A RELAXATION APPROACH TO UBPP BASED ON EQUIVALENT DC PENALIZED FACTORIZED MATRIX PROGRAMS∗

YITIAN QIAN† AND SHAOHUA PAN‡

Abstract. This paper is concerned with the unconstrained binary polynomial program (UBPP), which has a host of applications in many science and engineering fields. By leveraging the global exact penalty for its DC constrained SDP reformulation, we achieve an equivalent DC penalized SDP, and propose a continuous relaxation approach by seeking the critical point of the Burer-Monteiro factorization for a finite number of DC penalized SDPs with increasing penalty factors. A globally convergent majorization-minimization (MM) method with extrapolation is also developed to capture such critical points. Under a mild condition, we show that the rank-one projection of the output for the relaxation approach is an approximate feasible solution of the UBPP and quantify the upper bound of its objective value from the optimal value. Numerical comparisons with the SDP relaxation method armed with a special random rounding technique and the DC relaxation approach based on the solution of linear SDPs confirm the efficiency of the proposed relaxation approach, which can solve the instance of 20000 variables in 15 minutes and yield an upper bound to the optimal value and the known best value with a relative error at most 1.824% and 2.870%, respectively.

Key words. UBPP; DC global exact penalty; Burer-Monteiro factorization; relaxation approach

AMS subject classifications. 90C27, 90C22, 90C26

1. Introduction. We focus on the unconstrained binary polynomial program:

\[
\min_{x \in \{-1,1\}^N} \vartheta(x)
\]

where \( \vartheta : \mathbb{R}^N \to \mathbb{R} \) is a polynomial function. Without of loss generality, we assume that the degree of \( \vartheta \) is 2\( d \). Such a problem has a large spectrum of applications; for example, the classical unconstrained binary quadratic program (UBQP), i.e., problem (1.1) with \( \vartheta(x) = \langle x, Qx \rangle + c^T x \) for an \( N \times N \) real symmetric matrix \( Q \) and a vector \( c \in \mathbb{R}^N \), is frequently used to formulate the optimization problems on graphs, facility locations problems, resources allocation problems, clustering problems, set partitioning problems, and various forms of assignment problems (see, e.g., [5, 16, 27, 31, 35]).

A more general example is problem (1.1) with \( \vartheta \) taking the following form

\[
\vartheta(x) := \prod_{i=1}^{q} \left( x_i^T Q_i x_i + c_i^T x_i + a_i \right),
\]

where \( x = (x_1; x_2; \ldots; x_q) \in \mathbb{R}^{nq} \) and every \( Q_i \) is an \( n \times n \) real symmetric matrix. For the application of problem (1.1) with such \( \vartheta \), please refer to He et al. [10].

Over the past several decades, many solution methods have been developed for this class of NP-hard problems (see [16] for the survey), which can be roughly classified into the exact method (see, e.g., [17, 18, 21, 33]), the metaheuristic method (see, e.g., [44, 8]), and the continuous relaxation method (see, e.g., [1, 6, 5, 9]). Among the existing continuous relaxation methods, the semidefinite program (SDP) relaxation is the most popular one owing to the significant work [9], which states that for the max-cut and max 2-satisfiability problem, a special random rounding for the solution of the

∗ July 28, 2021

Funding: This work is funded by the National Natural Science Foundation of China under project No.11571120.

† School of Mathematics, South China University of Technology (mayttqian@mail.scut.edu.cn).

‡ School of Mathematics, South China University of Technology (shhpan@scut.edu.cn).
associated linear SDP problem can yield a feasible solution whose expected objective value is at least 0.87856 times the optimal value. The later research works in this line mainly focus on the relaxation improvement by adding valid inequalities [11, 37, 17] or using the double nonnegative cone relaxation (see [15, 7]). This work is also concerned with the SDP relaxation, but its aim is to design a relaxation approach that can yield a desirable approximate feasible solution without any rounding technique.

Notice that $z \in \{-1, 1\}^N$ iff $Z = zz^T$ is a rank-one PSD matrix with one diagonals. Then, it is not difficult to reformulate the UBPP (1.1) as the following rank-one SDP:

$$(1.3) \quad \min_{X \in S_p^p} \left\{ f(X) \mid \text{s.t. } \text{rank}(X) \leq 1, \text{diag}(X) = e, X \in S^p \right\},$$

where $p$ is a positive integer related to $N$ and $d$, $S_p^p$ denotes the set of all PSD matrices in $S^p$, the set of all $p \times p$ real symmetric matrices, and $f: S^p \to \mathbb{R}$ is a smooth function (determined by $\vartheta$) with gradient Lipschitz relative to a compact set $\mathbb{B}_\Omega$, a compact set containing $(1 + \tau)\Omega - \tau \Omega$ with $\Omega := \{ X \in S_p^p \mid \text{diag}(X) = e \}$ for any $\tau \in [0, 1]$. For example, by taking $v_{d}(x) := (1, x_1, x_2, \ldots, x_N, x_1^2, x_1x_2, \ldots, x_N^2, x_1^2, \ldots, x_N^2)^T$, it is immediate to reformulate (1.1) as (1.3) with $f(X) = \langle C_\vartheta, X \rangle$ and $p = (N + d)$ for a $p \times p$ real symmetric matrix $C_\vartheta$. Of course, one can reformulate (1.1) as (1.3) with a nonlinear $f$ but a small $p$ by leveraging the structure of $\vartheta$; see section 5.5.

Our relaxation approach is based on a global exact penalty for the difference of convex (DC) constrained SDP reformulation of the UBPP. Since $\text{rank}(X) \leq 1$ if and only if $\|X\|_* - \|X\| = 0$, where $\|X\|_*$ and $\|X\|$ denote the nuclear norm and spectral norm of $X$, problem (1.3) can be rewritten as the DC constrained SDP problem:

$$(1.4) \quad \min_{X \in S_p^p} \left\{ f(X) \mid \text{s.t. } (I, X) - \|X\| = 0, \text{diag}(X) = e, X \in S^p_+ \right\}.$$  

The feasible set of (1.4), denoted by $\mathcal{F}$, is still combinatorial, which is the set of all $p \times p$ rank-one PSD binary matrices. Since numerically it is more difficult to handle DC constraints than to handle DC cost functions, we focus on its penalty problem

$$(1.5) \quad \min_{X \in \Omega} \left\{ f(X) + \rho(\langle I, X \rangle - \|X\|) \right\}$$

where $\rho > 0$ is the penalty parameter. By [3, Proposition 2.3&Theorem 3.1], problem (1.5) associated to every $\rho \geq \rho^* := (1 + 2p)|\alpha_f|$ has the same global optimal solution set as the original (1.4) does, where $\alpha_f > 0$ is the Lipschitz constant of $f$ on $\Omega$.

In practice, the exact penalty methods based on model (1.5) still need to solve many DC penalty problems even if the threshold $\rho^*$ is known, because it is almost impossible to achieve a global optimal solution of (1.5) associated to a fixed $\rho \geq \rho^*$. Furthermore, the convex relaxation methods for the penalty problem (1.5) all require an eigenvalue decomposition in each iteration, which forms the major computational bottleneck and restricts their scalability to large-scale problems. Inspired by the recent renewed interest in the Burer-Monteiro factorization method [5, 4] for low-rank matrix recovery (see, e.g., [41, 22]), we consider the factorized form of problem (1.5):

$$(1.6) \quad \min_{V \in \mathbb{R}^{m \times p}} \left\{ f(V^TV) + \rho(\|V\|_F^2 - \|V\|_2^2) \right\}$$

s.t. $V \in \mathcal{S} := \{ V \in \mathbb{R}^{m \times p} \mid \|V_j\| = 1, j = 1, \ldots, p \}$

where $1 < m < p$ is an appropriate integer and $V_j$ means the $j$th column of $V$. It is easy to verify that if $X^*$ is a global optimizer of rank $r$ for problem (1.5) associated to
with those of \( P \times \). We notice that Pham Dinh and Le Thi \cite{pham1994dc} for which the simple MATLAB common “round” yields a feasible solution with the DC relaxation approach based on \( \text{Biq Mac Library} \) instances with 100 to 250 variables. Among others, \( \text{dcSDPT3} \) is the limit lies in a critical point set smaller than that of common DC programs for short), we compare its performance with that of \( \text{SDPRR} \) and \( \text{dcSDPT3} \) for rank-one and cannot provide a feasible solution to the UBQPs without rounding. A continuous approach by solving the factorization form of a linear SDP. Since their miss those outputs of high quality. For the UBQPs, Wen and Yin \cite{wen2015exact} of solution will become worse. This implies that solving a single penalty problem will only solves a single penalty problem associated to a well-chosen penalty parameter. There indeed exists the best penalty parameter but to capture it is almost impossible, and if the chosen penalty parameter is greater than the unknown best one, the quality of solution will become worse. This implies that solving a single penalty problem will miss those outputs of high quality. For the UBQPs, Wen and Yin \cite{wen2018dc} ever provided a continuous approach by solving the factorization form of a linear SDP. Since their factorized form neglects the rank-one constraint, the obtained critical point is far from rank-one and cannot provide a feasible solution to the UBQPs without rounding.

To confirm the efficiency of our relaxation approach based on model \( (1.6) \) \( \text{dcFAC} \) for short), we compare its performance with that of \( \text{SDPWR} \) and \( \text{dcSDPT3} \) for 119 Biq Mac Library instances with 100 to 250 variables. Among others, \( \text{dcSDPT3} \) is the DC relaxation approach based on \( (1.5) \) for which the involved linear SDP subproblems
are solved with the software SDPT3 [42], and SDPRR is the SDP relaxation method armed with the random rounding technique in [9] (see section 5 for the details). We also compare the performance of dcFAC with that of dcSDPT3 for 112 UBQPs from the G-set, OR-Library and Palubeckis instances with 800 to 20000 variables, and with that of dcSNCG for 26 UBPP examples constructed with $\vartheta$ from (1.2) for $q = 2$. Here, dcSNCG is the DC relaxation approach based on (1.5) for which the quadratic SDP subproblems are solved with the dual semismooth Newton method [36] (see section 5 for the details). Numerical comparisons show that dcFAC is comparable to dcSDPT3 in terms of the quality of the output if the latter uses the same adjusting rule of $\rho$ (only possible for $n \leq 500$) as dcFAC does, otherwise dcFAC is superior to dcSDPT3.

Moreover, dcFAC is significantly superior to SDPRR and dcSNCG by the quality of the outputs and the CUP time taken. For 119 Biq Mac Library instances, the outputs of dcFAC have the relative error with the optimal values at most 1.824% except the special gka9b and gka10b, and for 112 UBQP instances with $n \geq 800$, the relative gaps of its outputs from the known best values are at most 2.870%.

Notation: Let $\mathbb{R}^{n_1 \times n_2}$ denote the vector space of all $n_1 \times n_2$ real matrices, equipped with the trace inner product $\langle \cdot, \cdot \rangle$ and its induced Frobenius norm $\| \cdot \|_F$, and let $\mathcal{O}^p$ denote the set of all $p \times p$ orthonormal matrices. For $X \in \mathbb{R}^{n_1 \times n_2}$, $X_J$ with some $J \subseteq \{1, \ldots, n_2\}$ means the submatrix of $X$ consisting of those columns $X_j$ with $j \in J$, $\|X\|$ and $\|X\|_*$ denote the spectral norm and nuclear norm of $X$, respectively, and $\mathbb{B}(X, \varepsilon)$ means the closed ball on Frobenius norm centered at $X$ with radius $\varepsilon > 0$. For every $X \in \mathbb{S}^p$, write $\lambda(X) = (\lambda_1(X), \ldots, \lambda_p(X))$ with $\lambda_1(X) \geq \cdots \geq \lambda_p(X)$ and $\mathcal{O}(X) := \{ P \in \mathcal{O}^p \mid X = P \text{Diag}(\lambda(X)) P^T \}$, and for every $P \in \mathcal{O}(X)$, let $P_I \in \mathbb{R}^{p \times m}$ denote the submatrix consisting of the first $m$ columns of $P$. Let $I$ and $e$ denote an identity matrix and a vector of all ones, whose dimensions are known from the context. For a closed set $\Delta \subseteq \mathbb{R}^{n_1 \times n_2}$, $\delta_\Delta$ denotes the indicator of $\Delta$, i.e., $\delta_\Delta(z) = 0$ if $z \in \Delta$, otherwise $\delta_\Delta(z) = +\infty$. Write $\mathcal{R} := \{ Z \in \mathbb{S}^p \mid \text{rank}(Z) \leq 1 \}$.

2. Preliminaries. We first recall from [38] the notions of subdifferentials.

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

$$\partial h(x) := \left\{ v \in \mathbb{R}^n \mid \liminf_{x' \neq x \to x} \frac{h(x') - h(x) - \langle v, x' - x \rangle}{\|x' - x\|} \geq 0 \right\};$$

and the (basic) subdifferential (also known as the limiting subdifferential) of $h$ at $x$ is

$$\partial h(x) := \left\{ v \in \mathbb{R}^n \mid \exists x^k \to x \text{ with } h(x^k) \to h(x), v^k \in \partial h(x^k) \text{ s.t. } v^k \to v \text{ as } k \to \infty \right\}.$$ 

**Remark 2.2.** (a) The sets $\widehat{\partial} h(x)$ and $\partial h(x)$ are closed with $\widehat{\partial} h(x) \subseteq \partial h(x)$, and the former is also convex. When $h$ is convex, they reduce to the subdifferential of $h$ at $x$ in the convex analysis context. The vector $\overline{x}$ at which $0 \in \partial h(\overline{x})$ is called a critical point of $h$, and we denote by crit $h$ the critical point set of $h$.

(b) When $h$ is an indicator function of a closed set $\Delta$ in $\mathbb{R}^n$, $\widehat{\partial} h(x)$ and $\partial h(x)$ respectively reduce to the regular normal cone $\widehat{N}_\Delta(x)$ and the normal cone $N_\Delta(x)$.

The following lemma provides the subdifferential of the concave function $\psi$, whose proof is included in Supplementary Materials.
Lemma 2.3. Fix any $X \in \mathcal{S}^p$ with $J_X := \{ j \in \{1, \ldots, p\} \mid |\lambda_j(X)| = \|X\| \}$. Then the subdifferential of $\psi$ defined in (1.7) at $X$ has the following expression:
\[
\partial \psi(X) = \begin{cases} 
- \text{sign}(\lambda_j(X))P_jP_j^T & j \in J_X, P \in \mathcal{O}(X) \} \quad \text{if } X \neq 0; \\
\{ -e_je_j^T & j \in J_X, P \in \mathcal{O}(X) \} \quad \text{if } X = 0,
\end{cases}
\]
where $e_j \in \mathbb{R}^p$ is the vector with the $j$th entry being 1 and others being 0; and when $\text{rank}(X) = 1$, $\psi$ is regular at $X$ with $\partial \psi(X) = \partial \psi(X) = -\partial(\psi)(X)$.

3. Critical points of equivalent models. As well known, when nonconvex models are equivalent in a global sense, they generally have different critical point sets even local optimizer sets. Then, it is necessary to discuss the relations between the local optimizers of (1.5) and those of (1.6), and so as their critical points.

**Proposition 3.1.** The following statements hold for problems (1.5) and (1.6).

(i) Every feasible point of (1.4) is a local optimal solution, which is a strictly local optimizer of (1.5) with $\rho > \rho^*$, the threshold for the exactness of (1.5).

(ii) If $X^*$ is a local optimizer of rank $r$ for problem (1.5) associated to $\rho$, then $\sqrt[n]{V^*(P^*)^T}$ with $P^* \in \mathcal{O}(X^*)$ and $\Lambda^* = \text{Diag}(\lambda_1(X^*), \ldots, \lambda_m(X^*))$ is a local optimizer of (1.6) associated to this $\rho$ and $m \geq r$. Conversely, if $V^*$ is a local optimizer of (1.6) with $\text{rank}(V^*) = 1$ and $\rho > 0$, then $(V^*)^TV^* \in \mathcal{F}$ and is a local optimal solution to (1.5) with $\rho \geq \rho^*$.

**Proof.** (i) By the discreteness of $\mathcal{F}$, it is easy to verify that $\mathcal{F}$ coincides with the local optimizer set of (1.4). Pick any $X \in \mathcal{F}$. By the proof of [3, Theorem 3.1(b)], $X$ is a local optimizer of (1.5) associated to $\rho \geq \rho^*$. So, there exists $\varepsilon \in (0, 1)$ such that
\[
\| f(X^*) - f(X) \| \geq f(X^*) - f(X) \quad \text{for all } X \in \mathcal{F}(X, \varepsilon).
\]

Since $\mathcal{F}$ is the set of all $p \times p$ rank-one PSD binary matrices, by reducing $\varepsilon$ if necessary, we have $\| (I, Z) - \| \| \| > 0$ for all $Z \in \mathcal{B}(X, \varepsilon) \cap \Omega$. Together with the last inequality, for all $Z \in \mathcal{B}(X, \varepsilon) \cap \Omega$ and $\rho > \rho^*$, it holds that
\[
f(X) + \rho\langle I, X \rangle - \| X \| \geq f(X) + \rho\langle I, X \rangle - \| X \| > f(Z) + \rho\langle I, Z \rangle - \| Z \|.\]

(ii) Let $X^*$ be a local optimizer of rank $r$ for (1.5). Then there exists $\varepsilon > 0$ such that
\[
\| f(X) - \rho\langle I, X \rangle - \| X \| \| \geq f(X^*) - \rho\langle I, X^* \rangle - \| X^* \| \| \| \quad \text{for all } X \in \mathcal{B}(X^*, \varepsilon) \cap \Omega.
\]

From $\text{Diag}(X^*) = \varepsilon$, each column of $V^* = \sqrt[n]{V^*(P^*)^T}$ has a unit length, which implies that $V^*$ is feasible to (1.6). Take $\varepsilon' = \min(1, \varepsilon (\|V^*\| + 1))$. For any $V \in \mathcal{B}(V^*, \varepsilon') \cap \mathcal{S}$,
\[
\| V^TV - X^* \| \leq \| V^TV - (V^*)^TV^* \| \leq \| V^TV - (V^*)^TW^* \| \leq \varepsilon' \| V^TV - (V^*)^TV^* \| \leq \varepsilon,
\]
which along with $V^TV \in \Omega$ means that $V^TV \in \mathcal{B}(X^*, \varepsilon) \cap \Omega$. Thus, from (3.1) we get
\[
f(V^TV) + \rho(\| V^TV \| - \| V \|) \geq f(V^TV) + \rho(\| V^* \| - \| V \|)
\]
for all $V \in \mathcal{B}(V^*, \varepsilon') \cap \mathcal{S}$. So, $V^*$ is a local optimizer of (1.6). The converse of part (ii) is easy to obtain by using part (i) and the feasibility of $(V^*)^TV^*$ to (1.4).

**Definition 3.2.** We call $X \in \mathcal{S}^p$ a critical point of (1.4) if $0 \in \nabla f(X) + \mathcal{N}_\mathcal{F}(X)$, and a critical point of (1.5) with $\rho > 0$ if $0 \in \nabla f(X) + \rho[I + \partial \psi(X)] + \mathcal{N}_\mathcal{S}(X)$; and call $V \in \mathbb{R}^{m \times p}$ a critical point of (1.6) with $\rho > 0$ if $0 \in \nabla_f(V) + \rho(2V + \partial \psi(V)) + \mathcal{N}_\mathcal{S}(V)$.
Remark 3.3. The critical point defined in Definition 3.2 for the DC problems (1.5) and (1.6) are stronger than the common one in the reference (see [32, 33, 24]), where $\partial \psi(X)$ (or $\partial \tilde{\psi}(V)$) is replaced with its upper inclusion $-\partial(-\psi)(X)$ (or $-\partial(-\tilde{\psi})(V)$).

Proposition 3.4. Let $\tilde{F}$ denote the critical point set of (1.4), and let $\tilde{\Omega}_\rho$ and $\tilde{S}_\rho$ denote the critical point set of (1.5) and (1.6) associated to $\rho > 0$. Then,

(i) $F = \tilde{F} = \{X \in \mathbb{R}^p | 0 \in \nabla f(X) + N_{\Omega}(X) + N_{\mathbb{R}}(X)\}$.

(ii) For any $\rho > 0$, crit$\Phi_\rho \subseteq \tilde{\Omega}_\rho$ where $\Phi_\rho(Z) := f(Z) + \rho(\langle I, Z \rangle - \|Z\|) + \delta_{\Omega}(Z)$, and every rank-one critical point of (1.5) associated to any $\rho > 0$ is a rank-one strictly local optimizer of (1.5) associated to those $\rho > \rho^*$.

(iii) For each $X \in \mathcal{F}$, there is a neighborhood in which the critical points of (1.5) with $\rho > \rho^*$ are all rank-one if their objective values are not more than $\Phi_\rho(X)$.

(iv) If $X \in \tilde{\Omega}_\rho$ is rank $r \leq m$, then $\nabla P_f^T \in \tilde{S}_\rho$ where $\Lambda = \text{Diag}(\lambda_1(X), \ldots, \lambda_m(X))$ and $P \in \Omega(X)$. If $V \in \tilde{S}_\rho$ and there exists $(W, y) \in \partial \psi(V^TV) \times \mathbb{R}^p$ such that $\nabla f(V^TV) + \rho(I + W) + \text{Diag}(y) \in \mathbb{S}_+^p$, then $V^TV \in \tilde{\Omega}_\rho$.

Proof. (i) Pick any $X \in \mathcal{F}$. Since $\mathcal{F}$ coincides with the local optimizer set of (1.4), we have $0 \in \nabla f(X) + N_{\mathbb{R}}(X)$, which means that $X \in \tilde{F}$. Since $F = \Omega \cap \mathbb{R}$, from [3, Proposition 2.3] and [13, Section 3.1] we have $N_f(X) \subseteq N_{\Omega}(X) + N_{\mathbb{R}}(X)$. Since rank$(X) = 1$, from [26, Proposition 3.6], $N_{\mathbb{R}}(X) = N_{\mathbb{R}}(X)$, which by [38, Corollary 10.9] and the convexity of $\Omega$ implies that $N_{\Omega}(X) + N_{\mathbb{R}}(X) \subseteq N_f(X) \subseteq N_{\mathbb{R}}(X)$. The two sides show that $N_f(X) = N_{\Omega}(X) + N_{\mathbb{R}}(X)$. So, $0 \in \nabla f(X) + N_{\Omega}(X) + N_{\mathbb{R}}(X)$.

Thus, $F = \mathcal{F} \subseteq \{X \in \mathbb{R}^p | 0 \in \nabla f(X) + N_{\Omega}(X) + N_{\mathbb{R}}(X)\}$. Notice that if $X \in \mathbb{S}^p$ satisfies $0 \in \nabla f(X) + N_{\Omega}(X) + N_{\mathbb{R}}(X)$, $X \in \Omega \cap \mathbb{R} = \mathcal{F}$. The second equality holds.

(ii) Pick any $X \in \text{crit} \Phi_\rho$. By [38, Exercise 10.10] and the Lipschitz continuity of $\| \cdot \|$, $\partial \Phi_\rho(X) \subseteq \nabla f(X) + \rho[I + \partial \psi(X)] + N_{\Omega}(X)$, which implies that crit$\Phi_\rho \subseteq \tilde{\Omega}_\rho$. Since every rank-one critical point of (1.5) lies in $\mathcal{F}$, by Proposition 3.1 (i), it is a rank-one strictly local optimizer of (1.5) with $\rho > \rho^*$.

(iii) By the proof of [3, Theorem 3.1(b)], every $X \in \mathcal{F}$ is a local optimizer of (1.5) with $\rho > \rho^*$. Then, there exists $\epsilon > 0$ such that $\Phi_\rho(Z) \geq \Phi_\rho(X)$ for all $Z \in \text{B}(X, \epsilon) \cap \Omega$. Fix any $\rho > \rho^*$. Pick any $X_\rho \in \Omega_\rho \cap \text{B}(X, \epsilon)$. Then,

$$f(X_\rho) + \rho(\langle I, X_\rho \rangle - \|X_\rho\|) \leq \Phi_\rho(X) = f(X) = \Phi_\rho(X) \leq \Phi_\rho^{\rho^*}(X_\rho),$$

which by $\rho > \rho^*$ implies that $\langle I, X_\rho \rangle - \|X_\rho\| = 0$. Together with $X_\rho \in \mathbb{S}_+^p$, it follows that $\lambda_2(X_\rho) = \cdots = \lambda_p(X_\rho) = 0$. Hence, the matrix $X_\rho$ is rank-one.

(iv) Fix any $V \in \mathcal{S}$. By [38, Theorem 10.6], we have $\partial \tilde{\psi}(V) = 2V \partial \psi(V^TV)$. Since $N_{\mathbb{S}}(V) = \{V \text{Diag}(w) | w \in \mathbb{R}^p\} = N_{\mathbb{S}}(V)$, the set $\mathcal{S}$ is Clarke regular. Thus, $V$ is a critical point of (1.6) if and only if there exist $W \in \partial \psi(V^TV) \times \mathbb{R}^p$ such that

$$V[\nabla f(V^TV) + \rho(I + W) + \text{Diag}(y)] = 0. \quad (3.2)$$

Now let $X \in \tilde{\Omega}_\rho$ with rank$(X) \leq m$. Clearly, $X \in \Omega$. By [38, Theorem 6.14], we have $N_{\Omega}(X) = \{\text{Diag}(z) | z \in \mathbb{R}^p\} + N_{\mathbb{S}}(X)$. From Definition 3.2, there exist $W \in \partial \psi(X), y \in \mathbb{R}^p$ and $S \in N_{\mathbb{S}}(X)$ such that $0 = \nabla f(X) + \rho(I + W) + \text{Diag}(y) + S$. Let $V = \sqrt{\Lambda P_f^T}$ with $P = \text{O}(X)$ and $\Lambda = \text{Diag}(\lambda_1(X), \ldots, \lambda_m(X))$. Notice that $VS = 0$. Then $V[\nabla f(X) + \rho(I + W) + \text{Diag}(y)] = 0$. Since $X = V^TV$, from (3.2) we have $V \in \tilde{S}_\rho$. 

For the second part, by taking $X = V^T V$ and $S = \nabla f(V^T V) + \rho(I+W) + \text{Diag}(y) \in \mathcal{S}_+^r$, from (3.2) we have $XS = 0$. Hence, $-S \in \mathcal{N}_{\mathcal{S}_+^r}(X)$, and $X \in \hat{\Omega}_\rho$ by Definition 3.2. □

Remark 3.5. By Proposition 3.4 (i), for every $X \in \mathcal{F}$, there exists $H \in \mathcal{N}_{\mathcal{S}}(X)$ such that $-\rho^{-1}(\nabla f(X) + H) \in \mathcal{N}_{\mathcal{S}}(X)$, but $-\rho^{-1}(\nabla f(X) + H)$ may not belong to $I + \partial \psi(X)$ which is the singleton $\{I - X/\|X\|\}$ by Lemma 2.3. This means that the rank-one critical point set of (1.5) associated to any $\rho > 0$ is far smaller than $\mathcal{F}$, and from the last part of Proposition 3.4 (ii), it is also a local optimizer set when $\rho > \rho'$.

4. Relaxation approach based on model (1.6). Inspired by the relationship between (1.5) and (1.6), we propose the following continuous relaxation approach by seeking a finite number of critical points of (1.6) associated to increasing $\rho$.

Algorithm 4.1 (DC relaxation approach based on (1.6))

Select an integer $m > 1$, a small $\epsilon \in (0,1)$, and appropriately large $l_{\max} \in \mathbb{N}$ and $\rho_{\max} > 0$. Choose $\rho_0 > 0$, $\sigma > 1$ and a starting point $V^0 \in \mathcal{S}$.

for $l = 0, 1, 2, \ldots, l_{\max}$ do

Starting from $V^l \in \mathcal{S}$, seek a critical point $V^{l+1}$ of the nonconvex problem

\[
\min_{V \in \mathcal{S}} \{ \tilde{f}(V) + \rho_l(\|V\|_F^2 + \tilde{\psi}(V)) \}.
\]

If $\|V^{l+1}\|_F^2 - \|V^{l+1}\|_F^2 \leq \epsilon$, then stop. Otherwise, $\rho_{l+1} \leftarrow \min\{\sigma\rho_l, \rho_{\text{max}}\}$.

end for

The core of Algorithm 4.1 is to achieve a critical point of (4.1) efficiently. Notice that the function $\tilde{f}$ is smooth with gradient Lipschitz relative to $\mathbb{B}_\mathcal{S}$, a compact set containing $(1 + \tau)\mathcal{S} - \tau \mathcal{S}$ for any $\tau \in [0,1]$. We denote by $L_{\tilde{f}}$ the Lipschitz constant of $\nabla \tilde{f}$ relative to $\mathbb{B}_\mathcal{S}$. Fix any $Z \in \mathbb{B}_\mathcal{S}$. From the descent lemma, for any $V \in \mathbb{B}_\mathcal{S}$,

\[
(4.2a) \quad \tilde{f}(V) \leq \tilde{f}(Z) + \langle \nabla \tilde{f}(Z), V - Z \rangle + (L_{\tilde{f}}/2)\|V - Z\|_F^2,
\]

\[
(4.2b) \quad -\tilde{f}(V) \leq -\tilde{f}(Z) - \langle \nabla \tilde{f}(Z), V - Z \rangle + (L_{\tilde{f}}/2)\|V - Z\|_F^2.
\]

Notice that $\tilde{\psi}$ is concave since $\tilde{\psi}(V) = -\|V^T V\| = -\|V\|^2$ for any $V \in \mathbb{R}^{m \times p}$. Hence, $\tilde{\psi}(V) \leq \tilde{\psi}(Z) + \langle \Gamma, V - Z \rangle$ for any $\Gamma \in \partial \tilde{\psi}(Z)$. Together with (4.2a), we have

\[
\tilde{f}(V) + \rho_l(\|V\|_F^2 + \tilde{\psi}(V)) \leq \tilde{F}(V, Z) := \tilde{f}(Z) + \rho_l \tilde{\psi}(Z) - \langle \nabla \tilde{f}(Z), Z \rangle - \rho_l \langle \Gamma, Z \rangle
+
\langle \nabla \tilde{f}(Z) + \rho_l \Gamma, V \rangle + \rho_l \|V\|_F^2 + \frac{L_{\tilde{f}}}{2} \|V - Z\|_F^2.
\]

Along with $\tilde{F}(Z, Z) = \tilde{f}(Z) + \rho_l(\|Z\|_F^2 + \tilde{\psi}(Z))$, $\tilde{F}(\cdot, Z)$ is a majorization of the cost function of (4.1) at $Z$. By this, we propose an MM method with extrapolation, which is not affiliated to the DCA [32] due to $\Gamma_k \in \partial \tilde{\psi}(V^k)$ and the constraint $V \in \mathcal{S}$.

Remark 4.1. (a) Since $L_{\tilde{f}}$ may be unknown in practice, one can search a suitable $L_k$ by the descent lemma. When $L_k$ is known, it suffices to choose $L_k \equiv (1 + \tau)\tilde{L}_{\tilde{f}}$ for a tiny $\tau > 0$. As will be shown below, the restriction $L_0 > L_{\tilde{f}}$ is necessary for the global convergence of Algorithm 1 due to the nonconvexity of (4.3).

(b) By the proof of Proposition 3.4 (iv), $\partial \tilde{\psi}(V^k) = 2V^k \partial \psi(X^k)$ with $X^k = (V^k)^T V^k$. Thus, by Lemma 2.3, one can choose $\Gamma^k = -2V^k P^k (P^k)^T$ with $P^k \in \mathbb{O}(X^k)$. Clearly,
Algorithm 1 (MM method with extrapolation for (4.1))

Fix an integer \( l \geq 0 \). Choose \( \rho \in (0, 1) \), \( 0 \leq \beta_0 \leq \beta < 1 \) and \( L_0 \geq L > L_f \). Set \( \rho = \rho_l \) and \( V^{-1} = V^0 = V^l \).

for \( k = 0, 1, 2, \ldots \) do

Choose an element \( \Gamma^k \in \partial \varphi(V^k) \). Let \( U^k = V^k + \beta_k(V^k - V^{k-1}) \) and compute

\[
V^{k+1} = \arg \min_{V \in S} \left\{ \langle \nabla \tilde{f}(U^k) + \rho \Gamma^k, V \rangle + \rho \|V\|_F^2 + \frac{L_k}{2} \|V - U^k\|_F^2 \right\}.
\]

Update \( \beta_k \) by \( \beta_{k+1} \in [0, \beta] \) and \( L_k \) by \( L_{k+1} \in [L, L_0] \).

end for

---

\( P_k \) can be achieved by the SVD of \( V^k \), whose computation cost is cheaper since \( V^k \) has less rows. We stipulate that \( \Gamma_k \) in Algorithm 1 is always chosen in this way.

(c) Let \( G^k := \frac{1}{L_k + 2\rho} (L_k U^k + \rho \Gamma^k - \nabla \tilde{f}(U^k)) \). Write \( J_k := \{ j \mid \|G^k\| \neq 0 \} \). Then \( V^{k+1} \) with \( V^{k+1}_j = \frac{G^k}{\|G^k\|} \) for \( j \in J_k \) and \( V^{k+1}_j = (1, 0, \ldots, 0)^T \in \mathbb{R}^m \) for \( j \notin J_k \) is an optimal solution of (4.3). So, the computation cost in each step of Algorithm 1 is very cheap.

**Lemma 4.2.** Fix an integer \( l \geq 0 \). Suppose that there exists \( P \in \mathbb{C}((V^1)^T V^1) \) such that \( P_l \) has no zero entries. Then, \( \|V^1\|_B^2 - \|V^0\|_B^2 \leq c_0 \) for some \( c_0 \in (0, 1) \) when \( \rho_l > \overline{\rho} := \frac{L_0^p + \overline{\nu}}{L_1 + 2\rho} \), where \( |P_{1\kappa}| := \min_{1 \leq j \leq p} |P_{1j}| \) and \( \overline{\nu} := (L_0 + 2L_f) \overline{\nu} + \|\nabla \tilde{f}(\frac{1}{\sqrt{m}} E)\|_F \).

**Proof.** Recall that \( \Gamma^0 = -2V^0 P^T \) by Remark 4.1 (b). After a simple calculation, for each \( j \in \{1, 2, \ldots, p\} \), we have \( \|\Gamma^0_j\| = 2\|V^0\| |P_{1j}| \geq 2|P_{1j}| \geq 2|P_{1\kappa}| > 0 \), where the first inequality is due to \( \|V^0\| \geq 1 \) implied by \( \|V^0\|_F = \sqrt{\overline{\rho}} \). Notice that \( V^0 \in S \) and \( \|\nabla \tilde{f}(U^0)\|_F \leq L_f \|V^0 - \frac{1}{\sqrt{m}} E\|_F + \|\nabla \tilde{f}(\frac{1}{\sqrt{m}} E)\|_F \). It is not hard to verify that \( \|L_0 U^0 - \nabla \tilde{f}(U^0)\|_F \leq L_0 \|V^0\|_F + \|\nabla \tilde{f}(U^0)\|_F \leq \overline{\nu} \). Fix any \( \rho > \overline{\rho} \). By Remark 4.1 (c), \( G^0 = \frac{\rho}{L_0 + 2\rho} \Gamma^0 + \frac{L_0^p \overline{\nu}}{L_0 + 2\rho} \). Then, for every \( j \in \{1, \ldots, p\} \), it holds that

\[
\|G^0_j\| \geq \frac{\rho}{L_0 + 2\rho} \|\Gamma^0_j\| - \frac{1}{L_0 + 2\rho} \|L_0 U^0 - \nabla \tilde{f}(U^0)\|_j \geq \frac{\rho \|\Gamma^0\| - \overline{\nu}}{L_0 + 2\rho} \geq \frac{\rho |P_{1\kappa}|}{L_0 + 2\rho},
\]

where the third inequality is using \( \|\Gamma^0\| \geq 2|P_{1\kappa}| \) and \( \rho > \overline{\rho} \). This means that \( G^0 \) has no zero columns. Define \( \overline{G}^0 := \frac{1}{2} \Gamma^0 \) with \( D = \text{Diag}(\frac{1}{\|G_1\|}, \ldots, \frac{1}{\|G_p\|}) \). Clearly, \( V^{l+1} = \overline{G}^0 \Gamma^0 D + \frac{1}{L_0 + 2\rho} \{L_0 U^0 - \nabla \tilde{f}(U^0)\} D \). Define \( \overline{G}^0 := \frac{1}{2} \Gamma^0 D \). Clearly, rank(\( \overline{G}^0 \)) = rank(\( \Gamma^0 \)) = 1. Then, it follows that

\[
\text{dist}(\overline{G}^0, \mathcal{K}) \leq \left\| \overline{G}^0 \right\|_F \leq \frac{2L_0 \|V^0\|}{L_0 + 2\rho} \|D\|_F + \frac{L_0^p \overline{\nu}}{L_0 + 2\rho} \|L_0 U^0 - \nabla \tilde{f}(U^0)\|_F \leq \frac{L_0 \|V^0\|}{L_0 + 2\rho} \|D\|_F + \frac{\overline{\nu}}{L_0 + 2\rho} \|D\| \leq \frac{L_0 \rho}{\rho |P_{1\kappa}|} + \frac{\overline{\nu}}{\rho |P_{1\kappa}|} \leq c_0,
\]

where the second inequality is using \( \|\Gamma^0\| \leq 2\|V^0\| \), the third and the fourth are using (4.4), and the last one is due to \( \rho > \overline{\rho} \). The proof is completed.

To establish the convergence of Algorithm 1, we define the potential function

\[
\Theta_\rho(V, \Gamma, U) := \tilde{f}(V) + \rho \|V\|_F^2 + \rho \langle \Gamma, V \rangle + \rho (-\tilde{\varphi})^* (-\Gamma) + \delta_S(V) + \frac{\gamma L}{2} \|V - U\|_F^2.
\]
with a certain $\gamma \in (0, \frac{L_1 - L_2}{L_3})$ for $(V, \Gamma, U) \in \mathbb{R}^{m \times p} \times \mathbb{R}^{m \times p} \times \mathbb{R}^{m \times p}$. Then, the following conclusion holds for $\Theta_{\rho}$, whose proof is included in the supplementary materials.

**Proposition 4.3.** Let $\{(V^k, \Gamma^k)\}$ be the sequence given by Algorithm 1. Then,

(i) for each $k \in \mathbb{N} \cup \{0\}$, with $\nu_k := \frac{(\gamma k - 2L_2\beta L_3)(L_1 - L_2 - \gamma L_3)(L_1 - L_2)^2 \beta^2}{L_1 - L_2}$,

$$\Theta_{\rho}(V^{k+1}, \Gamma^k, V^k) \leq \Theta_{\rho}(V^k, \Gamma^k, V^{k-1}) - \frac{\nu_k}{2} \|V^k - V^{k-1}\|_F^2;$$

(ii) the sequence $\{(V^k, \Gamma^k)\}$ is bounded, and hence the accumulation point set of the sequence $\{(V^k, \Gamma^k, V^{k-1})\}$, denoted by $\Delta_\rho$, is nonempty and compact;

(iii) when $\beta < \sqrt{\frac{L_1 - L_2 - \gamma L_3}{L_1 - L_2 - \gamma L_3}}$, the limit $\lim_{k \to \infty} \Theta_{\rho}(V^k, \Gamma^k, V^{k-1})$ exists, and moreover, the function $\Theta_{\rho}$ keeps unchanged on the set $\Delta_\rho$;

(iv) for all $k \in \mathbb{N}$, with $\alpha = \sqrt{2(L_f + L_0 + \gamma L_1)^2 + \rho^2 + \gamma^2 L_2^2}$ it holds that

$$\text{dist}(0, \partial \Theta_{\rho}(V^k, \Gamma^k, V^{k-1})) \leq \alpha \|V^k - V^{k-1}\|_F + \|V^k - V^{k-2}\|_F.$$ 

**Remark 4.4.** (a) Let $L_0 = \kappa L_f$ for $\kappa > 1$ and $L_f = c L_f$ for $1 < c \leq \kappa$. If $\gamma > 0$ is such that $2\gamma c \leq \kappa - 1$, then

$$\sqrt{\frac{\gamma k(L_0 - L_f - \gamma L_0)}{L_0 - L_0 - L_f}} = \sqrt{\frac{2\gamma c(\kappa - 1 - \kappa)}{\kappa^2 - 2\kappa c - 1}}.$$ 

When $f$ is convex, the restriction on $\beta$ in part (iii) can be updated to $\beta < \sqrt{\frac{\gamma k(L_0 - L_f - \gamma L_0)}{L_0 - L_0 - L_f}} = \sqrt{\frac{2\gamma c(\kappa - 1 - \kappa)}{\kappa^2 - 2\kappa c - 1}}$, because the coefficient $\frac{\gamma}{\kappa}$ appearing in the first term of (SM2.2) can be removed.

(b) By Remark 2.2 and the proof of part (iv), we have $\Delta_\rho \subseteq \text{crit} \Theta_{\rho}$. While from (SM2.3) and Definition 3.2, one can check that $\Pi_1(\text{crit} \Theta_{\rho}) \subseteq \hat{S}_\rho$, where $\Pi_1(V, \Gamma, U) = V$ for $(V, \Gamma, U) \in \mathbb{R}^{m \times p} \times \mathbb{R}^{m \times p} \times \mathbb{R}^{m \times p}$. So, $\Pi_1(\Delta_\rho) \subseteq \Pi_1(\text{crit} \Theta_{\rho}) \subseteq \hat{S}_\rho$.

By [38, Proposition 11.21], $(-\psi)^*(U) = \frac{1}{\kappa} \|U\|_2^2$ for $U \in \mathbb{R}^{m \times p}$. Clearly, $(-\psi)^*$ is semialgebraic. By Remark 4.4 (b), we can check that $\Pi_1(\text{crit} \Theta_{\rho}) \subseteq \hat{S}_\rho$, where $\Pi_1(V, \Gamma, U) = V$ for $(V, \Gamma, U) \in \mathbb{R}^{m \times p} \times \mathbb{R}^{m \times p} \times \mathbb{R}^{m \times p}$. So, $\Pi_1(\Delta_\rho) \subseteq \Pi_1(\text{crit} \Theta_{\rho}) \subseteq \hat{S}_\rho$.

**Theorem 4.5.** Let $\{(V^k, \Gamma^k)\}$ be the sequence generated by Algorithm 1 for solving (1.6) associated to $\rho$ with $\beta$ satisfying the restriction in Proposition 4.3 (iii). Then, $\{V^k\}$ is convergent and its limit $V^*$ is a critical point of (1.6) associated to $\rho$. If the limit $V^*$ is rank-one, then $V^*$ is a local optimal solution of (1.4).

Next we focus on the stopping criterion of Algorithm 4.1. If it occurs at some $l < l_{\text{max}}$, we say that Algorithm 4.1 exits normally. The following result states that Algorithm 4.1 armed with Algorithm 1 can exit normally under a mild condition. In this case, $(V^{l+1})^T V^{l+1}$ provides an approximate rank-one feasible point of (1.4).

**Proposition 4.6.** Fix an integer $l \geq 0$. Let $\{V^k\}$ be the sequence generated by Algorithm 1 from $V^0 = V^l$. Then, for any $\varepsilon \in (0, c_0]$ with $c_0$ from Lemma 4.2, when $\rho_l \geq \bar{\rho} := \max \left\{ \frac{2\gamma L_f}{(1 - \sqrt{1 - 0.5p^{-1}c_0}) \varepsilon} \right\}$ with $\bar{\varepsilon} = 8(L_f + L_0)p + 2\sqrt{\rho_l \|V_f\|_F^2}$, then $\bar{\varepsilon} = 8(L_f + L_0)p + 2\sqrt{\rho_l \|V_f\|_F^2}$, $\bar{\epsilon} = 8(L_f + L_0)p + 2\sqrt{\rho_l \|V_f\|_F^2}$, $\bar{\epsilon} = 8(L_f + L_0)p + 2\sqrt{\rho_l \|V_f\|_F^2}$.

(i) for each integer $k \geq 0$ with $\frac{\bar{\varepsilon}}{p} \leq \rho_l - \|V^k\|_2 \leq c_0$,

$$\|V^{k+1}\|_2^2 \geq \|V^k\|_2^2 + (1 - \sqrt{1 - 0.5p^{-1}c_0}) \varepsilon - c_0;$$

(ii) if there is $P \in \mathbb{R}^{m \times p}$ such that $P_1$ has no zero entries and $\rho_l > \max(\rho, \bar{\rho})$ with $\bar{\rho}$ from Lemma 4.2, there is an integer $1 \leq k \leq \left[ \frac{c_0}{(1 - \sqrt{1 - 0.5p^{-1}c_0}) \varepsilon - c_0} \right] + 1$ such that $\|V^k\|_F^2 - \|V^{k+1}\|_2^2 \leq \epsilon$ for $k \geq \bar{k}$.
Proof. (i) For each integer \( k \geq 1 \), from the definition of \( V^{k+1} \), for any \( V \in S \),

\[
\rho(\Gamma^{k}, V^{k+1} - V) \leq \langle \nabla f(U^{k}), V - V^{k+1} \rangle + \frac{L_{k}}{2}\|V - U^{k}\|^{2}_{F} - \frac{L_{k}}{2}\|V^{k+1} - U^{k}\|^{2}_{F}
\]

\[
\leq \langle \nabla f(U^{k}), V - V^{k+1} \rangle + \frac{L_{k}}{2}[\|V\|^{2}_{F} + 2\|V^{k+1} - V\|_{F}\|U^{k}\|_{F}]
\]

\[
\leq \langle \nabla f(U^{k}) - \nabla f(\sqrt{m}), V - V^{k+1} \rangle + 6.5L_{0}p
\]

\[
\leq 2\sqrt{4L_{f}}\sqrt{p} + \|\nabla f(\sqrt{m})\|_{F} + 6.5L_{0}p \leq \omega,
\]

where the third inequality is by \( L_{k} \leq L_{0} \) for all \( k \) and \( \|V\|^{2}_{F} = p \) for \( V \in S \). Then

\[
(4.6) \quad -\langle T^{k}, V \rangle \leq \rho^{-1}\omega + \|T^{k}\|_{\infty}\|V^{k+1}\| = \rho^{-1}\omega + 2\|V^{k}\|\|V^{k+1}\|
\]

where the equality is by the choice of \( \Gamma^{k} \) in Remark 4.1 (ii). Let \( V^{k} \) have the SVD given by \( Q|\text{Diag}((\sigma_{1}(V^{k}), \ldots, \sigma_{m}(V^{k}))^{T})|U^{k} \) with \( \sigma_{1}(V^{k}) \geq \cdots \geq \sigma_{m}(V^{k}) \). Write \( Q = [q_{1} \cdots q_{m}] \in \Omega^{m} \) and \( U = [u_{1} \cdots u_{p}] \in \Omega^{p} \). Then, for every \( j \in \{1, \ldots, p\} \),

\[
(4.7) \quad \|\sigma_{i}(V^{k})\|^{2}u_{i}^{2}_{j} = 1 - \sum_{i=2}^{m}[\sigma_{i}(V^{k})]^{2}u_{i}^{2}_{j} \geq 1 - \sum_{i=2}^{m}[\sigma_{i}(V^{k})]^{2} \geq 1 - c_{0} > 0
\]

where the next to last inequality is using \( \|V^{k}\|^{2}_{F} - \|V^{k}\|^{2} \leq c_{0} \). Take \( V = q_{i}u_{i}^{T} \) with \( \hat{u}_{1j} = \frac{u_{i j}}{|u_{i j}|} \) for each \( j \). Clearly, \( \hat{V} \in S \). From (4.6) with \( V = \hat{V} \) and \( \Gamma^{k} = -2V^{k}u_{1j}u_{1j}^{T} \),

\[
\rho^{-1}\omega + 2\|V^{k}\|\|V^{k+1}\| \geq -\langle T^{k}, \hat{V} \rangle = 2\|V^{k}\|(u_{1j}^{T}\hat{V})_{1} = 2\|V^{k}\|\sum_{j=1}^{p} \frac{|u_{i j}|}{\|V\|}|u_{i j} - u_{i}^{2}_{j}|
\]

\[
(4.8) \quad = 2\|V^{k}\|^{2} + 2\|V^{k}\|\sum_{j=1}^{p} \left( \frac{u_{i j}}{\|V\|}|u_{i j}| - u_{i}^{2}_{j} \right)
\]

where the third equality is by \( \sum_{j=1}^{p} u_{i}^{2}_{j} = 1 \). Since \( \|u_{1}^{T}\| = 1 \), there is an index \( j \) such that \( u_{i}^{2}_{j} \leq \frac{1}{p} \). Note that \( \|V^{k}\|\|u_{1}^{T}\| \leq 1 \) for each \( j \) by the first equality of (4.7). So,

\[
\rho^{-1}\omega + 2\|V^{k}\|\|V^{k+1}\| \geq 2\|V^{k}\|^{2} + 2\|V^{k}\|^{2}(\frac{1}{\|V^{k}\|}\|u_{1}^{T}\| - 1)u_{i}^{2}_{j}
\]

\[
\geq 2\|V^{k}\|^{2} + 2(1 - \|V^{k}\|/\sqrt{p})\sqrt{1 - c_{0}}
\]

\[
\geq 2\|V^{k}\|^{2} + 2(1 - \sqrt{1 - 0.5p^{-1}\epsilon})\sqrt{1 - c_{0}},
\]

where the second inequality is by (4.7), and the last is since \( p - \|V^{k}\|^{2} \geq \frac{\epsilon}{2} \). Along with \( \rho \geq \frac{\omega}{(1 - \sqrt{1 - 0.5p^{-1}\epsilon})\sqrt{1 - c_{0}}} \), we get \( \|V^{k}\|\|V^{k+1}\| \geq \|V^{k}\|^{2} + \frac{1}{2}(1 - \sqrt{1 - 0.5p^{-1}\epsilon})\sqrt{1 - c_{0}} \).

Together with \( \|V^{k}\|\|V^{k+1}\| \leq \frac{\epsilon}{2}\|V^{k}\|^{2} + \frac{1}{2}\|V^{k+1}\|^{2} \), the desired result follows.

(ii) Let \( \eta := (1 - \sqrt{1 - 0.5p^{-1}\epsilon})\sqrt{1 - c_{0}} \) and \( \hat{\eta} := \frac{\eta}{(1 - \sqrt{1 - 0.5p^{-1}\epsilon})\sqrt{1 - c_{0}}} + 1 \). We first argue that there exists \( 1 \leq \underline{\eta} \leq \hat{\eta} \) such that \( p - \|V^{k}\|^{2} \leq \epsilon \). If not, for all \( 1 \leq k \leq \hat{\eta} \), we have \( p - \|V^{k}\|^{2} > \epsilon \). By Lemma 4.2, \( p - \|V^{1}\|^{2} \leq c_{0} \). Thus, from part (i), it follows that \( p - \|V^{k}\|^{2} \leq c_{0} \) for all \( 1 \leq k \leq \hat{\eta} \). Using part (i) again, it holds that \( \|V^{k+1}\|^{2} \geq \|V^{k}\|^{2} + \eta \) for all \( 1 \leq k \leq \hat{\eta} \). From this, \( \|\hat{V}\|^{2} \geq \|V^{k}\|^{2} + (\hat{\eta} - 1)\eta \geq p - c_{0} + (\hat{\eta} - 1)\eta \), which is impossible since \( \|\hat{V}\|^{2} < p - \epsilon \). So, the stated \( \underline{\eta} \) exists. Next we argue by induction that \( p - \|V^{k}\|^{2} \leq \epsilon \) for all \( k \geq \underline{\eta} \). Suppose that \( p - \|V^{j}\|^{2} \leq \epsilon \) for \( j \geq \underline{\eta} \). We show that \( p - \|V^{j+1}\|^{2} \leq \epsilon \) by two cases. If \( p - \|V^{j}\|^{2} < \epsilon/2 \), by invoking (4.6) with \( V = V^{j} \), we have \( 2\|V^{j}\|\|V^{j+1}\| \geq 2\|V^{j}\|^{2} + \rho^{-1}\omega \), which implies that \( \|V^{j+1}\|^{2} \geq \|V^{j}\|^{2} - \rho^{-1}\omega \). So, \( p - \|V^{j+1}\|^{2} \geq p - \|V^{j}\|^{2} + \rho^{-1}\omega \leq \epsilon/2 + \rho^{-1}\omega \leq \epsilon \). If \( p - \|V^{j}\|^{2} \geq \epsilon/2 \), since \( p - \|V^{j}\|^{2} \leq \epsilon < c_{0} \), from part (i) we have \( p - \|V^{j+1}\|^{2} \leq p - \|V^{j}\|^{2} - \eta \leq \epsilon \). \( \square \)
The following theorem states that the rank-one projection of the normal output of Algorithm 4.1 is also an approximately feasible solution of problem (1.1), and provides a quantitative bound estimation for its objective value to the optimal value of (1.1).

**Theorem 4.7.** Let $v^*$ be the optimal value of (1.1) and $V^{l*}$ be a normal output of Algorithm 4.1. For each $l \geq 0$, let $\{V^{l,k}, G^{l,k}\}$ be generated by Algorithm 1 with $V^{l,0} = V^l$ and $\beta_k \equiv 0$. If there exists $l^* \in \{0, 1, \ldots, l_f\}$ such that $f((V^{l^*})^TV^{l^*}) \leq v^*$, then the following relations hold with $r^* = \text{rank}(V^{l^*})$:

$$f(x^{l^*}(x^{l^*})^T) - v^* \leq \rho_l \|V^{l^*}\|_F^2 - \rho_l p/r^* + \sum_{j=1}^{l^*-1} (\rho_j - \rho_{j+1}) \|V^{j+1}\|_F^2 + \alpha_l \epsilon;$$

$$\|x^{l^*} \circ x^{l^*} - e\| \leq \epsilon \quad \text{with} \quad x^{l^*} = \|V^{l^*}\|P_1 \quad \text{for} \quad P \in \mathcal{O}((V^{l^*})^2V^{l^*}).$$

**Proof.** Fix any $l \in \{0, 1, \ldots, l_f\}$. For each $k \geq 0$, from $\beta_k \equiv 0$ and (SM2.1),

$$\langle \nabla \tilde{f}(V^{l,k}), P_{l,k}^T V^{l,k} - V^{l,k} - V^{l,k+1} \rangle + \rho_l \|V^{l,k+1}\|_F^2 - \|V^{l,k}\|_F^2 + \frac{L_{l,k}}{2} \|V^{l,k+1} - V^{l,k}\|_F^2 \leq 0. \quad (4.2a)$$

Notice that $\tilde{f}(V^{l,k+1}) \leq \tilde{f}(V^{l,k}) + \langle \nabla \tilde{f}(V^{l,k}), V^{l,k+1} - V^{l,k} \rangle + \frac{L_{l,k}}{2} \|V^{l,k+1} - V^{l,k}\|_F^2$ by using (4.2a) with $V = V^{l,k+1}$ and $Z = V^{l,k}$. From the last inequality and $L_{l,k} \geq L_l$, $\tilde{f}(V^{l,k+1}) - \tilde{f}(V^{l,k}) + \rho_l \langle P_{l,k}^T, V^{l,k+1} - V^{l,k} \rangle \leq 0,$

where $\|V^{l,k+1}\|_F^2 = \|V^{l,k}\|_F^2 = p$ is also used. Notice that $\langle P_{l,k}^T, V^{l,k+1} \rangle \geq -2\|P_{l,k}\|_F \|V^{l,k+1}\| \geq -\|V^{l,k}\|^2 - \|V^{l,k+1}\|^2$. Then, it holds that $\tilde{f}(V^{l,k+1}) - \rho_l \|V^{l,k+1}\|^2 \leq \tilde{f}(V^{l,k}) - \rho_l \|V^{l,k}\|^2$. By using this recursion formula,

$$\tilde{f}(V^{l,k+1}) - \rho_l \|V^{l,k+1}\|^2 \leq \cdots \leq \tilde{f}(V^{l,0}) - \rho_l \|V^0\|^2 = \tilde{f}(V^l) - \rho_l \|V^l\|^2. \quad (4.10)$$

By Theorem 4.5, the sequence $\{V^{l,k}\}$ is convergent as $k \to \infty$. Let $V^{l,*}$ denote its limit. Then $V^{l+1} = V^{l,*}$. From the last inequality, for each $l \in \{0, 1, \ldots, l_f\}$,

$$\tilde{f}(V^{l+1}) - \rho_l \|V^{l+1}\|^2 \leq \tilde{f}(V^l) - \rho_l \|V^l\|^2. \quad (4.10)$$

Notice that $X^{l^*} = \sum_{i=1}^p \lambda_i(X^{l^*}) P_i P_i^T \in \Omega$. From the Lipschitz continuity of $f$ relative to $\Omega$ with modulus $\alpha_l$, it follows that

$$f(x^{l^*}) = f\left(\sum_{i=1}^p \lambda_i(X^{l^*}) P_i P_i^T\right) = f\left(\lambda_1(X^{l^*}) P_1 P_1^T + \sum_{i=2}^p \lambda_i(X^{l^*}) P_i P_i^T\right)$$

$$\geq f(x^{l^*}(x^{l^*})^T) - \alpha_l \|\sum_{i=2}^p \lambda_i(X^{l^*}) P_i P_i^T\|_F \geq f(x^{l^*}(x^{l^*})^T) - \alpha_l \epsilon. \quad (4.11)$$

On the other hand, adding $(\rho_l - \rho_{l+1})\|V^{l+1}\|^2$ to the both sides of (4.10) yields that

$$\tilde{f}(V^{l+1}) - \rho_l \|V^{l+1}\|^2 \leq \tilde{f}(V^l) - \rho_l \|V^l\|^2 + (\rho_l - \rho_{l+1})\|V^{l+1}\|^2 \quad \forall l \in \{0, \ldots, l_f\}.$$
5. Numerical experiments. This section tests the performance of Algorithm 4.1 armed with Algorithm 1 (dcFAC for short). To confirm its efficiency, we compare its performance with that of the SDP relaxation method equipped with the random rounding technique in [9] (SDPRR for short); see section 5.2 for its introduction.

When \( f \) is a linear function (say, the instances in section 5.2-5.5), we also compare the performance of dcFAC with that of a DC relaxation approach based on model (1.5), i.e., Algorithm 5.1 below for which every penalty problem is solved by Algorithm 2 with \( L_k \equiv L_f = 0 \) and \( \beta_k \equiv 0 \). It is worthwhile to point out that there is no convergence certificate for such Algorithm 2. During the tests, the linear SDP subproblems in Algorithm 2 are solved with the software SDPT3 [42], and Algorithm 5.1 equipped with such Algorithm 2 and SDPT3 is abbreviated to dcSDPT3. The results of Proposition SM3.4 and Theorem SM3.5 are also applicable to dcSDPT3.

\begin{algorithm}
\caption{(DC relaxation approach based on (1.5))}
Choose \( \epsilon \in (0, 1), l_{\text{max}} \in \mathbb{N}, \rho_{\text{max}} > 0, \sigma > 1, \rho_0 > 0 \) and \( X^0 \in \Omega \).
\begin{algorithmic}
\For {\( l = 0, 1, 2, \ldots, l_{\text{max}} \)}
\State Starting from \( X^l \), seek a critical point \( X^{l+1} \) to the nonconvex problem
\begin{equation}
\min_{X \in \Omega} \left\{ f(X) + \rho_l(I, X) - \|X\| \right\}.
\end{equation}
\If {(\( I, X^{l+1} \) - \|X^{l+1}\| \leq \epsilon \), then stop. Otherwise, let \( \rho_{l+1} \leftarrow \min\{\sigma \rho_l, \rho_{\text{max}}\} \).}
\EndFor
\end{algorithmic}
\end{algorithm}

\begin{algorithm}
\caption{(An MM method with extrapolation for (5.1))}
Fix \( l \geq 0 \). Choose \( 0 \leq \beta_0 \leq \overline{\beta} < 1 \) and \( L_0 \geq L_f \). Set \( \rho = \rho_l \) and \( X^{-1} = X^0 = X^l \).
\begin{algorithmic}
\For {\( k = 0, 1, 2, \ldots \)}
\State Choose an element \( W^k \in \partial \psi(X^k) \).
\State Let \( Y^k = X^k + \beta_k(X^k - X^{k-1}) \). Compute an optimal solution of the convex SDP:
\begin{equation}
X^{k+1} = \arg\min_{X \in \Omega} \left\{ \|\nabla f(Y^k) + \rho(I + W^k), X \| + (L_k/2)\|X - Y^k\|_F^2 \right\}.
\end{equation}
\State Update \( \beta_k \) by \( \beta_{k+1} \in [0, \overline{\beta}] \) and \( L_k \) by \( L_{k+1} \in [L_f, L_0] \).
\EndFor
\end{algorithmic}
\end{algorithm}

When \( f \) is not linear, we employ Algorithm 2, an MM method with extrapolation, to solve subproblem (5.1), where \( L_f \) denotes the Lipschitz constant of \( \nabla f \) in \( B_\Omega \). For the convergence analysis of Algorithm 2, see the supplementary materials. Considering that QSDPNAL [23] is not well adapted to the quadratic SDP subproblems of Algorithm 2, we use the dual semismooth Newton method in [36] to solve them. In the sequel, Algorithm 5.1 armed with Algorithm 2 is abbreviated to dcSNCG. Our code can be downloaded from https://github.com/SCUT-OptGroup/rankone.UPPs.

All tests are performed in MATLAB on a workstation running on 64-bit Windows Operating System with an Intel Xeon(R) W-2245 CPU 3.90GHz and 128 GB RAM. We measure the performance of a solver by the relative gap and infeasibility of its outputs and the CPU time (in seconds) taken. Let \( x^* = \|V^*\|Q^*_V \) or \( \sqrt{\|X^*\|P^*_V} \) with \( Q^* \in \mathbb{D}(V^*)^T V^* \) and \( P^* \in \mathbb{D}(X^*) \), where \( V^* \) is the output of dcFAC and \( X^* \) is
an output for one of other three solvers. The relative gap and infeasibility of \( x^\star \) are defined by \( \text{gap} := \frac{\text{Obj} - \text{Bval}}{\text{Bval}} \) and \( \text{infeas} := \| (x^\star \otimes x^\star)^{1/2} - e \|_\infty \), where \(-\text{Bval}\) means the known best value of (1.1), and Obj denotes the objective value of (1.1) at \( x^\star \). For the subsequent tests, we use the default setting for the softwares SDPT3 and SDPNAL+.

5.1. Implementation of dcFAC and dcSNCG. We first focus on the choice of parameters in Algorithm 4.1 and 5.1. Preliminary tests indicate that smaller \( \rho_0 \) and \( \sigma \) often lead to better relative gaps for Algorithm 4.1 and 5.1. Since it is time consuming to search the best \( \rho \), an appropriately small \( \rho_0 \) becomes a reasonable choice. We choose \( \rho_0 = 0.001, \sigma = 1.005 \) for Algorithm 4.1, but \( \rho_0 = 0.1, \sigma = 1.05 \) for Algorithm 5.1 armed with SNCG since it requires much more time for those examples with \( n \geq 2000 \). For Algorithm 5.1 armed with SDPT3, we use \( \rho_0 = 0.001, \sigma = 1.005 \) for solving the examples with \( n < 500 \), but \( \rho_0 = 0.1, \sigma = 1.05 \) for solving the examples with \( n \geq 500 \). We set \( \epsilon = 10^{-8}, \rho_{\max} = 10^6 \) and \( l_{\max} = 10^4 \) for Algorithm 4.1 and 5.1.

In addition, we take \( m = \max(\min(50, \text{round}(p/2)), 2) \) by considering that a smaller \( m \) makes problem (1.6) vulnerable to much worse critical points, but a larger \( m \) requires more computation cost. The starting point \( V^0 \) of Algorithm 4.1 is chosen to be \( \bar{V} \text{Diag}((\| V_1 \|^{-1}, \ldots, \| V_p \|^{-1}) \) where \( \bar{V} \in \mathbb{R}^{m \times p} \) is generated in MATLAB command \( \text{randn}(m, p) \) with a fixed seed for all test problems; and the starting point \( X^0 \) of Algorithm 5.1 is chosen to be \( (V^0)^T V^0 \).

The parameter \( \beta_k \) in Algorithm 1 and 2 is given by Nesterov’s accelerated strategy [28]. Although their convergence analysis requires a restriction on \( \beta_k \), numerical tests indicate that they still converge without it. So, we do not impose any restriction on such \( \beta_k \) during their implementation, and leave this gap for a future topic. For the parameter \( L_k \) of Algorithm 1, when \( L_f \) is known (say, the instances in section 5.2-5.4), we set it to be a fixed constant, otherwise search a desired \( L_k \) by the descent lemma. Specifically, we set \( L_k \equiv 2.001 \| C \| \) for the instances in section 5.2-5.4 since \( L_f = 2 \| C \| \), and search a desired \( L_k \) with \( L_0 \equiv 6 \rho (\| C_1 \| \| C_2 \| F + \| C_1 \| F \| C_2 \|) \) for the instances in section 5.5. For the parameter \( L_k \) of Algorithm 2, we take \( \| C_1 \| \| C_2 \| F + \| C_1 \| F \| C_2 \| \) for the problems in section 5.5 since it is exactly the Lipschitz constant \( L_f \).

During the implementation of Algorithm 1, we seek an approximate critical point of subproblem (4.1). According to the optimality conditions of (4.1), we terminate Algorithm 1 whenever \( k \leq k_{\max} \) or the following condition is satisfied

\[
\| \nabla \bar{f}(V^{k+1}) - \nabla \bar{f}(U^k) - L_k (V^{k+1} - U^k) + \rho (\Gamma^{k+1} - \Gamma^k) \|_F \leq \tau_k \max(1, \eta),
\]

where \( \tau_{k+1} = \max(10^{-5}, 0.995 \tau_k) \) with \( \tau_0 = 0.005 \), and \( \eta > 0 \) is a constant related to test instances. Among others, \( \eta = \| C \|_F \) for the examples in Section 5.2-5.4, and \( \eta = \max_{1 \leq i \leq q} \| C_i \|_F \) for those in Section 5.5. A similar stopping condition, except \( \tau_{k+1} = \max(10^{-5}, 0.9 \tau_k) \), is used for Algorithm 2 and SDPT3 to solve (5.1). Consider that those penalty problems with smaller \( \rho \) are actually used to seek an appropriate \( \rho \). Hence, when \( \| V' \|_F^2 - \| V' \|_F^2 \) has a larger value (corresponding to a smaller \( \rho \)), we can calculate a very rough approximate critical point for (4.1). Inspired by this, during the testing, we take \( k_{\max} = 3 \) when \( \| V' \|_F^2 - \| V' \|_F^2 > 1 \) for Algorithm 1 and 2, but respectively set \( k_{\max} = 3000 \) and \( k_{\max} = 1000 \) for them when \( \| V' \|_F^2 - \| V' \|_F^2 \leq 1 \).

5.2. Comparisons with SDPRR and dcSDPT3 for Biq instances. In this part, we compare the performance of dcFAC with that of SDPRR and dcSDPT3 for the problem \( \max_{z \in \{0, 1\}^n} z^TAz \), which can be reformulated as (1.4) with \( p = n + 1 \) and \( f(X) = \langle C, X \rangle \) for \( C = -\frac{1}{4} \begin{pmatrix} 0 & e^TA \\ A^T & A \end{pmatrix} \). The matrix \( A \) is from the Biq Mac
Table 5.1: Numerical results of dcSDPT3, dcFAC and SDPRR for Biq Mac Library instances

| Name             | dcSDPT3 | dcFAC | Optval | Obj time | Obj time | Obj time | dcSDPT3 | dcFAC | Optval | Obj time | Obj time | Obj time |
|------------------|---------|-------|--------|----------|----------|----------|----------|-------|--------|----------|----------|----------|
| bqp100.1-1       | 100     | 7970  | 7848   | 333.3    | 7848     | 4.4      | 7938     | 408.6 | 100    | 7938     | 408.6    | 100      |
| bqp100.2-1       | 100     | 10721 | 10378 | 315.3    | 12722    | 8.7      | 12722    | 2.3   | 100    | 12722    | 2.3      | 100      |
| bqp100.3-1       | 100     | 11093 | 10226 | 10600    | 4.3      | 10121    | 462.2    | 100    | 10121   | 462.2    | 100      |
| bqp100.4-1       | 100     | 10038 | 10486 | 308.6    | 10106     | 4.5      | 10106    | 5.2   | 100    | 10106    | 5.2      | 100      |

Library in http://biqmac.uni-kl.de/biqmacLib.html. The SDPRR first uses the software SDPNAL+ [45, 40] to solve the SDP yields by removing the DC constraint in (1.3) but adding the valid inequalities \( X_{ij} + X_{ik} + X_{jk} \geq 1, X_{ij} - X_{ik} - X_{jk} \geq -1, X_{ij} + X_{ik} - X_{jk} \geq 1 \) and \( X_{ij} - X_{ik} + X_{jk} \geq -1 \) for all \( i \neq k \), and then impose 50 times random rounding technique [9] on the solution and select the best one from 50 feasible solutions. Table 5.1 reports the objective values of the outputs of the solvers and the CPU time taken by them, and the optimal values of these instances, where the gap value in red means the best for an instance.

We see that dcFAC and dcSDPT3 have much better performance than SDPRR does in terms of the quality of the outputs, and among the 119 instances, the outputs of dcSDPT3 and dcFAC respectively have 106 and 103 best ones, and their relative gaps to the optimal values are at most 1.824% except gka9b and gka10b. Since the data matrix from gka1b-gka10b has a special structure, i.e., the diagonal entries are from \([-63, 0]\) while the off-diagonal entries are from \([0, 100]\), the outputs of dcSDPT3 and dcFAC have a zero objective value for them. The CPU time of dcFAC is far less than that of dcSDPT3 and SDPRR, and for those examples with \( n = 250 \), dcFAC requires at most 6.0s but dcSDPT3 and SDPRR require at least 90s.
5.3. Comparisons with dcSDPT3 for G-set instances. Given a graph $G = (V, E)$ with $|V| = n$ and a weight matrix $W \in \mathbb{R}^n$, the max-cut problem partitions $V$ into two nonempty sets $(V, V\setminus V)$ so that the total weights of the edges in the cut is maximized. It can be reformulated as (1.4) with $p = n$ and $f(X) = (C, X)$ for $C = W - \frac{1}{n}\text{diag}(W)$. We solve the G-set instances with $W$ from http://www.stanford.edu/yeyey/yeyey/Gaet for 800 to 4000 variables. Table 5.1 reports the relative gap and infeasibility of the outputs of dcFAC and dcSDPT3 and the CPU time taken, where “-” means that the CPU time is more than 2 hours. Since it is impractical for an exact method, say BqCrunch [18], to yield the optimal values for these instances, Table 5.1 lists the known best values got with some advanced heuristic methods [44, 39].

| Name(s)   | Eval | dcSDPT3  | gap(%) | time infeas | dcFAC  | gap(%) | time infeas |
|-----------|------|----------|--------|-------------|--------|--------|-------------|
| G1(800)   | 11624| 0.052    | 442.7  | 2.6e-13     | 0.017  | 7.0     | 0.4e-10     |
| G2(800)   | 11620| 0.120    | 467.3  | 3.1e-11     | 0.060  | 7.2     | 5.8e-13     |
| G3(800)   | 11622| 0.124    | 440.2  | 3.1e-12     | 0.121  | 7.2     | 5.9e-10     |
| G4(800)   | 11640| 0.232    | 438.2  | 3.2e-12     | 0.180  | 7.2     | 5.9e-10     |
| G5(800)   | 11631| 0.146    | 441.5  | 3.2e-19     | 0.146  | 6.4     | 3.3e-10     |
| G6(800)   | 11614| 0.147    | 446.6  | 3.3e-19     | 0.147  | 6.4     | 3.3e-10     |
| G7(800)   | 2006 | 1.645    | 512.4  | 3.3e-12     | 1.047  | 6.5     | 2.2e-10     |
| G8(800)   | 2000 | 1.047    | 471.6  | 2.7e-11     | 1.047  | 6.5     | 2.2e-10     |
| G9(800)   | 2000 | 1.047    | 471.6  | 2.7e-11     | 1.047  | 6.5     | 2.2e-10     |
| G10(800)  | 2000 | 1.047    | 471.6  | 2.7e-11     | 1.047  | 6.5     | 2.2e-10     |
| G11(800)  | 2000 | 1.047    | 471.6  | 2.7e-11     | 1.047  | 6.5     | 2.2e-10     |
| G12(800)  | 2000 | 1.047    | 471.6  | 2.7e-11     | 1.047  | 6.5     | 2.2e-10     |
| G13(800)  | 2000 | 1.047    | 471.6  | 2.7e-11     | 1.047  | 6.5     | 2.2e-10     |
| G14(800)  | 2000 | 1.047    | 471.6  | 2.7e-11     | 1.047  | 6.5     | 2.2e-10     |
| G15(800)  | 2000 | 1.047    | 471.6  | 2.7e-11     | 1.047  | 6.5     | 2.2e-10     |
| G16(800)  | 2000 | 1.047    | 471.6  | 2.7e-11     | 1.047  | 6.5     | 2.2e-10     |
| G17(800)  | 2000 | 1.047    | 471.6  | 2.7e-11     | 1.047  | 6.5     | 2.2e-10     |
| G18(800)  | 2000 | 1.047    | 471.6  | 2.7e-11     | 1.047  | 6.5     | 2.2e-10     |
| G19(800)  | 2000 | 1.047    | 471.6  | 2.7e-11     | 1.047  | 6.5     | 2.2e-10     |
| G20(800)  | 2000 | 1.047    | 471.6  | 2.7e-11     | 1.047  | 6.5     | 2.2e-10     |
| G21(800)  | 2000 | 1.047    | 471.6  | 2.7e-11     | 1.047  | 6.5     | 2.2e-10     |
| G22(800)  | 2000 | 1.047    | 471.6  | 2.7e-11     | 1.047  | 6.5     | 2.2e-10     |
| G23(800)  | 2000 | 1.047    | 471.6  | 2.7e-11     | 1.047  | 6.5     | 2.2e-10     |
| G24(800)  | 2000 | 1.047    | 471.6  | 2.7e-11     | 1.047  | 6.5     | 2.2e-10     |
| G25(800)  | 2000 | 1.047    | 471.6  | 2.7e-11     | 1.047  | 6.5     | 2.2e-10     |
| G26(800)  | 2000 | 1.047    | 471.6  | 2.7e-11     | 1.047  | 6.5     | 2.2e-10     |
| G27(800)  | 2000 | 1.047    | 471.6  | 2.7e-11     | 1.047  | 6.5     | 2.2e-10     |
| G28(800)  | 2000 | 1.047    | 471.6  | 2.7e-11     | 1.047  | 6.5     | 2.2e-10     |
| G29(800)  | 2000 | 1.047    | 471.6  | 2.7e-11     | 1.047  | 6.5     | 2.2e-10     |
| G30(800)  | 2000 | 1.047    | 471.6  | 2.7e-11     | 1.047  | 6.5     | 2.2e-10     |
We see that the outputs of dcFAC have the least gap for almost all instances, though their infeasibility is a little worse than that of dcSDPT3. The relative gaps of the outputs for dcFAC and dcSDPT3 are respectively at most 2.870% and 3.091%. When \( n = 5000 \), the CPU time taken by dcSDPT3 is more than 2 hours, but dcFAC yields the desirable result for the instance with \( n = 20000 \) in 900 seconds. By comparing the results of dcSDPT3 with those in Table 5.1, we conclude that the use of \( \rho = 0.1, \sigma = 1.05 \) leads to its worse performance. We also compare the relative gaps of dcFAC with the relative gaps for the rounding of its outputs, and find that their maximal error is 8.20e-9. This means that the rounding of the final output has little influence on the objective value if the infeasibility is in the magnitude of 10^{-9}.

5.4. Comparisons with dcSDPT3 for OR-Library instances. This part compares the performance of dcFAC with that of dcSDPT3 for \( \max_{z \in \{0,1\}^n} z^T A z \), with \( A \in \mathbb{S}^n \) from the OR-Library \( \texttt{http://people.brunel.ac.uk/~mastjjb/jeb/orlib/bqpinfo.html} \). Table 5.2 reports the relative gap and infeasibility of their outputs, the CPU time taken, and the known best values obtained in [29] with an advanced heuristic method. We see that the relative gaps of the outputs for dcFAC and dcSDPT3 are respectively not more than 0.688% and 0.671%, and the outputs of dcFAC have the less relative gap for most instances with infeasibility less than 10^{-9}. When \( n = 2500 \), dcFAC yields the desired result in 40s but dcSDPT3 can not yield the result in 2h.

5.5. Numerical results of dcFAC for Palubeckis instances. This part provides the results of dcFAC for solving \( \max_{z \in \{0,1\}^n} z^T A z \) with \( A \in \mathbb{S}^n \) from the Palubeckis instances \( \texttt{https://www.personalas.ktu.lt/~ginpalu/} \), and the known best value obtained in [8] with an advanced heuristic method. Since these instances involve more than 3000 variables, and it is time consuming for dcSDPT3 to compute...
an instance, we do not compare the results of dcFAC with those of dcSDPT3. From Table 5.4, the outputs of dcFAC have the relative gaps at most 0.356% for the 21 instances.

5.6. Numerical comparisons with dcSNCG for UBPP instances. With $X = (1; x_1; \ldots; x_q)(1; x_1; \ldots; x_q)^T$, we can reformulate (1.1) with $\vartheta$ from (1.2) as (1.4) for $f(X) = \prod_{i=1}^q (C_i, X)$ with $C_i = \begin{pmatrix} a_i & b_i^T \\ b_i & B_i \end{pmatrix}$, where $b_i = (0_{n(i-1)}; \frac{1}{2}; 0_{n(q-i)}) \in \mathbb{R}^{nq}$ and $B_i = \text{BlkDiag}(0, \ldots, 0, Q_i, 0, \ldots, 0) \in \mathbb{S}^{nq}$ for $i = 1, \ldots, q$. We test the performance of dcFAC and dcSNCG for solving this class of examples with $q = 2$. To verify the efficiency of Algorithm 1 with varying $L_k$, we compare their performance with that of Algorithm 4.1 armed with [25, Algorithm 2] (dcFAC\$s for short), where Algorithm 2 of [25] is an MM method with linesearch technique for solving (4.1). In addition, we also compare their performance with that of GloptiPoly\$3 [12], a software for the Lasserre relaxation of polynomial programs. Let $N_k^r := \{ \alpha \in \mathbb{N}^k \mid \sum_{i=1}^k \alpha_i \leq r \}$ and $s(r) := (\frac{nq+r}{q})^k$ for $k, r \in \mathbb{N}$. The $r(q \geq q)$-order Lasserre relaxation of (1.1) is given by

$$\inf_{y} \left\{ \sum_{\alpha \in N_k^{s(r)}} p_{\alpha} y_{\alpha} \text{ s.t. } M_r(y) \in \mathbb{S}_+^n, M_{r-1}(h_i y) = 0, \ i = 1, \ldots, nq \right\},$$

where $p_{\alpha}$ is the component of the coefficient vector of $\vartheta(x)$, $h_i$ is the coefficient of $h_i(x) = x_i^2 - 1$ for $i = 1, 2, \ldots, n$, and $M_r(y)$ and $M_{r-1}(h_i y)$ are respectively the moment matrix of dimensions $s(r)$ and $s(r-1)$ (see [19] for the details).

The first group of problems is using $Q_i = \frac{1}{2} D_i$, $c_i = 2Q_i e$ and $a_i = e^T Q_i e + \omega_i$ with $D_i = \frac{1}{2} M_i$ for $i = 1, 2$, where the entries of each $M_i \in \mathbb{S}^q$ and $\omega_i$ are generated to obey the standard normal distribution. Such a problem is a reformulation of

$$(5.3) \min_{x, y \in \{0,1\}^l} \left\{ (x^T D_1 x + \omega_1)(y^T D_2 y + \omega_2) \right\}.$$ 

Table 5.5 reports the results of three solvers for solving (5.3) with different $l$ and those of GloptiPoly3 for solving its 2-order Lasserre relaxation. For $l = 2$ and 3, the three solvers deliver the same objective value as GloptiPoly3 does, which now becomes the optimal since the Lasserre relaxation provides a lower bound for the optimal value. For $l \geq 20$, GloptiPoly3 fails to deliver the result due to out of memory, while dcFAC, dcFAC\$s and dcSNCG can provide an approximate upper bound of the optimal value even for $l = 1000$ within 127$s$, 136$s$ and 627$s$, respectively. For 12 instances, the outputs of dcFAC and dcFAC\$s respectively have 8 and 6 best objective values, and dcFAC requires more CPU time than dcFAC\$s does due to the worse $L_0$.
### Table 5.5: Numerical results of dcFAC, dcFAC$_{ls}$, GloptiPoly3 and dcSNCG for (5.3)

| $\text{GPoly3}$ | dcSNCG | dcFAC | dcFAC$_{ls}$ |
|----------------|---------|--------|-------------|
| $l$ | obj | obj | time | infeas | obj | time | infeas | obj | time | infeas |
| 2 | -5.6696 | -5.6692 | 0.1 | 2.1e-5 | -5.6696 | 0.1 | 2.4e-9 | -5.6696 | 0.1 | 2.4e-9 |
| 8 | -22.3922 | -22.3923 | 0.6 | 1.8e-5 | -22.3922 | 1.7 | 1.5e-9 | -22.3922 | 5.6 | 1.5e-9 |
| 20 | * | -58.1403 | 1.3 | 4.7e-6 | -57.8066 | 6.0 | 1.1e-9 | 57.8066 | 28.6 | 1.0e-9 |

The second group of problems is using $Q_i = \frac{1}{4}(-1)^iW^i$, $a_i = \frac{1}{4}(-1)^{i+1}e^T W^i$ and $c_i = 0$ with $W^i = \overline{W}^i = \frac{W^i}{\|W^i\|}$ for $i = 1, 2$, where each $W^i \in S_l$ is chosen from the G-set and the Biq Mac Library. Such a problem is a reformulation of the generalized max-cut

$$\min_{x, y \in \{-1, 1\}^n} \frac{1}{4} \left( \sum_{i<j} m_{ij}^1 (1 - x_i x_j) \right) \left( \sum_{i<j} m_{ij}^2 (1 - y_i y_j) \right).$$

Table 5.5 reports the results of dcFAC, dcFAC$_{ls}$ and dcSNCG for solving problem (5.4) with different $(W^1, W^2)$. Among 14 instances, the outputs of dcFAC and dcFAC$_{ls}$ respectively have 8 and 7 best objective values, and dcFAC requires a little more CUP time than dcFAC$_{ls}$ does. When $n = 2000$, dcFAC$_{ls}$ and dcFAC can yield the result within 1500 s, but dcSNCG cannot yield the result within 2 hours.

### Table 5.6: Numerical results of dcFAC, dcFAC$_{ls}$ and dcSNCG for problem (5.4)

| $(W^1, W^2)$ | n | Obj | obj | Time | infeas | Obj | time | infeas | Obj | time | infeas |
|--------------|---|-----|-----|------|--------|-----|------|--------|-----|------|--------|
| (sing2.5, sing2.5) | 200 | -8.2944e+3 | 105.0 | 1.3e-5 | -8.4516e+3 | 96.0 | 2.3e-9 | -8.4686e+3 | 248.0 | 2.3e-9 |
| (sing3.0, sing3.0) | 200 | -8.1765e+3 | 93.6 | 1.9e-6 | -8.2312e+3 | 93.0 | 2.9e-9 | -8.2312e+3 | 148.9 | 2.8e-9 |
| (sing2.5, sing3.0) | 300 | -2.0570e+4 | 245.4 | 3.4e-5 | -2.0933e+4 | 115.4 | 1.8e-9 | -2.0877e+4 | 115.7 | 1.6e-9 |
| (sing3.0, sing3.0) | 300 | -2.0055e+4 | 246.9 | 3.4e-5 | -2.0933e+4 | 115.4 | 1.8e-9 | -2.0877e+4 | 115.7 | 1.6e-9 |

6. Conclusions. We have proposed a relaxation approach to the UBPP (1.1) by seeking a finite number of critical points for the DC penalized matrix program (1.6) with increasing penalty factors, and developed a globally convergent MM method.
with extrapolation to achieve such critical points. The rank-one projection of their outputs is shown under a mild condition to be an approximate feasible solution of the UBPP, and the upper bound of their objective values to the optimal value is also quantified. Numerical comparisons with SDPRR for 119 Biq Mac Library instances and with dcSNCG for 26 UBPP instances constructed with $q = 2$ show that dcFAC is remarkably superior to SDPRR and dcSNCG whether by the quality of the output or by the CPU time. The comparisons with dcSDPT3 for 119 Biq Mac Library instances and 112 UBQP instances indicate that dcFAC is comparable with dcSDPT3 if the latter is using the same updating rule of $\rho$ (only possible for small-scale instances), otherwise is superior to dcSDPT3 in the quality of solutions and the CPU time.

REFERENCES

[1] M. F. Anjos and H. Wolkowicz, Strengthened semidefinite relaxations via a second lifting for the max-cut problem, Discrete Applied Mathematics, 119 (2002), pp. 79–106.
[2] H. Krislock, J. Malick, P. Redont, and A. Soubyran, Proximal alternating minimization and projection methods for nonconvex problems: an approach based on the kurdyka-Łojasiewicz inequality, Mathematics of Operations Research, 35 (2010), pp. 438–457.
[3] S. J. Bi and S. H. Pan, Error bounds for rank constrained optimization problems and applications, Operations Research Letters, 44 (2016), pp. 336–341.
[4] S. Burer and R. D. C. Monteiro, A nonlinear programming algorithm for solving semidefinite programs via low-rank factorization, Mathematical Programming, 95 (2003), pp. 329–357.
[5] S. Burer, R. D. C. Monteiro, and Y. Zhang, Rank-two relaxation heuristics for max-cut and other binary quadratic programs, SIAM Journal on Optimization, 12 (2001), pp. 503–521.
[6] P. Chardaire and A. Sutter, A decomposition method for quadratic zero-one programming, Management Science, 41 (1994), pp. 704–712.
[7] T. R. Fu, D. D. Ge, and Y. Y. Ye, On doubly positive semidefinite programming relaxations, Journal of Computational Mathematics, 36 (2018), pp. 391–403.
[8] F. Glover, Z. P. Lü, and J. K. Hao, Diversification-driven tabu search for unconstrained binary quadratic problems, 4OR-A Quarterly Journal of Operations Research, 8 (2010), pp. 239–253.
[9] M. X. Goemans and D. P. Williamson, Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming, Journal of the Association for Computing Machinery, 42 (1995), pp. 1115–1145.
[10] S. M. He, Z. N. Li, and S. Z. Zhang, Approximation algorithms for discrete polynomial optimization, Journal of the Operations Research Society of China, 1 (2013), pp. 3–36.
[11] C. Helmberg and F. Rendl, Solving quadratic (0,1)-problems by semidefinite programs and cutting planes, Mathematical Programming, 82 (1998), pp. 291–395.
[12] D. Henrion, J. Lasserre, and J. Löfberg, Gloptipoly3: moments, optimization and semidefinite programming, Optimization Methods and Software, 24 (2009), pp. 761–779.
[13] A. D. Ioffe and J. V. Outrata, On metric and calmness qualification conditions in subdifferential calculus, Set-Valued Analysis, 16 (2008), pp. 199–227.
[14] Z. X. Jiang, X. Y. Zhao, and C. Ding, A proximal dc approach for quadratic assignment problem, Computational Optimization and Applications, (2021), https://doi.org/10.1007/s10589-020-00252-5.
[15] S. Y. Kim, M. Kohima, and K. C. Toh, A lagrangian-dnn relaxation: a fast method for computing tight lower bounds for a class of quadratic optimization problems, Mathematical Programming, 156 (2016), pp. 161–187.
[16] G. Kochenberger, J. K. Hao, F. Glover, M. Lewis, Z. P. Lü, H. B. Wang, and Y. Wang, The unconstrained binary quadratic programming problem: a survey, Journal of Global Optimization, 28 (2004), pp. 58–81.
[17] N. Krislock, J. Malick, and F. Roupin, Improved semidefinite bounding procedure for solving max-cut problems to optimality, Mathematical Programming, 143 (2014), pp. 61–86.
[18] N. Krislock, J. Malick, and F. Roupin, Biqcrunch: a semidefinite branch-and-bound method for solving binary quadratic problems, ACM Transactions on Mathematical Software, 43 (2017), pp. 1–23.
[19] J. B. Lasserre, Global optimization with polynomials and the problem of moments, SIAM Journal on Optimization, 11 (2001), pp. 796–817.
[20] H. A. Le Thi and T. Pham Dinh, Dc programming and dca: thirty years of developments, Mathematical Programming B, Special Issue dedicated to: DC Programming–Theory, Algo-
algorithms and Applications, 169 (2018), pp. 5–68.

[21] D. Li, X. L. Sun, and C. L. Liu, An exact solution method for unconstrained quadratic 0-1 programming: a geometric approach, Journal of Global Optimization, 52 (2012), pp. 797–829.

[22] 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), pp. 51–96.

[23] X. D. Li, D. F. Sun, and K. C. Toh, Qsdpnal: A two-phase augmented lagrangian method for convex quadratic semidefinite programming, Mathematical Programming Computation, 10 (2018), pp. 703–743.

[24] T. X. Liu, T. K. Pong, and A. Takeda, A refined convergence analysis of pdca, with applications to simultaneous sparse recovery and outlier detection, Computational Optimization and Applications, 73 (2019), pp. 69–100.

[25] T. X. Liu, T. K. Pong, and A. Takeda, A successive difference-of-convex approximation method for a class of nonconvex nonsmooth optimization problems, Mathematical Programming, 176 (2019), pp. 339–367.

[26] D. R. Luke, Prox-regularity of rank constraint sets and implications for algorithms, Journal of Mathematical Imaging and Vision, 47 (2013), pp. 231–238.

[27] J. Luo, K. Pattipati, P. Willett, and F. Hasegawa, Near-optimal multiuser detection in synchronous cdma using probabilistic data association, IEEE Communications Letters, 5 (2001), pp. 361–363.

[28] Y. Nesterov, A method of solving a convex programming problem with convergence rate o(1/k^2), Soviet Math. Dokl., 27 (1983), pp. 372–376.

[29] G. Palubeckis, Multistart tabu search strategies for the unconstrained binary quadratic optimization problem, Annals of Operations Research, 131 (2004), pp. 259–282.

[30] J. S. Pang, M. Razaviyayn, and A. Alvarado, Computing b-stationary points of nonsmooth dc programs, Mathematics of Operations Research, 42 (2017), pp. 95–118.

[31] P. M. Pardalos and G. R. Rodgers, A branch and bound algorithm for maximum clique problem, Computers & Operations Research, 19 (1992), pp. 363–375.

[32] T. Pham Din and H. A. Le Thi, Convex analysis approach to dc programming: theory, algorithms and applications, Acta mathematica vietnamica, 22 (1997), pp. 289–355.

[33] T. Pham Din and H. A. Le Thi, An efficient combined dca and b&b using dc/sdp relaxation for globally solving binary quadratic programs, Journal of Global Optimization, 48 (2010), pp. 595–632.

[34] T. Pham Din and H. A. Le Thi, Recent advances in dc programming and dca, Transactions on Computational Intelligence XIII, 8342 (2014), pp. 1–37.

[35] A. T. Phillips and J. B. Rosen, A quadratic assignment formulation of the molecular conformation problem, Journal of Global Optimization, 4 (1994), pp. 229–241.

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

[37] F. Rendl, G. Rinaldi, and A. Wiegele, Solving max-cut to optimality by intersecting semidefinite and polyhedral relaxations, Mathematical Programming, 121 (2010), pp. 307–335.

[38] R. T. Rockafellar and R. J.-B. Wets, Variational Analysis, Springer, 1998.

[39] V. P. Shylo, F. Glover, and I. V. Sergienko, Teams of global equilibrium search algorithms for solving the weighted maximum cut problem in parallel, Cybernetics and Systems Analysis, 51 (2015), pp. 16–24.

[40] D. F. Sun, K. C. Toh, Y. C. Yuan, and X. Y. Zhao, Sdpt3+: A matlab software for semidefinite programming with bound constraints (version 1.0), Optimization Methods & Software, 35 (2000), pp. 209–211.

[41] R. Y. Sun and Z. Q. Luo, Guaranteed matrix completion via non-convex factorization, IEEE Transactions on Information Theory, 62 (2016), pp. 6535–6579.

[42] K. C. Toh, M. J. Todd, and R. H. Tutuncu, Sdpt3—a matlab software package for semidefinite programming, version 4.0, Optimization Methods & Software, 11 (1999).

[43] Z. W. Wen and W. T. Yin, A feasible method for optimization with orthogonality constraints, Mathematical Programming, 142 (2013), pp. 397–434.

[44] Q. H. Wu, Y. Wang, and Z. P. Li, A tabu search based hybrid evolutionary algorithm for the max-cut problem, Applied Soft Computing, 34 (2015), pp. 827–837.

[45] L. Q. Yang, D. F. Sun, and K. C. Toh, Sdpt3+: A majorized semismooth newton-cg augmented lagrangian method for semidefinite programming with nonnegative constraints, Mathematical Programming Computation, 7 (2015), pp. 331–366.
SUPPLEMENTARY MATERIALS: A RELAXATION APPROACH TO UBPP BASED ON EQUIVALENT DC PENALIZED FACTORIZED MATRIX PROGRAMS

YITIAN QIAN† AND SHAOHUA PAN‡

SM1. Proof of Lemma 2.3.

Proof. Let \( h(z) := -\|z\|_\infty \) for \( z \in \mathbb{R}^p \). Notice that \( \psi \) is the spectral function associated to \( h \), i.e., \( \psi(Z) = h(\lambda(Z)) \) for any \( Z \in \mathcal{P} \). By [SM2, Theorem 6], we have

\[
\partial \psi(X) = \{ P \text{Diag}(\xi) P^T \mid P \in \mathcal{O}(X), \xi \in \partial h(\lambda(X)) \}.
\]

In addition, by [SM5, Corollary 9.21] and the expression of \( h \), it is easy to calculate that

\[
\partial h(\lambda(X)) = \begin{cases} 
- \text{sign}(\lambda_j(X))e_j & \text{if } \lambda(X) \neq 0; \\
\{e_j, -e_j\} & \text{if } \lambda(X) = 0.
\end{cases}
\]

From the last two equations, we obtain the first part. When \( \text{rank}(X) = 1 \), it is easy to check that \( h \) is differentiable at \( \lambda(X) \), and the result holds by [SM2, Theorem 6]. \( \square \)

SM2. Proof of Proposition 4.3.

Proof. (i) From the definition of \( V^{k+1} \) and the feasibility of \( V^k \) to problem (4.3),

\[
\langle \nabla \tilde{f}(U^k) + \rho \Gamma^k, V^{k+1} \rangle + \rho \| V^{k+1} - U^k \|_F^2 + (L_k/2)\| V^{k+1} - U^k \|_F^2 \\
\leq \langle \nabla \tilde{f}(U^k) + \rho \Gamma^k, V^k \rangle + \rho \| V^k - U^k \|_F^2 + (L_k/2)\| V^k - U^k \|_F^2.
\]

(4.2.1)

Notice that \( \Gamma^k \in \partial \tilde{\psi}(V^k) \subseteq -\partial(-\tilde{\psi})(V^k) \). From the convexity of \( -\tilde{\psi} \) and [SM4, Theorem 23.5], \( \tilde{\psi}(V^k) - (-\tilde{\psi})^\star(-\Gamma^k) = (\Gamma^k, V^k) \). Along with the expression of \( \Theta_\rho \),

\[
\Theta_\rho(V^{k+1}, \Gamma^k, V^k) \leq \tilde{f}(V^{k+1}) + \langle \nabla \tilde{f}(U^k), V^k - V^{k+1} \rangle + \rho \| V^k - V^{k+1} \|_F^2 + \rho \tilde{\psi}(V^k) \\
+ \frac{\gamma L}{2} \| V^{k+1} - V^k \|_F^2 + \frac{L_k}{2} \| V^k - U^k \|_F^2 - \frac{L_k}{2} \| V^{k+1} - U^k \|_F^2, \\
\leq \tilde{f}(V^k) + \rho \| V^k \|_F^2 + \rho \tilde{\psi}(V^k) + \frac{\gamma L}{2} \| V^{k+1} - V^k \|_F^2 \\
+ \frac{L_k + L_f}{2} \| V^k - U^k \|_F^2 - \frac{L_k - L_f}{2} \| V^{k+1} - U^k \|_F^2,
\]

(4.2.2)

where the last inequality is using (4.2a) with \( V = V^{k+1}, Z = U^k \) and (4.2b) with \( V = V^k, Z = U^k \). Notice that \( \tilde{\psi}(V^k) - \langle V^k, \Gamma^{k-1} \rangle \leq (-\tilde{\psi})^\star(-\Gamma^{k-1}) \). Together with

---

* July 28, 2021

**Funding**: This work is funded by the National Natural Science Foundation of China under project No.11571120.

†School of Mathematics, South China University of Technology (mayttqian@mail.scut.edu.cn).

‡School of Mathematics, South China University of Technology (shhpan@scut.edu.cn).

SM1
the definition of $\Theta_\rho$ and $U^k = V^k + \beta_k (V^k - V^{k-1})$, it follows that

$$
\Theta_\rho(V^{k+1}, \Gamma^k, V^k) \leq \Theta_\rho(V^k, \Gamma^{k-1}, V^{k-1}) + \frac{\gamma}{2} \|V^{k+1} - V^k\|_F^2 + \frac{L_k + L_f}{2} \|V^k - U^k\|_F^2 \\
- \frac{L_k - L_f}{2} \|V^{k+1} - U^k\|_F^2 - \frac{\gamma L}{2} \|V^k - V^{k-1}\|_F^2 \\
\leq \Theta_\rho(V^k, \Gamma^{k-1}, V^{k-1}) - \frac{L_k - \gamma L - L_f}{2} \|V^{k+1} - V^k\|_F^2 \\
- \frac{\gamma L - 2L_f \beta_k^2}{2} \|V^k - V^{k-1}\|_F^2 + (L_k - L_f) \beta_k (V^{k+1} - V^k, V^k - V^{k-1}).
$$

Since $2(L_k - L_f) \beta_k (V^{k+1} - V^k, V^k - V^{k-1}) \leq \mu (L_k - L_f)^2 \|V^{k+1} - V^k\|_F^2 + \frac{\beta_k^2}{\mu} \|V^k - V^{k-1}\|_F^2$ for any $\mu > 0$, the following inequality holds for any $\mu > 0$:

$$
\Theta_\rho(V^{k+1}, \Gamma^k, V^k) \leq \Theta_\rho(V^k, \Gamma^{k-1}, V^{k-1}) - \frac{\gamma L - 2L_f \beta_k^2}{2} \|V^k - V^{k-1}\|_F^2 \\
- \frac{(L_k - \gamma L - L_f)^2 \mu}{2} \|V^{k+1} - V^k\|_F^2.
$$

By taking $\mu = \frac{L_k - L_f}{(L_k - L_f)^2}$, the desired result follows from the last inequality.

(ii)-(iii) The boundedness of $\{V^k\}$ is trivial. Since $\Gamma^k \in \partial \psi(V^k)$, its boundedness is due to Remark 4.1 (b). So, it suffices to prove part (iii). By part (i), the sequence $\{\Theta_\rho(V^k, \Gamma^{k-1}, V^{k-1})\}$ is nonincreasing. Notice that $\Theta_\rho$ is proper lsc and level-bounded. From [SM5, Theorem 1.9], $\Theta_\rho$ is bounded below. This means that the limit $\varpi^*: = \lim_{k \to \infty} \Theta_\rho(V^k, \Gamma^k, V^{k-1})$ exists. From $\nu_k \geq \frac{\gamma L - 2L_f \beta_k^2}{(L_k - L_f)^2} > 0$ and part (i), we obtain $\lim_{k \to \infty} \|V^k - V^{k-1}\|_F = 0$. We next show that $\Theta_\rho \equiv \varpi^*$ on the set $\Delta_\rho$. Pick any $(\tilde{V}, \tilde{\Gamma}, \tilde{U}) \in \Delta_\rho$. From part (ii), there exists $K \subseteq \mathbb{N}$ such that $\lim_{k \to \infty} (V^k, \Gamma^k, V^{k-1}) = (\tilde{V}, \tilde{\Gamma}, \tilde{U})$. From the expression of $\Theta_\rho(V^k, \Gamma^{k-1}, V^{k-1})$,

$$
\varpi^* = \lim_{K \to \infty} \left[ \tilde{f}(V^k) + \rho(\Gamma^{k-1}, V^k) + \rho\|V^k\|_F^2 + \rho(-\tilde{\psi})^*(\Gamma^{k-1}) \right] \\
= \tilde{f}(\tilde{V}) + \rho(\tilde{\Gamma}, \tilde{V}) + \rho\|\tilde{V}\|_F^2 + \rho(-\tilde{\psi})^*(\tilde{\Gamma}) = \Theta_\rho(\tilde{V}, \tilde{\Gamma}, \tilde{U}) = \Theta_\rho(\tilde{V}, \tilde{\Gamma}, \tilde{U}),
$$

where the second equality is by the continuity of $(-\tilde{\psi})^*$ since $(-\tilde{\psi})^*(U) = \frac{1}{2}\|U\|_F^2$ by [SM5, Proposition 11.21], the third one is using $\tilde{V} \in S$ implied by $\{V^k\}_{k \in K} \subseteq S$, and the last one is using $\tilde{V} = \tilde{U}$ implied by $\lim_{k \to \infty} \|V^k - V^{k-1}\|_F = 0$.

(iv) By invoking [SM5, Exercise 8.8], for any $(V, \Gamma, U) \in S \times \mathbb{R}^{m \times p} \times \mathbb{R}^{m \times p}$,

$$
(\text{SM2.3}) \quad \partial \Theta_\rho(V, \Gamma, U) = \begin{bmatrix}
\nabla \tilde{f}(V) + 2\rho V + \rho \Gamma + \gamma L(V - U) + N_\delta(V) \\
\rho V - \rho (-\tilde{\psi})^*(\Gamma) \\
\gamma L(U - V)
\end{bmatrix}.
$$

From the definition of $V^k$, $0 \in \nabla \tilde{f}(U^{k-1}) + \rho \Gamma^{k-1} + 2\rho V^k + L_{k-1}(V^k - U^{k-1}) + N_\delta(V^k)$. Since $\Gamma^{k-1} \in \partial \psi(V^{k-1}) \subseteq \partial (-\psi)(V^{k-1})$, by invoking [SM4, Theorem 23.5] we have $V^{k-1} \in \partial (-\psi)^*(-\Gamma^{k-1})$. By combining with the last equality yields, it follows that

$$
\begin{bmatrix}
\nabla \tilde{f}(V^k) - \nabla \tilde{f}(U^{k-1}) - L_{k-1}(V^k - U^{k-1}) + \gamma L(V^k - V^{k-1}) \\
\rho (V^k - V^{k-1}) \\
\gamma L(V^{k-1} - V^k)
\end{bmatrix} \in \partial \Theta_\rho(V^k, \Gamma^{k-1}, V^{k-1}).
$$

This along with $U^{k-1} = V^{k-1} + \beta_{k-1}(V^{k-1} - V^{k-2})$ implies the desired result. \qed
SM3. Theoretical analysis of Algorithm 5.1. We first provide the convergence of Algorithm 2. From the Lipschitz continuity of $\nabla f$ on $\mathbb{B}_\Omega$, for every $X \in \mathbb{B}_\Omega$, 

\begin{align}
(SM3.1a) \quad & f(X) \leq f(Y) + \langle \nabla f(Y), X - Y \rangle + \frac{L_f}{2} \| X - Y \|_F^2; \\
(SM3.1b) \quad & -f(X) \leq -f(Y) - \langle \nabla f(Y), X - Y \rangle + \frac{L_f}{2} \| X - Y \|_F^2,
\end{align}

where $L_f$ denotes the Lipschitz constant of $\nabla f$ in $\mathbb{B}_\Omega$. Algorithm 2 is similar to the proximal DC algorithm proposed in [SM3], but the conclusion of [SM3, Theorem 3.1] cannot be directly applied to it since the convexity of $f$ is not required here. Inspired by the analysis technique in [SM3], we define the following potential function

$$
\Xi_\rho(X, W, Z) := f(X) + \rho \langle I + W, X \rangle + \delta_\Omega(X) + \rho \delta_B(-W) + \frac{L_f}{2} \| X - Z \|_F^2
$$

associated to $\rho > 0$, where $B := \{ Z \in S^p \mid \| Z \|_* \leq 1 \}$ is the nuclear norm unit ball.

**Proposition SM3.1.** Let $\{ (X^k, W^k) \}$ be the generated by Algorithm 2. Then,

(i) $\Xi_\rho(X^{k+1}, W^k, X^{k}) \leq \Xi_\rho(X^k, W^{k-1}, X^{k-1}) - \frac{L_f - (L_k + L_{k+1})}{2} \| X^k - X^{k-1} \|_F^2$;

(ii) the sequence $\{ (X^k, W^k) \}$ is bounded, and consequently, the cluster point set of $\{ (X^k, W^{k-1}, X^{k-1}) \}$, denoted by $\mathcal{Y}_\rho$, is nonempty and compact;

(iii) the limit $\omega^* := \lim_{k \to \infty} \Xi_\rho(X^k, W^{k-1}, X^{k-1})$ exists whenever $\beta < \sqrt{\frac{L_f}{L_0 + L_f}}$, and moreover, $\Xi_\rho(X', W', Z') = \omega^*$ for every $(X', W', Z') \in \mathcal{Y}_\rho$;

(iv) for all $k \in \mathbb{N}$, with $\eta = \sqrt{9L_f^2 + 4L_0^2 + \rho^2}$ it holds that

$$
\text{dist}(0, \partial \Xi_\rho(X^k, W^{k-1}, X^{k-1})) \leq \eta \left[ \| X^k - X^{k-1} \|_F + \| X^{k-1} - X^{k-2} \|_F \right].
$$

**Proof.** (i) By the definition of $X^{k+1}$, the strong convexity of the objective function of (5.2), and the feasibility of $X^k$ to the subproblem (5.2), it follows that

$$
\langle \nabla f(Y^k) + \rho \langle I + W^k, X^{k+1} \rangle + (L_k/2) \| X^{k+1} - Y^k \|_F^2 \\
\leq \langle \nabla f(Y^k) + \rho \langle I + W^k, X^k \rangle + (L_k/2) \| X^k - Y^k \|_F^2 - \rho \delta_B(-W) \rangle - (L_k/2) \| X^{k+1} - X^k \|_F^2,
$$

which, after a suitable rearrangement, can be equivalently rearranged as

$$
\rho \langle I + W^k, X^{k+1} \rangle \leq \langle \nabla f(Y^k), X^k - X^{k+1} \rangle + \rho \langle I + W^k, X^k \rangle + 0.5L_k \| X^k - Y^k \|_F^2 \\
- 0.5L_k \| X^{k+1} - X^k \|_F^2 - 0.5L_k \| X^{k+1} - Y^k \|_F^2.
$$

(SM3.2)

Since $W^k \in \partial \psi(X^k) \subseteq -\partial(-\psi)(X^k)$ and the spectral function is the support of $B$, we have $-W^k \in B$ and $-\langle W^k, X^k \rangle = \| X^k \| \geq -\langle W^{k-1}, X^k \rangle$ by [SM4, Corollary 23.5.3]. Thus, for each $k \in \mathbb{N}$, $\delta_B(W^k) = 0$ and $\langle I + W^k, X^k \rangle \leq \langle I + W^{k-1}, X^k \rangle$. Together with the definition of $\Xi_\rho$ and (SM3.2), it follows that

$$
\Xi_\rho(X^{k+1}, W^k, X^k) \leq f(X^{k+1}) + \langle \nabla f(Y^k), X^k - X^{k+1} \rangle + \rho \langle I + W^{k-1}, X^k \rangle \\
+ \frac{L_k}{2} \| X^k - Y^k \|_F^2 - \frac{L_k}{2} \| X^{k+1} - Y^k \|_F^2 - \frac{L_k - L_f}{2} \| X^k - X^k \|_F^2,
$$

(SM3.3)

$$
\leq f(X^k) + \rho \langle I + W^{k-1}, X^k \rangle + \frac{L_k + L_f}{2} \| X^k - Y^k \|_F^2 \\
- \frac{L_k - L_f}{2} \| X^{k+1} - Y^k \|_F^2 - \frac{L_k - L_f}{2} \| X^k - X^k \|_F^2,
$$
where the second inequality is obtained by using (SM3.1a) with \( X = X^{k+1}, Y = Y^k \), and (SM3.1b) with \( X = X^{k+1}, Y = Y^k \). Now substituting \( Y^k = X^k + \beta_k (X^k - X^{k-1}) \) into the last inequality and using \( L_k \geq L_f \) yields

\[
\Xi_\rho(X^{k+1}, W^k, X^k) \leq \Xi_\rho(X^k, W^{k-1}, X^{k-1}) - \frac{L_f - (L_k + L_f)\beta_k^2}{2} \|X^k - X^{k-1}\|_F^2 - \frac{L_k - L_f}{2} \|X^k - Y^k\|_F^2 - \frac{L_k - L_f}{2} \|X^k - X^k - X^{k-1}\|_F^2
\]

\[
\leq \Xi_\rho(X^k, W^{k-1}, X^{k-1}) - \frac{L_f - (L_k + L_f)\beta_k^2}{2} \|X^k - X^{k-1}\|_F^2.
\]

(ii)-(iii) Part (ii) is immediate by noting that \( \{X^k\} \subseteq \Omega \) and \( \{W^k\} \subseteq \mathbb{B} \). Next we prove part (iii). By part (i), the sequence \( \{\Xi_\rho(X^k, W^{k-1}, X^{k-1})\} \) is nonincreasing. Notice that \( \Xi_\rho \) is proper lsc and level-bounded. By [SM5, Theorem 1.9], it is bounded below. So, the limit \( \omega^* \) is well defined. By part (i) and \( L_f - (L_k + L_f)\beta_k^2 \geq L_f - (L_0 + L_f)\beta^2 > 0 \), we have \( \lim_{k \to \infty} \|X^k - X^{k-1}\|_F = 0 \). We next show that \( \Xi_\rho \equiv \omega^* \) on the set \( \Upsilon_\rho \). Pick any \((\hat{X}, \hat{W}, \hat{Z}) \in \Upsilon_\rho \). By part (ii), there exists an index set \( \mathcal{K} \subseteq \mathbb{N} \) such that

\[
\{X^k, W^{k-1}, X^{k-1}\} = (\hat{X}, \hat{W}, \hat{Z}).
\]

Along with the expression of \( \Xi_\rho \),

\[
\omega^* = \lim_{k \in \mathcal{K} \to \infty} \Xi_\rho(X^k, W^{k-1}, X^{k-1}) = \lim_{k \in \mathcal{K} \to \infty} \left[ f(X^k) + \rho((I + W^{k-1}, X^k) \right]
\]

\[
= f(\hat{X}) + \rho(\hat{I} + \hat{W}, \hat{X}) = \Xi_\rho(\hat{X}, \hat{W}, \hat{Z}) = \Xi_\rho(\hat{X}, \hat{W}, \hat{Z}),
\]

and the last one is due to \( \hat{X} = \hat{Z} \), implied by \( \lim_{k \to \infty} \{X^k, X^{k-1}\} = (\hat{X}, \hat{Z}) \).

(iv) Notice that \( 0 \in \nabla f(Y^{k-1}) + \rho(I + W^{k-1}) + L_{k-1}(X^k - Y^{k-1}) + N_\Omega(X^k) \) by the optimality condition of (5.2). Recall that \( W^{k-1} \in \partial \psi(X^{k-1}) \subseteq -\partial(\psi(X^{k-1})) \) and the conjugate of the spectral function is \( \delta_\beta \). By [SM4, Theorem 23.5], we have \( X^{k-1} \in \partial \delta_\beta(-W^{k-1}) = \mathcal{N}_\beta(-W^{k-1}) \). Together with the expression of \( \Xi_\rho \), we have

\[
\left[ \nabla f(X^k) - \nabla f(Y^{k-1}) + L_k(X^k - X^{k-1}) - L_{k-1}(X^k - Y^{k-1}) \rho(X^k - X^{k-1}) \right] \in \partial \Xi_\rho(X^k, W^{k-1}, X^{k-1}).
\]

This implies that the desired inequality holds. Thus, we complete the proof. 

Remark SM3.2. (a) When \( f \) is convex, the coefficient \( L_f \) appearing in (SM3.3) can be removed. So, the restriction on \( \mathcal{F} \) in part (iii) can be improved as \( \mathcal{F} < \sqrt{L_f/L_0} \). This coincides with the requirement of [SM3, Proposition 3.1] for the convex function \( f \).

(b) Let \((\hat{X}, \hat{W}) \) be an accumulation point of \( \{(X^k, W^k)\} \). By the outer semicontinuity of \( N_\Omega \) and \( \partial \psi \), we have \( \hat{W} \in \partial \psi(\hat{X}) \) and \( 0 \in \nabla f(\hat{X}) + \rho(I + \hat{W}) + N_\Omega(\hat{X}) \) which by the expression of \( \partial \Xi_\rho \) and Definition 3.2 implies that \( \Pi_1(\Upsilon_\rho) \subseteq \Pi_1(\text{crit } \Xi_\rho) \subseteq \Upsilon_{\bar{\rho}} \), where \( \Pi_1(\Upsilon_{\bar{\rho}}) := \{Z \in \mathcal{S}^p \mid \exists W \text{ s.t. } (Z, W, Z) \in \Upsilon_\rho \} \).

By [SM1, Section 4.3], the indicator functions \( \delta_\Omega \) and \( \delta_\beta \) are semialgebraic, which implies that \( \Xi_\rho \) is a KL function (see [SM1] for the detail). By using Proposition SM3.1 and the same arguments as those for [SM3, Theorem 3.1] (see also [SM1, Theorem 3.1]), we obtain the following conclusion.

Theorem SM3.3. Let \( \{(X^k, W^k)\} \) be the sequence generated by Algorithm 2 from \( X^0 = X^1 \) with \( \mathcal{F} < \sqrt{L_f/L_0} \) for solving (1.5) associated to \( \rho_1 \). Then, the sequence
$\{X^k\}$ is convergent, and its limit is a critical point of (1.5) associated to $\rho_1$. If this limit is rank-one, it is also a local optimal solution of (1.4).

We have provided a convergent algorithm to seek a critical point of the subproblems in Algorithm 5.1. Next we focus on the stopping criterion of Algorithm 5.1 which aims to seek an approximate rank-one critical point. When this criterion occurs at some $l < l_{\text{max}}$, we say that Algorithm 5.1 exits normally. The following proposition states that under a certain condition Algorithm 5.1 can exit normally.

**Proposition SM3.4.** Fix an arbitrary integer $l \geq 0$. Suppose that $X^l \in \Omega$ satisfies $\langle I, X^l \rangle - \|X^l\| \leq c_0$ for some $c_0 \in (0, 1)$. Then for any given $\varepsilon \in (0, c_0]$, the following results hold for the sequence $\{X^k\}$ generated by Algorithm 2 from $X^0 = X^l$ with $\rho_l \geq \max \left\{ \frac{2c_0}{(1 - \sqrt{1 - 0.5p^2})\sqrt{1 - c_0}}, \frac{2c_0}{\varepsilon} \right\}$ for $\varepsilon = 6.5(L_f + L_0)p^2 + 2p\|\nabla f(I)\|_F$:

(i) for each integer $k \geq 0$ with $\varepsilon \leq \langle I, X^k \rangle - \|X^k\| \leq c_0$,

$$\|X^{k+1}\| \geq \|X^k\| + 0.5(1 - \sqrt{1 - 0.5p^2})\sqrt{1 - c_0};$$

(ii) there exists $\bar{k} \leq \frac{2(c_0 - \varepsilon)}{(1 - \sqrt{1 - 0.5p^2})\sqrt{1 - c_0}} + 1$ such that $\langle I, X^k \rangle - \|X^k\| \leq \varepsilon$ for all $k \geq \bar{k}$.

**Proof.** (i) For each $k \in \mathbb{N}$, from the definition of $X^{k+1}$, for any $X \in \Omega$ we have

$$\rho_l \langle W^k, X - X^{k+1} \rangle \leq \langle \nabla f(Y^k), X - X^{k+1} \rangle + \frac{L_k}{2} \|X - Y^k\|_F^2 - \frac{L_k}{2} \|X^{k+1} - Y^k\|_F^2$$

$$\leq \langle \nabla f(Y^k), X - X^{k+1} \rangle + \frac{L_k}{2} \left( \|X\|_F^2 + 2\|X^{k+1} - X\|_F \|Y^k\|_F \right)$$

$$\leq \langle \nabla f(Y^k) - \nabla f(I) + \nabla f(I), X - X^{k+1} \rangle + 6.5L_0p^2$$

$$\leq (L_f)\langle Y^k - I, X - X^{k+1} \rangle + \|\nabla f(I)\|_F \|X - X^{k+1}\|_F + 6.5L_0p^2 \leq \bar{\varepsilon}$$

where the third inequality is by $L_k \leq L_0$ for all $k \in \mathbb{N}$ and $\|X\|_F \leq p$ for $X \in \Omega$. So

$$\bar{\varepsilon} \rho_l \langle W^k, X \rangle \leq \frac{\bar{\varepsilon}}{\rho_l} + \frac{\bar{\varepsilon}}{\rho_l} \|W^k\|_F \|X^{k+1}\| = \frac{\bar{\varepsilon}}{\rho_l} + \|X^{k+1}\|.$$ 

Let $X^k$ have the eigenvalue decomposition $UD_iag(\lambda(X))U^T$ with $U = [u_1 \cdots u_p] \in \mathbb{R}^p$. Since $\langle I, X^k \rangle - \|X^k\| \leq c_0 < 1$ and $\text{Diag}(X^k) = e$, for every $j \in \{1, 2, \ldots, p\}$,

$$\lambda_1(X^k) u_{i_j}^2 = 1 - \sum_{i=2}^p \lambda_i(X^k) u_{i_j}^2 \geq 1 - \sum_{i=2}^p \lambda_i(X^k) \geq 1 - c_0 > 0.$$

Take $\tilde{X} = \lambda_1(X^k) u_{i_1} u_{i_2}^T$ with $u_{i_j} = \frac{u_{i_j}}{\sqrt{\|X^k\|/u_{i_j}^2}}$ for $j = 1, \ldots, p$. It is easy to check that $\tilde{X} \in \Omega$. Now using (SM3.6) with $X = \tilde{X}$ and recalling that $W^k = u_{i_1} u_{i_2}^T$, we obtain

$$\frac{\bar{\varepsilon}}{\rho_l} + \|X^{k+1}\| \geq \langle W^k, \tilde{X} \rangle = \|X^k\|(u_{i_1}^T u_{i_2})^2 = \left( \sqrt{u_{i_1}^2 + \cdots + u_{i_2}^2} \right)^2$$

$$\geq \|X^k\| + \left( \sqrt{u_{i_1}^2 + \cdots + u_{i_2}^2} - \|X^k\| \right) \sqrt{u_{i_1}^2 + \cdots + u_{i_2}^2}$$

$$\geq \|X^k\| + \left( \sqrt{u_{i_1}^2 + \cdots + u_{i_2}^2} - \|X^k\| \right) \sqrt{\|X^k\|}$$

$$\|X^{k+1}\| \geq \|X^k\| + \left( \sqrt{u_{i_1}^2 + \cdots + u_{i_2}^2} - \|X^k\| \right) \sqrt{\|X^k\|}.$$ 

where the second inequality is using $\sqrt{u_{l1}^2 + \cdots + u_{lp}^2} \geq \sqrt{\|X^k\|}$ implied by (SM3.7),
and the last equality is by $\sum_{j=1}^p u_{lj}^2 = 1$. Since $\|u_l\| = 1$, there exists $j \in \{1, \ldots, p\}$ such that $u_{lj}^2 \leq \frac{1}{p}$. Note that $1 - \sqrt{\|X^k\|u_{lj}^2} \geq 0$ for all $j = 1, \ldots, p$. From (SM3.8),

$$
\rho_l^{-1}v + \|X^{k+1}\| \geq \|X^k\| + \left[\sqrt{u_{lj}^2} - \sqrt{\|X^k\|u_{lj}^2}\right] \sqrt{\|X^k\|}
= \|X^k\| + \left(1 - \sqrt{\|X^k\|u_{lj}^2}\right) \sqrt{\|X^k\|}
\geq \|X^k\| + \left(1 - \sqrt{\|X^k\|/p}\right) \sqrt{\|X^k\|}
\geq \|X^k\| + \left(1 - \sqrt{1 - 0.5p^{-1}c}\right) \sqrt{1 - c_0}
$$

where the second inequality is due to $u_{lj}^2 \leq \frac{1}{p}$, and the last one is using $p - \|X^k\| \geq \frac{c_0}{p}$
and (SM3.7). Together with $\rho_l \geq \frac{1}{(1 - \sqrt{1 - 0.5p^{-1}c}) \sqrt{1 - c_0}}$, we get the desired result.

(ii) Since the proof is similar to that of Proposition 4.6 (ii), we here delete it.

Unlike for Algorithm 4.1, now we can not provide a suitable condition to ensure that some $X^l \in \Omega$ with \(\langle I, X^l \rangle - \|X^l\| \leq c_0\) occurs, and then Algorithm 5.1 exists normally. We leave this question for a research topic. To close this section, we show that the rank-one projection of its normal output is an approximately feasible solution of (1.1), and provide an upper estimation of its objective value to the optimal one.

**Theorem SM3.5.** Let $v^*$ denote the optimal value of (1.1) and $X^{l_f}$ be the normal output of Algorithm 5.1. For each $l \geq 0$, let $\{(X^{l,k}, W^{l,k})\}$ be the sequence generated by Algorithm 2 with $X^{l,0} = X^l$ and $\beta_k \equiv 0$. If there exists $l^* \in \{0, 1, \ldots, l_f\}$ such that $f(X^{l^*}) \leq v^*$, then the following inequalities hold with $r^* = \text{rank}(X^{l^*})$:

(SM3.9a) $f(x^{l_f}(x^{l_f})^T) - v^* \leq \rho_l \|X^{l_f}\| - \rho_l \cdot p/r^* + \sum_{j=1}^{l_f-1} (\rho_j - \rho_{j+1}) \|X^{j+1}\| + \alpha_f \epsilon,$

(SM3.9b) $\|x^{l_f} \circ x^{l_f} - e\| \leq \epsilon$ with $x^{l_f} = \|X^{l_f}\|^{1/2} P_1$ for $P \in \Omega(X^{l_f})$.

**Proof.** Fix any $l \in \{0, 1, \ldots, l_f\}$. For each $k \geq 0$, from $\beta_k \equiv 0$ and (SM3.2),

$$
\langle \nabla f(X^{l,k}) + \rho_l (I + W^{l,k}), X^{l,k+1} - X^{l,k} \rangle + L_k \|X^{l,k+1} - X^{l,k}\|_F^2 \leq 0.
$$

Since $f(X^{l,k+1}) \leq f(X^{l,k}) + \langle \nabla f(X^{l,k}), X^{l,k+1} - X^{l,k} \rangle + \frac{L_k}{2} \|X^{l,k+1} - X^{l,k}\|_F^2$ by using (SM3.1a) with $X = X^{l,k+1}, Y = X^{l,k}$, from the last inequality it follows that

$$
f(X^{l,k+1}) - f(X^{l,k}) + \rho_l (\langle I + W^{l,k}, X^{l,k+1} - X^{l,k} \rangle \leq \frac{L_k}{2} - \frac{2L_k}{2} \|X^{l,k+1} - X^{l,k}\|_F^2.
$$

Notice that $-\|X^{l,k+1}\| + \|X^{l,k}\| \leq \langle W^{l,k}, X^{l,k+1} - X^{l,k} \rangle$ and $\langle I, X^{l,k+1} \rangle = \langle I, X^{l,k} \rangle = p$ by $X^{l,k+1}, X^{l,k} \in \Omega$, and $L_k \geq L_f$. Then, $f(X^{l,k+1}) - \rho_l \|X^{l,k+1}\| \leq f(X^{l,k}) - \rho_l \|X^{l,k}\|$. From this recursion formula, it immediately follows that

$$
f(X^{l,k+1}) + \rho_l \|X^{l,k+1}\| \leq \cdots \leq f(X^{l,0}) - \rho_l \|X^{l,0}\| = f(X^l) - \rho_l \|X^l\|.
$$

By Theorem SM3.3, the sequence $\{X^{l,k}\}$ is convergent as $k \to \infty$. Let $X^{l,*}$ denote its limit. Then $X^{l,k+1} = X^{l,*}$. From the last inequality, for each $l \in \{0, 1, \ldots, l_f\}$,

(SM3.10) $f(X^{l+1}) - \rho_l \|X^{l+1}\| \leq f(X^l) - \rho_l \|X^l\|.$
Notice that $X^l_j = \sum_{i=1}^p \lambda_i(X^l_j)P_iP_i^T \in \Omega$ where $P_i$ denotes the $i$th column of $P$. By the Lipschitz continuity of $f$ relative to $\Omega$ with modulus $\alpha_f$, it follows that

$$f(X^l_j) = f(\sum_{i=1}^p \lambda_i(X^l_j)P_iP_i^T) = f(\lambda_1(X^l_j)P_1P_1^T + \sum_{i=2}^p \lambda_i(X^l_j)P_iP_i^T) \geq f(x^j_l(x^l_j)^T) - \alpha_f \| \sum_{i=2}^p \lambda_i(X^l_j)P_iP_i^T \|_F \geq f(x^j_l(x^l_j)^T) - \alpha_f \epsilon_f.$$  

(SM3.11)

In addition, adding $(\rho_l - \rho_{l+1})\|X^{l+1}\|$ to the both sides of (SM3.10) yields that

$$f(X^{l+1}) - \rho_{l+1}\|X^{l+1}\| \leq f(X^l) - \rho_l\|X^l\| + (\rho_l - \rho_{l+1})\|X^{l+1}\|.$$  

Thus, $f(X^l) - \rho_j\|X^l\| \leq f(X^{l'}) - \rho_{l'}\|X^{l'}\| + \sum_{j=l}^{l'-1} (\rho_j - \rho_{j+1})\|X^{j+1}\|$. Combining this inequality with (SM3.11) and noting that $\|X^{l'}\| \geq p/r^*$ yields (SM3.9a). Since $\text{diag}(X^l) = e$, we have $\|x^j_l \circ x^j_l - e\| = \| \sum_{i=2}^p \lambda_i(X^l_j)P_i \circ P_i \| \leq \epsilon_f$. \hfill \Box

By Theorem SM3.5, when Algorithm 5.1 exits normally at the $l_f$th step and $f(X^{l'}) \leq v^*$ for some $l' \in \{0, 1, \ldots, l_f\}$, the rank-one projection $x^l_f$ of $X^l_f$ delivers an approximate feasible solution of problem (1.1), and the difference between its objective value and the optimal value of (1.1) is upper bounded by the right hand side of (SM3.9a), which becomes less if $l^*$ is closer to $l_f - 1$ or the rank of $X^{l_f}$ is close to 1. Clearly, if there exists $l^* \in \{0, 1, \ldots, l_f\}$ such that $f(X^{l^*})$ is close to the optimal value of (1.4) without the DC constraint, it is more likely for $f(X^{l^*}) \leq v^*$ to hold.

REFERENCES

[SM1] H. Attouch, J. Bolte, P. Redont, and A. Soubeyran, Proximal alternating minimization and projection methods for nonconvex problems: an approach based on the Kurdyka-
Lojasiewicz inequality, Mathematics of Operations Research, 35 (2010), pp. 438–457.

[SM2] A. S. Lewis, Nonsmooth analysis of eigenvalues, Mathematical Programming, 84 (1999), pp. 1–24.

[SM3] T. X. Liu, T. K. Pong, and A. Takeda, A refined convergence analysis of pdca, with applications to simultaneous sparse recovery and outlier detection, Computational Optimization and Applications, 73 (2019), pp. 69–100.

[SM4] R. T. Rockafellar, Convex Analysis, Princeton University Press, 1970.

[SM5] R. T. Rockafellar and R. J.-B. Wets, Variational Analysis, Springer, 1998.