Yokohama Publishers
ISSN 1349-8169 ONLINE JOURNAL

Pacific Journal of Optimization
Vol. 18, No. 1, pp. 1–26, 2022

KL PROPERTY OF EXPONENT 1/2 OF $\ell_{2,0}$-NORM AND DC REGULARIZED FACTORIZATIONS FOR LOW-RANK MATRIX RECOVERY*

SHUJUN BI, TING TAO† AND SHAOHUA PAN

Abstract: This paper is concerned with the factorization form of the rank regularized loss minimization problem. To cater for the scenario in which only a coarse estimation is available for the rank of the true matrix, an $\ell_{2,0}$-norm regularized term is added to the factored loss function to reduce the rank adaptively; and account for the ambiguities in the factorization, a balanced term is then introduced. For the least squares loss, under a restricted condition number assumption on the sampling operator, we establish the KL property of exponent 1/2 of the nonsmooth factored composite function and its equivalent DC regularized surrogates in the set of their global minimizers. We also confirm the theoretical findings by applying a proximal linearized alternating minimization method to the regularized factorizations.

Key words: *low-rank matrix recovery, rank regularized factored loss minimization, $\ell_{2,0}$-norm and equivalent DC regularized surrogates, KL property of exponent* 1/2

Mathematics Subject Classification: *15A83, 47A52, 90C26*

---

1 Introduction

Let $\mathbb{R}^{m\times n}$ be the vector space of all $m\times n$ real matrices, endowed with the trace inner product $\langle\cdot,\cdot\rangle$ and its induced Frobenius norm $\|\cdot\|_F$. Low-rank matrix recovery problems aim to recover a matrix $M\in\mathbb{R}^{m\times n}$ of rank r with $1\le r\ll\min(m,n)$ via some noiseless or noisy observations. Let $f:\mathbb{R}^{m\times n}\to\mathbb{R}_+$ denote an appropriate smooth empirical loss function. When a tight upper estimation, say the positive integer κ, for the rank of M is available, it is natural to model low-rank matrix recovery problems as

$$\min_{X\in\mathbb{R}^{m\times n}}\Big\{f(X)\ \ \text{s.t.}\ \ \text{rank}(X)\le\kappa\Big\};\tag{1.1}$$

if only a coarse estimation, say $\min(m,n)$, is available, it is reasonable to model it as

$$\min_{X\in\mathbb{R}^{m\times n}}\Big\{\nu f(X)+\text{rank}(X)\Big\},\tag{1.2}$$

*This work was supported by the National Natural Science Foundation of China (Nos. 11971177 and 11701186).

†Corresponding author

© 2022 Yokohama Publishers

where $\nu > 0$ is the regularization parameter. Without loss of generality, we stipulate that $m \leq n$ and assume that problem (1.2) has a nonempty global optimal solution set, denoted by $\mathcal{X}^*$. When the loss function f is specified as

$$f(X) := \frac{1}{2}\|\mathcal{A}(X) - b\|^2 \tag{1.3}$$

for a sampling operator $\mathcal{A}\colon \mathbb{R}^{m\times n} \to \mathbb{R}^p$ and an observation vector $b \in \mathbb{R}^p$, the above (1.1) and (1.2) become the popular rank constrained and rank regularized least squares problem, respectively. Rank optimization problems have a host of applications in a variety of fields such as system identification and control [14, 16], signal and image processing [20, 12], machine learning [40, 3, 26], statistics [35, 24], finance [38], and quantum tomography [19], just to name a few (see also the survey [13] for some further applications).

Due to the involved combinatorial property, rank optimization problems are generally NP-hard [15] and it is almost impossible to seek their global optima with an algorithm of polynomial-time complexity. Over the past decade, many convex relaxation approaches have been developed. A popular one is the nuclear norm relaxation approach [10, 39] that yields a low-rank solution via a single tractable nuclear norm optimization problem. Observe that the nuclear norm has a relatively weak ability even fails to promoting a low-rank solution in some scenarios [42, 33]. Some researchers have proposed convex relaxation approaches by solving a sequential convex program arising from a certain nonconvex surrogate of rank minimization problems (see, e.g., [15, 27, 34]). Recently, Bi and Pan [4] proposed a multi-stage convex relaxation approach to (1.2) by the global exact penalty for its equivalent MPEC reformulation, which shows that several convex relaxation problems are enough to yield a high-quality low-rank solution. Although convex relaxation methods have received well study from many fields such as information, computer science, statistics, optimization, and so on (see, e.g., [39, 11, 35, 24, 9, 45]), the computation of tractable convex optimization problems involved in them are very expensive since each iterate requires performing a singular value decomposition (SVD), which making them computationally prohibitive in large-scale settings [21].

To overcome the computational bottleneck of the convex relaxation approach, inspired by the work [7], a common way is to reparameterize the $m \times n$ matrix variable X as $UV^{\mathbb{T}}$ with $U \in \mathbb{R}^{m\times\kappa}$ and $V \in \mathbb{R}^{n\times\kappa}$, and achieve a desirable low-rank solution by solving a bi-factored nonconvex problem with less variables which is often regularized by a term to account for the ambiguities in the factorization $X = UV^{\mathbb{T}}$. Corresponding to the rank constrained problem (1.1), one usually solves the following smooth nonconvex optimization

$$\min_{U\in\mathbb{R}^{m\times\kappa}, V\in\mathbb{R}^{n\times\kappa}} \Big\{ f(UV^{\mathbb{T}}) + \frac{\mu}{4}\|U^{\mathbb{T}}U - V^{\mathbb{T}}V\|_F^2 \Big\} \tag{1.4}$$

where $\mu > 0$ is the regularization parameter. Such a reparametrization automatically enforces the low-rank structure and leads to low computational cost per iteration. Owing to this, the nonconvex factored approach is widely used in large-scale applications such as recommendation systems or collaborative filtering [25, 26]. Though the bi-factored problem is nonconvex, it is amenable to local search heuristics such as gradient descent or alternating minimization which can exploit its bilinear structure well. In fact, numerical experiments indicate that they tend to work well in practice [17, 28, 43].

Despite the superior empirical performance of the nonconvex factored approach, the understanding of its theoretical guarantees is rather limited in comparison with the convex relaxation approach since its analysis is well known to be notoriously difficult. In the recent years, some active progress has been made for the theory of the nonconvex factored approach.

One research line focuses on the (regularized) factorizations of rank optimization problems from a local view and aims to characterize the growth behavior of objective functions around the set of global optimal solutions (see, e.g., [22, 36, 43, 46, 48, 50, 51]). A direct consequence of this line is that standard methods such as gradient descent and alternating minimization, when initialized with a point in the stated neighborhood, will converge linearly to an optimal solution. Another line takes a global view and aims to establish the geometric landscape of the factorization models of rank optimization problems, especially strict saddle point property (see, e.g., [37, 17, 18, 29, 30, 49, 52]). As a result, suitably modified gradient descent methods can be shown to converge globally to an optimal solution, and their convergence rates can be established; see Jin et al. [23].

We notice that most of the above theoretical results focus on the exact-parametrization case $\kappa = r$ of the (regularized) factorizations, and consequently are applicable to the rank optimization problems (1.1) and (1.2) for which the rank of the optimal solutions is known. In fact, the most crucial and difficult part for most of rank optimization problems is to achieve the rank of global optima or its best estimation. This means that the factored model of the rank regularized problem (1.2) with a coarse estimation κ is more practical. Motivated by the efficiency of the nuclear norm relaxation approach, there are some works [40, 8, 30, 48] centering around the factorized nuclear norm surrogate problem

$$\min_{U\in\mathbb{R}^{m\times\kappa},V\in\mathbb{R}^{n\times\kappa}} \Big\{\nu f(UV^{\mathbb{T}}) + \frac{1}{2}(\|U\|_F^2 + \|V\|_F^2)\Big\} \tag{1.5}$$

where the positive integer κ is an upper estimation for r. Among others, Li et al. [30] took a global view and proved that each critical point of the smooth factored problem (1.5) either corresponds to the global optimum of the nuclear norm surrogate problem or is a strict saddle point; while Zhang et al. [48] took a local view and established the KL property of exponent 1/2 of the objective function over the set of global optima but only for the exact-parametrization case $\kappa = r$ and the loss f in (1.3) with a full sampling operator $\mathcal{A}$. As previously mentioned, the nuclear norm has a weak ability even fails to promoting a low-rank solution in some cases, which implies that the Fro-norm regularized factorization model will carry on this potential insufficiencey.

Motivated by this, in this work we concentrate on the following factored form of (1.2):

$$\min_{U\in\mathbb{R}^{m\times\kappa},V\in\mathbb{R}^{n\times\kappa}} \Big\{\Psi(U,V) := \nu f(UV^{\mathbb{T}}) + \frac{\mu}{4}\|U^{\mathbb{T}}U - V^{\mathbb{T}}V\|_F^2 + \frac{1}{2}\big(\|U\|_{2,0} + \|V\|_{2,0}\big)\Big\}, \tag{1.6}$$

where the positive integer κ is an upper estimation for r, and $\|U\|_{2,0}$ means the $\ell_{2,0}$-norm of U, i.e., the number of nonzero columns of U. As will be shown in Section 3.1, when the set $\mathcal{X}^*$ contains a matrix with rank no more than κ, the factored problem (1.6) is equivalent to (1.2) in the sense that their global optima can be obtained from each other. The balanced term $\frac{\mu}{4}\|U^{\mathbb{T}}U - V^{\mathbb{T}}V\|_F^2$ is introduced to remove the ambiguities caused by the bilinear form $UV^{\mathbb{T}}$. Consider that the discontinuity of $\ell_{2,0}$-norm may cause some inconvenience in designing effective algorithms for solving (1.6). We also reformulate (1.6) as a mathematical program with equilibrium constraint (MPEC) by the variational characterization of $\ell_{2,0}$-norm, and derive some equivalent DC (difference of convexity) regularized surrogates from the global exact penalty of the MPEC under a suitable restriction on the loss f. The main contribution of this work is, for the least squares loss with $b = \mathcal{A}(M)$, to establish the KL property of exponent 1/2 of the function Ψ and its equivalent DC regularized surrogates over the set of their global minimizers under a restricted condition number assumption on $\mathcal{A}$.

For nonconvex and nonsmooth functions, the KL property of exponent 1/2 over the set of critical points is a weak growth behavior to guarantee that many first-order algorithms for minimizing them converge to a critical point with a linear rate (see, e.g., [1, 2, 6]). From [1, Section 4] many classes of functions indeed satisfy the KL property, but generally it is not an easy task to verify whether these functions have the KL property of exponent 1/2 or not over the set of critical points. Though some useful rules are provided in [31] to identify the exponent of KL functions, in most cases one still needs to analyze them case by case. Though we establish the KL property of exponent 1/2 of the function Ψ and its equivalent DC regularized surrogates only in a subset of their critical point set, to the best of our knowledge, this work is the first to achieve such a growth behavior for the nonsmooth factored function with an over-parametrization case $\kappa \geq r$.

2 Notation and Preliminaries

Throughout this paper, $\mathbb{O}^n$ represents the set of all $n \times n$ orthonormal matrices, e denotes a vector of all ones whose dimension is known from the context, and $\mathrm{Diag}(z)$ for a vector z means a rectangular diagonal matrix. For a given $X \in \mathbb{R}^{m\times n}$, $\sigma(X) \in \mathbb{R}^m$ denotes the singular value vector of X arranged in a nonincreasing order, i.e., $\sigma_1(X) \geq \cdots \geq \sigma_m(X)$, $\mathbb{O}^{m,n}(X) := \{(P,Q) \in \mathbb{O}^m \times \mathbb{O}^n \mid X = P\mathrm{Diag}(\sigma(X))Q^{\mathbb{T}}\}$, $\|X\|$ means the spectral norm of X, X_j denotes the j-th column of X, and for an index set $J \subseteq \{1,\ldots,n\}$, X_J represents the matrix consisting of those X_j with $j \in J$. For a given $\delta > 0$, $\mathbb{B}(X,\delta)$ denotes the closed ball on the Frobenius norm centered at X of radius δ. For a linear operator $\mathcal{B}\colon \mathbb{R}^{m\times n} \to \mathbb{R}^p$, $\|\mathcal{B}\|$ means the spectral norm of $\mathcal{B}$, and $\mathcal{B}^*$ denotes the adjoint of $\mathcal{B}$. For a given integer $k \geq 1$, write $\Omega_k := \big\{X \in \mathbb{R}^{m\times n} \mid \mathrm{rank}(X) \leq k\big\}$. For a convex set $\mathcal{C}$, $\mathcal{C}^\infty$ denotes the recession cone of $\mathcal{C}$. In the sequel, we denote by $\mathscr{L}$ the family of proper lower semicontinuous (lsc) convex functions $\phi\colon \mathbb{R} \to (-\infty,+\infty]$ satisfying

$$[0,1] \subseteq \mathrm{int}(\mathrm{dom}\phi),\ \ \phi(1) = 1,\ \ t^*_\phi = \mathop{\arg\min}_{t\in[0,1]} \phi(t) \ \ \text{with } \phi(t^*_\phi) = 0. \tag{2.1}$$

For each $\phi \in \mathscr{L}$, let $\psi\colon \mathbb{R} \to (-\infty,+\infty]$ be the associated proper lsc convex function:

$$\psi(t) := \begin{cases} \phi(t) & \text{if } t \in [0,1]; \\ +\infty & \text{otherwise,} \end{cases} \tag{2.2}$$

and denote by ψ^* the conjugate function of ψ, i.e., $\psi^*(s) := \sup_{t\in\mathbb{R}}\{st - \psi(t)\}$ for $s \in \mathbb{R}$.

2.1 Generalized subdifferentials

Next we recall from [41, Definition 8.3] the concepts of the regular, limiting and horizon subdifferentials of an extended real-valued function at a finite-valued point. For an extended real-valued $g\colon \mathbb{R}^n \to (-\infty,+\infty]$, write $\mathrm{dom}\, g := \{x \in \mathbb{R}^n \mid g(x) < +\infty\}$.

Definition 2.1. Consider a function $g\colon \mathbb{R}^n \to [-\infty,+\infty]$ and a point x with $g(x)$ finite. The regular subdifferential of g at x, denoted by $\widehat{\partial} g(x)$, is defined as

$$\widehat{\partial} g(x) := \Big\{ v \in \mathbb{R}^n \ \Big|\ \liminf_{\substack{x'\to x\\ x'\neq x}} \frac{g(x') - g(x) - \langle v, x'-x\rangle}{\|x'-x\|} \geq 0 \Big\};$$

the (limiting) subdifferential of g at x, denoted by $\partial g(x)$, is defined as

$$\partial g(x) := \Big\{ v \in \mathbb{R}^n \mid \exists\, x^k \xrightarrow[g]{} x \text{ and } v^k \in \widehat{\partial} g(x^k) \text{ with } v^k \to v \Big\},$$

and the horizon subdifferential of g at x, denoted by $\partial^\infty g(x)$, is defined as

$$\partial^\infty g(x) := \Big\{ v \in \mathbb{R}^n \mid \exists\, x^k \underset{g}{\to} x \text{ and } v^k \in \widehat{\partial} g(x^k) \text{ with } \lambda_k v^k \to v \text{ for some } \lambda_k \downarrow 0 \Big\}.$$

Remark 2.2. **(i)** The regular and limit subdifferentials are all closed and satisfy $\widehat{\partial} g(x) \subseteq \partial g(x)$, and the set $\widehat{\partial} g(x)$ is convex but $\partial g(x)$ is generally nonconvex. When g is convex, $\widehat{\partial} g(x) = \partial g(x)$ and is precisely the subdifferential of g at x in the sense of convex analysis. When g is nonconvex, there may be a big difference between them. For example, for the function $g(z) = -\|z\|$ for $z \in \mathbb{R}^n$, we have $\widehat{\partial} g(0) = \emptyset$ but $\partial g(0) = \{v \in \mathbb{R}^n \mid \|v\| = 1\}$.

(ii) The point $\overline{x}$ at which $0 \in \partial g(\overline{x})$ (respectively, $0 \in \widehat{\partial} g(\overline{x})$) is called a limiting (respectively, regular) critical point of g. By [41, Theorem 10.1], a local minimizer of g is necessarily a regular critical point, and consequently a limit critical point.

(iii) By [41, Corollary 8.11], g is (subdifferentially) regular at $\overline{x}$ if and only if g is locally lsc at $\overline{x}$ with $\partial g(\overline{x}) = \widehat{\partial} g(\overline{x})$ and $\partial^\infty g(\overline{x}) = [\widehat{\partial} g(\overline{x})]^\infty$.

We also recall from [41, Chapter 6] the concepts of the regular normal cone and limiting normal cone. Let $\mathcal{C} \subseteq \mathbb{R}^n$ be a closed set. The regular normal cone to $\mathcal{C}$ at a point $\overline{x} \in \mathcal{C}$ is defined by

$$\widehat{\mathcal{N}}_{\mathcal{C}}(\overline{x}) := \Big\{ v \in \mathbb{R}^n : \limsup_{x \underset{\mathcal{C}}{\to} \overline{x}} \frac{\langle v, x - \overline{x}\rangle}{\|x - \overline{x}\|} \le 0 \Big\},$$

and the limiting normal cone to $\mathcal{C}$ at $\overline{x} \in \mathcal{C}$ is defined as the outer limit of $\widehat{\mathcal{N}}_{\mathcal{C}}(x)$ as $x \underset{\mathcal{C}}{\to} \overline{x}$, i.e.,

$$\mathcal{N}_{\mathcal{C}}(\overline{x}) := \Big\{ v \in \mathbb{R}^n : \exists\, x^k \underset{\mathcal{C}}{\to} \overline{x},\ v^k \to v \text{ with } v^k \in \widehat{\mathcal{N}}_{\mathcal{C}}(x^k) \Big\}.$$

The following lemmas focus on the subdifferential of two composite functions.

Lemma 2.3. *Let $h(z) := \mathrm{sign}(\|z\|)$ for $z \in \mathbb{R}^n$. Fix an arbitrary $\overline{z} \in \mathbb{R}^n$. Then, we have*

$$\widehat{\partial} h(\overline{z}) = \partial h(\overline{z}) = \begin{cases} \{0\}^n & \text{if } \overline{z} \neq 0; \\ \mathbb{R}^n & \text{if } \overline{z} = 0. \end{cases}$$

Proof. When $\overline{z} \neq 0$, since $\widehat{\partial}\mathrm{sign}(t)|_{t=\|\overline{z}\|} = \partial \mathrm{sign}(t)|_{t=\|\overline{z}\|} = \{0\}$, by [41, Exercise 10.7] it follows that $\widehat{\partial} h(\overline{z}) = \partial h(\overline{z}) = \{0\}^n$. When $\overline{z} = 0$, by Definition 2.1 it is easy to calculate that $\widehat{\partial} h(\overline{z}) = \mathbb{R}^n$, which means that $\partial h(\overline{z}) = \mathbb{R}^n$. So, the result holds. □

Lemma 2.4. *Let $h(z) = \vartheta(g(z))$ where $\vartheta : \mathbb{R} \to \mathbb{R}$ is a locally Lipschitz function and $g(z) \equiv \|z\|$ for $z \in \mathbb{R}^n$, and $\mathrm{gph} g$ be the graph of g. Consider any point $\overline{z} \in \mathbb{R}^n$. Then,*

$$\partial h(\overline{z}) = \partial \vartheta(\|\overline{z}\|) \frac{\overline{z}}{\|\overline{z}\|} \ \text{ if } \ \overline{z} \neq 0 \ \text{ and } \ \partial h(\overline{z}) \subseteq D^* g(\overline{z})[\partial \vartheta(\|\overline{z}\|)] \ \text{ if } \ \overline{z} = 0$$

where $D^* g(\overline{z})$ is the coderivative of g at $\overline{z}$, i.e., $u \in D^* g(\overline{z})(\tau)$ iff $(u, -\tau) \in \mathcal{N}_{\mathrm{gph} g}(\overline{z}, \|\overline{z}\|)$.

Proof. Since ϑ is a locally Lipschitz function, $\partial^\infty \vartheta(t) = \{0\}$ for any $t \in \mathbb{R}$. When $\overline{z} \neq 0$, the result follows from [41, Exercise 10.7]. When $\overline{z} = 0$, by [41, Theorem 10.49] we have $\partial h(\overline{z}) \subseteq D^* g(\overline{z}) \partial \vartheta(\|\overline{z}\|)$, where $D^* g(\overline{z})$ is the coderivative of g at $\overline{z}$. □

By Remark 2.2(iii), $-g$ is not regular since $\widehat{\partial}(-g)(0) \neq \partial(-g)(0)$. This means that sg for each $s \in \partial \vartheta(0)$ is not regular at the origin unless $\partial \vartheta(0) \subseteq \mathbb{R}_+$. Consequently, the inclusion in Lemma 2.4 generally can not become an equality.

2.2 Kurdyka-Łojasiewicz property

We recall from [1] the concept of the KL property of an extended real-valued function.

Definition 2.5. Let $g\colon \mathbb{R}^n \to (-\infty, +\infty]$ be a proper function. The function g is said to have the Kurdyka-Łojasiewicz (KL) property at $\overline{x} \in \operatorname{dom} \partial g$ if there exist $\eta \in (0, +\infty]$, a continuous concave function $\varphi\colon [0, \eta) \to \mathbb{R}_+$ satisfying the following two conditions

(i) $\varphi(0) = 0$ and φ is continuously differentiable on $(0, \eta)$;

(ii) for all $s \in (0, \eta)$, $\varphi'(s) > 0$,

and a neighborhood $\mathcal{U}$ of $\overline{x}$ such that for all $x \in \mathcal{U} \cap \big[g(\overline{x}) < g < g(\overline{x}) + \eta\big]$,

$$\varphi'(g(x) - g(\overline{x}))\mathrm{dist}(0, \partial g(x)) \geq 1.$$

If the corresponding φ can be chosen as $\varphi(s) = c\sqrt{s}$ for some $c > 0$, then g is said to have the KL property of exponent $1/2$ at $\overline{x}$. If g has the KL property of exponent $1/2$ at each point of $\operatorname{dom} \partial g$, then g is called a KL function of exponent $1/2$.

Remark 2.6. By [1, Lemma 2.1], a proper function has the KL property of exponent $1/2$ at any noncritical point. Hence, to show that it is a KL function of exponent $1/2$, it suffices to check whether it has the KL property of exponent $1/2$ at each critical point.

2.3 Restricted smallest and largest eigenvalues

When handling low-rank matrix recovery problems, restricted strong convexity and restricted smoothness are common requirement for loss functions (see, e.g., [35, 52, 30]). For the least squares loss function in (1.3), these properties essentially require that the restricted smallest and largest eigenvalues of $\mathcal{A}^*\mathcal{A}$ satisfy a certain condition.

Definition 2.7. Let $\mathcal{B}\colon \mathbb{R}^{m\times n} \to \mathbb{R}^p$ be a given linear mapping. The k-restricted smallest and largest eigenvalues of of $\mathcal{B}^*\mathcal{B}$ are respectively defined as follows:

$$\lambda_{k,\min}(\mathcal{B}^*\mathcal{B}) := \min_{X\in\Omega_k, \|X\|_F=1} \|\mathcal{B}(X)\|^2 \quad \text{and} \quad \lambda_{k,\max}(\mathcal{B}^*\mathcal{B}) := \max_{X\in\Omega_k, \|X\|_F=1} \|\mathcal{B}(X)\|^2,$$

and the ratio $\frac{\lambda_{k,\max}(\mathcal{B}^*\mathcal{B})}{\lambda_{k,\min}(\mathcal{B}^*\mathcal{B})}$ is called the k-restricted condition number of $\mathcal{B}^*\mathcal{B}$.

Clearly, the least squares loss function (1.3) has the k-restricted strong convexity in [52, 30] if and only if the k-restricted smallest eigenvalue of $\mathcal{A}^*\mathcal{A}$ is positive. When the k-restricted smallest and largest eigenvalues of $\mathcal{A}^*\mathcal{A}$ satisfy $\lambda_{k,\min}(\mathcal{A}^*\mathcal{A}) = 1 - \delta_k$ and $\lambda_{k,\max}(\mathcal{A}^*\mathcal{A}) = 1 + \delta_k$ for some $\delta_k \in [0, 1)$, the sampling operator $\mathcal{A}$ satisfies the restricted isometry property (RIP). Thus, we conclude from [39] that for many types of random sampling operators, there is a high probability for $\mathcal{A}^*\mathcal{A}$ to have a good restricted condition number, that is, the value of $\lambda_{r,\max}(\mathcal{A}^*\mathcal{A})/\lambda_{r,\min}(\mathcal{A}^*\mathcal{A})$ is not too large.

For the least squares loss function (1.3), the following lemma was appeared in [44], which improves a little the result of [30, Proposition 2.1].

Lemma 2.8. *For the loss function* (1.3)*, let α and β be the r-restricted smallest and largest eigenvalues of $\mathcal{A}^*\mathcal{A}$, respectively. If $\alpha > 0$, for any $X, Y \in \mathbb{R}^{m\times n}$ with* $\operatorname{rank}([X\ Y]) \leq r$,

$$\Big|\frac{2}{\alpha + \beta}\langle \mathcal{A}(X), \mathcal{A}(Y)\rangle - \langle X, Y\rangle\Big| \leq \frac{\beta - \alpha}{\beta + \alpha}\|X\|_F\|Y\|_F.$$

To close this section, we characterize the rank function in terms of the $\ell_{2,0}$-norm.

Lemma 2.9. *Given a matrix $X \in \mathbb{R}^{m\times n}$. If* $\mathrm{rank}(X) \le \kappa$ *for an integer* $\kappa \ge 1$*, then*

$$\mathrm{rank}(X) = \min_{R\in\mathbb{R}^{m\times\kappa},L\in\mathbb{R}^{n\times\kappa}} \Big\{\frac{1}{2}\big(\|R\|_{2,0} + \|L\|_{2,0}\big) \colon X = RL^{\mathbb{T}}\Big\}. \tag{2.3}$$

Proof. Take an arbitrary feasible point $(R, L) \in \mathbb{R}^{m\times\kappa} \times \mathbb{R}^{n\times\kappa}$ of (2.3). Notice that $X = RL^{\mathbb{T}} = \sum_{j=1}^{\kappa} R_j L_j^{\mathbb{T}}$ and there are at most $\min(\|R\|_{2,0}, \|L\|_{2,0})$ nonzero terms. Hence, $\mathrm{rank}(X) \le \min(\|R\|_{2,0}, \|L\|_{2,0})$ and $\mathrm{rank}(X) \le \frac{1}{2}\big(\|R\|_{2,0} + \|L\|_{2,0}\big)$. Since (R, L) is an arbitrary feasible point, this shows that $\mathrm{rank}(X)$ is a lower bound for the objective function of (2.3) over its feasible set. Also, by taking $(U, V) \in \mathbb{O}^{m,n}(X)$ and setting $\overline{R} = \big[\sqrt{\sigma_1(X)}U_1 \ \cdots \ \sqrt{\sigma_\kappa(X)}U_\kappa\big]$ and $\overline{L} = \big[\sqrt{\sigma_1(X)}V_1 \ \cdots \ \sqrt{\sigma_\kappa(X)}V_\kappa\big]$, we have

$$\|\overline{R}\|_{2,0} = \|\overline{L}\|_{2,0} = \mathrm{rank}(X) \ \text{ and } \ X = \overline{R}\,\overline{L}^{\mathbb{T}}.$$

This shows that the optimal value of (2.3) equals $\mathrm{rank}(X)$. The desired result holds. ☐

3 Factorized Reformulations

We provide several factored reformulations of the rank regularized problem (1.2) by means of the $\ell_{2,0}$-norm of matrices and its variational characterization. Recall the definition of the function family $\mathscr{L}$. With an arbitrary $\phi \in \mathscr{L}$, it is easy to check that for any $z \in \mathbb{R}^\kappa$,

$$\|z\|_0 = \min_{w\in\mathbb{R}^\kappa} \Big\{\textstyle\sum_{i=1}^{\kappa}\phi(w_i) :\ 0 \le w \le e,\ \langle e - w, |z|\rangle = 0\Big\}.$$

Consequently, for any $Z \in \mathbb{R}^{m\times n}$, with $\mathcal{G}(Z) := (\|Z_1\|, \|Z_2\|, \ldots, \|Z_n\|)^{\mathbb{T}}$ it holds that

$$\|Z\|_{2,0} = \min_{w\in\mathbb{R}^n} \Big\{\textstyle\sum_{i=1}^{n}\phi(w_i) :\ 0 \le w \le e,\ \langle e - w, \mathcal{G}(Z)\rangle = 0\Big\}. \tag{3.1}$$

This provides a variational characterization for the $\ell_{2,0}$-norm of matrices. Such a characterization was exploited in [5] to design a convex relaxation approach to group sparsity.

3.1 $\ell_{2,0}$-norm regularized factorization

We first argue that the rank regularized problem (1.2) can be reformulated as an equivalent $\ell_{2,0}$-norm regularized factorization model. This is implied by the following lemma.

Lemma 3.1. *If X^* is a global optimal solution of rank r for the problem* (1.2)*, then* (R^*, L^*) *with* $R^* = [\sqrt{\sigma_1(X^*)}U_1^* \ \cdots \ \sqrt{\sigma_\kappa(X^*)}U_\kappa^*]$ *and* $L^* = [\sqrt{\sigma_1(X^*)}V_1^* \ \cdots \ \sqrt{\sigma_\kappa(X^*)}V_\kappa^*]$ *for* $(U^*, V^*) \in \mathbb{O}^{m,n}(X^*)$ *is globally optimal to the following problem with* $\kappa \ge r$

$$\min_{U\in\mathbb{R}^{m\times\kappa},V\in\mathbb{R}^{n\times\kappa}} \Big\{\nu f(UV^{\mathbb{T}}) + \frac{1}{2}\big(\|U\|_{2,0} + \|V\|_{2,0}\big)\Big\}. \tag{3.2}$$

Conversely, if $\mathcal{X}^* \cap \Omega_\kappa \ne \emptyset$ *and* $(\overline{U}, \overline{V})$ *is a global optimal solution of* (3.2)*, then* $\overline{X} = \overline{U}\,\overline{V}^{\mathbb{T}}$ *is globally optimal to the problem* (1.2)*.*

Proof. Fix an arbitrary $(U, V) \in \mathbb{R}^{m\times\kappa} \times \mathbb{R}^{n\times\kappa}$ and write $X = UV^{\mathbb{T}}$. By Lemma 2.9, $\|U\|_{2,0} + \|V\|_{2,0} \ge 2\mathrm{rank}(X)$, which along with the global optimality of X^* implies that

$$\nu f(UV^{\mathbb{T}}) + \frac{1}{2}\big(\|U\|_{2,0} + \|V\|_{2,0}\big) \ge \nu f(X) + \mathrm{rank}(X) \ge \nu f(X^*) + \mathrm{rank}(X^*).$$

Notice that $R^*L^{*\mathbb{T}} = X^*$ and $2\text{rank}(X^*) = \|R^*\|_{2,0} + \|L^*\|_{2,0}$. From the last inequality,

$$\nu f(UV^{\mathbb{T}}) + \frac{1}{2}\big(\|U\|_{2,0} + \|V\|_{2,0}\big) \geq \nu f(R^*L^{*\mathbb{T}}) + \frac{1}{2}\big(\|R^*\|_{2,0} + \|L^*\|_{2,0}\big).$$

By the arbitrariness of (U, V), this shows that (R^*, L^*) is globally optimal to (3.2).

Conversely, let X^* be an arbitrary point from $\mathcal{X}^* \cap \Omega_\kappa$ and let $(U^*, V^*) \in \mathbb{O}^{m,n}(X^*)$. Write $R = \big[\sqrt{\sigma_1(X^*)}U_1^* \ \cdots \ \sqrt{\sigma_\kappa(X^*)}U_\kappa^*\big]$ and $L = \big[\sqrt{\sigma_1(X^*)}V_1^* \ \cdots \ \sqrt{\sigma_\kappa(X^*)}V_\kappa^*\big]$. Then, it holds that $\|R\|_{2,0} + \|L\|_{2,0} = 2\text{rank}(X^*)$. Consequently, we have

$$\begin{aligned}
\nu f(X^*) + \text{rank}(X^*) &= \nu f(RL^{\mathbb{T}}) + \frac{1}{2}\big(\|R\|_{2,0} + \|L\|_{2,0}\big) \\
&\geq \nu f(\overline{U}\overline{V}^{\mathbb{T}}) + \frac{1}{2}\big(\|\overline{U}\|_{2,0} + \|\overline{V}\|_{2,0}\big) \\
&\geq \nu f(\overline{X}) + \text{rank}(\overline{X}) \quad \text{with } \overline{X} = \overline{U}\overline{V}^{\mathbb{T}},
\end{aligned}$$

where the last inequality is by Lemma 2.9. So, $\overline{U}\overline{V}^{\mathbb{T}}$ is globally optimal to (1.2). □

Lemma 3.1 shows that if an upper bound κ is available for a low-rank global optimal solution of (1.2), seeking such a low-rank global optimal solution is equivalent to finding a global optimal solution of the $\ell_{2,0}$-norm regularized factorization model (3.2). Thus, to achieve a low-rank global optimal solution of (1.2) with balanced factors, i.e., $\overline{X} = \overline{U}\overline{V}^{\mathbb{T}}$ with $\overline{U}^{\mathbb{T}}\overline{U} = \overline{V}^{\mathbb{T}}\overline{V}$, one may solve the $\ell_{2,0}$-norm regularized factorization model (1.6). The following lemma builds a bridge for the global optimal solution set of (1.6) and (1.2).

Lemma 3.2. *Suppose $\mathcal{X}^* \cap \Omega_\kappa \neq \emptyset$. Then, the optimal solution set of* (1.6) *has the form*

$$\mathcal{W}^* := \Big\{(\overline{U}, \overline{V}) \mid \overline{U}\overline{V}^{\mathbb{T}} \in \mathcal{X}^* \cap \Omega_\kappa,\ \overline{U}^{\mathbb{T}}\overline{U} = \overline{V}^{\mathbb{T}}\overline{V},\ \|\overline{U}\|_{2,0} = \|\overline{V}\|_{2,0} = \text{rank}(\overline{U}\overline{V}^{\mathbb{T}})\Big\}.$$

Proof. Take an arbitrary $X^* \in \mathcal{X}^* \cap \Omega_\kappa$. By Lemma 2.9, for any $(U, V) \in \mathbb{R}^{m\times\kappa} \times \mathbb{R}^{n\times\kappa}$,

$$\begin{aligned}
&\nu f(UV^{\mathbb{T}}) + \frac{\mu}{4}\|U^{\mathbb{T}}U - V^{\mathbb{T}}V\|_F^2 + \frac{1}{2}(\|U\|_{2,0} + \|V\|_{2,0}) \\
&\geq \nu f(UV^{\mathbb{T}}) + \frac{1}{2}(\|U\|_{2,0} + \|V\|_{2,0}) \\
&\geq \nu f(UV^{\mathbb{T}}) + \text{rank}(UV^{\mathbb{T}}) \geq \nu f(X^*) + \text{rank}(X^*). \qquad (3.3)
\end{aligned}$$

Moreover, when $(U, V) = (\overline{U}, \overline{V})$ for an arbitrary $(\overline{U}, \overline{V})$ from $\mathcal{W}^*$, it holds that

$$\begin{aligned}
&\nu f(\overline{U}\overline{V}^{\mathbb{T}}) + \frac{\mu}{4}\|\overline{U}^{\mathbb{T}}\overline{U} - \overline{V}^{\mathbb{T}}\overline{V}\|_F^2 + \frac{1}{2}(\|\overline{U}\|_{2,0} + \|\overline{V}\|_{2,0}) \\
&= \nu f(\overline{U}\overline{V}^{\mathbb{T}}) + \text{rank}(\overline{U}\,\overline{V}^{\mathbb{T}}) = \nu f(X^*) + \text{rank}(X^*)
\end{aligned}$$

where the last equality is due to $\overline{U}\,\overline{V}^{\mathbb{T}} \in \mathcal{X}^*$. The last two equations show that the problems (1.6) and (1.2) have the same optimal value, and by the arbitrariness of $(\overline{U}, \overline{V})$ in $\mathcal{W}^*$, we conclude that $\mathcal{W}^*$ is included in the set of global optimal solutions to (1.6). Thus, it suffices to argue that the converse inclusion holds. For this purpose, let $(\overline{U}, \overline{V})$ be an arbitrary global optimal solution to (1.6). Then, by using the fact that the problems (1.6) and (1.2) have the same optimal value and Lemma 2.9, it follows that

$$\begin{aligned}
\nu f(X^*) + \text{rank}(X^*) &= \nu f(\overline{U}\,\overline{V}^{\mathbb{T}}) + \frac{\mu}{4}\|\overline{U}^{\mathbb{T}}\overline{U} - \overline{V}^{\mathbb{T}}\overline{V}\|_F^2 + \frac{1}{2}(\|\overline{U}\|_{2,0} + \|\overline{V}\|_{2,0}) \\
&\geq \nu f(\overline{U}\,\overline{V}^{\mathbb{T}}) + \text{rank}(\overline{U}\,\overline{V}^{\mathbb{T}}) \geq \nu f(X^*) + \text{rank}(X^*). \qquad (3.4)
\end{aligned}$$

This implies that $\overline{U}\,\overline{V}^{\mathbb{T}} \in \mathcal{X}^*$ and the inequalities in (3.4) become the equalities. Then,

$$\nu f(\overline{U}\,\overline{V}^{\mathbb{T}}) + \frac{\mu}{4}\|\overline{U}^{\mathbb{T}}\overline{U} - \overline{V}^{\mathbb{T}}\overline{V}\|_F^2 + \frac{1}{2}(\|\overline{U}\|_{2,0} + \|\overline{V}\|_{2,0}) = \nu f(\overline{U}\,\overline{V}^{\mathbb{T}}) + \mathrm{rank}(\overline{U}\,\overline{V}^{\mathbb{T}}).$$

Along with $\frac{1}{2}(\|\overline{U}\|_{2,0} + \|\overline{V}\|_{2,0}) \geq \mathrm{rank}(\overline{U}\,\overline{V}^{\mathbb{T}})$ by Lemma 2.9, we deduce that $\overline{U}^{\mathbb{T}}\overline{U} = \overline{V}^{\mathbb{T}}\overline{V}$ and $\frac{1}{2}(\|\overline{U}\|_{2,0} + \|\overline{V}\|_{2,0}) = \mathrm{rank}(\overline{U}\,\overline{V}^{\mathbb{T}})$, which implies that $\|\overline{U}\|_{2,0} = \|\overline{V}\|_{2,0} = \mathrm{rank}(\overline{U}\,\overline{V}^{\mathbb{T}})$. That is, $(\overline{U},\overline{V}) \in \mathcal{W}^*$. Hence, the desired converse inclusion holds. □

Remark 3.3. Lemma 3.1 and 3.2 show that a global optimal solution of (1.2) can be obtained from that of (3.2) or (1.6) when $\mathcal{X}^* \cap \Omega_\kappa \neq \emptyset$. It is easy to verify that every global optimal solution $(\overline{U},\overline{V})$ of (3.2) with $\overline{U}^{\mathbb{T}}\overline{U} = \overline{V}^{\mathbb{T}}\overline{V}$ is necessarily a global optimal solution of (1.6), and every global optimal solution of (1.6) is a global optimal solution of (3.2).

Lemma 3.2 implies that if the set $\mathcal{X}^* \cap \Omega_\kappa$ can be characterized, one may achieve the global optimal solution set of (1.6). The following proposition states that for the function f specified as in (1.3) with $b = \mathcal{A}(M)$, the set $\mathcal{X}^* \cap \Omega_k$ can be characterized under a suitable condition for the restricted smallest eigenvalue, and so is the set $\mathcal{W}^*$.

Proposition 3.4. *Suppose that the function f is given by* (1.3) *with $b = \mathcal{A}(M)$ for a matrix M of rank r, and $\mathcal{X}^* \cap \Omega_\kappa \neq \emptyset$. If the $2r$-restricted smallest eigenvalue α of the linear operator $\mathcal{A}^*\mathcal{A}$ satisfies $\alpha > \frac{2}{\nu\sigma_r^2(M)}$, then $\mathcal{X}^* \cap \Omega_k = \{M\}$, and consequently*

$$\mathcal{W}^* := \Big\{(\overline{U},\overline{V}) \mid \overline{U}\,\overline{V}^{\mathbb{T}} = M,\, \overline{U}^{\mathbb{T}}\overline{U} = \overline{V}^{\mathbb{T}}\overline{V},\, \|\overline{U}\|_{2,0} = \|\overline{V}\|_{2,0} = r\Big\}.$$

Proof. Take an arbitrary $\overline{X} \in \mathcal{X}^* \cap \Omega_\kappa$ and write $\overline{r} := \mathrm{rank}(\overline{X})$. By the expression of f,

$$\nu f(\overline{X}) + \mathrm{rank}(\overline{X}) \leq \nu f(M) + \mathrm{rank}(M) = \mathrm{rank}(M) = r. \tag{3.5}$$

This implies that $\overline{r} \leq r$. If $\overline{r} < r$, from the fact that α is the $2r$-restricted smallest eigenvalue of $\mathcal{A}^*\mathcal{A}$ and $\min_{\mathrm{rank}(X)\leq\overline{r}} \|X - M\|_F^2 = \sum_{i=\overline{r}+1}^{r}[\sigma_i(M)]^2$ it follows that

$$\begin{aligned}\nu f(\overline{X}) + \mathrm{rank}(\overline{X}) &\geq \frac{1}{2}\nu\alpha\big\|\overline{X} - M\big\|_F^2 + \mathrm{rank}(\overline{X})\\ &\geq \frac{1}{2}\nu\alpha(r-\overline{r})[\sigma_r(M)]^2 + \overline{r} > r,\end{aligned}$$

where the last inequality is due to $\alpha > \frac{2}{\nu\sigma_r^2(M)}$ and $r - \overline{r} > 0$. This gives a contradiction to the inequality (3.5). Consequently, $\mathrm{rank}(\overline{X}) = \overline{r} = r$, and $f(\overline{X}) = 0$ follows from (3.5). Together with $f(\overline{X}) \geq \frac{1}{2}\alpha\|\overline{X} - M\|^2$, we obtain $\overline{X} = M$. By the arbitrariness of $\overline{X} \in \mathcal{X}^* \cap \Omega_\kappa$, it follows that $\mathcal{X}^* \cap \Omega_\kappa = \{M\}$. The proof is completed. □

3.2 DC regularized factorizations

By Remark 3.3, one may achieve a global optimal solution of (1.2) by solving the $\ell_{2,0}$-norm regularized factorization model (3.2) or its balanced formulation (1.6). Write

$$F(U,V) := \nu f(UV^{\mathbb{T}}) + \frac{\mu}{4}\|U^{\mathbb{T}}U - V^{\mathbb{T}}V\|_F^2. \tag{3.6}$$

Fix an arbitrary $\phi\in\mathscr{L}$. By (3.1), the problem (1.6) is equivalent to the following problem

$$\begin{aligned}\min_{\substack{(U,u)\in\mathbb{R}^{m\times\kappa}\times\mathbb{R}^{\kappa}\\(V,v)\in\mathbb{R}^{n\times\kappa}\times\mathbb{R}^{\kappa}}} & F(U,V)+\frac{1}{2}\sum_{i=1}^{\kappa}\big(\phi(u_i)+\phi(v_i)\big)\\ \text{s.t.}\quad & 0\le u\le e,\ \langle e-u,\mathcal{G}(U)\rangle=0,\\ & 0\le v\le e,\ \langle e-v,\mathcal{G}(V)\rangle=0\end{aligned}\tag{3.7}$$

in the sense that if $(\overline{U},\overline{V})$ is a global optimal solution to (1.6), then $(\overline{U},\overline{V},\overline{u},\overline{v})$ with $\overline{u}=\max(\text{sign}(\mathcal{G}(\overline{U})),t_{\phi}^*e)$ and $\overline{v}=\max(\text{sign}(\mathcal{G}(\overline{V})),t_{\phi}^*e)$ is globally optimal to (3.7); and conversely, if $(\overline{U},\overline{V},\overline{u},\overline{v})$ is a global optimal solution of (3.7), then $(\overline{U},\overline{V})$ is globally optimal to (1.6). Furthermore, the problems (1.6) and (3.7) have the same optimal value. The problem (3.7) is an MPEC involving the equilibrium constraints $\langle e-u,\mathcal{G}(U)\rangle=0, e-u\ge 0$ and $\langle e-v,\mathcal{G}(V)\rangle=0, e-v\ge 0$. The equivalence between (1.6) and (3.7) reveals that the combinatorial property of $\|U\|_{2,0}$ and $\|V\|_{2,0}$ arises from the equilibrium constraints.

It is well known that handling nonconvex constraints is much harder than handling nonconvex objective functions. So, we consider the following penalized problem of (3.7)

$$\begin{aligned}\min_{\substack{(U,u)\in\mathbb{R}^{m\times\kappa}\times\mathbb{R}^{\kappa}\\(V,v)\in\mathbb{R}^{m\times\kappa}\times\mathbb{R}^{\kappa}}} & \Big\{F(U,V)+\frac{1}{2}\sum_{j=1}^{\kappa}\big[\phi(u_j)+\phi(v_j)+\rho(1-u_j)\|U_j\|+\rho(1-v_j)\|V_j\|\big]\Big\}\\ \text{s.t.}\quad & 0\le u\le e,\ 0\le v\le e.\end{aligned}\tag{3.8}$$

By using [32, Theorem 3.2], we can establish the following global exact penalty result.

Proposition 3.5. *Let $\phi\in\mathscr{L}$. If $\widetilde{f}(U,V):=f(UV^{\mathbb{T}})$ for $(U,V)\in\mathbb{R}^{m\times\kappa}\times\mathbb{R}^{n\times\kappa}$ is coercive and f is continuously differentiable in $\mathbb{R}^{m\times n}$, then there exists $\widehat{\rho}>0$ such that the problem* (3.8) *associated to each $\rho>\widehat{\rho}$ has the same optimal solution set as* (3.7) *does.*

Proof. By the coerciveness of $\widetilde{f}$, there exists a constant $\omega>0$ such that (3.7) and (3.8) are equivalent to their respective version in which the variables U and V are restricted to lie in $\mathbb{B}(0,\omega)$. Thus, the conclusion follows by [32, Theorem 3.2]. □

Recall the definition of ψ in (2.2). The problem (3.8) can be compactly written as

$$\min_{\substack{(U,u)\in\mathbb{R}^{m\times\kappa}\times\mathbb{R}^{\kappa}\\(V,v)\in\mathbb{R}^{m\times\kappa}\times\mathbb{R}^{\kappa}}}\Big\{F(U,V)+\frac{1}{2}\sum_{j=1}^{\kappa}\big[\psi(u_j)+\psi(v_j)+\rho(1-u_j)\|U_j\|+\rho(1-v_j)\|V_j\|\big]\Big\}$$

which, by introducing the function $\theta(t):=t-\psi^*(t)$ for $t\in\mathbb{R}$, is simplified as

$$\min_{U\in\mathbb{R}^{m\times\kappa},V\in\mathbb{R}^{n\times\kappa}}\Big\{\Theta_{\rho}(U,V):=F(U,V)+\frac{1}{2}\sum_{j=1}^{\kappa}\big[\theta(\rho\|U_j\|)+\theta(\rho\|V_j\|)\big]\Big\}.\tag{3.9}$$

Notice that $\sum_{j=1}^{\kappa}\big[\theta(\rho\|U_j\|)+\theta(\rho\|V_j\|)\big]$ is a DC function since $U_j\mapsto\psi^*(\rho\|U_j\|)$ is convex by the nondecreasing and convexity of ψ^*. By Theorem 3.5, the following result holds.

Corollary 3.6. *Let $\phi\in\mathscr{L}$. If the function $\widetilde{f}$ defined in Proposition 3.5 is coercive and f is continuously differentiable in $\mathbb{R}^{m\times n}$, then there exists $\widehat{\rho}>0$ such that the problem* (3.9) *associated to each $\rho>\widehat{\rho}$ has the same optimal solution set as* (1.6) *does.*

Combining the above discussions with Remark 3.3, we conclude that when f satisfies the requirement of Proposition 3.5 and $\mathcal{X}^*\cap\Omega_\kappa\neq\emptyset$, one may achieve a global optimal solution of (1.2) by solving the penalized problem (3.8) or its equivalent DC regularized surrogates (3.9). However, the function $\widetilde{f}$ associated to many f is not coercive, say, the least squares loss function in (1.3). In this case, it is natural to ask what conditions can ensure that the problem (3.8) is still a global exact penalty of (3.7). The following proposition provides such a condition when f is specified as in (1.3) with $b=\mathcal{A}(M)$.

Proposition 3.7. *Suppose that the function f is given by* (1.3) *with $b=\mathcal{A}(M)$ for a matrix M of rank r. If $\mathcal{X}^*\cap\Omega_\kappa\neq\emptyset$ and the $2r$-restricted smallest eigenvalue α of $\mathcal{A}^*\mathcal{A}$ satisfies $\alpha>\frac{2}{\nu\sigma_r^2(M)}$, then for each $\phi\in\mathscr{L}$ the problem* (3.8) *is a global exact penalty of* (3.7) *with threshold $\overline{\rho}:=\max\big(1,\frac{\sqrt{\nu}\|\mathcal{A}\|\sqrt{\kappa}}{\sqrt{\nu\alpha}\sigma_r(M)-\sqrt{2}}\sqrt{1+\frac{2\sqrt{r}}{\sqrt{\mu}}}\big)\phi'_{-}(1)$, and consequently the problem* (3.9) *associated to every $\rho>\overline{\rho}$ has the same global optimal solution set as* (1.6) *does.*

Proof. We first argue that for all $\rho>\overline{\rho}$ and $(U,V,u,v)\in\mathbb{R}^{m\times\kappa}\times\mathbb{R}^{n\times\kappa}\times[0,e]\times[0,e]$,

$$F(U,V)+\frac{1}{2}\sum_{j=1}^{\kappa}\big[\phi(u_j)+\rho(1-u_j)\|U_j\|+\phi(v_j)+\rho(1-v_j)\|V_j\|\big]\geq r. \tag{3.10}$$

Fix an arbitrary $\rho>\overline{\rho}$ and an arbitrary $(U,V,u,v)\in\mathbb{R}^{m\times\kappa}\times\mathbb{R}^{n\times\kappa}\times[0,e]\times[0,e]$. Write

$$J:=\big\{j\mid \phi(u_j)+\rho(1-u_j)\|U_j\|+\phi(v_j)+\rho(1-v_j)\|V_j\|\geq 2\big\}\ \text{and}\ \overline{J}=\{1,\ldots,\kappa\}\backslash J.$$

Clearly, if $|J|\geq r$ or $\frac{\mu}{4}\|U^{\mathbb{T}}U-V^{\mathbb{T}}V\|_F^2\geq r$, the stated conclusion automatically holds. We next consider the case that $|J|<r$ and $\frac{\mu}{4}\|U^{\mathbb{T}}U-V^{\mathbb{T}}V\|_F^2<r$. Notice that $(\overline{U},\overline{V},e,e)$ for any $(\overline{U},\overline{V})\in\mathcal{W}^*$ is a feasible solution of (3.7) with the objective value equal to κ, while the optimal value of (3.7) is r. Hence, $r\leq\kappa$, and we have $\overline{J}\neq\emptyset$. For each $j\in\overline{J}$,

$$\|U_jV_j^{\mathbb{T}}\|_F=\|U_j\|\|V_j\|\leq\frac{\phi'_{-}(1)}{\rho}\sqrt{\Big(\frac{\phi'_{-}(1)}{\rho}\Big)^2+\frac{2\sqrt{r}}{\sqrt{\mu}}}. \tag{3.11}$$

Indeed, for each $j\in\overline{J}$, it holds that $\phi(u_j)+\rho(1-u_j)\|U_j\|<1$ or $\phi(v_j)+\rho(1-v_j)\|V_j\|<1$. Together with $\phi(u_j)-1=\phi(u_j)-\phi(1)\geq\phi'_{-}(1)(u_j-1)$ and $\phi(v_j)-1\geq\phi'_{-}(1)(v_j-1)$, it follows that $\rho\|U_j\|<\phi'_{-}(1)$ or $\rho\|V_j\|<\phi'_{-}(1)$. Notice that $|\|U_j\|^2-\|V_j\|^2|<\frac{2\sqrt{r}}{\sqrt{\mu}}$ since $\frac{\mu}{4}\|U^{\mathbb{T}}U-V^{\mathbb{T}}V\|_F^2<r$. Then, the inequality (3.11) follows. Thus, we have

$$\begin{aligned}\big\|\mathcal{A}(UV^{\mathbb{T}}-M)\big\|&\geq\big\|\mathcal{A}(U_JV_J^{\mathbb{T}}-M)\big\|-\big\|\mathcal{A}(U_{\overline{J}}V_{\overline{J}}^{\mathbb{T}})\big\|\\&\geq\sqrt{\alpha}\big\|U_JV_J^{\mathbb{T}}-M\big\|_F-\|\mathcal{A}\|\big\|U_{\overline{J}}V_{\overline{J}}^{\mathbb{T}}\big\|_F\\&\geq\sqrt{\alpha(r-|J|)}\sigma_r(M)-\sqrt{\kappa}\|\mathcal{A}\|\max_{1\leq j\leq\kappa}\big\|U_jV_j^{\mathbb{T}}\big\|_F\\&\geq\sqrt{\alpha(r-|J|)}\sigma_r(M)-\frac{\|\mathcal{A}\|\sqrt{\kappa}\phi'_{-}(1)}{\rho}\sqrt{\Big(\frac{\phi'_{-}(1)}{\rho}\Big)^2+\frac{2\sqrt{r}}{\sqrt{\mu}}}\\&>\sqrt{\frac{2(r-|J|)}{\nu}}\end{aligned}$$

where the third inequality is using $\min_{\mathrm{rank}(X)\leq|J|}\|X-M\|_F^2=\sum_{i=|J|+1}^{r}\sigma_i^2(M)$, and the last one is due to $\rho>\phi'_{-}(1)$ and $\rho\big(\sqrt{\alpha}\sigma_r(M)-\sqrt{2\nu^{-1}}\big)>\|\mathcal{A}\|\sqrt{\kappa}\phi'_{-}(1)\sqrt{1+\frac{2\sqrt{r}}{\sqrt{\mu}}}$. The last inequality implies that (3.10) holds. Recall that the optimal value of (3.7) equals r. By [32, Definition 2.2], the inequality (3.10) implies that the MPEC (3.7) is uniformly partial calm over its global optimal solution set, which by [32, Proposition 2.1(a)] is equivalent to saying that (3.8) is a global exact penalty of (3.7). □

From the proof of Proposition 3.7, we see that the balanced term $\frac{\mu}{4}\|U^{\mathbb{T}}U - V^{\mathbb{T}}V\|_F^2$ in the function F plays a crucial role. When replacing F with $\widetilde{f}$, it is unclear whether the result of Proposition 3.7 holds or not, and we leave this topic for future study.

4 KL Property of Exponent $1/2$ of Ψ and Θ_ρ

In this section, for the function f specified as in (1.3) with $b = \mathcal{A}(M)$ for a matrix M of rank r, we shall establish the KL property of exponent $1/2$ for the functions Ψ and Θ_ρ over the set of their global minimizers. This often requires the following inequality

$$\sigma_\kappa(A)\|B\|_F \le \|AB^{\mathbb{T}}\|_F \le \|A\|\|B\|_F \quad \forall\, A \in \mathbb{R}^{m\times\kappa},\ B \in \mathbb{R}^{n\times\kappa}. \tag{4.1}$$

For convenience, in the subsequent analysis we write $\sigma_i = \sigma_i(M)$ for $i = 1, 2, \ldots, m$.

4.1 KL property of exponent $1/2$ of Ψ

To establish the KL property of exponent $1/2$ of Ψ over the set of its global minimizer set, we first characterize the subdifferential of Ψ at any point $(U, V) \in \mathbb{R}^{m\times\kappa} \times \mathbb{R}^{n\times\kappa}$.

Lemma 4.1. *Fix an arbitrary $(U, V) \in \mathbb{R}^{m\times\kappa} \times \mathbb{R}^{n\times\kappa}$ and write $J_U := \{j \mid U_j \ne 0\}$ and $J_V := \{j \mid V_j \ne 0\}$. Then, $\widehat{\partial}\Psi(U, V) = \partial\Psi(U, V) = \partial_U\Psi(U, V) \times \partial_V\Psi(U, V)$ with*

$$\partial_U\Psi(U,V) = \Big\{G \in \mathbb{R}^{m\times\kappa} \mid G_j = \nu[\mathcal{A}^*\mathcal{A}(UV^{\mathbb{T}} - M)]V_j + \mu U(U^{\mathbb{T}}U_j - V^{\mathbb{T}}V_j),\ j \in J_U\Big\},$$
$$\partial_V\Psi(U,V) = \Big\{H \in \mathbb{R}^{n\times\kappa} \mid H_j = \nu[\mathcal{A}^*\mathcal{A}(UV^{\mathbb{T}} - M)]^{\mathbb{T}}U_j + \mu V(V^{\mathbb{T}}V_j - U^{\mathbb{T}}U_j),\ j \in J_V\Big\}.$$

Proof. By the expression of F in (3.6), F is continuously differentiable in $\mathbb{R}^{m\times\kappa} \times \mathbb{R}^{n\times\kappa}$. Let $g(Z) := \|Z\|_{2,0}$ for $Z \in \mathbb{R}^{m\times\kappa}$, and $h(L) := \|L\|_{2,0}$ for $L \in \mathbb{R}^{n\times\kappa}$. From [41, Exercise 8.8(c)& Proposition 10.5],

$$\widehat{\partial}\Psi(U,V) = \widehat{\partial}_U\Psi(U,V) \times \widehat{\partial}_V\Psi(U,V) = \big(\nabla_U F(U,V) + \widehat{\partial}g(U)\big) \times \big(\nabla_V F(U,V) + \widehat{\partial}h(V)\big),$$
$$\partial\Psi(U,V) = \partial_U\Psi(U,V) \times \partial_V\Psi(U,V) = \big(\nabla_U F(U,V) + \partial g(U)\big) \times \big(\nabla_V F(U,V) + \partial h(V)\big).$$

By invoking [41, Proposition 10.5] and Lemma 2.3, it immediately follows that

$$\widehat{\partial}g(U) = \partial g(U) = S_1 \times \cdots \times S_\kappa \ \text{ with } S_j = \begin{cases} \{0\}^m & \text{if } j \in J_U; \\ \mathbb{R}^m & \text{if } j \notin J_U. \end{cases}$$

Similarly, $\widehat{\partial}h(V) = \partial h(V)$ has such a characterization. Thus, we get the result. □

The following lemma provides a kind of stability for the global minimizer set of Ψ, i.e., for each global minimizer $(\overline{U}, \overline{V})$ of Ψ, there exists a neighborhood such that every point pair in this neighborhood has the same nonzero columns.

Lemma 4.2. *Suppose that $\mathcal{X}^*\cap\Omega_\kappa\neq\emptyset$ and the $2r$-restricted smallest eigenvalue α of $\mathcal{A}^*\mathcal{A}$ satisfies $\alpha>\frac{2}{\nu\sigma_r^2}$. Fix an arbitrary $(\overline{U},\overline{V})\in\mathcal{W}^*$. Then, for any $(U,V)\in\mathbb{R}^{m\times\kappa}\times\mathbb{R}^{n\times\kappa}$ with $(U,V)\in\mathbb{B}((\overline{U},\overline{V}),\frac{\sqrt{\sigma_r}}{4})$ and $0<\Psi(U,V)-\Psi(\overline{U},\overline{V})<\frac{1}{2}$, it holds that*

$$\big\{j\mid \overline{U}_j\neq 0\big\}=\big\{j\mid \overline{V}_j\neq 0\big\}=\big\{j\mid U_j\neq 0\big\}=\big\{j\mid V_j\neq 0\big\}.$$

Proof. Write $J:=\big\{j\mid \overline{U}_j\neq 0\big\}, J_U:=\{j\mid U_j\neq 0\}$ and $J_V:=\{j\mid V_j\neq 0\}$. By using Proposition 3.4, we have $J=\big\{j\mid \overline{V}_j\neq 0\big\}$, $\|\overline{U}_j\|=\|\overline{V}_j\|$ for each $j\in J$, and $|J|=r$. Let $\{e_1,\ldots,e_r\}$ be the orthonormal basis of $\mathbb{R}^r$. Then, for each $j\in J$, we have

$$\|\overline{U}_j\|^2\geq\min_{1\leq i\leq r}(\overline{U}_J^{\mathbb{T}}\overline{U}_J)_{ii}=\min_{1\leq i\leq r}\langle e_ie_i^{\mathbb{T}},\overline{U}_J^{\mathbb{T}}\overline{U}_J\rangle\geq\sigma_r(\overline{U}_J^{\mathbb{T}}\overline{U}_J)=\sigma_r(\overline{U}^{\mathbb{T}}\overline{U}).$$

This implies that $\min_{j\in J}\|\overline{U}_j\|\geq\sigma_r(\overline{U})=\sqrt{\sigma_r}$, where the equality is using $\overline{U}^{\mathbb{T}}\overline{U}=\overline{V}^{\mathbb{T}}\overline{V}$ and $\overline{U}\overline{V}^{\mathbb{T}}=M$. Together with $\|\overline{U}_j\|=\|\overline{V}_j\|$ for each $j\in J$,

$$\min_{j\in J}\|\overline{U}_j\|=\min_{j\in J}\|\overline{V}_j\|\geq\sqrt{\sigma_r}.$$

For each $j\in J$, since $\|U_j\|=\|U_j-\overline{U}_j+\overline{U}_j\|\geq\|\overline{U}_j\|-\|U-\overline{U}\|_F\geq\sqrt{\sigma_r}-\frac{\sqrt{\sigma_r}}{4}>\frac{1}{2}\sqrt{\sigma_r}$, we have $J\subseteq J_U$. Similarly, $J\subseteq J_V$. In addition, from $0<\Psi(U,V)-\Psi(\overline{U},\overline{V})<1/2$ and $\Psi(\overline{U},\overline{V})=r$, it follows that $r<\Psi(U,V)<r+\frac{1}{2}$. By the expression of Ψ, we have $r=|J|\leq\frac{1}{2}(\|U\|_{2,0}+\|V\|_{2,0})<\Psi(U,V)<r+\frac{1}{2}$, which implies that $\|U\|_{2,0}+\|V\|_{2,0}=2r$. Together with $J\subseteq J_U$ and $J\subseteq J_V$, we conclude that $J=J_U=J_V$. □

Theorem 4.3. *Let α and β be the $2r$-restricted smallest and largest eigenvalues of $\mathcal{A}^*\mathcal{A}$, respectively. Suppose that $\mathcal{X}^*\cap\Omega_\kappa\neq\emptyset$ and $\frac{\beta}{\alpha}<\frac{128\sigma_1^2(4\sigma_1+\sigma_r)^2+\sigma_r^4}{128\sigma_1^2(4\sigma_1+\sigma_r)^2-\sigma_r^4}$ with $\alpha>\frac{4}{\nu\sigma_r^2}$. Fix an arbitrary $(\overline{U},\overline{V})\in\mathcal{W}^*$. Then, for any (U,V) with $(U,V)\in\mathbb{B}((\overline{U},\overline{V}),\frac{\sqrt{\sigma_r}}{4})$ and $0<\Psi(U,V)-\Psi(\overline{U},\overline{V})<\frac{1}{2}$, the following inequality holds*

$$\mathrm{dist}^2(0,\partial\Psi(U,V))\geq\gamma\big[\Psi(U,V)-\Psi(\overline{U},\overline{V})\big],\tag{4.3}$$

with $\gamma=\min\big(\frac{\nu}{\beta}\big[\frac{(\beta+\alpha)\sigma_r^4}{128\sqrt{\sigma_1^3}(4\sigma_1+\sigma_r)^2}-\sqrt{\sigma_1}(\beta-\alpha)\big]^2,2\mu\sigma_r\big)$.

Proof. Fix an arbitrary (U,V) with $(U,V)\in\mathbb{B}((\overline{U},\overline{V}),\frac{\sqrt{\sigma_r}}{4})$ and $0<\Psi(U,V)-\Psi(\overline{U},\overline{V})<\frac{1}{2}$. Write $\Delta_U=U-\overline{U}$ and $\Delta_V=V-\overline{V}$. Then $\max(\|\Delta_U\|_F,\ \|\Delta_V]\|_F)<\frac{1}{4}\sqrt{\sigma_r}$. Let J,J_U and J_V be the index sets defined as in the proof of Lemma 4.2. By Lemma 4.1,

$$\begin{aligned}\mathrm{dist}^2(0,\partial\Psi(U,V))&=\mathrm{dist}^2(0,\partial_U\Psi(U,V))+\mathrm{dist}^2(0,\partial_V\Psi(U,V))\\&=\nu^2\big[\|[\mathcal{A}^*\mathcal{A}(UV^{\mathbb{T}}-M)]V\|_F^2+\|U^{\mathbb{T}}[\mathcal{A}^*\mathcal{A}(UV^{\mathbb{T}}-M)]\|_F^2\big]\\&\quad+\mu^2\big[\|U(U^{\mathbb{T}}U-V^{\mathbb{T}}V)\|_F^2+\|(V^{\mathbb{T}}V-U^{\mathbb{T}}U)V^{\mathbb{T}}\|_F^2\big].\end{aligned}\tag{4.4}$$

Since $J_U=J$ by Lemma 4.2, we have $\|U(U^{\mathbb{T}}U-V^{\mathbb{T}}V)\|_F^2=\|U_J(U_J^{\mathbb{T}}U_J-V_J^{\mathbb{T}}V_J)\|_F^2$. Together with (4.1), we have $\|U(U^{\mathbb{T}}U-V^{\mathbb{T}}V)\|_F^2\geq\sigma_r^2(U_J)\|U_J^{\mathbb{T}}U_J-V_J^{\mathbb{T}}V_J\|_F^2$. Notice that $\sigma_r(U_J)=\sigma_r(\overline{U}_J+[\Delta_U]_J)\geq\sigma_r(\overline{U}_J)-\sigma_1(\Delta_U)>\sqrt{\sigma_r}-\frac{\sqrt{\sigma_r}}{4}>\sqrt{\sigma_r/2}$. Hence,

$$\|U(U^{\mathbb{T}}U-V^{\mathbb{T}}V)\|_F^2\geq\frac{\sigma_r}{2}\|U_J^{\mathbb{T}}U_J-V_J^{\mathbb{T}}V_J\|_F^2=\frac{\sigma_r}{2}\|U^{\mathbb{T}}U-V^{\mathbb{T}}V\|_F^2.$$

Following the same arguments, we also have $\|(V^{\mathbb{T}}V-U^{\mathbb{T}}U)V^{\mathbb{T}}\|_F^2\geq\frac{\sigma_r}{2}\|U^{\mathbb{T}}U-V^{\mathbb{T}}V\|_F^2$. By combining with these inequalities and the inequality (4.4), it follows that

$$\mathrm{dist}^2(0,\partial\Psi(U,V))\geq\mu^2\sigma_r\|U^{\mathbb{T}}U-V^{\mathbb{T}}V\|_F^2. \tag{4.5}$$

Next we proceed the arguments by two cases $UV^{\mathbb{T}}=M$ and $UV^{\mathbb{T}}\neq M$, respectively.

Case 1: $UV^{\mathbb{T}}=M$. In this case, $\Psi(U,V)-\Psi(\overline{U},\overline{V})=\frac{\mu}{4}\|U^{\mathbb{T}}U-V^{\mathbb{T}}V\|_F^2$. Together with the inequality (4.5), the desired inequality holds with $\gamma=4\mu\sigma_r$.

Case 2: $UV^{\mathbb{T}}\neq M$. Since $\|U\|\leq\|\overline{U}\|+\|\Delta_U\|\leq\frac{5}{4}\sqrt{\sigma_1}\leq\sqrt{2\sigma_1}$ and $\|V\|\leq\sqrt{2\sigma_1}$, we have

$$\begin{aligned}
&2\sqrt{\sigma_1}\sqrt{\|[\mathcal{A}^*\mathcal{A}(UV^{\mathbb{T}}-M)]V\|_F^2+\|U^{\mathbb{T}}[\mathcal{A}^*\mathcal{A}(UV^{\mathbb{T}}-M)]\|_F^2}\\
&\geq\sqrt{2\sigma_1}\big[\|[\mathcal{A}^*\mathcal{A}(UV^{\mathbb{T}}-M)]V\|_F+\|U^{\mathbb{T}}[\mathcal{A}^*\mathcal{A}(UV^{\mathbb{T}}-M)]\|_F\big]\\
&\geq\big[\|[\mathcal{A}^*\mathcal{A}(UV^{\mathbb{T}}-M)]VV^{\mathbb{T}}\|_F+\|UU^{\mathbb{T}}[\mathcal{A}^*\mathcal{A}(UV^{\mathbb{T}}-M)]\|_F\big]\\
&\geq\frac{1}{\|UV^{\mathbb{T}}-M\|_F}\langle[\mathcal{A}^*\mathcal{A}(UV^{\mathbb{T}}-M)]VV^{\mathbb{T}},UV^{\mathbb{T}}-M\rangle\\
&\quad+\frac{1}{\|UV^{\mathbb{T}}-M\|_F}\langle UU^{\mathbb{T}}[\mathcal{A}^*\mathcal{A}(UV^{\mathbb{T}}-M)],UV^{\mathbb{T}}-M\rangle\\
&=\frac{1}{\|UV^{\mathbb{T}}-M\|_F}\langle\mathcal{A}(UV^{\mathbb{T}}-M),\mathcal{A}[(UV^{\mathbb{T}}-M)VV^{\mathbb{T}}]\rangle\\
&\quad+\frac{1}{\|UV^{\mathbb{T}}-M\|_F}\langle\mathcal{A}(UV^{\mathbb{T}}-M),\mathcal{A}[UU^{\mathbb{T}}(UV^{\mathbb{T}}-M)]\rangle.
\end{aligned}$$

Notice that $\mathrm{rank}([UV^{\mathbb{T}}-M\ \ (UV^{\mathbb{T}}-M)VV^{\mathbb{T}}])\leq 2r$ and $\mathrm{rank}([UV^{\mathbb{T}}-M\ \ UU^{\mathbb{T}}(UV^{\mathbb{T}}-M)])\leq 2r$. Applying Lemma 2.8 to the two terms on the right hand side, we obtain

$$\begin{aligned}
&2\sqrt{\sigma_1}\sqrt{\|[\mathcal{A}^*\mathcal{A}(UV^{\mathbb{T}}-M)]V\|_F^2+\|U^{\mathbb{T}}[\mathcal{A}^*\mathcal{A}(UV^{\mathbb{T}}-M)]\|_F^2}\\
&\geq\frac{\beta+\alpha}{2\|UV^{\mathbb{T}}-M\|_F}\langle UV^{\mathbb{T}}-M,(UV^{\mathbb{T}}-M)VV^{\mathbb{T}}\rangle-\frac{\beta-\alpha}{2}\|(UV^{\mathbb{T}}-M)VV^{\mathbb{T}}\|_F\\
&\quad+\frac{\beta+\alpha}{2\|UV^{\mathbb{T}}-M\|_F}\langle UV^{\mathbb{T}}-M,UU^{\mathbb{T}}(UV^{\mathbb{T}}-M)\rangle-\frac{\beta-\alpha}{2}\|UU^{\mathbb{T}}(UV^{\mathbb{T}}-M)\|_F\\
&\geq\frac{(\beta+\alpha)[\|(UV^{\mathbb{T}}-M)V\|_F^2+\|U^{\mathbb{T}}(UV^{\mathbb{T}}-M)\|_F^2]}{2\|UV^{\mathbb{T}}-M\|_F}-2\sigma_1(\beta-\alpha)\|UV^{\mathbb{T}}-M\|_F
\end{aligned}$$

where the last inequality is due to (4.1), $\|U\|\leq\sqrt{2\sigma_1}$ and $\|V\|\leq\sqrt{2\sigma_1}$. Along with (4.4),

$$\begin{aligned}
&\frac{4\sqrt{\sigma_1}}{\nu(\beta+\alpha)}\mathrm{dist}(0,\partial\Psi(U,V))\\
&\geq\frac{\|(UV^{\mathbb{T}}-M)V\|_F^2+\|U^{\mathbb{T}}(UV^{\mathbb{T}}-M)\|_F^2}{\|UV^{\mathbb{T}}-M\|_F}-\frac{4\sigma_1(\beta-\alpha)}{\beta+\alpha}\|UV^{\mathbb{T}}-M\|_F.
\end{aligned}\tag{4.6}$$

To further deal with the right hand side of (4.6), we take $(P,Q)\in\mathbb{O}^{m,n}(M)$ and write

$$\Sigma_+=\mathrm{diag}(\sigma_1,\ldots,\sigma_r),\ \widehat{U}=P^{\mathbb{T}}U,\ \widehat{V}=Q^{\mathbb{T}}V,\ \widehat{U}_J=\begin{pmatrix}\widehat{U}_{IJ}\\ \widehat{U}_{\overline{I}J}\end{pmatrix}\ \text{and}\ \widehat{V}_J=\begin{pmatrix}\widehat{V}_{IJ}\\ \widehat{V}_{\overline{I}J}\end{pmatrix}$$

where $\widehat{U}_{IJ} \in \mathbb{R}^{r\times r}$ and $\widehat{U}_{\overline{I}J} \in \mathbb{R}^{(m-r)\times r}$ are the matrix consisting of the first r rows and the last $m-r$ rows of $\widehat{U}_J$, respectively, and similar is for $\widehat{V}_{IJ} \in \mathbb{R}^{r\times r}$ and $\widehat{V}_{\overline{I}J} \in \mathbb{R}^{(n-r)\times r}$. Then, by using (4.1), $\|\widehat{U}\| = \|U\| \le \sqrt{2\sigma_1}$ and $\|\widehat{V}\| = \|V\| \le \sqrt{2\sigma_1}$, we obtain

$$\begin{aligned} \|UV^{\mathbb{T}} - M\|_F = \|\widehat{U}\widehat{V}^{\mathbb{T}} - \mathrm{Diag}(\sigma(M))\|_F &= \left\| \begin{pmatrix} \widehat{U}_{IJ}\widehat{V}_{IJ}^{\mathbb{T}} - \Sigma_+ & \widehat{U}_{IJ}\widehat{V}_{\overline{I}J}^{\mathbb{T}} \\ \widehat{U}_{\overline{I}J}\widehat{V}_{IJ}^{\mathbb{T}} & \widehat{U}_{\overline{I}J}\widehat{V}_{\overline{I}J}^{\mathbb{T}} \end{pmatrix} \right\|_F \\ &\le \|\widehat{U}_{IJ}\widehat{V}_{IJ}^{\mathbb{T}} - \Sigma_+\|_F + \sqrt{2\sigma_1}\|\widehat{U}_{\overline{I}J}\|_F + \sqrt{2\sigma_1}\|\widehat{V}_{\overline{I}J}\|_F. \end{aligned} \tag{4.7}$$

Notice that $\sigma_r(\widehat{V}_J) = \sigma_r(V_J) > \sqrt{\frac{\sigma_r}{2}}$ and $\mathrm{rank}(\widehat{V}_J) = \mathrm{rank}(V_J) = r$. Hence, it holds that

$$\begin{aligned} \|(UV^{\mathbb{T}} - M)V\|_F = \|(\widehat{U}\widehat{V}^{\mathbb{T}} - \mathrm{Diag}(\sigma(M)))\widehat{V}\|_F &= \left\| \begin{pmatrix} (\widehat{U}_{IJ}\widehat{V}_{IJ}^{\mathbb{T}} - \Sigma_+)\widehat{V}_{IJ} + \widehat{U}_{IJ}\widehat{V}_{\overline{I}J}^{\mathbb{T}}\widehat{V}_{\overline{I}J} \\ \widehat{U}_{\overline{I}J}\widehat{V}_J^{\mathbb{T}}\widehat{V}_J \end{pmatrix} \right\|_F \\ &\ge \frac{1}{\sqrt{2}}\big[\|(\widehat{U}_{IJ}\widehat{V}_{IJ}^{\mathbb{T}} - \Sigma_+)\widehat{V}_{IJ} + \widehat{U}_{IJ}\widehat{V}_{\overline{I}J}^{\mathbb{T}}\widehat{V}_{\overline{I}J}\|_F + \|\widehat{U}_{\overline{I}J}\widehat{V}_J^{\mathbb{T}}\widehat{V}_J\|_F\big] \\ &\ge \frac{\sigma_r}{2\sqrt{2}}\|\widehat{U}_{\overline{I}J}\|_F + \frac{1}{\sqrt{2}}\big\|(\widehat{U}_{IJ}\widehat{V}_{IJ}^{\mathbb{T}} - \Sigma_+)\widehat{V}_{IJ} + \widehat{U}_{IJ}\widehat{V}_{\overline{I}J}^{\mathbb{T}}\widehat{V}_{\overline{I}J}\big\|_F. \end{aligned} \tag{4.8}$$

Since $\|\widehat{U}_{IJ}\widehat{V}_{IJ}^{\mathbb{T}} - \Sigma_+\|_F \le \|\widehat{U}\widehat{V}^{\mathbb{T}} - \mathrm{Diag}(\sigma(M))\|_F = \|UV^{\mathbb{T}} - M\|_F \le \frac{1}{\sqrt{\alpha}}\|\mathcal{A}(UV^{\mathbb{T}} - M)\| \le \frac{1}{\sqrt{\alpha\nu}}$ where the last inequality is due to $\Psi(U,V) - \Psi(\overline{U},\overline{V}) < 1/2$, using $\alpha \ge \frac{4}{\nu\sigma_r^2}$ then yields $\|\widehat{U}_{IJ}\widehat{V}_{IJ}^{\mathbb{T}} - \Sigma_+\|_F \le \frac{\sigma_r}{2}$, which in turn implies that $\sigma_r(\widehat{U}_{IJ}\widehat{V}_{IJ}^{\mathbb{T}}) \ge \frac{\sigma_r}{2}$. Consequently,

$$\begin{aligned} \|(UV^{\mathbb{T}} - M)V\|_F &\ge \frac{\sigma_r}{2\sqrt{2}}\|\widehat{U}_{\overline{I}J}\|_F + \frac{1}{\sqrt{2}}\big\|(\widehat{U}_{IJ}\widehat{V}_{IJ}^{\mathbb{T}} - \Sigma_+)\widehat{V}_{IJ}\big\|_F - \frac{1}{\sqrt{2}}\big\|\widehat{U}_{IJ}\widehat{V}_{\overline{I}J}^{\mathbb{T}}\widehat{V}_{\overline{I}J}\big\|_F \\ &\ge \frac{\sigma_r}{2\sqrt{2}}\|\widehat{U}_{\overline{I}J}\|_F + \frac{1}{\sqrt{2}\|\widehat{U}_{IJ}\|}\big\|(\widehat{U}_{IJ}\widehat{V}_{IJ}^{\mathbb{T}} - \Sigma_+)\widehat{V}_{IJ}\widehat{U}_{IJ}^{\mathbb{T}}\|_F - \frac{\|\widehat{U}_{IJ}\widehat{V}_{\overline{I}J}^{\mathbb{T}}\widehat{V}_{\overline{I}J}\|_F}{\sqrt{2}} \\ &\ge \frac{\sigma_r}{2\sqrt{2}}\|\widehat{U}_{\overline{I}J}\|_F + \frac{\sigma_r}{4\sqrt{\sigma_1}}\|\widehat{U}_{IJ}\widehat{V}_{IJ}^{\mathbb{T}} - \Sigma_+\|_F - \sqrt{2}\sigma_1\|\widehat{V}_{\overline{I}J}^{\mathbb{T}}\|_F \end{aligned} \tag{4.9}$$

where the last inequality is also using $\|\widehat{U}_{IJ}\| \le \sqrt{2\sigma_1}$ and $\|\widehat{V}_{\overline{I}J}\| \le \sqrt{2\sigma_1}$. Similarly,

$$\|U^{\mathbb{T}}(UV^{\mathbb{T}} - M)\|_F \ge \frac{\sigma_r}{2\sqrt{2}}\|\widehat{V}_{\overline{I}J}\|_F + \frac{1}{\sqrt{2}}\big\|\widehat{U}_{IJ}^{\mathbb{T}}(\widehat{U}_{IJ}\widehat{V}_{IJ}^{\mathbb{T}} - \Sigma_+) + \widehat{U}_{\overline{I}J}^{\mathbb{T}}\widehat{U}_{\overline{I}J}\widehat{V}_{IJ}^{\mathbb{T}}\big\|_F, \tag{4.10a}$$

$$\|U^{\mathbb{T}}(UV^{\mathbb{T}} - M)\|_F \ge \frac{\sigma_r}{2\sqrt{2}}\|\widehat{V}_{\overline{I}J}\|_F + \frac{\sigma_r}{4\sqrt{\sigma_1}}\|\widehat{U}_{IJ}\widehat{V}_{IJ}^{\mathbb{T}} - \Sigma_+\|_F - \sqrt{2}\sigma_1\|\widehat{U}_{\overline{I}J}\|_F. \tag{4.10b}$$

From (4.8), we have $\|\widehat{U}_{\overline{I}J}\|_F \le \frac{2\sqrt{2}}{\sigma_r}\|(UV^{\mathbb{T}} - M)V\|_F$. Together with (4.10b), it follows that

$$\frac{4\sigma_1}{\sigma_r}\|(UV^{\mathbb{T}} - M)V\|_F + \|U^{\mathbb{T}}(UV^{\mathbb{T}} - M)\|_F \ge \frac{\sigma_r}{2\sqrt{2}}\|\widehat{V}_{\overline{I}J}\|_F + \frac{\sigma_r}{4\sqrt{\sigma_1}}\|\widehat{U}_{IJ}\widehat{V}_{IJ}^{\mathbb{T}} - \Sigma_+\|_F.$$

Similarly, from (4.10a) and (4.9), we get

$$\|(UV^{\mathbb{T}} - M)V\|_F + \frac{4\sigma_1}{\sigma_r}\|U^{\mathbb{T}}(UV^{\mathbb{T}} - M)\|_F \ge \frac{\sigma_r}{2\sqrt{2}}\|\widehat{U}_{\overline{I}J}\|_F + \frac{\sigma_r}{4\sqrt{\sigma_1}}\|\widehat{U}_{IJ}\widehat{V}_{IJ}^{\mathbb{T}} - \Sigma_+\|_F.$$

Combining the last two inequalities, we have

$$\begin{aligned}&\Big(\frac{4\sigma_1}{\sigma_r}+1\Big)\big[\|U^{\mathbb{T}}(UV^{\mathbb{T}}-M)\|_F+\|(UV^{\mathbb{T}}-M)V\|_F\big]\\ &\ge \frac{\sigma_r}{2\sqrt{2}}\|\widehat{V}_{\overline{I}J}\|_F+\frac{\sigma_r}{2\sqrt{\sigma_1}}\|\widehat{U}_{IJ}\widehat{V}_{IJ}^{\mathbb{T}}-\Sigma_+\|_F+\frac{\sigma_r}{2\sqrt{2}}\|\widehat{U}_{\overline{I}J}\|_F\\ &\ge \frac{\sigma_r}{4\sqrt{\sigma_1}}\|UV^{\mathbb{T}}-M\|_F,\end{aligned}$$

where the last inequality is due to (4.7). From this, we immediately obtain that

$$\begin{aligned}&\|U^{\mathbb{T}}(UV^{\mathbb{T}}-M)\|_F^2+\|(UV^{\mathbb{T}}-M)V\|_F^2\\ &\ge \frac{1}{2}\big[\|U^{\mathbb{T}}(UV^{\mathbb{T}}-M)\|_F+\|(UV^{\mathbb{T}}-M)V\|_F\big]^2\\ &\ge \frac{\sigma_r^4}{32\sigma_1(4\sigma_1+\sigma_r)^2}\|UV^{\mathbb{T}}-M\|_F^2.\end{aligned}\tag{4.11}$$

Combining the last inequality and (4.6) and using $\frac{1}{\sqrt{\beta}}\|\mathcal{A}(UV^{\mathbb{T}}-M)\|\le\|UV^{\mathbb{T}}-M\|_F$ gives

$$\begin{aligned}\frac{4\sqrt{\sigma_1}}{\nu(\beta+\alpha)}\mathrm{dist}(0,\partial\Psi(U,V))&\ge\Big[\frac{\sigma_r^4}{32\sigma_1(4\sigma_1+\sigma_r)^2}-\frac{4\sigma_1(\beta-\alpha)}{\beta+\alpha}\Big]\|UV^{\mathbb{T}}-M\|_F\\ &\ge\Big[\frac{\sigma_r^4}{32\sigma_1(4\sigma_1+\sigma_r)^2}-\frac{4\sigma_1(\beta-\alpha)}{\beta+\alpha}\Big]\frac{1}{\sqrt{\beta}}\|\mathcal{A}(UV^{\mathbb{T}}-M)\|\end{aligned}$$

where the last inequality is using $\frac{\beta}{\alpha}<\frac{128\sigma_1^2(4\sigma_1+\sigma_r)^2+\sigma_r^4}{128\sigma_1^2(4\sigma_1+\sigma_r)^2-\sigma_r^4}$. Together with (4.5), we have

$$\begin{aligned}\mathrm{dist}^2(0,\partial\Psi(U,V))&\ge\frac{\nu^2}{2\beta}\Big[\frac{(\beta+\alpha)\sigma_r^4}{128\sigma_1\sqrt{\sigma_1}(4\sigma_1+\sigma_r)^2}-\sqrt{\sigma_1}(\beta-\alpha)\Big]^2\|\mathcal{A}(UV^{\mathbb{T}}-M)\|^2\\ &\quad+\frac{\mu^2\sigma_r}{2}\|U^{\mathbb{T}}U-V^{\mathbb{T}}V\|_F^2\\ &\ge\gamma\Big[\frac{\nu}{2}\|\mathcal{A}(UV^{\mathbb{T}}-M)\|^2+\frac{\mu}{4}\|U^{\mathbb{T}}U-V^{\mathbb{T}}V\|_F^2\Big].\end{aligned}$$

Notice that $\Psi(U,V)-\Psi(\overline{U},\overline{V})=\frac{\nu}{2}\|\mathcal{A}(UV^{\mathbb{T}}-M)\|^2+\frac{\mu}{4}\|U^{\mathbb{T}}U-V^{\mathbb{T}}V\|_F^2$. Together with the result of Case 1, we obtain the desired conclusion. □

Remark 4.4. Theorem 4.3 establishes the KL property of exponent 1/2 of the function Ψ over the set of global minimizers under a suitable assumption on the $2r$-restricted condition number of $\mathcal{A}^*\mathcal{A}$. Along with Proposition 3.4, under the assumptions of Theorem 4.3 one may seek the unique global optimal solution M of rank not more than κ in a linear rate when solving the problem (1.6) with a starting point not far from the set $\mathcal{W}^*$.

Inspired by the result of Theorem 4.3, it is natural to ask whether the function Ψ has the KL property of exponent 1/2 in the set of critical points. The following example shows that Ψ does not have the KL property of exponent 1/2 at those critical points for which the number of nonzero columns is greater than the rank of M.

Example 4.5. Take $\mathcal{A}(X)=\mathrm{vec}(X)$ for $X\in\mathbb{R}^{4\times 4}$, where $\mathrm{vec}(X)$ represents the vector obtained by arranging X in terms of its columns, $M=4E$ with $E\in\mathbb{R}^{4\times 4}$ being a matrix of all ones, and $\kappa=4$. Now the problem (1.6) is specified as follows

$$\min_{U,V\in\mathbb{R}^{4\times 4}}\Big\{\Psi(U,V):=\frac{\nu}{2}\big\|UV^{\mathbb{T}}-M\big\|_F^2+\frac{\mu}{4}\|U^{\mathbb{T}}U-V^{\mathbb{T}}V\|_F^2+\frac{1}{2}(\|U\|_{2,0}+\|V\|_{2,0})\Big\}.$$

Consider $\overline{U} = \overline{V} = E$. It is easy to check that $(\overline{U}, \overline{V})$ is a regular critical point of Ψ. For any $t \in (0,1)$, define $U(t) = V(t) := \overline{U} + t\Delta$ with $\Delta \in \mathbb{R}^{4\times 4}$ defined as follows

$$\Delta = \begin{bmatrix} -E+2I & 0_{2\times 2} \\ 0_{2\times 2} & 0_{2\times 2} \end{bmatrix}.$$

We calculate that $\Psi(U(t),V(t)) - \Psi(\overline{U},\overline{V}) = 8\nu t^4$ and $\mathrm{dist}(0, \partial\Psi(U(t),V(t))) = 8\sqrt{2}\nu t^3$. This shows that Ψ can not have the KL property of exponent $1/2$ at those critical points with the number of nonzero columns greater than r.

4.2 KL property of exponent $1/2$ of Θ_ρ

In this part, we establish the KL property of exponent $1/2$ for the function Θ_ρ over its global minimizer set, where Θ_ρ is defined by (3.9) with θ satisfying Assumption 4.1 below. Such an assumption is mild and the functions in [32, Example 3-5] all satisfy it.

Assumption 4.1. The function $\theta(t) := t - \psi^*(t)$ with $\phi \in \mathscr{L}$ is concave and nondecreasing on $[0,+\infty)$, and there exist constants $\varpi > 0$ and $c > 0$ such that $\theta'(t) \ge c$ for all $t \in (0,\varpi)$.

To achieve the KL property of exponent of such Θ_ρ, we need the following lemma. Since its proof is similar to that of Lemma 4.1 by Lemma 2.4, we here omit it.

Lemma 4.6. *Let Θ_ρ be the function in equation* (3.9) *associated to $\rho > 0$. Fix an arbitrary $(U,V) \in \mathbb{R}^{m\times\kappa} \times \mathbb{R}^{n\times\kappa}$. Write $J_U := \{j \mid U_j \ne 0\}, J_V := \{j \mid V_j \ne 0\}, \overline{J}_U = \{1,\ldots,\kappa\}\backslash J_U$ and $\overline{J}_V = \{1,\ldots,\kappa\}\backslash J_V$. Then, $\partial\Theta_\rho(U,V) = \partial_U\Theta_\rho(U,V) \times \partial_V\Theta_\rho(U,V)$ with*

$$\begin{aligned}\partial_U\Theta_\rho(U,V) \subseteq \Big\{ & G \in \mathbb{R}^{m\times\kappa} \mid G_j = \nu[\mathcal{A}^*\mathcal{A}(UV^{\mathbb{T}} - M)]V_j + \mu U(U^{\mathbb{T}}U_j - V^{\mathbb{T}}V_j) + S_j, \\ & S_j = \frac{\theta'(\rho\|U_j\|)\rho U_j}{2\|U_j\|} \text{ for } j \in J_U, S_j \in \frac{\rho}{2}D^*g(0)[\partial\theta(0)] \text{ for } j \in \overline{J}_U \Big\}, \\ \partial_V\Theta_\rho(U,V) \subseteq \Big\{ & H \in \mathbb{R}^{n\times\kappa} \mid H_j = \nu[\mathcal{A}^*\mathcal{A}(UV^{\mathbb{T}} - M)]^{\mathbb{T}}U_j + \mu V(V^{\mathbb{T}}V_j - U^{\mathbb{T}}U_j) + T_j, \\ & T_j = \frac{\theta'(\|\rho V_j\|)\rho V_j}{2\|V_j\|} \text{ for } j \in J_V, T_j \in \frac{\rho}{2}D^*g(0)[\partial\theta(0)] \text{ for } j \in \overline{J}_V \Big\}\end{aligned}$$

where $g(z) = \|z\|$ for $z \in \mathbb{R}^m$ and $z \in \mathbb{R}^n$ respectively, and $D^*g(z)$ is the coderivative of g at z.

Theorem 4.7. *Let Θ_ρ be the function in* (3.9) *associated to $\rho > \overline{\rho}$ with θ satisfying Assumption 4.1 and $\overline{\rho} := \max\big(1, \frac{4\sqrt{\nu}\|\mathcal{A}\|\sqrt{\kappa}}{\sqrt{\nu\alpha}\sigma_r(M) - \sqrt{2}}\sqrt{1 + \frac{2\sqrt{r}}{\sqrt{\mu}}}\big)\phi'_-(1)$, and let α and β be the $2r$-restricted smallest and largest eigenvalues of $\mathcal{A}^*\mathcal{A}$, respectively. Suppose that $\mathcal{X}^* \cap \Omega_\kappa \ne \emptyset$ and $\frac{\beta}{\alpha} < \frac{128\sigma_1^2(4\sigma_1+\sigma_r)^2 + \sigma_r^4}{128\sigma_1^2(4\sigma_1+\sigma_r)^2 - \sigma_r^4}$ with $\alpha > \frac{4}{\nu\sigma_r^2}$. Fix an arbitrary $(\overline{U},\overline{V}) \in \mathcal{W}^*$. Then, for any (U,V) with $0 < \Theta_\rho(U,V) - \Theta_\rho(\overline{U},\overline{V}) < \frac{1}{2}$ and $(U,V) \in \mathbb{B}((\overline{U},\overline{V}), \min(\frac{\sqrt{\sigma_r}}{4}, \frac{\varpi}{\rho}, \frac{c\rho}{4\sqrt{\nu}\|\mathcal{A}\| + 16\mu\sigma_1}))$, the following inequality*

$$\mathrm{dist}^2(0, \partial\Theta_\rho(U,V)) \ge \gamma'(\Theta_\rho(U,V) - \Theta_\rho(\overline{U},\overline{V}))$$

holds with $\gamma' = \min\big(\frac{\nu}{\beta}\big[\frac{(\beta+\alpha)\sigma_r^4}{128\sqrt{\sigma_1^3}(4\sigma_1+\sigma_r)^2} - \sqrt{\sigma_1}(\beta-\alpha)\big]^2, 2\mu\sigma_r, \frac{c^2\rho^2}{8}\big)$.

Proof. Fix an arbitrary (U,V) with $(U,V)\in\mathbb{B}((\overline{U},\overline{V}),\min(\frac{\sqrt{\sigma_r}}{4},\frac{\varpi}{\rho},\frac{c\rho}{4\sqrt{\nu}\|\mathcal{A}\|+16\mu\sigma_1}))$ and $0<\Theta_\rho(U,V)-\Theta_\rho(\overline{U},\overline{V})<\frac{1}{2}$. Write $J:=\{j\mid \overline{U}_j\neq 0\},\Delta_U=U-\overline{U}$ and $\Delta_V=V-\overline{V}$. By Proposition 3.7, it follows that $\mathcal{W}^*$ is exactly the global minimizer set of Θ_ρ. Using the same arguments as those for Lemma 4.2, we have $\|U_j\|\geq\frac{1}{2}\sqrt{\sigma_r}$ and $\|V_j\|\geq\frac{1}{2}\sqrt{\sigma_r}$ for each $j\in J$, and $J\subseteq J_U$ and $J\subseteq J_V$. Moreover, from $(U,V)\in\mathbb{B}((\overline{U},\overline{V}),\frac{\sqrt{\sigma_r}}{4})$, it follows that $\|U\|\leq\sqrt{2\sigma_1}$ and $\|V\|\leq\sqrt{2\sigma_1}$. In addition, for each $j\in J$ it holds that

$$\theta(\rho\|U_j\|)=\theta(\rho\|V_j\|)=1\ \text{ and }\ \theta'(\rho\|U_j\|)=\theta'(\rho\|V_j\|)=0. \tag{4.13}$$

Indeed, by the definition of $\overline{\rho}$, it is immediate to have

$$\rho>\overline{\rho}\geq\frac{4\|\mathcal{A}\|\sqrt{\kappa}}{\sqrt{\alpha}\sigma_r-\sqrt{2\nu^{-1}}}\phi'_-(1)\geq\frac{4\|\mathcal{A}\|\sqrt{\kappa}\phi'_-(1)}{\sqrt{\alpha}\sigma_r-\sqrt{2\nu^{-1}}}\cdot\frac{\phi'_-(1)}{\rho}$$

where the last inequality is by $\rho>\phi'_-(1)$. So, $\rho^2\sigma_r>\rho^2\big(\sigma_r-\sqrt{\frac{2}{\nu\alpha}}\big)\geq\frac{4\|\mathcal{A}\|\sqrt{\kappa}[\phi'_-(1)]^2}{\sqrt{\alpha}}$. Since $\|A\|\geq\sqrt{\beta}\geq\sqrt{\alpha}$, we have $\frac{1}{2}\rho\sqrt{\sigma_r}>\phi'_-(1)$. Together with $\|U_j\|\geq\frac{1}{2}\sqrt{\sigma_r}$ for each $j\in J$, we have $\rho\|U_j\|>\phi'_-(1)$ for each $j\in J$. Similarly, $\rho\|V_j\|>\phi'_-(1)$ for each $j\in J$. By the definition of θ, we calculate that $\theta(\rho\|U_j\|)=\theta(\rho\|V_j\|)=\psi(1)=1$ for each $j\in J$. By noting that $\theta(t)\equiv 1$ over the set $\{t:\ t>\phi'_-(1)\}$, we have $\theta'(\rho\|U_j\|)=0$ for each $j\in J$. Similarly, $\theta'(\rho\|V_j\|)=0$ for each $j\in J$. Thus, the equalities in (4.13) hold. Note that $0<\Theta_\rho(U,V)-\Theta_\rho(\overline{U},\overline{V})<\frac{1}{2}$ and $\Theta_\rho(\overline{U},\overline{V})=r$. By the expression of $\Theta_\rho(U,V)$ and the first equality in (4.13), we have $\frac{\nu}{2}\|\mathcal{A}(UV^{\mathbb{T}}-M)\|^2\leq\frac{1}{2}$. Now we proceed the arguments by three cases.

Case 1: $J\neq J_U$. Now $U_j=[\Delta_U]_j$ and $V_j=[\Delta_V]_j$ for each $j\in J_U\backslash J$. By using Assumption 4.1, $\|\mathcal{A}(UV^{\mathbb{T}}-M)\|\leq 1/\sqrt{\nu}$ and $\max(\|U\|,\|V\|)\leq\sqrt{2\sigma_1}$, for each $j\in J_U\backslash J$,

$$\begin{aligned}
&\big\|\nu[\mathcal{A}^*\mathcal{A}(UV^{\mathbb{T}}-M)]V_j+\mu U(U^{\mathbb{T}}U_j-V^{\mathbb{T}}V_j)+\frac{\rho U_j}{2\|U_j\|}\theta'(\rho\|U_j\|)\big\|\\
&\geq\frac{1}{2}c\rho-\sqrt{\nu}\|\mathcal{A}^*\|\|V_j\|-2\mu\sigma_1(\|U_j\|+\|V_j\|)\\
&\geq\frac{1}{2}c\rho-(\sqrt{\nu}\|\mathcal{A}\|+4\mu\sigma_1)\max(\|\Delta_U\|_F,\ \|\Delta_V\|_F)\geq\frac{c\rho}{2}-\frac{c\rho}{4}=\frac{c\rho}{4}
\end{aligned}$$

where the last inequality is using $\max(\|\Delta_U\|_F,\ \|\Delta_V\|_F)\leq\frac{c\rho}{4\sqrt{\nu}\|\mathcal{A}\|+16\mu\sigma_1}$. Along with Lemma 4.6,

$$\mathrm{dist}^2(0,\partial_U\Theta_\rho(U,V))\geq\sum_{j\in J_U\backslash J}\big[\frac{c\rho}{2}-\sqrt{\nu}\|\mathcal{A}\|\|V_j\|-2\mu\sigma_1(\|U_j\|+\|V_j\|)\big]^2\geq\frac{c^2\rho^2}{16},$$

which together with $\Theta_\rho(U,V)-\Theta_\rho(\overline{U},\overline{V})<1/2$ implies that

$$\mathrm{dist}^2(0,\partial\Theta_\rho(U,V))\geq\mathrm{dist}^2(0,\partial_U\Theta_\rho(U,V))\geq\frac{c^2\rho^2}{8}\big[\Theta_\rho(U,V)-\Theta_\rho(\overline{U},\overline{V})\big]. \tag{4.14}$$

Case 2: $J\neq J_V$. Using the same arguments as those for Case 1 yields (4.14).

Case 3: $J=J_U=J_V$. In this case, by invoking (4.13) and Lemma 4.6, we have

$$\begin{aligned}
\mathrm{dist}^2(0,\partial\Psi(U,V))&=\mathrm{dist}^2(0,\partial_U\Psi(U,V))+\mathrm{dist}^2(0,\partial_V\Psi(U,V))\\
&=\nu^2\big[\|[\mathcal{A}^*\mathcal{A}(UV^{\mathbb{T}}-M)]V\|_F^2+\|U^{\mathbb{T}}[\mathcal{A}^*\mathcal{A}(UV^{\mathbb{T}}-M)]\|_F^2\big]\\
&\quad+\mu^2\big[\|U(U^{\mathbb{T}}U-V^{\mathbb{T}}V)\|_F^2+\|(V^{\mathbb{T}}V-U^{\mathbb{T}}U)V^{\mathbb{T}}\|_F^2\big].
\end{aligned}$$

In addition, $\Theta_\rho(U,V)-\Theta_\rho(\overline{U},\overline{V})=\frac{\nu}{2}\|\mathcal{A}(UV^{\mathbb{T}}-M)\|^2+\frac{\mu}{4}\|U^{\mathbb{T}}U-V^{\mathbb{T}}V\|_F^2$. Hence, using the same arguments as those for Theorem 4.3, we can obtain $\mathrm{dist}^2(0,\partial\Theta_\rho(U,V))\ge\gamma(\Theta_\rho(U,V)-\Theta_\rho(\overline{U},\overline{V}))$. Together with Case 1 and Case 2, the desired result follows. □

Finally, it is worthwhile to point out that when the linear operator $\mathcal{A}$ is obtained by full sampling, i.e., $\mathcal{A}^*\mathcal{A}(X)=X$ for all $X\in\mathbb{R}^{m\times n}$, the functions Ψ and Θ_ρ associated to $\nu>\frac{4}{\sigma_r^2}$ have the KL property of exponent 1/2 over the set of their global minimizers.

5 Numerical Experiments

In this section, we illustrate the KL property of exponent 1/2 of Ψ and Θ_ρ by applying the proximal linearized alternating minimization (PLAM) method to the problems (1.6) and (3.9), respectively, where f is specified as in (1.3) with $b=\mathcal{A}(M)$ for a low-rank M. For the function Θ_ρ, we choose $\phi(t):=\frac{a-1}{a+1}t^2+\frac{2}{a+1}t$ $(a>1)$ for $t\in\mathbb{R}$. An elementary calculation yields that the conjugate function of ψ takes the following form

$$\psi^*(s)=\begin{cases}0 & \text{if } s\le\frac{2}{a+1};\\ \frac{((a+1)s-2)^2}{4(a^2-1)} & \text{if } \frac{2}{a+1}<s\le\frac{2a}{a+1};\\ s-1 & \text{if } s>\frac{2a}{a+1}.\end{cases} \tag{5.1}$$

So, Assumption 4.1 holds with $\varpi=\frac{2}{a+1}$ and $c=1$. Set $\lambda=1/\nu$ and $\widetilde{\mu}=\mu/\nu$. Write $g_{\lambda,\rho}(t)\equiv\lambda\theta(\rho t)+\frac{\tau}{2}t^2$ with $\tau=\frac{\lambda(a+1)\rho^2}{2(a-1)}$. Along with $\theta(t)=t-\psi^*(t)$, it is easy to check that $g_{\lambda,\rho}$ is convex. For the problem (1.6), let $\Phi(U,V)\equiv f(U,V)+\frac{\widetilde{\mu}}{4}\|U^{\mathbb{T}}U-V^{\mathbb{T}}V\|_F^2$ and $h(t)\equiv\lambda\mathrm{sign}(|t|)$; while for (3.9), let $\Phi(U,V)\equiv f(U,V)+\frac{\widetilde{\mu}}{4}\|U^{\mathbb{T}}U-V^{\mathbb{T}}V\|_F^2-\frac{\tau}{4}\big[\|U\|_F^2+\|V\|_F^2\big]$ and $h(t)\equiv g_{\lambda,\rho}(t)$. Then, the problems (1.6) and (3.9) can be written as

$$\min_{U\in\mathbb{R}^{m\times\kappa},V\in\mathbb{R}^{n\times\kappa}}\Big\{\Phi(U,V)+\frac{1}{2}\sum_{j=1}^{\kappa}\big[h(\|U_j\|)+h(\|V_j\|)\big]\Big\}. \tag{5.2}$$

We apply the PLAM method [6, 47] to solving (5.2) and its iterate steps are as follows.

Remark 5.1. **(i)** The constants L_U and L_V in Algorithm 1 are an upper estimation for the Lipschitz constant of $\nabla_U\Phi(\cdot,V)$ and $\nabla_V\Phi(U,\cdot)$, respectively, over a certain compact set of (U,V) which includes the iterate sequence $\{(\widetilde{U}^k,\widetilde{V}^k)\}$.

(ii) Whether $h(t)\equiv\mathrm{sign}(|t|)$ or $h(t)\equiv g_{\lambda,\rho}(t)$, one may easily achieve a global optimal solution of the subproblems since their proximal operators have a closed form. Notice that Algorithm 1 is an accelerated type of the PLAM proposed in [6], and for its global convergence and linear rate of convergence analysis, the reader may refer to [47].

(iii) By comparing the optimal conditions of the two subproblems with that of (5.2), when

$$\begin{cases}\dfrac{\|\nabla_U\Phi(\widetilde{U}^k,V^k)-\nabla_U\Phi(U^{k+1},V^{k+1})+L_U(U^{k+1}-\widetilde{U}^k)\|}{1+\|b\|}\le\epsilon, & \text{(5.4a)}\\[2ex] \dfrac{\|\nabla_V\Phi(U^{k+1},\widetilde{V}^k)-\nabla_V\Phi(U^{k+1},V^{k+1})+L_V(V^{k+1}-\widetilde{V}^k)\|}{1+\|b\|}\le\epsilon & \text{(5.4b)}\end{cases}$$

holds for a pre-given tolerance $\epsilon>0$, we terminate Algorithm 1 at the iterate (U^{k+1},V^{k+1}).

Algorithm 1 (**Accelerated PLAM method for solving** (5.2))

Initialization: Select an integer $\kappa \geq 1$, an appropriate $\lambda > 0$, and constants $L_U, L_V > 0$. Choose $(U^{-1}, V^{-1}) = (U^0, V^0) \in \mathbb{R}^{m\times\kappa} \times \mathbb{R}^{n\times\kappa}$ and $t_0 = t_{-1} = 1$. Set $k = 0$.

while the stopping conditions are not satisfied **do**

Set $\widetilde{U}^k = U^k + \frac{t_{k-1}-1}{t_k}(U^k - U^{k-1})$ and $\widetilde{V}^k = V^k + \frac{t_{k-1}-1}{t_k}(V^k - V^{k-1})$;

Solve the following two minimization problems

$$U^{k+1} \in \mathop{\arg\min}_{U\in\mathbb{R}^{m\times\kappa}} \Big\{ \langle \nabla_U \Phi(\widetilde{U}^k, V^k), U - \widetilde{U}^k\rangle + \frac{L_U}{2}\|U - \widetilde{U}^k\|_F^2 + \frac{\lambda}{2}\sum_{j=1}^{\kappa} h(\|U_j\|) \Big\},$$

$$V^{k+1} \in \mathop{\arg\min}_{V\in\mathbb{R}^{n\times\kappa}} \Big\{ \langle \nabla_V \Phi(U^{k+1}, \widetilde{V}^k), V - \widetilde{V}^k\rangle + \frac{L_V}{2}\|V - \widetilde{V}^k\|_F^2 + \frac{\lambda}{2}\sum_{j=1}^{\kappa} h(\|V_j\|) \Big\}.$$

Set $t_{k+1} := \frac{1+\sqrt{1+4t_k^2}}{2}$ and $k \leftarrow k+1$.

end while

For the subsequent testing, the starting point (U^0, V^0) of Algorithm 1 is always chosen as

$$(P\mathrm{Diag}(\sqrt{\sigma^\kappa(X^0)}), Q\mathrm{Diag}(\sqrt{\sigma^\kappa(X^0)}))$$

where $(P, Q) \in \mathbb{O}^{m,n}(X^0)$ for $X^0 = \mathcal{A}^*(\mathcal{A}(M))$, and $\sigma^\kappa(X^0) \in \mathbb{R}^\kappa$ is the vector consisting of the first $\kappa \geq r$ components of $\sigma(X^0)$. It should be emphasized that such a starting point is not close to the bi-factors of M unless $\kappa = r$. Unless otherwise stated, the tolerance ϵ in (5.4a)-(5.4b) is chosen as 10^{-10}. All numerical results are computed by a laptop computer running on 64-bit Windows Operating System with an Intel(R) Core(TM) i7-7700 CPU 2.8GHz and 16 GB RAM.

5.1 Illustration of the linear convergence

We take $M \in \mathbb{R}^{m\times n}$ with $m = n = 4000$ and $r = 10$ for example to illustrate the linear convergence of the iterate sequence, where M is generated as follows: to generate the matrices $X_L^* \in \mathbb{R}^{m\times r}$, $X_R^* \in \mathbb{R}^{n\times r}$ with independently and identically distributed (i.i.d.) standard Gaussian entries and then set $M = X_L^*(X_R^*)^{\mathbb{T}}$. For this purpose, we apply Algorithm 1 to the problem (5.2) associated to (1.6) with $\kappa = 30r$, where the linear operator $\mathcal{A}$ is obtained by the uniform sampling with the sample ratio 8.99%, and the parameters $\widetilde{\mu}$ and λ are set as 10^{-3} and $150\|\mathcal{A}^*(\mathcal{A}(M))\|$, respectively. Figure 1 plots the iteration error curves and the time curve, where (U^f, V^f) is the final output of Algorithm 1. Since the relative error $\|U^f(V^f)^{\mathbb{T}} - M\|_F/\|M\|_F \leq 2.04 \times 10^{-12}$, we conclude that (U^f, V^f) is a global optimal solution of (1.6). The subfigure on the left hand shows that the sequence $\{(U^k, V^k)\}$ generated by Algorithm 1 indeed converges linearly to (U^f, V^f). We also use Algorithm 1 to solve the problem (5.2) associated to (3.9) with $\kappa = 30r$, where $\mathcal{A}$ is obtained by the uniform sampling with the sample ratio 7.49%, and the parameters $\widetilde{\mu}, \lambda$ and ρ are respectively set as 10^{-3}, $\frac{a+1}{2}(0.05\|\mathcal{A}^*(\mathcal{A}(M))\|)^2$ and $\frac{40}{(a+1)\|\mathcal{A}^*(\mathcal{A}(M))\|}$ for $a = 3.7$. Figure 2 plots the iteration error curves and the time curve, and the iteration error curves show that the sequence $\{(U^k, V^k)\}$ is linearly convergent. Since the relative error

$\|U^f(V^f)^{\mathbb{T}} - M\|_F/\|M\|_F \leq 2.69 \times 10^{-12}$, we conclude that the sequence $\{(U^k, V^k)\}$ also converges linearly to a global optimal solution (U^f, V^f).

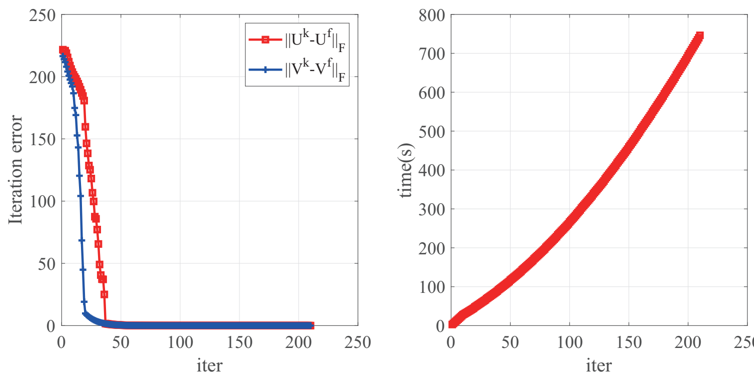

Figure 1: The iterate errors and computing time of Algorithm 1 for minimizing Ψ

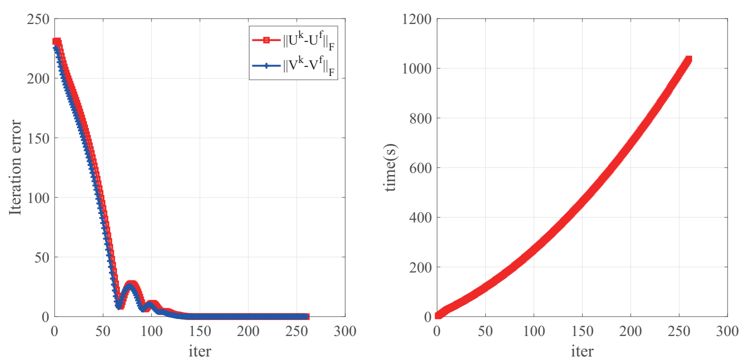

Figure 2: The iterate errors and computing time of Algorithm 1 for minimizing Θ_ρ

5.2 Influence of λ on the linear convergence

We take $M \in \mathbb{R}^{m\times n}$ with $m = n = 2000$ and $r = 10$ for example to illustrate the influence of λ on the linear convergence of the iterates, where M is generated as before. To that end, we use Algorithm 1 to solve the problem (5.2) with $\kappa = 20r$ and $\widetilde{\mu} = 10^{-3}$, where the linear operator $\mathcal{A}$ is obtained by the uniform sampling with the sample ratio 13.3%. For the problem (5.2) associated to (1.6), we set $\lambda_i = c_i\|\mathcal{A}^*(\mathcal{A}(M))\|$, while for the problem (5.2) associated to (3.9) set $\lambda_i = \frac{a+1}{2}(c_i\|\mathcal{A}^*(\mathcal{A}(M))\|)^2$ with $a = 3.7$ and $\rho_i = \frac{2}{(a+1)c_i\|\mathcal{A}^*(\mathcal{A}(M))\|}$. Figure 3 plots the iteration error curve $\|U^k - U^f\|_F$ corresponding to four different λ_i ($i = 0, 1, 2, 3$).

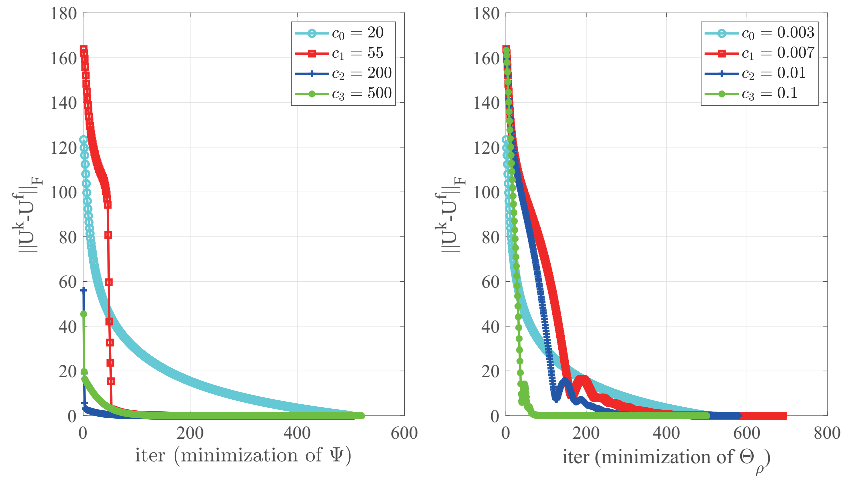

Figure 3: The iterate errors of Algorithm 1 for minimizing Ψ and Θ_ρ with different λ

We see that, when minimizing Ψ and Θ_ρ with $\lambda \geq \lambda_1$, the iterate sequence displays a linear convergence, but it does not have the linear convergence when minimizing Ψ and Θ_ρ with $\lambda = \lambda_0$. We check that for $\lambda = \lambda_0$ the final output (U^f, V^f) still has a full rank 200 since $\lambda = \lambda_0$ is too small. This does not contradicts the result of Theorem 4.3 since our starting point is not close to the bi-factors of M. In addition, the iterate sequence also displays a linear convergence when minimizing Ψ and Θ_ρ with $\lambda = \lambda_3$, though the final output does not recover M since the relative error $\|U^f(V^f)^{\mathbb{T}} - M\|_F/\|M\|_F \geq 0.13$. Since the nonzero-columns of U^f and V^f are equal to r, this performance of this example does not contradicts the result of Example 4.5. In other words, the functions Ψ and Θ_ρ may have the KL property of exponent $1/2$ over the set of stationary points for which the number of nonzero columns equals the rank of the true matrix M.

6 Conclusions

For the rank regularized loss minimization model, we have proposed an $\ell_{2,0}$-norm regularized factored formulation and derived some equivalent DC regularized surrogates from the global exact penalty of its MPEC reformulation. For these nonsmooth factored models, we take a local view to study the KL property of exponent $1/2$ of their objective functions over the set of their global minimizers, which means that gradient descent or alternating minimization methods with a well-chosen starting point will converge linearly to an optimal solution. Numerical testing in Section 5 indicates that the proposed regularized factored models are successful even if the chosen starting point is far from the set of global optima. This implies that a good geometric landscape may exist for these nonsmooth factored functions. We will leave this topic for our future study. In addition, we will focus on the KL property of exponent $1/2$ for Ψ and Θ_ρ under the noisy setting.

References

[1] H. Attouch, J. Bolte, P. Redont and A. Soubeyran, Proximal alternating minimization and projection methods for nonconvex problems: an approach based on the Kerdyka-

Łojasiewicz inequality, *Math. Oper. Res.* 35 (2010) 438–457.

[2] H. Attouch, J. Bolte and B.F. Svaiter, Convergence of descent methods for semi-algebraic and tame problems: proximal algorithms, forward-backward splitting, and reguarlized Gauss-Seidel methods, *Math. Program.* 137 (2013) 91–129.

[3] F.R. Bach, Consistency of trace norm minimization, *J. Mach. Learn. Res.* 8 (2008) 1019–1048.

[4] S.J. Bi and S.H. Pan, Multistage convex relaxation approach to rank regularized minimization problems based on equivalent mathematical program with a generalized complementarity constraint, *SIAM J. Control Optim.* 55 (2017) 2493–2518.

[5] S.J. Bi and S.H. Pan, GEP-MSCRA for computing the group zero-norm regularized least squares estimator, [Online], (2018), available at https://arxiv.org/pdf/1804.09887.

[6] J. Bolte, S. Sabach and M. Teboulle, Proximal alternating linearized minimization for nonconvex and nonsmooth problems, *Math. Program.* 146 (2014) 459–494.

[7] S. Burer and R.D. Monteiro, A nonlinear programming algorithm for solving semidefinite programs with low-rank factorization, *Math. Program.* 95 (2003) 329–357.

[8] R. Cabral, F. De la Torre, J. Costeira and A. Bernardino, Unifying nuclear norm and bilinear factorization approaches for low-rank matrix decomposition, in: *Proc. IEEE Int. Conf. Comput. Vis.*, 2013, pp. 2488–2495.

[9] J.F. Cai, E.J. Candès and Z. Shen, A singular value thresholding algorithm for matrix completion, *SIAM J. Optim.* 20 (2010) 1956–1982.

[10] E.J. Candès and B. Recht, Exact matrix completion via convex optimization, *Found. Comput. Math.* 9 (2009) 717–772.

[11] E.J. Candès and Y. Plain, Tight oracle inequalities for low-rank matrix recovery from a minimal number of noisy random measurements, *IEEE Trans. Inform. Theory* 57 (2011) 2342–2359.

[12] Y.X. Chen and Y.J. Chi, Robust spectral compressed sensing via structured matrix completion, *IEEE Trans. Inform. Theory* 60 (2014) 6576–6601.

[13] M.A. Davenport and J. Romberg, An overview of low-rank matrix recovery from incomplete observations, *IEEE J. Sel. Topics Signal Process.* 10 (2016) 608–622.

[14] M. Fazel, Matrix rank minimization with applications, PhD thesis, Stanford University, 2002.

[15] M. Fazel, H. Hindi and S. Boyd, Log-det heuirstic formatrix rank minimization with applications to Hankel and Euclidean distance matrices, in: *Proc. American Control Conf.*, vol. 3, 2003, pp. 2156–2162.

[16] M. Fazel, T.K. Pong, D.F. Sun and P. Tseng, Hankel matrix rank minimization with applications to system identification and realization, *SIAM J. Matrix Anal. Appl.* 34 (2013) 946–977.

[17] R. Ge, J.D. Lee and T. Ma, Matrix completion has no spurious local minimum, in: *Adv. Neural Inf. Process. Syst.*, 2016, pp. 2973–2981.

[18] R. Ge, C. Jin and Y. Zheng, No spurious local minima in nonconvex low rank problems: A unified geometric analysis, in: *Proc. 34th Int. Conf. Mach. Learn.*, 2017, pp. 1233–1242.

[19] D. Gross, Recovering low-rank matrices from few coefficients in any basis, *IEEE Trans. Inform. Theory* 57 (2011) 1548–1566.

[20] B.D. Haeffele, E. Yang and R. Vidal, Structured low-rank matrix factorization: optimality, algorithm and applications to image processing, in: *Proc. 31th Int. Conf. Mach. Learn.*, 2014, pp. 2007–2015.

[21] C.J. Hsieh and P. Olsen, Nuclear norm minimization via active subspace selection, in: *Proc. 31th Int. Conf. Mach. Learn.*, 2014, pp. 575–583.

[22] P. Jain, P. Netrapalli and S. Sanghavi, Low-rank matrix completion using alternating minimization, in: *Proc. of the 45th annual ACM symposium on Theory of Computing*, 2013, 665–674.

[23] C. Jin, R. Ge, P. Netrapalli, S.M. Kakade and M.I. Jordan, How to escape saddle points efficiently, in: *Proc. 34th Int. Conf. Mach. Learn.*, 2017, pp. 1724–1732.

[24] V. Koltchinskii, K. Lounici and A.B. Tsybakov, Nuclear-norm penalization and optimal rates for noisy low-rank matrix completion, *Ann. Statist.* 39 (2011) 2302–2329.

[25] Y. Koren, The bellkor solution to the netflix grand prize, Netflix Prize Documentation, 2009.

[26] Y. Koren, R.M. Bell and C. Volinsky, Matrix factorization techniques for recommender systems, *Computer* 42 (2009) 30–37.

[27] M.J. Lai, Y.Y. Xu and W.T. Yin, Improved iteratively reweighted least squares for unconstrained smoothed l_q minimization, *SIAM J. Numer. Anal.* 51 (2013) 927–957.

[28] J.D. Lee, M. Simchowitz, M.I. Jordan and B. Recht, Gradient descent only converges to minimizers, in: *Conference on Learning Theory*, 2016, pp. 1246–1257.

[29] X.G. Li, J.W. Lu, R. Arora, J. Haupt, H. Liu, Z.R. Wang and T. Zhao, Symmetry, saddle Points, and global geometry of nonconvex matrix factorization, in: *Information Theory and Applications Workshop (ITA)* (2018), doi: 10.1109/ITA.2018.8503215.

[30] Q.W. Li, Z.H. Zhu and G.G. Tang, The non-convex geometry of low-rank matrix optimization, *Information and Inference: A Journal of the IMA* 8 (2018) 51–96.

[31] G.Y. Li and T.K. Pong, Calculus of the exponent of Kurdyka-Łojasiewicz inequality and its applications to linear convergence of first-order methods, *Found. Comput. Math.* 18 (2018) 1199–1232.

[32] Y.L. Liu, S.J. Bi and S.H. Pan, Equivalent lipschitz surrogates for zero-norm and rank optimization problems, *J. Global Optim.* 72 (2018) 679–704.

[33] W.M. Miao, S.H. Pan and D.F. Sun, *A rank-corrected procedure for matrix completion with fixed basis coefficients*, *Math. Program.* 159 (2016), 289–338.

[34] K. Mohan and M. Fazel, Iterative reweighted algorithm for matrix rank minimization, *J. Mach. Learn. Res.* 13 (2012) 3441–3473.

[35] S. Negahban and M.J. Wainwright, Estimation of (near) low-rank matrices with noise and high-dimensional scaling, *Ann. Statist.* 39 (2011) 1069–1097.

[36] D. Park, A. Kyrillidis, C. Caramanis and S. Sanghavi, Finding low-rank solution via non-convex matrix factorization efficiently and provably, *SIAM J. Imaging Sci.* 11 (2018), 2165–2204.

[37] D. Park, A. Kyrillidis, C. Caramanis and S. Sanghavi, Non-square matrix sensing without spurious local minima via the Burer-Monteiro approach, in: *Proc. 20th Int. Conf. Artif. Intell. Statist.*, 2017, pp. 65–74.

[38] R. Pietersz and P. Groenen, Rank reduction of correlation matrices by majorization, *Quant. Finance* 4 (2004) 649–662.

[39] B. Recht, M. Fazel and P.A. Parrilo, Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization, *SIAM Rev.* 52 (2010) 471–501.

[40] J. Rennie and N. Srebro, in: Fast maximum margin matrix factorization for collaborative prediction, *Proc. 22nd Int. Conf. Mach. Learn.*, 2005, pp. 713–719.

[41] R.T. Rockafellar and R. J-B. Wets, *Variational Analysis*, Springer, New York, 1998.

[42] S. Ruslan and N. Srebro, Collaborative filtering in a non-uniform world: learning with the weighted trace norm, in: *Adv. Neural Inf. Process. Syst.*, 2010, pp. 2056–2064.

[43] R.Y. Sun and Z.Q. Luo, Guaranteed matrix completion via non-convex factorization, *IEEE Trans. Inform. Theory* 62 (2016) 6535–6579.

[44] T. Tao, S.H. Pan and S.J. Bi, Error bound of critical points and KL property of exponent 1/2 for squared F-norm regularized factorization, *J. Global Optim.* (2021), available: https://doi.org/10.1007/s10898-021-01077-0.

[45] K.C. Toh and S.W. Yun, An accelerated proximal gradient algorithm for nuclear norm regularized least squares problems, *Pacific J. Optim.* 6 (2010) 615–640.

[46] S. Tu, R. Boczar, M. Simchowitz, M. Soltanolkotabi and B. Recht, Low-rank solution of linear matrix equations via procrustes flow, in: *Proc. 33rd Int. Conf. Mach. Learn.*, 2016, pp. 424–446.

[47] Y.Y. Xu and W.T. Yin, A block coordinate descent method for regularized multiconvex optimization with applications to nonnegative tensor factorization and completion, *SIAM J. Imaging Sci.* 6 (2013) 1758–1789.

[48] Q. Zhang, C.H. Chen, H.K. Liu and A.M.C. So, On the linear convergence of the ADMM for regularized non-convex low-rank matrix recovery, [Online], (2018), available at https://www1.se.cuhk.edu.hk/ manchoso/admm_MF.pdf.

[49] X. Zhang, L.X. Wang, Y.D. Yu and Q.Q. Gu, A primal-dual analysis of global optimality in nonconvex low-rank matrix recovery, in: *Proc. 35th Int. Conf. Mach. Learn.*, 2018, pp. 5862–5871.

[50] T. Zhao, Z. Wang and H. Liu, in: A nonconvex optimization framework for low rank matrix estimation, *Adv. Neural Inf. Process. Syst.*, 2015, pp. 559–567.

[51] Q. Zheng and J. Lafferty, A convergent gradient descent algorithm for rank minimization and semidefinite programming from random linear measurements, in: *Adv. Neural Inf. Process. Syst.*, 2015, pp. 109–117.

[52] Z.H. Zhu, Q.W. Li, G.G. Tang and M.B. Wakin, Global optimization in low-rank matrix optimization, *IEEE Trans. Signal Process.* 66 (2018) 3628–3641.

Manuscript received 15 April 2021
revised 20 June 2021
accepted for publication 11 October 2021

Shujun Bi
Shujun Bi
School of Mathematics, South China University of Technology
Guangzhou, 510640, China
E-mail address: bishj@scut.edu.cn

Ting Tao
School of Mathematics, South China University of Technology
Guangzhou, 510640, China
E-mail address: mattao@mail.scut.edu.cn

Shaohua Pan
School of Mathematics, South China University of Technology
Guangzhou, 510640, China
E-mail address: shhpan@scut.edu.cn