# Quasinormality and pseudonormality for nonlinear semidefinite programming*

Daiana Oliveira dos Santos†

February 8, 2026

**Abstract**

Quasinormality is a classical constraint qualification originally introduced by Hestenes in 1975 and subsequently extensively studied in nonlinear programming and in problems with abstract constraints. In this paper, we extend this concept to the setting of nonlinear semidefinite programming (NSDP). We show that the proposed condition is strictly weaker than Robinson's constraint qualification, while still guaranteeing the existence of exact penalty functions, local error bounds, and boundedness of dual sequences generated by augmented Lagrangian methods. As a consequence, convergence to Karush–Kuhn–Tucker points can be established for a broader class of NSDPs under mild regularity assumptions. In addition, a pseudonormality condition is introduced and explored.

**Key words:** Nonlinear Semidefinite Programming, Augmented Lagrangian Methods, Constraint Qualifications, Quasinormality, Pseudonormality

**AMS subject classifications:** 90C30, 90C22, 65K05.

## 1 Introduction

Nonlinear semidefinite programming (NSDP) generalizes classical nonlinear programming by incorporating semidefinite constraints. These problems arise in numerous practical applications, including control theory, structural and material optimization, robust estimation, and eigenvalue problems (see, for instance, [16–18, 21, 22, 27, 31, 32]). From a theoretical perspective, key topics such as optimality conditions, duality theory, and constraint qualifications have received considerable attention in the literature, alongside the development of numerical algorithms for solving NSDPs [1, 24, 30, 33, 35].

A common feature of modern methods for NSDP is their primal-dual nature: they generate iterates of both the primal variable and Lagrange multipliers. To guarantee convergence, classical regularity conditions such as Robinson's constraint qualification or nondegeneracy are often imposed [29, 33]. However, these conditions may be too restrictive for many practical problems, particularly in the presence of redundancies or degeneracies in the constraints.

To address these limitations, weaker constraint qualifications have been introduced in recent years. Notably, constant rank-type conditions and their sequential variants [5] have proven effective in analyzing convergence under milder assumptions. Nonetheless, much remains to be explored

*The author received financial support from FAPESP 2023/08706-1 and CNPq grants 407147/2023-3

†Paulista School of Politics, Economics and Business, Federal University of São Paulo, Osasco-SP, Brazil. e-mail: daiana.santos@unifesp.br

regarding the theoretical foundation of NSDPs, especially when compared to classical nonlinear programming (NLP), where a broad array of constraint qualifications such as the Mangasarian-Fromovitz condition (MFCQ), linear independence (LICQ), and constant positive linear dependence (CPLD) has been extensively studied [1, 11].

In the NLP literature, quasinormality was originally introduced by Hestenes [20] as a relatively weak constraint qualification that nonetheless guarantees the existence of Lagrange multipliers and the validity of the KKT conditions at local minima. While weaker than classical conditions such as LICQ or MFCQ, quasinormality plays a fundamental role in the theoretical analysis of constrained optimization by ruling out certain pathological multiplier sequences. Later, Bertsekas and Ozdaglar [11] proposed the related concept of pseudonormality, which admits a geometric interpretation and is strictly stronger than quasinormality.

More recently, quasinormality has also been recognized as a key condition in the convergence analysis of primal-dual algorithms. In particular, it has been shown that, under quasinormality, the sequences of approximate Lagrange multipliers generated by augmented Lagrangian methods remain bounded, a crucial stability property both for the practical performance of these algorithms and for their complexity analysis.

Motivated by the fact that classical quasinormality may fail in the presence of equality constraints written in redundant or degenerate forms, Andreani, Haeser, Schuverdt, and Secchin introduced in [9] a relaxed version of quasinormality, called relaxed quasinormality (RQN). This condition is designed to treat equality constraints in a more flexible manner, inspired by the notion of informative Lagrange multipliers, while preserving the main stability properties of quasinormality. In particular, they proved that RQN is sufficient to guarantee the boundedness of dual sequences generated by safeguarded augmented Lagrangian methods.

These notions have since been explored and generalized in several directions. In particular, Guo, Ye, and Zhang [19] extended quasinormality and pseudonormality to mathematical programs with complementarity constraints (MPCCs) and to problems with geometric constraints in Banach spaces, emphasizing their relevance to sensitivity analysis, exact penalization, and refined stationarity concepts. Their work establishes a connection between quasinormality-type conditions and the existence of enhanced KKT (eKKT) multipliers, which not only satisfy standard stationarity conditions but also allow for the computation of directional derivatives of the value function; see also Ye, Zhang, and Zhang [34]. These enhanced multipliers provide deeper insight into problem sensitivity and lead to sharper optimality conditions in generalized frameworks.

Inspired by these developments, this work investigates the extension of quasinormality-type conditions to the semidefinite setting. Our main goal is to identify regularity conditions that ensure the boundedness of dual sequences generated by augmented Lagrangian methods, even in the absence of classical assumptions such as Robinson's condition.

**Main Contributions.**

The main contributions of this work are summarized as follows:

- We introduce a spectral notion of quasinormality for nonlinear semidefinite programming (NSDP) that extends the classical nonlinear programming concept in a way that is intrinsic to the semidefinite structure and aligned with the eigenspaces of the constraints and their associated multipliers.

- We compare our notion of quasinormality with the componentwise quasinormality proposed by Guo et al. [19], providing strong evidence that the two notions capture different regularity mechanisms. In particular, we show that the componentwise condition may fail even in very simple (diagonal) NSDPs, whereas the proposed spectral formulation remains satisfied and offers a more faithful geometric interpretation in the semidefinite setting.

- We show that quasinormality guarantees the existence of exact penalty functions and local error bounds for NSDP.
- We establish that, under quasinormality, dual sequences generated by augmented Lagrangian methods, including a scaled stopping-criterion variant, are bounded, and that every accumulation point satisfies the Karush–Kuhn–Tucker conditions.
- We introduce a relaxed version of quasinormality for NSDP and show that boundedness of dual sequences and convergence properties of augmented Lagrangian methods can still be obtained under this weaker assumption.
- We also discuss pseudonormality conditions in the NSDP setting, and analyze their relationship with the proposed quasinormality condition.

The paper is organized as follows. In Section 2, we present the notation and preliminaries required for our analysis and review classical and recent constraint qualifications for NSDP. Section 3 introduces quasinormality-type conditions and investigates their properties. In Section 4, we establish boundedness of the dual sequence generated by augmented Lagrangian methods and propose a scaled variant of the method, showing that boundedness is preserved. Section 5 introduces a relaxed notion of quasinormality for NSDP and investigates its implications for boundedness of dual sequences and convergence of augmented Lagrangian methods. Section 6 discusses theoretical applications of quasinormality, including exact penalization and local error bound results for NSDP. Finally, Section 7 provides concluding remarks and outlines directions for future research.

# 2 Preliminaries

## 2.1 Notations

Let $\mathbb{R}^n$ denote the $n$-dimensional Euclidean space equipped with the standard inner product $\langle\cdot,\cdot\rangle$ and the associated Euclidean norm $\|\cdot\|$. Let $\mathbb{S}^m$ be the space of real symmetric $m\times m$ matrices, endowed with the Frobenius inner product

$$\langle A,B\rangle := \operatorname{tr}(AB), \quad A,B\in\mathbb{S}^m,$$

and the corresponding Frobenius norm $\|A\| := \sqrt{\langle A,A\rangle}$. We denote by $\mathbb{S}^m_+$ (resp. $\mathbb{S}^m_-$) the cone of positive semidefinite (resp. negative semidefinite) matrices in $\mathbb{S}^m$. For $A,B\in\mathbb{S}^m$, we write $A\succeq B$ or $B\preceq A$ to indicate that $A-B\in\mathbb{S}^m_+$ and $A\succ B$ or $B\prec A$ when $A-B$ is positive definite. Given a matrix $A\in\mathbb{S}^m$ and an orthogonal matrix $U$ that diagonalizes $A$, we denote by $\lambda^U_1(A),\dots,\lambda^U_m(A)$ the diagonal entries of $U^\top AU$ in the order that they appear. That is, $U^\top AU = \operatorname{diag}(\lambda^U_1(A),\dots,\lambda^U_m(A))$. The superscript $U$ indicates the ordering induced by this diagonalization. When the superscript is omitted, the eigenvalues $\lambda_1(A),\dots,\lambda_m(A)$ are assumed to be listed in nondecreasing order. In this setting, the projection of $A$ onto the cone $\mathbb{S}^m_+$ can be computed as:

$$[A]_+ := U\operatorname{diag}\big(\max\{\lambda^U_1(A),0\},\dots,\max\{\lambda^U_m(A),0\}\big)\,U^\top.$$

The projection $[\cdot]_+$ is nonexpansive with respect to the Frobenius norm, that is,

$$\big\|[A]_+-[B]_+\big\| \le \|A-B\| \quad \text{for all } A,B\in\mathbb{S}^m.$$

Let $T\colon\mathbb{R}^n\to\mathbb{S}^m$ be a smooth mapping. The *derivative* of $T$ at a point $x\in\mathbb{R}^n$ is the linear operator $DT(x)\colon\mathbb{R}^n\to\mathbb{S}^m$ defined by

$$DT(x)h := \sum_{i=1}^n T_i(x)h_i, \quad h\in\mathbb{R}^n,$$

where $T_i(x) := \frac{\partial T(x)}{\partial x_i} \in \mathbb{S}^m$ denotes the matrix of partial derivatives of $T(x)$ with respect to the $i$-th component of $x$. The *adjoint operator* of $DT(x)$, denoted $DT(x)^* \colon \mathbb{S}^m \to \mathbb{R}^n$, is given by

$$DT(x)^*\Xi := \big(\langle T_1(x), \Xi\rangle, \ldots, \langle T_n(x), \Xi\rangle\big)^T, \quad \Xi \in \mathbb{S}^m.$$

Finally, for a function $f \colon \mathbb{R}^n \to \mathbb{R}$, we denote by $\nabla f(x)$ its gradient at $x$. Moreover, given a cone $K \subset \mathbb{R}^n$, its polar cone is defined by

$$K^\circ := \{w \in \mathbb{R}^n \mid \langle w, d\rangle \le 0 \text{ for all } d \in K\}.$$

## 2.2 Constraint qualifications for nonlinear semidefinite programming

In this work, we consider the following nonlinear semidefinite programming problem:

$$\begin{aligned} \min_{x\in\mathbb{R}^n} \quad & f(x), \\ \text{s.t.} \quad & H(x) = 0, \\ & G(x) \preceq 0, \end{aligned} \tag{NSDP}$$

where $f \colon \mathbb{R}^n \to \mathbb{R}$, $H \colon \mathbb{R}^n \to \mathbb{S}^\ell$, and $G \colon \mathbb{R}^n \to \mathbb{S}^m$ are continuously differentiable functions. We denote by $\mathcal{F}$ the feasible set of problem (NSDP). In our formulation, matrix equality constraints are explicitly included. There are several motivations for adopting this setting. First, in the NSDP literature it is common to reformulate the negative semidefinite constraint $G(x) \preceq 0$ as an equality by introducing symmetric slack variables, as in [19, 23]. Moreover, matrix equality constraints naturally arise in convergence analyses of interior-point methods for conic constraints, as discussed in [7]. These examples indicate that including constraints of the form $H(x) = 0$, with $H(x) \in \mathbb{S}^\ell$, not only generalizes the modeling framework but is also consistent with existing theoretical and numerical approaches in semidefinite optimization.

The Lagrangian function associated with problem (NSDP) is defined as

$$\mathcal{L} \colon \mathbb{R}^n \times \mathbb{S}^\ell \times \mathbb{S}^m \to \mathbb{R}, \quad \mathcal{L}(x, \Lambda, \Omega) := f(x) + \langle H(x), \Lambda\rangle + \langle G(x), \Omega\rangle.$$

The Karush–Kuhn–Tucker (KKT) conditions for NSDP are given as follows. A point $\bar{x} \in \mathbb{R}^n$ is called a *KKT point* if there exist multipliers $\Lambda \in \mathbb{S}^\ell$ and $\Omega \in \mathbb{S}^m_+$ such that

$$\nabla f(\bar{x}) + DH(\bar{x})^*\Lambda + DG(\bar{x})^*\Omega = 0, \tag{1}$$
$$H(\bar{x}) = 0, \tag{2}$$
$$G(\bar{x}) \preceq 0, \tag{3}$$
$$\langle G(\bar{x}), \Omega\rangle = 0. \tag{4}$$

Condition (4) is known as the *complementarity condition*. The following lemma shows that complementarity can be expressed in several equivalent forms.

**Lemma 1** ([10, Lemma 2.2]). *Let $X \in \mathbb{S}^m_-$ and $Y \in \mathbb{S}^m_+$. The following statements are equivalent:*

*(a)* $\langle X, Y\rangle = 0$;

*(b)* $XY = 0$;

(c) *There exists an orthogonal matrix* $U$ *such that*

$$U^\top X U = \operatorname{diag}\big(\lambda_1^U(X), \ldots, \lambda_m^U(X)\big) \quad \text{and} \quad U^\top Y U = \operatorname{diag}\big(\lambda_1^U(Y), \ldots, \lambda_m^U(Y)\big),$$

*and*

$$\lambda_i^U(X)\,\lambda_i^U(Y) = 0, \qquad i = 1, \ldots, m.$$

*That is,* $X$ *and* $Y$ *are simultaneously diagonalizable by* $U$*, and their eigenvalues are complementary according to the ordering induced by* $U$*.*

When $\langle X, Y\rangle = 0$ with $X \in \mathbb{S}^m_-$ and $Y \in \mathbb{S}^m_+$, the matrix $U$ in Lemma 1(c) may be taken such that the eigenvalues are ordered, that is, $\lambda_i^U(X) = \lambda_i(X)$ and $\lambda_i^U(Y) = \lambda_i(Y)$ for all $i$. To guarantee that the KKT conditions hold at a local minimizer of (NSDP), appropriate *constraint qualification* (CQ) conditions are required. Such conditions also play a central role in the global convergence analysis of iterative algorithms, as they ensure that every accumulation point of a sequence generated by the method satisfies the KKT conditions. We next recall the most commonly used and computationally relevant CQs for nonlinear semidefinite programming.

**Definition 1.** Let $\bar{x} \in \mathbb{R}^n$ be a feasible point. We say that $\bar{x}$ satisfies the *nondegeneracy condition* if the following holds. Let $r = \operatorname{rank} G(\bar{x})$, and let $\{e_1, \ldots, e_{m-r}\} \subset \mathbb{R}^m$ be an orthonormal basis for the null space of $G(\bar{x})$. Let $\{s_1, \ldots, s_\ell\} \subset \mathbb{R}^\ell$ be an orthonormal basis of $\mathbb{R}^\ell$. Then the set of vectors

$$\Big\{\big(e_i^\top G_1(\bar{x})e_j, \ldots, e_i^\top G_n(\bar{x})e_j\big)\Big\}_{i,j=1}^{m-r} \bigcup \Big\{\big(s_i^\top H_1(\bar{x})s_j, \ldots, s_i^\top H_n(\bar{x})s_j\big)\Big\}_{i,j=1}^{\ell}$$

is linearly independent in $\mathbb{R}^n$.

**Definition 2.** A feasible point $\bar{x} \in \mathbb{R}^n$ is said to satisfy *Robinson's constraint qualification* if the operator $DH(\bar{x})\colon \mathbb{R}^n \to \mathbb{S}^\ell$ is surjective and there exists a direction $d \in \mathbb{R}^n$ such that

(i) $DH(\bar{x})d = 0$;

(ii) $G(\bar{x}) + DG(\bar{x})d \prec 0$.

Analogously to the classical NLP setting, Robinson's condition is equivalent to the nonemptiness and boundedness of the set of Lagrange multipliers at a local minimizer. In contrast, the nondegeneracy condition, also known as the *transversality condition*, ensures the uniqueness of the multipliers. Thus, it plays a role similar to the linear independence constraint qualification (LICQ) in standard nonlinear programming, although the analogy is not exact. For instance, when $G(x)$ is diagonal, nondegeneracy does not reduce to LICQ. Lourenço, Fukuda, and Fukushima [23] showed that if the NSDP is reformulated as an NLP by introducing slack variables, then the nondegeneracy of $x$ for the original NSDP is equivalent to LICQ for the reformulated problem.

Recent works have introduced new CQ conditions for NSDP inspired by constant-rank-type (CRCQ) assumptions from NLP [4]. These include weak-nondegeneracy, weak-Robinson, and weak-CRCQ conditions, which capture sparsity patterns in eigenvectors and provide flexible frameworks for ensuring algorithmic convergence. Variants such as weak-nondegeneracy and weak-CRCQ explore matrix sparsity, offering more relaxed alternatives to classical assumptions. For diagonal problems, weak-nondegeneracy condition is equivalent to LICQ.

Moreover, Andreani et al. first introduced in [6] a *facial constant rank constraint qualification* for NSDP. Subsequently, in [3], the authors proposed a *minimal face constant rank* CQ for reducible conic programming. This latter approach provides a geometric interpretation based on facial reduction, leading to stronger second-order necessary optimality conditions, even in situations where

Robinson's condition fails. The use of facial reduction allows for a local reformulation of the problem in terms of appropriate cone faces, thereby enhancing robustness in NSDP formulations. In this framework, the *facial-CRCQ* condition ensures a second-order necessary optimality condition stronger than the one obtained under Robinson's condition, while the *constant rank of the subspace component* (CRSC) and *strong-CRSC* conditions [3] characterize unique constant-rank subsets, further enriching the theoretical landscape of NSDP and related conic optimization problems.

Building on these recent developments, the present paper extends the notion of *quasinormality* to NSDP. This extension provides new insights into the structure of feasible and optimal points and aims to reinforce both optimality conditions and convergence analyses, bridging classical and emerging theories in nonlinear conic programming.

# 3 Quasinormality and pseudonormality conditions for NSDP

In NLP, constraint qualifications play a fundamental role in ensuring the existence of Lagrange multipliers, the validity of optimality conditions, and the convergence of numerical methods. Among the various conditions proposed in the literature, the *quasinormality condition* (QN) stands out as a relevant and relatively weak qualification that has found applications in exact penalization, sensitivity analysis, and algorithmic stability. Originally introduced by Hestenes [20], quasinormality was later explored and extended in several directions [11, 34], particularly in settings where stronger assumptions such as MFCQ or LICQ fail. In this work, we extend Hestenes' definition to the setting of NSDP by incorporating the spectral structure of the semidefinite constraints into the notion of constraint activity.

**Definition 3.** A feasible point $\bar{x}$ of problem (NSDP) is said to satisfy the *quasinormality condition* if there does not exist a pair $(\Lambda, \Omega) \in \mathbb{S}^\ell \times \mathbb{S}^m_+$, with $(\Lambda, \Omega) \neq 0$ and a sequence $\{x^k\} \to \bar{x}$, such that

$$DH(\bar{x})^*\Lambda + DG(\bar{x})^*\Omega = 0, \tag{5}$$

and

(i) $\lambda_i^{V_k}(H(x^k))\lambda_i^V(\Lambda) > 0$ for all $i$ such that $\lambda_i^V(\Lambda) \neq 0$,

(ii) $\lambda_i^{U_k}(G(x^k)) > 0$ for all $i$ such that $\lambda_i^U(\Omega) > 0$,

where $V_k$, $U_k$, $V$ and $U$ are orthogonal matrices that diagonalize $H(x^k)$, $G(x^k)$, $\Lambda$ and $\Omega$, respectively, such that $U_k \to U$ and $V_k \to V$.

We will show in Section 4 that the quasinormality condition is indeed a constraint qualification for NSDP.

*Remark* 1. We will prove later that Definition 3 reduces to the classical notion of quasinormality in NLP when $G$ and $H$ are diagonal mappings. In this case, $G$, $H$, as well as the multipliers $\Omega$ and $\Lambda$, are all diagonal matrices, and conditions (i)–(ii) reduce to componentwise inequalities. More precisely, for indices $i$ such that $\Omega_{ii} > 0$, one has $(G(x^k))_{ii} > 0$, while for indices $i$ such that $\Lambda_{ii} \neq 0$, one has $(H(x^k))_{ii}\Lambda_{ii} > 0$. These sign conditions coincide with those appearing in the classical NLP definition of quasinormality.

The requirements $U_k \to U$ and $V_k \to V$ enforce a spectral pairing between $G(x^k)$ and $\Omega$, and between $H(x^k)$ and $\Lambda$, ensuring that the sign control is checked along eigendirections that are consistent in the limit.

*Remark* 2. The multipliers whose existence is ruled out by Definition 3 necessarily satisfy the complementarity relation with the limiting constraints. Indeed, since $G(x^k) \to G(\bar{x})$ and $U_k \to U$, we have that $G(\bar{x})$ and $\Omega$ are simultaneously diagonalizable by $U$. For any index with $\lambda_i^U(\Omega) > 0$, condition (ii) gives $\lambda_i^{U_k}\big(G(x^k)\big) > 0$ for large $k$, hence by continuity $\lambda_i^U\big(G(\bar{x})\big) \geq 0$. Feasibility of $\bar{x}$ implies $\lambda_i^U\big(G(\bar{x})\big) = 0$. Thus, by Lemma 1, we have $\langle G(\bar{x}), \Omega \rangle = 0$.

An earlier extension of quasinormality was proposed by Guo, Ye, and Zhang [19] within the broader framework of Mathematical Programming with Geometric Constraints (MPGC). Their approach is formulated in a fixed coordinate system and does not act directly on the underlying geometric constraint. This is a natural and powerful idea in the general MPGC setting, where the constraints do not necessarily exhibit a specific algebraic or geometric structure, thus allowing the analysis to proceed in a manner similar to NLP. However, when one restricts attention to NSDP, which constitutes a particular class of MPGCs, this formulation does not fully exploit the rich spectral structure inherent to semidefinite constraints. Also, their condition includes a slack variable that reduces significantly its applicability, making it fail in very simple examples, as we will show next.

More specifically, when the MPGC-based definition is specialized to NSDP, a slack variable $Y \in \mathbb{S}^m_-$ is introduced so that $G(x) - Y = 0$. The resulting conditions are then written componentwise using a fixed basis $\{E_i\}_{i=1}^d$ of $\mathbb{S}^m$ and $\{\tilde{E}_i\}_{i=1}^{\tilde{d}}$ of $\mathbb{S}^\ell$, where $d = \dfrac{m(m+1)}{2}$ and $\tilde{d} = \dfrac{\ell(\ell+1)}{2}$.

**Definition 4** ([19])**.** Let $\{E_i\}_{i=1}^d$ be a basis of $\mathbb{S}^m$ and $\{\tilde{E}_i\}_{i=1}^{\tilde{d}}$ of $\mathbb{S}^\ell$. A feasible point $\bar{x}$ of problem (NSDP) is said to satisfy the *componentwise quasinormality* if there does not exist $(\Lambda, \Omega) \in \mathbb{S}^\ell \times \mathbb{S}^m_+$, with $(\Lambda, \Omega) \neq 0$, and a sequence $\{(x^k, Y_k, \Omega_k)\} \subset \mathbb{R}^n \times \mathbb{S}^m_- \times \mathbb{S}^m_+$ converging to $(\bar{x}, G(\bar{x}), \Omega)$, such that (5) holds together with the following conditions:

1. $\langle \Omega_k, Y_k \rangle = 0$;
2. For every $i$ such that $\langle \Lambda, \tilde{E}_i \rangle \neq 0$, it holds that
$$\langle \Lambda, \tilde{E}_i \rangle \cdot \langle H(x^k), \tilde{E}_i \rangle > 0, \quad \text{for all } k.$$
3. For every $i$ such that $\langle \Omega, E_i \rangle \neq 0$, it holds that
$$\langle \Omega, E_i \rangle \cdot \langle G(x^k) - Y_k, E_i \rangle > 0, \quad \text{for all } k.$$

In contrast, the quasinormality condition introduced in Definition 3 differs structurally, as it is formulated in a spectral basis dynamically adapted to the problem data. In our approach, the directions along which the sign control is evaluated vary with $x^k$ and are determined by the eigenvectors of $G(x^k)$ and by the multipliers $\Omega$. This allows complementarity to be analyzed within eigenspaces that accurately reflect the local geometry of the matrix constraint. Consequently, even if one were to replace the fixed basis used in the framework of Guo et al. with a spectral basis, the lack of alignment between the basis directions and the multipliers would prevent their condition from recovering ours. Thus, the formulation proposed in this work provides a more natural and geometrically meaningful extension of quasinormality to the NSDP setting, as it aligns the notion of constraint activity with the spectral structure of $G(x)$ and with the directions defined by $\Omega$.

The following example illustrates that the componentwise quasinormality condition may fail even in a very simple setting, whereas the proposed spectral quasinormality condition remains satisfied. In particular, the example is fully diagonal and involves no structural complexity beyond that of a basic NSDP formulation. This highlights a potential limitation of the componentwise definition, showing that it may exclude benign situations where no genuine pathological behavior is present.

**Example 1.** Consider the NSDP constraint

$$G(x) = \begin{pmatrix} -x^2 & 0 \\ 0 & 0 \end{pmatrix} \preceq 0, \qquad x \in \mathbb{R},$$

and the feasible point $\bar{x} = 0$. Fix the Frobenius–orthonormal basis of $\mathbb{S}^2$ given by

$$E_1 = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \quad E_2 = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad E_3 = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}.$$

(i) *Quasinormality holds at* $\bar{x}$. Since $G(x) \preceq 0$ for all $x$, there is no sequence $x^k \to \bar{x}$ along which $G(x^k)$ has a positive eigenvalue in any direction associated with a positive eigenvalue of $\Omega \succeq 0$. Hence the violating configuration in our definition cannot occur, and our condition holds at $\bar{x}$.

(ii) *Componentwise quasinormality fails at* $\bar{x}$. Let $\Omega = E_1 = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \succeq 0$, $\Omega \neq 0$. Then, $DG(\bar{x})^* \Omega = 0$ and $\langle G(\bar{x}), \Omega \rangle = 0$. Define $x^k = 1/k^2$, $t_k = 1/k$,

$$\Omega_k = \begin{pmatrix} 1 & t_k \\ t_k & t_k^2 \end{pmatrix} \succeq 0 \to \Omega, \qquad Y_k := -\frac{1}{k} \begin{pmatrix} t_k^2 & -t_k \\ -t_k & 1 \end{pmatrix} \preceq 0 \to 0 = G(\bar{x}).$$

Then

$$\langle \Omega_k, Y_k \rangle = 1 \cdot \Big( -\frac{t_k^2}{k} \Big) + t_k \cdot \Big( \frac{t_k}{k} \Big) + t_k \cdot \Big( \frac{t_k}{k} \Big) + t_k^2 \cdot \Big( -\frac{1}{k} \Big) = 0.$$

Moreover, since $\langle \Omega, E_1 \rangle = 1$ is the only nonzero coordinate of $\Omega$ in the fixed basis, the sign condition reduces to

$$\langle G(x^k) - Y_k, E_1 \rangle = (G(x^k) - Y_k)_{11} = -\frac{1}{k^4} + \frac{1}{k^3} > 0$$

for all large $k$. Therefore, the nonexistence statement in the definition of componentwise quasinormality is violated, and therefore it does not hold at $\bar{x}$.

Although we believe that the proposed spectral quasinormality condition and the componentwise quasinormality condition are independent, we were not able to provide the converse counterexample. This is mainly due to the fact that it is hard to find an example such that componentwise quasinormality holds. We recall that the componentwise formulation relies on auxiliary slack variables and enforces sign conditions in a fixed basis chosen a priori, which may introduce artificial restrictions unrelated to the intrinsic geometry of the semidefinite constraint. In contrast, in our approach the directions along which the sign control is evaluated vary with $x^k$ and are determined by the eigenvectors of $G(x^k)$ and by the associated multipliers $\Omega$.

Note that the previous example does not satisfy Robinson's condition at $\bar{x} = 0$. Indeed, $G(\bar{x}) + DG(\bar{x})[d] = 0$ for all $d$. Moreover, Robinson's condition can be equivalently expressed in a dual form. More precisely, following the characterization given in [14, Prop. 2.97], Robinson's condition holds at a feasible point $\bar{x}$ if and only if the implication

$$DH(\bar{x})^* \Lambda + DG(\bar{x})^* \Omega = 0, \quad \langle G(\bar{x}), \Omega \rangle = 0, \quad \Omega \succeq 0 \implies (\Lambda, \Omega) = (0, 0), \tag{6}$$

is satisfied. In this form, it becomes apparent that Robinson's condition implies quasinormality. Furthermore, the previous example shows that this implication is strict.

Before proceeding further, it is instructive to clarify the relationship between the proposed quasinormality condition for NSDP and the classical notion of quasinormality in NLP. Since diagonal semidefinite constraints naturally reduce to a collection of scalar inequalities, this setting provides a convenient framework in which the spectral formulation introduced here can be directly compared with its NLP counterpart. The next result shows that, in the diagonal case, our spectral quasinormality condition is equivalent to the classical quasinormality condition for NLP.

*Remark* 3. We do not claim that treating an NLP as a diagonal NSDP would be a reasonable approach to deal with an NLP. Rather, more conveniently would be to treat an NLP as multiple one-dimensional semidefinite constraints. Our claim here is that our notion of quasinormality in NSDP is reasonable enough in order to recover the NLP definition in the diagonal setting. This is not the case of the nondegeneracy condition, for instance, which motivates alternative definitions. See [4].

**Theorem 1.** *Let NSDP be a diagonal nonlinear semidefinite program, that is, assume that $H(x)$ and $G(x)$ are diagonal matrices for all $x$. Then the quasinormality condition for NSDP is equivalent to the classical quasinormality condition for NLP.*

*Proof.* We present the proof only for the semidefinite constraint, since the equality constraints can be handled analogously. Let

$$G(x) := \operatorname{diag}(g_1(x), \ldots, g_m(x))$$

be the semidefinite constraint function. Both implications are proved by contraposition. We first show that quasinormality for NSDP implies quasinormality for the NLP problem with constraints $g_1(x) \leq 0, \ldots, g_m(x) \leq 0$. Assume that the quasinormality condition for NLP fails at $\bar{x}$. Then there exist a nonzero multiplier $\mu \in \mathbb{R}^m$, $\mu \geq 0$, and a sequence $\{x^k\}$ with $x^k \to \bar{x}$ such that the stationarity condition $DG(\bar{x})^*\Omega = 0$ holds, where $0 \neq \Omega := \operatorname{diag}(\mu) \in \mathbb{S}^m_+$, and the sign requirements are met:

$$\mu_i > 0 \;\Rightarrow\; g_i(x^k) > 0 \quad \text{for all } i = 1, \ldots, m.$$

Choosing $U_k = U = I$, this implication can be written equivalently as

$$\lambda_i^U(\Omega) > 0 \;\Rightarrow\; \lambda_i^{U_k}(G(x^k)) > 0 \quad \text{for all } i = 1, \ldots, m,$$

which violates the NSDP definition. We now prove the converse implication. Assume that the NSDP quasinormality condition fails at $\bar{x}$. Then there exist $\Omega \in \mathbb{S}^m_+$, $\Omega \neq 0$, a sequence $\{x^k\}$ with $x^k \to \bar{x}$, and orthogonal matrices $U_k$ and $U$ diagonalizing $G(x^k)$ and $\Omega$, respectively, for all $k$, where $U_k \to U$ and $DG(\bar{x})^*\Omega = 0$, together with

$$\lambda_i^U(\Omega) > 0 \implies \lambda_i^{U_k}(G(x^k)) > 0 \quad \text{for all } i = 1, \ldots, m. \tag{7}$$

By Remark 2, $\langle G(\bar{x}), \Omega \rangle = 0$ can be concluded and the diagonal structure of $G(\bar{x})$ implies $g_i(\bar{x})\Omega_{ii} = 0$ for all $i$. By applying a similarity transformation $P(\cdot)P^T$ to $\Omega$ and $G$, where $P$ is a permutation matrix, let us assume that the diagonal entries of $\Omega$ appear in non-decreasing order. Since $\Omega \in \mathbb{S}^m_+$, the corresponding rows and columns of a zero diagonal entry are also zero. Thus $\Omega$ has a $2 \times 2$ block structure where only the southeast block is non-zero (with positive diagonal) while the corresponding block of $G(\bar{x})$ is equal to zero due to complementarity. Focusing only on this block, we assume without loss of generality that all diagonal entries of $\Omega$ are positive and $G(\bar{x}) = 0$. Since the stationarity condition $DG(\bar{x})^*\Omega = 0$ depends only on the diagonal elements of $\Omega$, the diagonal matrix $\tilde{\Omega} = \operatorname{diag}(\Omega_{11}, \ldots, \Omega_{mm})$ together with the sequence $\{x^k\}$ will attest that quasinormality for NLP fails as long as $g_i(x^k) > 0$ for all $i$ and all $k$. Equivalently, let us prove that $\lambda_i^{U_k}(G(x^k)) > 0$ for all $i$ and all $k$.

Assume that this is not the case, that is, for each $k$, there is an index $i_k$ such that $\lambda_k := \lambda_{i_k}^{U_k}(G(x^k)) \leq 0$. For all $k$, let $I_k$ be the non-empty subset of $\{1, \ldots, m\}$ consisting of all indexes $i$ such that $\lambda_i^{U_k}(G(x^k)) = \lambda_k$. By the infinite pigeonhole principle, let us take a subsequence such that $I_k$ is constant, that is, $I_k = \{i_1, \ldots, i_p\} \subseteq \{1, \ldots, m\}$ for some non-empty set $\{i_1, \ldots, i_p\}$. But for all $k$ in the subsequence, $G(x^k)$ is diagonal and the eigenvalues $\lambda_i^{U_k}(G(x^k)) = \lambda_k$ for $i = i_1, \ldots, i_p$

appear in the diagonal of $G(x^k)$ at some correspondent positions $g_j(x^k) = \lambda_k, j = j_1, \ldots, j_p$, with $g_j(x^k) \neq \lambda_k$ for $j \neq j_1, \ldots, j_p$. Thus, the only possibility is that the correspondent eigenvectors $u^k_{i_1}, \ldots, u^k_{i_p}$ (columns $i_1, \ldots, i_p$ of $U^k$) of $G(x^k)$ span the eigenspace associated with $\lambda_k$, which clearly coincides with the space generated by the canonical vectors $e_{j_1}, \ldots, e_{j_p}$, given that $G(x^k)$ is diagonal.

By (7), since $\lambda_i^{U_k}(G(x^k)) \leq 0$ for all $i \in \{i_1, \ldots, i_p\}$ and all $k$ in the subsequence, we must have $\lambda_i^U(\Omega) = 0$ for all $i \in \{i_1, \ldots, i_p\}$. By the definition of quasinormality for NSDP we have that $U_k \to U$, that is, $u^k_{i_1} \to u_{i_1}, \ldots, u^k_{i_p} \to u_{i_p}$ where the limits are eigenvectors of $\Omega$ associated with the zero eigenvalue. Thus the space generated by $u_{i_1}, \ldots, u_{i_p}$ is contained in the eigenspace of $\Omega$ associated with the zero eigenvalue. Since the span of $e_{j_1}, \ldots, e_{j_p}$ is a closed set, we conclude that $e_{j_1}, \ldots, e_{j_p}$ must be eigenvectors of $\Omega$ associated with the zero eigenvalue. Therefore, since $\{j_1, \ldots, j_p\}$ is non-empty, for some canonical vector, say, $e_{j_1}$, we have $\Omega e_{j_1} = 0$ and in particular $\Omega_{j_1 j_1} = 0$, which is a contradiction.

Note that we did not use the full convergence $U^k \to U$, but rather only the limit of the columns associated with the zero eigenvalue of $\Omega$ (similar to Remark 2, one still has $\langle G(\bar{x}), \Omega \rangle = 0$ under this weaker requirement). Thus quasinormality for NLP implies this stronger variation of quasinormality for NSDP in the diagonal setting. □

In contrast to the approach adopted here, where the notion of quasinormality proposed by Guo et al. [19] is directly specialized to the NSDP setting, the authors in [19] follow a different strategy to handle semidefinite constraints. Rather than working directly with the semidefinite constraint $G(x) \preceq 0$, they reformulate the NSDP by replacing this constraint with the scalar inequality $\lambda_m(G(x)) \leq 0$, where $\lambda_m(G(x))$ is the largest eigenvalue of $G(x)$. This reformulation allows the problem to be treated within the framework of nonlinear programming, since the function $\lambda_m(\cdot)$ is locally Lipschitz continuous and directionally differentiable, thereby enabling the application of standard NLP tools. The formal definition is given below.

**Definition 5** (Corollary 3.6 of [19])**.** A feasible point $\bar{x}$ of problem (NSDP), with the equality constraints omitted, is said to satisfy the *maximum-eigenvalue quasinormality* condition if there does not exist $\Omega \in \mathbb{S}^m_+$, with $\Omega \neq 0$, such that

$$DG(\bar{x})^*\Omega = 0, \qquad \langle G(\bar{x}), \Omega \rangle = 0, \tag{8}$$

and a sequence $\{x^k\} \to \bar{x}$ satisfying $\lambda_m(G(x^k)) > 0$.

Although this reformulation simplifies the analysis and allows the direct application of classical NLP techniques, it focuses on a single spectral direction, namely that associated with the dominant eigenvalue. As a result, it does not explicitly expose the full range of spectral activity directions that may be relevant in the NSDP framework.

We now show that this approach is more naturally related to a pseudonormality condition, as introduced by Bertsekas and Ozdaglar [11] for NLP, which we extend next to NSDP.

**Definition 6.** A feasible point $\bar{x}$ of problem (NSDP) is said to satisfy the *pseudonormality condition* if there do not exist a pair $(\Lambda, \Omega) \in \mathbb{S}^\ell \times \mathbb{S}^m_+$, with $(\Lambda, \Omega) \neq (0, 0)$, and a sequence $\{x^k\} \to \bar{x}$ such that (5) holds and,

$$\sum_{i=1}^{\ell} \lambda_i^V(\Lambda)\, \lambda_i^{V_k}\big(H(x^k)\big) \;+\; \sum_{j=1}^{m} \lambda_j^U(\Omega)\, \lambda_j^{U_k}\big(G(x^k)\big) \;>\; 0 \quad \text{for all } k, \tag{9}$$

where $V_k$, $U_k$, $V$ and $U$ are orthogonal matrices that diagonalize $H(x^k)$, $G(x^k)$, $\Lambda$ and $\Omega$, respectively, such that $U_k \to U$ and $V_k \to V$.

*Remark* 4. Note that the pseudonormality condition implies the classical complementarity relation between $\Omega$ and $G(\bar{x})$, namely, $\langle G(\bar{x}), \Omega\rangle = 0$. Indeed, let $\bar{x}$ be feasible, so that $H(\bar{x}) = 0$ and $G(\bar{x}) \preceq 0$, and assume that there exist $(\Lambda, \Omega) \in \mathbb{S}^\ell \times \mathbb{S}^m_+$, $(\Lambda, \Omega) \neq (0, 0)$, a sequence $x^k \to \bar{x}$, and orthogonal matrices $V_k, U_k, V$ and $U$, diagonalizing $H(x^k)$, $G(x^k)$, $\Lambda$ and $\Omega$, respectively, with $V_k \to V$ and $U_k \to U$ such that the sign condition (9) holds for all $k$. Taking the limit in (9) we arrive at

$$\sum_{j=1}^{m} \lambda_j^U(\Omega)\lambda_j^U(G(\bar{x})) = \sum_{i=1}^{\ell} \lambda_i^V(\Lambda)\lambda_i^V(H(\bar{x})) + \sum_{j=1}^{m} \lambda_j^U(\Omega)\lambda_j^U(G(\bar{x})) \geq 0,$$

which implies $\lambda_j^U(\Omega)\lambda_j^U(G(\bar{x})) = 0$ for all $j = 1, \ldots, m$ given that $\Omega \succeq 0$ and $G(\bar{x}) \preceq 0$. We conclude that $\langle G(\bar{x}), \Omega\rangle = 0$ due to Lemma 1.

We now investigate the relationship between maximum-eigenvalue quasinormality and pseudonormality.

**Theorem 2.** *Let $\bar{x}$ be a feasible point of problem* (NSDP)*, with the equality constraints omitted. Then, the maximum-eigenvalue quasinormality condition implies the pseudonormality condition.*

*Proof.* Suppose, by contradiction, that $\bar{x}$ does not satisfy the pseudonormality condition. Then there exist $\Omega \succeq 0$, $\Omega \neq 0$, and a sequence $x^k \to \bar{x}$ such that

$$DG(\bar{x})^*\Omega = 0 \quad \text{and} \quad \sum_{j=1}^{m} \lambda_j^U(\Omega)\, \lambda_j^{U_k}\big(G(x^k)\big) > 0 \quad \text{for all } k,$$

where $U_k$ and $U$ diagonalize $G(x^k)$ and $\Omega$, respectively, with $U_k \to U$. By Remark 4, the sign condition implies the classical complementarity relation $\langle G(\bar{x}), \Omega\rangle = 0$. The inequality above implies that for each $k$ there exists $j_k$ such that $\lambda_{j_k}^{U_k}(G(x^k)) > 0$. In particular, $G(x^k)$ has a strictly positive eigenvalue for all $k$, that is, $\lambda_m(G(x^k)) > 0$ for all $k$. Thus, $\bar{x}$ violates the maximum-eigenvalue quasinormality condition, which is a contradiction. □

The next example shows that the relationship between Definitions 5 and 6 is strict.

**Example 2.** Consider the NSDP constraint

$$G(x) = \begin{pmatrix} x_1 & x_2 \\ x_2 & -x_1 \end{pmatrix} \preceq 0, \qquad x \in \mathbb{R}^2,$$

and the feasible point $\bar{x} = 0$.

(i) $\bar{x}$ *satisfies pseudonormality.* Since $0 = DG(\bar{x})^*\Omega = (\Omega_{11} - \Omega_{12}, 2\Omega_{12})$ we must have $\Omega_{12} = 0$ and $\Omega_{11} = \Omega_{22} > 0$ are the eigenvalues of $\Omega$. But for any sequence $\{x^k\}$, the eigenvectors of $G(x^k)$ are $\pm\sqrt{(x_1^k)^2 + (x_2^k)^2}$ hence

$$\lambda_1^U(\Omega)\lambda_1^{U_k}(G(x^k)) + \lambda_2^U(\Omega)\lambda_2^{U_k}(G(x^k)) = 0,$$

which violates (9).

(ii) $\bar{x}$ *does not satisfy maximum-eigenvalue quasinormality.* Take $\Omega = I$ and $x^k = (1/k, 1/k)$. We have $DG(\bar{x})^*\Omega = 0, \langle G(\bar{x}), \Omega\rangle = 0$, and $\lambda_2(G(x^k)) > 0$ for all $k$.

This example shows that pseudonormality is strictly weaker than maximum-eigenvalue quasinormality, which makes it a more general and flexible constraint qualification. Another important feature of our formulation is that complementarity is intrinsically built into the condition itself, whereas in the maximum-eigenvalue framework this property has to be imposed explicitly. We now show that pseudonormality implies our definition of quasinormality.

**Theorem 3.** *Let $\bar{x}$ be a feasible point of (NSDP). If $\bar{x}$ satisfies the pseudonormality condition, then it also satisfies the quasinormality condition.*

*Proof.* The proof is a consequence of the fact that items (i) and (ii) of Definition 3 imply that both sums in (9) consist of non-negative summands. The fact that $(\Lambda, \Omega) \neq (0, 0)$ implies that at least one summand is positive, which concludes the proof. $\square$

The relationship between quasinormality and pseudonormality is strict, as illustrated by the following example:

**Example 3.** Consider the NSDP constraint

$$G(x) = \begin{pmatrix} \frac{x^2}{2} & -\left(x + \frac{x^2}{2}\right) \\ -\left(x + \frac{x^2}{2}\right) & \frac{x^2}{2} \end{pmatrix} \preceq 0, \qquad x \in \mathbb{R},$$

and the feasible point $\bar{x} = 0$.

(i) *Quasinormality holds at $\bar{x}$.* Suppose, by contradiction, that quasinormality fails at $\bar{x}$. Then there exist $\Omega \succeq 0$, $\Omega \neq 0$, and a sequence $x^k \to \bar{x}$ such that

$$DG(\bar{x})^*\Omega = 0 \quad \text{and} \quad \lambda_i^{U_k}\big(G(x^k)\big) > 0 \quad \text{for all } i \text{ with } \lambda_i^U(\Omega) > 0, \tag{10}$$

where $U_k \to U$ and $U_k$ and $U$ diagonalize $G(x^k)$ and $\Omega$, respectively. The eigenvalues of $G(x^k)$ are $-x^k$ and $x^k + (x^k)^2$. Since (10) implies $x^k \neq 0$, the eigenvalues must be distinct for all sufficiently large $k$. This implies that $U_k$ is uniquely determined and a simple computation gives $U_k = U = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}$. Now, by the identity $U^\top \left(\begin{smallmatrix} 0 & -1 \\ -1 & 0 \end{smallmatrix}\right) U = \operatorname{diag}(-1, 1)$ and cyclic trace invariance, we obtain

$$\begin{aligned} 0 = DG(\bar{x})^*\Omega = \left\langle \begin{pmatrix} 0 & -1 \\ -1 & 0 \end{pmatrix}, \Omega \right\rangle &= \left\langle \operatorname{diag}(-1, 1), \operatorname{diag}\big(\lambda_1^U(\Omega), \lambda_2^U(\Omega)\big) \right\rangle \\ &= -\lambda_1^U(\Omega) + \lambda_2^U(\Omega). \end{aligned}$$

Since $\Omega \succeq 0$ and $\Omega \neq 0$, it follows that $\lambda_1^U(\Omega) = \lambda_2^U(\Omega) > 0$. Consequently, the sign control requirement forces

$$-x^k > 0 \quad \text{and} \quad x^k + (x^k)^2 > 0 \qquad \text{for all sufficiently large } k,$$

which is a contradiction. Hence quasinormality holds at $\bar{x}$.

(ii) *Pseudonormality fails at $\bar{x}$.* Let $\Omega = I \succ 0$. Then

$$DG(\bar{x})^*\Omega = \left\langle \begin{pmatrix} 0 & -1 \\ -1 & 0 \end{pmatrix}, I \right\rangle = 0,$$

so (5) holds. Moreover, take any sequence $x^k \to 0$ with $x^k \neq 0$. Since $\Omega = I$, we may choose $U_k = U = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}$ for all sufficiently large $k$, and thus

$$\sum_{j=1}^{2} \lambda_j^U(\Omega)\, \lambda_j^{U_k}\big(G(x^k)\big) = \lambda_1^{U_k}\big(G(x^k)\big) + \lambda_2^{U_k}\big(G(x^k)\big) = (x^k)^2 > 0 \quad \text{for all } k,$$

where we used that the eigenvalues of $G(x^k)$ are $-x^k$ and $x^k + (x^k)^2$. Hence (9) is satisfied for all $k$, providing a violating configuration for pseudonormality at $\bar{x}$. Therefore, pseudonormality does not hold at $\bar{x}$.

A result similar to Theorem 1 can be proved for pseudonormality, which we omit. We summarize the concepts introduced so far with the strict hierarchy below:

$$\text{Maximum-eigenvalue quasinormality} \;\Rightarrow\; \text{Pseudonormality} \;\Rightarrow\; \text{Quasinormality}.$$

In the next section, it will be shown that quasinormality is a constraint qualifications for NSDP, which implies the same is true for the other conditions.

## 4 Augmented Lagrangian Methods for NSDP

The augmented Lagrangian method is a powerful framework for solving constrained optimization problems, including NSDP; see, for instance, our previous work [10] for a detailed analysis in this setting. It simultaneously penalizes violations of both equality and semidefinite constraints by introducing appropriate quadratic terms into the augmented Lagrangian function. From a theoretical perspective, most existing global convergence results for augmented Lagrangian methods in NSDP rely on Robinson's condition, which ensures boundedness of the dual sequence and convergence to a KKT point. However, this assumption is often too restrictive in practice. Robinson's condition imposes a strong regularity requirement that guarantees bounded Lagrange multipliers at the limit point, but it may fail in many realistic applications, as is the case for NSDP problems with complementarity constraints; see [15].

Recent developments, such as the sequential optimality frameworks known as the Approximate KKT (AKKT) and Complementarity AKKT (CAKKT) conditions [2, 8], have shown that the primal sequence generated by augmented Lagrangian-type methods converges to points satisfying sequential stationarity conditions. These results hold under weak assumptions, such as AKKT-regularity and CAKKT-regularity, which are independent of the quasinormality condition considered in this work. However, they only ensure that the limit point satisfies the KKT conditions without identifying a corresponding Lagrange multiplier. In many applications, it is not sufficient to know that a KKT point exists because one also needs to compute the associated multiplier. Moreover, when the dual sequence is unbounded, numerical instabilities may arise in the evaluation of stopping criteria. Our result, by establishing the boundedness of the dual sequence, enables a relaxed variant of the algorithm in which subproblems are solved under scaled stopping criteria, an approach that is not supported by the AKKT or CAKKT theory.

In contrast to previous works, the main contribution of this paper is to show that quasinormality, a relatively weak constraint qualification, suffices to guarantee the boundedness of the dual sequence generated by the augmented Lagrangian method for NSDP. Furthermore, we prove that every accumulation point of the primal-dual sequence satisfies the KKT conditions under quasinormality. This result considerably extends the class of NSDP problems for which strong global convergence properties can be rigorously established. Let

$$\mathcal{L}_\rho\colon \mathbb{R}^n \times \mathbb{S}^\ell \times \mathbb{S}^m \to \mathbb{R}$$

denote the standard Powell–Hestenes–Rockafellar augmented Lagrangian function, defined as

$$\mathcal{L}_\rho(x, \Lambda, \Omega) = f(x) + \frac{\rho}{2}\Big(\big\|H(x) + \tfrac{\Lambda}{\rho}\big\|^2 + \big\|\big[G(x) + \tfrac{\Omega}{\rho}\big]_+\big\|^2\Big), \tag{11}$$

where $\rho > 0$ is the penalty parameter, $\Lambda \in \mathbb{S}^\ell$ and $\Omega \in \mathbb{S}^m$ are the Lagrange multipliers approximations associated with the equality and semidefinite constraints, respectively.

**Algorithm 1** Augmented Lagrangian Algorithm

**Step 0 (Initialization):** Let $\tau \in (0,1)$, $\gamma > 1$, $\rho_1 > 0$, $\Omega_{\max} \in \mathbb{S}^m_+$, $\Lambda_{\min}, \Lambda_{\max} \in \mathbb{S}^\ell$ with $\Lambda_{min} \preceq \Lambda_{max}$.

Take a sequence $\{\varepsilon_k\} \subset \mathbb{R}$ such that $\varepsilon_k > 0$ and $\lim_{k\to\infty} \varepsilon_k = 0$.

Choose an initial point $x^0 \in \mathbb{R}^n$ and initial multipliers $0 \preceq \bar{\Omega}^1 \preceq \Omega_{\max}$ and $\Lambda_{\min} \preceq \bar{\Lambda}^1 \preceq \Lambda_{\max}$. $k := 1$.

**Step 1 (Subproblem Solution):** Using $x^{k-1}$, compute an approximate minimizer $x^k$ of $\mathcal{L}_{\rho_k}(x, \bar{\Lambda}^k, \bar{\Omega}^k)$ satisfying:

$$\|\nabla_x \mathcal{L}_{\rho_k}(x^k, \bar{\Lambda}^k, \bar{\Omega}^k)\| \le \varepsilon_k.$$

**Step 2 (Penalty Parameter Update):** Define

$$V^k := \left[\frac{\bar{\Omega}^k}{\rho_k} + G(x^k)\right]_+ - \frac{\bar{\Omega}^k}{\rho_k}.$$

**if** $k = 1$ or $\max\{\|H(x^k)\|, \|V^k\|\} \le \tau \max\{\|H(x^{k-1})\|, \|V^{k-1}\|\}$ **then**

  Set $\rho_{k+1} := \rho_k$.

**else**

  Set $\rho_{k+1} := \gamma\rho_k$.

**end if**

**Step 3 (Multiplier Update):**

$$\Omega^k := \left[\bar{\Omega}^k + \rho_k G(x^k)\right]_+, \quad \bar{\Omega}^{k+1} := \mathrm{proj}_{\mathcal{S}_1}(\Omega^k),$$

$$\Lambda^k := \bar{\Lambda}^k + \rho_k H(x^k), \quad \bar{\Lambda}^{k+1} := \mathrm{proj}_{\mathcal{S}_2}(\Lambda^k),$$

where $\mathcal{S}_1 := \{X \in \mathbb{S}^m \mid 0 \preceq X \preceq \Omega_{\max}\}$ and $\mathcal{S}_2 := \{Y \in \mathbb{S}^\ell \mid \Lambda_{\min} \preceq Y \preceq \Lambda_{\max}\}$.

Set $k := k + 1$ and return to Step 1.

---

*Remark* 5. We emphasize several important aspects of Algorithm 1:

i) Each subproblem solved at iteration $k$ is unconstrained, and the augmented Lagrangian function $\mathcal{L}_\rho$ simultaneously penalizes both equality and semidefinite constraint violations in a smooth and unified manner. In order to guarantee that a solution exists, one may add large enough box constraints. See [13].

ii) The projection operator $[\cdot]_+$ onto the cone of positive semidefinite matrices guarantees that the dual variables associated with the semidefinite constraint remain feasible throughout the iterations, preserving the structure of the problem.

iii) The matrix $V^k$ quantifies the degree of constraint violation at iteration $k$, serving as an indicator of the solution quality with respect to both feasibility and complementarity. A large value of $\|V^k\|$ indicates that the current iterate poorly satisfies the constraints or complementarity conditions. This measure is fundamental for assessing convergence toward feasibility and for monitoring the progressive reduction of violations.

iv) The penalty parameter $\rho_k$ is updated adaptively. It remains unchanged when $\|V^k\|$ and

$\|H(x^k)\|$ decreases sufficiently, and it is increased otherwise, in order to enforce feasibility and complementarity more strongly.

v) The projection $\text{proj}_{\mathcal{S}_i}(\cdot), i = 1, 2$ ensures that the matrix of multipliers remain bounded and within a prescribed feasible set, which is essential for establishing convergence under weak constraint qualifications.

vi) The algorithm allows inexact solutions of the subproblems, as controlled by the decreasing sequence $\{\varepsilon_k\}$. This feature enables practical implementations to balance computational efficiency with the theoretical convergence guarantees.

**Proposition 1.** *Let $f\colon \mathbb{R}^n \to \mathbb{R}$, $H\colon \mathbb{R}^n \to \mathbb{S}^\ell$, and $G\colon \mathbb{R}^n \to \mathbb{S}^m$ be continuously differentiable functions, and let $\mathcal{L}_\rho$ be the augmented Lagrangian function defined in* (11)*. Then, any accumulation point $\bar{x}$ of a sequence $\{x^k\}$ generated by Algorithm 1 is a stationary point of the infeasibility minimization problem*

$$\min_{x \in \mathbb{R}^n} \Phi(x) := \|H(x)\|^2 + \|[G(x)]_+\|^2. \tag{12}$$

The proof of Proposition 1 follows directly from Theorem 4.1 of [10] and Theorem 6.5 of [13]. These results establish that, under mild assumptions, any accumulation point of the sequence generated by an augmented Lagrangian algorithm satisfies first-order stationarity conditions for the infeasibility minimization problem (12). This implies that, even in the absence of standard constraint qualifications, the algorithm drives the iterates toward points where the equality constraints are nearly satisfied and violations of the semidefinite constraint are minimized. Consequently, stationary points of $\Phi$ can be interpreted as approximately feasible solutions to the original problem (NSDP), which demonstrates the robustness of the method in handling infeasibility. With this in mind, we proceed with the main global convergence result.

**Theorem 4.** *Let $\{x^k\}$ be a sequence generated by Algorithm 1, and suppose there exists an infinite subset $K \subset \mathbb{N}$ such that $x^k \to \bar{x}$ as $k \in K \to \infty$, where $\bar{x} \in \mathbb{R}^n$ is feasible. If $\bar{x}$ satisfies the quasinormality condition, then the sequence of dual variables $\{(\Lambda^k, \Omega^k)\}_{k \in K}$ is bounded, and $\bar{x}$ satisfies the KKT conditions. Moreover, any accumulation point of the dual sequence is a Lagrange multiplier associated with $\bar{x}$.*

*Proof.* By Step 1 of the algorithm

$$\nabla f(x^k) + DH(x^k)^*\Lambda^k + DG(x^k)^*\Omega^k \to 0, \tag{13}$$

where $\Lambda^k = \bar{\Lambda}^k + \rho_k H(x^k)$ and $\Omega^k = [\bar{\Omega}^k + \rho_k G(x^k)]_+$. Assume, by contradiction, that the sequence of multipliers $\{(\Lambda^k, \Omega^k)\}_{k \in K}$ is unbounded. Then, there exists an infinite subset $K_0 \subset K$ and a normalization sequence $\eta_k := \max\{\|\Lambda^k\|, \|\Omega^k\|\} \to \infty$ such that

$$\left(\frac{\Lambda^k}{\eta_k}, \frac{\Omega^k}{\eta_k}\right) \to (\Lambda^*, \Omega^*),$$

with $0 \neq (\Lambda^*, \Omega^*) \in \mathbb{S}^\ell \times \mathbb{S}^m_+$. Now, dividing (13) by $\eta_k$ and taking limits along $K_0$ yields

$$DH(\bar{x})^*\Lambda^* + DG(\bar{x})^*\Omega^* = 0. \tag{14}$$

We now analyze the behavior of the constraint violations relative to the limiting multipliers, where we restrict our attention to $k \in K_0$.

(a) Suppose that $\Omega^* \neq 0$. Since

$$\Omega^k = \left[\bar{\Omega}^k + \rho_k G(x^k)\right]_+ \quad \text{and} \quad \frac{\bar{\Omega}^k}{\eta_k} \to 0,$$

we obtain $\frac{\Omega^k}{\eta_k} = \left[\frac{\bar{\Omega}^k}{\eta_k} + \frac{\rho_k}{\eta_k} G(x^k)\right]_+$. Define

$$A_k := \frac{\bar{\Omega}^k}{\eta_k} + \frac{\rho_k}{\eta_k} G(x^k), \qquad B_k := \frac{\rho_k}{\eta_k} G(x^k).$$

Then $A_k - B_k = \bar{\Omega}^k/\eta_k \to 0$ due to the fact that $\{\bar{\Omega}^k\}$ is bounded. Since the projection $[\cdot]_+$ is nonexpansive, we have

$$\|[A_k]_+ - [B_k]_+\| \le \|A_k - B_k\| = \left\| \frac{\bar{\Omega}^k}{\eta_k} \right\| \to 0.$$

Consequently, $[A_k]_+ - [B_k]_+ \to 0$. But $[A_k]_+ = \Omega^k/\eta_k \to \Omega^*$, and therefore $[B_k]_+ \to \Omega^*$. That is,

$$\frac{\rho_k}{\eta_k} [G(x^k)]_+ \to \Omega^*. \tag{15}$$

Let $U_k$ be an orthogonal matrix such that

$$G(x^k) = U_k \operatorname{diag}\big(\lambda_1^{U_k}(G(x^k)), \ldots, \lambda_m^{U_k}(G(x^k))\big)\, U_k^\top.$$

Then the same matrix $U_k$ diagonalizes $\frac{\rho_k}{\eta_k}[G(x^k)]_+$. By (15), we may take an orthogonal matrix $U$ that diagonalizes $\Omega^*$ such that $U_k \to U$, ensuring the spectral pairing required in the definition of quasinormality. Now fix an index $i$ such that $\lambda_i^U(\Omega^*) > 0$. Then

$$\frac{\rho_k}{\eta_k} \lambda_i^{U_k}\Big([G(x^k)]_+\Big) \;\to\; \lambda_i^U(\Omega^*) > 0.$$

Hence,

$$\lambda_i^{U_k}\Big([G(x^k)]_+\Big) > 0 \quad \text{for all sufficiently large } k.$$

Since $\lambda_i^{U_k}\big([G(x^k)]_+\big) = \max\big\{\lambda_i^{U_k}(G(x^k)), 0\big\}$, it follows that $\lambda_i^{U_k}(G(x^k)) > 0 \quad \text{for all sufficiently large } k$, which proves the desired property.

(b) Suppose that $\Lambda^* \neq 0$. Since

$$\Lambda^k = \bar{\Lambda}^k + \rho_k H(x^k) \quad \text{and} \quad \frac{\bar{\Lambda}^k}{\eta_k} \to 0,$$

we obtain

$$\frac{\Lambda^k}{\eta_k} = \frac{\bar{\Lambda}^k}{\eta_k} + \frac{\rho_k}{\eta_k} H(x^k) \to \Lambda^*, \text{ which implies that } \frac{\rho_k}{\eta_k} H(x^k) \to \Lambda^*. \tag{16}$$

Let $V_k$ be an orthogonal matrix diagonalizing $H(x^k)$. Since $V_k$ also diagonalizes $(\rho_k/\eta_k)H(x^k)$, by (16) we may choose an orthogonal matrix $V$ diagonalizing $\Lambda^*$ such that $V_k \to V$, ensuring the spectral pairing in the definition of quasinormality. Fix $i$ with $\lambda_i^V(\Lambda^*) \neq 0$. By continuity of the paired eigenvalues we have

$$\lambda_i^{V_k}\left(\frac{\rho_k}{\eta_k} H(x^k)\right) \to \lambda_i^V(\Lambda^*),$$

so for all sufficiently large $k$, $\lambda_i^{V_k}\left(\frac{\rho_k}{\eta_k} H(x^k)\right)$ has the same sign of $\lambda_i^V(\Lambda^*)$, that is,

$$\lambda_i^{V_k}\left(\frac{\rho_k}{\eta_k} H(x^k)\right) \lambda_i^V(\Lambda^*) > 0.$$

Since $\rho_k/\eta_k > 0$, this is equivalent to

$$\lambda_i^{V_k}\big(H(x^k)\big)\, \lambda_i^V(\Lambda^*) > 0 \quad \text{for all sufficiently large } k,$$

which is the desired sign condition.

We have constructed $(\Lambda^*, \Omega^*) \neq (0,0)$, a sequence $x^k \to \bar{x}$, and diagonalizations $U_k \to U$, $V_k \to V$ such that

$$DH(\bar{x})^*\Lambda^* + DG(\bar{x})^*\Omega^* = 0,$$

and conditions (i)–(ii) of Definition 3 are satisfied. This contradicts the quasinormality assumption at $\bar{x}$. Therefore, the sequence $\{(\Lambda^k, \Omega^k)\}_{k\in K}$ must be bounded. The remaining conclusions follow simply by taking the limit of an appropriate subsequence in (13), where complementarity follows as in the proof of [10, Theorem 4].

□

The previous result shows that, under the assumption of quasinormality at a feasible point $\bar{x}$, the dual sequences generated by the augmented Lagrangian algorithm remain bounded, and any accumulation point satisfies the KKT conditions. We now proceed to demonstrate that quasinormality can indeed be regarded as a constraint qualification for NSDP. To that end, we first recall the construction in the proof of [10, Theorem 5], which guarantees the existence of asymptotic primal-dual sequences that satisfy approximate stationarity and complementarity conditions.

**Theorem 5** ([10])**.** *Let $\bar{x} \in \mathbb{R}^n$ be a local minimizer of* (NSDP)*. Then, there exist sequences $\rho_k \to \infty$, $x^k \to \bar{x}$, and*

$$\Omega^k := [\rho_k G(x^k)]_+ \in \mathbb{S}^m_+, \quad \Lambda^k := \rho_k H(x^k) \in \mathbb{S}^\ell,$$

*such that*

$$\nabla f(x^k) + DG(x^k)^*\Omega^k + DH(x^k)^*\Lambda^k \to 0 \quad \text{and} \quad \langle G(x^k), \Omega^k\rangle \to 0.$$

We observe that the sequences $\Lambda^k$ and $\Omega^k$ appearing in Theorem 5 match the structure of those produced by Algorithm 1 when the safeguarded multipliers are set to zero (i.e., $\bar{\Lambda}^k = 0$ and $\bar{\Omega}^k = 0$). This observation motivates the next result, in which we apply the quasinormality condition to these sequences to deduce that $\bar{x}$ satisfies the KKT conditions.

**Theorem 6.** *Let $\bar{x} \in \mathbb{R}^n$ be a local minimizer of* (NSDP)*, and suppose that the quasinormality condition holds at $\bar{x}$. Then $\bar{x}$ satisfies the KKT conditions for the same problem.*

*Proof.* According to Theorem 5, there exist sequences $\rho_k \to \infty$, $x^k \to \bar{x}$, and

$$\Omega^k := [\rho_k G(x^k)]_+ \in \mathbb{S}^m_+, \quad \Lambda^k := \rho_k H(x^k) \in \mathbb{S}^\ell,$$

such that

$$\nabla f(x^k) + DG(x^k)^*\Omega^k + DH(x^k)^*\Lambda^k \to 0, \qquad \langle G(x^k), \Omega^k\rangle \to 0.$$

Observe that the sequences $\{\Omega^k\}$ and $\{\Lambda^k\}$ coincide with those generated by Algorithm 1 when the safeguarded multipliers are set to zero, i.e., $\bar{\Omega}^k = 0$ and $\bar{\Lambda}^k = 0$. Then, by an argument analogous to that used in the proof of Theorem 4, and invoking the quasinormality assumption at $\bar{x}$, we conclude that the sequence of dual variables $\{(\Lambda^k, \Omega^k)\}$ is bounded. Hence, there exists a convergent subsequence with limit $(\Lambda, \Omega) \in \mathbb{S}^\ell \times \mathbb{S}^m_+$. Moreover, passing to the limit in the approximate stationarity condition and using the continuity of $\nabla f$, $DG$, and $DH$, we obtain

$$\nabla f(\bar{x}) + DG(\bar{x})^*\Omega + DH(\bar{x})^*\Lambda = 0.$$

Since $x^k \to \bar{x}$ and $G$ is continuous, $G(x^k) \to G(\bar{x})$. Together with $\langle G(x^k), \Omega^k\rangle \to 0$ and the convergence $\Omega^k \to \Omega$, this yields $\langle G(\bar{x}), \Omega\rangle = 0$. Therefore, $\bar{x}$ satisfies the KKT conditions for (NSDP) with Lagrange multipliers $\Lambda$ and $\Omega$. □

**Corollary 1.** *Let $\bar{x}$ be a local minimizer of* (NSDP)*. If $\bar{x}$ satisfies* pseudonormality*, then $\bar{x}$ satisfies the KKT conditions.*

*Proof.* By Theorem 3, pseudonormality implies quasinormality at $\bar{x}$. The conclusion then follows from Theorem 6. □

We now present a refined version of the augmented Lagrangian framework, inspired by Birgin, Haeser, and Martínez [12]. Their approach introduces a scaled stopping criterion for the inner minimization problems, designed to improve numerical performance by adapting the optimality tolerance to the magnitude of the updated multipliers.

Let $\mathcal{L}_\rho(x, \Lambda, \Omega)$ be the Powell–Hestenes–Rockafellar augmented Lagrangian defined in (11). In the classical scheme, the subproblem solution $x^k$ is required to satisfy

$$\|\nabla_x \mathcal{L}_\rho(x^k, \bar{\Lambda}^k, \bar{\Omega}^k)\| \le \varepsilon_k.$$

In the scaled version, this fixed tolerance is replaced by one proportional to the norms of the prospective multipliers,

$$\Lambda^k := \bar{\Lambda}^k + \rho_k H(x^k), \qquad \Omega^k := [\,\bar{\Omega}^k + \rho_k G(x^k)\,]_+,$$

so that the stopping test becomes

$$\|\nabla_x \mathcal{L}_{\rho_k}(x^k, \bar{\Lambda}^k, \bar{\Omega}^k)\| \le \varepsilon_k \, \max\{1, \|\bar{\Lambda}^k + \rho_k H(x^k)\|, \|[\,\bar{\Omega}^k + \rho_k G(x^k)\,]_+\|\}. \tag{17}$$

This scaling reflects that the accuracy required to solve the inner subproblems is measured relative to the magnitude of the dual quantities generated by the method. In particular, the scaled stopping rule allows one to solve the subproblems less accurately when the chosen scaling factor is large, while enforcing progressively higher accuracy through the prescribed decay of the tolerance sequence. We denote by *scaled augmented Lagrangian algorithm* the variant obtained from Algorithm 1 by replacing Step 1 with the scaled condition (17).

**Theorem 7.** *Let $\{x^k\}$ be a sequence generated by the scaled augmented Lagrangian algorithm, where each subproblem solution satisfies the scaled first-order condition* (17)*. Assume that $\bar{x}$ is a feasible accumulation point of the sequence, that is, there exists an infinite index set $K \subset \mathbb{N}$ such that $x^k \to \bar{x}$ as $k \in K$. If $\bar{x}$ satisfies the* quasinormality condition*, then the sequence of dual variables $\{(\Lambda^k, \Omega^k)\}_{k\in K}$ is bounded.*

*Proof.* The proof follows the same reasoning as in Theorem 4. Specifically, recall that the first step in that argument consists of normalizing the approximate stationarity condition

$$\|\nabla_x \mathcal{L}_{\rho_k}(x^k, \bar{\Lambda}^k, \bar{\Omega}^k)\| \le \varepsilon_k \max\{1, \|\Lambda^k\|, \|\Omega^k\|\},$$

by dividing both sides by the scaling factor $\max\{\|\Lambda^k\|, \|\Omega^k\|\}$. Since $\varepsilon_k \to 0$, the right-hand side converges to zero, implying that any normalized limit of the dual sequences satisfies the same limiting stationarity system obtained in Theorem 4. Consequently, the contradiction argument used there applies verbatim: if the multipliers were unbounded, one would obtain a nonzero limiting pair $(\Lambda^*, \Omega^*)$ violating the quasinormality condition at $\bar{x}$. Hence, both $\{\Lambda^k\}_{k\in K}$ and $\{\Omega^k\}_{k\in K}$ are bounded. □

# 5 Relaxed Quasinormality for NSDP

In nonlinear programming, quasinormality-type conditions arise naturally from refined optimality frameworks that go beyond the classical Fritz–John theory. These ideas led to *relaxed quasinormality* (RQN), see, e.g., [9], a condition that allows controlled degeneracy in constraint violations while preserving meaningful stationarity properties. We now propose a natural extension of this condition for NSDP.

**Definition 7.** Let $\bar{x} \in \mathbb{R}^n$ be a feasible point of (NSDP). We say that $\bar{x}$ satisfies the *relaxed quasinormality (RQN)* condition if there is no pair $(\Lambda, \Omega) \in \mathbb{S}^\ell \times \mathbb{S}^m_+$, with $(\Lambda, \Omega) \neq 0$, such that:

1. $DH(\bar{x})^*\Lambda + DG(\bar{x})^*\Omega = 0$;
2. There exists a sequence $\{x^k\}$ converging to $\bar{x}$ such that:
   (a) $\lambda_i^{V_k}(H(x^k))\lambda_i^V(\Lambda) > 0$ for all $i$ such that $\lambda_i^V(\Lambda) \neq 0$,
   (b) $\lambda_i^{U_k}(G(x^k)) > 0$ for all $i$ such that $\lambda_i^U(\Omega) > 0$,
   (c) $|\lambda_i^{V_k}(H(x^k))| = o(w(x^k))$ for all $i$ with $\lambda_i^V(\Lambda) = 0$, and $[\lambda_j^{U_k}(G(x^k))]_+ = o(w(x^k))$ for all $j$ with $\lambda_j^U(\Omega) = 0$, where the function $w(x^k)$ is defined as

$$w(x^k) := \min\left\{ \min_{\lambda_i^V(\Lambda)\neq 0} |\lambda_i^{V_k}(H(x^k))|, \min_{\lambda_j^U(\Omega)>0} [\lambda_j^U(G(x^k))]_+ \right\},$$

and $U_k$, $V_k$, $V$ and $U$ are orthogonal matrices that diagonalize $G(x^k)$, $H(x^k)$, $\Lambda$ and $\Omega$, respectively, such that $U_k \to U$ and $V_k \to V$.

The function $w(x^k)$ plays a crucial role in the relaxed quasinormality condition by acting as a reference scale to evaluate the relative significance of constraint violations. Specifically, $w(x^k)$ is defined as the minimum between the absolute values of the violated equality-matrix eigenvalue constraints $\lambda_i^{V_k}(H(x^k))$, for which the corresponding eigenvalues $\lambda_i^V(\Lambda)$ of the multiplier matrix are nonzero, and the positive eigenvalues of the matrix-valued constraint $G(x^k)$ associated with the positive eigenvalues of the multiplier matrix $\Omega$. This means that $w(x^k)$ reflects the smallest among the most relevant violations, those for which the associated multipliers are non-zero. The requirement that other violations, corresponding to constraints with vanishing multipliers, must satisfy $|\lambda_i^{V_k}(H(x^k))| = o(w(x^k))$ and $[\lambda_j^U(G(x^k))]_+ = o(w(x^k))$, ensures that these secondary violations remain asymptotically negligible in comparison. Consequently, the relaxed quasinormality condition allows a form of controlled degeneracy in the constraint system, as long as the primary violations dominate the asymptotic behavior of the sequence.

The relationship between quasinormality and relaxed quasinormality is straightforward, and we state it as follows.

**Theorem 8.** *Let $\bar{x} \in \mathbb{R}^n$ be a feasible point for (NSDP). If the quasinormality condition holds at $\bar{x}$, then $\bar{x}$ satisfies the relaxed quasinormality condition.*

*Proof.* We argue by contraposition. Suppose that $\bar{x}$ does not satisfy the relaxed quasinormality condition. Then, by Definition 7, there exist $(\Lambda, \Omega) \neq (0,0)$ and a sequence $x^k \to \bar{x}$ such that

$$DH(\bar{x})^*\Lambda + DG(\bar{x})^*\Omega = 0,$$

and the sign control conditions (a) and (b) in Definition 7 hold along $\{x^k\}$. Since these conditions coincide with the corresponding sign requirements in the definition of quasinormality, the same pair $(\Lambda, \Omega)$ and sequence $\{x^k\}$ provide a certificate that quasinormality fails at $\bar{x}$. Therefore, failure of relaxed quasinormality implies failure of quasinormality, and the result follows by contraposition. □

The following example shows that the implication from quasinormality to relaxed quasinormality is strict.

**Example 4.** Consider the matrix-valued equality constraint

$$H(x) = \frac{x}{2}\begin{pmatrix} 1+x & x-1 \\ x-1 & 1+x \end{pmatrix}.$$

and the feasible point $\bar{x} = 0$.

(i) *The relaxed quasinormality condition holds at $\bar{x} = 0$.* Suppose, by contradiction, that RQN fails at $\bar{x}$. Then, by Definition 7, there exist a nonzero multiplier $\Lambda \in \mathbb{S}^2$, $\Lambda \neq 0$ and a sequence $\{x_k\} \to \bar{x}$ such that the stationarity condition

$$DH(\bar{x})^*\Lambda = 0$$

holds, and items (a)–(c) of Definition 7 are satisfied. The eigenvalues of $H(x)$ are $x$ and $x^2$. In particular, for all sufficiently small $x \neq 0$, the eigenvalues must be distinct. This implies that $V_k$ is uniquely determined up to a permutation and a simple computation gives $V_k = V = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}$ with $\lambda_1^{V_k}(H(x^k)) = (x^k)^2$ and $\lambda_2^{V_k}(H(x^k)) = x^k$. Now, by cyclic trace invariance, we obtain

$$\begin{aligned} 0 = DH(\bar{x})^*\Lambda &= \left\langle \frac{1}{2}\begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}, \Lambda \right\rangle = \frac{1}{2}\left\langle V^\top \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} V,\, V^\top \Lambda V \right\rangle \\ &= \left\langle \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} \lambda_1^V(\Lambda) & 0 \\ 0 & \lambda_2^V(\Lambda) \end{pmatrix} \right\rangle = \lambda_2^V(\Lambda). \end{aligned}$$

Since $\Lambda \neq 0$, it follows that $\lambda_1^V(\Lambda) \neq 0$. Therefore,

$$w(x_k) = \min_{\lambda_i^V(\Lambda) \neq 0} \left|\lambda_i^{V_k}(H(x_k))\right| = |\lambda_1^{V_k}(H(x_k))| = (x^k)^2.$$

On the other hand, for the index $i = 2$ satisfying $\lambda_2^V(\Lambda) = 0$, item (c) of Definition 7 requires

$$|\lambda_2^{V_k}(H(x_k))| = o\big(w(x_k)\big), \quad \text{that is,} \quad |x_k| = o((x_k)^2).$$

This is impossible for any sequence $x_k \to 0$ with $x_k \neq 0$, since

$$\frac{|x_k|}{x_k^2} = \frac{1}{|x_k|} \to +\infty.$$

This contradiction shows that the relaxed quasinormality condition holds at $\bar{x} = 0$.

(ii) *The quasinormality condition fails at* $\bar{x} = 0$*.* Consider the sequence $x^k = 1/k \to 0$ and $V_k = V = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}$. For the nonzero multiplier $\Lambda = V\mathrm{diag}(1,0)V^T$, we have $DH(\bar{x})^*\Lambda = 0$ while $\lambda_1^V(\Lambda) = 1$ and $\lambda_2^V(\Lambda) = 0$, with

$$\lambda_1^{V_k}\big(H(x^k)\big) = (x^k)^2 > 0.$$

Therefore $\Lambda$ and the sequence $\{x^k\}$ satisfy the requirements in Definition 3, and quasinormality does not hold at $\bar{x}$.

We now show that, under the assumption of relaxed quasinormality, the sequence of Lagrange multipliers generated by the augmented Lagrangian algorithm is also bounded.

**Theorem 9.** *Let* $\{x^k\}$ *be a sequence generated by Algorithm 1 and let* $\bar{x}$ *be a feasible accumulation point, that is, there exists an infinite index set* $K \subset \mathbb{N}$ *such that* $x^k \to \bar{x}$ *as* $k \in K$*. If* $\bar{x}$ *satisfies the relaxed quasinormality condition, then the corresponding dual sequences* $\{\Lambda^k\}_{k\in K}$ *and* $\{\Omega^k\}_{k\in K}$ *generated in Step 2 are bounded. In particular, every accumulation point of the dual sequence is a Lagrange multiplier associated with* $\bar{x}$*.*

*Proof.* Assume, by contradiction, that the dual sequence $\{(\Lambda^k, \Omega^k)\}_{k\in K}$ is unbounded. Then, there exists an infinite subset $K_0 \subset K$ such that

$$\eta_k := \max\{\|\Lambda^k\|, \|\Omega^k\|\} \to +\infty \quad \text{as } k \in K_0,$$

and dividing the approximate stationarity condition in Step 1 by $\eta_k$ and taking limits along $K_0$, we obtain

$$DH(\bar{x})^*\Lambda^* + DG(\bar{x})^*\Omega^* = 0,$$

with $0 \neq (\Lambda^*, \Omega^*) \in \mathbb{S}^\ell \times \mathbb{S}^m_+$, where

$$\frac{\Lambda^k}{\eta_k} = \frac{\bar{\Lambda}^k + \rho_k H(x^k)}{\eta_k} \longrightarrow \Lambda^*, \qquad \frac{\Omega^k}{\eta_k} = \frac{[\bar{\Omega}^k + \rho_k G(x^k)]_+}{\eta_k} \longrightarrow \Omega^*. \tag{18}$$

By Theorem 4, the sequences generated by Algorithm 1 satisfy conditions (a) and (b) of Definition 7, where $V_k, U_k, V$ and $U$ are orthogonal matrices diagonalizing $H(x^k), G(x^k), \Lambda^*$ and $\Omega^*$, respectively, such that $V_k \to V$ and $U_k \to U$. It remains to verify condition (c).

If all eigenvalues of $\Lambda^*$ and $\Omega^*$ are nonzero, then condition (c) holds trivially. Otherwise, let $i$ and $j$ be any index such that

$$(\lambda_i^V(\Lambda^*) = 0 \text{ or } \lambda_i^U(\Omega^*) = 0) \quad \text{and} \quad (\lambda_j^V(\Lambda^*) \neq 0 \text{ or } \lambda_j^U(\Omega^*) \neq 0).$$

Assume that $\lambda_i^V(\Lambda^*) = 0$ and $\lambda_j^V(\Lambda^*) \neq 0$ (the other cases are analogous). From (18), it follows that

$$\lambda_i^{V_k}\left(\tfrac{\rho_k}{\eta_k}H(x^k)\right) \to \lambda_i^V(\Lambda^*) = 0, \qquad \lambda_j^{V_k}\left(\tfrac{\rho_k}{\eta_k}H(x^k)\right) \to \lambda_j^V(\Lambda^*) \neq 0.$$

Consequently,

$$\frac{\lambda_i^{V_k}(H(x^k))}{\lambda_j^{V_k}(H(x^k))} = \frac{\lambda_i^{V_k}\left(\frac{\rho_k}{\eta_k}H(x^k)\right)}{\lambda_j^{V_k}\left(\frac{\rho_k}{\eta_k}H(x^k)\right)} \to 0.$$

Recalling the definition of $w(x^k)$ and considering that the limit above holds for any $j$, we conclude that

$$\frac{|\lambda_i^{V_k}(H(x^k))|}{w(x^k)} \to 0,$$

which implies

$$|\lambda_i^{V_k}(H(x^k))| = o\big(w(x^k)\big) \quad \text{for all } i \text{ such that } \lambda_i^V(\Lambda^*) = 0.$$

The argument is similar when $\lambda_i^U(\Omega^*) = 0$. Combining the above arguments, we conclude that the nonzero pair $(\Lambda^*, \Omega^*)$ together with the sequence $\{x^k\}_{k\in K_0}$ satisfies items (1) and (2)(a)–(c) of Definition 7, which contradicts the relaxed quasinormality of $\bar{x}$. Hence, the assumption that $\{(\Lambda^k, \Omega^k)\}_{k\in K}$ is unbounded is false, and both dual sequences are bounded. The proof now concludes as in the proof of Theorem 4. □

# 6 Applications of Quasinormality

In this section, we derive two fundamental implications of the quasinormality condition introduced in this work. First, we show that quasinormality yields a local error bound for NSDP, thereby quantifying the relation between feasibility and constraint violation. Second, we demonstrate that the same condition ensures the existence of an exact penalty formulation. Together, these results reveal, in the semidefinite setting, the classical interplay between geometric regularity and error–bound behavior. We recall that the feasible set of (NSDP) is given by $\mathcal{F} := \{x : H(x) = 0,\ G(x) \preceq 0\}$.

**Definition 8.** A feasible point $\bar{x} \in \mathcal{F}$ satisfies a local error bound if there exist $c > 0$ and $\varepsilon > 0$ such that

$$\operatorname{dist}(x, \mathcal{F}) \ \le\ c\,(\|H(x)\| + \|[G(x)]_+\|) \quad \text{for all } x \text{ with } \|x - \bar{x}\| \le \varepsilon,$$

where $\operatorname{dist}(x, \mathcal{F})$ is the Euclidean distance from $x$ to $\mathcal{F}$.

In order to prove that quasinormality indeed yields a local error bound for NSDP, we will need some preliminary results. We begin with the following geometric lemma, which is standard and corresponds to Lemma 2.1 in [25].

**Lemma 2.** *Let $C \subset \mathbb{R}^n$ be a closed set, $x \notin C$, and let $y \in C$ satisfy $\|x - y\| = \operatorname{dist}(x, C)$. Then*

$$\frac{x - y}{\|x - y\|} \in [T_C(y)]^\circ,$$

*where $T_C(y)$ is the contingent (Bouligand) cone to $C$ at $y$, defined by*

$$T_C(y) := \{d \in \mathbb{R}^n : \exists\, t_k \downarrow 0,\ d_k \to d \text{ such that } y + t_k d_k \in C\}.$$

Now, we establish a preliminary lemma showing that the quasinormality condition is stable under small perturbations of the feasible point. This result extends the corresponding local stability property proved for nonlinear programming problems in [26].

**Lemma 3.** *Let $\bar{x} \in \mathcal{F}$ satisfy the quasinormality condition. Then there exists $\varepsilon > 0$ such that every feasible point $x \in \mathcal{F}$ with $\|x - \bar{x}\| < \varepsilon$ also satisfies quasinormality.*

*Proof.* Suppose the claim is false. Then there exists a sequence $\{x^k\} \subset \mathcal{F}$ such that $x^k \to \bar{x}$ and, for every $k$, the point $x^k$ violates the quasinormality condition. Hence, for each $k$, there exist multipliers $(\Lambda^k, \Omega^k) \in \mathbb{S}^\ell \times \mathbb{S}^m_+$ with $(\Lambda^k, \Omega^k) \neq (0, 0)$ and a sequence $\{x^{k,l}\}_l \subset \mathbb{R}^n$ with $x^{k,l} \to x^k$ such that

$$DH(x^k)^*\Lambda^k + DG(x^k)^*\Omega^k = 0, \tag{19}$$

and there exist orthogonal matrices $V_k, U_k$ diagonalizing $\Lambda^k, \Omega^k$, respectively, and orthogonal matrices $V_{k,l}, U_{k,l}$ diagonalizing $H(x^{k,l})$ and $G(x^{k,l})$, respectively, with

$$V_{k,l} \to V_k, \qquad U_{k,l} \to U_k \qquad (l \to \infty),$$

such that the sign conditions in Definition 3 hold along $\{x^{k,l}\}_l$, namely:

$$\lambda_i^{V_{k,l}}\big(H(x^{k,l})\big)\,\lambda_i^{V_k}(\Lambda^k) > 0 \quad \text{for all } i \text{ with } \lambda_i^{V_k}(\Lambda^k) \neq 0, \tag{20}$$

$$\lambda_i^{U_{k,l}}\big(G(x^{k,l})\big) > 0 \quad \text{for all } i \text{ with } \lambda_i^{U_k}(\Omega^k) > 0. \tag{21}$$

Normalize the multipliers by setting

$$M_k := \|(\Lambda^k, \Omega^k)\|, \qquad (\widehat{\Lambda}^k, \widehat{\Omega}^k) := \frac{1}{M_k}(\Lambda^k, \Omega^k),$$

so that $\|(\widehat{\Lambda}^k, \widehat{\Omega}^k)\| = 1$ and $\widehat{\Omega}^k \succeq 0$ for all $k$. Thus, there exists a subsequence such that

$$(\widehat{\Lambda}^k, \widehat{\Omega}^k) \to (\Lambda, \Omega) \neq (0,0), \qquad V_k \to V, \qquad U_k \to U,$$

for some orthogonal matrices $V \in \mathbb{R}^{\ell\times\ell}$ and $U \in \mathbb{R}^{m\times m}$. Dividing (19) by $M_k$ and letting $k \to \infty$ in such subsequence, using continuity of $DH(\cdot)^*$ and $DG(\cdot)^*$, yields

$$DH(\bar{x})^*\Lambda + DG(\bar{x})^*\Omega = 0, \text{ with } 0 \neq (\Lambda, \Omega) \in \mathbb{S}^\ell \times \mathbb{S}^m_+. \tag{22}$$

We now build a single sequence converging to $\bar{x}$ that violates Definition 3 at $\bar{x}$. For each $k$, since $x^{k,l} \to x^k$, and $U_{k,l} \to U_k$, $V_{k,l} \to V_k$ as $l \to \infty$, we may choose an index $l(k)$ sufficiently large so that, with

$$z^k := x^{k,l(k)}, \qquad \widetilde{U}_k := U_{k,l(k)}, \qquad \widetilde{V}_k := V_{k,l(k)},$$

we have $z^k \to \bar{x}$, $\widetilde{U}_k \to U$, and $\widetilde{V}_k \to V$. Since $U_k \to U$ and $\widehat{\Omega}^k \to \Omega$, for every index $i$ with $\lambda_i^U(\Omega) > 0$ we have $\lambda_i^{U_k}(\widehat{\Omega}^k) > 0$ for all sufficiently large $k$. Then (21) implies

$$\lambda_i^{\widetilde{U}_k}\big(G(z^k)\big) > 0 \quad \text{for all sufficiently large } k \text{ and all } i \text{ with } \lambda_i^U(\Omega) > 0.$$

Similarly, since $V_k \to V$ and $\widehat{\Lambda}^k \to \Lambda$, for every index $i$ with $\lambda_i^V(\Lambda) \neq 0$ we have $\lambda_i^{V_k}(\widehat{\Lambda}^k) \neq 0$ for all sufficiently large $k$, and (20) yields

$$\lambda_i^{\widetilde{V}_k}\big(H(z^k)\big)\,\lambda_i^V(\Lambda) > 0 \quad \text{for all sufficiently large } k \text{ and all } i \text{ with } \lambda_i^V(\Lambda) \neq 0.$$

Together with (22), this shows that $(\Lambda, \Omega)$ and the sequence $\{z^k\}$ contradicts quasinormality at $\bar{x}$. □

Before establishing the final auxiliary result required for the proof of the error bound, we need one additional lemma providing a gradient representation of normal vectors to the feasible set. This result is classical in variational analysis; see, for instance, [28, p. 205].

**Lemma 4.** *Let $\bar{x} \in \mathcal{F}$ be a feasible point. For every vector $y \in T_{\mathcal{F}}(\bar{x})^\circ$, there exists a function $F\colon \mathbb{R}^n \to \mathbb{R}$ of class $\mathcal{C}^1$ such that*

$$-\nabla F(\bar{x}) = y,$$

*and $\bar{x}$ is a strict global minimizer of $F$ over $\mathcal{F}$.*

We now establish the final auxiliary result needed for the error bound analysis. This proposition extends to the NSDP setting a representation result originally proved for nonlinear programming in [26].

**Proposition 2.** *Let $\bar{x} \in \mathcal{F}$ be a feasible point and suppose that the quasinormality condition holds at $\bar{x}$. Then, for any vector $y \in T_{\mathcal{F}}(\bar{x})^{\circ}$, there exist matrices $\Lambda \in \mathbb{S}^{\ell}$ and $\Omega \in \mathbb{S}^m_+$ such that*

$$y = DH(\bar{x})^*\Lambda + DG(\bar{x})^*\Omega.$$

*Moreover, there exists a sequence $\{x^k\} \subset \mathbb{R}^n$ with $x^k \to \bar{x}$ and orthogonal matrices $U_k, V_k$ diagonalizing $G(x^k)$ and $H(x^k)$, respectively, together with orthogonal matrices $U, V$ diagonalizing $\Omega$ and $\Lambda$, respectively, such that $U_k \to U$, $V_k \to V$, and*

$$\lambda_i^{V_k}\big(H(x^k)\big)\,\lambda_i^{V}(\Lambda) > 0 \quad \text{for all } i \text{ with } \lambda_i^{V}(\Lambda) \neq 0, \tag{23}$$

$$\lambda_i^{U_k}\big(G(x^k)\big) > 0 \quad \text{for all } i \text{ with } \lambda_i^{U}(\Omega) > 0. \tag{24}$$

*Proof.* Let $y \in T_{\mathcal{F}}(\bar{x})^{\circ}$ be arbitrary. By Lemma 4, there exists a function $F\colon \mathbb{R}^n \to \mathbb{R}$ of class $\mathcal{C}^1$ that achieves a strict global minimum over $\mathcal{F}$ at $\bar{x}$ and satisfies $-\nabla F(\bar{x}) = y$. The proof now follows similarly to the proof of Theorem 6. We present the full argument below for completeness. Fix $\varepsilon > 0$ and let $B[\bar{x}, \varepsilon]$ denote the closed ball centered at $\bar{x}$ with radius $\varepsilon$. For each integer $k \geq 1$, consider the penalized problem

$$\min_{x \in B[\bar{x},\varepsilon]} \; F_k(x) := F(x) + k\,\|H(x)\|^2 + k\,\|[G(x)]_+\|^2,$$

Since $B[\bar{x}, \varepsilon]$ is compact and $F_k$ is continuous, there exists a minimizer $x^k$ of $F_k$. We first show that $x^k \to \bar{x}$. Indeed, for all $k$,

$$F(x^k) + k\,\|H(x^k)\|^2 + k\,\|[G(x^k)]_+\|^2 = F_k(x^k) \leq F_k(\bar{x}) = F(\bar{x}),$$

Since $F$ is bounded on $B[\bar{x}, \varepsilon]$, it follows that

$$\lim_{k\to\infty} \|H(x^k)\| = 0 \quad \text{and} \quad \lim_{k\to\infty} \|[G(x^k)]_+\| = 0.$$

Hence, every accumulation point of the sequence $\{x^k\}$ is feasible. We now show that $\bar{x}$ is in fact the unique accumulation point. Assume by contradiction that there exists a subsequence $\{x^{k_j}\}$ converging to some $\tilde{x} \neq \bar{x}$. By feasibility of accumulation points, we have $\tilde{x} \in \mathcal{F}$. Since $F$ attains a strict global minimum over $\mathcal{F}$ at $\bar{x}$, it follows that $F(\tilde{x}) > F(\bar{x})$. On the other hand, from the penalized problem we have $F(x^k) \leq F(\bar{x})$ for all $k$, and by continuity of $F$,

$$\lim_{j\to\infty} F(x^{k_j}) = F(\tilde{x}),$$

which yields a contradiction. Therefore, $\bar{x}$ is the unique accumulation point of $\{x^k\}$, and $x^k \to \bar{x}$. In particular, for all $k$ sufficiently large, $x^k$ lies in the interior of $B[\bar{x}, \varepsilon]$. For such $k$, the first-order necessary optimality condition for the penalized problem yields

$$\nabla F_k(x^k) = 0.$$

Let $V_k$ and $U_k$ be orthogonal matrices diagonalizing $H(x^k)$ and $G(x^k)$, respectively. Define

$$\Lambda_k := 2k\,H(x^k), \qquad \Omega_k := 2k\,[G(x^k)]_+.$$

so that

$$\nabla F(x^k) + DH(x^k)^*\Lambda_k + DG(x^k)^*\Omega_k = 0.$$

Consider the normalization

$$\delta_k := \max\{1, \|\Lambda^k\|, \|\Omega^k\|\}, \quad \widehat{\Lambda}_k := \frac{\Lambda_k}{\delta_k}, \quad \widehat{\Omega}_k := \frac{\Omega_k}{\delta_k}.$$

Thus, there exists a subsequence such that

$$\widehat{\Lambda}_k \to \Lambda \in \mathbb{S}^\ell, \qquad \widehat{\Omega}_k \to \Omega \in \mathbb{S}^m_+.$$

Passing to the limit and using $x^k \to \bar{x}$, we obtain

$$\nabla F(\bar{x}) + DH(\bar{x})^*\Lambda + DG(\bar{x})^*\Omega = 0.$$

Recalling that $-\nabla F(\bar{x}) = y$, we conclude that

$$y = DH(\bar{x})^*\Lambda + DG(\bar{x})^*\Omega.$$

Finally, conditions (23)–(24) follows from the definitions of $\Lambda^k$ and $\Omega^k$.

□

We are now in a position to establish a local error bound for NSDP under the quasinormality condition.

**Theorem 10.** *Let $\bar{x} \in \mathcal{F}$ be a feasible point. If $\bar{x}$ satisfies the quasinormality condition, then a local error bound holds at $\bar{x}$.*

*Proof.* Assume, by contradiction, that the local error bound fails at $\bar{x}$. Then, for every $k \in \mathbb{N}$, there exists $x^k \to \bar{x}$ such that

$$\operatorname{dist}(x^k, \mathcal{F}) \ > \ k\big(\|H(x^k)\| + \|[G(x^k)]_+\|\big). \tag{25}$$

In particular, $x^k \notin \mathcal{F}$ for all $k$. Let $y^k \in \mathcal{F}$ satisfy

$$\|x^k - y^k\| = \operatorname{dist}(x^k, \mathcal{F}).$$

Since $\bar{x} \in \mathcal{F}$, we have

$$\|x^k - y^k\| = \operatorname{dist}(x^k, \mathcal{F}) \leq \|x^k - \bar{x}\| \to 0,$$

and therefore

$$\|y^k - \bar{x}\| \leq \|y^k - x^k\| + \|x^k - \bar{x}\| \to 0, \qquad \text{i.e.,} \qquad y^k \to \bar{x}.$$

Define

$$\eta_k := \frac{x^k - y^k}{\|x^k - y^k\|}.$$

By Lemma 2, it follows that

$$\eta_k \in [T_{\mathcal{F}}(y^k)]^\circ. \tag{26}$$

Since $y^k \to \bar{x}$ and $\bar{x}$ satisfies the quasinormality condition, Lemma 3 ensures that, for all $k$ sufficiently large, $y^k$ also satisfies quasinormality. Therefore, Proposition 2 applies at $y^k$ and yields the existence of matrices $\Lambda^k \in \mathbb{S}^\ell$ and $\Omega^k \in \mathbb{S}^m_+$ such that

$$\eta_k = DH(y^k)^*\Lambda^k + DG(y^k)^*\Omega^k, \tag{27}$$

and there exists a sequence $y^{k,l} \to y^k$, and $U_{k,l} \to U_k$, $V_{k,l} \to V_k$ as $l \to \infty$, such that

$$\lambda_i^{U_{k,l}}\big(G(y^{k,l})\big) > 0 \quad \text{for all sufficiently large } k \text{ and all } i \text{ with } \lambda_i^{U_k}(\Omega^k) > 0, \tag{28}$$

and

$$\lambda_i^{V_{k,l}}\big(H(y^{k,l})\big)\,\lambda_i^{V_k}(\Lambda^k) > 0 \quad \text{for all sufficiently large } k \text{ and all } i \text{ with } \lambda_i^{V_k}(\Lambda^k) \neq 0. \tag{29}$$

Notice that by Remark 2, $\langle G(y^k), \Omega^k\rangle = 0$. We claim that the sequence $\{(\Lambda^k, \Omega^k)\}$ is bounded. Assume by contradiction that there is a subsequence such that $M_k := \|(\Lambda^k, \Omega^k)\| \to \infty$ and the normalized multipliers

$$(\widehat{\Lambda}^k, \widehat{\Omega}^k) := \frac{1}{M_k}(\Lambda^k, \Omega^k)$$

are convergent, say,

$$(\widehat{\Lambda}^k, \widehat{\Omega}^k) \to (\Lambda, \Omega) \neq (0,0), \qquad \Omega \succeq 0.$$

We consider only $k$ in this subsequence from now on. Dividing (27) by $M_k$ and letting $k \to \infty$, we obtain

$$DH(\bar{x})^*\Lambda + DG(\bar{x})^*\Omega = 0. \tag{30}$$

Moreover, if $\lambda_i^U(\Omega) > 0$, then $\lambda_i^{U_k}(\Omega^k) > 0$ for all sufficiently large $k$. Hence, for such indices, inequality (28) yields

$$\lambda_i^{U_{k,l}}\big(G(y^{k,l})\big) > 0 \quad \text{for all sufficiently large } k.$$

For each $k$, choose an index $l(k) \to +\infty$ as $k \to +\infty$ so that the spectral sign conditions above hold. Define the sequence $z^k := y^{k,l(k)}$. Since $y^{k,l} \to y^k$ as $l \to \infty$ and $y^k \to \bar{x}$, it follows that $z^k \to \bar{x}$ and

$$\lambda_i^{U_{k,l(k)}}\big(G(z^k)\big) > 0.$$

Similarly, if $\lambda_i^V(\Lambda) \neq 0$, then $\lambda_i^{V_k}(\Lambda^k) \neq 0$ for all sufficiently large $k$. Hence, for such indices and increasing $l(k)$ if needed, inequality (29) yields

$$\lambda_i^{V_{k,l(k)}}\big(H(z^k)\big)\,\lambda_i^V(\Lambda) > 0.$$

Moreover, since $U_{k,l} \to U_k$ and $U_k \to U$, and similarly $V_{k,l} \to V_k$ and $V_k \to V$, it follows that

$$U_{k,l(k)} \to U, \qquad V_{k,l(k)} \to V.$$

Thus, the spectral pairing required in the definition of quasinormality at $\bar{x}$ is satisfied. Together with (30), this contradicts the quasinormality condition at $\bar{x}$. Hence, $\{(\Lambda^k, \Omega^k)\}$ must be bounded. Consequently, there exists $M > 0$ such that

$$\max\{\|\Lambda^k\|, \|\Omega^k\|\} \leq M \qquad \text{for all } k. \tag{31}$$

Using (27) and the definition of $\eta_k$, we obtain

$$\|x^k - y^k\| = \langle \eta_k, x^k - y^k\rangle = \langle \Lambda^k, DH(y^k)(x^k - y^k)\rangle + \langle \Omega^k, G(y^k) + DG(y^k)(x^k - y^k)\rangle,$$

where we used that $\langle G(y^k), \Omega^k\rangle = 0$. By diagonalizing $G(y^k) + DG(y^k)(x^k - y^k)$ and using cyclic trace invariance, one can see that, since $\Omega^k$ is positive semidefinite,

$$\langle \Omega^k, G(y^k) + DG(y^k)(x^k - y^k)\rangle \leq \langle \Omega^k, [G(y^k) + DG(y^k)(x^k - y^k)]_+\rangle.$$

Thus, by Cauchy–Schwarz and (31),

$$\|x^k - y^k\| \leq M\Big(\|DH(y^k)(x^k - y^k)\| + \|[G(y^k) + DG(y^k)(x^k - y^k)]_+\|\Big). \tag{32}$$

Recalling that $y^k \in \mathcal{F}$, by differentiability of $H$ and $G$ at $y^k$, it follows that

$$H(x^k) = DH(y^k)(x^k - y^k) + r_H^k, \qquad G(x^k) = G(y^k) + DG(y^k)(x^k - y^k) + r_G^k,$$

where $\|r_H^k\| + \|r_G^k\| = o(\|x^k - y^k\|)$. Consequently,

$$\|DH(y^k)(x^k - y^k)\| \leq \|H(x^k)\| + \|r_H^k\|.$$

Similarly,

$$\begin{aligned}&\big\|[G(y^k) + DG(y^k)(x^k - y^k)]_+\big\| \leq \\ &\leq \|[G(x^k)]_+\| + \big\|[G(y^k) + DG(y^k)(x^k - y^k)]_+ - [G(x^k)]_+\big\| \leq \|[G(x^k)]_+\| + \|r_G^k\|,\end{aligned}$$

where we used the triangular inequality and nonexpansiveness of the projection. Substituting into (32) yields

$$\|x^k - y^k\| \leq M\Big(\|H(x^k)\| + \|[G(x^k)]_+\| + \|r_H^k\| + \|r_G^k\|\Big).$$

Since $\|r_H^k\| + \|r_G^k\| = o(\|x^k - y^k\|)$, we have $\|r_H^k\| + \|r_G^k\| \leq 1/(2M)\|x^k - y^k\|$ for $k$ sufficiently large, therefore

$$(1 - M/(2M))\|x^k - y^k\| \leq M\Big(\|H(x^k)\| + \|[G(x^k)]_+\|\Big).$$

That is

$$\operatorname{dist}(x^k, \mathcal{F}) = \|x^k - y^k\| \leq 2M\big(\|H(x^k)\| + \|[G(x^k)]_+\|\big),$$

contradicting (25).

□

**Definition 9.** Let $\bar{x} \in \mathcal{F}$ be a feasible point. We say that $\bar{x}$ admits a *(strict) exact penalty* if, for every continuously differentiable function $f$ for which $\bar{x}$ is a (strict) local minimizer over $\mathcal{F}$, there exists a constant $c > 0$ such that $\bar{x}$ is also a (strict) local minimizer of the penalized problem

$$\min_{x \in \mathbb{R}^n} f(x) + c\left(\|H(x)\| + \|[G(x)]_+\|\right).$$

We next derive the existence of an exact penalty function for $\mathcal{F}$ under the quasinormality condition. This result follows from the validity of a local error bound.

**Theorem 11.** *Let $\bar{x} \in \mathcal{F}$ be a feasible point. If $\bar{x}$ satisfies the quasinormality condition, then $\bar{x}$ admits an exact penalty and a scrict exact penalty.*

*Proof.* Let $f\colon \mathbb{R}^n \to \mathbb{R}$ be a continuously differentiable function for which $\bar{x}$ is a (strict) local minimizer over the feasible set $\mathcal{F}$. By Theorem 10, quasinormality at $\bar{x}$ implies the existence of constants $\kappa > 0$ and $\varepsilon > 0$ such that

$$\operatorname{dist}(x, \mathcal{F}) \;\leq\; \kappa\big(\|H(x)\| + \|[G(x)]_+\|\big) \quad \text{for all } x \text{ with } \|x - \bar{x}\| \leq \varepsilon. \tag{33}$$

Define the violation function

$$\phi(x) := \|H(x)\| + \|[G(x)]_+\|.$$

Since $f$ is $C^1$, it is Lipschitz continuous on the closed ball $B[\bar{x}, \varepsilon]$. Hence there exists $L > 0$ such that

$$|f(x) - f(y)| \;\leq\; L\|x - y\| \qquad \text{for all } x, y \in B[\bar{x}, \varepsilon].$$

Because $\bar{x}$ is a (strict) local minimizer of $f$ over $\mathcal{F}$, there exists $\rho \in (0, \varepsilon]$ such that

$$f(x) \geq f(\bar{x}), (f(x) > f(\bar{x}), \text{ resp.}) \qquad \text{for all } x \in \mathcal{F} \text{ with } 0 < \|x - \bar{x}\| \leq \rho.$$

Fix any penalty parameter $c > \kappa L$, and consider the penalized objective

$$F_c(x) := f(x) + c\,\phi(x) = f(x) + c\big(\|H(x)\| + \|[G(x)]_+\|\big).$$

We now show that $\bar{x}$ is a (strict) local minimizer of $F_c$. Let $x \in B[\bar{x}, \rho/2]$ be arbitrary with $x \neq \bar{x}$. Choose $y \in \mathcal{F}$ such that

$$\|x - y\| = \operatorname{dist}(x, \mathcal{F}).$$

Since $\bar{x} \in \mathcal{F}$, we have

$$\operatorname{dist}(x, \mathcal{F}) \leq \|x - \bar{x}\|,$$

and therefore

$$\|y - \bar{x}\| \leq \|y - x\| + \|x - \bar{x}\| \leq 2\|x - \bar{x}\| \leq \rho.$$

Hence $y \in \mathcal{F} \cap B[\bar{x}, \rho]$. By the Lipschitz continuity of $f$,

$$f(x) \;\geq\; f(y) - L\|x - y\| \;=\; f(y) - L \operatorname{dist}(x, \mathcal{F}).$$

Using the error bound (33), we obtain

$$f(x) \;\geq\; f(y) - L\kappa\,\phi(x).$$

Hence

$$F_c(x) = f(x) + c\,\phi(x) \;\geq\; f(y) + (c - \kappa L)\,\phi(x). \tag{34}$$

Let us show that $F_c(x) \geq F_c(\bar{x})$ ($F_c(x) > F_c(\bar{x})$, resp.). We distinguish two cases:
*Case 1:* $x \in \mathcal{F}$. Then $\phi(x) = 0$ and $F_c(x) = f(x)$. The (strict) local minimality of $\bar{x}$ over $\mathcal{F}$ implies $F_c(x) = f(x) \geq f(\bar{x}) = F_c(\bar{x})$ $(F_c(x) > F_c(\bar{x}), \text{ resp.})$.
*Case 2:* $x \notin \mathcal{F}$. Then $\phi(x) > 0$, and since $c - \kappa L > 0$, inequality (34) yields

$$F_c(x) \;\geq\; f(y) + (c - \kappa L)\,\phi(x) \;>\; f(y) \;\geq\; f(\bar{x}) = F_c(\bar{x}),$$

where the last inequality follows from the local minimality of $\bar{x}$ over $\mathcal{F}$. This concludes the proof. □

# 7 Conclusions

In this work, we have developed and analyzed new quasinormality-type regularity conditions tailored to NSDP. Motivated by the need for theoretically robust and geometrically meaningful constraint qualifications, we introduced a novel definition of quasinormality that generalizes the classical NLP concept by incorporating eigenvalue-based criteria consistent with the spectral structure of the matrix constraints.

Unlike previous definitions, such as the componentwise quasinormality of Guo et al. [19], which rely on fixed arbitrary bases and slack variables, our approach dynamically adapts to the eigenspaces of the constraint and multiplier matrices. This provides a more accurate interpretation of constraint activity, preserves the geometric intuition underlying NLP constraint qualifications, and ensures that our regularity condition remains applicable even in degenerate or ill-conditioned scenarios.

We also reformulated the concept of pseudonormality for NSDP and clarified its role as a stronger regularity condition, closely related to the maximum-eigenvalue-based formulation. While this notion retains the favorable theoretical properties known from NLP and MPGC contexts, our analysis shows that such a strong condition is not required to obtain stability properties in NSDP. In particular, we proved that the existence of exact penalty formulations and local error bounds can already be guaranteed under the proposed quasinormality condition, thereby extending classical results from nonlinear programming to the semidefinite setting under significantly weaker assumptions.

Furthermore, we established that under the proposed (relaxed) quasinormality condition, dual sequences generated by standard and scaled augmented Lagrangian methods remain bounded, which is an important stability property related with a consistent stopping criterion, yielding primal-dual global convergence.

## References


[1] Andreani, R., Birgin, E.G., Martínez, J.M., Schuverdt, M.L.: On augmented lagrangian methods with general lower-level constraint. SIAM Journal on Optimization **18**(4), 1286–1309 (2007)

[2] Andreani, R., Haeser, G., Martínez, J.M.: On sequential optimality conditions for smooth constrained optimization. Optimization **60**(5), 627–641 (2011)

[3] Andreani, R., Haeser, G., Mito, L.M., Ramírez, H.: A minimal face constant rank constraint qualification for reducible conic programming. Mathematical Programming (2023)

[4] Andreani, R., Haeser, G., Mito, L.M., Ramírez, H.: Weak notions of nondegeneracy in nonlinear semidefinite programming. Mathematical Programming **205**, 1–32 (2024)

[5] Andreani, R., Haeser, G., Mito, L.M., Ramírez, H.: Sequential constant rank constraint qualifications for nonlinear semidefinite programming with applications. Set-Valued and Variational Analysis **31**(3), 555–589 (2023)

[6] Andreani, R., Haeser, G., Mito, L.M., Ramírez, H., Silveira, T.P.: First- and second-order optimality conditions for second-order cone and semidefinite programming under a constant rank condition. Mathematical Programming **202**(1–2), 473–513 (2023)

[7] Andreani, R., Haeser, G., Ramos, A., Santos, D.O., Secchin, L.D., Serranoni, A.: Strong global convergence properties of algorithms for nonlinear symmetric cone programming. Computational Optimization and Applications (2024)

[8] Andreani, R., Haeser, G., Ramos, A., Silva, P.J.S.: A sequential optimality condition and its consequences. SIAM Journal on Optimization **29**(1), 643–672 (2019)

[9] Andreani, R., Haeser, G., Schuverdt, M.L., Secchin, L.D.: A relaxed quasinormality condition and the boundedness of dual augmented lagrangian sequences. SIAM Journal on Optimization **35**(4), 2474–2489 (2025)

[10] Andreani, R., Haeser, G., Viana, D.S.: Optimality conditions and global convergence for nonlinear semidefinite programming. Mathematical Programming **180**(1), 203–235 (2020)

[11] Bertsekas, D.P., Ozdaglar, A.E.: Pseudonormality and a lagrange multiplier theory for constrained optimization. Journal of Optimization Theory and Applications **114**(2), 287–343 (2002)

[12] Birgin, E.G., Haeser, G., Martínez, J.M.: Safeguarded augmented lagrangian algorithms with scaled stopping criterion for the subproblems. Computational Optimization and Applications **91**, 491–509 (2026)

[13] Birgin, E.G., Martínez, J.M.: Practical Augmented Lagrangian Methods for Constrained Optimization. SIAM, Philadelphia (2014)

[14] Bonnans, J.F., Shapiro, A.: Perturbation Analysis of Optimization Problems. Springer Verlag, New York (2000)

[15] Ding, C., Sun, D., Ye, J.J.: First order optimality conditions for mathematical programs with semidefinite cone complementarity constraints. Mathematical Programming, Series A **147**(1–2), 539–579 (2014)

[16] Fares, B., Apkarian, P., Noll, D.: An augmented lagrangian method for a class of lmi-constrained problems in robust control theory. International Journal of Control **74**(4), 348–360 (2001)

[17] Fares, B., Noll, D., Apkarian, P.: Robust control via sequential semidefinite programming. SIAM Journal on Control and Optimization **40**(6), 1791–1820 (2002)

[18] Freund, R.W., Jarre, F., Vogelbusch, C.H.: Nonlinear semidefinite programming: sensitivity, convergence, and an application in passive reduced-order modeling. Mathematical Programming **109**, 581–611 (2007)

[19] Guo, L., Ye, J.J., Zhang, J.: Mathematical programs with geometric constraints in banach spaces: Enhanced optimality, exact penalty, and sensitivity. SIAM Journal on Optimization **23**(4), 2295–2319 (2013)

[20] Hestenes, M.R.: Optimization Theory: The Finite Dimensional Case. John Wiley & Sons, New York (1975)

[21] Kanno, Y., Takewaki, I.: Sequential semidefinite program for maximum robustness design of structures under load uncertainty. Journal of Optimization Theory and Applications **130**, 265–287 (2006)

[22] Konno, H., Kawadai, N., Wu, D.: Estimation of failure probability using semi-definite logit model. Computational Management Science **1**(1), 59–73 (2003)

[23] Lourenço, B.F., Fukuda, E.H., Fukushima, M.: Optimality conditions for nonlinear semidefinite programming via squared slack variables. Mathematical Programming **166**, 1–24 (2016)

[24] Luo, H.Z., Wu, H.X., Chen, G.T.: On the convergence of augmented lagrangian methods for nonlinear semidefinite programming. Journal of Global Optimization **54**(3), 599–618 (2012)

[25] Minchenko, L.I., Tarakanov, A.V.: On error bounds for quasinormal programs. Journal of Optimization Theory and Applications **148**(3), 571–579 (2011)

[26] Ozdaglar, A.E., Bertsekas, D.P.: The relation between pseudonormality and quasiregularity in constrained optimization. Optimization Methods and Software **19**(5), 493–506 (2004)

[27] Qi, H., Sun, D.: A quadratically convergent newton method for computing the nearest correlation matrix. SIAM Journal on Matrix Analysis and Applications **28**(2), 360–385 (2006)

[28] Rockafellar, R.T., Wets, R.J.B.: Variational Analysis. Springer-Verlag, Berlin (1998)

[29] Shapiro, A.: First and second order analysis of nonlinear semidefinite programs. SIAM Journal on Optimization **77**(1), 301–320 (1997)

[30] Shapiro, A., Sun, J.: Some properties of the augmented lagrangian in cone constrained optimization. Mathematics of Operations Research **29**(3), 479–491 (2004)

[31] Stingl, M., Kočvara, M., Leugering, G.: A sequential convex semidefinite programming algorithm with an application to multiple-load free material optimization. SIAM Journal on Optimization **20**(1), 130–155 (2009)

[32] Vandenberghe, L., Boyd, S., Wu, S.P.: Determinant maximization with linear matrix inequality constraints. SIAM Journal on Matrix Analysis and Applications **19**(2), 499–533 (1998)

[33] Yamashita, H., Yabe, H.: A survey of numerical methods for nonlinear semidefinite programming. Journal of the Operations Research Society of Japan **58**(1), 24–60 (2015)

[34] Ye, J., Zhang, D., Zhang, J.Z.: Enhanced karush-kuhn-tucker conditions and their applications. Mathematics of Operations Research **38**(4), 693–714 (2013)

[35] Zhu, Z.B., Zhu, H.L.: A filter method for nonlinear semidefinite programming with global convergence. Acta Mathematica Sinica **30**(10), 1810–1826 (2014)