5e)} \qquad \left(\boldsymbol{A}(\boldsymbol{p})[i,:]\boldsymbol{\xi} + \boldsymbol{B}(\boldsymbol{p})[i,:]\boldsymbol{q}\right)\boldsymbol{\eta}[i] = 0,\ i \in \mathcal{A}^1(\boldsymbol{p}) \setminus \mathcal{A}^0(\boldsymbol{p}).$$

Under SSOSC and LICQ for $P(\boldsymbol{p}^\star)$ at $\boldsymbol{z}^\star$, these conditions are necessary and sufficient for any solution of $QP(\boldsymbol{p},\boldsymbol{q})$. From the complementarity condition (3.5e), one can observe that there exists $\mathcal{A}^0(\boldsymbol{p}) \subseteq \mathcal{A}'(\boldsymbol{p},\boldsymbol{q}) \subseteq \mathcal{A}^1(\boldsymbol{p})$ such that:

$$\text{(3.6a)} \qquad \boldsymbol{A}(\boldsymbol{p})[i,:]\boldsymbol{\xi} + \boldsymbol{B}(\boldsymbol{p})[i,:]\boldsymbol{q} = 0,\ i \in \mathcal{A}'(\boldsymbol{p},\boldsymbol{q})$$

$$\text{(3.6b)} \qquad \boldsymbol{\eta}[i] = 0,\ i \in \mathcal{A}^1(\boldsymbol{p}) \setminus \mathcal{A}'(\boldsymbol{p},\boldsymbol{q}).$$

are satisfied at $\boldsymbol{\xi}^\dagger(\boldsymbol{p},\boldsymbol{q})$, $\boldsymbol{\eta}^\dagger(\boldsymbol{p},\boldsymbol{q})$. Thus, from (3.5), (3.6), and $\boldsymbol{\eta}[i] = 0$ for $i \in$

$\mathbb{I}_{[1,\boldsymbol{m}]} \setminus \mathcal{A}^1(\boldsymbol{p})$ (from Lemma 3.3), we have:

$$\begin{bmatrix} \boldsymbol{Q}(\boldsymbol{p}) & \boldsymbol{A}(\boldsymbol{p})[\mathcal{A}'(\boldsymbol{p},\boldsymbol{q}),:]^\top \\ \boldsymbol{A}(\boldsymbol{p})[\mathcal{A}'(\boldsymbol{p},\boldsymbol{q}),:] & \end{bmatrix} \begin{bmatrix} \boldsymbol{\xi} \\ \boldsymbol{\eta}[\mathcal{A}'(\boldsymbol{p},\boldsymbol{q})] \end{bmatrix} = -\begin{bmatrix} \boldsymbol{S}(\boldsymbol{p}) \\ \boldsymbol{B}(\boldsymbol{p})[\mathcal{A}'(\boldsymbol{p},\boldsymbol{q}),:] \end{bmatrix} \boldsymbol{q} \tag{3.7}$$

$$\boldsymbol{\eta}[\mathbb{I}_{[1,\boldsymbol{m}]} \setminus \mathcal{A}'(\boldsymbol{p},\boldsymbol{q})] = \boldsymbol{0}.$$

The linear equation (3.7) provides a relationship between $[\boldsymbol{\xi}^\dagger(\boldsymbol{p},\boldsymbol{q});\boldsymbol{\eta}^\dagger(\boldsymbol{p},\boldsymbol{q})]$ and $\boldsymbol{q}$; however, it does not provide an explicit relationship because $\mathcal{A}'(\boldsymbol{p},\boldsymbol{q})$ depends on $\boldsymbol{q}$.

By rearranging $[\boldsymbol{\xi}^\dagger(\boldsymbol{p},\boldsymbol{q});\boldsymbol{\eta}^\dagger(\boldsymbol{p},\boldsymbol{q})]$ one can obtain $\boldsymbol{\zeta}^\dagger(\boldsymbol{p}) = \{[\xi_i^\dagger(\boldsymbol{p},\boldsymbol{q});\eta_i^\dagger(\boldsymbol{p},\boldsymbol{q})]\}_{i\in\mathcal{V}}$, where $\boldsymbol{\xi}^\dagger(\boldsymbol{p},\boldsymbol{q}) = \{\xi_i^\dagger(\boldsymbol{p},\boldsymbol{q})\}_{i\in\mathcal{V}}$, $\boldsymbol{\eta}^\dagger(\boldsymbol{p},\boldsymbol{q}) = \{\eta_i^\dagger(\boldsymbol{p},\boldsymbol{q})\}_{i\in\mathcal{V}}$. To perform such rearrangement, we consider a permutation $\phi : \mathbb{I}_{[1,\boldsymbol{n}]} \to \mathbb{I}_{[1,\boldsymbol{n}]}$ that achieves $\boldsymbol{z}[\phi(i)] = [\boldsymbol{\xi};\boldsymbol{\eta}][i]$. This permutation enables the following definition:

$$\mathcal{B}^0(\boldsymbol{p}) := \phi(\mathbb{I}_{[1,\boldsymbol{r}]} \cup (\mathcal{A}^0(\boldsymbol{p}) + \boldsymbol{r})), \quad \mathcal{B}^1(\boldsymbol{p}) := \phi(\mathbb{I}_{[1,\boldsymbol{r}]} \cup (\mathcal{A}^1(\boldsymbol{p}) + \boldsymbol{r})) \tag{3.8a}$$

$$\mathcal{B}'(\boldsymbol{p},\boldsymbol{q}) := \phi(\mathbb{I}_{[1,\boldsymbol{r}]} \cup (\mathcal{A}'(\boldsymbol{p},\boldsymbol{q}) + \boldsymbol{r})). \tag{3.8b}$$

Finally, $[\boldsymbol{\xi};\boldsymbol{\eta}]$ can be rearranged in such a way that the relationship $D_{\boldsymbol{q}}\boldsymbol{z}^\dagger(\boldsymbol{p}) = \boldsymbol{\zeta}^\dagger(\boldsymbol{p})$ (from Lemma 3.3) can be used; this yields:

$$D_{\boldsymbol{q}}\boldsymbol{z}^\dagger(\boldsymbol{p})[\mathcal{B}'(\boldsymbol{p},\boldsymbol{q})] = -\left(\boldsymbol{H}(\boldsymbol{p})[\mathcal{B}'(\boldsymbol{p},\boldsymbol{q}),\mathcal{B}'(\boldsymbol{p},\boldsymbol{q})]\right)^{-1} \boldsymbol{R}(\boldsymbol{p})[\mathcal{B}'(\boldsymbol{p},\boldsymbol{q}),:]\boldsymbol{q}, \tag{3.9a}$$

$$D_{\boldsymbol{q}}\boldsymbol{z}^\dagger(\boldsymbol{p})[\mathbb{I}_{[1,\boldsymbol{n}]} \setminus \mathcal{B}'(\boldsymbol{p},\boldsymbol{q})] = \boldsymbol{0}, \tag{3.9b}$$

where:

$$\boldsymbol{H}(\boldsymbol{p}) := \nabla_{\boldsymbol{zz}}\mathcal{L}(\boldsymbol{z}^\dagger(\boldsymbol{p});\boldsymbol{p}); \quad \boldsymbol{R}(\boldsymbol{p}) := \nabla_{\boldsymbol{zp}}\mathcal{L}(\boldsymbol{z}^\dagger(\boldsymbol{p});\boldsymbol{p}). \tag{3.10}$$

Here, the nonsingularity of $\boldsymbol{H}(\boldsymbol{p})[\mathcal{B}'(\boldsymbol{p},\boldsymbol{q}),\mathcal{B}'(\boldsymbol{p},\boldsymbol{q})]$ is guaranteed by the fact that $\boldsymbol{Q}(\boldsymbol{p})$ is positive definite on the null-space of $\boldsymbol{A}(\boldsymbol{p})[\mathcal{A}^0(\boldsymbol{p}),:]$. This follows from the satisfaction of LICQ and SSOSC (from Lemma 3.3) and [25, Lemma 16.1].

3.2. Nodal Sensitivity Result. We now observe that $\boldsymbol{H}(\boldsymbol{p})$ and $\boldsymbol{R}(\boldsymbol{p})$ are graph-induced matrices; these have bandwidth not greater than two, induced by $(\mathcal{G},\mathcal{I},\mathcal{I})$ and $(\mathcal{G},\mathcal{I},\mathcal{K})$, and where

$$I_i := \mathbb{I}_{\sum_{j\in\mathcal{V},j<i} n_j + [1,n_i]}, \quad K_i := \mathbb{I}_{\sum_{j\in\mathcal{V},j<i} l_j + [1,l_i]}, \quad i \in \mathcal{V}. \tag{3.11}$$

Note that $\mathcal{I} := \{I_i\}_{i\in\mathcal{V}}$ and $\mathcal{K} := \{K_i\}_{i\in\mathcal{V}}$ partition $\mathbb{I}_{[1,\boldsymbol{n}]}$ and $\mathbb{I}_{[1,\boldsymbol{l}]}$, respectively. We now observe that:

$$H_{ij}(\boldsymbol{p}) := \nabla^2_{z_i z_j}\mathcal{L}(\boldsymbol{z}^\dagger(\boldsymbol{p});\boldsymbol{p}) = (\boldsymbol{H}(\boldsymbol{p}))_{[i][j]} \tag{3.12a}$$

$$R_{ij}(\boldsymbol{p}) := \nabla^2_{z_i p_j}\mathcal{L}(\boldsymbol{z}^\dagger(\boldsymbol{p});\boldsymbol{p}) = (\boldsymbol{R}(\boldsymbol{p}))_{[i][j]}. \tag{3.12b}$$

From this we can see see that, if $d_{\mathcal{G}}(i,j) > 2$ holds, then $(\boldsymbol{H}(\boldsymbol{p}))_{[i][j]} = \boldsymbol{0}$ and $(\boldsymbol{R}(\boldsymbol{p}))_{[i][j]} = \boldsymbol{0}$ hold.

The submatrices of $\boldsymbol{H}(\boldsymbol{p})$ and $\boldsymbol{R}(\boldsymbol{p})$ are also graph-structured (induced by properly chosen index sets). In particular, $\boldsymbol{H}(\boldsymbol{p})[\mathcal{B},\mathcal{B}]$ and $\boldsymbol{R}[\mathcal{B},:]$ with $\mathcal{B} \subseteq \mathbb{I}_{[1,\boldsymbol{n}]}$ have bandwidth not greater than two induced by $(\mathcal{G},\mathcal{I}^{\mathcal{B}},\mathcal{I}^{\mathcal{B}})$ and $(\mathcal{G},\mathcal{I}^{\mathcal{B}},\mathcal{K})$, where $\mathcal{I}^{\mathcal{B}} := \{I_i \cap \mathcal{B}\}_{i\in\mathcal{V}}$.

We thus see that (3.7) involves the inverse of a graph-induced matrix; as such, Theorem 2.5 can be used for establishing the desired sensitivity bounds. By combining Theorem 2.5 and Lemma 3.3, one can establish the following result.

Lemma 3.4. *Suppose Assumptions 3.1, 3.2 hold, and suppose that, for given* $\boldsymbol{p} \in \mathbb{P}$ *(defined in Lemma 3.3) and* $\boldsymbol{q} := \{q_i\}_{i\in\mathcal{V}} \in \mathbb{R}^{\boldsymbol{l}}$*, we have* $\overline{\sigma}_{\boldsymbol{H}}(\boldsymbol{p},\boldsymbol{q})$*,* $\overline{\sigma}_{\boldsymbol{R}}(\boldsymbol{p},\boldsymbol{q})$*, and* $\underline{\sigma}_{\boldsymbol{H}}(\boldsymbol{p},\boldsymbol{q})$ *such that:*

$$\overline{\sigma}_{\boldsymbol{H}}(\boldsymbol{p},\boldsymbol{q}) \geq \overline{\sigma}(\boldsymbol{H}(\boldsymbol{p})[\mathcal{B}'(\boldsymbol{p},\boldsymbol{q}),\mathcal{B}'(\boldsymbol{p},\boldsymbol{q})]) \tag{3.13a}$$

$$\overline{\sigma}_{\boldsymbol{R}}(\boldsymbol{p},\boldsymbol{q}) \geq \overline{\sigma}(\boldsymbol{R}(\boldsymbol{p})[\mathcal{B}'(\boldsymbol{p},\boldsymbol{q}),:]) \tag{3.13b}$$

$$0 < \underline{\sigma}_{\boldsymbol{H}}(\boldsymbol{p},\boldsymbol{q}) \leq \underline{\sigma}(\boldsymbol{H}(\boldsymbol{p})[\mathcal{B}'(\boldsymbol{p},\boldsymbol{q}),\mathcal{B}'(\boldsymbol{p},\boldsymbol{q})]) \tag{3.13c}$$

hold; then the following holds for any $i \in \mathcal{V}$*:*

$$\|D_{\boldsymbol{q}} z_i^\dagger(\boldsymbol{p},\boldsymbol{q})\| \leq \sum_{j\in\mathcal{V}} \frac{\overline{\sigma}_{\boldsymbol{H}}(\boldsymbol{p},\boldsymbol{q})\overline{\sigma}_{\boldsymbol{R}}(\boldsymbol{p},\boldsymbol{q})}{\underline{\sigma}_{\boldsymbol{H}}(\boldsymbol{p},\boldsymbol{q})^2} \left(\frac{\overline{\sigma}_{\boldsymbol{H}}(\boldsymbol{p},\boldsymbol{q})^2 - \underline{\sigma}_{\boldsymbol{H}}(\boldsymbol{p},\boldsymbol{q})^2}{\overline{\sigma}_{\boldsymbol{H}}(\boldsymbol{p},\boldsymbol{q})^2 + \underline{\sigma}_{\boldsymbol{H}}(\boldsymbol{p},\boldsymbol{q})^2}\right)^{\left\lceil \frac{d_{\mathcal{G}}(i,j)}{4} - 1\right\rceil_+} \|q_j\|. \tag{3.14}$$

Proof. For simplicity, we denote $\mathcal{B}'(\boldsymbol{p},\boldsymbol{q})$ (defined in (3.8)) as $\mathcal{B}'$. From the fact that $\boldsymbol{H}(\boldsymbol{p})[\mathcal{B}',\mathcal{B}']$ is always nonsingular (as discussed after (3.10)) we have that $\underline{\sigma}_{\boldsymbol{H}}(\boldsymbol{p},\boldsymbol{q})$ satisfying (3.13c) always exists. By inspecting the block structure of (3.9) we can see that:

$$D_{\boldsymbol{q}} \boldsymbol{z}^\dagger(\boldsymbol{p})[I_i \cap \mathcal{B}'] = \sum_{j\in\mathcal{V}} -((\boldsymbol{H}(\boldsymbol{p})[\mathcal{B}',\mathcal{B}'])^{-1}\boldsymbol{R}(\boldsymbol{p})[\mathcal{B}',:])_{[i][j]} q_j, \tag{3.15}$$

where $\mathcal{I}^{\mathcal{B}'} := \{I_i \cap \mathcal{B}'\}_{i\in\mathcal{V}}$ and $\mathcal{K} := \{K_i\}_{i\in\mathcal{V}}$ (defined in (3.11)) are used for index sets. We have already established that $\boldsymbol{H}(\boldsymbol{p})[\mathcal{B}',\mathcal{B}']$ has bandwidth not greater than two, induced by $(\mathcal{G},\mathcal{I}^{\mathcal{B}'},\mathcal{I}^{\mathcal{B}'})$ and that $\boldsymbol{R}(\boldsymbol{p})[\mathcal{B}',:]$ has bandwidth not greater than two, induced by $(\mathcal{G},\mathcal{I}^{\mathcal{B}'},\mathcal{K})$. By applying Theorem 2.5, we obtain:

$$\begin{aligned}&\|((\boldsymbol{H}(\boldsymbol{p})[\mathcal{B}',\mathcal{B}'])^{-1}\boldsymbol{R}(\boldsymbol{p})[\mathcal{B}',:])_{[i][j]}\| \\ &\quad\leq \frac{\overline{\sigma}_{\boldsymbol{H}}(\boldsymbol{p},\boldsymbol{q})\overline{\sigma}_{\boldsymbol{R}}(\boldsymbol{p},\boldsymbol{q})}{\underline{\sigma}_{\boldsymbol{H}}(\boldsymbol{p},\boldsymbol{q})^2} \left(\frac{\overline{\sigma}_{\boldsymbol{H}}(\boldsymbol{p},\boldsymbol{q})^2 - \underline{\sigma}_{\boldsymbol{H}}(\boldsymbol{p},\boldsymbol{q})^2}{\overline{\sigma}_{\boldsymbol{H}}(\boldsymbol{p},\boldsymbol{q})^2 + \underline{\sigma}_{\boldsymbol{H}}(\boldsymbol{p},\boldsymbol{q})^2}\right)^{\left\lceil \frac{d_{\mathcal{G}}(i,j)}{4} - 1\right\rceil_+}.\end{aligned} \tag{3.16}$$

Now note that $\|D_{\boldsymbol{q}} z_i^\dagger(\boldsymbol{p})\| \leq \|D_{\boldsymbol{q}} \boldsymbol{z}^\dagger(\boldsymbol{p})[I_i \cap \mathcal{B}']\| + \|D_{\boldsymbol{q}} \boldsymbol{z}^\dagger(\boldsymbol{p})[I_i \setminus \mathcal{B}']\|$, and recall from (3.9b) that $D_{\boldsymbol{q}} \boldsymbol{z}^\dagger(\boldsymbol{p})[I_i \setminus \mathcal{B}'] = \boldsymbol{0}$. Hence, by applying (3.16) to (3.15), and applying triangle inequality, we obtain (3.14). ∎

Lemma 3.4 establishes that the dependence of $\|D_{\boldsymbol{q}} z_i^\dagger(\boldsymbol{p})\|$ on perturbation p_j decays with $d_{\mathcal{G}}(i,j)$. However, the right-hand side of (3.16) still depends on $\boldsymbol{p},\boldsymbol{q}$ and this complicates the use of Lemma 3.4 (needed to quantify sensitivity behavior). To express this result as in (1.2), wherein the expressions on the right-hand side are independent of $\boldsymbol{p},\boldsymbol{q}$, we exploit the continuity of singular values [16, Corollary 8.6.2]. This gives us the main result of the paper.

Theorem 3.5 (Exponential Decay of Sensitivity (EDS)). *Under Assumptions 3.1 and 3.2, and for given* $\epsilon > 0$, $\overline{\sigma}_{\boldsymbol{H}} \geq \overline{\overline{\sigma}}_{\boldsymbol{H}}(\boldsymbol{p}^\star)$*,* $\overline{\sigma}_{\boldsymbol{R}} \geq \overline{\overline{\sigma}}_{\boldsymbol{R}}(\boldsymbol{p}^\star)$*, and* $0 < \underline{\sigma}_{\boldsymbol{H}} \leq \underline{\underline{\sigma}}_{\boldsymbol{H}}(\boldsymbol{p}^\star)$*, where*

$$\overline{\overline{\sigma}}_{\boldsymbol{H}}(\boldsymbol{p}) := \max\{\overline{\sigma}(\boldsymbol{H}(\boldsymbol{p})[\mathcal{B},\mathcal{B}]) : \mathcal{B}^0(\boldsymbol{p}^\star) \subseteq \mathcal{B} \subseteq \mathcal{B}^1(\boldsymbol{p}^\star)\}$$
$$\overline{\overline{\sigma}}_{\boldsymbol{R}}(\boldsymbol{p}) := \max\{\overline{\sigma}(\boldsymbol{R}(\boldsymbol{p})[\mathcal{B},:]) : \mathcal{B}^0(\boldsymbol{p}^\star) \subseteq \mathcal{B} \subseteq \mathcal{B}^1(\boldsymbol{p}^\star)\}$$
$$\underline{\underline{\sigma}}_{\boldsymbol{H}}(\boldsymbol{p}) := \min\{\underline{\sigma}(\boldsymbol{H}(\boldsymbol{p})[\mathcal{B},\mathcal{B}]) : \mathcal{B}^0(\boldsymbol{p}^\star) \subseteq \mathcal{B} \subseteq \mathcal{B}^1(\boldsymbol{p}^\star)\},$$

there exists a neighborhood $\mathbb{P}_\epsilon$ *of* $\boldsymbol{p}^\star$ *such that the following holds for any* $\boldsymbol{p}, \boldsymbol{p}' \in \mathbb{P}_\epsilon$*:*

$$\|z_i^\dagger(\boldsymbol{p}) - z_i^\dagger(\boldsymbol{p}')\| \leq \sum_{j\in\mathcal{V}} \Upsilon \rho^{\left\lceil \frac{d_{\mathcal{G}}(i,j)}{4} - 1 \right\rceil_+} \|p_j - p_j'\|, \quad i \in \mathcal{V}, \tag{3.17}$$

with $\Upsilon := \dfrac{\overline{\sigma}_{\boldsymbol{H}}\overline{\sigma}_{\boldsymbol{R}}}{\underline{\sigma}_{\boldsymbol{H}}^2} + \epsilon$ *and* $\rho := \dfrac{\overline{\sigma}_{\boldsymbol{H}}^2 - \underline{\sigma}_{\boldsymbol{H}}^2}{\overline{\sigma}_{\boldsymbol{H}}^2 + \underline{\sigma}_{\boldsymbol{H}}^2} + \epsilon$.

Proof. From the continuity of $\boldsymbol{z}^\dagger(\cdot)$ in the neighborhood of $\boldsymbol{p}^\star$ and the continuity of $\boldsymbol{c}(\cdot,\cdot)$ in the neighborhood of $[\boldsymbol{z}^\star; \boldsymbol{p}^\star]$, there exists a neighborhood $\widetilde{\mathbb{P}} \subseteq \mathbb{P}$ of $\boldsymbol{p}^\star$ such that, for $\boldsymbol{p} \in \widetilde{\mathbb{P}}$ and $i \in \mathbb{I}_{[1,\boldsymbol{m}]}$,

$$\boldsymbol{c}(\boldsymbol{x}^\star)[i] > 0 \;\Rightarrow\; \boldsymbol{c}(\boldsymbol{x}^\dagger(\boldsymbol{p}))[i] > 0, \quad \boldsymbol{y}^\star[i] \neq 0 \;\Rightarrow\; \boldsymbol{y}^\dagger(\boldsymbol{p})[i] \neq 0.$$

These conditions and complementarity slackness imply that for $\boldsymbol{p} \in \widetilde{\mathbb{P}}$, we have $\mathcal{A}^0(\boldsymbol{p}^\star) \subseteq \mathcal{A}^0(\boldsymbol{p})$ and $\mathcal{A}^1(\boldsymbol{p}) \subseteq \mathcal{A}^1(\boldsymbol{p}^\star)$; that is, $\mathcal{B}^0(\boldsymbol{p}^\star) \subseteq \mathcal{B}^0(\boldsymbol{p})$ and $\mathcal{B}^1(\boldsymbol{p}) \subseteq \mathcal{B}(\boldsymbol{p}^\star)$. From this result and the fact that $\mathcal{B}^0(\boldsymbol{p}) \subseteq \mathcal{B}'(\boldsymbol{p},\boldsymbol{q}) \subseteq \mathcal{B}^1(\boldsymbol{p})$, we have that:

$$\overline{\sigma}(\boldsymbol{H}(\boldsymbol{p})[\mathcal{B}'(\boldsymbol{p},\boldsymbol{q}), \mathcal{B}'(\boldsymbol{p},\boldsymbol{q})]) \leq \overline{\overline{\sigma}}_{\boldsymbol{H}}(\boldsymbol{p}) \tag{3.18a}$$

$$\overline{\sigma}(\boldsymbol{R}(\boldsymbol{p})[\mathcal{B}'(\boldsymbol{p},\boldsymbol{q}), :]) \leq \overline{\overline{\sigma}}_{\boldsymbol{R}}(\boldsymbol{p}) \tag{3.18b}$$

$$\underline{\sigma}(\boldsymbol{H}(\boldsymbol{p})[\mathcal{B}'(\boldsymbol{p},\boldsymbol{q}), \mathcal{B}'(\boldsymbol{p},\boldsymbol{q})]) \geq \underline{\underline{\sigma}}_{\boldsymbol{H}}(\boldsymbol{p}). \tag{3.18c}$$

By the twice-continuous differentiability of $\mathcal{L}(\cdot,\cdot)$ and the continuity of $\boldsymbol{z}^\dagger(\cdot)$, we have that $\boldsymbol{H}(\cdot)$ is continuous. The same holds true for its submatrices: $\boldsymbol{H}(\cdot)[\mathcal{B},\mathcal{B}]$ with $\mathcal{B}^0(\boldsymbol{p}^\star) \subseteq \mathcal{B} \subseteq \mathcal{B}^1(\boldsymbol{p}^\star)$. From the continuity of singular values with respect to its entries [16, Corollary 8.6.2], we have that $\overline{\sigma}(\boldsymbol{H}(\cdot)[\mathcal{B},\mathcal{B}])$ and $\underline{\sigma}(\boldsymbol{H}(\cdot)[\mathcal{B},\mathcal{B}])$ are continuous for any $\mathcal{B}^0(\boldsymbol{p}^\star) \subseteq \mathcal{B} \subseteq \mathcal{B}^1(\boldsymbol{p}^\star)$; accordingly, since a maximum and a minimum of a fixed and finite number of continuous functions is continuous, we have that $\overline{\overline{\sigma}}_{\boldsymbol{H}}(\boldsymbol{p})$, $\overline{\overline{\sigma}}_{\boldsymbol{R}}(\boldsymbol{p})$, $\underline{\underline{\sigma}}_{\boldsymbol{H}}(\boldsymbol{p})$ are continuous with respect to $\boldsymbol{p}$ in $\widetilde{\mathbb{P}}$. Thus, there exists a convex neighborhood $\mathbb{P}_\epsilon \subseteq \widetilde{\mathbb{P}}$ of $\boldsymbol{p}^\star$ wherein the following are satisfied:

$$\frac{\overline{\overline{\sigma}}_{\boldsymbol{H}}(\boldsymbol{p})\overline{\overline{\sigma}}_{\boldsymbol{R}}(\boldsymbol{p})}{\underline{\underline{\sigma}}_{\boldsymbol{H}}(\boldsymbol{p})^2} \leq \frac{\overline{\overline{\sigma}}_{\boldsymbol{H}}(\boldsymbol{p}^\star)\overline{\overline{\sigma}}_{\boldsymbol{R}}(\boldsymbol{p}^\star)}{\underline{\underline{\sigma}}_{\boldsymbol{H}}(\boldsymbol{p}^\star)^2} + \epsilon \tag{3.19a}$$

$$\frac{\overline{\overline{\sigma}}_{\boldsymbol{H}}(\boldsymbol{p})^2 - \underline{\underline{\sigma}}_{\boldsymbol{H}}(\boldsymbol{p})^2}{\overline{\overline{\sigma}}_{\boldsymbol{H}}(\boldsymbol{p})^2 + \underline{\underline{\sigma}}_{\boldsymbol{H}}(\boldsymbol{p})^2} \leq \frac{\overline{\overline{\sigma}}_{\boldsymbol{H}}(\boldsymbol{p}^\star)^2 - \underline{\underline{\sigma}}_{\boldsymbol{H}}(\boldsymbol{p}^\star)^2}{\overline{\overline{\sigma}}_{\boldsymbol{H}}(\boldsymbol{p}^\star)^2 + \underline{\underline{\sigma}}_{\boldsymbol{H}}(\boldsymbol{p}^\star)^2} + \epsilon. \tag{3.19b}$$

Here, note that $\underline{\underline{\sigma}}_{\boldsymbol{H}}(\boldsymbol{p}^\star) > 0$ because $\underline{\sigma}(\boldsymbol{H}[\mathcal{B},\mathcal{B}]) > 0$ holds for any $\mathcal{B}^0(\boldsymbol{p}^\star) \subseteq \mathcal{B} \subseteq \mathcal{B}^1(\boldsymbol{p}^\star)$ (as discussed in the proof of Lemma 3.4). By applying (3.18) and (3.19) to Lemma 3.4, we obtain:

$$\begin{aligned}
\|D_{\boldsymbol{q}} z_i^\dagger(\boldsymbol{p})\| &\leq \sum_{j\in\mathcal{V}} \left(\frac{\overline{\overline{\sigma}}_{\boldsymbol{H}}(\boldsymbol{p}^\star)\overline{\overline{\sigma}}_{\boldsymbol{R}}(\boldsymbol{p}^\star)}{\underline{\underline{\sigma}}_{\boldsymbol{H}}(\boldsymbol{p}^\star)^2} + \epsilon\right)\left(\frac{\overline{\overline{\sigma}}_{\boldsymbol{H}}(\boldsymbol{p}^\star)^2 - \underline{\underline{\sigma}}_{\boldsymbol{H}}(\boldsymbol{p}^\star)^2}{\overline{\overline{\sigma}}_{\boldsymbol{H}}(\boldsymbol{p}^\star)^2 + \underline{\underline{\sigma}}_{\boldsymbol{H}}(\boldsymbol{p}^\star)^2} + \epsilon\right)^{\left\lceil \frac{d_{\mathcal{G}}(i,j)}{4} - 1 \right\rceil_+} \|q_j\| \\
&\leq \sum_{j\in\mathcal{V}} \left(\frac{\overline{\sigma}_{\boldsymbol{H}}\overline{\sigma}_{\boldsymbol{R}}}{\underline{\sigma}_{\boldsymbol{H}}^2} + \epsilon\right)\left(\frac{\overline{\sigma}_{\boldsymbol{H}}^2 - \underline{\sigma}_{\boldsymbol{H}}^2}{\overline{\sigma}_{\boldsymbol{H}}^2 + \underline{\sigma}_{\boldsymbol{H}}^2} + \epsilon\right)^{\left\lceil \frac{d_{\mathcal{G}}(i,j)}{4} - 1 \right\rceil_+} \|q_j\| \\
&\leq \sum_{j\in\mathcal{V}} \Upsilon \rho^{\left\lceil \frac{d_{\mathcal{G}}(i,j)}{4} - 1 \right\rceil_+} \|q_j\|, \qquad (3.20)
\end{aligned}$$

for any $\boldsymbol{p} \in \mathbb{P}_\epsilon$ and $\boldsymbol{q} \in \mathbb{R}^{\boldsymbol{l}}$. Finally, since we have chosen $\mathbb{P}_\epsilon$ to be convex, for any $\boldsymbol{p}, \boldsymbol{p}' \in \mathbb{P}_\epsilon$, the line segment between $\boldsymbol{p}, \boldsymbol{p}'$ is within $\mathbb{P}_\epsilon$. Thus, we have:

$$
\begin{aligned}
\|z_i^\dagger(\boldsymbol{p}) - z_i^\dagger(\boldsymbol{p}')\| &\leq \|\int_0^1 D_{\boldsymbol{p}'-\boldsymbol{p}} z_i^\dagger((1-t)\boldsymbol{p} + t\boldsymbol{p}')dt\| \\
&\leq \int_0^1 \|D_{\boldsymbol{p}'-\boldsymbol{p}} z_i^\dagger((1-t)\boldsymbol{p} + t\boldsymbol{p}')\|dt \leq \sum_{j \in \mathcal{V}} \Upsilon \rho^{\left\lceil \frac{d_{\mathcal{G}}(i,j)}{4} - 1 \right\rceil_+} \|p_j - p_j'\|,
\end{aligned}
$$

where the first inequality is from Newton-Leibniz, the second inequality follows from triangle inequality for integrals, and the last inequality follows from (3.20). ∎

Theorem 3.5 establishes the sensitivity bounds $\{C_{ij} = \Upsilon \rho^{\lceil d_{\mathcal{G}}(i,j)/4 - 1 \rceil_+}\}_{i,j \in \mathcal{V}}$. One can observe that $\Upsilon > 0$ and $\rho \in (0,1)$ hold; consequently, we have that the upper bound of the nodal sensitivity decays exponentially as $d_{\mathcal{G}}(i,j)$ increases. We can also see that (Υ, ρ) depend on the singular values of the submatrices of $\boldsymbol{H}(\boldsymbol{p}^\star)$, $\boldsymbol{R}(\boldsymbol{p}^\star)$, which are submatrices of the full Hessian matrix $\nabla^2 \mathcal{L}(\boldsymbol{z}^\star; \boldsymbol{p}^\star)$. Therefore, the singular values play important roles in sensitivity behavior.

Remark 3.6. One can establish EDS for a more general version of NLP (1.1), in which coupling is allowed within the expanded neighborhood $N_{\mathcal{G}}^B[i] := \{j \in \mathcal{V} : d_{\mathcal{G}}(i,j) \leq B\}$ with $B > 1$. Such an NLP arises when algebraic coupling between nodes extends beyond immediate neighbors. In such a case, the matrices $\boldsymbol{H}(\boldsymbol{p})$ and $\boldsymbol{R}(\boldsymbol{p})$ (and their submatrices) have bandwidths not greater than $2B$. For this more general setting, the corresponding results for Theorem 3.5 can be established; in particular, if the rest of the assumptions in Theorem 3.5 remain the same, the following holds:

$$
\|z_i^\dagger(\boldsymbol{p}) - z_i^\dagger(\boldsymbol{p}')\| \leq \sum_{j \in \mathcal{V}} \Upsilon \rho^{\left\lceil \frac{d_{\mathcal{G}}(i,j)}{4B} - 1 \right\rceil_+} \|p_j - p_j'\|. \tag{3.21}
$$

We can observe that the exponential decay rate increases (the decay becomes slower) as the constant B increases. This implies that we require a small coupling radius B in order to have fast decay of sensitivity (which makes intuitive sense).

4. Uniform Sensitivity Bounds. An interesting class of graph-structured NLPs is that in which the underlying graph is a subgraph of an infinite-dimensional graph. Examples include time-dependent problems (in which we might want to extend the horizon) and discretized PDE optimization (in which we might want to expand the domain). To analyze this setting, we consider a family of problems $\{P_{(k)}(\cdot)\}_{k \in K}$ with a potentially infinite problem index set K. The associated quantities are introduced accordingly; $\{\mathcal{G}_{(k)} = (\mathcal{V}_{(k)}, \mathcal{E}_{(k)})\}_{k \in K}$, $\{\boldsymbol{f}_{(k)}(\cdot)\}_{k \in K}$, $\{\boldsymbol{c}_{(k)}(\cdot)\}_{k \in K}$. Also, a set of data $\{\boldsymbol{p}_{(k)}^\star\}_{k \in K}$ and the associated base solutions $\{\boldsymbol{z}_{(k)}^\star\}_{k \in K}$ are considered. The submatrices $\{\boldsymbol{H}_{(k)}(\boldsymbol{p}^\star)\}_{k \in K}$, $\{\boldsymbol{R}_{(k)}(\boldsymbol{p}^\star)\}_{k \in K}$ of the full Hessian matrix can be defined as in (3.10) for each k.

This section aims to establish sufficient conditions for:

$$
\sup_{k \in K} \overline{\overline{\sigma}}_{\boldsymbol{H},(k)}(\boldsymbol{p}_{(k)}^\star) < +\infty; \quad \sup_{k \in K} \overline{\overline{\sigma}}_{\boldsymbol{R},(k)}(\boldsymbol{p}_{(k)}^\star) < +\infty; \quad \inf_{k \in K} \underline{\underline{\sigma}}_{\boldsymbol{H},(k)}(\boldsymbol{p}_{(k)}^\star) > 0, \tag{4.1}
$$

where $\overline{\overline{\sigma}}_{\boldsymbol{H},(k)}(\boldsymbol{p}_{(k)}^\star)$, $\overline{\overline{\sigma}}_{\boldsymbol{R},(k)}(\boldsymbol{p}_{(k)}^\star)$, and $\underline{\underline{\sigma}}_{\boldsymbol{H},(k)}(\boldsymbol{p}_{(k)}^\star)$ are defined in Theorem 3.5, but the problem index k is added. One can observe that, if (4.1) is violated, $\Upsilon_{(k)}$ may become indefinitely large and $\rho_{(k)}$ may approach one (thus making the bounds derived in

Theorem 3.5 not particularly useful). Hence, ensuring (4.1) is crucial for guaranteeing a moderately bounded sensitivity magnitude $\Upsilon_{(k)}$ and a fast sensitivity decay rate $\rho_{(k)}$.

We call (4.1) *uniform* boundedness conditions; furthermore, we call a quantity to be *uniform* in k if the quantity is independent of the index k. Note that (4.1) holds trivially if K is finite and Assumptions 3.1 and 3.2 hold for each $k \in K$. However, even if K is finite, it is necessary for Theorem 3.5 to be practically useful that $\inf_{k\in K} \underline{\underline{\sigma}}_{\boldsymbol{H},(k)}(\boldsymbol{p}^\star_{(k)})$ is sufficiently bounded away from zero and that $\sup_{k\in K} \overline{\overline{\sigma}}_{\boldsymbol{H},(k)}(\boldsymbol{p}^\star_{(k)})$ and $\sup_{k\in K} \overline{\overline{\sigma}}_{\boldsymbol{R},(k)}(\boldsymbol{p}^\star_{(k)})$ are bounded above by a moderately large number. As such, the results in this section provide useful information even if K is finite (and even if K is a singleton). Hereafter, we will consistently use k to denote the problem index and drop the notation for dependency on $\boldsymbol{p}$, since it is fixed to $\boldsymbol{p}^\star$ for the rest of the discussion in this section (e.g., $\boldsymbol{H}_{(k)} \leftarrow \boldsymbol{H}_{(k)}(\boldsymbol{p}^\star)$).

4.1. Sufficient Conditions for Uniform Boundedness. We now state assumptions that enable uniform boundedness (4.1). These assumptions provide basic uniform parameters from which we can establish explicit bounds for the quantities in (4.1).

ASSUMPTION 4.1 (Uniformly Bounded Degree of Graphs). *There exists a uniform upper bound $D \in \mathbb{I}_{>0}$ (uniform in k) of the degrees of nodes in $\mathcal{G}_{(k)}$. That is, $|N_{\mathcal{G}_{(k)}}[i]| \leq D$ for any $i \in \mathcal{V}_{(k)}$ and $k \in K$.*

ASSUMPTION 4.2 (Uniformly Bounded Second Derivatives). *There exists $L \geq 0$ (uniform in k) such that $\|H_{(k),ij}\|, \|R_{(k),ij}\| \leq L$ for any $i, j \in \mathcal{V}_{(k)}$, $k \in K$, where $H_{(k),ij}$ and $R_{(k),ij}$ are defined in (3.12).*

ASSUMPTION 4.3 (Uniform SSOSC). *There exists $\gamma > 0$ (uniform in k) such that $\mathrm{ReH}(\boldsymbol{Q}_{(k)}, \boldsymbol{A}^0_{(k)}) \succeq \gamma \boldsymbol{I}$ for any $k \in K$, where $\boldsymbol{A}^0_{(k)} := \boldsymbol{A}_{(k)}[\mathcal{A}^0_{(k)}, :]$, and $\mathcal{A}^0_{(k)}$, $\boldsymbol{Q}_{(k)}$, and $\boldsymbol{A}_{(k)}$ are defined in (3.4).*

ASSUMPTION 4.4 (Uniform LICQ). *There exists $\beta > 0$ (uniform in k) such that $\boldsymbol{A}^1_{(k)}(\boldsymbol{A}^1_{(k)})^\top \succeq \beta \boldsymbol{I}$ for $k \in K$, where $\boldsymbol{A}^1_{(k)} := \boldsymbol{A}_{(k)}[\mathcal{A}^1_{(k)}, :]$, and $\mathcal{A}^1_{(k)}$ and $\boldsymbol{A}_{(k)}$ are defined in (3.4).*

Assumptions 4.3, 4.4 are extensions of Assumption 3.2; in particular, the assumptions are strengthened by introducing additional uniform paramters, $\gamma, \beta > 0$. With these parameters, we can establish the uniform bounds in (4.1). In the next theorem we establish upper bounds for $\overline{\overline{\sigma}}_{\boldsymbol{H},(k)}$ and $\overline{\overline{\sigma}}_{\boldsymbol{R},(k)}$.

LEMMA 4.5. *Under Assumptions 4.1, 4.2, we have that $\overline{\overline{\sigma}}_{\boldsymbol{H},(k)}, \overline{\overline{\sigma}}_{\boldsymbol{R},(k)} \leq D^2 L$ for $k \in K$.*

In order to prove this lemma, we first need to establish a general inequality for matrix norms. The following lemma is a generalization of inequality $\|M\| \leq (\|M\|_1 \|M\|_\infty)^{1/2}$ [16, Corollary 2.3.2].

LEMMA 4.6. *Consider $M \in \mathbb{R}^{m\times n}$ with index set families $\mathcal{I} := \{I_i\}_{i\in\mathcal{V}}$ and $\mathcal{J} = \{J_i\}_{i\in\mathcal{V}}$ that partition $\mathbb{I}_{[1,m]}$ and $\mathbb{I}_{[1,n]}$, respectively. The following holds:*

$$\overline{\sigma}(M) \leq \left(\left(\max_{i\in\mathcal{V}} \sum_{j\in\mathcal{V}} \|M_{[i][j]}\|\right)\left(\max_{j\in\mathcal{V}} \sum_{i\in\mathcal{V}} \|M_{[i][j]}\|\right)\right)^{1/2}, \tag{4.2}$$

where $M_{[i][j]} = M[I_i][J_j]$ for any $i, j \in \mathcal{V}$.

Proof. The inequality holds trivially if $M = \mathbf{0}$; we thus assume $M \neq \mathbf{0}$. Consider the left singular vector $v \in \mathbb{R}^n$ of M with singular value $\overline{\sigma}(M)$. We have that

$\overline{\sigma}(M)^2 v = MM^\top v$. We let $u = M^\top v$, which yields $\overline{\sigma}(M)^2 v = Mu$; accordingly,

$$
\begin{aligned}
\overline{\sigma}(M)^2 \sum_{i\in\mathcal{V}} \|v_{[i]}\| &= \sum_{i\in\mathcal{V}} \| \sum_{j\in\mathcal{V}} M_{[i][j]} u_{[j]} \| \\
&\leq \sum_{j\in\mathcal{V}} (\sum_{i\in\mathcal{V}} \|M_{[i][j]}\|) \|u_{[j]}\| \leq \Big(\max_{j\in\mathcal{V}} \sum_{i\in\mathcal{V}} \|M_{[i][j]}\| \Big) \Big(\sum_{j\in\mathcal{V}} \|u_{[j]}\| \Big), \qquad (4.3)
\end{aligned}
$$

where the first inequality is obtained by applying the triangle inequality and the submultiplicativity of the matrix norm, and by switching the order of summation; the second inequality is obtained from $\sum_{i\in\mathcal{V}} \|M_{[i][j]}\| \leq \max_{j\in\mathcal{V}} \sum_{i\in\mathcal{V}} \|M_{[i][j]}\|$. Using the same logic, we obtain: $\sum_{j\in\mathcal{V}} \|u_{[j]}\| \leq (\max_{i\in\mathcal{V}} \sum_j \|M_{[i][j]}\|)(\sum_{i\in\mathcal{V}} \|v_{[i]}\|)$. From these results, (4.3), and the fact that $v \neq \boldsymbol{0}$ (by $M \neq \boldsymbol{0}$), we obtain (4.2). ∎

Proof of Lemma 4.5. Within this proof, we omit the subscript (k) for conciseness. Since $\mathcal{B}^1 \subseteq \mathbb{I}_{[1,\boldsymbol{n}]}$, $\overline{\sigma}(\boldsymbol{H}) \geq \overline{\overline{\sigma}}_{\boldsymbol{H}}$ and $\overline{\sigma}(\boldsymbol{R}) \geq \overline{\overline{\sigma}}_{\boldsymbol{R}}$. Thus, it suffices to show that $\overline{\sigma}(\boldsymbol{H}), \overline{\sigma}(\boldsymbol{R}) \leq D^2 L$. As observed in Section 3.2, $\boldsymbol{H}$ and $\boldsymbol{R}$ have bandwidth not greater than two since H_{ij} and R_{ij} equal zero if $d_{\mathcal{G}}(i,j) > 2$. Hence, the number of nonzero blocks on one-block rows or on one-block columns of $\boldsymbol{H}$ and $\boldsymbol{R}$ is at most D^2, since $|N^2_{\mathcal{G}}[i]| \leq D^2$ for any $i \in \mathcal{V}$ (i.e., for any node, there exist at most D^2 nodes within distance two). As such, we have:

$$
\max_{i\in\mathcal{V}} \sum_{j\in\mathcal{V}} \|H_{ij}\| \vee \max_{j\in\mathcal{V}} \sum_{i\in\mathcal{V}} \|H_{ij}\| \leq D^2 L; \quad \max_{i\in\mathcal{V}} \sum_{j\in\mathcal{V}} \|R_{ij}\| \vee \max_{j\in\mathcal{V}} \sum_{i\in\mathcal{V}} \|R_{ij}\| \leq D^2 L.
$$

By Lemma 4.6, $\overline{\sigma}(\boldsymbol{H}) \leq D^2 L$; and $\overline{\sigma}(\boldsymbol{R}) \leq D^2 L$. ∎

From Lemma 4.5, we can see that two upper bounds in (4.1) hold under Assumption 4.1, 4.2. The following lemma establishes a lower bound for $\underline{\underline{\sigma}}_{\boldsymbol{H},(k)}$.

Lemma 4.7. *Under Assumptions 4.1, 4.2, 4.3, and 4.4 we have that:*

$$
\underline{\underline{\sigma}}_{\boldsymbol{H},(k)} \geq \left(\frac{2}{\gamma} + \frac{8\alpha D^4 L^2}{\gamma^3 \beta} + \frac{4 D^2 L}{\gamma^2 \beta} \right)^{-1} \left(1 + \alpha D^2 L\right)^{-1}, \qquad (4.4)
$$

for any $k \in K$ and where $\alpha := (2D^4 L^2/\gamma + \gamma + D^2 L)/\beta$.

We first prove that $\boldsymbol{Q}_{(k)} + \alpha (\boldsymbol{A}^0_{(k)})^\top \boldsymbol{A}^0_{(k)}$ is positive definite (recall that SSOSC does not necessarily guarantee positive definiteness of $\boldsymbol{Q}_{(k)}$).

Lemma 4.8. $\boldsymbol{Q}_{(k)} + \alpha (\boldsymbol{A}^0_{(k)})^\top \boldsymbol{A}^0_{(k)} \succeq (\gamma/2) \boldsymbol{I}$.

Proof. Within this proof, we omit the subscript (k) for conciseness. From Lemma 4.5, $\|\boldsymbol{H}\| \leq D^2 L$; this implies that its submatrices $\boldsymbol{Q}, \boldsymbol{A}^0$ satisfy $\overline{\sigma}(\boldsymbol{Q}), \overline{\sigma}(\boldsymbol{A}^0) \leq D^2 L$. The smallest eigenvalue of $\boldsymbol{Q} + \alpha (\boldsymbol{A}^0)^\top \boldsymbol{A}^0$ is obtained from:

$$
\begin{aligned}
&\min_w \; w^\top (\boldsymbol{Q} + \alpha (\boldsymbol{A}^0)^\top \boldsymbol{A}^0) w && (4.5a) \\
&\text{s.t. } \|w\| = 1. && (4.5b)
\end{aligned}
$$

Any $w \in \mathbb{R}^{\boldsymbol{r}}$ can be expressed as $w = Z w_Z + Y w_Y$, where the columns of Z form an orthonormal basis for the null space of $\boldsymbol{A}^0$ and the columns of Y form an orthonormal basis for the row space of $\boldsymbol{A}^0$. We have that

$$
1 = \|w_Z\|^2 + \|w_Y\|^2; \quad \|Z w_Z\| = \|w_Z\|; \quad \|Y w_Y\| = \|w_Y\|, \qquad (4.6)
$$

which follows from (4.5b) and orthogonality of Z and Y. The objective (4.5a) satisfies:

(4.7)
$$\begin{aligned}
&w^\top(\boldsymbol{Q}+\alpha(\boldsymbol{A}^0)^\top\boldsymbol{A}^0)w\\
&= w_Z^\top Z^\top \boldsymbol{Q} Z w_Z + 2w_Y^\top Y^\top \boldsymbol{Q} Z w_Z + w_Y^\top Y^\top \boldsymbol{Q} Y w_Y + \alpha w_Y^\top Y^\top (\boldsymbol{A}^0)^\top \boldsymbol{A}^0 Y w_Y\\
&\geq \gamma\|w_Z\|^2 - 2\|\boldsymbol{Q}\|\|Zw_Z\|\|Yw_Y\| - \|\boldsymbol{Q}\|\|Yw_Y\|^2 + \alpha\underline{\lambda}(Y^\top(\boldsymbol{A}^0)^\top\boldsymbol{A}^0 Y)\|w_Y\|^2\\
&\geq \gamma(1-\|w_Y\|^2) - 2\overline{\sigma}(\boldsymbol{Q})\|w_Z\|\|w_Y\| - \overline{\sigma}(\boldsymbol{Q})\|w_Y\|^2 + \alpha\underline{\lambda}(\boldsymbol{A}^0 Y Y^\top(\boldsymbol{A}^0)^\top)\|w_Y\|^2\\
&\geq \gamma(1-\|w_Y\|^2) - 2D^2L\|w_Y\| - D^2L\|w_Y\|^2 + \alpha\underline{\lambda}(\boldsymbol{A}^0(\boldsymbol{A}^0)^\top)\|w_Y\|^2\\
&\geq \gamma - 2D^2L\|w_Y\| + (\alpha\beta-\gamma-D^2L)\|w_Y\|^2,
\end{aligned}$$

where $\underline{\lambda}(\cdot)$ denotes the smallest eigenvalue of the symmetric matrix argument. The equality follows from $\boldsymbol{A}^0 Z = \boldsymbol{0}$; the first inequality follows from (i) Assumption 4.3, (ii) submultiplicativity of matrix norms, and (iii) the fact that $w^\top M w \geq \underline{\lambda}(M)\|w\|^2$ for positive definite M; the second inequality comes from (i) Equation (4.6), (ii) the fact that the induced 2-norm is equal to the largest singular value, and (iii) the equality $\underline{\lambda}(MM^\top) = \underline{\lambda}(M^\top M)$ for square M; the third inequality follows from (i) Lemma 4.5 and (ii) $\boldsymbol{A}^0 Y Y^\top = \boldsymbol{A}^0$ since Y is an orthogonal matrix whose columns span the row space of $\boldsymbol{A}^0$; and the last inequality follows from Assumption 4.4.

Since $\mathcal{V}$ is nonempty, we have that $D > 0$; furthermore, we have that $L \neq 0$ from SSOSC and LICQ and thus $D^2L \neq 0$ holds. This implies that $\alpha\beta - \gamma - D^2L = 2D^4L^2/\gamma > 0$. Accordingly, the quadratic expression on the right-hand side of the last inequality of (4.7) is lower-bounded by:

$$w^\top(\boldsymbol{Q}+\alpha(\boldsymbol{A}^0)^\top\boldsymbol{A}^0)w \geq \gamma - \frac{D^4L^2}{\alpha\beta-\gamma-D^2L} = \frac{\gamma}{2}. \qquad \square$$

Proof of Lemma 4.7. Within this proof, we omit the subscript (k) for conciseness. It suffices to show that $\underline{\sigma}(\boldsymbol{H}[\mathcal{B},\mathcal{B}])$ for any $\mathcal{B}^0 \subseteq \mathcal{B} \subseteq \mathcal{B}^1$ is lower bounded by the right-hand-side of (4.4). Furthermore, we know that $\boldsymbol{H}[\mathcal{B},\mathcal{B}]$ is a permutation of:

$$(4.8) \qquad \begin{bmatrix} \boldsymbol{Q} & (\boldsymbol{A}')^\top \\ \boldsymbol{A}', & \end{bmatrix}$$

where $\boldsymbol{A}' := \boldsymbol{A}[\mathcal{A},:]$, and $\mathcal{A} := (\phi^{-1}(\mathcal{B}) \setminus \mathbb{I}_{[1,\boldsymbol{r}]}) - \boldsymbol{r}$; here, $\phi : \mathbb{I}_{[1,\boldsymbol{n}]} \to \mathbb{I}_{[1,\boldsymbol{n}]}$ is a permutation that achieves $\boldsymbol{z}[\phi(i)] = [\boldsymbol{\xi};\boldsymbol{\eta}][i]$. It thus suffices to show that the lowest singular value of the matrix in (4.8) with $\mathcal{A}^0 \subseteq \mathcal{A} \subseteq \mathcal{A}^1$ is lower bounded by the right-hand side of (4.4).

We now make the following observation:

$$(4.9) \qquad (\boldsymbol{A}')^\top\boldsymbol{A}' \succeq (\boldsymbol{A}^0)^\top\boldsymbol{A}^0; \quad \underline{\lambda}(\boldsymbol{A}'(\boldsymbol{A}')^\top) \geq \underline{\lambda}(\boldsymbol{A}^1(\boldsymbol{A}^1)^\top);$$

here, the first inequality results from $(\boldsymbol{A}')^\top\boldsymbol{A}' - (\boldsymbol{A}^0)^\top\boldsymbol{A}^0 = \boldsymbol{A}[\mathcal{A}\setminus\mathcal{A}^0,:]^\top\boldsymbol{A}[\mathcal{A}\setminus\mathcal{A}^0,:] \succeq \boldsymbol{0}$. To establish the second inequality, we consider a unit vector $w \in \mathbb{R}^{\boldsymbol{m}}$ such that $w[\mathcal{A}]$ is the eigenvector of $\boldsymbol{A}'(\boldsymbol{A}')^\top$ associated with the smallest eigenvalue and $w[\mathbb{I}_{[1,\boldsymbol{m}]}\setminus\mathcal{A}] = \boldsymbol{0}$. We can see that:

$$(4.10) \qquad \underline{\lambda}(\boldsymbol{A}^1(\boldsymbol{A}^1)^\top) \leq w[\mathcal{A}^1]^\top\boldsymbol{A}^1(\boldsymbol{A}^1)^\top w[\mathcal{A}^1] = \underline{\lambda}(\boldsymbol{A}'(\boldsymbol{A}')^\top);$$

here, the first inequality follows from the fact that $\lambda(\boldsymbol{A}^1(\boldsymbol{A}^1)^\top)$ is the smallest eigenvalue, and the equality follows from the fact that $w[\mathcal{A}^1\setminus\mathcal{A}] = \boldsymbol{0}$. This establishes the second inequality in (4.9).

We now study the inverse of the matrix in (4.8); note that $\underline{\sigma}(\boldsymbol{H}) = \|\boldsymbol{H}^{-1}\|$ and

$$\begin{bmatrix} \boldsymbol{Q} & (\boldsymbol{A}')^\top \\ \boldsymbol{A}' & \end{bmatrix}^{-1} = \begin{bmatrix} \boldsymbol{Q} + \alpha(\boldsymbol{A}')^\top \boldsymbol{A}' & (\boldsymbol{A}')^\top \\ \boldsymbol{A}' & \end{bmatrix}^{-1} \begin{bmatrix} \boldsymbol{I} & \alpha(\boldsymbol{A}')^\top \\ & \boldsymbol{I} \end{bmatrix} \tag{4.11a}$$

$$= \begin{bmatrix} T + \alpha T(\boldsymbol{A}')^\top S\boldsymbol{A}'T & T(\boldsymbol{A}')^\top S \\ S\boldsymbol{A}'T & \end{bmatrix} \begin{bmatrix} \boldsymbol{I} & \alpha(\boldsymbol{A}')^\top \\ & \boldsymbol{I} \end{bmatrix}, \tag{4.11b}$$

where $T := (\boldsymbol{Q}+\alpha(\boldsymbol{A}')^\top \boldsymbol{A}')^{-1}$, and $S := (\boldsymbol{A}'(\boldsymbol{Q}+\alpha(\boldsymbol{A}')^\top \boldsymbol{A}')(\boldsymbol{A}')^\top)^{-1}$; here, the first equality can be easily verified; and the second equality follows from [3, Proposition 2.8.7]. Now observe that:

$$\underline{\lambda}(\boldsymbol{Q} + \alpha(\boldsymbol{A}')^\top \boldsymbol{A}') \geq \underline{\lambda}(\boldsymbol{Q} + \alpha(\boldsymbol{A}^0)^\top \boldsymbol{A}^0) \geq \gamma/2; \tag{4.12}$$

here, the first inequality follows from (4.9) and the second inequality follows from Lemma 4.8. Furthermore,

$$\underline{\lambda}(\boldsymbol{A}'(\boldsymbol{Q} + \alpha(\boldsymbol{A}')^\top \boldsymbol{A}')(\boldsymbol{A}')^\top) \geq \underline{\lambda}(\boldsymbol{Q} + \alpha(\boldsymbol{A}')^\top \boldsymbol{A}')\underline{\lambda}(\boldsymbol{A}'(\boldsymbol{A}')^\top) \geq \gamma\beta/2;$$

here, the first inequality follows from

$$\begin{aligned} \min_{\|w\|\leq 1} w^\top(\boldsymbol{A}'(\boldsymbol{Q} + \alpha(\boldsymbol{A}')^\top \boldsymbol{A}')(\boldsymbol{A}')^\top)w &\geq \min_{\|w\|\leq 1} \underline{\lambda}(\boldsymbol{Q} + \alpha(\boldsymbol{A}')^\top \boldsymbol{A}')\|(\boldsymbol{A}')^\top w\|^2 \\ &\geq \underline{\lambda}(\boldsymbol{Q} + \alpha(\boldsymbol{A}')^\top \boldsymbol{A}')\underline{\lambda}(\boldsymbol{A}'(\boldsymbol{A}')^\top), \end{aligned}$$

and the second inequality follows from (4.12) and (4.9). Thus, $\|T\| \leq 2/\gamma$ and $\|S\| \leq 2/\gamma\beta$. By using Lemma 4.6, the triangle inequality, the submultiplicativity of matrix norms, and the fact that $\boldsymbol{Q}$, $\boldsymbol{A}'$ are submatrices of $\boldsymbol{H}$, we have:

$$\left\| \begin{bmatrix} T + \alpha T(\boldsymbol{A}')^\top S\boldsymbol{A}'T & T(\boldsymbol{A}')^\top S \\ S\boldsymbol{A}'T & \end{bmatrix} \right\| \leq \frac{2}{\gamma} + \frac{8\alpha D^4 L^2}{\gamma^3 \beta} + \frac{4D^2 L}{\gamma^2 \beta} \tag{4.13a}$$

$$\left\| \begin{bmatrix} \boldsymbol{I} & \alpha(\boldsymbol{A}')^\top \\ & \boldsymbol{I} \end{bmatrix} \right\| \leq 1 + \alpha D^2 L. \tag{4.13b}$$

Therefore, from (4.11) and (4.13), we obtain:

$$\left\| \begin{bmatrix} \boldsymbol{Q} & (\boldsymbol{A}')^\top \\ \boldsymbol{A}' & \end{bmatrix}^{-1} \right\| \leq \left(\frac{2}{\gamma} + \frac{8\alpha D^4 L^2}{\gamma^3 \beta} + \frac{4D^2 L}{\gamma^2 \beta} \right) \left(1 + \alpha D^2 L\right). \tag{4.14}$$

Because (4.14) holds for any $\mathcal{A}^0 \subset \mathcal{A} \subset \mathcal{A}^1$, the desired condition is obtained; the proof is complete. ∎

We have established in Lemmas 4.5, 4.7 that Assumptions 4.1, 4.2, 4.3, 4.4 guarantee the uniform boundedness condition (4.1). The result is summarized as follows.

THEOREM 4.9. *Under Assumptions 4.1, 4.2, 4.3, and 4.4 we have that:*

$$\overline{\overline{\sigma}}_{\boldsymbol{H},(k)} \leq D^2 L; \quad \overline{\overline{\sigma}}_{\boldsymbol{R},(k)} \leq D^2 L; \tag{4.15a}$$

$$\underline{\sigma}_{\boldsymbol{H},(k)} \geq \left(\frac{2}{\gamma} + \frac{8\alpha D^4 L^2}{\gamma^3 \beta} + \frac{4D^2 L}{\gamma^2 \beta} \right)^{-1} \left(1 + \alpha D^2 L\right)^{-1}, \tag{4.15b}$$

for $k \in K$ and where α is defined in Lemma 4.7; that is, (4.1) *holds.*

Proof. The result follows directly from Lemma 4.5, 4.7. ∎

All the quantities in Theorem 3.5 can be expressed using uniform parameters; therefore, we can uniformly bound the exponential decay parameters (Υ, ρ) using Theorem 4.9.

4.2. Sufficient Conditions for Uniform SSOSC and LICQ. Verifying Assumptions 4.3, 4.4 can be challenging if the size of $\boldsymbol{Q}_{(k)}$ and $\boldsymbol{A}_{(k)}$ grows indefinitely with k. Thus, in this section, we provide sufficient conditions for uniform SSOSC and LICQ that do not require checking singular values of indefinitely large matrices. The problems of interest can have arbitrarily large graphs (e.g., dynamic optimization with infinite horizons and PDE optimization with an unbounded domains). One key characteristic of such problems is that there exists a recurrent structure (as depicted in Figure 1). As such, we can construct sufficient conditions based on uniform SSOSC and LICQ over *blocks* of the Hessian and Jacobian matrices, defined by a partition $\mathscr{J}_{(k)} := \{J_{(k)(q)}\}_{q=1}^{\overline{q}_{(k)}}$ of the primal variable index set $\mathbb{I}_{[1,\boldsymbol{r}_{(k)}]}$. To state these assumptions, we define the following submatrices of $\boldsymbol{Q}_{(k)}$ and $\boldsymbol{A}_{(k)}$ for $q \in \mathbb{I}_{[1,\overline{q}_{(k)}]}$:

$$\begin{aligned} \boldsymbol{Q}_{(k)(q)} &:= \boldsymbol{Q}_{(k)}[J_{(k)(q)}, J_{(k)(q)}]; & \boldsymbol{Q}_{(k)(-q)} &:= \boldsymbol{Q}_{(k)}[J_{(k)(q)}, \mathbb{I}_{[1,\boldsymbol{r}_{(k)}]} \setminus J_{(k)(q)}]; \\ \boldsymbol{A}^-_{(k)(q)} &:= \boldsymbol{A}_{(k)}[\mathcal{A}^-_{(k)(q)}, J_{(k)(q)}]; & \boldsymbol{A}^+_{(k)(q)} &:= \boldsymbol{A}_{(k)}[\mathcal{A}^+_{(k)(q)}, J_{(k)(q)}], \end{aligned}$$

where:

$$\begin{aligned} \mathcal{A}^-_{(k)(q)} &:= \{i \in \mathcal{A}^0_{(k)} : \boldsymbol{A}_{(k)}[i, \mathbb{I}_{[1,\boldsymbol{r}_{(k)}]} \setminus J_{(k)(q)}] = \boldsymbol{0}\}; \\ \mathcal{A}^+_{(k)(q)} &:= \{i \in \mathcal{A}^1_{(k)} : \boldsymbol{A}_{(k)}[i, J_{(k)(q)}] \neq \boldsymbol{0}\}. \end{aligned}$$

Assumption 4.10 (Block Diagonal $\boldsymbol{Q}_{(k)}$). *$\boldsymbol{Q}_{(k)(-q)} = \boldsymbol{0}$ for $k \in K, q \in \mathbb{I}_{[1,\overline{q}_{(k)}]}$.*

Assumption 4.11 (Nonzero Rows of $\boldsymbol{A}_{(k)}$). *$\boldsymbol{A}_{(k)}[i,:] \neq \boldsymbol{0}$ for $k \in K, i \in \mathcal{A}^1_{(k)}$.*

Assumption 4.12 (Block SSOSC). *There exists $\gamma > 0$ (uniform in k, q) such that $\mathrm{ReH}(\boldsymbol{Q}_{(k)(q)}, \boldsymbol{A}^-_{(k)(q)}) \succeq \gamma \boldsymbol{I}$. holds for $k \in K$ and $q \in \mathbb{I}_{[1,\overline{q}_{(k)}]}$, where reduced Hessian $\mathrm{ReH}(\cdot,\cdot)$ is defined in Assumption 4.3.*

Assumption 4.13 (Block LICQ). *There exists $\beta > 0$ (uniform in k, q) such that $\boldsymbol{A}^+_{(k)(q)}(\boldsymbol{A}^+_{(k)(q)})^\top \succeq \beta \boldsymbol{I}$. holds for $k \in K$ and $q \in \mathbb{I}_{[1,\overline{q}_{(k)}]}$.*

We emphasize that Assumption 4.10 does not assume separability of the problem; a block-diagonal structure in $\boldsymbol{Q}_{(k)}$ is obtained when coupling across blocks exist only via linear constraints. This is not a restrictive assumption since any problem of the form (1.1) can be reformulated into a form with linear coupling by introducing auxiliary variables (i.e., via a *lifting* procedure). Assumption 4.11 is not difficult to satisfy.

In the following lemmas, we show that the above assumptions guarantee uniform SSOSC and LICQ for the original NLP (1.1).

Lemma 4.14. *Under Assumptions 4.10 and 4.12 we have $\mathrm{ReH}(\boldsymbol{Q}_{(k)}, \boldsymbol{A}^0_{(k)}) \succeq \gamma \boldsymbol{I}$*

Proof. Within this proof, we omit the subscript (k) for conciseness. From the block diagonal structure of $\boldsymbol{Q}$ (Assumption 4.10), $\boldsymbol{x}^\top \boldsymbol{Q} \boldsymbol{x} = \sum_{q=1}^{\overline{q}_{(k)}} \boldsymbol{x}[J_{(q)}]^\top \boldsymbol{Q}_{(q)} \boldsymbol{x}[J_{(q)}]$. If $\boldsymbol{A}^0 \boldsymbol{x} = \boldsymbol{0}$, $\boldsymbol{A}^-_{(q)} \boldsymbol{x}[J_{(q)}] = \boldsymbol{0}$ holds for $q \in \mathbb{I}_{[1,\overline{q}_{(k)}]}$; therefore, by Assumption 4.12, the following can be obtained: if $\boldsymbol{A}^0 \boldsymbol{x} = \boldsymbol{0}$,

$$\sum_{q=1}^{\overline{q}_{(k)}} \boldsymbol{x}[J_{(q)}]^\top \boldsymbol{Q}_{(q)} \boldsymbol{x}[J_{(q)}] \geq \sum_{q=1}^{\overline{q}_{(k)}} \gamma \|\boldsymbol{x}[J_{(q)}]\|^2 = \gamma \|\boldsymbol{x}\|^2. \tag{4.16}$$

Here, the last equality follows from the fact that $\{J_{(q)}\}_{q=1}^{\overline{q}_{(k)}}$ partitions $\mathbb{I}_{[1,\boldsymbol{r}_{(k)}]}$. From (4.16), we obtain the result. ∎

LEMMA 4.15. *Under Assumptions 4.11 and 4.13 we have* $\boldsymbol{A}^1_{(k)}(\boldsymbol{A}^1_{(k)})^\top \succeq \beta \boldsymbol{I}$.

Proof. Within this proof, we omit the subscript (k) for conciseness. We have that for any $\boldsymbol{y} \in \mathbb{I}_{[1,\boldsymbol{m}]}$,

$$\boldsymbol{y}[\mathcal{A}^1]^\top \boldsymbol{A}^1(\boldsymbol{A}^1)^\top \boldsymbol{y}[\mathcal{A}^1] = \boldsymbol{y}[\mathcal{A}^1]^\top (\sum_{q=1}^{\overline{q}_{(k)}} \boldsymbol{A}[\mathcal{A}^1, J_{(q)}](\boldsymbol{A}[\mathcal{A}^1, J_{(q)}])^\top)\boldsymbol{y}[\mathcal{A}^1]$$

$$= \sum_{q=1}^{\overline{q}_{(k)}} (\boldsymbol{y}[\mathcal{A}^+_{(q)}])^\top \boldsymbol{A}^+_{(q)}(\boldsymbol{A}^+_{(q)})^\top \boldsymbol{y}[\mathcal{A}^+_{(q)}].$$

Here the first equality follows from block multiplication formula and the second equality follows from the fact that $\boldsymbol{A}[\mathcal{A}^1 \setminus \mathcal{A}^+_{(q)}, J_{(q)}] = \boldsymbol{0}$. By Assumption 4.13,

$$\boldsymbol{y}[\mathcal{A}^1]^\top \boldsymbol{A}^1(\boldsymbol{A}^1)^\top \boldsymbol{y}[\mathcal{A}^1] \geq \sum_{q=1}^{\overline{q}_{(k)}} \beta \|\boldsymbol{y}[\mathcal{A}^+_{(q)}]\|^2 \geq \beta \|\boldsymbol{y}[\mathcal{A}^1]\|^2. \tag{4.17}$$

where the second inequality follows from the fact that $\bigcup_{q=1}^{\overline{q}_{(k)}} \mathcal{A}^+_{(q)} = \mathcal{A}^1$, which follows from Assumption 4.11. Inequality (4.17) implies the desired result. ∎

We now summarize the developments in this subsection in the following theorem.

THEOREM 4.16. *Under Assumptions 4.1, 4.2, 4.10, 4.11, 4.12, 4.13, we have that* (4.15) *holds for* $k \in K$*; that is,* (4.1) *holds.*

Proof. The result follows from Theorem 4.9 and Lemmas 4.14, 4.15. ∎

The results in Section 4.1-4.2 are useful for different problems of interest but might not be applicable to certain problem classes. For instance, it is difficult to derive uniform boundedness conditions for multi-stage stochastic programs because the probability of a given stage decays asymptotically over time (this prevents Assumption 4.12 to hold). This indicates that these types of problems might exhibit parasitic behavior that might manifest as extreme sensitivity (associated with non-uniqueness of the solution). We will leave specialized treatment for those problems as a topic of future work. Also, we have not discussed how the sensitivity behavior changes when the discretization resolution changes; such behavior can be used to understand sensitivity behavior of the continuous-time (infinite-dimensional) optimization problems studied in [17–19]. This is also left as a topic of future work.

4.3. Discussion. We now illustrate the practical applicability of our results using simple examples. With the intuition obtained from the examples, we discuss qualitative conditions under which the problem is likely to exhibit EDS.

Example 4.17. We first consider a PDE optimization problem for a steady-state thin-plate system described in [1]:

$$PDEO_k(p(\cdot)): \quad \min_{s(\cdot),u(\cdot)} \int_{w\in\Omega_{(k)}} \frac{1}{2}a(s(w) - s_{\text{ref}}(w))^2 + \frac{1}{2}u(w)^2 dw \tag{4.18a}$$

$$\text{s.t. } \Delta s(w) = \frac{2h_c}{\kappa t_z}(s(w) - \overline{T}) + \frac{2\epsilon\sigma}{\kappa t_z}(s(w)^4 - \overline{T}^4) \tag{4.18b}$$

$$- \frac{1}{\kappa t_z}(bu(w) + d(w)),\ w \in \Omega_{(k)}$$

$$\nabla s(w) \cdot \hat{\boldsymbol{n}}(w) = 0,\ w \in \partial\Omega_{(k)}, \tag{4.18c}$$

where $\Omega_{(k)} = [0,k] \times [0,k] \subseteq \mathbb{R}^2$ is the 2-dimensional domain of interest; $\partial\Omega_{(k)}$ is the boundary of $\Omega_{(k)}$; $s : \Omega_{(k)} \to \mathbb{R}$ is the temperature; $u : \Omega_{(k)} \to \mathbb{R}$ is the control; $d : \Omega_{(k)} \to \mathbb{R}$ is the disturbance; Δ is the Laplacian operator; $\hat{\boldsymbol{n}}$ is the unit normal vector; $\cdot$ is the inner product; (4.18b) is the heat equation whose right-hand-side consists of convection, radiation, and forcing terms by control and disturbance; (4.18c) is the Neumann boundary condition (i.e., insulated); $s_{\text{ref}} : \Omega_{(k)} \to \mathbb{R}$ is the desired temperature; $\kappa = 400$, $t_z = .01$, $h_c = 1$, $\epsilon = .5$, $\sigma = 5.67 \times 10^{-8}$, and $\overline{T} = 300$ are the constant parameters. We define the variable vector to be $x(w) = [s(w); u(w)]$; the data vector to be $p(w) := [s_{\text{ref}}(w); d(w)]$ for $w \in \Omega_{(k)}$. We consider a discretized version of $PDEO_k(p(\cdot))$ (e.g., each $[i,j] \times [i+1, j+1]$ cell for $i, j = 0, \cdots, k-1$); we consider the discretization mesh as the graph $\mathcal{G}$ and let $K = \mathbb{I}_{>0}$. Observe that the domain expands with k but the fineness of the discretization mesh remains the same.

We assume that $s_{\text{ref}}(w) = d(w) = \overline{T}$ for $w \in \Omega$; in this case, the problem admits a trivial solution $s^\star(w) = \overline{T}$ and $u^\star(w) = 0$. Since there are no inequality constraints, $\boldsymbol{A}^0_{(k)} = \boldsymbol{A}^1_{(k)} = \boldsymbol{A}_{(k)}$. Furthermore, there exist permutations of the Lagrangian Hessian $\boldsymbol{Q}_{(k)}$ and of constraint Jacobian $\boldsymbol{A}_{(k)}$ that can be expressed as:

$$\widetilde{\boldsymbol{Q}}_{(k)} = \Pi_{(k)} \boldsymbol{Q}_{(k)} \Pi_{(k)} = \begin{bmatrix} \boldsymbol{I} & \\ & a\boldsymbol{I} \end{bmatrix}; \quad \widetilde{\boldsymbol{A}}_{(k)} := \boldsymbol{A}_{(k)} \Pi_{(k)} = \begin{bmatrix} \boldsymbol{L}_{(k)} + c_0 \boldsymbol{I} & (b/b_0)\boldsymbol{I} \end{bmatrix},$$

where $\Pi_{(k)}$ is the permuation operator, $\boldsymbol{L}_{(k)}$ is the graph Laplacian matrix induced by the mesh graph, $c_0 = (2h_c + 8\epsilon\sigma\overline{T}^3)/\kappa t_z$ and $b_0 = -1/\kappa t_z$. It is easy to see that Assumptions 4.1, 4.2 hold; thus, upper bounds in (4.1) are satisfied.

We now show that uniform SSOSC and LICQ hold. One can observe that the uniform LICQ condition holds if $b \neq 0$ from $\widetilde{\boldsymbol{A}}_{(k)} (\widetilde{\boldsymbol{A}}_{(k)})^\top = (\boldsymbol{L}_{(k)} + c_0 \boldsymbol{I})^2 + (b/b_0)^2 \boldsymbol{I} \succeq (b/b_0)^2 \boldsymbol{I}$. Furthermore, since the smallest eigenvalue of $Z^\top \widetilde{\boldsymbol{Q}}_{(k)} Z$ with orthogonal Z is always greater than or equal to the smallest eigenvalue of $\widetilde{\boldsymbol{Q}}_{(k)}$, we have that $\mathrm{ReH}(\widetilde{\boldsymbol{Q}}_{(k)}, \widetilde{\boldsymbol{A}}_{(k)}) \succeq aI$. Thus, if $a > 0$, uniform SSOSC holds. This example demonstrates that the developments of Section 4.1 can be applied to check uniform boundedness.

Example 4.18. Consider a dynamic optimization problem for energy storage:

$$OCP_{(k)}(\boldsymbol{p}) : \min_{\{s_i, u_i, v_i\}_{i=1}^k} \sum_{i=1}^k \frac{1}{2} a(s_i)^2 + \frac{1}{2}(u_i)^2 + \pi_i v_i \tag{4.19a}$$

$$\text{s.t. } s_1 = \overline{s},\ (\mu_1) \tag{4.19b}$$

$$s_i = s_{i-1} + b u_{i-1} + w_i,\ i \in \mathbb{I}_{[2,k]},\ (\mu_i) \tag{4.19c}$$

$$v_i = u_i + d_i,\ i \in \mathbb{I}_{[1,k]},\ (\nu_i). \tag{4.19d}$$

Here, $s_i \in \mathbb{R}$ is the stored energy (state) at time i; $u_i \in \mathbb{R}$ is the charge/discharge of energy (control); $v_i \in \mathbb{R}$ are the transactions with the grid; $\overline{s} = w_1$ is the initial storage; $\pi_i \in \mathbb{R}$ is the energy price forecast; d_i is the energy demand forecast; and w_i is the disturbance forecast. We define $x_i := [s_i, u_i, v_i]$ as the primal variables; $y_i := [\mu_i, \nu_i]$ as the dual variables; $p_i = [\pi_i; w_i; d_i]$ as the data. We consider $\mathcal{G}_{(k)} = (\mathcal{V}_{(k)}, \mathcal{E}_{(k)})$ as a linear graph that represents time domain, $\mathcal{V}_{(k)} := \mathbb{I}_{[1,k]}$ and $K = \mathbb{I}_{>0}$. Practical problems have inequality constraints for s_i, u_i, but here we neglect them for simplicity. The index k is the length of the horizon.

Given any $k \in K$, problem $OCP_{(k)}(\boldsymbol{p})$ satisfies SSOSC and LICQ, regardless of the choice of (a, b), but (4.1) can be violated. For example, if $b = 0$, $w_1 = 1$, and $w_2 = w_3 = \cdots = 0$, one can see that $s_i = 1$ for $i \in \mathbb{I}_{[1,k]}$; thus, $\underline{\sigma}(\boldsymbol{H}_{(k)}) \to 0$. On the other hand, if $a > 0$ and $b \neq 0$, we will see below that $\underline{\sigma}(\boldsymbol{H}_{(k)})$ can be uniformly bounded below by a strictly positive number. Other parts of (4.1) can be easily confirmed from Theorem 4.16. Since there is no inequality constraint, $\boldsymbol{A}^0 = \boldsymbol{A}^1 = \boldsymbol{A}$.

It is clear that Assumptions 4.1, 4.2 hold for Example 4.18, but it is difficult to check whether Assumptions 4.3, 4.4 (uniform SSOSC and LICQ) hold. Specifically, we need to show that there exist some uniform constants $\gamma, \beta > 0$ such that $\mathrm{ReH}(\boldsymbol{Q}_{(k)}, \boldsymbol{A}_{(k)}) \succeq \gamma \boldsymbol{I}$ and $\boldsymbol{A}_{(k)}(\boldsymbol{A}_{(k)})^\top \succeq \beta I$ hold, where

$$\boldsymbol{Q}_{(k)} = \begin{bmatrix} a & & & & & & \\ & 1 & & & & & \\ & & 0 & & & & \\ & & & \ddots & & & \\ & & & & a & & \\ & & & & & 1 & \\ & & & & & & 0 \end{bmatrix}_{3k \times 3k} \tag{4.20a}$$

$$\boldsymbol{A}_{(k)} = \begin{bmatrix} 1 & & & & & & & & & \\ & -1 & 1 & & & & & & & \\ -1 & -b & & 1 & & & & & & \\ & & & -1 & 1 & & & & & \\ & & & & & \ddots & & & & \\ & & & & & & -1 & -b & & 1 \\ & & & & & & & & -1 & 1 \end{bmatrix}_{2k \times 3k}. \tag{4.20b}$$

The difficulty arises from the sizes of $\boldsymbol{Q}_{(k)}$ and $\boldsymbol{A}_{(k)}$ (which grow indefinitely with k); hence, we leverage the result in Section 4.2. If we choose $J_{(k)(q)} = \mathbb{I}_{[3(q-1)+1, 3q]}$ for $q \in \mathbb{I}_{[1,k]}$,

(4.21)

$$\boldsymbol{Q}_{(k)(q)} = \begin{bmatrix} a & & \\ & 1 & \\ & & 0 \end{bmatrix}; \qquad \boldsymbol{Q}_{(k)(-q)} = \boldsymbol{0};$$

$$\boldsymbol{A}^-_{(k)(q)} = \begin{cases} \begin{bmatrix} 1 & & \\ 0 & -1 & 1 \end{bmatrix}, & q = 1 \\ \begin{bmatrix} 0 & -1 & 1 \end{bmatrix}, & q \in \mathbb{I}_{[2,k]} \end{cases}; \quad \boldsymbol{A}^+_{(k)(q)} = \begin{cases} \begin{bmatrix} 1 & & \\ & -1 & 1 \\ -1 & -b & \end{bmatrix} & q \in \mathbb{I}_{[1,k-1]} \\ \begin{bmatrix} 1 & & \\ & -1 & 1 \end{bmatrix} & q = k \end{cases}.$$

One can confirm from (4.21) that Assumptions 4.12, 4.13 are satisfied if $a > 0$ and $b \neq 0$ hold. Thus, Example 4.18 with $a > 0$ and $b \neq 0$ satisfies (4.1); accordingly, (Υ, ρ) are uniformly bounded. This demonstrates the usefulness of the results in Section 4.2 in checking the uniform boundedness conditions.

The above results provide *qualitative* conditions under which quantities in (4.1) are likely to be moderately bounded (and thus the problem exhibits EDS). The first is having *sufficient positive curvature* in the objective function (related to Assumptions 4.3, 4.12), and the second is having a *sufficient flexibility* in the constraints (related to Assumptions 4.4, 4.13). Here, flexibility is defined in the sense that a small perturbation on the data does not require a big adjustment of the decision variables to restore feasibility (thus is related to the smallest non-trivial singular value of the active constraint Jacobian). For Example 4.17-4.18, having larger a produces stronger positive curvature and larger b yields more flexibility in the constraints (makes the control more impactful). Intuitively, the first qualitative condition helps the decay of sensitivity because positive curvature produces a direction to which the solution tends and the second qualitative condition helps the decay of sensitivity as it enables the solutions to dampen the impact of perturbations. These conditions can be related to specific properties of particular problem classes; for example, for the dynamic optimization problems analyzed in [22], it can be seen that uniform LICQ is related to uniform controllability. Similarly, it is known that observability is directly related to SSOSC for state and parameter estimation problems [35]. Establishing such connections for specific problem instances is an interesting topic of future work. The results provided in this work are general and provide a unified view of different problem classes.

5. Numerical Studies.

In this section, we illustrate the theoretical developments with different classes of graph-structured problems. We are particularly interested in exploring the effect of positive curvature of the objective function and flexibility of the constraints on the exponential decay of sensitivity. To this end, we use parameters (a, b) to manipulate such behavior. We conduct case studies for four different classes of graph-structured optimization problems: dynamic optimization, stochastic optimization, PDE-constrained optimization, and network optimization. In what follows, we describe the particular problem instances under study. Throughout the instances, a and b are parameters that control positive objective curvature and flexibility, respectively, and j represents the node where the data perturbation will be introduced.

5.1. Dynamic Optimization.

The study is performed for Example 4.18. We set $w_i = d_i = \pi_i = 0$ for $i \in \mathcal{V}_{(k)}$, $k = 9$, and $j = 5$. Numerical sensitivity study with more sophisticated dynamic optimization problems (e.g., with nonlinearities) are provided in [24,31].

5.2. Stochastic Optimization.

We consider a stochastic program:

$$\min_{\{s_i, u_i, v_i \in \mathbb{R}\}_{i\in\mathcal{V}}} \sum_{i\in\mathcal{V}} p_i\left(a\frac{1}{2}s_i^2 + \frac{1}{2}u_i^2 + \pi_i v_i\right) \tag{5.1a}$$

$$\text{s.t. } s_1 = \overline{s},\ (\mu_1) \tag{5.1b}$$

$$s_i = s_{an(i)} + b u_{an(i)} + w_i,\ i \in \mathcal{V}\setminus\{1\},\ (\mu_i), \tag{5.1c}$$

$$v_i = u_i + d_i,\ i \in \mathcal{V},\ (\nu_i). \tag{5.1d}$$

Here $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ represents the scenario tree; $1 \in \mathcal{V}$ is the root node; $an(i) \in N_{\mathcal{G}}[i]$ denotes the parent node; $p_i \in \mathbb{R}_{\geq 0}$ denotes the probability of node i; $s_i \in \mathbb{R}$ is the stored energy at node i; $u_i \in \mathbb{R}$ is the charge/discharge of energy at node i; $v_i \in \mathbb{R}$ is the transactions with grid at node i; $\overline{s} = w_1$ is the initial energy storage; $\pi_i \in \mathbb{R}$ is the forecast energy price at node i; d_i is the energy demand forecast at node i; w_i is the disturbance forecast at node i. We set $x_i := [s_i, u_i, v_i]$ as the primal variable vector at node i; $y_i := [\mu_i, \nu_i]$ as the dual variable vector at node i; $p_i = [\pi_i; w_i; d_i]$ as the data at node i. Furthermore, we set $|c(i)| = 3$, where $c(i) \subseteq N_{\mathcal{G}}[i]$ is the set of children nodes, $d_1 = 0$; $\{d_j\}_{j\in c(i)} = [-1; 0; 1]$; $w_i = \pi_i = 0$ for $i \in \mathcal{V}$; $k = 6$, where k denotes the number of stages. We choose $j = 3$ (a node in the second stage).

5.3. PDE Optimization. The study is performed for Example 4.17 with $k = 9$ (9×9 grid). We choose j to be the node at the center $(5, 5)$. For the base data, we set $s_{\text{ref}}(w) = d(w) = \overline{T}$ for $w \in \Omega$.

5.4. Network Optimization. We consider the alternating current (AC) optimal power flow problem:

$$\min_{\substack{\{v_i\in\mathbb{C}\}_{i\in\mathcal{V}}\\ \{s^g_k\in\mathbb{C}\}_{k\in\mathcal{W}}\\ \{s_{ij}\in\mathbb{C}\}_{i,j\in\mathcal{V}}}} a\Big(\sum_{i\in\mathcal{V}}(|v_i| - v_{\text{ref}})^2 + \sum_{\{i,j\}\in\mathcal{E}}(\angle v_i v_j^*)^2\Big) + \sum_{k\in\mathcal{W}} c^1_{(k)}\Re(s^g_{(k)}) + c^2_{(k)}\Re(s^g_{(k)})^2 \tag{5.2a}$$

$$\text{s.t. } \angle v_i = 0,\ i \in \mathcal{V}_{\text{ref}} \tag{5.2b}$$

$$s^{gL}_{(k)} - b(1+\sqrt{-1}) \le s^g_{(k)} \le s^{gU}_{(k)} + b(1+\sqrt{-1}),\ k \in \mathcal{W} \tag{5.2c}$$

$$\sum_{k\in\mathcal{W}_i} s^g_k - s^d_i = \sum_{j\in N_{\mathcal{G}}[i]} s_{ij},\ v^L_i \le |v_i| \le v^U_i,\ i \in \mathcal{V} \tag{5.2d}$$

$$s_{ij} = (Y_{ij} + Y^c_{ij})^* \frac{|v_i|^2}{|T_{ij}|} v_i^* - Y_{ij}\frac{v_i v_j^*}{T_{ij}},\ |s_{ij}| \le s^U_{ij},\ i,j \in \mathcal{V} \tag{5.2e}$$

$$\theta^{\Delta L}_{ij} \le \angle v_i v_j^* \le \theta^{\Delta U}_{ij},\ \{i,j\} \in \mathcal{E}. \tag{5.2f}$$

Here, $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ represents the power network, $\mathbb{C}$ denotes the set of complex numbers; $\Re(\cdot)$ and $\Im(\cdot)$ denotes the real and imaginary part of the argument; $(\cdot)^*$ denotes the complex conjugate of the argument; $p \ge q \iff \Re(p) \ge \Re(q)$ and $\Im(p) \ge \Im(q)$ for $p, q \in \mathbb{C}$; $\mathcal{W}_i$ is the set of generators connected to node i; $\mathcal{W} := \bigcup_{i\in\mathcal{V}} \mathcal{W}_i$; $\mathcal{V}_{\text{ref}}$ is the set of reference nodes; $v_i \in \mathbb{C}$ is the voltage at node i; $s^g_{(k)} \in \mathbb{C}$ is the power generation at generator k; $\{v^L_i, v^U_i, s^d_i \in \mathbb{C}\}_{i\in\mathcal{V}}$,$\{\theta^{\Delta,L}_{ij}, \theta^{\Delta,U}_{ij} \in \mathbb{R}\}_{\{i,j\}\in\mathcal{E}}$, $\{s^U_{ij}, Y_{ij}, Y^c_{ij}, T_{ij} \in \mathbb{C}\}_{i,j\in\mathcal{V}}$, $\{c^1_{(k)}, c^2_{(k)} \in \mathbb{R}, s^{gL}_{(k)}, s^{gU}_{(k)} \in \mathbb{C}\}_{k\in\mathcal{W}}$ are the data. The readers are pointed to the documentation of `PowerModels.jl` [10] for the details of Problem (5.2). Here we modified the problem by adding the regularization term (the first term in (5.2a)) and by introducing the additional terms in constraint (5.2c) to examine the effect of positive curvature and flexibility in the constraints; the problem reduces to the original problem when $(a, b) = 0$. We treat the edge variables, constraints, and the objective terms by treating them as node terms for one of the connected node (in particular, the one with lower node index), as explained in Remark 1.1; note that this manipulation only alters indexing and does not change the problem. We set z_i as all the primal/dual variable that are associated with node $i \in \mathcal{V}$ (including generator and edge variables/constraints), and we set $p_i = [\Re(s^d_i), \Im(s^d_i)]$. We choose $j = 1$ as the perturbation location. We use test case `pglib_opf_case500_tamu` available at `pglib-opf` v18.08 [2, 8] (the problem data $c^1_{(k)}$, $c^2_{(k)}$, V^L_i, etc are available therein). The problem is modeled using modeling library `PowerModels.jl`.

5.5. Methods. We conduct the following numerical study for each problem instance. We consider a problem $P(\boldsymbol{p}^\star)$ with the base data $\boldsymbol{p}^\star$. Then, we consider perturbed problems $\{P(\boldsymbol{p}^{(q)})\}_{q=1}^{\overline{q}}$ in which the data are perturbed as $\boldsymbol{p}^{(q)} = \boldsymbol{p}^\star + \Delta\boldsymbol{p}^{(q)}$, where $\Delta\boldsymbol{p}^{(q)}$ are i.i.d samples drawn from $\Delta p_j \sim \mathcal{U}([-\sigma, \sigma]^{l_j})$, and $\Delta p_i = \mathbf{0}$ if $i \neq j$. Here, $j \in \mathcal{V}$ is a selected perturbation point and $\mathcal{U}(\Omega)$ denotes the multivariate uniform distribution on Ω. We choose $\sigma = 10^{-3}$ and $\overline{q} = 30$ for all instances. Then, the

Table 1
Variation of (a, b) in numerical studies.

	Case 1	Case 2	Case 3	Case 4
Dynamic Optimization	$(1,1)$	$(10^{-2},1)$	$(1,10^{-2})$	$(10^{-2},10^{-2})$
Stochastic Optimization	$(1,1)$	$(10^{-2},1)$	$(1,10^{-2})$	$(10^{-2},10^{-2})$
PDE Optimization	$(1,1)$	$(10^{-2},1)$	$(1,10^{-2})$	$(10^{-2},10^{-2})$
Network Optimization	$(10^6,10)$	$(0,10)$	$(10^6,0)$	$(0,0)$

empirical sensitivity coefficients:

$$\overline{C}_{ij} = \max_{q\in\mathbb{I}_{[1,\overline{q}]}} \|z_i^{\ddagger}(\boldsymbol{p}^{(q)})\|/\|\Delta\boldsymbol{p}^{(q)}\|, \quad i,j\in\mathcal{V}$$

are computed and visualized. The empirical sensitivity $\overline{C}_{ij}$ converges to $\|\nabla_{p_j} z_i^{\dagger}(\boldsymbol{p}^{\star})\|$ as $\sigma\to 0$ and $\overline{q}\to\infty$; thus, these empirical sensitivities are suitable quantities for study of sensitivity coefficients. We recall that (a, b) are the key parameters that control the positive curvature and flexibility. We vary these parameters as shown in Table 1, and see how they affect the decay (spread) of the sensitivity coefficients. Here, Case 1 has sufficiently large (a, b); Case 2 has low a; Case 3 has low b; and Case 4 has low (a, b). The results can be reproduced using the scripts provided in https://github.com/zavalab/JuliaBox/tree/master/SensitivityNLP.

5.6. Results. The sensitivity results are illustrated as heat maps of the empirical coefficients (Figure 2) and as scatter plots of the coefficients against distance $d_{\mathcal{G}}(i, j)$ (Figure 3). From Figure 2 we see that, with sufficiently large (a, b) (Case 1), the empirical sensitivity coefficients decay as they move away from the perturbation location. Furthermore, from Figure 3, one can confirm that the sensitivity coefficients decay exponentially with distance (i.e., $\log C_{ij} \propto d_{\mathcal{G}}(i, j)$). This verifies the theoretical results in Section 3. If either one or both of (a, b) are not sufficiently large (Case 2, 3, 4), the decay of sensitivity is weaker or not observed (except for the PDE optimization problem). This is because, without strong curvature or flexibility, $\underline{\sigma}(\boldsymbol{H}_{(k)})$ may be close to zero, and the coefficients in Theorem 3.5 do not exhibit sufficient decay. The reason that the PDE optimization problem exhibits decay of sensitivity even in the absence of positive curvature and flexibility is that the system itself has a strong dissipative property (temperature naturally tends towards ambient temperature via convection and radiation). From these results, we can confirm that it is sufficient for problems to have strong positive curvature and flexibility in the constraints to exhibit decay of sensitivity (this confirms the theoretical results in Section 4). Notably, even though we cannot guarantee uniform boundedness of the multi-stage stochastic programs, we can observe EDS for sufficiently large (a, b).

6. Conclusions. We have presented sensitivity results for graph-structured nonlinear programs (problems whose structure is induced by a graph). Our results indicate that the sensitivity of the solution at a given location decays exponentially with respect to the distance to the perturbation point. This result holds under the strong second-order sufficiency condition (SSOSC) and the linear independence constraint qualification (LICQ). We show that sensitivity decay depends on the singular values of the submatrices of the Lagrangian Hessian matrix; as such, uniform boundedness conditions for such singular values are essential. We have also shown that the singular values can be uniformly bounded under bounded graph degree, bounded second-order derivatives, and uniform SSOSC and LICQ conditions. Furthermore,

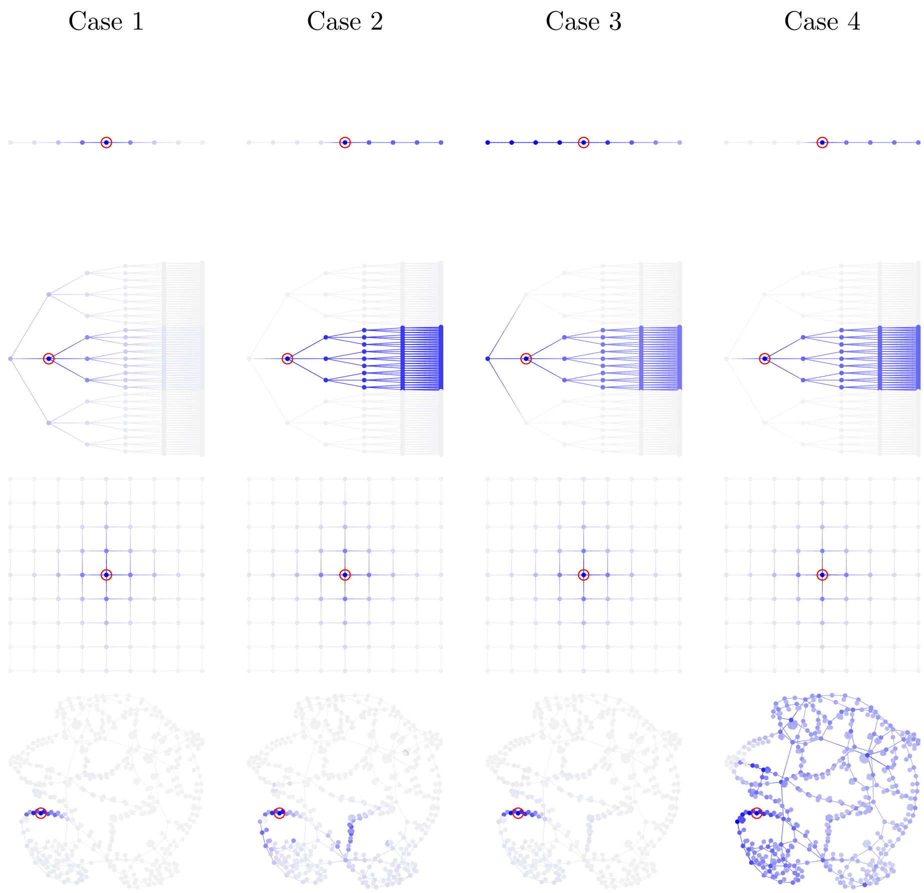

FIG. 2. *Spread of empirical sensitivity coefficients $\overline{C}_{ij}/\overline{C}_{jj}$ on $\mathcal{G}$ for dynamic optimization (top), stochastic optimization (second row), PDE optimization (third row), and network optimization (bottom) problem for different values of (a,b). Red circles denote perturbation point, dark blue approaches one, and white approaches zero.*

we have shown that uniform SSOSC and LICQ conditions can be guaranteed under uniform SSOSC and LICQ at problem blocks. Qualitatively, these conditions can be interpreted as having sufficiently strong positive curvature in the objective and flexibility in the constraints. Our numerical studies (for dynamic optimization, stochastic optimization, network optimization, and PDE optimization) confirm that, if the graph-structured problem exhibits such properties, sensitivity indeed decays exponentially. As part of future work, we are interested in exploring the propagation of sensitivity in specific problem classes. A sensitivity decay property for continuous-time (infinite-dimensional) dynamic optimization problems has been recently established in [17–19]. We are interested in studying the limiting behavior of our graph sensitivity results to establish a similar result for such problems.

Acknowledgments. 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. We also acknowledge support from the Grainger Wisconsin Distin-

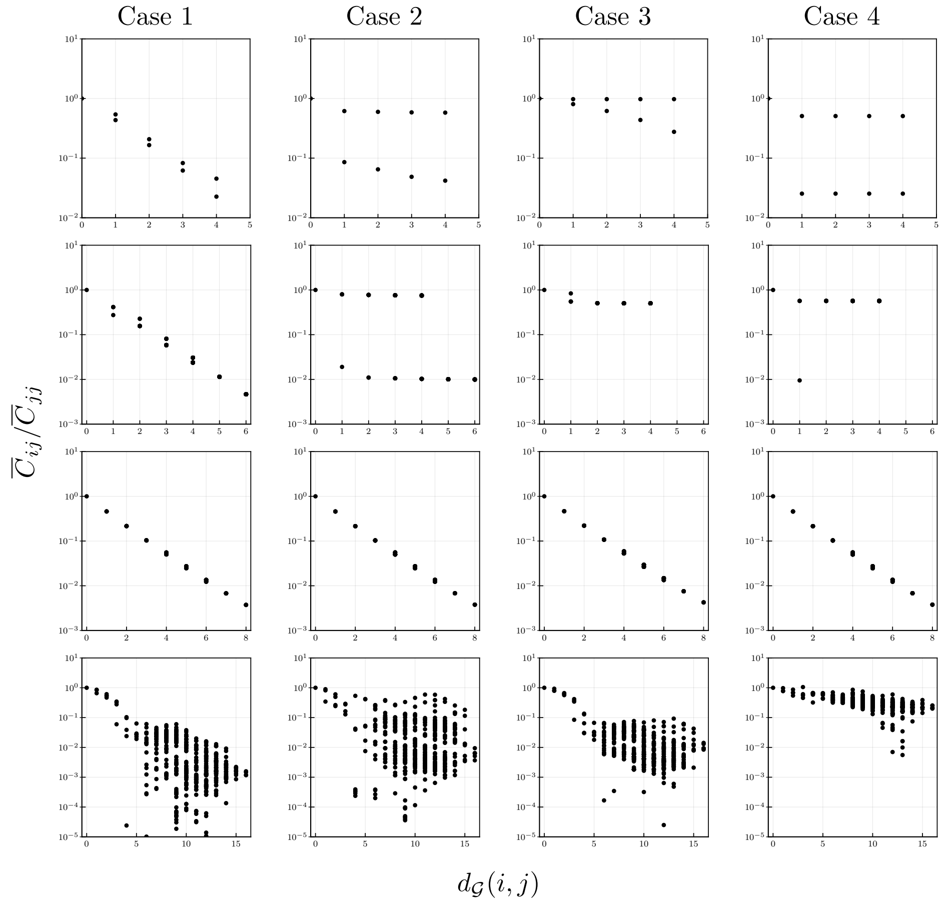

FIG. 3. *Scatter plots of sensitivity coefficients* $\overline{C}_{ij}/\overline{C}_{jj}$ *versus* $d_{\mathcal{G}}(i,j)$ *for dynamic optimization (top), stochastic optimization (second row), PDE optimization (third row), and network optimization (bottom) for different values of* (a,b).

guished Graduate Fellowship.

REFERENCES

[1] *Nonlinear heat transfer in thin plate*, https://www.mathworks.com/help/pde/ug/nonlinear-heat-transfer-in-a-thin-plate.html. https://www.mathworks.com/help/pde/ug/nonlinear-heat-transfer-in-a-thin-plate.html.

[2] S. Babaeinejadsarookolaee, A. Birchfield, R. D. Christie, C. Coffrin, C. DeMarco, R. Diao, M. Ferris, S. Fliscounakis, S. Greene, R. Huang, et al., *The power grid library for benchmarking ac optimal power flow algorithms*, arXiv preprint arXiv:1908.02788, (2019).

[3] D. S. Bernstein, *Matrix mathematics: theory, facts, and formulas*, Princeton university press, 2009.

[4] J. T. Betts, *Practical methods for optimal control and estimation using nonlinear programming*, vol. 19, Siam, 2010.

[5] L. T. Biegler, *Nonlinear programming: concepts, algorithms, and applications to chemical processes*, vol. 10, Siam, 2010.

[6] L. T. Biegler, *A survey on sensitivity-based nonlinear model predictive control*, IFAC Proceedings Volumes, 46 (2013), pp. 499–510.

[7] L. T. Biegler, O. Ghattas, M. Heinkenschloss, and B. van Bloemen Waanders, *Large-scale pde-constrained optimization: an introduction*, in Large-Scale PDE-Constrained Optimization, Springer, 2003, pp. 3–13.
[8] A. B. Birchfield, T. Xu, K. M. Gegner, K. S. Shetye, and T. J. Overbye, *Grid structural characteristics as validation criteria for synthetic networks*, IEEE Transactions on power systems, 32 (2016), pp. 3258–3265.
[9] J. F. Bonnans and A. Shapiro, *Perturbation analysis of optimization problems*, Springer Science & Business Media, 2013.
[10] C. Coffrin, R. Bent, K. Sundar, Y. Ng, and M. Lubin, *Powermodels. jl: An open-source framework for exploring power flow formulations*, in 2018 Power Systems Computation Conference (PSCC), IEEE, 2018, pp. 1–8.
[11] R. S. Dembo, J. M. Mulvey, and S. A. Zenios, *OR practice-large-scale nonlinear network models and their application*, Operations Research, 37 (1989), pp. 353–372.
[12] S. Demko, W. F. Moss, and P. W. Smith, *Decay rates for inverses of band matrices*, Mathematics of computation, 43 (1984), pp. 491–499.
[13] M. Diehl, H. G. Bock, J. P. Schlöder, R. Findeisen, Z. Nagy, and F. Allgöwer, *Real-time optimization and nonlinear model predictive control of processes governed by differential-algebraic equations*, Journal of Process Control, 12 (2002), pp. 577–585.
[14] A. L. Dontchev and R. T. Rockafellar, *Implicit functions and solution mappings*, vol. 543, Springer, 2009.
[15] A. V. Fiacco, *Sensitivity analysis for nonlinear programming using penalty methods*, Mathematical programming, 10 (1976), pp. 287–311.
[16] G. H. Golub and C. F. Van Loan, *Matrix computations*, vol. 3, JHU press, 2012.
[17] L. Grüne, M. Schaller, and A. Schiela, *Sensitivity analysis of optimal control for a class of parabolic pdes motivated by model predictive control*, SIAM Journal on Control and Optimization, 57 (2019), pp. 2753–2774.
[18] L. Grüne, M. Schaller, and A. Schiela, *Abstract nonlinear sensitivity and turnpike analysis and an application to semilinear parabolic pdes*, arXiv preprint arXiv:2008.13001, (2020).
[19] L. Grüne, M. Schaller, and A. Schiela, *Exponential sensitivity and turnpike analysis for linear quadratic optimal control of general evolution equations*, Journal of Differential Equations, 268 (2020), pp. 7311–7341.
[20] R. A. Horn and C. R. Johnson, *Matrix analysis*, Cambridge university press, 2012.
[21] J. Jalving, Y. Cao, and V. M. Zavala, *Graph-based modeling and simulation of complex systems*, Computers & Chemical Engineering, 125 (2019), pp. 134–154.
[22] S. Na and M. Anitescu, *Exponential decay in the sensitivity analysis of nonlinear dynamic programming*, To appear in SIAM Journal on Optimization, (2019), https://arxiv.org/abs/1912.06734.
[23] S. Na and M. Anitescu, *Superconvergence of online optimization for model predictive control*, arXiv preprint arXiv:2001.03707, (2020).
[24] S. Na, S. Shin, M. Anitescu, and V. M. Zavala, *Overlapping schwarz decomposition for nonlinear optimal control*, (2020). Under Review.
[25] J. Nocedal and S. Wright, *Numerical optimization*, Springer Science & Business Media, 2006.
[26] M. V. Pereira and L. M. Pinto, *Multi-stage stochastic optimization applied to energy planning*, Mathematical programming, 52 (1991), pp. 359–375.
[27] S. M. Robinson, *Perturbed Kuhn-Tucker points and rates of convergence for a class of nonlinear-programming algorithms*, Mathematical programming, 7 (1974), pp. 1–16.
[28] S. M. Robinson, *Strongly regular generalized equations*, Mathematics of Operations Research, 5 (1980), pp. 43–62.
[29] A. Shapiro, D. Dentcheva, and A. Ruszczyński, *Lectures on stochastic programming: modeling and theory*, SIAM, 2014.
[30] S. Shin, M. Anitescu, and V. M. Zavala, *Overlapping schwarz decomposition for constrained quadratic programs*, arXiv preprint arXiv:2003.07502, (2020).
[31] S. Shin, T. Faulwasser, M. Zanon, and V. M. Zavala, *A parallel decomposition scheme for solving long-horizon optimal control problems*, in 2019 IEEE 58th Conference on Decision and Control (CDC), 2019, pp. 5264–5271.
[32] S. Shin and V. M. Zavala, *Diffusing-horizon model predictive control*, arXiv preprint arXiv:2002.08556, (2020).
[33] S. Shin, V. M. Zavala, and M. Anitescu, *Decentralized schemes with overlap for solving graph-structured optimization problems*, IEEE Transactions on Control of Network Systems, (2020).
[34] W. Xu and M. Anitescu, *Exponentially accurate temporal decomposition for long-horizon*

linear-quadratic dynamic optimization, SIAM Journal on Optimization, 28 (2018), pp. 2541–2573.

[35] V. M. Zavala, *Stability analysis of an approximate scheme for moving horizon estimation*, Computers & Chemical Engineering, 34 (2010), pp. 1662–1670.

[36] A. Zlotnik, M. Chertkov, and S. Backhaus, *Optimal control of transient flow in natural gas networks*, in 2015 54th IEEE conference on decision and control (CDC), IEEE, 2015, pp. 4563–4570.