Calmness of partial perturbation to composite rank constraint systems and its applications

Yitian Qian · Shaohua Pan · Yulan Liu

Received: 2 January 2022 / Accepted: 18 August 2022 / Published online: 3 October 2022
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2022

Abstract
This paper is concerned with the calmness of a partial perturbation to the composite rank constraint system, an intersection of the rank constraint set and a general closed set, which is shown to be equivalent to a local Lipschitz-type error bound and also a global Lipschitz-type error bound under a certain compactness. Based on its lifted formulation, we derive two criteria for identifying those closed sets such that the associated partial perturbation possesses the calmness, and provide a collection of examples to demonstrate that the criteria are satisfied by common nonnegative and positive semidefinite rank constraint sets. Then, we use the calmness of this perturbation to obtain several global exact penalties for rank constrained optimization problems, and a family of equivalent DC surrogates for rank regularized problems.

Keywords Composite rank constraint systems · Calmness · Error bound · Exact penalty

Mathematics Subject Classification 90C31 · 54C60 · 49K40

1 Introduction

Let \( \mathbb{R} \) represent \( \mathbb{R}^{n \times m} \) or \( \mathbb{S}^n \), where \( \mathbb{R}^{n \times m} \) and \( \mathbb{S}^n \) respectively denote the space of all \( n \times m \) \((n \leq m)\) real matrices and the space of all \( n \times n \) real symmetric matrices, equipped with the trace inner product \( \langle \cdot, \cdot \rangle \) and its induced Frobenius norm \( \| \cdot \|_F \). Fix any integer \( r \in \{1, 2, \ldots, n\} \). Consider the rank constrained optimization problem

\[
\min_{X \in \Omega} \left\{ f(X) \text{ s.t. } \text{rank}(X) \leq r \right\}
\]

1 School of Mathematics, South China University of Technology, Guangzhou, China
2 School of Mathematics and Statistics, Guangdong University of Technology, Guangzhou, China
where $\Omega \subset \mathbb{X}$ is a closed set and $f : \mathbb{X} \to (-\infty, \infty]$ is a proper lsc and lower bounded function with $\Omega \subseteq \text{dom } f$. This model is suitable for the scenario where $r$ is a tight upper estimation for the rank of the target matrix. If such $r$ is unavailable, one would prefer the rank regularized model

$$
\min_{X \in \Omega} \left\{ f(X) + \nu \text{rank}(X) \right\}
$$

(2)

where $\nu > 0$ is the regularization parameter. Throughout this paper, we assume that $\Gamma_r := \{ X \in \Omega \mid \text{rank}(X) \leq r \}$ is nonempty, and $f$ is coercive and locally Lipschitz continuous relative to the set $\Omega$. The coercive assumption of $f$ relative to $\Omega$ is very mild and guarantees that (1) has a nonempty global optimal solution set. When $f$ does not satisfy this assumption, one may consider replacing $f$ with $f + \frac{\nu}{2} \| \cdot \|^2_F$ for a tiny $\mu > 0$.

Models (1) and (2) have a host of applications in statistics [46], system identification and control [13], finance [49], machine learning [20, 30], and quantum tomography [19]. In particular, model (1) with $\mathbb{X} = S^n$ frequently arises from the positive semidefinite (PSD) relaxations for combinational and graph optimization problems (see, e.g., [9, 18, 21]). Note that $\text{rank}(X) \leq r$ if and only if $\| X \|_* - \| X \|_r = 0$, where $\| \cdot \|_*$ and $\| \cdot \|_r$ denote the nuclear norm and the Ky-Fan $r$-norm of matrices, respectively. We are interested in the calmness of the following perturbation to the composite rank constraint system $\Gamma_r$:

$$
S_r(\tau) := \{ X \in \Omega \mid \| X \|_* - \| X \|_r = \tau \} \quad \text{for } \tau \in \mathbb{R}.
$$

(3)

Clearly, $\text{dom } S_r := \{ \tau \in \mathbb{R} \mid S_r(\tau) \neq \emptyset \} \subseteq \mathbb{R}$, because $\| X \|_* \geq \| X \|_r$ for all $X \in \mathbb{X}$.

Motivated by the fact that the difference-of-convex (DC) algorithm has been extensively explored (see, e.g., [24, 32]), when the set $\Omega$ does not have a simple structure (say, the projection onto $\Omega$ has no closed form), it is natural to deal with problem (1) by penalizing the DC reformulation $\| X \|_* - \| X \|_r = 0$ of the rank constraint, and then develop effective algorithms for solving the obtained DC penalized problem

$$
\min_{X \in \Omega} \left\{ f(X) + \rho \| X \|_* - \| X \|_r \right\}
$$

(4)

or the factorized form of the penalized problem (4), where $\rho > 0$ is the penalty parameter. As far as we know, the idea to penalize the DC reformulation of the rank constraint first appeared in the technical report [14]. Recently, for the quadratic assignment problem, Jiang et al. [27] developed a proximal DC approach by the penalized problem (4) of its equivalent rank constrained doubly nonnegative reformulation; and for the unconstrained binary polynomial program, Qian and Pan [50] developed a relaxation approach by the factorized formulation of the penalized problem (4) of its equivalent PSD program. The encouraging numerical results in [27, 50] inspire us to explore the relation between global (or local) optimal solutions of the penalized problem (4) and those of the origin problem (1) for more closed sets $\Omega$. As will be shown in Sect. 4.1, the calmness of $S_r$ is the key to achieve the goal. This is a motivation for us to study the calmness of $S_r$ at 0.

Another motivation for studying the calmness of $S_r$ at 0 is to derive equivalent DC surrogates for the rank regularized problem (2). It is well known that nonconvex surrogate methods are more effective than the nuclear norm convex surrogate method (see, e.g., [6, 38, 42]). Take into account that the efficiency of some nonconvex surrogates, such as the Schatten $p$-norm [31, 42] and the log-determinant [12], depends on their approximation level to the rank function. The authors in [36] derived a class of equivalent DC surrogates by the uniformly partial calmness of the MPEC reformulation of (2), which includes the matrix version of the popular SCAD [11] and MCP [58] surrogates. However, the assumption there (see [36, Theorem 4.2]) is very restrictive on the set $\Omega$ and it may not hold even for a closed ball on...
the elementwise norm of matrices. Then, it is natural to ask if there is a practical criterion for identifying more classes of $\Omega$ to obtain such surrogates.

The last but not least one is to characterize the normal cone to the set $\Gamma_r$, which plays a significant role in deriving the optimality conditions of (1) (see [35]) and verifying the KL property of exponent $1/2$ for its extended objective function. Indeed, the two tasks involve the characterization on the normal cone to $\Gamma_r$. By [26, Sect. 3.1], the calmness of $S_r$ at 0 or the equivalent metric qualification is enough to achieve an upper inclusion for the normal cone to $\Gamma_r$ in terms of the normal cones to $\Omega$ and the rank constraint set.

The Aubin property (often under the name of the Lipschitz-like property) of $S_r$ at 0 for $X \in S_r(0)$ implies its calmness at 0 for $X \in S_r(0)$ and the Mordukhovich criterion [44] (see also [52, Theorem 9.40]) provides a convenient tool to identify the Aubin property. Unfortunately, it is impossible for $S_r$ to have the Aubin property at 0 because $S_r(\tau) = \emptyset$ for $\tau < 0$. In the past few decades, there have been a large number of research works on the calmness of a multifunction or equivalently the subregularity of its inverse mapping (see, e.g., [2, 8, 16, 22, 59]) and the closely related error bounds of a general lsc function (see, e.g., [10, 29, 41, 48, 54]). A collection of criteria have been proposed in these literatures for identifying the calmness of a multifunction, but most of them are neighborhood-type and to check if they hold or not is not an easy task for a specific $\Omega$. One contribution of this work is to present two practical criteria for identifying those closed $\Omega$ such that the associated perturbation $S_r$ is calm at 0 for any $X \in S_r(0)$; see Sect. 3.1. Although our criteria are stronger than those coming from the above works, they are point-type and as will be illustrated in Sect. 3.2, they hold for many common nonnegative and PSD composite rank constraint systems. Interestingly, the two criteria are precisely the linear regularity of $\Gamma_r$ when it is regarded as an intersection of the set $\Omega$ and the rank constraint set or an intersection of a closed set and the positive semidefinite rank constraint set. Linear regularity of collections of sets was earliest introduced in [43] as the generalized nonseparation property, and was recently employed in [33] to achieve the local linear convergence for alternating and averaged nonconvex projections. For more discussions on the linear regularity of collections of sets, refer to [28, 47].

As will be shown in Sect. 3.1, the calmness of $S_r$ at 0 for any $X \in S_r(0)$ is equivalent to a local Lipschitz-type error bound and also a global Lipschitz-type error bound under the compactness of $\Omega$. To the best of our knowledge, few works discuss the error bounds for rank constrained optimization problems except [5, 37]. In [5] the error bound was obtained for $\Gamma_r$ only involving three special $\Omega$ by constructing a feasible point technically, while in [37] the error bound was established only for the spectral norm unit ball $\Omega$. The two papers did not provide a criterion to identify the set $\Omega$ such that $\Gamma_r$ has this property.

The other contribution of this work is to apply the calmness of $S_r$ to establishing several classes of global exact penalties for the rank constrained problem (1), and deriving a family of equivalent DC surrogates for the rank regularized problem (2). For the former, we show that the penalized problem (4), the Schatten $p$-norm penalized problem, and the truncated difference penalized problem of $\| \cdot \|_*$ and $\| \cdot \|_F$ are all the global exact penalty for problem (1), which not only generalizes the exact penalty result of [37] to more types of $\Omega$, but also first verifies the exact penalization for the truncated difference of $\| \cdot \|_*$ and $\| \cdot \|_F$ introduced in [40]. For the latter, we greatly improve the result of [36, Theorem 4.2] by weakening the restriction there on the set $\Omega$; see Sect. 4.2.
2 Notation and preliminaries

Throughout this paper, \( \mathbb{R} \) represents the real number set, \( \mathbb{R}^{n \times m}_+ \) denotes the nonnegative matrix set of \( \mathbb{R}^{n \times m} \), and \( \mathbb{S}^n_+ \) denotes the set of all positive semidefinite matrices of \( \mathbb{S}^n \). For each integer \( k \geq 1 \), write \( [k] := \{1, 2, \ldots, k\} \). For each \( r \in [n] \), let \( \Lambda_r \) denote the rank constraint set \( R_r \) or \( S_r \), where \( R_r := \{ X \in \mathbb{R}^{n \times m} \ | \ \text{rank}(X) \leq r \} \), \( S_r := \{ X \in \mathbb{S}^n \ | \ \text{rank}(X) \leq r \} \), and write \( S^+_r := S_r \cap S^n_+ \). The notation \( \mathbb{O}^n \) represents the set of all \( n \times n \) matrices with orthonormal columns, and \( I \) and \( e \) denote an identity matrix and a vector of all ones, respectively, whose dimensions are known from the context. For a given \( X \in \mathbb{X} \), \( \lambda(X) = (\lambda_1(X), \ldots, \lambda_n(X))^\top \) and \( \sigma(X) = (\sigma_1(X), \ldots, \sigma_n(X))^\top \) denote the eigenvalue and singular value vectors of \( X \) arranged in a nonincreasing order. For \( X \in \mathbb{S}^n \), \( \mathbb{O}^n(X) := \{ P \in \mathbb{O}^n \ | \ X = P \text{Diag}(\lambda(X)) P^\top \} \). For a closed set \( C \subset \mathbb{X} \), \( \Pi_C \) denotes the projection mapping onto \( C \), and for a given \( X \in \mathbb{X} \), if \( \Pi_C(X) \) is non-unique, then \( \Pi_C(X) \) represents an arbitrary point of this set; \( \text{dist}(X, C) \) means the distance from \( X \) to the set \( C \) in terms of the Frobenius norm; and \( \delta_C \) denotes the indicator function of \( C \), i.e., \( \delta_C(x) = 0 \) if \( x \in C \) and \( \infty \) otherwise. The notation \( \mathbb{B}(\overline{X}, \delta) \) denotes a closed ball of radius \( \delta > 0 \) centered at \( \overline{X} \) with interior \( \mathbb{B}(\overline{X}, \delta) \), and \( \mathbb{B}_\mathbb{X} \) means the unit ball in \( \mathbb{X} \). For a linear mapping \( \mathcal{A} : \mathbb{X} \rightarrow \mathbb{R}^m \), the notation \( \mathcal{A}^* \) denotes its adjoint.

2.1 Calmness and subregularity

The notion of calmness of a multifunction was first introduced in [56] under the term “pseudo upper-Lipschitz continuity” owing to the fact that it is a combination of Aubin’s pseudo-Lipschitz continuity and Robinson’s upper-Lipschitz continuity [51], and the term “calmness” was later coined in [52]. Let \( \mathbb{Y} \) and \( \mathbb{Z} \) respectively represent a finite dimensional real vector space equipped with the inner product \( \langle \cdot, \cdot \rangle \) and its induced norm \( \| \cdot \| \). A multifunction \( \mathcal{M} : \mathbb{Y} \rightharpoonup \mathbb{Z} \) is said to be calm at \( \overline{y} \) for \( \overline{z} \in \mathcal{M}(\overline{y}) \) if there exists a constant \( \gamma \geq 0 \) together with \( \varepsilon > 0 \) and \( \delta > 0 \) such that for all \( y \in \mathbb{B}(\overline{y}, \varepsilon) \),

\[
\mathcal{M}(y) \cap \mathbb{B}(\overline{z}, \delta) \subseteq \mathcal{M}(\overline{y}) + \gamma \| y - \overline{y} \| \mathbb{B}_\mathbb{Z}. \tag{5}
\]

By [8, Exercise 3H.4], the neighborhood restriction \( \mathbb{B}(\overline{y}, \varepsilon) \) on \( y \) in (5) can be removed. As observed by Henrion and Outrata [22], the calmness of \( \mathcal{M} \) at \( \overline{y} \) for \( \overline{z} \in \mathcal{M}(\overline{y}) \) is equivalent to the (metric) subregularity of its inverse at \( \overline{z} \) for \( \overline{y} \in \mathcal{M}^{-1}(\overline{z}) \). Subregularity was introduced by Ioffe in [25] (under a different name) as a constraint qualification related to equality constraints in nonsmooth optimization problems, and was later extended to generalized equations. Recall that a multifunction \( \mathcal{F} : \mathbb{Z} \rightharpoonup \mathbb{Y} \) is called (metrically) subregular at \( \overline{z} \) for \( \overline{y} \in \mathcal{F}(\overline{z}) \) if there exist a constant \( \kappa \geq 0 \) along with \( \varepsilon > 0 \) such that

\[
\text{dist}(z, \mathcal{F}^{-1}(\overline{y})) \leq \kappa \text{dist}(\overline{y}, \mathcal{F}(z)) \quad \text{for all } z \in \mathbb{B}(\overline{z}, \varepsilon). \tag{6}
\]

The calmness and subregularity have already been studied by many authors under various names (see, e.g., [8, 16, 22, 23, 26, 59] and the references therein).
2.2 Normal and tangent cones

Next we recall from the monographs [45, 52] the normal and tangent cones to a closed set \( C \subseteq \mathbb{X} \). The Fréchet (regular) normal cone to \( C \) at \( \bar{x} \in C \) is defined as

\[
\hat{N}_C(\bar{x}) := \left\{ v \in \mathbb{X} \mid \limsup_{\bar{x} \neq x} \frac{\langle v, x - \bar{x} \rangle}{\| x - \bar{x} \|} \leq 0 \right\},
\]

and the limiting (also called Mordukhovich) normal cone to \( C \) at \( \bar{x} \) is defined by

\[
N_C(\bar{x}) := \left\{ v \in \mathbb{X} \mid \exists x^k \to \bar{x}, v^k \to v \text{ with } v^k \in \hat{N}_C(x^k) \text{ for all } k \right\}.
\]

The following lemmas provide the characterization on the normal cone to \( S_r \).

**Lemma 2.1** (see [7, Example 2.65]) Fix any \( X \in S^+_r \) with \( \text{rank}(X) = r \). Let \( X \) have the eigenvalue decomposition as \( P \text{Diag}(\lambda(X))P^\top \) with \( P \in \mathbb{O}^n(X) \) and let \( P_1 \) be the submatrix consisting of the first \( r \) columns of \( P \). Then, \( \hat{N}_{S^+_r}(X) = \left\{ W \in S^+_n \mid P_1^\top WP = 0 \right\} \).

**Lemma 2.2** (see [39, Proposition 3.6]) Fix any \( r \in [n] \). Consider any \( X \in \Lambda_r \) with the SVD as \( U \text{Diag}(\sigma(X))V^\top \). Let \( \beta := \{ i \in [n] \mid \sigma_i(X) = 0 \} \). If \( \text{rank}(X) = r \), then

\[
\begin{align*}
\hat{N}_{S_r}(X) &= N_{S_r}(X) = \left\{ U_\beta HU_\beta^\top \mid H \in \mathbb{S}^{[\beta]} \right\}, \\
\hat{N}_{R_r}(X) &= N_{R_r}(X) = \left\{ U_\beta HV_\beta^\top \mid H \in \mathbb{R}^{[\beta] \times [\beta]} \right\};
\end{align*}
\]

and if \( \text{rank}(X) < r \), then it holds that

\[
\begin{align*}
\hat{N}_{S_r}(X) \cap \hat{N}_{R_r}(X) &= \left\{ W \in \mathbb{S}^n \mid \text{rank}(W) \leq n - r \right\} \cap \left\{ U_\beta HU_\beta^\top \mid H \in \mathbb{S}^{[\beta]} \right\}, \\
\hat{N}_{R_r}(X) &= \left\{ W \in \mathbb{R}^{n \times m} \mid \text{rank}(W) \leq n - r \right\} \cap \left\{ U_\beta HV_\beta^\top \mid H \in \mathbb{R}^{[\beta] \times [\beta]} \right\}.
\end{align*}
\]

By Lemmas 2.1 and 2.2, when \( X \in S^+_r \) with \( \text{rank}(X) \leq r \), it is easy to check that every \( W \in \hat{N}_{S^+_r}(X) \) satisfies \( W = U_\beta U_\beta^\top W U_\beta U_\beta^\top \), which implies that \( \hat{N}_{S^+_r}(X) \subseteq N_{S_r}(X) \) and consequently \( \hat{N}_{S^+_r}(X) = N_{S_r}(X) \). The following lemma provides a characterization on the normal cone to the composite set \( S^+_r \).

**Lemma 2.3** Fix any \( r \in [n] \) and \( X \in S^+_r \). When \( \text{rank}(X) = r \), \( \hat{N}_{S^+_r}(X) = N_{S^+_r}(X) = N_{S_r}(X) \); when \( \text{rank}(X) < r \), \( N_{S^+_r}(X) \subseteq \hat{N}_{S^+_r}(X) \subseteq N_{S^+_r}(X) \subseteq N_{S^+_r}N_{S^+_r}(X) \).

**Proof** Note that for any \( Y \in S^+_r \), \( \text{dist}(Y, S^+_r) = \text{dist}(Y, S_r) \). Hence, for any \( Z \in \mathbb{S}^n \),

\[
\begin{align*}
\text{dist}(Z, S^+_r) &\leq \| Z - \Pi_{S^+_r}(Z) \|_F + \text{dist}(\Pi_{S^+_r}(Z), S^+_r) \\
&= \| Z - \Pi_{S^+_r}(Z) \|_F + \text{dist}(\Pi_{S^+_r}(Z), S_r) \\
&\leq 2\text{dist}(Z, S^+_r) + \text{dist}(Z, S_r) \tag{9}
\end{align*}
\]

By [26, Sect. 3.1], \( \hat{N}_{S^+_r}(X) \subseteq N_{S^+_r}(X) + N_{S_r}(X) \). Along with [52, Theorem 6.42],

\[
N_{S^+_r}(X) + \hat{N}_{S_r}(X) \subseteq \hat{N}_{S^+_r}(X) \subseteq N_{S^+_r}(X) \subseteq N_{S^+_r}(X) + N_{S_r}(X) \tag{10}
\]

When \( \text{rank}(X) = r \), since \( \hat{N}_{S_r}(X) = N_{S_r}(X) \) and \( N_{S^+_r}(X) + N_{S_r}(X) = N_{S_r}(X) \), the last inclusions become the desired equalities. When \( \text{rank}(X) < r \), since \( \hat{N}_{S_r}(X) = \{ 0 \} \), we have \( N_{S^+_r}(X) + \hat{N}_{S_r}(X) = N_{S^+_r}(X) \), which by (10) yields the desired inclusions. \( \square \)
Next we recall from [17, 57] the directional version of limiting normal cone to a set.

**Definition 2.1** Given a set $C \subseteq \mathbb{X}$, a point $z \in C$ and a direction $d \in \mathbb{X}$, the limiting normal cone to $C$ in direction $d$ at $z$ is defined by

$$N_C(z; d) := \left\{ v \in \mathbb{X} \mid \exists t_k \downarrow 0, d^k \to d, v^k \to v \text{ with } v^k \in \tilde{N}_C(z + t_k d^k) \right\},$$

and the inner limiting normal cone to $C$ in direction $d$ at $z$ is defined by

$$\tilde{N}_C^i(z; d) := \left\{ v \in \mathbb{X} \mid \forall t_k \downarrow 0, d^k \to d, v^k \to v \text{ with } v^k \in \tilde{N}_C(z + t_k d^k) \right\}.$$

By Definition 2.1, it is obvious that $N_C(z; d) = \emptyset$ if $d \notin T_C(z)$, $N_C(z; d) \subseteq \tilde{N}_C(z)$, and $\tilde{N}_C(z; 0) = N_C(z)$. When $C$ is convex and $d \in T_C(z)$, $N_C(z; d) = \tilde{N}_C(z; d)$.

**Proposition 2.1** Fix any $r \in [n]$ and any $X \in S^+_r$ with rank$(X) = r$. Then,

$$N^i_S(X; H) = \tilde{N}^i_S(X; H) = N^i_S(X) \text{ for all } H \in T^+_S(X).$$

In particular, for any $X \in \Lambda_r$ with rank$(X) = r$, it also holds that

$$N^i_R_r(X; H) = \tilde{N}^i_R_r(X; H) = N^i_R_r(X) \text{ for all } H \in T^+_R_r(X).$$

**Proof** Fix any $0 \neq H \in T^+_S(X)$. Then $N^i_S(X; H) \subseteq N^i_S(X; H) \subseteq \tilde{N}^i_S(X; H)$. Pick any $W \in N^i_S(X)$. Let $X$ have the eigenvalue decomposition as $P \text{Diag}(\lambda(X)) P^\top$ and $\beta := \{ i \in [n] \mid \lambda_i(X) = 0 \}$. By Lemma 2.2, $W = P_\beta P^\top_\beta W P_\beta P^\top_\beta$. Since $S^+_r$ is Clarke regular at $X$ by Lemma 2.2, from $H \in T^+_S(X)$ it follows that for any $t_k \downarrow 0$, there exists a sequence $\{X^k\} \subseteq S^+_r$ with $X^k = X + t_k (H + \frac{\alpha(t_k)}{t_k})$. Since rank$(X) = r$ and $X^k \to X$, there exists $\tilde{k} \in \mathbb{N}$ such that rank$(X^k) \geq r$ for all $k \geq \tilde{k}$. Along with $\{X^k\} \subseteq S^+_r$, we have rank$(X^k) = r$ for each $k \geq \tilde{k}$, which implies that $\lambda_i(X^k) \neq 0$ for $i \notin \beta$ and $\lambda_i(X^k) = 0$ for $i \in \beta$ when $k$ is large enough. For each $k$, let $W^k = P^\top_\beta (P^\top_\beta W P_\beta) P^\top_\beta$ with $P^\top_\beta \in \mathbb{O}^n(X^k)$. By Lemma 2.2, we have $W^k \in \tilde{N}^i_S(X^k)$ for all $k$ large enough. Since the sequence $\{P^\top_\beta\}$ is bounded, we may assume that (if necessary taking a subsequence) that $P^\top_k \to \tilde{P}$. Clearly, $\tilde{P} \in \mathbb{O}^n(X)$ and $W^k \to \tilde{W} := \tilde{P} \tilde{P}^\top \tilde{W} \tilde{P} \tilde{P}^\top \tilde{P}$. Let $\mu_1 > \mu_2 > \cdots > \mu_l$ be the distinct eigenvalues of $X$ and $a_k := \{ i \mid \lambda_i(X) = \mu_k \}$ for $k = 1, 2, \ldots, l$. Since $\tilde{P} \in \mathbb{O}^n(X)$ and $P \in \mathbb{O}^n(X)$, there exists $Q = \text{BlkDiag}(Q_1, \ldots, Q_l)$ with $Q_k \in \mathbb{O}^{n_k} \text{ for } k = 1, 2, \ldots, l$ such that $\tilde{P} = PQ$, which implies that $\tilde{P} \tilde{P}^\top = P \tilde{P} \tilde{P}^\top$. Thus, $\tilde{W} = W$. By Definition 2.1, we conclude that $W \in N^i_S(X; H)$. Using the same arguments, we obtain the second part. \hfill \Box

When rank$(X) < r$, for every $H \in T^+_S(X)$, a tighter upper estimation than $N^i_S(X; H)$ for $N^i_S(X; H)$ cannot be achieved because the exact expression of $\tilde{N}^i_S(X)$ for $Z \in S^+_r$ with rank$(Z) < r$ is unavailable. Such a difficulty also appears in sparsity constraint sets.

### 3 Calmness of $S^r$ and examples

In this section we establish the calmness of the mapping $S^r$ under a regularity condition, and illustrate that this condition can be satisfied via a collection of common examples.
3.1 Calmness of $S_r$

First, we achieve the calmness of $S_r$ at 0 for any $\overline{X} \in S_r(0)$ or equivalently a Lipschitz-type local error bound for the set $\Gamma_r$ at any $\overline{X} \in \Gamma_r$, under a condition coming from the metric regularity of a lifted formulation of $S_r$ at $(\overline{X}, \overline{X})$ for the origin.

**Theorem 3.1** Consider any $\overline{X} \in \Gamma_r$. The mapping $S_r$ is calm at 0 for $\overline{X}$ if and only if either of the following equivalent conditions holds:

(i) there exists a constant $\gamma \geq 0$ along with $\delta > 0$ such that for all $X \in \mathbb{B}(\overline{X}, \delta)$,

$$\text{dist}(X, \Gamma_r) \leq \gamma \left[ \text{dist}(X, \Omega) + \text{dist}(X, \Lambda_r) \right];$$

(ii) $\mathcal{F}_r(X, Y) := \begin{cases} \{X - Y\} \text{ if } (X, Y) \in \Omega \times \Lambda_r \\
\emptyset \text{ otherwise} \end{cases}$ is subregular at $(\overline{X}, \overline{X})$ for the origin.

Consequently, the calmness of $S_r$ at 0 for any $\overline{X} \in \Gamma_r$ is implied by the following condition

$$[- \mathcal{N}_\Omega(\overline{X})] \cap \mathcal{N}_{\Lambda_r}(\overline{X}) = \{0\}. \quad (12)$$

**Proof** By the definition, the calmness of the mapping $S_r$ at 0 for $\overline{X}$ is equivalent to the existence of $\gamma' \geq 0$ and $\delta' > 0$ such that for all $Z \in \mathbb{B}(\overline{X}, \delta')$,

$$\text{dist}(Z, S_r(0)) \leq \gamma' \text{dist}(0, S_r^{-1}(Z)) \leq \begin{cases} \gamma' ||Z||_* - ||Z||_{(r)} \text{ if } Z \in \Omega, \\
\infty \text{ otherwise} \end{cases}$$

(13)

where the equality is due to the definition of $S_r$ and the fact that dom$S_r = \mathbb{R}_+^r$.

(i) If there exist $\gamma \geq 0$ and $\delta > 0$ such that inequality (11) holds for all $X \in \mathbb{B}(\overline{X}, \delta)$, then inequality (13) obviously holds, and the mapping $S_r$ is calm at 0 for $\overline{X}$. Now assume that $S_r$ is calm at 0 for $\overline{X}$, i.e., there exist $\gamma' \geq 0$ and $\delta' > 0$ such that inequality (13) holds for all $X \in \mathbb{B}(\overline{X}, \delta')$. We will show that inequality (11) holds with $\delta = \delta'/2$ and $\gamma = 1 + \sqrt{n}\gamma'$. Pick any $X \in \mathbb{B}(\overline{X}, \delta)$. If $X \not\in \Omega$, inequality (11) holds with $\gamma = \sqrt{n}\gamma'$ by (13). If $X \in \Omega$, since $||\Pi_\Omega(X) - \overline{X}||_F \leq 2||X - \overline{X}||_F \leq \delta'$, from (13) we have

$$\text{dist}(X, \Gamma_r) \leq ||X - \Pi_\Omega(X)||_F + \text{dist}(\Pi_\Omega(X), S_r(0)) \leq \text{dist}(X, \Omega) + \gamma'(\sum_{r=1}^n \sigma_i(\Pi_\Omega(X)))$$

$$= \text{dist}(X, \Omega) + \gamma' \min_{Z \in \Lambda_r} ||Z - \Pi_\Omega(X)||_*$$

$$\leq \text{dist}(X, \Omega) + \gamma'||X - \Pi_\Omega(X)||_* + \gamma' \min_{Z \in \Lambda_r} ||Z - X||_*$$

$$\leq \text{dist}(X, \Omega) + \sqrt{n}\gamma'||X - \Pi_\Omega(X)||_F + \gamma'(\sum_{r=1}^n \sigma_i(\Pi_\Omega(X)))$$

$$\leq (1 + \sqrt{n}\gamma')\text{dist}(X, \Omega) + \gamma'(\sum_{r=1}^n \sigma_i(\Pi_\Omega(X)))$$

where the equality is are by Lemma 1 in Appendix. Then (11) holds with $\gamma = 1 + \sqrt{n}\gamma'$.

(ii) It suffices to argue that $\mathcal{F}_r$ is subregular at $(\overline{X}, \overline{X})$ for the origin iff part (i) holds. 

$\implies$. Since the mapping $\mathcal{F}_r$ is subregular at $(\overline{X}, \overline{X})$ for the origin, there exists a constant $\kappa \geq 0$ along with $\epsilon > 0$ such that for all $(X, Y) \in \mathbb{B}((\overline{X}, \overline{X}), \epsilon) \cap (\Omega \times \Lambda_r)$,

$$\text{dist}((X, Y), \mathcal{F}_r^{-1}(0)) \leq \kappa \text{dist}(0, \mathcal{F}_r(X, Y)). \quad (14)$$

Pick any $X \in \mathbb{B}(\overline{X}, \epsilon/4)$. Obviously, $(X, X) \in \mathbb{B}((\overline{X}, \overline{X}), \epsilon)$. If $X \in \Omega \cap \Lambda_r = \Gamma_r$, part (i) automatically holds. If $X \in \Omega \cap \Lambda_r$, by noting that $||\Pi_{\Lambda_r}(X) - \overline{X}||_F \leq 2||X - \overline{X}||_F \leq \epsilon/2$, we have $(X, \Pi_{\Lambda_r}(X)) \in \mathbb{B}((\overline{X}, \overline{X}), \epsilon) \cap (\Omega \times \Lambda_r)$, and from (14) it follows that

$$\text{dist}(X, \Gamma_r) \leq \text{dist}((X, X), \mathcal{F}_r^{-1}(0)) \leq \text{dist}((X, \Pi_{\Lambda_r}(X)), \mathcal{F}_r^{-1}(0)) + ||\Pi_{\Lambda_r}(X) - X||_F$$

$$\leq \kappa \text{dist}(0, \mathcal{F}_r(X, \Pi_{\Lambda_r}(X))) + \text{dist}(X, \Lambda_r) = (1 + \kappa)\text{dist}(X, \Lambda_r)$$
where the first inequality is due to \( \Gamma_r \times \Gamma_r \supseteq \mathcal{F}_r^{-1}(0) \). If \( X \in \Lambda_r \setminus \Gamma_r \), using the similar arguments yields that \( \text{dist}(X, \Gamma_r) \leq (1 + \kappa)\text{dist}(X, \Omega) \). Finally, we consider the case that \( X \not\in \Omega \cup \Lambda_r \). Since \( \|\Pi \Omega(X) - \mathbf{X}\|_{F} \leq \varepsilon / 2 \) and \( \|\Pi \Lambda_r(X) - \mathbf{X}\|_{F} \leq \varepsilon / 2 \), we have \( (\Pi \Omega(X), \Pi \Lambda_r(X)) \in B(\mathbf{X}, \varepsilon) \cap (\Omega \times \Lambda_r) \), which along with (14) implies that

\[
\text{dist}(X, \Gamma_r) \leq \text{dist}((X, X), \mathcal{F}_r^{-1}(0)) \\
\leq \text{dist}((\Pi \Omega(X), \Pi \Lambda_r(X)), \mathcal{F}_r^{-1}(0)) + \|(\Pi \Omega(X), \Pi \Lambda_r(X)) - (X, X)\|_{F} \\
\leq \kappa \text{dist}(0, \mathcal{F}_r(\Pi \Omega(X), \Pi \Lambda_r(X))) + \text{dist}(X, \Omega) + \text{dist}(X, \Lambda_r) \\
\leq (1 + \kappa)\left[\text{dist}(X, \Omega) + \text{dist}(X, \Lambda_r)\right].
\]

The above arguments for the four cases demonstrate that part (i) holds.

\[\iff\]

Since part (i) holds, there exist \( r \geq 0 \) and \( \delta > 0 \) such that inequality (11) holds for all \( Z \in B(\mathbf{X}, \delta) \). Fix any \((X, Y) \in B((\mathbf{X}, \mathbf{X}), \delta) \cap (\Omega \times \Lambda_r)\). Since \( X \in B(\mathbf{X}, \delta) \cap \Omega \) and \( Y \in B(\mathbf{X}, \delta) \cap \Lambda_r \), from inequality (11) it immediately follows that

\[
\max\{\text{dist}(X, \Gamma_r), \text{dist}(Y, \Gamma_r)\} \leq \gamma\left[\text{dist}(Y, \Omega) + \sum_{i=r+1}^{n} \sigma_i(X)\right].
\]

Observe that \( \text{dist}((X, Y), \mathcal{F}_r^{-1}(0)) \leq \min_{Z \in \Gamma_r} \text{dist}(Z, \mathcal{F}_r^{-1}(0)) \leq \text{dist}(X, \Omega) + \text{dist}(Y, \Omega) \leq 2\gamma\left[\text{dist}(Y, \Omega) + \sum_{i=r+1}^{n} \sigma_i(X)\right].
\]

This shows that the mapping \( \mathcal{F}_r \) is metrically subregular at \((\mathbf{X}, \mathbf{X})\) for the origin.

From the equivalence between (i) and (ii), the local error bound in part (i) is implied by the regularity of \( \mathcal{F}_r \) at \((\mathbf{X}, \mathbf{X})\) for the origin or the Aubin property of its inverse at the origin for \((\mathbf{X}, \mathbf{X})\). The latter is equivalent to \([-N_{\Omega}(\mathbf{X})] \cap N_{\Lambda_r}(\mathbf{X}) = \{0\}\). Indeed, since \( gph\mathcal{F}_r^{-1} = \mathcal{L}^{-1}(\Omega \times \Lambda_r \times \{0\}) \) with \( \mathcal{L}(G, X, Y) := (X; Y; G - X + Y) \) for \( G, X, Y \in \mathbb{X} \), from the surjectivity of the mapping \( \mathcal{L} \) and [52, Exercise 6.7 & Proposition 6.41],

\[
N_{gph, \mathcal{F}_r^{-1}}(0, \mathbf{X}, \mathbf{X}) = \mathcal{L}^*[N_{\Omega}(\mathbf{X}) \times N_{\Lambda_r}(\mathbf{X}) \times \mathbb{X}] \\
= \left\{(\Delta W, \Delta S - \Delta W, \Delta Z + \Delta W) \mid \Delta S \in N_{\Omega}(\mathbf{X}), \Delta Z \in N_{\Lambda_r}(\mathbf{X})\right\}.
\]

From [44, Proposition 3.5] or [52, Theorem 9.40] it follows that the mapping \( \mathcal{F}_r \) has the Aubin property at the origin for \( \mathbf{X} \) iff \( D^\circ \mathcal{F}_r((0, 0)\mathbf{X}) = \{0, 0\}\), or equivalently

\[
(\Delta G, 0, 0) \in N_{gph, \mathcal{F}_r^{-1}}(0, \mathbf{X}, \mathbf{X}) \implies \Delta G = 0,
\]

where \( D^\circ \mathcal{F}_r((0, 0)\mathbf{X}) \) is the coderivative mapping of \( \mathcal{F}_r \) at \((0, 0, \mathbf{X})\) (see [45, Sect. 1.2.2]). This by (15) is equivalent to saying that \([-N_{\Omega}(\mathbf{X})] \cap N_{\Lambda_r}(\mathbf{X}) = \{0\}\). \( \square \)

When \( \Omega = \mathbf{S}_{\beta}^n \cap \Xi \) for a closed set \( \Xi \subset \mathbb{S}^n \), condition (12) does not hold because, by letting \( \mathbf{X} \) have the eigenvalue decomposition as \( \mathbf{X} = \mathbf{P}_{\Omega}(\lambda(\mathbf{X}))\mathbf{P}_{\Omega}^T \) and taking \( \mathbf{Z} = \mathbf{P}_{\beta_1} \mathbf{P}_{\beta_1}^T \) with \( \beta_1 \subset \beta := \{i \in [n] \mid \lambda_i(\mathbf{X}) = 0\} \) for \( |\beta_1| \leq n - r \), from Lemma 2.1 and 2.2 we have \( \mathbf{Z} \in [-N_{\Omega}(\mathbf{X})] \cap N_{S_{\beta}}(\mathbf{X}) \), which together with \( N_{\Omega}(\mathbf{X}) \supseteq \overline{N}_{\Xi}(\mathbf{X}) + N_{S_{\beta}}(\mathbf{X}) \) means that \( 0 \neq \mathbf{Z} \in [-N_{\Omega}(\mathbf{X})] \cap N_{S_{\beta}}(\mathbf{X}) \). The reason is that the separation of \( \mathbf{S}_{\beta}^n \) from \( S_{\beta} \) makes it difficult to hold by recalling that \( N_{S_{\beta}}(X) \subseteq N_{S_{\beta}}(X) \). Inspired by this, for this class of \( \Omega \), we achieve the calmness of \( S_{\beta} \) at \( 0 \) for \( \mathbf{X} \in \Gamma_r \) by combining \( \mathbf{S}_{\beta}^n \) and \( S_{\beta} \).

**Theorem 3.2** Let \( \Omega = \mathbf{S}_{\beta}^n \cap \Xi \) for a closed set \( \Xi \subset \mathbb{S}^n \). Consider any \( \mathbf{X} \in \Gamma_r \). The mapping \( S_{\beta} \) is calm at \( 0 \) for \( \mathbf{X} \) under either of the equivalent conditions:

\( \square \)

**Springer**
(i) there exists a constant $\beta \geq 0$ along with $\varepsilon > 0$ such that for all $X \in \mathbb{B}(\overline{X}, \varepsilon)$,
\[
\text{dist}(X, \Gamma_r) \leq \beta \left( \text{dist}(X, \Xi) + \text{dist}(X, S_r^+) \right);
\]
(ii) $G_r(X, Y) := \begin{cases} 
\{X - Y\} & \text{if } (X, Y) \in \Xi \times S_r^+ \\
\emptyset & \text{otherwise}
\end{cases}$ is subregular at $(\overline{X}, \overline{X})$ for the origin;

and the calmness of $S_r$ at $0$ for $\overline{X}$ is equivalent to either of conditions (i) and (ii) if in addition there exists a constant $\kappa' > 0$ along with $\varepsilon' > 0$ such that for all $X \in \mathbb{B}(\overline{X}, \varepsilon')$
\[
\text{dist}(X, \Omega) \leq \kappa' \left( \text{dist}(X, \Xi) + \text{dist}(X, S_r^+) \right).
\]

Consequently, the calmness of $S_r$ at $0$ for any $\overline{X} \in \Gamma_r$ is implied by the following condition
\[
[ - \mathcal{N}_\Xi(\overline{X})] \cap \mathcal{N}_{S_r^+} (\overline{X}) = \{0\}. \tag{18}
\]

**Proof** Pick any $X \in \mathbb{B}(\overline{X}, \varepsilon)$. By combining inequality (16) with (9), it follows that
\[
\text{dist}(X, \Gamma_r) \leq \beta \left[ \text{dist}(X, \Xi) + 2 \text{dist}(X, S_r^n) + \sum_{i=r+1}^{n} \sigma_i(X) \right] \\
\leq \beta \left[ 3 \text{dist}(X, \Omega) + \sum_{i=r+1}^{n} \sigma_i(X) \right].
\]

Hence, part (i) of Theorem 3.1 holds with $\delta = \varepsilon$ and $\gamma = 3\beta$, and $S_r$ is calm at $0$ for $\overline{X}$. By following the same arguments as those for part (ii) of Theorem 3.1, it is not hard to verify that $G_r$ is subregular at $(\overline{X}, \overline{X})$ for the origin if and only if part (i) holds.

Next under inequality (17) we argue that the calmness of $S_r$ at $0$ for $\overline{X}$ implies part (i). Indeed, since $S_r$ is calm at $0$ for $\overline{X}$, there exist $\gamma' \geq 0$ and $\delta' > 0$ such that (13) holds for all $X \in \mathbb{B}(\overline{X}, \delta')$. Let $\varepsilon = \min(\delta', \varepsilon')$ and pick any $X \in \mathbb{B}(\overline{X}, \varepsilon)$. By following the same arguments as those for part (i) of Theorem 3.1, we have
\[
\text{dist}(X, \Gamma_r) \leq (1 + \sqrt{n}\gamma') \text{dist}(X, \Omega) + \gamma' \sum_{i=r+1}^{n} \sigma_i(X) \\
\leq (1 + \sqrt{n}\gamma') \kappa' \left( \text{dist}(X, \Xi) + \text{dist}(X, S_r^n) \right) + \gamma' \sum_{i=r+1}^{n} \sigma_i(X) \\
\leq (1 + \sqrt{n}\gamma') \kappa' \left( \text{dist}(X, \Xi) + \text{dist}(X, S_r^+) \right) + \sqrt{n}\gamma' \text{dist}(X, S_r^+).
\]

This means that part (i) holds with $\beta = (1 + \sqrt{n}\gamma')(1 + \kappa')$ and $\varepsilon = \min(\delta', \varepsilon')$.

From the equivalence between (i) and (ii), the local error bound in part (i) is implied by the metric regularity of $G_r$ at $(\overline{X}, \overline{X})$ for the origin or the Aubin property of its inverse at the origin for $(\overline{X}, \overline{X})$. The latter is equivalent to $[-\mathcal{N}_\Xi(\overline{X})] \cap \mathcal{N}_{S_r^+} (\overline{X}) = \{0\}$ by following the similar arguments as those for the last part of Theorem 3.1. \hfill \Box

**Remark 3.1** (a) Condition (17) is equivalent to the calmness at $\overline{X}$ of the mapping
\[
\mathcal{M}(X, Y) := \{ W \in S_r^n \mid X + W \in \Xi, Y + W \in S_r^n \} \quad \text{for } X, Y \in S_r^n.
\]

When the set $\Xi$ is convex, from [3, Corollary 3] the condition $\text{ri}(\Xi) \cap \text{int}(S_r^n) \neq \emptyset$ is enough for condition (17) to hold. Clearly, there are many classes of closed convex sets $\Xi$ to satisfy this constraint qualification. When the set $\Xi$ is nonconvex, by noting that the Aubin property of $\mathcal{M}$ at $(\overline{X}, \overline{X})$ for the origin is equivalent to $[-\mathcal{N}_\Xi(\overline{X})] \cap \mathcal{N}_{S_r^+} (\overline{X}) = \{0\}$. So, in this case, $[-\mathcal{N}_\Xi(\overline{X})] \cap \mathcal{N}_{S_r^+} (\overline{X}) = \{0\}$ is enough for condition (17) to hold.

(b) The conditions in (12) and (18) are pointed, i.e., they depends only on the reference point. As will be illustrated in Sect. 3.2, by using the characterization on $\mathcal{N}_{\lambda_r}(\overline{X})$ and $\mathcal{N}_{S_r^+} (\overline{X})$, it is convenient to check if they hold or not. Although many weaker conditions are available to guarantee the calmness of $S_r$ at $0$ for any $\overline{X} \in S_r(0)$ (see, e.g., [10, 29, 41, 48]), they are all neighborhood-type and hard to check in practice. Gfrerer [16] proposed a point-type
criterion to identify the subregularity of a mapping, but as demonstrated below his criterion is only applicable to those $\bar{X}$ with rank($\bar{X}$) = $r$. By [16, Proposition 3.8], the mapping $G_r$ is subregular at ($\bar{X}$, $\bar{X}$) if $(0, 0, 0) \notin Cr_0G_r(\bar{X}, \bar{X}, 0)$ where

$$
Cr_0G_r(\bar{X}, \bar{X}, 0) := \{(F, S, T) \in S^n \times S^n \times S^n \mid \exists (G^k, H^k) \in S_{S^n} \times S_{S^n}, R^k \in S_{S^n}, (F^k, S^k, T^k) \to (F, S, T), t_k \downarrow 0,
\quad (-S^k, -T^k, R^k) \in \tilde{N}_{gphG_r}((\bar{X}, \bar{X}) + t_k(G^k, H^k), t_k F^k)\},
$$

where $S_{S^n} \times S_{S^n}$ and $S_{S^n}$ respectively denote the unit sphere in the space $S^n \times S^n$ and $S^n$. By Definition 2.1, it is not hard to verify that Gfrerer’s criterion is equivalent to

$$(0, 0, W) \notin N_{gphG_r}((\bar{X}, \bar{X}, 0); (G, H, 0)) \quad \text{for all} \quad ((G, H), W) \in S_{S^n} \times S_{S^n} \times S_{S^n},$$

but unfortunately the exact characterization for the directional normal cone to $gphG_r$ is unavailable. By [4, Theorem 3.1] and [57, Proposition 3.3], if rank($\bar{X}$) = $r$ or the set $\Xi$ is convex, one may obtain a verifiable but stronger version of Gfrerer’s criterion

$$(W, -W) \notin N_{\Xi}(\bar{X}; G) \times N_{\Xi^+}(\bar{X}; H) \quad \text{for all} \quad W \neq 0, (G, H) \in [T_{\Xi}(\bar{X}) \times T_{\Xi^+}(\bar{X})]\{0, 0\}.$$  

Similarly, by applying Gfrerer’s criterion to the mapping $F_r$, if rank($\bar{X}$) = $r$ or the set $\Omega$ is convex, one may obtain a verifiable but stronger version of Gfrerer’s criterion

$$(W, -W) \notin N_{\Omega}(\bar{X}; G) \times N_{\Lambda}(\bar{X}; H) \quad \text{for all} \quad W \neq 0, (G, H) \in [T_{\Omega}(\bar{X}) \times T_{\Lambda}(\bar{X})]\{0, 0\}.$$  

Recall that $N_{\Xi}(\bar{X}; G) = N_{T_{\Xi}(\bar{X})}(G)$ if $\Xi$ is convex. When rank($\bar{X}$) = $r$ and the set $\Xi$ or $\Omega$ is convex, by Proposition 2.1, the above two conditions are respectively equivalent to

$$\begin{align*}
\{ -N_{T_{\Xi}(\bar{X})}(G) \} \cap N_{\Xi^+}(\bar{X}) = \{0\} \quad &\text{for all} \quad G \in T_{\Xi}(\bar{X}), \quad (19a) \\
\{ -N_{T_{\Omega}(\bar{X})}(G) \} \cap N_{\Lambda}(\bar{X}) = \{0\} \quad &\text{for all} \quad G \in T_{\Omega}(\bar{X}). \quad (19b)
\end{align*}$$

When $\Xi$ or $\Omega$ is not an affine set, it is possible for (19a) or (19b) to be weaker than the criterion (18) or (12), but the former is only applicable to those $\bar{X}$ with rank($\bar{X}$) = $r$.

To close this part, we show that under the compactness of the set $\Omega$, the calmness of the mapping $S_r$ at 0 for all $X \in S_r(0)$ is equivalent to a global error bound for $\Gamma_r$.

**Theorem 3.3** Let $\Delta \subseteq \mathbb{X}$ be a compact set. If the mapping $S_r$ is calm at 0 for any $X \in S_r(0)$, then there exists a constant $\kappa' > 0$ such that for all $X \in \Delta \cap \Omega$

$$\text{dist}(X, \Gamma_r) \leq \kappa' \|X\|_* - \|X\|_{(r)}.$$

**Proof** Since the mapping $S_r$ is calm at 0 for all $X \in \Gamma_r$, for every $X \in \Gamma_r$ there exist $\kappa_X \geq 0$ and $\varepsilon_X > 0$ such that for all $Z \in \Omega \cap \mathbb{B}(X, \varepsilon_X)$,

$$\text{dist}(Z, \Gamma_r) \leq \kappa_X \|Z\|_* - \|Z\|_{(r)}.$$

Since $\bigcup_{X \in \Gamma_r \cap \Delta} \mathbb{B}(X, \varepsilon_X)$ is an open covering of the compact set $\Gamma_r \cap \Delta$, by Heine-Borel covering theorem, there exist $X^1, \ldots, X^p \in \Gamma_r \cap \Delta$ such that $\Gamma_r \cap \Delta \subseteq \bigcup_{i=1}^p \mathbb{B}(X^i, \varepsilon_{X^i})$. Write $\tilde{\kappa} := \max_{1 \leq i \leq p} \kappa_{X^i}$. From the last inequality, it then follows that

$$\text{dist}(Z, \Gamma_r) \leq \tilde{\kappa} \|Z\|_* - \|Z\|_{(r)} \quad \text{for all} \quad Z \in \bigcup_{i=1}^p \left[\Omega \cap \Delta \cap \mathbb{B}(X^i, \varepsilon_{X^i})\right].$$

Let $D = \bigcup_{i=1}^p \left[\Omega \cap \Omega \cap \mathbb{B}(X^i, \varepsilon_{X^i})\right]$. Consider the set $\tilde{\Omega} = cl[\Omega \cap \Delta \cap D]$. Then, there exists a constant $\tilde{\kappa} > 0$ such that $\min_{Z \in \tilde{\Omega}} \|Z\|_* - \|Z\|_{(r)} \geq \tilde{\kappa}$. If not, there exists a sequence
{Z^k} \subseteq \tilde{\Omega} such that \|Z^k\|_* - \|Z^k\|_{(r)} \leq 1/k, which by the compactness of the set \tilde{\Omega} and the continuity of the function Z \mapsto \|Z\|_* - \|Z\|_{(r)} means that there is a cluster point, say \( \tilde{Z} \in \tilde{\Omega} \), of \( \{Z^k\} \) such that \|\tilde{Z}\|_* - \|\tilde{Z}\|_{(r)} = 0. Then \( \tilde{Z} \in \Gamma_r \cap \Delta \subseteq \bigcup_{i=1}^p \mathbb{B}_\epsilon(X_i, \epsilon_i), \) a contradiction to the fact that \( \tilde{Z} \in \tilde{\Omega} \). In addition, since the sets \( \tilde{\Omega} \) and \( \Gamma_r \cap \Delta \) are compact, there exists a constant \( c > 0 \) such that dist(\( \tilde{Z}, \Gamma_r \cap \Delta \)) \( \leq c \) for all \( \tilde{Z} \in \tilde{\Omega} \). Together with \( \min_{x \in \mathbb{S}^r} (\|Z\|_* - \|Z\|_{(r)}) \geq \tilde{c} \), for any \( Z \in \tilde{\Omega} \) we have dist(\( Z, \Gamma_r \cap \Delta \)) \( \leq (c/\tilde{c})(\|Z\|_* - \|Z\|_{(r)}) \). Consequently, for all \( Z \in \tilde{\Omega} \), dist(\( Z, \Gamma_r \)) \( \leq (c/\tilde{c})(\|Z\|_* - \|Z\|_{(r)}) \). Along with the last inequality, the desired result holds with \( c' = \max(\tilde{c}, c/\tilde{c}) \).

\[ \square \]

### 3.2 Some examples

In this part we use the criteria in (12) and (18) to identify some closed sets \( \Omega \) for which the associated mapping \( S_r \) with any \( r \in [n] \) is calm at 0 for all \( \bar{X} \in S_r(0) \).

**Example 3.1** Fix any \( \varrho > 0 \). Let \( \Omega = \{Z \in \mathbb{X} | ||Z|| \leq \varrho\} \) where \( || \cdot || \) is an arbitrary matrix norm with dual norm \( || \cdot ||_* \). Fix any \( \bar{X} \in \Gamma_r \). Pick any \( H \in [-N_\Omega(\bar{X})] \cap N_{S_r}(\bar{X}) \). From \( H \in N_{\lambda_r}(\bar{X}) \) and Lemma 2.2, we have \( \langle H, \bar{X} \rangle = 0 \). From \( -H \in N_\Omega(\bar{X}) \) and the convexity of \( \Omega \), for any \( Z \in \Omega \) we have \( 0 \leq \langle H, Z - \bar{X} \rangle = : \langle H, Z \rangle - \langle H, \bar{X} \rangle \), which implies that

\[ \varrho ||Z - H||_* = \max_{||Z|| \leq \varrho} \langle Z, -H \rangle = \max_{Z \in \Omega} \langle Z, -H \rangle \leq 0. \]

Hence, \( H = 0 \) and criterion (12) holds at \( \bar{X} \). When \( \Omega = \{Z \in \mathbb{X} | ||Z||_F = \varrho\} \), by noting that \( N_\Omega(\bar{X}) = \{\alpha \bar{X} | \alpha \in \mathbb{R}\} \), one can check that criterion (12) holds at any \( \bar{X} \in \Gamma_r \).

Let \( \Xi = \{Z \in \mathbb{S}^n | ||Z|| \leq \varrho\} \). Since \( \lambda_r(\bar{X}) = N_{S_r}(\bar{X}) + N_{S_r^+}(\bar{X}) \) and argue that \( H = 0 \). Indeed, since \( H \in N_{S_r}(\bar{X}) \) and \( H^2 \in N_{S_r^+}(\bar{X}) \), such that \( H = H^1 + H^2 \). By Lemma 2.1 and 2.2, we have \( \langle H^1, \bar{X} \rangle = 0 \) and \( \langle H^2, \bar{X} \rangle = 0 \), so \( \langle \bar{X}, H \rangle = 0 \). Using the same arguments as above yields that \( H = 0 \).

**Example 3.2** Let \( \Xi = \{X \in \mathbb{S}^n | A(X) = e\} \) where \( A : \mathbb{S}^n \rightarrow \mathbb{R}^n \) is a linear mapping defined by \( A(X) := \text{diag}(X) \). Consider any \( \bar{X} \in \Gamma_r \). Since \( \lambda_r(\bar{X}) = N_{S_r^+}(\bar{X}) + N_{S_r^+}(\bar{X}) \) and argue that \( H = 0 \). Clearly, there exist \( y \in \mathbb{R}^n \), \( H^1 \in N_{S_r^+}(\bar{X}) \) and \( H^2 \in N_{S_r^+}(\bar{X}) \) such that \( H = \text{Diag}(y) = H^1 + H^2 \). From \( H^1 \in N_{S_r^+}(\bar{X}) \), we have \( H^1 \bar{X} = 0 \). By Lemma 2.2, \( H^2 \bar{X} = 0 \). Hence, \( \text{Diag}(y) \bar{X} = 0 \). Along with \( \text{Diag}(\bar{X}) = e \), we get \( y = 0 \) and then \( H = 0 \). For the set \( \Omega = \Xi \cap \mathbb{S}^n \), the associated \( \Gamma_r \) is the composite rank constraint set in [49], and when \( r = 1 \), it is the PSD matrix reformulation of the max-cut problem [18].

**Example 3.3** Fix any \( b_1 \in \mathbb{R} \) and \( b_2 \in \mathbb{R} \) \{0\}. Let \( B, C \in \mathbb{S}^n \) such that \( B - (b_1/b_2)C \) is nonsingular. Let \( \Xi = \{X \in \mathbb{S}^n | A(X) = b\} \) where \( A : \mathbb{S}^n \rightarrow \mathbb{R}^2 \) is a linear mapping defined by \( A(X) := \begin{pmatrix} B & X \\ C & X \end{pmatrix} \). Consider any \( \bar{X} \in \Gamma_r \). To verify that criterion (18) holds at \( \bar{X} \), we pick any \( H \in \text{Range}(A^*) \cap \big[ N_{S_r^+}(\bar{X}) + N_{S_r^+}(\bar{X}) \big] \) and argue that \( H = 0 \). Since \( \text{Range}(A^*) = \{y_1 B + y_2 C | y_1 \in \mathbb{R}, y_2 \in \mathbb{R}\} \), there exist \( y_1 \in \mathbb{R}, y_2 \in \mathbb{R}, \) \( H^1 \in N_{S_r^+}(\bar{X}) \) and \( H^2 \in N_{S_r^+}(\bar{X}) \) such that \( H = y_1 B + y_2 C = H^1 + H^2 \). Since \( \langle \bar{X}, H^1 + H^2 \rangle = 0 \), we have \( y_1 b_1 + y_2 b_2 = 0 \), which implies that \( H = y_1 (B - (b_1/b_2)C) \). Together with \( H^1 \bar{X} = 0 \) and \( H^2 \bar{X} = 0 \), we obtain \( y_1 (B - (b_1/b_2)C) \bar{X} = 0 \), which by the assumption on \( B \) and \( C \).
implies $y_1 = 0$ and then $y_2 = 0$. Consequently, $H = 0$. For the set $\Xi = S^n \cap S_1^n$, when $r = 1$, the set $\Gamma_r$ is exactly the feasible set of the generalized eigenvalue problem \cite{15}.

When $B = 0$ and $b_1 = 0$, $\Xi = \{X \in S^n \mid \langle C, X \rangle = b_2\}$ for an arbitrary nonsingular $C \in S^n$. The above arguments show that the associated $S_r$ is calm at 0 for all $\overline{X} \in \Gamma_r$. The set $\Gamma_r$ with $\Omega = \Xi \cap S_1^n$ for $C = I$ often appears in quantum state tomography \cite{19}.

**Example 3.4** Let $\Xi = \{X \in S^n \mid X_{11} = 1, X_{ii} - \frac{1}{2}(X_{1i} + X_{i1}) = 0 \text{ for } i = 2, \ldots, n\}$. Consider any $\overline{X} \in \Gamma_r$. In order to verify that criterion (18) holds at $\overline{X}$, we pick any $H \in \text{Range}(A^\ast) \cap \{N_{S_1^n}(\overline{X}) + N_{S_r}(\overline{X})\}$ and argue that $H = 0$, where for any $y \in \mathbb{R}^n$

$$A^\ast(y) = \begin{bmatrix} y_1 & -\frac{1}{2}y_2 & -\frac{1}{2}y_3 & \cdots & -\frac{1}{2}y_n \\ -\frac{1}{2}y_2 & y_2 & 0 & \cdots & 0 \\ -\frac{1}{2}y_3 & 0 & y_3 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -\frac{1}{2}y_n & 0 & 0 & \cdots & y_n \end{bmatrix} \in S^n.$$  

Clearly, there exist $y \in \mathbb{R}^n$, $H_1 \in N_{S_1^n}(\overline{X})$ and $H^2 \in N_{S_r}(\overline{X})$ such that $H = H_1 + H^2 = A^\ast(y)$. Since $H\overline{X} = 0$, for $i = 2, 3, \ldots, n$ we have $0 = \langle H\overline{X}\rangle_{ii} = y_i(\overline{X}_{ii} - \overline{X}_{1i}/2)$ and $0 = \langle H\overline{X}\rangle_{i1} = y_i(\overline{X}_{i1} - \overline{X}_{11}/2)$. Along with $\overline{X}_{11} = 1$ and $\overline{X}_{ii} - \overline{X}_{1i} = 0$ for $i = 2, \ldots, n$, we obtain $y_i = 0$ for all $i = 2, \ldots, n$, which implies that $0 = \langle H, \overline{X} \rangle = y_1\overline{X}_{11} = y_1$. Consequently, $H = 0$. When $r = 1$, the set $\Gamma_r$ associated to $\Omega = S_1^n \cap \Xi$ is precisely the PSD matrix reformation for the unconstrained 0-1 quadratic program.

**Example 3.5** Let $\Xi = \{X \in S^n \mid \langle I, X_{ii} \rangle = 1, \langle I, X_{ij} \rangle = 0, i \neq j \in \{1, \ldots, k\}\}$ with $n = kp$, where $X_{ij} \in S^n$ is the $(i, j)$th block of $X$. Fix any $\overline{X} \in \Gamma_r$. Let $A : S^n \rightarrow \mathbb{R}^{k^2}$ be a linear mapping given by $[A(X)]_{(i-1)k+j} = \langle I, X_{ij} \rangle$ for $i, j = 1, \ldots, k$. To verify that (18) holds at $\overline{X}$, we pick any $H \in \text{Range}(A^\ast) \cap \{N_{S_1^n}(\overline{X}) + N_{S_r}(\overline{X})\}$ and argue $H = 0$, where

$$A^\ast(y) = \begin{bmatrix} y_{11}I & y_{12}I & \cdots & y_{1k}I \\ y_{21}I & y_{22}I & \cdots & y_{2k}I \\ \vdots & \vdots & \ddots & \vdots \\ y_{k1}I & y_{k2}I & \cdots & y_{kk}I \end{bmatrix} \text{ for } y \in \mathbb{R}^{k^2}.$$  

Clearly, there exist $y_{ij} \in \mathbb{R}$ for $i, j = 1, \ldots, k$, $H_1 \in N_{S_1^n}(\overline{X})$ and $H^2 \in N_{S_r}(\overline{X})$ such that $H = A^\ast(y) = H_1 + H^2$. Multiplying this equality by $\overline{X}$ yields $H\overline{X} = H_1^\top\overline{X} + H^2\overline{X}$. Notice that $H_1^\top\overline{X} + H^2\overline{X} = 0$. So, for all $i$, $j = 1, 2, \ldots, k$, $0 = \langle H\overline{X}\rangle_{ij} = \sum_{i=1}^k y_{ij}\overline{X}_{ij}$, and consequently $0 = \langle I, \sum_{i=1}^k y_{ij}X_{ij} \rangle = y_{ij}$. This means that $H = 0$.

When $r = 1$, the set $\Gamma_r$ associated to $\Omega = \Xi \cap S_1^n$ is precisely the PSD reformation for the orthogonal matrix set $\{Y \in \mathbb{R}^{p \times k} \mid Y^\top Y = I\}$ with $X = \text{vec}(Y)\text{vec}(Y)^\top$.

**Example 3.6** Let $\Omega = \mathbb{R}_+^{n \times n} \cap \Delta$ with $\Delta = \{X \in \mathbb{R}_+^{n \times n} \mid Xe = e\}$. Consider any $\overline{X} \in \Gamma_r$. To verify that criterion (12) holds, we pick any $H \in [-N_{\Omega}(\overline{X})] \cap N_{\Lambda_r}(\overline{X})$ and argue that $H = 0$. Since $N_{\Omega}(\overline{X}) = N_{\mathbb{R}_+^{n \times n}}(\overline{X}) + N_{\Lambda_r}(\overline{X})$, there exist $y \in \mathbb{R}^n$, $H_1 \in N_{\mathbb{R}_+^{n \times n}}(\overline{X})$ and $H^2 \in N_{\Lambda_r}(\overline{X})$ such that $H = y e^\top - H_1 = H^2$. Note that $\langle H_1, \overline{X} \rangle = 0$ and $\langle H^2, \overline{X} \rangle = 0$. Hence, $0 = \langle ye^\top, \overline{X} \rangle = \langle ye^\top, \overline{X} \rangle = (H_1^\top + H_1^2)^{\top} e = H_1^{\top} X^\top e \leq 0$, which along with $\langle y, e \rangle = 0$ implies that $y = 0$. Together with $X^\top H^2 = 0$, we have $\overline{X}^\top H^1 = 0$. Thus, $0 = e^\top \overline{X}^\top H^1 = e^\top H^1$.  

\(\square\) Springer
Combining this with $-H^1 \in \mathbb{R}^{n \times n}_+$ yields that $H^1 = 0$, so $H = 0$. The set $\Gamma_r$ associated to such $\Omega$ appears in the transition matrix estimation in low-rank Markov chains [34].

**Example 3.7** Let $\Omega = \mathbb{R}^{n \times n}_+ \cap \Delta$ with $\Delta = \{X \in \mathbb{R}^{n \times n} | X e = e, X^T e = e\}$. Consider any $\overline{X} \in \Gamma_r$. To verify that the criterion (12) holds, we pick any $H \in \{-N_\Omega(\overline{X})\} \cap N_\Delta(\overline{X})$ and argue that $H = 0$. Since $N_\Omega(\overline{X}) = N_{\mathbb{R}^{n \times n}_+}(\overline{X}) + N_\Delta(\overline{X})$, by the expression of $N_\Delta(\overline{X})$, there exist $y \in \mathbb{R}^n, z \in \mathbb{R}^n, H^1 \in N_{\mathbb{R}^{n \times n}_+}(\overline{X})$ and $H^2 \in N_{\Lambda_r}(\overline{X})$ such that

$$H = ye^T + e z^T - H^1 = H^2.$$  

Since $H^2 \overline{X}^T = 0$, we have $ye^T + e z^T \overline{X}^T = H^1 \overline{X}^T$. Note that $H^1 \in N_{\mathbb{R}^{n \times n}_+}(\overline{X})$. Hence, $(H^1 \overline{X}^T)_{ii} = 0$, and $(H^1 \overline{X}^T)_{ij} \leq 0$ for all $i, j \in [n]$. Thus,

$$
\begin{align*}
\left\{ \begin{array}{l}
y_i + X_{i1}z_1 + \ldots + X_{in}z_n = 0 \quad \text{for all } i = 1, \ldots, n; \\
y_j + X_{j1}z_1 + \ldots + X_{jn}z_n \leq 0 \quad \text{for all } i, j = 1, \ldots, n.
\end{array} \right.
\end{align*}
$$

(21a) (21b)

Adding the inequalities in (21b) from $j = 1$ to $n$ yields that $ny_i + z_1 + \ldots + z_n \leq 0$ for $i = 1, \ldots, n$. Notice that $(H^1, \overline{X}) = 0$ and $(H^2, \overline{X}) = 0$. From (20), we have $0 = \langle ye^T + e z^T, \overline{X} \rangle = \langle y + z, e \rangle$. From the two sides, we have $y_i \leq \frac{1}{n} \langle y, e \rangle$ for $i = 1, \ldots, n$. This means that $y = e = y_n$ (if not, there is an index $k$ such that $y_k < \frac{1}{n} \langle y, e \rangle$, and then $\frac{1}{n} \langle y, e \rangle = \frac{1}{n} \langle \sum_{i \neq k} y_i + y_k \rangle < \max_{1 \leq i \leq n} y_i \leq \frac{1}{n} \langle y, e \rangle$). Since $\overline{X}^T H^2 = 0$, by (20),

$$
\overline{X}^T ye^T + e z^T = \overline{X}^T H^1 \overline{X}^T \text{.}
$$

Since $H^1 \in N_{\mathbb{R}^{n \times n}_+}(\overline{X})$, $(\overline{X}^T H^1)_{ij} \leq 0$ for all $i, j = 1, \ldots, n$. Thus, $z_i + X_{ij}y_j + \ldots + X_{nj}y_n \leq 0$ for all $i, j = 1, \ldots, n$. Adding these inequalities from $j = 1$ to $n$ yields that $nz_i + y_1 + \ldots + y_n \leq 0$ for $i = 1, \ldots, n$. Together with $\langle y + z, e \rangle = 0$, it implies that $z_i \leq \frac{1}{n} \langle z, e \rangle$ for all $i = 1, \ldots, n$. This means that $z_i = z_2 = \ldots = z_n$. Combining with $y_1 = y_2 = \ldots = y_n$ and (21a), we have $y_1 + z_1 = 0$. This implies that $ye^T + e z^T = 0$. Together with (20), $-H^1 = H^2$. Multiplying this equality by $\overline{X}^T$ yields that $H^1 \overline{X}^T = 0$. Thus, $0 = H^1 \overline{X}^T e = H^1 e$. Note that $H^1 \leq 0$ since $H^1 \in N_{\mathbb{R}^{n \times n}_+}(\overline{X})$. Consequently, we have $H^1 = 0$ and $H = 0$. The set $\Gamma_r$ associated to such $\Omega$ often appears in those problems aiming to seek a low-rank doubly stochastic matrix [55].

**Remark 3.2** (a) For Example 3.1 and 3.2, the calmness of the mapping $S_r$ at 0 for any $\overline{X} \in \Gamma_r$ was shown in [5] by constructing a point in $\Gamma_r$ technically. Here we achieve it by checking the criterion (12) directly. For Examples 3.3–3.7, to the best of our knowledge, there is no work to discuss the calmness of the associated rank constraint system.

(b) For the above examples, the criterion (12) or (18) is shown to hold at any $\overline{X} \in \Gamma_r$. By [1, Proposition 4.1], for Example 3.1, 3.6 and 3.7, the associated function $F(X, Y) = \frac{1}{2} \|X - Y\|_F^2 + \delta_\Omega(\overline{X}) + \delta_\Lambda_r(Y)$ for $X, Y \in \overline{X}$ has the KL property of exponent 1/2 at $(\overline{X}, \overline{X})$; while for Examples 3.2–3.5, the function $G(X, Y) = \frac{1}{2} \|X - Y\|_F^2 + \delta_\Omega(\overline{X}) + \delta_\Lambda_r^+(Y)$ for $X, Y \in \mathbb{S}^n$ has the KL property of exponent 1/2 at $(\overline{X}, \overline{X})$. Thus, for these examples, the proximal alternating minimization method [1] can seek a point of $\Gamma_r$ in a linear rate.

(c) The above $\Omega$ except the one in Example 3.3 are all compact. Then, the calmness of the associated $S_r$ implies the global error bound as in Theorem 3.3. In addition, since the set $\Omega$ in the above examples are regular, when $\Omega$ is from Example 3.1, 3.6 and 3.7, for any $X \in \Gamma_r$ with rank($X$) = $r$, $\overline{N}_r(X) = \overline{N}_r(\overline{X}) = N_\Omega(X) + N_\Lambda_r(X)$, and for any $X \in \Gamma_r$ with rank($X$) < $r$, $N_\Omega(X) \subseteq \overline{N}_r(X) \subseteq \overline{N}_r(\overline{X}) \subseteq N_\Omega(X) + N_\Lambda_r(X)$; when
Ω is from Examples 3.2–3.5, for any \( X \in \Gamma_r \) with \( \text{rank}(X) = r \), \( \tilde{N}_{\Gamma_r}(X) = N_{\Gamma_r}(X) = N_{\Xi}(X) + N_{\Sigma^{+}}(X) + N_{\Sigma^{+}}(X) + N_{\Sigma^{+}}(X) \), and for any \( X \in \Gamma_r \) with \( \text{rank}(X) < r \), \( \tilde{N}_{\Xi}(X) + N_{\Sigma^{+}}(X) \subseteq \tilde{N}_{\Gamma_r}(X) \subseteq \tilde{N}_{\Xi}(X) \subseteq N_{\Xi}(X) + N_{\Sigma^{+}}(X) + N_{\Sigma^{+}}(X) \).

To close this part, we demonstrate via an example that the calmness of \( \mathcal{S}_r \) associated to the above \( \Omega \) can be used to achieve the calmness of \( \mathcal{S}_r \) with a more complicated \( \Omega \).

**Example 3.8** Let \( \Xi = \{ X \in S^{n+1} | \text{diag}(X) = e, A(X) \leq b \} \), where \( b = (b_1, \ldots, b_m)^\top \) is a given vector, and the linear mapping \( \mathcal{A} : S^{n+1} \to \mathbb{R}^m \) is defined as follows:

\[
\mathcal{A}(X) := (\langle A_1, X \rangle, \ldots, \langle A_m, X \rangle)^\top \quad \text{with} \quad A_i = \begin{pmatrix} 0 & e_i^\top \\ c_i & Q_i \end{pmatrix} \quad \text{for} \quad i = 1, \ldots, m.
\]

When \( r = 1 \), the set \( \Gamma_r \) associated to \( \Omega = \Xi \cap S^{n+1}_+ \) is precisely the feasible set of the PSD matrix reformulation for the following binary quadratic programming problem

\[
\begin{align*}
\min_{x \in [-1,1]^n} & \quad c_0^\top x + 2\langle c_1, x \rangle \\
\text{s.t.} & \quad \langle x, Q_i x \rangle + 2\langle c_i, x \rangle \leq b_i, \quad i = 1, 2, \ldots, m.
\end{align*}
\]

Next we use Example 3.2 to argue that the mapping \( \mathcal{S}_1 \) associated to \( \Gamma_1 \) is calm at 0 for any \( \overline{X} \in \Gamma_1 \). Write \( \tilde{\Gamma} := \{ X \in S^{n+1}_+ | \text{rank}(X) \leq 1, \text{diag}(X) = e \} \). Note that \( \tilde{\Gamma} \) is a discrete set, so there exists \( \delta_1 > 0 \) such that for all \( X \in \mathbb{B}(\overline{X}, \delta_1) \), \( \text{dist}(X, \tilde{\Gamma}) = \| X - \overline{X} \|_F \). Since \( \Gamma_1 \subseteq \tilde{\Gamma} \), \( \Gamma_1 \) is also a discrete set in \( S^{n+1}_+ \). Then, there exists \( \delta_2 \in (0, \delta_1) \) such that for all \( X \in \mathbb{B}(\overline{X}, \delta_2) \), \( \text{dist}(X, \Gamma_1) = \| X - \overline{X} \|_F = \text{dist}(X, \tilde{\Gamma}) \). Let \( \Delta := \{ X \in S^{n+1}_+ | \text{diag}(X) = e \} \). By Example 3.2 and Theorem 3.2 (i), there exist \( \beta \geq 0 \) and \( \delta_3 > 0 \) such that

\[
\text{dist}(X, \tilde{\Gamma}) \leq \beta \left[ \text{dist}(X, \Delta) + \text{dist}(X, \Lambda^+_1) \right] \quad \text{for all} \quad X \in \mathbb{B}(\overline{X}, \delta_3).
\]

Take \( \delta := \min(\delta_2, \delta_3) \). From the last inequality it follows that for all \( X \in \mathbb{B}(\overline{X}, \delta) \),

\[
\text{dist}(X, \Gamma_1) = \text{dist}(X, \tilde{\Gamma}) \leq \beta \left[ \text{dist}(X, \Delta) + \text{dist}(X, \Lambda^+_1) \right] \leq \beta \left[ \text{dist}(X, \Xi) + \text{dist}(X, \Lambda^+_1) \right].
\]

By Theorem 3.2, the mapping \( \mathcal{S}_1 \) associated to \( \Omega = \Xi \cap S^{n+1}_+ \) is calm at 0 for any \( \overline{X} \in \Gamma_1 \).

### 4 Applications of calmness of \( \mathcal{S}_r \)

We use the calmness of \( \mathcal{S}_r \) at 0 to achieve several global exact penalties for the rank constrained problem (1) and a family of equivalent DC surrogates for the rank regularized problem (2). Among others, the former covers the penalty problem (4), the Schatten p-norm penalty in [37] and the truncated difference penalty of \( \| \cdot \|_p \) and \( \| \cdot \|_F \) in [40].

#### 4.1 Global exact penalties for problem (1)

Recall that problem (1) is equivalent to the DC constrained optimization problem

\[
\min_{X \in \Omega} \left\{ f(X) \quad \text{s.t.} \quad \| X \|_p - \| X \|_F = 0 \right\}. \tag{22}
\]

Recall that \( f \) is assumed to be coercive and locally Lipschitz continuous relative to \( \Omega \). By [36, Lemma 2.1&Proposition 2.1], we have the following global exact penalty result.

 Springer
Theorem 4.1 If the mapping $S_r$ is calm at 0 for every $X \in S_r(0)$, then problem (22) is partially calm at every local optimal solution $X^*$, i.e., there exist $\varepsilon > 0$ and $\overline{p} > 0$ such that for all $\tau \in \mathbb{R}$ and all $X \in \mathbb{B}(X^*, \varepsilon) \cap S_r(\tau)$, $f(X) - f(X^*) + \overline{p}[\|X\|_{\sigma} - \|X^*\|_{\sigma}] \geq 0$, and consequently, there exists a threshold $\overline{p} > 0$ such that problem (4) associated to every $\rho \geq \overline{p}$ has the same global optimal solution set as problem (1) does.

Remark 4.1 (a) From Theorem 4.1 and Sect. 3.2, we conclude that problem (4) is a global exact penalty of (22) or (1) with $\Omega$ from Examples 3.1–3.7. (b) The partial calmness of problem (22) at every local optimal solution implies that every local optimal solution of (1) is locally optimal to problem (4) associated to $\rho \geq \overline{p}$. Conversely, when a local optimal solution of problem (4) associated to any $\rho > 0$ has rank not more than $r$, it must be locally optimal to problem (22) or (1).

By Theorem 4.1 and Sect. 3.2, the following theorem shows that for the set $\Omega$ from Examples 3.1–3.7, the Schatten $p$-norm penalty for problem (1) is a global exact one. When $\Omega$ is the spectral norm unit ball, we recover the exact penalty result in [37].

Theorem 4.2 If problem (4) is a global exact penalty for problem (1), then for any $p \in (0, 1)$ the following problem is also a global exact penalty of (1):

$$\min_{X \in \Omega} \left\{ f(X) + \rho \sum_{i=r+1}^{n} [\lambda_i(X)]^p \right\}. \quad (23)$$

Proof Fix any $p \in (0, 1)$. Since problem (4) is a global exact penalty for (1), there exists $\overline{p} > 0$ such that problem (4) associated to every $\rho \geq \overline{p}$ has the same global optimal solution set as problem (1) does. For each $\rho > 0$, let $X_\rho$ be a global optimal solution of problem (23) associated to $\rho$. Pick a global optimal solution $X^*$ of (1). Then,

$$f(X_\rho) + \rho \sum_{i=r+1}^{n} [\sigma_i(X_\rho)]^p \leq f(X^*) + \rho \sum_{i=r+1}^{n} [\sigma_i(X^*)]^p = f(X^*).$$

Recall that $f$ is lower bounded. There exists a constant $c_0 > -\infty$ such that $f(X_\rho) > c_0$. Together with the last inequality, we have $\rho \sum_{i=r+1}^{n} [\sigma_i(X_\rho)]^p \leq f(X^*) - c_0$, which implies that there exists $\tilde{\rho} > 0$ such that for all $\rho > \tilde{\rho}$, $\sigma_i(X_\rho) < 1$ for $i = r+1, \ldots, n$. Now fix any $\rho > \max(\overline{p}, \tilde{\rho})$. Let $X^*_\rho$ be a global optimal solution of (4) associated to $\rho$. Then,

$$f(X_\rho) + \rho \sum_{i=r+1}^{n} [\sigma_i(X_\rho)]^p \leq f(X^*_\rho) + \rho \sum_{i=r+1}^{n} [\sigma_i(X^*_\rho)]^p = f(X^*_\rho) + \rho \sum_{i=r+1}^{n} [\sigma_i(X^*_\rho)]^p$$

where the first inequality is by the feasibility of $X^*_\rho$ to (23), the second one is by the feasibility of $X_\rho$ to (4), and the equality is using $\sigma_i(X^*_\rho) = 0$ for all $i = r+1, \ldots, n$, since $X^*_\rho$ is a global optimal solution of (1). From the last inequalities, it follows that $\rho \sum_{i=r+1}^{n} [\sigma_i(X^*_\rho)]^p - \rho \sum_{i=r+1}^{n} [\sigma_i(X^*_\rho)]^p \leq 0$. Together with $0 \leq \sigma_i(X_\rho) < 1$ for all $i = r+1, \ldots, n$, we deduce that $\sigma_i(X_\rho) = 0$ for $i = r+1, \ldots, n$, and hence $X_\rho \in \Gamma_r$. Substituting this into the last inequalities yields that $f(X^*_\rho) = f(X_\rho)$. This means that every global optimal solution of problem (23) associated to $\rho > \max(\overline{p}, \tilde{\rho})$ is globally optimal to (1). In addition, it is easy to argue that every global optimal solution of (1) is globally optimal to (23) associated to any $\rho > 0$. Thus, problem (23) is a global exact penalty of (1). \qed

For each $k \in \{n\}$, let $H_k(X) := \sum_{i=1}^{k-1} \sigma_i(X) + \sqrt{\sum_{i=k}^{n} \sigma_i^2(X)}$ for $X \in \mathbb{X}$, a truncated difference reformulation of $\| \cdot \|_*$ and $\| \cdot \|_p$. By Lemma 2 in Appendix, problem (1) is also equivalent to the following DC constrained problem

$$\min_{X \in \Omega} \left\{ f(X) \text{ s.t. } \|X\_* - H_r(X) = 0 \right\}. \quad (24)$$
Recently, for the least squared function $f$, Ma et al. [40] studied the penalty problem

$$\min_{X \in \Omega} f(X) + \rho \left[ \|X\|_* - H_f(X) \right],$$  \hspace{1cm} (25)$$

but did not verify its global exactness. By Theorem 4.1 and Sect. 3.2, the following theorem shows that (25) is a global exact penalty of (1) when $\Omega$ is from Examples 3.1–3.7.

**Theorem 4.3** For each $k \in [n]$, let $M_k : \mathbb{R} \to \mathbb{R}$ be the multifunction defined by

$$M_k(\tau) := \{ X \in \Omega \mid \|X\|_* - H_k(X) = \tau \}.$$  

Fix any $\overline{X} \in \Gamma_r$. Then $S_r$ is calm at 0 for $\overline{X}$ if and only if $M_r$ is calm at 0 for $\overline{X}$. Consequently, when $\Omega$ satisfies criterion (12) or (18), there exists $\overline{\rho} > 0$ such that problem (25) associated to every $\rho \geq \overline{\rho}$ has the same global optimal solution set as problem (1) does.

**Proof** Suppose that $S_r$ is calm at 0 for $\overline{X}$. Then there exist $\delta > 0$ and $\kappa \geq 0$ such that

$$S_r(\tau) \cap B(\overline{X}, \delta) \subseteq S_r(0) + \kappa |\tau| \quad \text{for all } \tau \in \mathbb{R}. \hspace{1cm} (26)$$

Fix any $\tau \in \mathbb{R}$. If $M_r(\tau) = \emptyset$, then the following inclusion holds for any $\gamma \geq 0:

$$M_r(\tau) \cap B(\overline{X}, \delta) \subseteq M_r(0) + \gamma |\tau|. \hspace{1cm} (27)$$

Now assume that $M_r(\tau) \neq \emptyset$. Pick any $X \in M_r(\tau) \cap B(\overline{X}, \delta)$. Clearly, $\tau = \|X\|_* - H_r(X)$. Note that $X \in S_r(\omega) \cap B(\overline{X}, \delta)$ with $\omega = \|X\|_* - \|X\|_r$. From inclusion (26),

$$\text{dist}(X, M_r(0)) = \text{dist}(X, S_r(0)) \leq \kappa |\omega| = \kappa [\|X\|_* - \|X\|_r] \leq 2\kappa [\|X\|_* - H_r(X)] = 2\kappa |\tau|$$

where the second inequality is by Lemma 2 in Appendix. This, by the arbitrariness of $X$ in $M_r(\tau) \cap B(\overline{X}, \delta)$, implies that $M_r(\tau) \cap B(\overline{X}, \delta) \subseteq M_r(0) + 2\kappa |\tau|$. Together with (27) and the arbitrariness of $\tau$, we conclude that $M_r$ is calm at 0 for $\overline{X}$. Conversely, suppose that $M_r$ is calm at 0 for $\overline{X}$. Then, there exist $\epsilon > 0$ and $\gamma \geq 0$ such that

$$M_r(\tau) \cap B(\overline{X}, \epsilon) \subseteq M_r(0) + \gamma |\tau| \quad \text{for all } \tau \in \mathbb{R}. \hspace{1cm} (28)$$

Fix any $\tau \in \mathbb{R}$. If $S_r(\tau) = \emptyset$, then the following inclusion holds for any $\kappa \geq 0:

$$S_r(\tau) \cap B(\overline{X}, \epsilon) \subseteq S_r(0) + \kappa |\tau|. \hspace{1cm} (29)$$

Now assume that $S_r(\tau) \neq \emptyset$. Pick any $X \in S_r(\tau) \cap B(\overline{X}, \epsilon)$. Clearly, $\tau = \|X\|_* - \|X\|_r$. Note that $X \in M_r(\omega) \cap B(\overline{X}, \epsilon)$ with $\omega = \|X\|_* - H_r(X)$. From inclusion (28), we have

$$\text{dist}(X, S_r(0)) = \text{dist}(X, M_r(0)) \leq \gamma |\omega| = \gamma [\|X\|_* - H_r(X)] \leq \gamma [\|X\|_* - \|X\|_r] = \gamma |\tau|,$$

where the second inequality is due to Lemma 2 in Appendix. This, by the arbitrariness of $X$ in $S_r(\tau) \cap B(\overline{X}, \epsilon)$, implies that $S_r(\tau) \cap B(\overline{X}, \epsilon) \subseteq S_r(0) + \gamma |\tau|$. Together with (29) and the arbitrariness of $\tau$, we conclude that the mapping $S_r$ is calm at 0 for $\overline{X}$. \hfill \Box

**4.2 Equivalent DC surrogates for problem (2)**

Let $\mathcal{S}$ denote the family of proper lsc convex functions $\phi$ on $\mathbb{R}$ satisfying the conditions:

$$\text{int}(\text{dom } \phi) \supseteq [0, 1], \ i^* := \arg \min_{0 \leq t \leq 1} \phi(t) \text{ with } \phi(i^*) = 0, \text{ and } \phi(1) = 1.$$
Many proper lsc convex functions $\phi$ belong to $\mathcal{L}$; see [36, Appendix] for the examples. For each $\phi \in \mathcal{L}$, let $\psi : \mathbb{R} \to (-\infty, +\infty]$ be the closed proper convex function defined by

$$
\psi(t) := \begin{cases} 
\phi(t) & \text{if } t \in [0, 1], \\
+\infty & \text{otherwise.}
\end{cases}
$$

Pick any $\phi \in \mathcal{L}$. It is easy to verify that problem (2) is equivalent to problem

$$
\min_{X \in \Omega, \; W \in \mathbb{X}} \left\{ f(X) + v \sum_{i=1}^{n} \phi(\sigma_i(W)) \; \text{s.t.} \; \|X\|_\Omega - \langle W, X \rangle = 0, \; \|W\| \leq 1 \right\} \tag{31}
$$

in the sense that if $X^*$ is a global (local) optimal solution of (2), then $(X^*, W^*)$ with $W^* = U_1^*V_1^\top + r^*U_2^*V_2^\top$ is globally (locally) optimal to (31), where $U_1^*$ and $V_1^*$ are the matrix consisting of the first $r^*$ columns of $U^*$ and $V^*$, and $U_2^*$ and $V_2^*$ are the matrix consisting of the last $n - r^*$ and $m - r^*$ columns of $U^*$ and $V^*$; and if $(X^*, W^*)$ is a global (local) optimal solution of (31), then $X^*$ is globally (locally) optimal to (2). Let $\mathcal{B}$ denote the spectral norm unit ball in $\mathbb{X}$. Note that $\|X\|_\Omega - \langle W, X \rangle = 0$ and $\|W\| \leq 1$ if and only if $X \in \mathcal{N}_\mathcal{B}(W)$. Hence, problem (31) is a mathematical program with the equilibrium constraint $X \in \mathcal{N}_\mathcal{B}(W)$. In fact, when $\Omega \subseteq S^n_+$, it reduces to

$$
\min_{X \in \Omega, \; W \in S^n} \left\{ f(X) + v \sum_{i=1}^{n} \phi(\lambda_i(W)) \; \text{s.t.} \; \langle I - W, X \rangle = 0, \; 0 \leq W \leq I \right\}.
$$

Let $\mathcal{X}^*$ denote the global optimal solution set of (2). For each $X \in \mathcal{X}$, denoted by $\sigma_{nz}(X)$ the smallest nonzero singular value of $X$. The following lemma characterizes the uniform lower boundedness of $\sigma_{nz}(X)$ for every $X \in \mathcal{X}^*$.

**Lemma 4.1** There exists a constant $\alpha > 0$ such that for all $X \in \mathcal{X}^*$, $\sigma_{nz}(X) > \alpha$.

**Proof** Suppose on the contradiction that the conclusion does not hold. There exists a sequence $\{X^k\}_{k \in \mathbb{N}} \subseteq \mathcal{X}^*$ such that for each $k \in \mathbb{N}$, $\sigma_{nz}(X^k) \leq 1/k$. Clearly, there exist an index set $K \subseteq \mathbb{N}$ and an index $j \in [n]$ such that $\sigma_j(X^k) \leq 1/k$ for all $k \in K$. Recall that $\mathcal{X}^*$ is compact due to the coerciveness of $f$ on the set $\Omega$. By taking a subsequence if necessary, we may assume that $\{X^k\}_{k \in K}$ is convergent, say, $\lim_{k \to +\infty} X^k = X^* \in \mathcal{X}^*$. Together with $\sigma_j(X^k) \leq 1/k$ for all $k \in K$, we obtain $\sigma_j(X^*) = 0$. Thus, for all sufficiently large $k \in K$, $\text{rank}(X^k) \geq \text{rank}(X^*) + 1$. Since $f$ is lsc, for all sufficiently large $k \in K$, $f(X^k) \geq f(X^*) - v/2$. From $\{X^k\}_{k \in K} \subseteq \mathcal{X}^*$ and $X^* \in \mathcal{X}^*$, we obtain

$$
f(X^*) + v\text{rank}(X^*) = f(X^k) + v\text{rank}(X^k) \geq f(X^*) + v/2 + v\text{rank}(X^*),
$$

which is impossible. This shows that the desired conclusion holds. \qed

The following lemma implies that the MPEC reformulation (31) for problem (2) with $\Omega$ from Examples 3.1–3.7 is partially calm at every global optimal solution.

**Lemma 4.2** If for each $r \in [n]$ the mapping $S_r$ is calm at $0$ for any $X \in \Gamma_r$, then the MPEC (31) is partially calm at every global optimal solution $(X^*, W^*)$, i.e., there exist $\delta > 0$ and $\overline{\rho} > 0$ such that for all $\varepsilon \geq 0$ and all $(X, W) \in \mathbb{B}((X^*, W^*), \delta) \cap \{(X, W) \in \Omega \times \mathcal{B} | \|X\|_\Omega - \langle X, W \rangle = \varepsilon\}$,

$$
f(X) + v \sum_{i=1}^{n} \phi(\sigma_i(W)) - \left[ f(X^*) + v \sum_{i=1}^{n} \phi(\sigma_i(W^*)) \right] + \overline{\rho}\varepsilon(\|X\|_\Omega - \langle X, W \rangle) \geq 0.
$$
Proof Recall that $f$ is assumed to be locally Lipschitz continuous relative to $\Omega$. There exist $\delta' > 0$ and $L_f > 0$ such that for any $Z, Z' \in B(X^*, \delta') \cap \Omega$,

$$|f(Z) - f(Z')| \leq L_f \|Z - Z'\|_F. \quad (32)$$

By Lemma 4.1, there exists $\alpha > 0$ such that $\sigma_{nz}(X^*) > \alpha$. Write $k := \text{rank}(X^*)$. From the continuity, for all $X \in B(X^*, \delta')$ (if necessary by reducing $\delta'$), $\sigma_{nz}(X) > \alpha$. Let $\beta := B(X^*, \delta') f(X) < +\infty$, which is well defined by the continuity of $f$ relative to $\Omega$. Then, by the coerciveness of $f + \delta \Omega$, the set $\mathcal{L}_\beta := \{X \in \Omega \mid f(X) \leq \beta\}$ is compact. By the given assumption and Theorem 3.3, for each $r \in [n]$ there exists $\kappa_r > 0$ such that

$$\text{dist}(Z, S_r(0)) \leq \kappa_r \left[\|Z\|_* - \|Z\|_{(r)}\right] \quad \text{for all } Z \in \mathcal{L}_\beta. \quad (33)$$

Let $\kappa := \max\{\kappa_1, \ldots, \kappa_n\}$ and $\delta := \delta'/2$. Take $\overline{\rho} := \max\left\{\phi'(1) \alpha, \frac{k \phi'(1)(1-t^*) L_f}{v(1-t_0)}\right\}$ where $t_0 \in [0, 1)$ is such that $\frac{1}{1-t} \in \partial \phi(t)$ and its existence is due to [36, Lemma 1]. Fix any $\epsilon \geq 0$ and pick any $(X, W) \in B((X^*, W^*), \delta) \cap \{\delta \mid (X, W) \in \Omega \times B \mid \|X\|_* - (X, W) = \epsilon\}$. Clearly, $X \in \mathcal{L}_\beta$. Define the index set $\overline{J} := \{i \in [n] \mid \overline{\rho} \sigma_i(X) > \phi'_-(1)\}$ and $\overline{\tau} := |\overline{J}|$. By using (33) with $Z = X$, there necessarily exists a point $\overline{X} \in \Pi_{S_r(0)}(X)$ such that

$$\|X - \overline{X}\|_F \leq \kappa \tau \sum_{i=\tau+1}^n \sigma_i(X). \quad (34)$$

Since $\overline{\rho} \geq \frac{\phi'_-(1)}{\alpha}$, we have $k \leq \overline{\tau}$. Together with $k = \text{rank}(X^*)$ and $X^* \in \Omega$, we deduce that $X^* \in S_r(0)$, and consequently, $\|\overline{X} - X^*\|_F \leq 2\|X - X^*\| \leq \delta'$. By invoking (32),

$$|f(X) - f(\overline{X})| \leq L_f \|X - \overline{X}\|_F. \quad (35)$$

Let $J_1 = \{i \mid \frac{1}{1-t^*} \leq \overline{\rho} \sigma_i(X) \leq \phi'_-(1)\}$ and $J_2 = \{i \mid 0 \leq \overline{\rho} \sigma_i(X) < \frac{1}{1-t^*}\}$. Note that

$$\sum_{i=1}^n \phi(\sigma_i(W)) + \overline{\rho}(\|X\|_* - (W, X)) \geq \sum_{i=1}^n \min_{t \in [0, 1]} \phi(t) + \overline{\rho} \phi'_-(1)(1-t)\).$$

By invoking [36, Lemma 1] with $\omega = \sigma_i(X)$, we obtain the following inequalities

$$\sum_{i=1}^n \phi(\sigma_i(W)) + \overline{\rho}(\|X\|_* - (W, X))$$

$$\geq \overline{\tau} + \frac{\overline{\rho}(1-t_0)}{\phi'_-(1)(1-t')} \sum_{j \in J_1} \sigma_j(X) + \overline{\rho}(1-t_0) \sum_{j \in J_2} \sigma_j(X)$$

$$\geq \text{rank}(|\overline{X}|) + \frac{\overline{\rho}(1-t_0)}{\phi'_-(1)(1-t')} \sum_{j \in J_1 \cup J_2} \sigma_j(X)$$

$$\geq \text{rank}(|\overline{X}|) + \frac{\overline{\rho}(1-t_0)}{\kappa \phi'_-(1)(1-t')} \|X - \overline{X}\|_F$$

$$\geq \text{rank}(|\overline{X}|) + v^{-1} L_f \|X - \overline{X}\|_F \geq \text{rank}(|\overline{X}|) + v^{-1}[f(\overline{X}) - f(X)]$$

where the second inequality is since $\overline{X} \in \Gamma_\tau$ and $\phi(t^*) - \phi(1) \geq \phi'_-(1)(t^* - 1)$, the third one is due to $J_1 \cup J_2 = \{j \in [n] \mid \sigma_j(X) < \sigma_j(X)\}$ and inequality (34), and the last one is using (35). Now assume that $\overline{X}$ has the SVD given by $\overline{U} \text{Diag}(\sigma(\overline{X})) \overline{V}^\top$ with $\overline{U} \in \mathcal{O}^n$ and $\overline{V} \in \mathcal{O}^m$. Let $\overline{W} = \overline{U}_1 \overline{V}_1^\top + t^* \overline{U}_2 \overline{V}_2^\top$, where $\overline{U}_1$ and $\overline{U}_2$ are the matrix consisting of the first $\overline{\tau}$ columns and the rest $n - \overline{\tau}$ columns of $\overline{U}$, and $\overline{V}_1$ and $\overline{V}_2$ are the matrix consisting of the first $\overline{\tau}$ columns and the rest $m - \overline{\tau}$ columns of $\overline{V}$. Clearly, $(\overline{X}, \overline{W})$ is a feasible point.
of the MPEC (31) and \( \sum_{i=1}^{n} \phi(\sigma_i(\bar{W})) = \text{rank}(\bar{X}) \). From the last inequality, it immediately follows that

\[
\begin{align*}
f(X) + \nu \sum_{i=1}^{n} \phi(\sigma_i(W)) + \bar{\nu} \|X\|_* - \langle W, X \rangle & \geq f(\bar{X}) + \nu \sum_{i=1}^{n} \phi(\sigma_i(\bar{W})) \\
& \geq f(X^*) + \nu \sum_{i=1}^{n} \phi(\sigma_i(W^*)).
\end{align*}
\]

This gives the desired inequality. The MPEC (31) is partially calm at \((X^*, W^*)\).

Now we are in a position to provide a family of equivalent DC surrogates for the rank regularized problem (2), which greatly improves the result of [6, Corollary 4.2] where the equivalent surrogates are only achieved for the unitarily invariant matrix norm ball.

**Theorem 4.4** If for each \( r \in [n] \) the mapping \( S_r \) is calm at 0 for all \( X \in S_r(0) \), then there exists a threshold \( \bar{\nu} > 0 \) such that the following problem associated to \( \rho \geq \bar{\nu} \)

\[
\min_{X \in \Omega} \left\{ f(X) + \nu \rho \left[ \|X\|_* - \rho^{-1} \sum_{i=1}^{n} \psi^*(\rho \sigma_i(X)) \right] \right\}
\]

(36)

has the same global optimal solution set as problem (2) does, where \( \psi^* \) is the conjugate of \( \psi \). Also, for the set \( \Omega \) from Examples 3.1–3.7, the assumption automatically holds.

**Proof** By combining Lemma 4.2 with [36, Proposition 2.1] and using the expression of \( \psi \) in (30), there exists a threshold \( \bar{\nu} > 0 \) such that the following penalized problem

\[
\min_{X \in \Omega, W \in \mathbb{X}} \left\{ f(X) + \nu \sum_{i=1}^{n} \psi(\sigma_i(W)) + \nu \rho \left[ \|X\|_* - \langle W, X \rangle \right] \right\}
\]

associated to every \( \rho \geq \bar{\nu} \) has the same global optimal solution set as the MPEC (31) does. From the definition of conjugate functions and von Neumann’s trace inequality, for every \( X \in \Omega \) it holds that \( \sup_{W \in \mathbb{X}} \left\{ \langle W, \rho X \rangle - \sum_{i=1}^{n} \psi(\sigma_i(W)) \right\} = \sum_{i=1}^{n} \psi^*(\rho \sigma_i(X)) \). Consequently, the last penalized problem is simplified as the one in (36). The conclusion then follows from the equivalence between (31) and (2) in a global sense.

From Theorem 4.4 and the examples in [36, Appendix], it follows that for the set \( \Omega \) from Examples 3.1–3.7, the matrix version of the capped-\( \ell_1 \), the SCAD and MCP, and the truncated \( \ell_p \) (0 < p < 1) are all the equivalent DC surrogates for problem (2).

## 5 Conclusions

For the composite rank constraint system \( X \in \Gamma_r \), we obtained two criteria for identifying those closed \( \Omega \) such that the associated partial perturbation \( S_r \) possesses the calmness at 0, and also illustrated their practicality by a collection of common nonnegative and PSD composite rank constraint sets. The calmness of \( S_r \) was also used to achieve several global exact penalties for problem (1) and a family of equivalent DC surrogates for problem (2) involving more types of \( \Omega \). Notice that the results in Sect. 3 are easily extended to the mapping \( S_r(\tau) \times \Upsilon_s(\sigma \tau) \), where \( \Upsilon_s(\sigma \tau) = \{ S \in \Delta \mid \|\text{vec}(S)\|_1 - \|\text{vec}(S)\|_0 = \sigma \} \) for \( \sigma \in \mathbb{R} \) is a partial perturbation to the zero-norm constraint \( \{ S \in \Delta \mid \|\text{vec}(S)\|_0 \leq s \} \). Then, for the rank plus zero-norm constrained or regularized problem, one can obtain the corresponding global exact penalties and equivalent DC surrogates. Our future work will explore other practical criteria to establish error bounds for more structured sets.
Acknowledgements The authors would like to thank the anonymous referee and the associated editor for their helpful comments. This work is supported by the National Natural Science Foundation of China under projects The Guangdong Basic and Applied Basic Research Foundation No. 2020A1515010408 and No. 2021A1515010210.

Appendix

Lemma 1 Let $G \in \mathbb{X}$ have the SVD as $U \text{Diag}(\sigma(G))V^\top$. Then, for any $r \in [n]$, $$U_1 \Sigma_r(G)V_1^\top \in \arg\min_{Z \in \Lambda_r} \|Z - G\|_*,$$

where $U_1$ and $V_1$ are the matrix consisting of the first $r$ columns of $U$ and $V$, respectively, and $\Sigma_r(G) = \text{diag}(\sigma_1(G), \ldots, \sigma_r(G))$ with $\sigma_1(G) \geq \sigma_2(G) \geq \cdots \geq \sigma_r(G)$.

Proof By Mirsky’s theorem (see [53, IV Theorem 4.11]), $\|Z - G\|_* \geq \|\sigma(Z) - \sigma(G)\|_1$ for any $Z \in \mathbb{X}$. Then it is easy to argue that if $Z^*$ is an optimal solution of $\min_{Z \in \Lambda_r} \|Z - G\|_*$, then $\sigma(Z^*)$ is optimal to $\min_{\|z\|_0 \leq r} \|z - \sigma(G)\|_1$. Conversely, if $z^*$ is an optimal solution of $\min_{\|z\|_0 \leq r} \|z - \sigma(G)\|_1$, then $U \text{Diag}(|z^*|^1)V^\top$ is optimal to $\min_{Z \in \Lambda_r} \|Z - G\|_*$. Clearly, $\text{diag}(\Sigma_r(G))$ is an optimal solution of $\min_{\|z\|_0 \leq r} \|z - \sigma(G)\|_1$. The result then holds. $\square$

Lemma 2 Fix an integer $r \in [n]$. Then, for any $X \in \mathbb{X}$, it holds that

$$\frac{1}{2}\|X\|_* - \|X\|_{(r)} \leq \sum_{i=r}^n \sigma_i(X) - \sqrt{\sum_{i=r}^n \sigma_i^2(X)} \leq \|X\|_* - \|X\|_{(r)}. \quad (37)$$

Proof Fix any $X \in \mathbb{X}$. Since $\sqrt{\sum_{i=r}^n \sigma_i^2(X)} \geq \sigma_r(X)$, it immediately follows that $\sum_{i=r}^n \sigma_i(X) - \sqrt{\sum_{i=r}^n \sigma_i^2(X)} \leq \sum_{i=r}^n \sigma_i(X) = \|X\|_* - \|X\|_{(r)}$, and the second inequality in (37) holds. Next we prove that the first inequality in (37) holds by two cases.

Case 1: $\sigma_{r+1}(X) \neq 0$. Recall that $\sigma_1(X) \geq \sigma_2(X) \geq \cdots \geq \sigma_n(X)$. Therefore,

$$\sum_{i=r}^n \sigma_i(X) - \sqrt{\sum_{i=r}^n \sigma_i^2(X)} = \sum_{i=r+1}^n \sigma_i(X) + \left[\sigma_r(X) - \sqrt{\sigma_r^2(X) + \sum_{i=r+1}^n \sigma_i^2(X)}\right]$$

$$= \sum_{i=r+1}^n \sigma_i(X) - \frac{\sum_{i=r+1}^n \sigma_i^2(X)}{\sigma_r(X) + \sqrt{\sigma_r^2(X) + \sum_{i=r+1}^n \sigma_i^2(X)}}$$

$$\geq \sum_{i=r+1}^n \sigma_i(X) - \frac{1}{2} \sum_{i=r+1}^n \frac{\sigma_i(X)}{\sigma_r(X)} \sigma_i(X)$$

$$\geq \sum_{i=r+1}^n \sigma_i(X) - \frac{1}{2} \sum_{i=r+1}^n \sigma_i(X) = \frac{1}{2}\|X\|_* - \|X\|_{(r)}.$$
References

1. Attouch, H., Bolte, J., Redont, P., Soubeyran, A.: Proximal alternating minimization and projection methods for nonconvex problems: an approach based on the Kurdyka-Łojasiewicz inequality. Math. Oper. Res. 35, 438–457 (2010)
2. Bai, K., Ye, J.J., Zhang, J.: Directional quasi-/pseudo-normality as sufficient conditions for metric sub-regularity. SIAM J. Optim. 29, 2625–2649 (2019)
3. Bauschke, H.H., Borwein, J.M., Li, W.: Strong conical hull intersection property, bounded linear regularity, Jameson’s property (G), and error bounds in convex optimization. Math. Program. 86, 135–160 (1999)
4. Benko, M., Gfrerer, H., Outrata, J.V.: Calculus for directional limiting normal cones and subdifferentials. Set-Valued Variat. Anal. 27, 713–745 (2019)
5. Bi, S.J., Pan, S.H.: Error bounds for rank constrained optimization problems and applications. Oper. Res. Lett. 44, 336–341 (2016)
6. Bi, S.J., Pan, S.H.: Multistage convex relaxation approach to rank regularized minimization problems based on equivalent mathematical program with a generalized complementarity constraint. SIAM J. Control. Optim. 55, 2493–2518 (2017)
7. Bonnans, J.F., Shapiro, A.: Perturbation Analysis of Optimization Problems. Springer, New York (2000)
8. Dontchev, A.L., Rockafellar, R.T.: Implicit Functions and Solution Mappings. Springer Monographs in Mathematics, LLC, New York (2009)
9. Dukanovic, I., Rendl, F.: Semidefinite programming relaxations for graph coloring and maximal clique problems. Math. Program. 109, 345–365 (2007)
10. Fabian, M.J., Henrion, R., Kruger, A.Y., Outrata, J.V.: Error bounds: necessary and sufficient conditions. Set-Valued Anal. 18, 121–149 (2010)
11. Fan, J., Li, R.: Variable selection via nonconcave penalized likelihood and its oracle properties. J. Am. Stat. Assoc. 9, 1348–1360 (2001)
12. Fazel, M., Hindi, H., Boyd, S.P.: Log-det heuristic for matrix rank minimization with applications to hankel and Euclidean distance matrices. In: American Control Conference 3, pp. 2156–2162 (2003)
13. Fazel, M.: Matrix Rank Minimization with Applications, Ph.D. thesis, Stanford University (2002)
14. Gao, Y., Sun, D.F.: A majorized penalty approach for calibrating rank constrained correlation matrix problems. Technical report, Department of Mathematics, National University of Singapore (2010)
15. Ge, R., Jin, C., Netrapalli, P., Sidford, A.: Efficient algorithms for large-scale generalized eigenvector computation and canonical correlation analysis. In: International Conference on Machine Learning, pp. 2741–2750 (2016)
16. Gfrerer, H.: First order and second order characterizations of metric subregularity and calmness of constraint set mappings. SIAM J. Optim. 21, 1439–1474 (2011)
17. Gfrerer, H.: Optimality conditions for disjunctive programs based on generalized differentiation with application to mathematical programs with equilibrium constraints. SIAM J. Optim. 24, 898–931 (2014)
18. Goemans, M.X., Williamson, D.P.: Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming. J. Assoc. Comput. Math. 42, 1115–1145 (1995)
19. Gross, D., Liu, Y.K., Flammia, S.T., Becker, S., Eisert, J.: Quantum state tomography via compressed sensing. Phys. Rev. Lett. 105, 150401 (2011)
20. Hajek, B., Wu, Y.H., Xu, J.M.: Semidefinite programs for exact recovery of a hidden community. In: 29th Annual Conference on Learning Theory, Proceedings of Machine Learning Research 49, pp. 1051–1095 (2016)
21. Helmberg, C., Rendl, F.: Solving quadratic (0,1)-problems by semidefinite programs and cutting planes. Math. Program. 82, 291–315 (1998)
22. Henrion, R., Outrata, J.: Calmness of constraint systems with applications. Math. Program. 104, 437–464 (2005)
23. Henrion, R., Jourani, A., Outrata, J.: On the calmness of a class of multifunctions. SIAM J. Optim. 13, 520–534 (2002)
24. Horst, R., Thoai, N.V.: DC programming: overview. J. Optim. Theory Appl. 103, 1–43 (1999)
25. Ioffe, A.D.: Regular points of Lipschitz functions. Trans. Am. Math. Soc. 251, 61–69 (1979)
26. Ioffe, A.D., Outrata, J.V.: On metric and calmness qualification conditions in subdifferential calculus. Set-Valued Anal. 16, 199–227 (2008)
27. Jiang, Z.X., Zhao, X.Y., Ding, C.: A proximal DC approach for quadratic assignment problem. Comput. Optim. Appl. 78, 825–851 (2021)
28. Kruger, A.Y.: About regularity of collections of sets. Set-Valued Anal. 4, 187–206 (2006)
29. Kruger, A.Y.: Error bounds and metric subregularity. Optimization 64, 49–79 (2015)
30. Kulis, B., Sustik, M.A., Dhillon, I.S.: Low-rank Kernel learning with Bregman matrix divergences. J. Mach. Learn. Res. 10, 341–376 (2009)
31. Lai, M.J., Xu, Y.Y., Yin, W.T.: Improved iteratively reweighted least squares for unconstrained smoothed $\ell_q$ minimization. SIAM J. Numer. Anal. 5, 927–957 (2013)
32. Le Thi, H.A., Pham Dinh, T.: DC programming and DCA: thirty years of developments. Math. Program. 169, 5–68 (2018)
33. Lewis, A.S., Luke, D.R., Malick, J.: Local linear convergence for alternating and averaged nonconvex projections. Found. Comput. Math. 9, 485–513 (2009)
34. Li, X.D., Wang, M.D., Zhang, A.R.: Estimation of Markov chain via rank-constrained likelihood. In: Proceedings of the 35th International Conference on Machine Learning, PMLR 80, 3033–3042 (2018)
35. Li, X.R., Xiu, N.H., Zhou, S.L.: Matrix optimization over low-rank spectral sets: stationary points and local and global minimizers. J. Optim. Theory Appl. 184, 895–930 (2020)
36. Liu, Y.L., Bi, S.J., Pan, S.H.: Equivalent Lipschitz surrogates for zero-norm and rank optimization problems. J. Global Optim. 72, 679–704 (2018)
37. Liu, T.X., Lu, Z.S., Chen, X.J., Dai, Y.H.: An exact penalty method for semidefinite-box-constrained low-rank matrix optimization problems. IMA J. Numer. Anal. 40, 563–86 (2020)
38. Lu, C.Y., Tang, J.H., Yan, S.C., Lin, Z.C.: Generalized nonconvex nonsmooth low-rank minimization. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 4130–4137
39. Luke, D.R.: Prox-regularity of rank constraint sets and implications for algorithms. J. Math. Imag. Vis. 47, 231–238 (2013)
40. Ma, T.H., Hou, Y.F., Huang, T.Z.: Truncated $\ell_{1-2}$ models for sparse recovery and rank minimization. SIAM J. Imaging Sci. 10, 1346–1380 (2017)
41. Mohan, K., Fazel, M.: Iterative reweighted algorithms for matrix rank minimization. J. Mach. Learn. Res. 13, 3441–3473 (2012)
42. Mordukhovich, B.S.: Nonsmooth analysis with nonconvex generalized differentials and adjoint mappings. Doklady Akademia Nauk BSSR (Belorussian Academy of Sciences) 28, 976–979 (1984). (Russian)
43. Mordukhovich, B.S.: Stability theory for parametric generalized equations and variational inequalities via nonsmooth analysis. Trans. Am. Math. Soc. 343, 609–656 (1994)
44. Pietersz, R., Groenen, P.J.F.: Rank reduction of correlation matrices by majorization. Quant. Finance 4, 649–662 (2004)
45. Qian, Y.T., Pan, S.H.: A relaxation approach to UBPP based on equivalent DC penalized factorized matrix programs. J. Optim. Theory Appl. 159, 915–949 (2013)
46. Robinson, S.M.: Some continuity properties of polyhedral multifunctions. Math. Program. Study 14, 206–214 (1981)
47. Rockafellar, R.T., Wets, R.J.-B.: Variational Analysis. Springer, Berlin (1998)
48. Stewart, G.W., Sun, J.G.: Matrix Perturbation Theory. Academic Press, London (1990)
49. Wu, Z.H., Ye, J.J.: First-order and second-order conditions for error bounds. SIAM J. Optim. 14, 621–645 (2003)
50. Ye, J.J., Zhou, J.C.: Variational analysis and applications. Springer, New York (2010)
51. Ye, J.J., Zhou, J.C.: Nearly unbiased variable selection under minimax concave penalty. Ann. Stat. 38, 894–942 (2010)
Publisher's Note  Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.