An inexact alternating projection method with application to matrix completion

Stefania Bellavia[1,4*], Simone Rebegoldi[2,4] and Mattia Silei[3,4]

[1*]Dipartimento di Ingegneria Industriale (DIEF), Università degli Studi di Firenze, Viale G.B. Morgagni 40, Firenze, 50134, Italia.
[2]Dipartimento di Scienze Fisiche, Informatiche e Matematiche, Università degli Studi di Modena e Reggio Emilia, Via Campi 213/B, Modena, 41125, Italia.
[3]Dipartimento di Matematica e Informatica (DIMAI), Università degli Studi di Firenze, viale G.B. Morgagni 65, Firenze, 50134, Italia.
[4]INdAM Research Group GNCS.

*Corresponding author(s). E-mail(s): stefania.bellavia@unifi.it;
Contributing authors: simone.rebegoldi@unimore.it; mattia.silei@unifi.it;

Abstract

We develop and analyze an inexact regularized alternating projection method for nonconvex feasibility problems. Such a method employs inexact projections on one of the two sets, according to a set of well-defined conditions. We prove the global convergence of the algorithm, provided that a certain merit function satisfies the Kurdyka-Łojasiewicz property on its domain. The method is then specialized to the class of affine rank minimization problems, which includes matrix completion as a special case. We approximate the truncated Singular Value Decomposition of the matrix that has to be projected by means of a Krylov solver, and provide suitable stopping criteria for the Krylov method complying with the theoretical inexactness conditions. The information needed to implement such stopping criteria do not require an extra computational effort as they are a by-product of the Krylov method itself and avoid the so called oversolving phenomena. Results of the numerical validation of the algorithm on matrix completion problems are presented.

Keywords: affine rank minimization, alternating projections, inexact projections, krylov subspaces, matrix completion

1 Introduction

This paper is concerned with regularized alternating projection methods for computing a point in the intersection of two nonconvex sets A and B, namely

$$\text{find } x \text{ such that } x \in A \cap B \qquad A, B \subset \mathbb{R}^n. \tag{1.1}$$

The alternating projection method (APM) dates back to the seminal paper [1] by J. von Neumann in 1950. Starting from a point in the first set, APM generates a sequence of points by alternately projecting onto the two sets. The method converges globally when A, B are convex [2], and a rich literature has been devoted to the convergence analysis of APM in the convex case [2–6]. When one of the sets is nonconvex, APM converges only locally and further hypotheses on the sets and their intersection are required [7–11].

1.1 Contribution of the paper

We focus on a regularized variant of APM (RAPM) proposed in [12], for which global convergence holds even in the case of nonconvex sets. RAPM requires exact projections on both sets; hence, it is not applicable when the projection on one of the sets is exact, while it is not possible or too expensive to compute the exact projection onto the other set. We design a theoretically well-founded inexact variant of RAPM, denoted by iRAPM, that allows to use inexact projections on the latter set. Inspired by [13–15], we introduce suitable conditions that the inexact projection has to satisfy in order to preserve the theoretical properties of the exact counterpart RAPM.

Our main motivation for this study lies in the application of alternating projection methods to Matrix Completion (MC) problems and, more generally, affine rank minimization problems, which consist in finding a matrix $Z \in \mathbb{R}^{n_1 \times n_2}$ of minimum rank that satisfies a given system of linear equations:

$$\begin{aligned} &\min \quad \operatorname{rank}(Z), \\ &\text{s.t. } \mathcal{A}(Z) = b, \end{aligned} \tag{1.2}$$

where $\mathcal{A} : \mathbb{R}^{n_1 \times n_2} \to \mathbb{R}^q$ is a linear map and $b \in \mathbb{R}^q$. Problem (1.2) encompasses the MC problem, system identification and control problems, Euclidean embedding, and collaborative filtering [16].

Assuming to know the sought rank of the matrix Z, say r, we reformulate the affine rank minimization problem as the problem of finding a matrix in the intersection of two sets: the set $\mathcal{C}$ of matrices satisfying the linear constraints and the *rank level set*, that is, the nonconvex set $\mathcal{C}_r$ of matrices of rank less than or equal to r. Then, we specialize APM, RAPM and iRAPM to this latter class of problems. First, we show that the two sets $\mathcal{C}$ and $\mathcal{C}_r$ satisfy the assumptions that guarantee the convergence of the three methods, thus providing a theoretical support to the employment of APMs for solving affine rank minimization problems. Second, we focus on the computation of the projection onto the rank level set, namely the truncated Singular Value Decomposition (SVD) of the matrix that has to be projected. This is carried out by an iterative method, with an inherent truncation error. In the large scale setting, this is an expensive task and iterative Krylov methods are used [17] in order to avoid full matrix factorizations. Suitable stopping criteria for this inner iterative process are therefore of crucial importance in order to save Krylov inner iterations, while preserving the same convergence properties as RAPM. We propose novel stopping criteria for the Krylov method that comply with the theoretical conditions imposed in iRAPM. The information needed to implement such stopping criteria does not require extra computational effort as it is a by-product of the Krylov method itself.

Some numerical experiments on matrix completion problems to validate the convergence results, the proposed stopping criteria and the performance of the studied methods will be also shown.

1.2 Related works

Early general results on inexact APM schemes in [18] emerge naturally from the convergence theory of the exact algorithm, but their application to common feasibility problems is not investigated. In the works [19, 20], inexact APMs are proposed for problems where, similarly to our setting, the projection onto one of the two sets can be computed exactly while the other projection is inexact. On the one hand, in [19], the authors propose an APM where the projection onto one of the sets is approximately computed by means of the conditional gradient method. On the other hand, in [20], inexact projections are considered onto two possible sets; the set of the solution of mixed equations and inequalities and a certain smooth manifold given by the image under a nonlinear map of an open set containing the iterates. Inexact projections are also considered in [21] where the author proposes a theoretical inexactness criterion based on normal cones. We also highlight that both [20] and [21] consider easy to project onto the rank level set since the projection of a matrix Z is given by its truncated SVD. Our point of view is different, as the computation of the SVD is expensive and inherently inexact, and differently from [21], we provide inexactness conditions and corresponding stopping criteria for the inner Krylov solver that can be implemented without any additional cost, allowing for the reduction of computational times. Such criteria are inspired by the works [15, 22, 23], where implementable inexactness conditions for computing proximal maps are considered in the context of inexact forward–backward algorithms. Unlike these previous works, we cannot rely on the convexity of the sets involved in problem (1.1); for that reason, our definition of inexact projection will be slightly more restrictive.

The APM is widely applied to the Matrix Completion problem under different names. For example, the Expectation-Maximization method is used in statistics for parameters estimation with missing data,

and, when data have a matrix structure, some of its variants are equivalent to APM [24]. In the context of denoising and recovering missing data in time series, the Cadzow's algorithm is well studied [25]. It consists in wrapping the time series in a Hankel matrix, applying an SVD truncation to denoise the data, and imputing the missing values. In its basic form, it consists in alternating the same two steps of APM. Note that the convergence results for RAPM when applied to the rank affine minimization problem provide also a theoretical support to regularized variants of the above approaches.

1.3 Outline

The paper is organised as follows. After an introduction to APM and RAPM given in Section 2, the iRAPM is proposed and its convergence analyzed in Section 3. Then, the method is specialized to the affine rank minimization problem in Section 4, where the new stopping criteria for the Krylov method used to compute the truncated SVD are presented, see Sections 4.2 and 4.3. In Section 5, numerical results on matrix completion problems are provided showing that the inexact approach is more efficient than the exact counterpart.

2 Preliminaries

In this section, we first introduce the notations and basic definitions used in the present work. Then, we recall the APM and RAPM algorithms and state their main convergence results in the case when one of the two sets is nonconvex.

2.1 Notations and basic definitions

The symbol $\|\cdot\|$ denotes either the standard Euclidean norm on $\mathbb{R}^n$ or the Frobenius norm on $\mathbb{R}^{n_1\times n_2}$, depending on the context. Analogously, $\langle\cdot,\cdot\rangle$ denotes either the Euclidean or the Frobenius inner product. Given two vectors $x, y \in \mathbb{R}^n$, we will use the notation (x, y) as a shorthand for $(x^T, y^T)^T$. Given a non-empty, closed subset $A \subseteq \mathbb{R}^n$, the indicator function $\iota_A : \mathbb{R}^n \to \bar{\mathbb{R}}$ is given by

$$\iota_A(x) = \begin{cases} 0, & \text{if } x \in A \\ +\infty, & \text{otherwise,} \end{cases}$$

where $\bar{\mathbb{R}} = \mathbb{R} \cup \{+\infty, \infty\}$ is the extended real numbers set.

Given a function $f : \mathbb{R}^n \to \bar{\mathbb{R}}$ and a set $A \subseteq \mathbb{R}^n$, $\underset{x\in A}{\operatorname{argmin}}\ f(x)$ denotes the set of global minimum points of f over the set A. The projection operator $\mathcal{P}_A : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ onto the set A is defined by

$$\mathcal{P}_A(x) = \underset{y\in A}{\operatorname{argmin}}\ \frac{1}{2}\|y - x\|^2 = \underset{y\in\mathbb{R}^n}{\operatorname{argmin}}\ \frac{1}{2}\|y - x\|^2 + \iota_A(y).$$

We say that $A \subseteq \mathbb{R}^n$ is *prox-regular* at a point $\bar{x} \in A$ if $\mathcal{P}_A$ is single-valued around $\bar{x}$ [21, Definition 2.6]. A set $A \subseteq \mathbb{R}^n$ is called *semialgebraic* if it assumes the form [26]

$$A = \bigcup_{i=1}^{p} \bigcap_{j=1}^{q} \{x \in \mathbb{R}^n :\ f_{i,j}(x) = 0,\ g_{i,j}(x) < 0\}, \tag{2.1}$$

where $f_{i,j}, g_{i,j} : \mathbb{R}^n \to \mathbb{R}$ are polynomial functions for all $1 \le i \le p$, $1 \le j \le q$. Likewise, a function $f : \mathbb{R}^n \to \bar{\mathbb{R}}$ is called semialgebraic if its graph is a semialgebraic subset of $\mathbb{R}^n \times \mathbb{R}$.

Given a proper, lower semicontinuous function $f : \mathbb{R}^n \to \bar{\mathbb{R}}$, the proximal operator of f is denoted by $\operatorname{prox}_f : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ and defined as

$$\operatorname{prox}_f(x) = \underset{y\in\mathbb{R}^n}{\operatorname{argmin}}\ \frac{1}{2}\|y - x\|^2 + f(y).$$

Note that, for any non-empty, closed subset $A \subseteq \mathbb{R}^n$, we have $\mathcal{P}_A(x) = \operatorname{prox}_{\iota_A}(x)$.

Let $f : \mathbb{R}^n \to \bar{\mathbb{R}}$ and $\bar{x} \in \mathrm{dom}(f)$. The *Fréchet subdifferential* and the *limiting subdifferential* of f at $\bar{x}$ are defined, respectively, as the sets [27, Definition 8.3]

$$\hat{\partial} f(\bar{x}) = \left\{ w \in \mathbb{R}^n : \liminf_{x \to \bar{x}, x \neq \bar{x}} \frac{f(x) - f(\bar{x}) - (x - \bar{x})^T w}{\|x - \bar{x}\|} \geq 0 \right\},$$
$$\partial f(\bar{x}) = \{ w \in \mathbb{R}^n : \exists\, x^k \to \bar{x},\ f(x^k) \to f(\bar{x}),\ w^k \in \hat{\partial} f(x^k) \to w \text{ as } k \to \infty \}.$$

A point $\bar{x} \in \mathbb{R}^n$ is said to be *stationary* for $f : \mathbb{R}^n \to \bar{\mathbb{R}}$ if $0 \in \partial f(\bar{x})$. We denote with $\|\partial f(x)\|_- = \inf_{v \in \partial f(x)} \|v\|$ the lazy slope of f at x.

Given a matrix $G \in \mathbb{R}^{n_1 \times n_2}$, we denote its Singular Value Decomposition (SVD) as

$$G = U \Sigma V^T,$$

where $U \in \mathbb{R}^{n_1 \times n_1}$, $V \in \mathbb{R}^{n_2 \times n_2}$ are orthonormal matrices and $\Sigma \in \mathbb{R}^{n_1 \times n_2}$ is the rectangular diagonal matrix whose diagonal entries $\sigma_1, \ldots, \sigma_{\min(n_1, n_2)}$ are the singular values of G. We will assume that the singular values are in decreasing order, that is, $\sigma_i \geq \sigma_{i+1},\ i = 1, \ldots, \min(n_1, n_2) - 1$. Given $1 \leq r \leq \min\{n_1, n_2\}$, we also consider the following partitions of matrices U, V, Σ: $U = [U_1\ U_2]$, $V = [V_1\ V_2]$, and $\Sigma = \mathrm{diag}(\Sigma_1, \Sigma_2)$, with $U_1 \in \mathbb{R}^{n_1 \times r}, U_2 \in \mathbb{R}^{n_1 \times (n_1 - r)}, V_1 \in \mathbb{R}^{n_2 \times r}, V_2 \in \mathbb{R}^{n_2 \times (n_2 - r)}, \Sigma_1 \in \mathbb{R}^{r \times r}, \Sigma_2 \in \mathbb{R}^{(n_1 - r) \times (n_2 - r)}$. With these partitions we can write

$$G = U_1 \Sigma_1 V_1^T + U_2 \Sigma_2 V_2^T = [G]_1 + [G]_2. \tag{2.2}$$

The matrix $[G]_1 = U_1 \Sigma_1 V_1^T$ is the truncated SVD to the first r singular values, and in the following we will refer to it as r-truncated SVD.

2.2 Alternating projection methods for nonconvex feasibility problems

Given two non empty, closed subsets $A \subseteq \mathbb{R}^n$, $B \subseteq \mathbb{R}^n$, we consider the feasibility problem (1.1) and the related minimization problem

$$\operatorname*{argmin}_{(x,y) \in \mathbb{R}^n \times \mathbb{R}^n} L(x, y) \equiv \frac{1}{2}\|x - y\|^2 + \iota_A(x) + \iota_B(y). \tag{2.3}$$

The classical *alternating projection method* (APM) for solving (1.1) was first presented by J. von Neumann [1]. Such a method can be seen as the standard alternating minimization (or Gauss-Seidel) method applied to problem (2.3), where the function is cyclically minimized with respect to each block of variables, while keeping the other one fixed. Then, given $y^0 \in B$, it generates two sequences $\{x^k\}_{k \in \mathbb{N}} \subseteq A$ and $\{y^k\}_{k \in \mathbb{N}} \subseteq B$ defined as follows:

$$x^{k+1} \in \operatorname*{argmin}_{x \in \mathbb{R}^n} L(x, y^k) = \operatorname*{argmin}_{x \in \mathbb{R}^n} \frac{1}{2}\|x - y^k\|^2 + \iota_A(x) = \mathcal{P}_A(y^k) \tag{2.4}$$

$$y^{k+1} \in \operatorname*{argmin}_{y \in \mathbb{R}^n} L(x^{k+1}, y) = \operatorname*{argmin}_{y \in \mathbb{R}^n} \frac{1}{2}\|x^{k+1} - y\|^2 + \iota_B(y) = \mathcal{P}_B(x^{k+1}) \tag{2.5}$$

Although global convergence of method (2.4)-(2.5) to a solution of problem (1.1) holds whenever A and B are convex, the same property can not be guaranteed in the case where either one of or both sets are nonconvex [11, 28]. On the other hand, the local convergence properties of APM in the nonconvex setting have been investigated under additional assumptions on the sets of interest. In [9], the authors prove the local linear convergence rate of APM whenever two smooth manifolds intersect transversally. This result is further generalized in [10] for any two sets intersecting transversally, provided that at least one of them is superregular at an intersecting point. In [7] (see also the subsequent work [8]), the authors remove the superregularity assumption and prove the local linear convergence of APM by only requiring transversality (or a slightly weaker version of it). We note that, in practice, checking whether transversality holds requires the knowledge of a point in the intersection, even for semialgebraic sets [8, p. 1648]. In [11], the authors prove the local convergence of APM at a sublinear rate, by assuming that the two sets intersect separably and at least one of them is $\sigma-$*Hölder regular*, with $\sigma \in [0, 1)$. Although the convergence rate in [11] is weaker than those obtained in [7–10], the separable intersection condition

automatically holds when the sets are semialgebraic [11, Theorem 3], which is the case of interest for affine rank minimization problems (see Section 4 for an in-depth analysis).

The definitions of separable intersection and Hölder regularity, as well as the local convergence result derived in [11], are reported below.

Definition 2.1 [11, Definition 1] Two sets A and B intersect separably at $\bar{x} \in A \cap B$ with exponent $\omega \in [0, 2)$ and constant $\gamma > 0$ if there exists a neighbourhood $\mathcal{U}$ of $\bar{x}$ such that, for any points $a \in A$, $b \in \mathcal{P}_B(a)$, $a^+ \in \mathcal{P}_A(b)$, $b^+ \in \mathcal{P}_B(a^+)$ belonging to $\mathcal{U}$, the following condition holds

$$\langle b - a^+, b^+ - a^+ \rangle \leq (1 - \gamma \|b^+ - a^+\|^\omega) \|b - a^+\| \|b^+ - a^+\|.$$

Definition 2.2 [11, Definition 2] Let $\sigma \in [0, 1)$. The set B is *$\sigma-$Hölder regular* with respect to A at $b^* \in A \cap B$ if there exists a neighborhood $\mathcal{U}$ of b^* and a constant $c > 0$ such that for every $a^+ \in A \cap \mathcal{U}$ and $b \in \mathcal{P}_B(a^+) \cap \mathcal{U}$, setting $r = \|a^+ - b^+\|$, one has

$$\mathcal{B}(a^+, (1 + c)r) \cap \{b \in \mathcal{P}_A(a^+)^{-1} : \ \langle a^+ - b^+, b - b^+ \rangle > \sqrt{c} r^{\sigma+1} \|b - b^+\|\} \cap B = \emptyset,$$

where $\mathcal{B}(a^+, (1 + c)r)$ denotes the ball of center a^+ and radius $(1 + c)r$.

Theorem 2.1 *[11, Theorem 1, Corollary 4] (Local convergence) Suppose B intersects A separably at $\bar{x} \in A \cap B$ with exponent $\omega \in (0, 2)$ and constant $\gamma > 0$. Assume also that B is $\omega/2-$Hölder regular at $\bar{x}$ with respect to A and constant $c < \frac{\gamma}{2}$. Then there exists a neighborhood $\mathcal{V}$ of $\bar{x}$ such that any sequences $\{x^k\}_{k\in\mathbb{N}}$ and $\{y^k\}_{k\in\mathbb{N}}$ generated by (2.4)-(2.5) with $y^0 \in \mathcal{V}$ converges to a point $\bar{b} \in A \cap B$. Furthermore, $\|x^k - \bar{b}\| = \mathcal{O}(k^{-\frac{2-\omega}{2\omega}})$ and $\|y^k - \bar{b}\| = \mathcal{O}(k^{-\frac{2-\omega}{2\omega}})$.*

In [12, 29], the authors investigate on the convergence of a regularized modification of APM, obtained by proximal modification of the two alternating steps. Given two strictly positive sequences $\{\lambda_k\}_{k\in\mathbb{N}}$ and $\{\mu_k\}_{k\in\mathbb{N}}$, the regularized APM generates two sequences $\{x^k\}_{k\in\mathbb{N}} \subseteq A$ and $\{y^k\}_{k\in\mathbb{N}} \subseteq B$ such that:

$$\begin{aligned} x^{k+1} &\in \operatorname{prox}_{\lambda_k L(\cdot, y^k)}(x^k) = \operatorname*{argmin}_{x\in\mathbb{R}^n} \frac{1}{2}\|x - y^k\|^2 + \frac{1}{2\lambda_k}\|x - x^k\|^2 + \iota_A(x), \\ y^{k+1} &\in \operatorname{prox}_{\mu_k L(x^{k+1}, \cdot)}(y^k) = \operatorname*{argmin}_{y\in\mathbb{R}^n} \frac{1}{2}\|y - x^{k+1}\|^2 + \frac{1}{2\mu_k}\|y - y^k\|^2 + \iota_B(y). \end{aligned}$$

By means of some easy algebraic manipulations, we can interpret the points x^{k+1} and y^{k+1} as projections onto A and B, respectively, of weighted means of the current iterates, i.e.,

$$x^{k+1} \in \mathcal{P}_A\left(\frac{1}{\lambda_k + 1}x^k + \frac{\lambda_k}{\lambda_k + 1}y^k\right), \quad y^{k+1} \in \mathcal{P}_B\left(\frac{1}{\mu_k + 1}y^k + \frac{\mu_k}{\mu_k + 1}x^{k+1}\right).$$

The regularized APM (RAPM) is sketched in Algorithm 1.

Algorithm 1 Regularized Alternating Projection Method (RAPM)[12, 29]

Choose $(x^0, y^0) \in A \times B$, $0 < \lambda_- < \lambda_+$, $0 < \mu_- < \mu_+$, $\{\lambda_k\}_{k\in\mathbb{N}} \subseteq [\lambda_-, \lambda_+]$, $\{\mu_k\}_{k\in\mathbb{N}} \subseteq [\mu_-, \mu_+]$.
For $k = 0, 1, \ldots$

Step 1 Compute $x^{k+1} \in \mathcal{P}_A\left(\frac{1}{\lambda_k+1}x^k + \frac{\lambda_k}{\lambda_k+1}y^k\right)$.
Step 2 Compute $y^{k+1} \in \mathcal{P}_B\left(\frac{1}{\mu_k+1}y^k + \frac{\mu_k}{\mu_k+1}x^{k+1}\right)$.

In [12], the authors carry out a convergence analysis for RAPM in the nonconvex setting, by just assuming that the function L satisfies the so-called Kurdyka-Łojasiewicz inequality, which is reported below for the reader's convenience.

Definition 2.3 Let $F : \mathbb{R}^n \longrightarrow \bar{\mathbb{R}}$ be a proper, lower semicontinuous function. The function F is said to have the Kurdyka-Łojasiewicz (KL) property at $\bar{x} \in \text{dom}(\partial F)$ if there exist $\upsilon \in (0, +\infty]$, a neighborhood $\mathcal{U}$ of $\bar{x}$ and a continuous concave function $\phi : [0, \upsilon) \longrightarrow [0, +\infty)$ with $\phi(0) = 0$, $\phi \in C^1(0, \upsilon)$, $\phi'(s) > 0$ for all $s \in (0, \upsilon)$, such that

$$\phi'(F(z) - F(\bar{x}))\|\partial F(z)\|_- \geq 1,$$

for all $z \in \mathcal{U} \cap \{z \in \mathbb{R}^n : \ F(\bar{x}) < F(z) < F(\bar{x}) + \upsilon\}$. If F satisfies the KL property at each point of $\text{dom}(\partial F)$, then F is called a KL function.

The convergence results proved in [12] are reported in Theorem 2.2 for the reader's convenience. Unlike APM, the RAPM algorithm shows both local and global convergence, and its convergence rate depends on the geometrical shape of the desingularizing function ϕ in Definition 2.3.

Theorem 2.2 *[12, Corollary 3.1, Theorem 3.4] Let $\{x^k\}_{k\in\mathbb{N}}$ and $\{y^k\}_{k\in\mathbb{N}}$ be the sequences generated by Algorithm RAPM.*

(i) (Global convergence) If the function L defined in (2.3) is a KL function, then either $\|(x^k, y^k)\| \to \infty$, or the sequence $\{(x^k, y^k)\}_{k\in\mathbb{N}}$ converges to a stationary point of L.

(ii) (Local convergence) Suppose that the function L defined in (2.3) has the KL property at a point $(\bar{x}, \bar{y})$ and that $\|\bar{x}-\bar{y}\| = \min\{\|x-y\| : \ x \in A, \ y \in B\}$. If (x^0, y^0) is sufficiently close to $(\bar{x}, \bar{y})$, then the whole sequence $\{(x^k, y^k)\}_{k\in\mathbb{N}}$ converges to a point (x_∞, y_∞) such that $\|x_\infty - y_\infty\| = \min\{\|x - y\| : \ x \in A, \ y \in B\}$.

(iii) (Rate of convergence) Suppose that $\{(x^k, y^k)\}_{k\in\mathbb{N}}$ converges to a point $(x^*, y^*) \in A\times B$ and that the function L defined in (2.3) has the KL property at (x^*, y^*) with $\phi(t) = ct^{1-\theta}$, being $c > 0$ and $\theta \in [0, 1)$. Then, if $\theta = 0$, the sequence $\{(x^k, y^k)\}_{k\in\mathbb{N}}$ terminates in a finite number of iterations; otherwise, we have

$$\|(x^k, y^k) - (x^*, y^*)\| = \begin{cases} \mathcal{O}(\tau^k), \ \text{for some } \tau \in [0, 1), & \text{if } \theta \in (0, \frac{1}{2}], \\ \mathcal{O}\left(k^{-(1-\theta)/(2\theta-1)}\right), & \text{if } \theta \in (\frac{1}{2}, 1). \end{cases}$$

3 An inexact regularized alternating projection method

In this section, we propose an inexact version of Algorithm RAPM, which is designed for solving the feasibility problem (1.1) whenever the projection onto B is costly or does not admit a closed-form implementation. After presenting the method in full detail, we discuss its practical implementation and analyze its convergence in a nonconvex scenario. Regarding the latter aspect, we prove the convergence of the iterates to a stationary limit point of the function L and investigate on the convergence rates of the distance to the limit, by assuming that a specific merit function related to L satisfies the KL property on its domain.

3.1 The proposed method

We call our proposed method iRAPM - inexact Regularized Alternating Projection Method. iRAPM alternates between an exact (regularized) projection onto A and an inexact (regularized) projection onto B, generating two sequences $\{x^k\}_{k\in\mathbb{N}} \subseteq A$ and $\{y^k\}_{k\in\mathbb{N}} \subseteq B$. At each iteration $k \in \mathbb{N}$, given x^k, y^k and two positive parameters λ_k, μ_k, the next iterate $x^{k+1} \in A$ is computed as an (exact) projected point of the convex combination $(x^k + \lambda_k y^k)/(\lambda_k + 1)$ onto A, i.e.,

$$x^{k+1} \in \mathcal{P}_A\left(\frac{1}{\lambda_k + 1}x^k + \frac{\lambda_k}{\lambda_k + 1}y^k\right),$$

whereas the next iterate $y^{k+1} \in B$ represents an inexact projection of the point $(y^k + \mu_k x^{k+1})/(\mu_k + 1)$ onto B. In order to introduce the inexactness criteria, we let Q^{k+1} be the function defined as

$$Q^{k+1}(y) := \frac{1}{2\mu_k}\|y - y^k\|^2 + L(x^{k+1}, y) - L(x^{k+1}, y^k). \tag{3.1}$$

Let us denote with y^k_{reg} the point $(y^k + \mu_k x^{k+1})/(\mu_k + 1)$. Then, we can conveniently write Q^{k+1} in (3.1) as

$$Q^{k+1}(y) = \frac{1}{2\mu_k}\|y - y^k\|^2 + \frac{1}{2}\|x^{k+1} - y\|^2 - \frac{1}{2}\|x^{k+1} - y^k\|^2 + \iota_B(y)$$

$$
\begin{aligned}
&= \frac{(1+\mu_k)}{2\mu_k}\|y-y^k\|^2 - \frac{1}{2}\|y-y^k\|^2 + \frac{1}{2}\|x^{k+1}-y\|^2 - \frac{1}{2}\|x^{k+1}-y^k\|^2 + \iota_B(y)\\
&= \frac{(1+\mu_k)}{2\mu_k}\|y-y^k\|^2 + \langle x^{k+1}-y^k, y^k-y\rangle + \iota_B(y)\\
&= \frac{(1+\mu_k)}{2\mu_k}\|y-y^k\|^2 + \frac{1+\mu_k}{\mu_k}\left\langle \frac{\mu_k}{1+\mu_k}(y^k-x^{k+1}), y-y^k\right\rangle + \iota_B(y)\\
&= \frac{(1+\mu_k)}{2\mu_k}\left\|y-y^k+\frac{\mu_k}{1+\mu_k}\left(y^k-x^{k+1}\right)\right\|^2 - \frac{(1+\mu_k)}{2\mu_k}\left\|\frac{\mu_k}{1+\mu_k}(y^k-x^{k+1})\right\|^2 + \iota_B(y)\\
&= \frac{(1+\mu_k)}{2\mu_k}\left\|y-y^k_{reg}\right\|^2 - \frac{(1+\mu_k)}{2\mu_k}\left\|y^k-y^k_{reg}\right\|^2 + \iota_B(y), \quad \forall\, y\in\mathbb{R}^n, \qquad (3.2)
\end{aligned}
$$

where the first equality follows from the form of L given in (2.3), the third and fifth are both due to the application of the basic relation $\|a+b\|^2 = \|a\|^2 + \|b\|^2 + 2\langle a,b\rangle$, and the sixth follows by using the definition of y^k_{reg}. Hence the set of minimum points of Q^{k+1} is the projection set of y^k_{reg} onto B. Letting $\hat{y}^{k+1}\in\mathcal{P}_B(y^k_{reg})$, we observe that the value $Q^{k+1}(\hat{y}^{k+1})$ assumes negative or null sign, since

$$
Q^{k+1}(\hat{y}^{k+1}) = \min_{y\in\mathbb{R}^n} Q^{k+1}(y) \le Q^{k+1}(y^k) = 0. \qquad (3.3)
$$

Then, given a prefixed tolerance parameter $\zeta\in(0,1]$, iRAPM computes an approximate projection $y^{k+1}\in B$ such that the following conditions are satisfied:

$$
Q^{k+1}(y^{k+1}) \le \zeta Q^{k+1}(\hat{y}^{k+1}) \qquad (3.4)
$$

$$
\|y^{k+1}-\hat{y}^{k+1}\| \le \sqrt{-\left(\frac{1-\zeta}{\zeta}\right)Q^{k+1}(y^{k+1})}, \qquad (3.5)
$$

We report iRAPM in its entirety in Algorithm 2.

Algorithm 2 inexact Regularized Alternating Projection Method (iRAPM)

Choose $(x^0,y^0)\in A\times B$, $0<\lambda_-<\lambda_+$, $0<\mu_-<\mu_+$, $\{\lambda_k\}_{k\in\mathbb{N}}\subseteq[\lambda_-,\lambda_+]$, $\{\mu_k\}_{k\in\mathbb{N}}\subseteq[\mu_-,\mu_+]$, $\zeta\in(0,1]$.

For $k=0,1,\ldots$

Step 1 Compute $x^{k+1}\in\mathcal{P}_A\left(\frac{1}{\lambda_k+1}x^k+\frac{\lambda_k}{\lambda_k+1}y^k\right)$.
Step 2 Compute $y^{k+1}\in B$ such that (3.4)-(3.5) hold.

Some comments on the inexactness conditions (3.4)-(3.5) employed in Algorithm 2 are in order.

- Condition (3.4) is well defined due to (3.3). Furthermore, $Q^{k+1}(y^{k+1})\le 0$ and if $\zeta=1$, the point y^{k+1} belongs to the (exact) projection set of y^k_{reg} onto B; otherwise, if $\zeta\in(0,1)$, y^{k+1} represents an appropriate approximation that is "sufficiently close" to a point of the projection set. The parameter ζ controls the approximation level of the iterates y^{k+1}; the closer ζ is to 1, the closer y^{k+1} is to the projection set, whereas the closer ζ is to 0, the coarser y^{k+1} is as an inexact projection. Condition (3.4) has been frequently used for computing inexact projections [13, 14] and inexact proximal points [23] in the convex setting;
- A bound similar to the one in (3.5) is automatically satisfied if condition (3.4) holds and B is convex; indeed, in this case, we can combine the $\frac{1+\mu_k}{\mu_k}-$strong convexity of Q^{k+1} and (3.4) to obtain

$$
\|y^{k+1}-\hat{y}^{k+1}\| \le \sqrt{\frac{2\mu_k}{1+\mu_k}\left(Q^{k+1}(y^{k+1})-Q^{k+1}(\hat{y}^{k+1})\right)} \le \sqrt{-\frac{2\mu_k(1-\zeta)}{(1+\mu_k)\zeta}Q^{k+1}(y^{k+1})}. \qquad (3.6)
$$

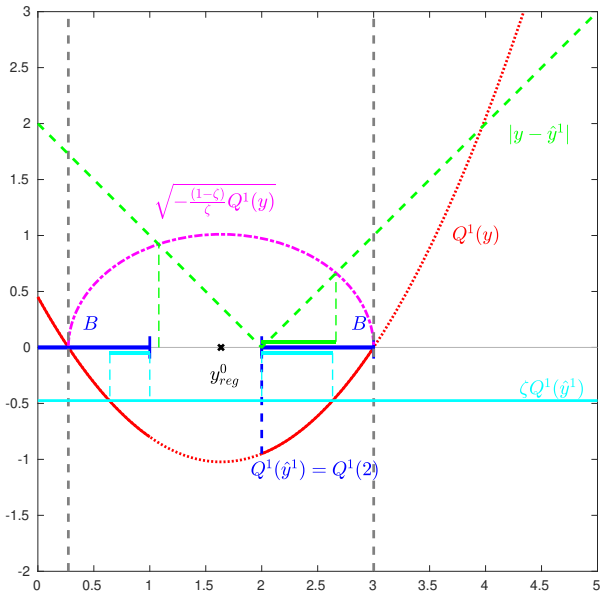

Fig. 1 One-dimensional example showing that the two inexactness conditions (3.4)-(3.5) are independent one of each other in the nonconvex case

However, if B is not convex, then Q^{k+1} is not strongly convex and condition (3.5) needs to be numerically imposed in order to guarantee the convergence of Algorithm 2.

- In Figure 1, we present an example showing that the two inexactness conditions (3.4)-(3.5) may be independent one of each other whenever B is nonconvex. Let $A = [0, 2] \subseteq \mathbb{R}$ and $B = [0, 1] \cup [2, 3] \subseteq \mathbb{R}$. Note that B is the union of two disjoint intervals, hence it is nonconvex. Given $(x^0, y^0) = (0, 3) \in A \times B$, $\lambda_0 = 1$, $\mu_0 = 10$ and $\zeta = \frac{1}{2}$, it is easy to show that $x^0_{reg} = \frac{3}{2}$, $y^0_{reg} = \frac{18}{11} \approx 1.6364$, and $\hat{y}^1 = 2$. In the figure, the red solid parabola represents the function $Q^1(y)$ for $y \in B$, whereas the red dotted parabola is its extension for $y \notin B$ obtained by neglecting $\iota_B(y)$, the cyan solid line is the right-hand term of condition (3.4), while the green dashed and magenta dash-dotted lines represent the left and right-hand sides of condition (3.5), respectively. In this case, the minimum of Q^1 is attained at a point $\hat{y}^1$ that is different from y^0_{reg}. The subintervals of B satisfying conditions (3.4)-(3.5) are represented as cyan and green segments on the horizontal axis, respectively. We see that there are no points of B to the left of y^0_{reg} that satisfy condition (3.5). This means that there exists a point $y \in B$ such that it satisfies condition (3.4) but not condition (3.5), showing that the second condition does not automatically follow from the first one in the nonconvex case. Conversely, the cyan interval on the right interval is strictly included in the green one, which implies that the first condition does not immediately follow from the second one. Furthermore, in these settings, it is possible to show that condition (3.4) is also independent of the bound (3.6), which would hold if the set B were convex, as discussed in the previous bullet.

3.2 Practical computation of the inexact projections

We now discuss how we can practically implement the regularized inexact projection appearing at Step 2 of Algorithm 2. At first glance, it looks like enforcing the inexactness conditions (3.4)-(3.5) requires the knowledge of the projection set of y^k_{reg} onto B, which we assumed we do not have access to. Nonetheless, we show how to get rid of such an information and still compute a point $y^{k+1} \in B$ satisfying (3.4)-(3.5).

3.2.1 Enforcing condition (3.4)

The following discussion is borrowed from [14, Section 2.1.1], which in turn revises a procedure that has been formerly proposed in [13]. By employing the form of Q^{k+1} given in (3.2), we can easily rewrite the

inexactness condition (3.4) as

$$\|y^{k+1} - y^k_{reg}\|^2 \leq \zeta \|\hat{y}^{k+1} - y^k_{reg}\|^2 + (1-\zeta)\|y^k - y^k_{reg}\|^2. \tag{3.7}$$

Condition (3.7) is identical to the concept of *feasible inexact projection of y^k_{reg} onto B relative to the point $y^k \in B$ and forcing parameter $\zeta \in (0,1]$ with respect to the norm $\|\cdot\|$*, which is proposed in [14, Definition 2] for convex sets. According to [14, Section 2.1.1], it is possible to define a practical procedure for computing a point y^{k+1} satisfying (3.7) if there exist two computable, iteration-dependent sequences $\{c^\ell\}_{\ell\in\mathbb{N}} \subseteq \mathbb{R}$, $\{w^\ell\}_{\ell\in\mathbb{N}} \subseteq B$ complying with the following conditions

$$c^\ell \leq \|\hat{y}^{k+1} - y^k_{reg}\|^2, \quad \forall\ \ell \in \mathbb{N} \tag{3.8a}$$

$$\lim_{\ell\to+\infty} c^\ell = \|\hat{y}^{k+1} - y^k_{reg}\|^2 \tag{3.8b}$$

$$\lim_{\ell\to+\infty} w^\ell = \hat{y}^{k+1}. \tag{3.8c}$$

More precisely, the following result holds. The proof, in the context of inexact projection onto convex sets is given in [14, Section 2.1.1]; we report the proof in the appendix for sake to completeness and to enlighten that the result is valid also if B is not convex.

Lemma 3.1 *Let $\{c^\ell\}_{\ell\in\mathbb{N}} \subseteq \mathbb{R}$, $\{w^\ell\}_{\ell\in\mathbb{N}} \subseteq B$ be two sequences complying with (3.8a)-(3.8b)-(3.8c). Then, there exists $\hat{\ell} \in \mathbb{N}$ such that*

$$\|w^\ell - y^k_{reg}\|^2 \leq \zeta c^\ell + (1-\zeta)\|y^k - y^k_{reg}\|^2, \quad \forall\ \ell \geq \hat{\ell}, \tag{3.9}$$

which implies that condition (3.4) holds with $y^{k+1} = w^\ell$ for $\ell \geq \hat{\ell}$.

Proof See Appendix A.1. □

The above procedure has been previously applied only to convex sets with a specific structure [13, 14]. In Section 4, we show how we can find the sequences $\{c^\ell\}_{\ell\in\mathbb{N}}$, $\{w^\ell\}_{\ell\in\mathbb{N}}$ for computing inexact projections over the rank level set $\mathcal{C}_r = \{Z \in \mathbb{R}^{n_1\times n_2} :\ \text{rank}(Z) \leq r\}$, which is nonconvex.

3.2.2 Enforcing condition (3.5)

We can implement condition (3.5) whenever we have at our disposal a vanishing sequence $\{a^\ell\}_{\ell\in\mathbb{N}} \subseteq \mathbb{R}$ representing an upper bound for the quantity $\|w^\ell - \hat{y}^{k+1}\|$, i.e.

$$\|w^\ell - \hat{y}^{k+1}\| \leq a^\ell, \quad \lim_{\ell\to\infty} a^\ell = 0. \tag{3.10}$$

In this case, the following result holds.

Lemma 3.2 *Let $\{w^\ell\}_{\ell\in\mathbb{N}} \subseteq B$ and $\{a^\ell\}_{\ell\in\mathbb{N}} \subseteq \mathbb{R}$ be sequences complying with (3.8c) and (3.10), respectively. Then there exists $\tilde{\ell} \in \mathbb{N}$ such that*

$$Q^{k+1}(w^\ell) < 0, \quad a^\ell \leq \sqrt{-\left(\frac{1-\zeta}{\zeta}\right) Q^{k+1}(w^\ell)} \quad \forall \ell \geq \tilde{\ell}, \tag{3.11}$$

which implies that condition (3.5) holds with $y^{k+1} = w^\ell$ for $\ell \geq \tilde{\ell}$.

Proof If $y^k \in \mathcal{P}_B(y^k_{reg})$, one can always set $w^\ell = y^k$ and $a^\ell = 0$ for any ℓ, so that conditions (3.8c), (3.10) and (3.11) are trivially satisfied for all $\ell \in \mathbb{N}$ with $\hat{y}^{k+1} = y^k$. Then, without loss of generality, we assume that $y^k \notin \mathcal{P}_B(y^k_{reg})$. Since $y^k \notin \mathcal{P}_B(y^k_{reg})$, we have $Q^{k+1}(\hat{y}^{k+1}) < 0$ by definition of Q^{k+1}. Then, since condition (3.8c) holds and Q^{k+1} is a continuous function on B, we have

$$\lim_{\ell\to\infty} Q^{k+1}(w^\ell) = Q^{k+1}(\hat{y}^{k+1}) < 0.$$

Then, for any $0 < \epsilon < -Q^{k+1}(\hat{y}^{k+1})$, there exists $\ell_\epsilon > 0$ such that

$$0 < -Q^{k+1}(\hat{y}^{k+1}) - \epsilon \leq -Q^{k+1}(w^\ell), \quad \forall\ \ell \geq \ell_\epsilon.$$

Likewise, due to the right part of (3.10), there exists $\tilde{\ell} \geq \ell_\epsilon$ sufficiently large such that

$$a^\ell \leq \sqrt{\left(\frac{1-\zeta}{\zeta}\right)(-Q^{k+1}(\hat{y}^{k+1}) - \epsilon)} \leq \sqrt{-\left(\frac{1-\zeta}{\zeta}\right) Q^{k+1}(w^{\tilde{\ell}})}, \quad \forall\, \ell > \tilde{\ell}$$

so that (3.11) is met and the inexactness condition (3.5) holds with $y^{k+1} = w^\ell$ for $\ell \geq \tilde{\ell}$. □

Once again, it is possible to compute a sequence $\{a^\ell\}_{\ell\in\mathbb{N}} \subseteq \mathbb{R}$ complying with (3.10) if B is the rank level set; see Section 4 for the details.

3.2.3 Computation of points satisfying (3.4)-(3.5)

Lemma 3.1 and Lemma 3.2 provide us with an implementable procedure for computing a point y^{k+1} satisfying simultaneously the inexactness conditions (3.4)-(3.5) employed in Algorithm 2. Indeed, given three sequences $\{c^\ell\}_{\ell\in\mathbb{N}}$, $\{w^\ell\}_{\ell\in\mathbb{N}}$, $\{a^\ell\}_{\ell\in\mathbb{N}}$ complying with (3.8a)-(3.8b)-(3.8c)-(3.10) at any given iteration $k \in \mathbb{N}$, we compute such sequences until the occurrence of the first iteration $\bar{\ell}$ satisfying

$$\begin{cases} \|w^{\bar{\ell}} - y^k_{reg}\|^2 \leq \zeta c^{\bar{\ell}} + (1-\zeta)\|y^k - y^k_{reg}\|^2 \\ a^{\bar{\ell}} \leq \sqrt{-\left(\frac{1-\zeta}{\zeta}\right) Q^{k+1}(w^{\bar{\ell}})}. \end{cases} \tag{3.12}$$

Such a procedure is well-defined thanks to Lemma 3.1 and Lemma 3.2. Then, the inexact projection complying with (3.4)-(3.5) is given by $y^{k+1} = w^{\bar{\ell}}$.

3.3 Convergence analysis

In the following, we prove the convergence of the iterates sequence generated by Algorithm 2 towards a stationary point of L, and derive some convergence rates on the gap between the iterates and their limit point. Our analysis is inspired by [12] and will be carried out by assuming that a certain merit function related to L is a KL function. The presence of an inexact projection complying to conditions (3.4)-(3.5) leads to some major changes in the convergence proof of Algorithm 2 in comparison to the one available in [12].

For our purposes, we introduce the merit function $F : \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}$ defined as

$$F(x, y, \rho) = L(x, y) + \frac{1}{2}\rho^2. \tag{3.13}$$

Our analysis relies on the following set of properties, which hold for the sequences generated by iRAPM and involve both the objective function L and the related merit function F.

Lemma 3.3 *Let $\{(x^k, y^k)\}_{k\in\mathbb{N}}$ be the sequence generated by Algorithm 2 and let $\{\hat{y}^k\}_{k\in\mathbb{N}}$ be any sequence with $\hat{y}^{k+1} \in \mathcal{P}_B(y^k_{reg})$ for all $k \in \mathbb{N}$. Furthermore, define the sequences $\{d_k\}_{k\in\mathbb{N}}$, $\{\gamma_k\}_{k\in\mathbb{N}}$ and $\{\rho_k\}_{k\in\mathbb{N}}$ such that, for all $k \in \mathbb{N}$,*

$$d_k = \sqrt{\|(x^{k+1} - x^k, y^{k+1} - y^k)\|^2 - Q^{k+1}(y^{k+1})}, \tag{3.14}$$

$$\gamma_k = -\left(\frac{1-\zeta}{\zeta}\right) Q^{k+1}(\hat{y}^{k+1}), \tag{3.15}$$

$$\rho_k = \sqrt{\frac{1}{\mu_{k-1}}\|\hat{y}^k - y^{k-1}\|^2 - 2\left(\frac{1-\zeta}{\zeta}\right) Q^k(y^k)}. \tag{3.16}$$

The following statements hold true.

(i) We have

$$\|(x^{k+1} - x^k, y^{k+1} - y^k)\| \leq d_k, \quad \sqrt{-Q^{k+1}(y^{k+1})} \leq d_k, \quad \forall\, k \in \mathbb{N}. \tag{3.17}$$

(ii) There exists $a > 0$ such that

$$L(x^{k+1}, y^{k+1}) + a d_k^2 \leq L(x^k, y^k), \quad \forall\, k \in \mathbb{N}.$$

(iii) We have

$$L(x^{k+1}, y^{k+1}) \leq F(x^{k+1}, \hat{y}^{k+1}, \rho_{k+1}) \leq L(x^k, y^k) + \gamma_k, \quad \forall\, k \in \mathbb{N}. \tag{3.18}$$

Furthermore $\lim_{k\to\infty} \rho_k = \lim_{k\to\infty} \gamma_k = 0$, and

$$\lim_{k\to\infty} \|\hat{y}^{k+1} - y^{k+1}\| = 0. \tag{3.19}$$

(iv) There exists $b > 0$ such that

$$\|\partial F(x^{k+1}, \hat{y}^{k+1}, \rho_{k+1})\|_- \leq b d_k, \quad \forall\, k \in \mathbb{N}. \tag{3.20}$$

Proof See Appendix A.2. □

Remark 3.1 Item (ii) of Lemma 3.3 shows that the function values $\{L(x^k, y^k)\}_{k\in\mathbb{N}}$ monotonically decrease along the iterations, although y^{k+1} is computed inexactly. An analogous property is also proved for the RAPM algorithm, see [12, Lemma 3.1(i)].

The bound in Item (iv) is a modification of the relative error condition used in the analysis of iterative algorithms under the KL property, which for the RAPM algorithm is written as [12, Eq. (24)]

$$\|\partial L(x^{k+1}, y^{k+1})\|_- \leq b\|(x^{k+1} - x^k, y^{k+1} - y^k)\|, \quad \forall\, k \in \mathbb{N}.$$

Differently from [12], we bound the norm of the subdifferential of the merit function F in place of L, at the point $(x^{k+1}, \hat{y}^{k+1}, \rho_{k+1})$ rather than at the actual iterate (x^{k+1}, y^{k+1}); furthermore, the norm is bounded with a term proportional to d_k in place of the norm $\|(x^{k+1} - x^k, y^{k+1} - y^k)\|$. These differences are due to the fact that the inexactness conditions (3.4)-(3.5) do not provide any useful information on the subdifferential of Q^{k+1} at the point y^{k+1}, while combining the optimality of $\hat{y}^{k+1}$ and conditions (3.4)-(3.5) we obtain (3.20). Similar bounds in the literature have been obtained for inexact proximal–gradient methods adopting analogous inexactness conditions, see e.g. [15, 23].

The bounds in item (iii) are needed to bridge the gap between the actual objective function L and the merit function F, whereas the limit in (3.19) is needed to guarantee that the KL inequality is applicable to F at the point $(x^k, \hat{y}^k, \rho_k)$ and that each limit point of $\{(x^k, y^k)\}_{k\in\mathbb{N}}$ is stationary for L (see the upcoming proofs of Lemma 3.4 and Theorem 3.6).

Let us define the set of all limit points of the sequence $\{(x^k, y^k)\}_{k\in\mathbb{N}}$, namely

$$\Gamma^*(x^0, y^0) = \{(x^*, y^*) \in A \times B : \exists\, \{k_j\}_{j\in\mathbb{N}} \subseteq \mathbb{N} \text{ s.t. } (x^{k_j}, y^{k_j}) \to (x^*, y^*)\}. \tag{3.21}$$

Analogously, $\Xi^*(x^0, y^0, \rho_0)$ denotes the set of all limit points of the sequence $\{(x^k, y^k, \rho_k)\}_{k\in\mathbb{N}}$. By Lemma 3.3(iii), we have that $\rho_k \to 0$. Then, we can write $\Xi^*(x^0, y^0, \rho_0)$ as

$$\Xi^*(x^0, y^0, \rho_0) = \Gamma^*(x^0, y^0) \times \{0\}. \tag{3.22}$$

The following lemma anticipates the main convergence result and contains some basic properties related to the set $\Xi^*(x^0, y^0, \rho_0)$.

Lemma 3.4 *Suppose that $\{(x^k, y^k)\}_{k\in\mathbb{N}}$ is bounded. The following statements hold true.*

(i) $\Xi^*(x^0, y^0, \rho_0)$ is non-empty and compact.

(ii) There exists $L^* \in [0, \infty)$ such that

$$F(x^*, y^*, 0) = L(x^*, y^*) = L^*, \quad \forall\, (x^*, y^*) \in \Gamma^*(x^0, y^0).$$

(iii) We have

$$\lim_{k\to\infty} \operatorname{dist}((x^k, y^k, \rho_k), \Xi^*(x^0, y^0, \rho_0)) = \lim_{k\to\infty} \operatorname{dist}((x^k, \hat{y}^k, \rho_k), \Xi^*(x^0, y^0, \rho_0)) = 0.$$

Proof (i) Since $\{(x^k, y^k, \rho_k)\}_{k\in\mathbb{N}}$ is bounded, the set $\Xi^*(x^0, y^0, \rho_0)$ is nonempty. As it is a countable intersection of compact sets, the set is also compact [30, Lemma 5].

(ii) By Lemma 3.3(ii) and the continuity of the function L on the set $A \times B$, it follows that there exists $L^* \in \mathbb{R}$ such that

$$\lim_{k\to\infty} L(x^k, y^k) = L(x^*, y^*) = L^*, \quad \forall\, (x^*, y^*) \in \Gamma^*(x^0, y^0). \tag{3.23}$$

Then, recalling the definition of F yields the expected result.

(iii) Since $\{(x^k, y^k)\}_{k\in\mathbb{N}}$ is bounded and by definition of $\Gamma^*(x^0, y^0)$ and $\Xi^*(x^0, y^0, \rho_0)$, we have

$$\lim_{k\to\infty} \operatorname{dist}((x^k, y^k), \Gamma^*(x^0, y^0)) = \lim_{k\to\infty} \operatorname{dist}((x^k, y^k, \rho_k), \Xi^*(x^0, y^0, \rho_0)) = 0.$$

Since (3.19) holds and $\{(x^k, \hat{y}^k)\}_{k\in\mathbb{N}}$ is bounded, we also conclude that

$$\lim_{k\to\infty} \operatorname{dist}((x^k, \hat{y}^k, \rho_k), \Xi^*(x^0, y^0, \rho_0)) = 0.$$

□

We are now ready to state the main convergence result for the sequence generated by iRAPM, under the assumption that the sequence itself is bounded and F is a KL function. The statement is analogous to the one obtained in Theorem 2.2(i) for the RAPM algorithm. Before stating and proving our result, we require the following uniformized version of the KL property.

Lemma 3.5 *[30, Lemma 6] Let $F : \mathbb{R}^n \to \bar{\mathbb{R}}$ be a proper, lower semicontinuous function and $\Xi \subset \mathbb{R}^n$ a compact set. Suppose that F satisfies the KL property at each point belonging to Ξ and that F is constant over Ξ, i.e. $F(z) = \bar{F} \in \mathbb{R}$ for all $z \in \Xi$. Then, there exist $\mu, \upsilon > 0$ and a continuous concave function $\phi : [0, \upsilon) \longrightarrow [0, +\infty)$ with $\phi(0) = 0$, $\phi \in C^1(0, \upsilon)$, $\phi'(s) > 0$ for all $s \in (0, \upsilon)$ such that*

$$\phi'(F(z) - \bar{F})\|\partial F(z)\|_- \geq 1, \quad \forall\, z \in \bar{\Xi} \tag{3.24}$$

where the set $\bar{\Xi}$ is defined as

$$\bar{\Xi} = \{z \in \mathbb{R}^n : \operatorname{dist}(z, \Xi) < \mu \text{ and } \bar{F} < F(z) < \bar{F} + \upsilon\}. \tag{3.25}$$

Theorem 3.6 *Let $\{(x^k, y^k)\}$ be generated by Algorithm 2. If the function F given in (3.13) is a KL function, then either $\|(x^k, y^k)\| \to \infty$, or $\{(x^k, y^k)\}_{k\in\mathbb{N}}$ converges to a stationary point of L.*

Proof By Lemma 3.3, it follows that Algorithm 2 is a special case of the abstract descent algorithm described in [15, Condition 3]. Then, one could apply [15, Theorem 5] to deduce the thesis. For the sake of completeness, we report a simplified convergence proof especially tailored for Algorithm 2. Assume that the sequence $\{(x^k, y^k)\}_{k\in\mathbb{N}}$ generated by Algorithm 2 is bounded. Let $\{d_k\}_{k\in\mathbb{N}}$ and $\{\rho_k\}_{k\in\mathbb{N}}$ be defined as in Lemma 3.3, and L^* as in Lemma 3.4. By Lemma 3.4(i)-(ii), the function F is constant and equal to L^* over the compact set $\Xi = \Xi^*(x^0, y^0, \rho_0)$, therefore we can apply Lemma 3.5. In what follows $\upsilon, \mu, \phi, \bar{\Xi}$ are defined in Lemma 3.5, a, b in Lemma 3.3.

Lemma 3.3(iii), Lemma 3.4(iii) and equation (3.23) guarantee the existence of a positive integer k_0 such that

$$L(x^k, y^k) + \gamma_k < L^* + \upsilon, \quad \operatorname{dist}((x^k, \hat{y}^k, \rho_k), \Xi^*(x^0, y^0, \rho_0)) < \mu, \quad \forall\, k \geq k_0. \tag{3.26}$$

Without loss of generality, up to a translation of the iteration index, we can assume $k_0 = 0$.

Since $\{L(x^k, y^k)\}_{k\in\mathbb{N}}$ is a monotone nonincreasing sequence and L^* is its limit (see Lemma 3.3(iii) and (3.23)), we have $L(x^k, y^k) \geq L^*$ for all $k \geq 0$, which implies that ϕ can be applied to both values $L(x^k, y^k) - L^*$ and $L(x^{k+1}, y^{k+1}) - L^*$. Then, one can define the following quantity

$$\phi_k = \frac{b}{a}(\phi(L(x^k, y^k) - L^*) - \phi(L(x^{k+1}, y^{(k+1)}) - L^*)), \quad \forall\, k \geq 0. \tag{3.27}$$

Since ϕ is a monotone increasing function and $\{L(x^k, y^k)\}_{k\in\mathbb{N}}$ a monotone nonincreasing sequence, we have $\phi_k \geq 0$, for all $k \geq 0$. Let us show that the following inequality holds:

$$2d_k \leq \phi_k + d_{k-1}, \quad \forall\, k \geq 1. \tag{3.28}$$

If $d_k = 0$, inequality (3.28) holds trivially. Otherwise, for any iteration index $k \geq 1$ such that $d_k > 0$ and using Lemma 3.3(ii)-(iii), we can write

$$L^* \leq L(x^{k+1}, y^{k+1}) < L(x^k, y^k) \leq F(x^k, \hat{y}^k, \rho_k) \leq L(x^{k-1}, y^{k-1}) + \gamma_{k-1},$$

which, in view of (3.26), implies that $(x^k, \hat{y}^k, \rho_k) \in \bar{\Xi}$. Then, combining the KL inequality related to the function F at $(x^k, \hat{y}^k, \rho_k)$ with Lemma 3.3(iv) yields:

$$\phi'(F(x^k, \hat{y}^k, \rho_k) - L^*) \geq \frac{1}{\|\partial F(x^k, \hat{y}^k, \rho_k)\|_-} \geq \frac{1}{bd_{k-1}}. \tag{3.29}$$

By employing the concavity of the function ϕ, we obtain

$$\begin{aligned}\phi(L(x^k, y^k) - L^*) - \phi(L(x^{k+1}, y^{k+1}) - L^*) &\geq \phi'(L(x^k, y^k) - L^*)(L(x^k, y^k) - L(x^{k+1}, y^{k+1})) \\ &\geq \frac{ad_k^2}{bd_{k-1}},\end{aligned}$$

where the latter inequality follows by applying Lemma 3.3(ii)-(iii), the monotonicity of ϕ' and (3.29). By recalling the definition of ϕ_k in (3.27), the above inequality implies $d_k^2 \leq \phi_k d_{k-1}$. Now taking the square root of both sides and using the inequality $2\sqrt{uv} \leq u + v$ on the right-hand side we get inequality (3.28). By summing (3.28) over $k = 0, \ldots, K$ and taking the limit for $K \to \infty$, following the same arguments as in [15, Theorem 5], we easily get

$$\sum_{k=0}^{\infty} d_k < \infty. \tag{3.30}$$

Finally, by combining (3.30) and (3.17), we conclude that $\sum_{k=0}^{\infty} \|(x^{k+1}-x^k, y^{k+1}-y^k)\| < \infty$. Hence the sequence $\{(x^k, y^k)\}_{k\in\mathbb{N}}$ converges to a point $(x^*, y^*) \in A \times B$.

Since $\{(x^k, y^k)\}_{k\in\mathbb{N}}$ converges to (x^*, y^*) and $\{\rho_k\}_{k\in\mathbb{N}}$ converges to 0 (due to Lemma 3.3(iii)), we have $(x^k, y^k, \rho_k) \to (x^*, y^*, 0)$; furthermore, thanks to (3.19) and the continuity of F in $A \times B$, it also follows that $(x^k, \hat{y}^k, \rho_k) \to (x^*, y^*, 0)$ and $F(x^k, \hat{y}^k, \rho_k) \to F(x^*, y^*, 0)$. Finally, by summing the inequality (3.20) for $k = 0, \ldots, K$, taking its limit for $K \to \infty$ and recalling that $\{d_k\}_{k\in\mathbb{N}}$ is summable (see (3.30)), we get

$$\sum_{k=0}^{\infty} \|\partial F(x^k, \hat{y}^k, \rho_k)\|_- < \infty \quad \Rightarrow \quad \lim_{k\to\infty} \|\partial F(x^k, \hat{y}^k, \rho_k)\|_- = 0.$$

To summarize, the sequence $\{(x^k, \hat{y}^k, \rho_k)\}_{k\in\mathbb{N}}$ satisfies the following limits:

$$\lim_{k\to\infty} (x^k, \hat{y}^k, \rho_k) = (x^*, y^*, 0), \quad \lim_{k\to\infty} F(x^k, \hat{y}^k, \rho_k) = F(x^*, y^*, 0),$$

and

$$\lim_{k\to\infty} \|\partial F(x^k, \hat{y}^k, \rho_k)\|_- = 0.$$

Then, it satisfies the hypotheses of [31, Lemma 2.1], which ensures that $0 \in \partial F(x^*, y^*, 0)$. Since $\partial F(x^*, y^*, 0) = \partial L(x^*, y^*) \times \{0\}$, it follows $0 \in \partial L(x^*, y^*)$, i.e., (x^*, y^*) is a stationary point of L. □

Inspired by [32], we now investigate the convergence rate of $\{(x^k, y^k)\}_{k\in\mathbb{N}}$ when the merit function F given in (3.13) satisfies the KL property at the limit point (x^*, y^*) with $\phi(t) = ct^{1-\theta}$, with $c > 0$, $\theta \in [0, 1[$. Similarly to the result obtained for the RAPM algorithm (see Theorem 2.2(iii)), the convergence rate of iRAPM depends on the value of θ appearing in the desingularizing function ϕ.

Theorem 3.7 *Suppose that $\{(x^k, y^k)\}_{k\in\mathbb{N}}$ converges to a point $(x^*, y^*) \in A \times B$. Additionally, assume that F defined in (3.13) satisfies the KL property at (x^*, y^*) with $\phi(t) = ct^{1-\theta}$, with $c > 0$, $\theta \in [0, 1)$. Then the following convergence rates hold.*

(i) If $\theta = 0$, the sequence (x^k, y^k) terminates in a finite number of iterations.

(ii) If $\theta \in]0, \frac{1}{2}]$, there exists $C > 0$ and $\tau \in (0, 1)$ such that

$$\|(x^k, y^k) - (x^*, y^*)\| \le C\tau^k. \tag{3.31}$$

(iii) If $\theta \in]\frac{1}{2}, 1[$, there exists $C > 0$ such that

$$\|(x^k, y^k) - (x^*, y^*)\| \le Ck^{-\frac{1-\theta}{2\theta-1}}. \tag{3.32}$$

Proof (i) If $\theta = 0$, then $\phi(t) = ct$ and $\phi'(t) = c$ for $t \in [0, \upsilon)$. By contradiction, assume that there exists an infinite subset $K \subseteq \mathbb{N}$ such that $(x^{k+1}, y^{k+1}) \neq (x^k, y^k)$ for all $k \in K$. From Lemma 3.3(i), this means that $d_k > 0$ for all $k \in K$, where d_k is given in (3.14). By using the same reasoning as in the proof of Theorem 3.6, we conclude that $(x^k, \hat{y}^k, \rho_k) \in \bar{\Xi}$ for all $k \in K$. Then, for all $k \in K$, the uniformized KL property (i.e. Lemma 3.5) applies as follows

$$\|\partial F(x^k, \hat{y}^k, \rho_k)\|_- \ge \frac{1}{\phi'(F(x^k, \hat{y}^k, \rho_k) - L^*)} = \frac{1}{c}, \quad \forall\, k \in K. \tag{3.33}$$

By combining (3.33) with (3.20), we get $d_{k-1} \ge \frac{1}{bc}$ $\forall\, k \in K$ and Lemma 3.3(ii) yields

$$\frac{a}{b^2c^2} \le L(x^{k-1}, y^{k-1}) - L(x^k, y^k), \quad \forall\, k \in K,$$

which is absurd since the right-hand side of the inequality is converging to zero. Then item (i) follows.

(ii)-(iii) If $\theta \in (0, 1)$, then let $(x^*, y^*) \in A \times B$ be the unique limit point of $\{(x^k, y^k)\}_{k\in\mathbb{N}}$. By summing inequality (3.28) for all indexes $k, k+1, \ldots, K$ with $k < K$, we get

$$2\sum_{\ell=k}^{K} d_\ell \le \sum_{\ell=k}^{K} \phi_\ell + \sum_{\ell=k-1}^{K-1} d_\ell. \tag{3.34}$$

Due to the definition of ϕ_k in (3.27) and the nonnegative sign of the desingularizing function ϕ, one can also write the inequality below:

$$\sum_{\ell=k}^{K} \phi_\ell \le \frac{b}{a}\phi(L(x^k, y^k) - L^*). \tag{3.35}$$

By applying inequality (3.35) to (3.34), we get

$$2\sum_{\ell=k}^{K} d_\ell \le \frac{b}{a}\phi(L(x^k, y^k) - L^*) + \sum_{\ell=k-1}^{K-1} d_\ell.$$

By passing to the limit for $K \to \infty$ and letting

$$\Delta_k = \sum_{\ell=k}^{\infty} d_\ell, \quad \forall\, k \in \mathbb{N},$$

we finally obtain

$$\Delta_k \leq \frac{b}{a}\phi(L(x^k, y^k) - L^*) + \Delta_{k-1} - \Delta_k, \quad \forall\, k \in \mathbb{N}. \tag{3.36}$$

Let $\bar{\Xi} \subseteq \mathbb{R}^n$ be defined as in the proof of Theorem 3.6. Since $(x^k, \hat{y}^k, \rho_k) \to (x^*, y^*, 0)$, we can assume without loss of generality that $(x^k, \hat{y}^k, \rho_k) \in \bar{\Xi}$ for all $k \geq 0$. Then, the following inequalities hold

$$\phi'(L(x^k, y^k) - L^*) \geq \phi'(F(x^k, \hat{y}^k, \rho_k) - L^*) \geq \frac{1}{\|\partial F(x^k, \hat{y}^k, \rho_k)\|_-},$$

where the first inequality follows from (3.18) and the monotonicity of ϕ', whereas the second one is the application of Lemma 3.5 at the point $(x^k, \hat{y}^k, \rho_k)$. Recalling that $\phi(t) = ct^{1-\theta}$ yields

$$c(1-\theta)(L(x^k, y^k) - L^*)^{-\theta}\|\partial F(x^k, \hat{y}^k, \rho_k)\|_- \geq 1,$$

and by rearranging terms in the previous inequality and applying Lemma 3.3(iv), we get

$$(L(x^k, y^k) - L^*)^\theta \leq c(1-\theta)\|\partial F(x^k, \hat{y}^k, \rho_k)\|_- \leq bc(1-\theta)d_{k-1} = bc(1-\theta)(\Delta_{k-1} - \Delta_k).$$

Hence, the previous inequality guarantees the existence of a positive constant $q > 0$ such that

$$\phi(L(x^k, y^k) - L^*) \leq q(\Delta_{k-1} - \Delta_k)^{\frac{1-\theta}{\theta}}.$$

Then, inequality (3.36) leads to

$$\Delta_k \leq \frac{bq}{a}(\Delta_{k-1} - \Delta_k)^{\frac{1-\theta}{\theta}} + \Delta_{k-1} - \Delta_k, \quad \forall\, k \in \mathbb{N}. \tag{3.37}$$

Any nonnegative sequence $\{\Delta_k\}_{k\in\mathbb{N}}$ complying with an inequality of form (3.37) converges to zero with rate either $\mathcal{O}(\tau^k)$, $\tau \in (0,1)$ or $\mathcal{O}(k^{-(1-\theta)/(2\theta-1)})$, depending on the value of θ, see the proof of [32, Theorem 2]. By Lemma 3.3(i) and the triangular inequality, it is immediate to verify that $\Delta_k \geq \sum_{\ell=k}^{\infty} \|(x^{\ell+1} - x^\ell, y^{\ell+1} - y^\ell)\| \geq \|(x^k - x^*, y^k - y^*)\|$. Hence, the sequence $\{\|(x^k - x^*, y^k - y^*)\|\}_{k\in\mathbb{N}}$ shares the same convergence rates as $\{\Delta_k\}_{k\in\mathbb{N}}$ and the proof is complete. □

4 Application to affine rank minimization

In this section, we focus on problem (1.2). Assuming that the rank r of the sought matrix Z is given, the affine rank minimization problem can be formulated as

$$\text{find } Z \in \mathbb{R}^{n_1 \times n_2} \text{ such that } Z \in \mathcal{C} \cap \mathcal{C}_r, \tag{4.1}$$

where

$$\mathcal{C} = \{Z \in \mathbb{R}^{n_1 \times n_2} :\ \mathcal{A}(Z) = b\}, \tag{4.2}$$

$$\mathcal{C}_r = \{Z \in \mathbb{R}^{n_1 \times n_2} : \operatorname{rank}(Z) \leq r\}. \tag{4.3}$$

Since $\mathcal{C}$ is an affine subspace, the projection onto it reduces to a least squares problem; however, in applications such as Matrix Completion [33], it consists in trivial element-wise operations on the matrix. Then, in the following, we consider the projection onto $\mathcal{C}$ as known in closed form and cheap to compute. Regarding the projection onto the set $\mathcal{C}_r$, we can rely on the truncated SVD of the matrix we would like to project, as the well-known result by Eckart and Young [34, Theorem 2.4.8] states that, given a matrix $G \in \mathbb{R}^{n_1 \times n_2}$ with $r < \operatorname{rank}(G) \leq \min\{n_1, n_2\}$, the nearest matrix of rank r to G (with respect to $\|\cdot\|$) is given by the r-truncated SVD; using the notation given in (2.2) we can write

$$\mathcal{P}_{\mathcal{C}_r}(G) = [G]_1 = U_1 \Sigma_1 V_1^T. \tag{4.4}$$

We now establish the prox-regularity of C_r at any matrix of rank r and the semialgebraicity of $\mathcal{C}$ and $\mathcal{C}_r$. These properties allow us to prove the convergence of APM, RAPM and iRAPM when applied to affine rank minimization problems.

Lemma 4.1 *[21, Proposition 3.8] The set $\mathcal{C}_r$ is prox-regular at all points $\bar{X} \in \mathbb{R}^{n_1 \times n_2}$ with rank$(\bar{X}) = r$.*

Lemma 4.2 *The sets $\mathcal{C}$ and $\mathcal{C}_r$ are closed semialgebraic subsets of $\mathbb{R}^{n_1\times n_2}$.*

Proof Note that the set $\mathcal{C}$ is an affine subspace, as it is defined by a finite number of affine equations. Hence, $\mathcal{C}$ complies with the definition of semialgebraic set of $\mathbb{R}^{n_1\times n_2}$ given in equation (2.1). Likewise, given $I \subseteq \{1,\dots,n_1\},\ J \subseteq \{1,\dots,n_2\}$, we denote with Z_{IJ} the submatrix of Z obtained by keeping the rows and the columns with indexes in I and J, respectively. Observe that

$$\operatorname{rank}(Z) \le r \quad \Leftrightarrow \quad \det(Z_{IJ}) = 0,\ \forall I \subseteq \{1,\dots,n_1\},\ J \subseteq \{1,\dots,n_2\},\ \text{s.t } |I| = |J| = r+1.$$

The function $\det : \mathbb{R}^{(r+1)\times(r+1)} \to \mathbb{R}$ is polynomial with respect to the elements of Z. Therefore we can conclude that $\mathcal{C}_r$ is a semialgebraic set, as it is defined by a finite number of polynomial equations. □

4.1 Convergence Results

We now specialize the convergence properties of APM and RAPM described in Section 2 and iRAPM proposed in Section 3 when applied to the affine rank minimization problem (4.1). We first need to adjust the notation. The Euclidean norm naturally becomes the Frobenius norm, the sets A and B are now $\mathcal{C}$ and $\mathcal{C}_r$, respectively, the alternating projection methods generate two sequences $\{X^k\}_{k\in\mathbb{N}},\ \{Y^k\}_{k\in\mathbb{N}}$ of matrices and the "regularized" points are defined as follows:

$$X^k_{reg} = \frac{1}{1+\lambda_k}X^k + \frac{\lambda_k}{1+\lambda_k}Y^k, \quad Y^k_{reg} = \frac{\mu_k}{1+\mu_k}X^{k+1} + \frac{1}{1+\mu_k}Y^k. \tag{4.5}$$

Then, the function L in (2.3) takes the form

$$L(X,Y) = \frac{1}{2}\|X-Y\|^2 + \iota_{\mathcal{C}}(X) + \iota_{\mathcal{C}_r}(Y).$$

From the analysis carried out in [11] (see Theorem 2.1) and the analytical properties of the sets $\mathcal{C}$ and $\mathcal{C}_r$, we easily derive the following result for APM applied to the feasibility problem (4.1).

Corollary 4.3 *Suppose that there exists $\bar{X} \in \mathcal{C}\cap\mathcal{C}_r$ such that rank$(\bar{X}) = r$. Let $\{X^k\}_{k\in\mathbb{N}} \subseteq \mathbb{R}^{n_1\times n_2}$ and $\{Y^k\}_{k\in\mathbb{N}} \subseteq \mathbb{R}^{n_1\times n_2}$ be the sequences generated by APM with $A=\mathcal{C}$ and $B=\mathcal{C}_r$. Then $\{X^k\}_{k\in\mathbb{N}}$ and $\{Y^k\}_{k\in\mathbb{N}}$ locally converge to a matrix $Z \in \mathbb{R}^{n_1\times n_2}$ with $Z \in \mathcal{C}\cap\mathcal{C}_r$ with rates $\|X^k - Z\| = \mathcal{O}(k^{-\frac{2-\omega}{2\omega}})$ and $\|Y^k - Z\| = \mathcal{O}(k^{-\frac{2-\omega}{2\omega}})$ for some $\omega \in (0,2)$.*

Proof Lemma 4.2 ensures that $\mathcal{C}$ and $\mathcal{C}_r$ are closed semialgebraic subsets of $\mathbb{R}^{n_1\times n_2}$, and every semialgebraic set is subanalytic, see e.g. [26, p. 1208]. Since $\mathcal{C}, \mathcal{C}_r$ are closed and subanalytic, it follows that they intersect separably with $\omega \in (0,2)$ [11, Theorem 3]. Being $\mathcal{C}_r$ prox-regular at $\bar{X}$ by Lemma 4.1, it follows that $\mathcal{C}_r$ is $\sigma-$Hölder regular with respect to $\mathcal{C}$ at $\bar{X}$ for any $\sigma \in [0,1)$ and any arbitrarily small constant c [11, Corollary 3]. Then, the hypotheses of Theorem 2.1 hold, and the thesis follows. □

Likewise, the following Corollary follows for RAPM when applied to (4.1).

Corollary 4.4 *Let $\{X^k\}_{k\in\mathbb{N}} \subseteq \mathbb{R}^{n_1\times n_2}$ and $\{Y^k\}_{k\in\mathbb{N}} \subseteq \mathbb{R}^{n_1\times n_2}$ be the sequences generated by RAPM (Algorithm 1) with $A=\mathcal{C}$, $B=\mathcal{C}_r$.*

(i) Either $\|(X^k,Y^k)\| \to \infty$, or the sequence $\{(X^k,Y^k)\}_{k\in\mathbb{N}}$ converges to a stationary point of the function L.

(ii) Let $(\bar{X},\bar{Y})$ be any point such that $\|\bar{X}-\bar{Y}\| = \min\{\|X-Y\| :\ X\in\mathcal{C},\ Y\in\mathcal{C}_r\}$. If (X^0,Y^0) is sufficiently close to $(\bar{X},\bar{Y})$, then the whole sequence $\{(X^k,Y^k)\}_{k\in\mathbb{N}}$ converges to a point (X_∞,Y_∞) such that $\|X_\infty - Y_\infty\| = \min\{\|X-Y\| :\ X\in\mathcal{C},\ Y\in\mathcal{C}_r\}$.

(iii) Suppose that the sequence $\{(X^k,Y^k)\}_{k\in\mathbb{N}}$ converges to $(X^*,Y^*) \in \mathcal{C}\times\mathcal{C}_r$. Then, $\{(X^k,Y^k)\}_{k\in\mathbb{N}}$ satisfies one of the following: it terminates in a finite number of iterations; it converges with a linear rate of the form (3.31); it converges with a sublinear rate of the form (3.32).

Proof The function L is given by the sum of three semialgebraic functions, which implies that L is itself semialgebraic [30, Example 2]. Furthermore, L is a semialgebraic function that is continuous on its closed domain, which implies that it satisfies the KL property at any point with $\phi(t) = ct^{1-\theta}$, being $c > 0$ and $\theta \in [0, 1)$ two parameters depending on the given point [26, Theorem 3.1]. Hence, the thesis follows by applying Theorem 2.2 with the choice $A = \mathcal{C}$, $B = \mathcal{C}_r$. □

In analogy with Corollary 4.3, we can be more precise on the convergence rate of RAPM for affine rank minimization problems if we assume that the iterate sequence converges to the intersection of the two sets. In particular, we state and prove the following Corollary, which is, to the best of our knowledge, new, and whose proof employs a result on inf-projection taken from [35] in combination with a property holding for subanalytic sets [11].

Corollary 4.5 *Suppose that the sequence $\{(X^k, Y^k)\}_{k\in\mathbb{N}}$ generated by RAPM (Algorithm 1) converges to (X^*, X^*) with $X^* \in \mathcal{C} \cap \mathcal{C}_r$. Then, there exists $\theta \in (\frac{1}{2}, 1)$ such that $\|(X^k, Y^k) - (X^*, X^*)\| = \mathcal{O}(k^{-\frac{1-\theta}{2\theta-1}})$.*

Proof We let $m : \mathbb{R}^{n_1\times n_2} \to \mathbb{R} \cup \{\infty\}$ be defined as the inf-projection

$$\begin{aligned} m(X) &= \frac{1}{2} \min_{Y\in\mathcal{C}_r} \|X - Y\|^2 + \iota_{\mathcal{C}}(X) \\ &= \min_{Y\in\mathbb{R}^{n_1\times n_2}} \left(\frac{1}{2}\|X - Y\|^2 + \iota_{\mathcal{C}}(X) + \iota_{\mathcal{C}_r}(Y)\right) = \min_{Y\in\mathbb{R}^{n_1\times n_2}} L(X, Y). \end{aligned} \tag{4.6}$$

Let us also denote with M the following set-valued function

$$M(X) = \operatorname*{argmin}_{Y\in\mathbb{R}^{n_1\times n_2}} L(X, Y). \tag{4.7}$$

We observe that the hypotheses of [35, Theorem 3.1] hold, indeed:

- If $\bar{X} \in \mathcal{C} \cap \mathcal{C}_r$, then $M(\bar{X})$ is single-valued and $M(\bar{X}) = \{\bar{X}\}$.
- $\partial L(X, Y) \neq \emptyset$, $\forall\ X \in \mathcal{C}$, $\forall\ Y \in \mathcal{C}_r$.
- The function L is lower-bounded with respect to Y locally uniformly in X, which means that, for all $\bar{X} \in \mathbb{R}^{n_1\times n_2}$ and for all $\alpha \in \mathbb{R}$, the lower-level sets $\mathcal{L}_{\alpha,X} = \{Y \in \mathbb{R}^{n_1\times n_2} :\ L(X, Y) \leq \alpha\}$ are uniformly bounded in a neighbourhood of $\bar{X}$. Indeed, note that
$$\|Y\| \leq \|Y - X\| + \|X\| \leq \sqrt{2\alpha} + \|X\|, \quad \forall\ Y \in \mathcal{L}_{\alpha,X}.$$
Then, if $X \in \mathcal{B}(\bar{X}, \epsilon)$, namely X belongs to the ball of center $\bar{X}$ and radius ϵ, it follows that the sets $\mathcal{L}_{\alpha,X}$ are uniformly bounded by $C = \sqrt{2\alpha} + \|\bar{X}\| + \epsilon$ for all $X \in B(\bar{X}, \epsilon)$.

Therefore, we can apply [35, Theorem 3.1], which states that, if $\bar{X} \in \mathcal{C} \cap \mathcal{C}_r$ and L satisfies the KL property at the point $(\bar{X}, \bar{X})$ with desingularizing function $\phi(t) = Ct^{1-\theta}$, then the inf-projection m defined in (4.6) satisfies the KL property at $(\bar{X}, \bar{X})$ with the same exponent $\theta \in [0, 1)$.

Now, we recall that $\mathcal{C}, \mathcal{C}_r$ are both subanalytic subsets of $\mathbb{R}^{n_1\times n_2}$. Then, from [11, Lemma 3], it follows that the KL exponent θ of m at the point $(\bar{X}, \bar{X})$ must be strictly greater than $\frac{1}{2}$. Therefore, we can conclude that the KL exponent of L at $(\bar{X}, \bar{X})$ is greater than $\frac{1}{2}$, otherwise there would be a contradiction due to [35, Theorem 3.1]. At this point, the thesis follows by applying Theorem 2.2(iii) with the choice $A = \mathcal{C}$, $B = \mathcal{C}_r$, and $(x^*, y^*) = (X^*, X^*)$. □

For our proposed iRAPM applied to (4.1), we can prove the following convergence result.

Corollary 4.6 *Let $\{X^k\}_{k\in\mathbb{N}} \subseteq \mathbb{R}^{n_1\times n_2}$ and $\{Y^k\}_{k\in\mathbb{N}} \subseteq \mathbb{R}^{n_1\times n_2}$ be the sequences generated by iRAPM (Algorithm 2) with $A = \mathcal{C}$, $B = \mathcal{C}_r$. Then either $\|(X^k, Y^k)\| \to \infty$, or the sequence $\{(X^k, Y^k)\}_{k\in\mathbb{N}}$ converges to a stationary point $(X^*, Y^*) \in \mathcal{C} \times \mathcal{C}_r$ of the function L. In the latter case, the sequence satisfies one of the following: it terminates in a finite number of iterations; it converges with a linear rate of the form (3.31); it converges with a sublinear rate of the form (3.32).*

Proof Since the merit function

$$F(X, Y, \rho) = L(X, Y) + \frac{1}{2}\rho^2, \tag{4.8}$$

is the sum of two semialgebraic functions, it follows that F is itself semialgebraic; furthermore, F is continuous on its closed domain. Hence, by [26, Theorem 3.1], F satisfies the KL property at any point with $\phi(t) = ct^{1-\theta}$, being $c > 0$ and $\theta \in [0, 1)$. Then the thesis follows by applying Theorems 3.6-3.7 to Algorithm 2 with the choice $A = \mathcal{C}$, $B = \mathcal{C}_r$. □

As done for RAPM, we also specify the convergence rate of iRAPM when the iterates sequence converges to an intersection point.

Corollary 4.7 *Suppose that the sequence* $\{(X^k, Y^k)\}_{k\in\mathbb{N}}$ *generated by iRAPM (Algorithm 2) converges to* (X^*, X^*) *with* $X^* \in \mathcal{C} \cap \mathcal{C}_r$. *Then, there exists* $\theta \in (\frac{1}{2}, 1)$ *such that* $\|(X^k, Y^k) - (X^*, X^*)\| = \mathcal{O}(k^{-\frac{1-\theta}{2\theta-1}})$.

Proof Letting $Z_1 = (X, Y)$, $Z_2 = \rho$ and $Z = (Z_1, Z_2)$, the function F given in (4.8) can be seen as a block separable sum of the form $F(Z) = F_1(Z_1) + F_2(Z_2)$, where $F_1 = L$ and $F_2(Z_2) = \frac{1}{2}Z_2^2$. Note that F_1 satisfies the KL property at any point $\bar{X} \in \mathcal{C} \cap \mathcal{C}_r$ with exponent $\theta > \frac{1}{2}$ (see proof of Theorem 4.5), whereas F_2 has the same property with exponent $\frac{1}{2}$. By [36, Theorem 3.3], we conclude that F has the KL property at any point $\bar{X} \in \mathcal{C} \cap \mathcal{C}_r$ with exponent $\theta > \frac{1}{2}$. Finally, the thesis follows by applying Theorem 3.7(iii) to Algorithm 2 with the choice $A = \mathcal{C}$, $B = \mathcal{C}_r$, and $(x^*, y^*) = (X^*, X^*)$. □

4.2 Inexact Projection

In the following, we fix $k \in \mathbb{N}$ and let Y^k_{reg} be the matrix to be projected and defined in (4.5), $\hat{Y}^{k+1}$ any matrix belonging to the projection set $\mathcal{P}_{\mathcal{C}_r}(Y^k_{reg})$, and Y_{k+1} the inexact projection satisfying (3.4)-(3.5).

As outlined in Section 3.2, the accuracy requirements in (3.4)-(3.5) can be enforced without the knowledge of the exact projection provided that it is possible to define two sequences of scalars c^ℓ, a^ℓ and a sequence of matrices W^ℓ satisfying (3.8a)-(3.8c) and (3.10). Then, the inexact projection is given by $W^{\bar{\ell}}$, where $\bar{\ell}$ is the first iteration such that (3.12) is met. In this section, by exploiting the Lanczos process [17, 37], we are going to show how to compute an approximation of the r-truncated SVD of Y^k_{reg} that represents an inexact projection satisfying (3.4)-(3.5). We first briefly describe how the Lanczos method works.

4.2.1 The Lanczos method

We employ the Lanczos procedure for approximate the r largest singular values of the matrix $Y^k_{reg} \in \mathbb{R}^{n_1\times n_2}$. Given an integer ℓ with $r \le \ell \le \min\{n_1, n_2\}$, such a procedure uses Lanczos bidiagonalization, also known as Paige-Saunders bidiagonalization [38], to find two orthonormal matrices $P_{\ell+1} \in \mathbb{R}^{n_1\times(\ell+1)}$, $Q_\ell \in \mathbb{R}^{n_2\times\ell}$, such that

$$P^T_{\ell+1} Y^k_{reg} Q_\ell = B_\ell = \begin{pmatrix} \alpha_1 & & & \\ \beta_2 & \alpha_2 & & \\ & \beta_3 & \ddots & \\ & & \ddots & \alpha_\ell \\ & & & \beta_{\ell+1} \end{pmatrix} \in \mathbb{R}^{(\ell+1)\times\ell}.$$

If we consider the columns partitioning $P_{\ell+1} = [p_1, \dots, p_{\ell+1}]$ and $Q_\ell = [q_1, \dots, q_\ell]$, the column vectors p_j, q_j are the so-called *Lanczos vectors* and can be built by recurrence, starting from two vectors q_0, p_1; indeed the following recurrence holds for $\ell = 1, 2, \dots$:

$$\alpha_\ell q_\ell = (Y^k_{reg})^T p_\ell - \beta_\ell q_{\ell-1} \tag{4.9}$$

$$\beta_{\ell+1} p_{\ell+1} = Y^k_{reg} q_\ell - \alpha_\ell p_\ell, \tag{4.10}$$

where α_ℓ, β_ℓ are chosen such that the Lanczos vectors have unitary norm. Moreover, we can rewrite the recurrence in matrix form as:

$$Y^k_{reg} Q_\ell = P_{\ell+1} B_\ell \tag{4.11}$$

$$(Y^k_{reg})^T P_{\ell+1} = Q_\ell B_\ell^T + \alpha_{\ell+1} q_{\ell+1} e^T_{\ell+1}, \tag{4.12}$$

where $e_{\ell+1}$ is the $(\ell+1)$-th vector of the canonical basis. Given the bidiagonal matrix B_ℓ at hand, its SVD $B_\ell = U_B \Sigma_B V_B^T$ can be cheaply computed and the following matrix $G_\ell \in \mathbb{R}^{n_1\times n_2}$ is defined

$$G_\ell = P_{\ell+1} B_\ell Q_\ell^T = \underbrace{P_{\ell+1} U_B}_{\tilde{U}_\ell} \underbrace{\Sigma_B}_{\tilde{\Sigma}_\ell} \underbrace{V_B^T Q_\ell^T}_{\tilde{V}_\ell^T} = \tilde{U}_\ell \tilde{\Sigma}_\ell \tilde{V}_\ell^T, \tag{4.13}$$

where $\tilde{\Sigma}_\ell \in \mathbb{R}^{(\ell+1)\times\ell}$ is a rectangular diagonal matrix whose diagonal entries $\tilde{\sigma}_{1,\ell}, \dots, \tilde{\sigma}_{\ell,\ell}$ can be seen as approximations of the largest ℓ singular values $\sigma_1, \dots, \sigma_\ell$ of Y^k_{reg}. Note also that the first ℓ singular values of B_ℓ and G_ℓ are the same and G_ℓ has $\min\{n_1, n_2\} - \ell$ extra null singular values. With the aim of approximating the r-truncated SVD of Y^k_{reg}, using the notation (2.2), we write

$$B_\ell = \underbrace{(U_B)_1(\Sigma_B)_1(V_B)_1^T}_{=[B_\ell]_1} + \underbrace{(U_B)_2(\Sigma_B)_2(V_B)_2^T}_{=[B_\ell]_2} = [B_\ell]_1 + [B_\ell]_2,$$

with $(U_B)_1 \in \mathbb{R}^{(\ell+1)\times r}, (U_B)_2 \in \mathbb{R}^{(\ell+1)\times(\ell+1-r)}, (V_B)_1 \in \mathbb{R}^{\ell\times r}, (V_B)_2 \in \mathbb{R}^{\ell\times(\ell-r)}, \Sigma_1 \in \mathbb{R}^{r\times r}, \Sigma_2 \in \mathbb{R}^{(\ell+1-r)\times(\ell-r)}$. Then, using the notation $G_\ell = [G_\ell]_1 + [G_\ell]_2$, where $[G_\ell]_1 = P_{\ell+1}[B_\ell]_1 Q_\ell^T$ and $[G_\ell]_2 = P_{\ell+1}[B_\ell]_2 Q_\ell^T$, we can truncate the decomposition in (4.13) and define

$$[G_\ell]_1 = \underbrace{P_{\ell+1}(U_B)_1}_{\tilde{U}_r} \underbrace{(\Sigma_B)_1}_{\tilde{\Sigma}_r} \underbrace{(V_B)_1^T Q_\ell^T}_{\tilde{V}_r^T} = \tilde{U}_r \tilde{\Sigma}_r \tilde{V}_r.$$

The preceding discussion indicates how singular values estimates can be obtained via the Lanczos method, but it reveals nothing about the approximation quality of the singular values and corresponding singular vectors. We refer to [34, Section 10.4] and references therein for the convergence results of the Lanczos procedure. Here, we highlight the following properties that are crucial for our analysis:

- in exact arithmetic the method has finite termination; indeed, when $\ell = \min\{n_1, n_2\}$, the SVD in (4.13) is the exact SVD of Y^k_{reg}. However, the procedure is generally implemented as an iterative procedure due to round-off errors and to avoid to perform a large number of steps;
- the inexact singular values $\tilde{\sigma}_{i,\ell}$ approximate the corresponding σ_i from below [17, Th. 6.4];
- letting $(U_B)_{\ell+1,j}$ be the last element of the j-th column of U_B, as shown in [37], we have that

$$\min_{\sigma\in\sigma(Y^k_{reg})} |\sigma - \tilde{\sigma}_{j,\ell}| \le |\alpha_{\ell+1}||(U_B)_{\ell+1,j}|. \tag{4.14}$$

In the classic implementations [37], the index ℓ is adaptively increased in order to reach the desired accuracy in the approximation of the first r singular values, using the error bound given in (4.14), that is based only on a byproduct of the Lanczos method. In the next subsection, we show that also the accuracy requirements (3.12) in iRAPM can be checked with no extra computational cost using information already provided by the Lanczos process.

4.2.2 Inexactness Criteria

Letting G_ℓ be the matrix given in (4.13), note that its first ℓ singular values are $\tilde{\sigma}_{1,\ell}, \dots, \tilde{\sigma}_{\ell,\ell}$, while the remaining are zero. Although the singular values $\tilde{\sigma}_{i,\ell}$, $i = 1, \dots, \ell$, are well-defined only for $\ell = 1, \dots, \min\{n_1, n_2\}$, we can artificially set $\tilde{\sigma}_{i,\ell} = \sigma_i$, $i = 1, \dots, \min\{n_1, n_2\}$ for $\ell > \min\{n_1, n_2\}$. We assume that at step $\ell > r$ of the Lanczos procedure it holds $\tilde{\sigma}_{r,\ell} > \tilde{\sigma}_{r+1,\ell}$, i.e., the r and $r+1$ singular values of G_ℓ are distinct. This is a reasonable assumption since the method is expected to converge to a r-rank matrix. Then, let $\kappa_\ell > 0$ be given by

$$\kappa_\ell = \frac{2}{1-\gamma} \cdot \frac{(1-\gamma)\tilde{\sigma}_{r,\ell} + \gamma\tilde{\sigma}_{r+1,\ell}}{\tilde{\sigma}_{r,\ell} - \tilde{\sigma}_{r+1,\ell}}, \tag{4.15}$$

where $\gamma \in (0, 1)$. For any $\ell > r$, we define the sequences

$$W^\ell = \begin{cases} [G^\ell]_1, & \text{if } \ell \le \min\{n_1, n_2\}, \\ [Y^k_{reg}]_1, & \text{if } \ell > \min\{n_1, n_2\} \end{cases} \tag{4.16}$$

$$c^\ell = \sum_{i=r+1}^{\min\{\ell, n_1, n_2\}} \tilde{\sigma}^2_{i,\ell}, \tag{4.17}$$

$$a^\ell = \begin{cases} \kappa_\ell \|Y^k_{reg} - G_\ell\| & \text{for } \ell \le \min\{n_1, n_2\}, \\ 0 & \text{for } \ell > \min\{n_1, n_2\} \end{cases}. \tag{4.18}$$

In the following, we show that these sequences satisfy (3.8a)-(3.8c) and (3.10).

Lemma 4.8 *The sequences* (4.16) *and* (4.17) *satisfy conditions* (3.8a)-(3.8c).

Proof From [17, Theorem 6.4], we have that $\tilde{\sigma}_{i,\ell} \leq \sigma_i$ for all $\ell < \min\{n_1, n_2\}$, and from the finite termination of the method we have $\tilde{\sigma}_{i,\ell} = \sigma_i$ for $\ell \geq \min\{n_1, n_2\}$. Then,

$$c_\ell = \sum_{i=r+1}^{\min\{\ell,n_1,n_2\}} \tilde{\sigma}_{i,\ell}^2 \leq \sum_{i=r+1}^{\min\{\ell,n_1,n_2\}} \sigma_i^2 \leq \sum_{i=r+1}^{\min\{n_1,n_2\}} \sigma_i^2 = \|\hat{Y}^{k+1} - Y_{reg}^k\|^2, \quad \forall\, \ell \in \mathbb{N},$$

$$\lim_{\ell\to\infty} c_\ell = \lim_{\ell\to\infty} \sum_{i=r+1}^{\min\{\ell,n_1,n_2\}} \tilde{\sigma}_{i,\ell}^2 = \sum_{i=r+1}^{\min\{n_1,n_2\}} \sigma_i^2 = \|\hat{Y}^{k+1} - Y_{reg}^k\|^2.$$

Then, (3.8a) and (3.8b) holds. Condition (3.8c) follows by the definition of W^ℓ in (4.2.1). □

Lemma 4.9 *Let $\tilde{\sigma}_{i,\ell}$, $i = 1, \ldots, \ell$, denote the singular values of G_ℓ and assume that $\tilde{\sigma}_{r,\ell} > \tilde{\sigma}_{r+1,\ell}$. Then, the sequences given in* (4.2.1) *and* (4.18) *satisfy condition* (3.10).

Proof The limit in (3.10) holds by the finite termination of the Lanczos algorithm and the definitions in (4.2.1) and (4.18). We now show that the inequality in (3.10) holds. We limit ourselves to the case $\ell < \min\{n_1, n_2\}$, as otherwise the thesis holds by (4.2.1) and (4.18). For simplicity, we omit the index k and let $m = \min\{n_1, n_2\}$. Using the notation given in (2.2), we consider the block partitions

$$Y_{reg} = [Y_{reg}]_1 + [Y_{reg}]_2, \quad G_\ell = [G_\ell]_1 + [G_\ell]_2. \tag{4.19}$$

Then, we have that $\hat{Y}^{k+1} = [Y_{reg}]_1$ and $W^\ell = [G_\ell]_1$. The SVD of Y_{reg} and G_ℓ can be written as

$$Y_{reg} = U_{Y_{reg}} \mathrm{diag}(\sigma_1, \ldots, \sigma_m) V_{Y_{reg}}^T, \quad G_\ell = U_{G_\ell} \mathrm{diag}(\tilde{\sigma}_{1,\ell}, \ldots, \tilde{\sigma}_{\ell,\ell}, 0, \ldots, 0) V_{G_\ell}^T.$$

Given $b_1 \geq 0$ and $b_2 \geq 0$, we observe that the SVD of $[Y_{reg}]_1 + b_1[Y_{reg}]_2$ and $[G_\ell]_1 + b_2[G_\ell]_2$ are

$$[Y_{reg}]_1 + b_1[Y_{reg}]_2 = U_{Y_{reg}} \mathrm{diag}(\sigma_1, \ldots, \sigma_r, b_1\sigma_{r+1}, \ldots, b_1\sigma_m) V_{Y_{reg}}^T$$
$$[G_\ell]_1 + b_2[G_\ell]_2 = U_{G_\ell} \mathrm{diag}(\tilde{\sigma}_{1,\ell}, \ldots, \tilde{\sigma}_{r,\ell}, b_2\tilde{\sigma}_{r+1,\ell}, \ldots, b_2\tilde{\sigma}_{\ell,\ell}, 0, \ldots, 0) V_{G_\ell}^T.$$

Then, we have:

$$[Y_{reg}]_1 = \mathcal{P}_{\mathcal{C}_r}([Y_{reg}]_1 + b_1[Y_{reg}]_2), \quad \forall\, b_1 \in [0, \sigma_r/\sigma_{r+1}) \tag{4.20}$$
$$[G_\ell]_1 = \mathcal{P}_{\mathcal{C}_r}([G_\ell]_1 + b_2[G_\ell]_2), \quad \forall\, b_2 \in [0, \tilde{\sigma}_{r,\ell}/\tilde{\sigma}_{r+1,\ell}). \tag{4.21}$$

Equation (4.20) yields

$$\|[Y_{reg}]_1 - ([Y_{reg}]_1 + b_1[Y_{reg}]_2)\|^2 \leq \|[G_\ell]_1 - ([Y_{reg}]_1 + b_1[Y_{reg}]_2)\|^2.$$

The above inequality implies

$$\|[G_\ell]_1 - [Y_{reg}]_1\|^2 + \|b_1[Y_{reg}]_2\|^2 - 2\langle [G_\ell]_1 - [Y_{reg}]_1, b_1[Y_{reg}]_2 \rangle \geq \|b_1[Y_{reg}]_2\|^2,$$

which yields

$$\frac{1}{2b_1}\|[G_\ell]_1 - [Y_{reg}]_1\|^2 - \langle [G_\ell]_1 - [Y_{reg}]_1, [Y_{reg}]_2 \rangle \geq 0. \tag{4.22}$$

Using (4.21) and following a similar reasoning as before, we get

$$\frac{1}{2b_2}\|[Y_{reg}]_1 - [G_\ell]_1\|^2 - \langle [Y_{reg}]_1 - [G_\ell]_1, [G_\ell]_2 \rangle \geq 0. \tag{4.23}$$

By summing up (4.22) and (4.23), we obtain:

$$0 \leq \left(\frac{1}{2b_1} + \frac{1}{2b_2}\right)\|[Y_{reg}]_1 - [G_\ell]_1\|^2 - \langle [G_\ell]_1 - [Y_{reg}]_1, [Y_{reg}]_2 - [G_\ell]_2 \rangle. \tag{4.24}$$

Recalling that $[Y_{reg}]_2 = Y_{reg} - [Y_{reg}]_1$ and $[G_\ell]_2 = G_\ell - [G_\ell]_1$, we have:

$$-\left(\frac{1}{2b_1} + \frac{1}{2b_2}\right)\|[Y_{reg}]_1 - [G_\ell]_1\|^2 \leq \langle [Y_{reg}]_1 - [G_\ell]_1, Y_{reg} - G_\ell \rangle - \|[Y_{reg}]_1 - [G_\ell]_1\|^2,$$

which implies

$$\left(1 - \frac{1}{2b_1} - \frac{1}{2b_2}\right)\|[Y_{reg}]_1 - [G_\ell]_1\|^2 \leq \langle [Y_{reg}]_1 - [G_\ell]_1, Y_{reg} - G_\ell \rangle.$$

By limiting ourselves to the case where $\left(1-\frac{1}{2b_1}-\frac{1}{2b_2}\right)>0$, which is verified by imposing $b_1, b_2 \geq 1$ (where equality cannot hold simultaneously for both), and using Cauchy-Schwarz inequality, we have:

$$\left(1-\frac{1}{2b_1}-\frac{1}{2b_2}\right)\|[Y_{reg}]_1-[G_\ell]_1\| \leq \|Y_{reg}-G_\ell\|. \tag{4.25}$$

Let us set $b_1=1$ and $b_2=\gamma+(1-\gamma)\tilde{\sigma}_{r,\ell}/\tilde{\sigma}_{r+1,\ell}=\frac{(1-\gamma)\tilde{\sigma}_{r,\ell}+\gamma\tilde{\sigma}_{r+1,\ell}}{\tilde{\sigma}_{r+1,\ell}}$, with $\gamma\in(0,1)$. With this choice of b_1 and b_2, the constant in (4.25) becomes:

$$\begin{aligned}\left(1-\frac{1}{2b_1}-\frac{1}{2b_2}\right) &= \left(1-\frac{1}{2}-\frac{\tilde{\sigma}_{r+1,\ell}}{2((1-\gamma)\tilde{\sigma}_{r,\ell}+\gamma\tilde{\sigma}_{r+1,\ell})}\right)\\ &= \frac{1}{2}\left(1-\frac{\tilde{\sigma}_{r+1,\ell}}{(1-\gamma)\tilde{\sigma}_{r,\ell}+\gamma\tilde{\sigma}_{r+1,\ell}}\right)=\frac{(1-\gamma)}{2}\cdot\frac{\tilde{\sigma}_{r,\ell}-\tilde{\sigma}_{r+1,\ell}}{(1-\gamma)\tilde{\sigma}_{r,\ell}+\gamma\tilde{\sigma}_{r+1,\ell}}=\frac{1}{\kappa_\ell},\end{aligned}$$

where κ_ℓ is given in (4.15). Then, from (4.25), we have:

$$\|[Y_{reg}]_1-[G_\ell]_1\| \leq \kappa_\ell\|Y_{reg}-G_\ell\| = a^\ell,$$

and the first part of (3.10) is verified. □

4.2.3 Computation of the right-hand side of the stopping criteria (3.12)

The criteria (3.12) require to compute additional quantities, in particular the Frobenius norms $\|W^\ell - Y^k_{reg}\|$ and $\|G_\ell - Y^k_{reg}\|$. We show that these two quantities can be computed without additional cost as a byproduct of the Lanczos process.

We start recalling few useful relations. By (4.11) it follows $P^T_{\ell+1}Y^k_{reg}Q_\ell = B_\ell$ and by (4.13) we have $G_\ell = P_{\ell+1}B_\ell Q^T_\ell$. Moreover, since the columns of $P_{\ell+1}$ and Q_ℓ are orthonormal, it holds:

$$p_i^T p_j = 0, \quad 1\leq i,j\leq \ell+1,\ i\neq j, \tag{4.26}$$

$$q_i^T q_j = 0, \quad 1\leq i,j\leq \ell,\ i\neq j, \tag{4.27}$$

$$(I-P_{\ell+1}P^T_{\ell+1})p_i = 0, \quad 1\leq i\leq \ell+1. \tag{4.28}$$

Inspired by [39], in the next Lemma we show that $\|Y^k_{reg}-G_\ell\|^2$ can be computed by recurrence exploiting the entries of B_ℓ.

Lemma 4.10 *Let us denote $\omega_\ell = \|Y^k_{reg}-G_\ell\|^2$ and let α_ℓ, $\beta_{\ell+1}$ be given in (4.9)-(4.10). Then,*

$$\omega_{\ell-1} = \omega_\ell + \alpha_\ell^2 + \beta^2_{\ell+1}.$$

Proof For the sake of simplicity, we omit the index k. From (4.12) we have:

$$\begin{aligned}Y_{reg}-G_\ell = Y_{reg}-P_{\ell+1}B_\ell Q_\ell^T &= Y_{reg}-P_{\ell+1}P^T_{\ell+1}Y_{reg}+\alpha_{\ell+1}P_{\ell+1}e_{\ell+1}q^T_{\ell+1}\\ &= (I-P_{\ell+1}P^T_{\ell+1})Y_{reg}+\alpha_{\ell+1}p_{\ell+1}q^T_{\ell+1}.\end{aligned} \tag{4.29}$$

We observe that:

$$(I-P_{\ell+1}P^T_{\ell+1})Y_{reg} = (I-P_\ell P_\ell^T)Y_{reg} - p_{\ell+1}p^T_{\ell+1}Y_{reg}. \tag{4.30}$$

Then, (4.29), (4.30) and (4.9), yield:

$$\begin{aligned}\omega_{\ell-1} &= \|Y_{reg}-G_{\ell-1}\|^2 = \|(I-P_\ell P_\ell^T)Y_{reg}+\alpha_\ell p_\ell q_\ell^T\|^2\\ &= \|(I-P_{\ell+1}P^T_{\ell+1})Y_{reg}+p_{\ell+1}p^T_{\ell+1}Y_{reg}+\alpha_\ell p_\ell q_\ell^T\|^2\\ &= \|(I-P_{\ell+1}P^T_{\ell+1})Y_{reg}+\alpha_{\ell+1}p_{\ell+1}q^T_{\ell+1}+\beta_{\ell+1}p_{\ell+1}q_\ell^T+\alpha_\ell p_\ell q_\ell^T\|^2\\ &= \omega_\ell + \|\beta_{\ell+1}p_{\ell+1}q_\ell^T+\alpha_\ell p_\ell q_\ell^T\|^2 + 2\langle (I-P_{\ell+1}P^T_{\ell+1})Y_{reg}+\alpha_{\ell+1}p_{\ell+1}q^T_{\ell+1}, \beta_{\ell+1}p_{\ell+1}q_\ell^T+\alpha_\ell p_\ell q_\ell^T\rangle.\end{aligned}$$

By using (4.26), (4.27), and (4.28), it follows that the Frobenius inner product is null. In addition, both $p_{\ell+1}q_\ell^T$ and $p_\ell q_\ell^T$ are rank-one matrices and the vectors $p_{\ell+1}, p_\ell$ and q_ℓ are unitary vectors, then $\|p_{\ell+1}q_\ell^T\|=1$, $\|p_\ell q_\ell^T\|=1$ and $\|\beta_{\ell+1}p_{\ell+1}q_\ell^T+\alpha_\ell p_\ell q_\ell^T\|^2 = \alpha_\ell^2+\beta^2_{\ell+1}$ and the thesis follows. □

The following Lemma shows the relationship between $\|Y^k_{reg}-W^\ell\|^2$ and $\|Y^k_{reg}-G_\ell\|^2$.

Lemma 4.11 *Let $\tilde{\sigma}_{i,\ell}$, $i = 1, \dots, \ell$, denote the singular values of G_ℓ. Then we have*

$$\|Y_{reg}^k - W^\ell\|^2 = \|Y_{reg}^k - G_\ell\|^2 + \sum_{i=r+1}^{\ell} \tilde{\sigma}_{i,\ell}^2.$$

Proof We omit the iteration index k, for the sake of simplicity. We recall that $W^\ell = [G_\ell]_1$, in addition (4.19) holds and $[G_\ell]_2 = P_{\ell+1}[B_\ell]_2 Q_\ell^T$, where $[B_\ell]_2 = (U_B)_2(\Sigma_B)_2(V_B)_2$. Then:

$$\begin{aligned}\|Y_{reg} - W^\ell\|^2 &= \|Y_{reg} - [G_\ell]_1\|^2 = \|Y_{reg} - G_\ell + [G_\ell]_2\|^2 \\ &= \|Y_{reg} - G_\ell\|^2 + \|[G_\ell]_2\|^2 + 2\langle Y_{reg} - G_\ell, P_{\ell+1}[B_\ell]_2 Q_\ell^T\rangle.\end{aligned}$$

We can show that the Frobenius inner product is null. Indeed, by rewriting $Y_{reg} - G_\ell$ as in (4.29), we have:

$$\begin{aligned}\langle Y_{reg} - G_\ell, P_{\ell+1}[B_\ell]_2 Q_\ell^T\rangle &= \langle (I - P_{\ell+1}P_{\ell+1}^T)Y_{reg} + \alpha_{\ell+1} p_{\ell+1} q_{\ell+1}^T, P_{\ell+1}[B_\ell]_2 Q_\ell^T\rangle \\ &= \langle (I - P_{\ell+1}P_{\ell+1}^T)Y_{reg}, P_{\ell+1}[B_\ell]_2 Q_\ell^T\rangle + \alpha_{\ell+1}\langle p_{\ell+1} q_{\ell+1}^T, P_{\ell+1}[B_\ell]_2 Q_\ell^T\rangle \\ &= \mathrm{tr}\Big(Y_{reg}^T (I - P_{\ell+1}P_{\ell+1}^T) P_{\ell+1}[B_\ell]_2 Q_\ell^T\Big) + \alpha_{\ell+1}\mathrm{tr}\left(q_{\ell+1} p_{\ell+1}^T P_{\ell+1}[B_\ell]_2 Q_\ell^T\right). \quad (4.31)\end{aligned}$$

The first term in (4.31) is null by (4.28), while the second, for the properties of the trace, becomes $\mathrm{tr}\left(Q_\ell^T q_{\ell+1} p_{\ell+1}^T P_{\ell+1}[B_\ell]_2\right)$ and then it is null for (4.27). On the other hand, by the definition of Frobenius norm, we have $\|[G_\ell]_2\|^2 = \sum_{i=r+1}^{\ell} \tilde{\sigma}_{i,\ell}^2$. Then, the thesis follows. □

5 Numerical illustration

In this section, we provide numerical tests which illustrate the behaviour of the iRAPM procedure. We focused on the Matrix Completion (MC) problem, which consists in recovering a matrix $M \in \mathbb{R}^{n_1 \times n_2}$ from the knowledge of a subset of its entries; namely, the entries with indexes in the index set $\Omega = \{(i,j) : M_{i,j} \text{ is known}\}$. In the following, we will write $(M)_\Omega \in \mathbb{R}^{n_1 \times n_2}$ to denote the matrix obtained by keeping the entries of M with indices in Ω and setting to 0 the others, and analogously for $(M)_{\bar{\Omega}} \in \mathbb{R}^{n_1 \times n_2}$ with $\bar{\Omega}$ the complement of Ω. Without further constraints, there exists an infinite number of matrices $Z \in \mathbb{R}^{n_1 \times n_2}$ that comply with the condition $Z_{i,j} = M_{i,j}$ for all $(i,j) \in \Omega$. On the other hand, in most applications, one can assume that the matrix to be recovered has low rank, for more details, see [33]. Problems that can be approached with MC arise in various fields, such as recommendation systems [40], image reconstruction [41], and data imputation [42, 43].

Then, the MC problem is generally formulated as the affine rank minimization problem

$$\begin{aligned}\min \quad & \mathrm{rank}(Z) \\ \text{s.t.} \quad & (Z)_\Omega = (M)_\Omega\end{aligned}. \quad (5.1)$$

This is a combinatorial optimization problem that requires a worst-case exponential running time.

In the literature, problem (5.1) has been approached in different ways, for example, replacing the rank function with a convex function like the nuclear norm [33], that is, the sum of the singular values of the matrix, or reformulating it as Semidefinite Programming Problem [44].

We are interested in reformulating the MC problem as the feasibility problem (4.1). In fact, given an estimate r of the rank of the sought matrix, we can reformulate it as the problem of finding a matrix in the intersection of the rank level set and the set of matrices that match the observed entries, i.e.:

$$\text{find } Z \in \mathbb{R}^{n_1 \times n_2} \text{ such that } Z \in \mathcal{C} \cap \mathcal{C}_r, \quad (5.2)$$

where

$$\mathcal{C} = \{Z \in \mathbb{R}^{n_1 \times n_2} : \ Z_{i,j} = M_{i,j}, \ \forall\ (i,j) \in \Omega\}, \quad (5.3)$$

$$\mathcal{C}_r = \{Z \in \mathbb{R}^{n_1 \times n_2} : \mathrm{rank}(Z) \le r\}. \quad (5.4)$$

As the function that associates M with $(M)_\Omega$ is a linear map, $\mathcal{C}$ is an affine set and the projection onto it is given by

$$\mathcal{P}_{\mathcal{C}}(Z) = (M)_\Omega + (Z)_{\bar{\Omega}}. \quad (5.5)$$

The inexact projection onto $\mathcal{C}_r$ in iRAPM is computed via the Lanczos method described in Section 4.2, equipped with the stopping criteria (3.12), while in APM and RAPM we use the standard stopping criterion based on the bound (4.14). We note that also in APM and RAPM the projection is not exact; however, we expect it to be more accurate than in iRAPM, more expensive and not justified to preserve the overall convergence behaviour.

For all the three methods under investigation, our implementation of the Lanczos procedure is based on the PROPACK code by Larsen [37], where we changed the stopping criterion when used in iRAPM. In APM and RAPM we used the standard stopping criterion and default tolerances, that is, $16\varepsilon_m$, where ε_m denotes the machine precision.

5.1 Projection operators: computational costs

Before presenting the numerical results, we analyze the computational costs of the projection operators involved in the MC problem (5.2).

As observed above, projecting onto $\mathcal{C}$ consists in an element-wise replacement and can be computed in $n_1 n_2$ operations. On the other hand, the projection onto $\mathcal{C}_r$ via the Lanczos method described in Section 4.2 has two main parts that concur to the final cost: updating the Lanczos vectors (4.9)-(4.10), and checking the criterion (4.14) for APM and RAPM (the criteria (3.12) for iRAPM, respectively).

The Lanczos procedure requires at each iteration two matrix-vector products, that is, $O(n_1 n_2)$ operations. Then, assuming that the final Krylov subspace dimension at the k-th outer iteration is $\bar{\ell}_k$, the final cost for building the Krylov subspace is of the order $O(n_1 n_2 \bar{\ell}_k)$.

The standard criterion (4.14), employed in APM and RAPM, requires to check r inequalities at most, then it costs $O(r)$ each time that the stopping criterion is checked. The evaluation of the two Frobenius norms required in the criteria (3.12) in the iRAPM procedure can be significantly reduced exploiting Lemma 4.10 and 4.11. In fact, the two norms can be computed as follows:

$$\begin{aligned} \|Y^k_{reg} - G_\ell\|^2 &= \|Y^k_{reg} - G_0\|^2 - \textstyle\sum_{i=1}^{\ell} \alpha_i^2 - \sum_{j=2}^{\ell+1} \beta_j^2, \\ \|Y^k_{reg} - W^\ell\|^2 &= \|Y^k_{reg} - G_\ell\|^2 + \textstyle\sum_{i=r+1}^{\ell} \tilde{\sigma}^2_{i,\ell}. \end{aligned} \tag{5.6}$$

where G_ℓ is given in (4.13). Then, using (5.6), the inexact criteria require to compute the Frobenius norm $\|Y^k_{reg} - G_0\|$, whose cost is $O(n_1 n_2)$, only once at the beginning of the Lanczos procedure, $3\bar{\ell}_k - r$ sums and the check of only 2 inequalities. Then, in all the alternating projection approaches under investigation, the cost of checking the stopping criteria is dominated by the cost of the Lanczos iterations, and we can conclude that, at a generic iteration k, the projection onto $\mathcal{C}_r$ costs $O(n_1 n_2 \bar{\ell}_k)$.

We stress that the strategy for increasing the dimension of the Krylov subspace has a crucial impact on the final cost of the projection onto $\mathcal{C}_r$. The PROPACK code implements an adaptive strategy to reduce the number of criteria checks. This strategy can result in selecting unnecessarily large values of $\bar{\ell}_k$. In our implementation of the three alternating projection methods, we chose to check the stopping criteria at each Lanczos iteration in order to select the minimum Krylov dimension for which the criteria are satisfied.

5.2 Experiments on Matrix Completion

For our experiments, we consider two different types of matrices: random Gaussian matrices of rank r obtained as $M = LR^T \in \mathbb{R}^{n_1 \times n_2}$, where $L \in \mathbb{R}^{n_1 \times r}$, $R \in \mathbb{R}^{n_2 \times r}$ are matrices whose entries are sampled from a normal distribution, and the r-truncated SVD of the grayscale 512×512 `lake` image, see Figure 2 (left). We choose $n_1 = n_2 = 512$ for the Gaussian matrices, and in both cases we choose $r = 30$. Then, we built the Gaussian matrices and truncated the image accordingly.

The Gaussian matrices and the truncated image have different singular value distributions as illustrated in Figure 2 (right), where we clearly see that the singular values of the truncated image decade faster than those of the Gaussian matrix. Then, this choice of the matrices allows us to test iRAPM against different singular value distributions.

The set Ω is built sampling q entries uniformly at random, where the number of observed entries q is chosen such that the ratio $q/((n_1 + n_2 - r)r) = 2.6$ in both cases. We consider the choice $X^0 = (M)_\Omega$, $Y^0 = \mathcal{P}_{\mathcal{C}_r}(X^0)$ as starting points for RAPM and iRAPM, while APM only needs Y^0. We compare the performance of the methods by performing a prefixed number of iterations $k_f = 200$. In this way, we can observe the effective saving of inexact projections in iRAPM along iterations while comparing the final reconstruction error.

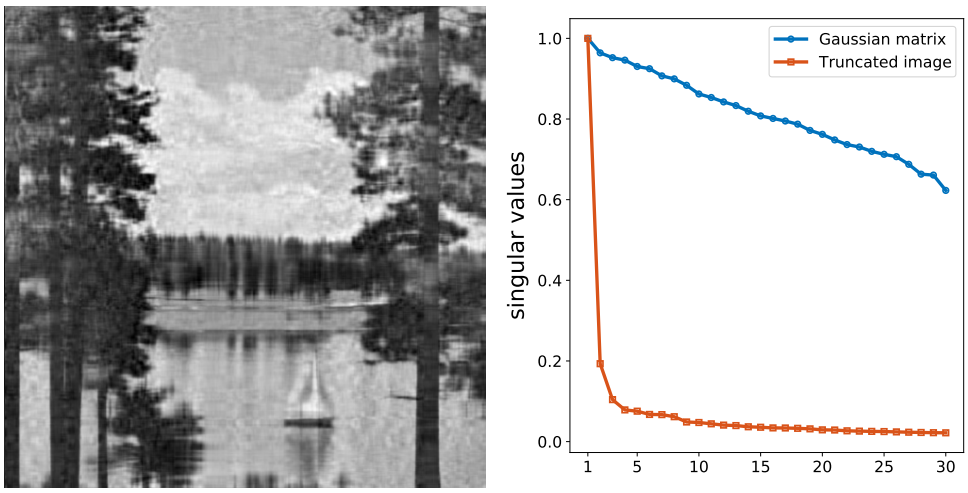

Fig. 2 Truncated `lake` image (left) and normalized singular values of a Gaussian matrix (blue circles) and of the truncated image (red squares) (right).

Two metrics are considered to monitor the performance of the methods:

$$e_\Omega(Y^k) = \frac{\|(M)_\Omega - (Y^k)_\Omega\|}{\|(M)_\Omega\|}, \tag{5.7}$$

$$e_{mse}(Y^{k_f}) = \frac{\|M - Y^{k_f}\|^2}{n_1 n_2}. \tag{5.8}$$

The former measures the error on the observed entries along the iterations, while the latter is the mean squared error on all the entries at termination and it measures the quality of the final reconstruction.

As a measure of the overall computational cost needed to reach iteration k and to compare the three approaches under investigation, we consider the quantity

$$\texttt{cost}(k) = \sum_{i=1}^{k} (\bar{\ell}_i - r),$$

as at each iteration k, the Lanczos process is started with a Krylov subspace of dimension r, then we can obtain computational saving only reducing $\bar{\ell}_k - r$. In Tables 1 and 2 we provide statistics of the three considered methods; for the regularized methods we used $\lambda_k = \mu_k = 16$ and iRAPM was tested with different values of ζ in (3.12). The reported statistics are the means over 5 random runs. Columns 3 and 4 show the values of the two metrics (5.7) and (5.8) at termination, and we also report in brackets the standard deviation of the mean-squared error. Finally, column 5, reports the quantity $\texttt{cost}(k_f)$. As a first comment, we observe that the choice of ζ in iRAPM does not seem to be critical, and any value smaller than 10^{-3} provides an overall reduction of the computational cost as the dimension of the Krylov subspaces employed in iRAPM is generally smaller than those employed in RAPM.

In Figures 3-5 we report information on the behavior of the procedures along the iterations; iRAPM was run using $\zeta = 10^{-7}$. In Figure 3, we plot the average over all the runs of the metric (5.7) versus the iterations (line), while the colored bands represent the minimum-maximum range of the metric. In the remaining two figures we focus on a representative single run: in Figure 4 we plot for each method the ratio between $\texttt{cost}(k)$ and the corresponding quantity for APM. This way, we compare the effective computational costs of RAPM and iRAPM as a fraction of the cost of APM. The plots in Figure 5 show two different quantities: the solid lines represent $\bar{\ell}_k$ versus the iterations, while the dashed lines display the number of singular values that satisfy the criterion (4.14), i.e. the number of singular values computed by iRAPM that are approximated with the standard accuracy requirement.

Focusing on the Gaussian matrices, we observe that RAPM and iRAPM are comparable in terms of accuracy and only slightly worse than APM. This is outlined in the statistics in Table 1 and in Figure 3. Figure 4 clearly shows that RAPM and APM require approximately the same computational effort, while the saving provided by the use of the new accuracy requirements (3.12) in iRAPM is relevant in the first phase of the convergence history (the first 20 iterations), while from the 20th iteration onward the

Method	ζ	e_Ω	e_{mse}	cost(k_f)
APM		2.968e-06	3.309e-09 (5.609e-09)	836
RAPM		7.638e-06	1.733e-08 (2.378e-08)	870
iRAPM	1e-09	8.229e-06	2.101e-08 (3.046e-08)	715
iRAPM	1e-07	8.105e-06	1.889e-08 (2.462e-08)	713
iRAPM	1e-05	8.204e-06	2.054e-08 (2.925e-08)	712
iRAPM	1e-03	7.606e-06	1.699e-08 (2.307e-08)	728

Table 1 Gaussian matrices: statistics of the runs, $maxit = 200$.

Method	ζ	e_Ω	e_{mse}	cost(k_f)
APM		5.009e-04	1.606e-06 (8.758e-07)	1654
RAPM		7.785e-04	3.092e-06 (1.374e-06)	1722
iRAPM	1e-09	5.633e-04	2.061e-06 (1.118e-06)	956
iRAPM	1e-07	5.181e-04	1.733e-06 (1.089e-06)	969
iRAPM	1e-05	5.056e-04	1.609e-06 (1.388e-06)	936
iRAPM	1e-03	7.652e-04	2.934e-06 (1.520e-06)	1244

Table 2 `lake` image: statistics of the runs, $maxit = 200$

savings settle around the 15%. This is also evident in Figure 5 where we can see that iRAPM employes Krylov subspaces of dimension lower than 40 in the first 20 outer iterations, while higher dimensions are needed by RAPM and APM. On the other hand, we observe that the Lanczos procedure provides accurate singular values in the final phase of the convergence history employing Krylov subspaces of dimension slightly larger than r; this behaviour can be ascribed to the fact that Y_{reg}^k becomes closer and closer to a rank r matrix, and ideally, in case of rank r matrices, the Lanczos process stops with ℓ no greater than $r + 1$ [17]. It is also interesting to observe that the approximated singular values used to compute the projection in iRAPM are, as expected, less accurate than in RAPM and APM, except for approximately 50 consecutive outer iterations starting from the 100th, and few more isolated ones. However, the overall convergence behavior and the final accuracy are comparable to those of APM and RAPM, showing that the standard criterion, except when the iterates are close to the sought solution, yields to the *oversolving* phenomena, i.e. the singular values are approximated with an accuracy higher than those needed for the convergence of the overall procedure.

Concerning the `lake` image, as before, the performances of the three methods are comparable in terms of accuracy, but this time iRAPM achieves an error similar to APM outperforming the RAPM variant. Figure 4 (right) shows that in this case the saving of the inexact approach in term of computational cost is even more relevant than in the case of Gaussian matrices. Indeed, in the first iterations, saving is over the 50%, and later on the savings are approximately the 40%. From Figure 5, we can observe that the number of singular values satisfying (4.14) is less than r along the whole process. This confirms the *oversolving* phenomena noticed for the Gaussian matrices.

The experiments offer different insights on the behavior of alternating projection approaches and of iRAPM. First, all the considered methods manage to approximate the sought matrix with good accuracy and the error on the known entries steadily decreases along the iterations. Second, it is evident the reduction in the overall computational cost provided by iRAPM. Remarkably, this computational saving is also obtained on the realistic case of the `lake` image.

6 Conclusions and perspectives

We have introduced an inexact regularized alternating projection method for computing a point in the intersection of two nonconvex sets. Inexact projections on one of the two sets are allowed. The convergence properties of the methods have been proved in a general setting and then specialized to affine rank minimization problems. For the latter class of problems the inexact projection on the rank level sets amounts to compute an approximate r-truncated SVD. Suitable accuracy requirements and corresponding stopping criteria for the Krylov inner solver have been proposed. Numerical results on

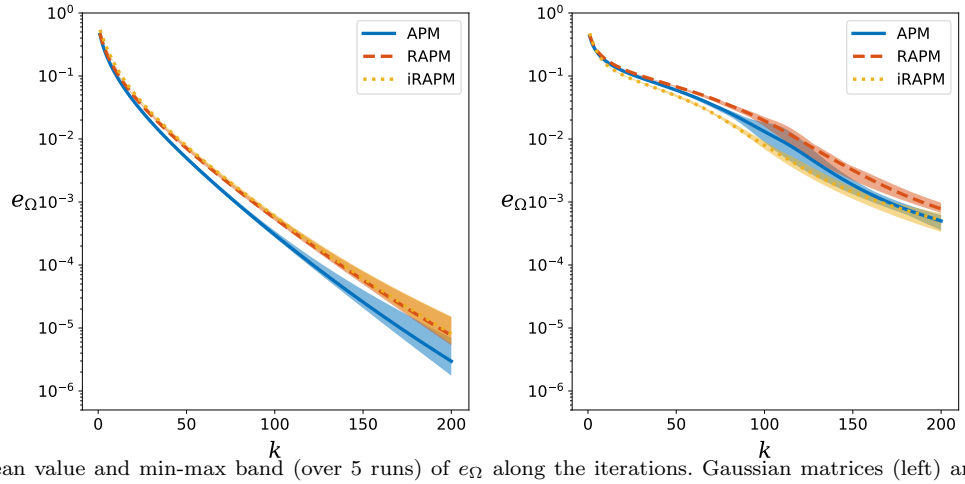

Fig. 3 Mean value and min-max band (over 5 runs) of e_Ω along the iterations. Gaussian matrices (left) and truncated `lake` (right).

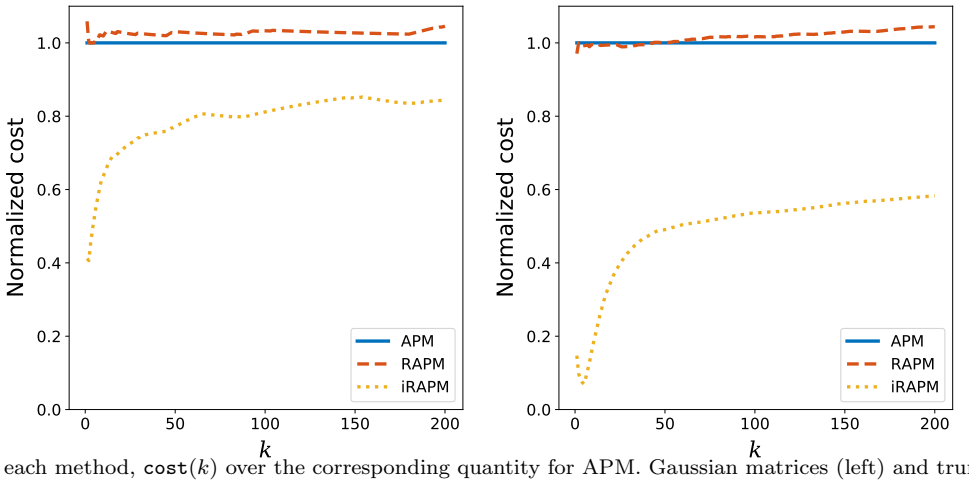

Fig. 4 For each method, `cost`(k) over the corresponding quantity for APM. Gaussian matrices (left) and truncated `lake` (right).

matrix completion problems showed that the inexact method is in general less expensive than the exact counterpart. We observed that the savings in computational cost are concentrated in the first phase of the convergence process. Then, it could be of interest to devise an adaptive strategy for choosing the parameter ζ that rules the inexactness level. A further extension is to devise adaptive strategies for the rank selection when the rank is unknown.

Acknowledgements. The authors are all members of the INdAM research group GNCS, which is kindly acknowledged.

Author Contributions. All authors contributed to the study conception and design, and the theoretical content of the paper. Code implementation, data collection and analysis were performed by MS. All authors read and approved the final manuscript.

Funding. The research that led to the present paper was partially supported by INDAM-GNCS through Progetti di Ricerca 2025 (CUP_E53C24001950001) and by PNRR - Missione 4 Istruzione e Ricerca - Componente C2 Investimento 1.1, Fondo per il Programma Nazionale di Ricerca e Progetti di Rilevante Interesse Nazionale (PRIN) funded by the European Commission under the NextGeneration EU programme, project "Inverse problems in PDE: theoretical and numerical analysis", code: 2022B32J5C, MUR D.D. financing decree n. 973 of 30th June 2023 (CUP B53D23009200006), project "Advanced optimization METhods for automated central veIn Sign detection in multiple sclerosis from magneTic resonAnce imaging (AMETISTA)", code: P2022J9SNP, MUR D.D. financing decree n. 1379 of 1st September 2023 (CUP E53D23017980001), project "Numerical Optimization with Adaptive Accuracy and Applications to Machine Learning", code: 2022N3ZNAX MUR D.D. financing decree n. 973

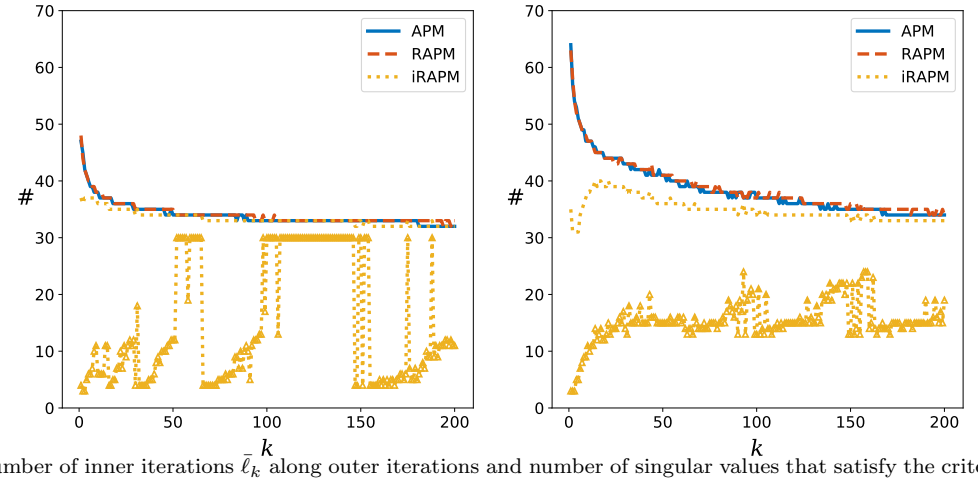

Fig. 5 Number of inner iterations $\bar{\ell}_k$ along outer iterations and number of singular values that satisfy the criterion (4.14). Gaussian matrices (left) and truncated `lake` (right).

of 30th June 2023 (CUP B53D23012670006), and by Partenariato esteso FAIR "Future Artificial Intelligence Research" SPOKE 1 Human-Centered AI. Obiettivo 4, Project "Mathematical and Physical approaches to innovative Machine Learning technologies (MaPLe)".

Data availability statement. The `lake` image can be downloaded from http://www.imageprocessingplace.com

Declarations

Conflict of interest. All authors certify that they have no affiliations with or involvement in any organization or entity with any financial interest or non-financial interest in the subject matter or materials discussed in this manuscript. The authors have no competing interests to declare that are relevant to the content of this article.

Ethics approval. The authors declare that research ethics approval was not required for this study.

Informed consent. The authors declare that informed consents were not required for this study.

Open Access. This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

Appendix A Additional proofs

A.1 Proof of Lemma 3.1

If $y^k \in \mathcal{P}_B(y^k_{reg})$, one can always set $w^\ell = y^k$ and $c^\ell = \|y^k - y^k_{reg}\|^2$ for any $\ell \in \mathbb{N}$, so that conditions (3.8a)-(3.8c) and (3.9) are trivially satisfied for all $\ell \in \mathbb{N}$ with $\hat{y}^{k+1} = y^k$. Then, without loss of generality, we assume that $y^k \notin \mathcal{P}_B(y^k_{reg})$. Hence, there holds $\|\hat{y}^{k+1} - y^k_{reg}\|^2 - \|y^k - y^k_{reg}\|^2 < 0$. First, we prove that there exists $\ell_\epsilon \in \mathbb{N}$ such that

$$\zeta(c^\ell - \|y^k - y^k_{reg}\|^2) - (\|\hat{y}^{k+1} - y^k_{reg}\|^2 - \|y^k - y^k_{reg}\|^2) > 0 \quad \forall\, \ell \geq \ell_\epsilon. \tag{A1}$$

Indeed, by (3.8b), we have

$$\forall\, \varepsilon > 0, \exists\, \ell_\varepsilon > 0 \,:\, c^\ell - \|\hat{y}^{k+1} - y^k_{reg}\|^2 > -\varepsilon, \quad \forall\, \ell \geq \ell_\varepsilon.$$

Setting $\varepsilon = \frac{1-\zeta}{\zeta}(\|y^k - y^k_{reg}\|^2 - \|\hat{y}^{k+1} - y^k_{reg}\|^2) > 0$, we have that there exists $\ell_\epsilon \in \mathbb{N}$ such that

$$\zeta(c^\ell - \|\hat{y}^{k+1} - y^k_{reg}\|^2) > -(1-\zeta)(\|y^k - y^k_{reg}\|^2 - \|\hat{y}^{k+1} - y^k_{reg}\|^2)$$

that yields (A1). Now, we show that there exists $\hat{\ell} \geq \ell_\epsilon$ such that (3.9) holds. Indeed, it holds

$$\|w^\ell - y^k_{reg}\|^2 = \|w^\ell - \hat{y}^{k+1}\|^2 + \|\hat{y}^{k+1} - y^k_{reg}\|^2 + 2\langle w^\ell - \hat{y}^{k+1}, \hat{y}^{k+1} - y^k_{reg}\rangle, \tag{A2}$$

and we observe that there exists $\hat{\ell} \geq \ell_\epsilon$ such that, for all $\ell \geq \hat{\ell}$, it holds

$$\begin{aligned}\|w^\ell - \hat{y}^{k+1}\|^2 + 2\langle w^\ell - \hat{y}^{k+1}, \hat{y}^{k+1} - y^k_{reg}\rangle &\leq \zeta(c^{\hat{\ell}} - \|y^k - y^k_{reg}\|^2) - \left(\|\hat{y}^{k+1} - y^k_{reg}\|^2 - \|y^k - y^k_{reg}\|^2\right) \\ &\leq \zeta(c^{\ell} - \|y^k - y^k_{reg}\|^2) - \left(\|\hat{y}^{k+1} - y^k_{reg}\|^2 - \|y^k - y^k_{reg}\|^2\right),\end{aligned} \tag{A3}$$

where the former inequality holds given that the left-hand side is converging to zero (see (3.8c)) and the right-hand side is positive (see (A1)), whereas the latter inequality is a consequence of (3.8a)-(3.8b) that ensures $c^{\hat{\ell}} < c^\ell$. By plugging (A3) into (A2) we obtain (3.9). Applying (3.8a) to (3.9) yields

$$\|w^\ell - y^k_{reg}\|^2 \leq \zeta\|\hat{y}^{k+1} - y^k_{reg}\|^2 + (1-\zeta)\|y^k - y^k_{reg}\|^2,$$

which is equivalent to condition (3.4) with $y^{k+1} = w^\ell$.

A.2 Proof of Lemma 3.3

Item (i). We recall that $Q(y^{k+1}) \leq 0$ due to equations (3.3) and (3.4). Then, from the definition of $\{d_k\}_{k\in\mathbb{N}}$ in (3.14), we obtain the thesis.

Item (ii). We observe that the inexactness condition (3.4) can be reformulated by means of the following equivalences

$$\begin{aligned}&Q^{k+1}(y^{k+1}) \leq \zeta Q^{k+1}(\hat{y}^{k+1}) \Leftrightarrow Q^{k+1}(y^{k+1}) \leq Q^{k+1}(\hat{y}^{k+1}) - \frac{1-\zeta}{\zeta}Q^{k+1}(y^{k+1}) \\ &\Leftrightarrow \frac{\|y^{k+1} - y^k\|^2}{2\mu_k} + L(x^{k+1}, y^{k+1}) \leq \frac{\|\hat{y}^{k+1} - y^k\|^2}{2\mu_k} + L(x^{k+1}, \hat{y}^{k+1}) - \frac{1-\zeta}{\zeta}Q^{k+1}(y^{k+1}).\end{aligned} \tag{A4}$$

Since $\hat{y}^{k+1} \in \mathcal{P}_B((y^k + \mu_k x^{k+1})/(1+\mu_k))$, we have $\hat{y}^{k+1} \in \operatorname{prox}_{\mu_k L(x^{k+1},\cdot)}(y^k)$, which yields

$$\frac{1}{2\mu_k}\|\hat{y}^{k+1} - y^k\|^2 + L(x^{k+1}, \hat{y}^{k+1}) \leq L(x^{k+1}, y^k). \tag{A5}$$

By applying the previous inequality to (A4), we obtain

$$\frac{1}{2\mu_k}\|y^{k+1} - y^k\|^2 + L(x^{k+1}, y^{k+1}) \leq L(x^{k+1}, y^k) - \frac{1-\zeta}{\zeta}Q^{k+1}(y^{k+1}). \tag{A6}$$

Writing $Q^{k+1}(y^{k+1})$ as in (3.1) and rearranging the terms of (A6) yields

$$\frac{1}{2\mu_k}\left(1 + \frac{1}{\zeta}\right)\|y^{k+1} - y^k\|^2 - Q^{k+1}(y^{k+1}) + \left(1 + \frac{1}{\zeta}\right)L(x^{k+1}, y^{k+1}) \leq \left(1 + \frac{1}{\zeta}\right)L(x^{k+1}, y^k).$$

By dividing both terms of the previous inequality by a factor $1 + 1/\zeta$ and summing the term $\frac{1}{2\lambda_k}\|x^{k+1} - x^k\|^2$, we get

$$\begin{aligned}&\frac{1}{2\mu_k}\|y^{k+1} - y^k\|^2 + \frac{1}{2\lambda_k}\|x^{k+1} - x^k\|^2 - \left(\frac{\zeta}{\zeta+1}\right)Q^{k+1}(y^{k+1}) + L(x^{k+1}, y^{k+1}) \\ &\qquad \leq \frac{1}{2\lambda_k}\|x^{k+1} - x^k\|^2 + L(x^{k+1}, y^k) \leq L(x^k, y^k),\end{aligned}$$

where the second inequality follows from $x^{k+1} \in \operatorname{prox}_{\lambda_k L(\cdot,y^k)}(x^k)$ and the definition of proximal operator. By observing that $\lambda_k \leq \lambda_+$ and $\mu_k \leq \mu_+$, we arrive to

$$\frac{1}{2\mu_+}\|y^{k+1}-y^k\|^2+\frac{1}{2\lambda_+}\|x^{k+1}-x^k\|^2-\left(\frac{\zeta}{\zeta+1}\right)Q^{k+1}(y^{k+1})+L(x^{k+1},y^{k+1})\leq L(x^k,y^k). \tag{A7}$$

Then the thesis of item (ii) follows from (A7) by setting $a = \min\{1/(2\mu_+), 1/(2\lambda_+), \zeta/(\zeta+1)\}$ and recalling the definition of d_k in (3.14).

Item (iii). The left-hand side of (3.18) follows immediately from (A4) by recalling the definition of F in (3.13) and $\{\rho_k\}_{k\in\mathbb{N}}$ in (3.16). Regarding the right-hand side, we sum γ_k given in (3.15) to both sides of (A5), thus obtaining

$$\begin{aligned}F(x^{k+1},\hat{y}^{k+1},\rho_{k+1}) &\leq \frac{1}{2\mu_k}\|\hat{y}^{k+1}-y^k\|^2+L(x^{k+1},\hat{y}^{k+1})+\gamma_k\\ &\leq L(x^{k+1},y^k)+\gamma_k \leq L(x^{k+1},y^k)+\frac{1}{2\lambda_k}\|x^{k+1}-x^k\|^2+\gamma_k \leq L(x^k,y^k)+\gamma_k,\end{aligned}$$

where the first inequality follows from $Q^{k+1}(\hat{y}^{k+1}) = \min_{y\in\mathbb{R}^n} Q^{k+1}(y) \leq Q^{k+1}(y^{k+1})$, and the last one holds by $x^{k+1} \in \operatorname{prox}_{\lambda_k L(\cdot,y^k)}(x^k)$ and the definition of proximal operator. We observe that the sequence $\{\gamma_k\}_{k\in\mathbb{N}}$ is non-negative due to (3.3). Furthermore, by summing (A7) for $k = 0,\dots,K$, it follows that

$$\frac{1}{2\mu_+}\sum_{k=0}^{K}\|y^{k+1}-y^k\|^2+\frac{1}{2\lambda_+}\sum_{k=0}^{K}\|x^{k+1}-x^k\|^2+\left(\frac{\zeta}{\zeta+1}\right)\sum_{k=0}^{K}-Q^{k+1}(y^{k+1})\leq L(x^0,y^0)-L(x^{K+1},y^{K+1}).$$

By taking the limit over $K\to\infty$ and recalling the definition of a, we get

$$a\sum_{k=0}^{\infty}\|x^{k+1}-x^k\|^2+\|y^{k+1}-y^k\|^2-Q^{k+1}(y^{k+1})<\infty. \tag{A8}$$

As a result, the sequence $\{-Q^{k+1}(y^{k+1})\}_{k\in\mathbb{N}}$ is summable and, thanks to the inexactness condition (3.4), we have that $\{-Q^{k+1}(\hat{y}^{k+1})\}_{k\in\mathbb{N}}$ is also summable. Then, (3.19) holds due to condition (3.5), and $\{\gamma_k\}_{k\in\mathbb{N}}$ is summable, which yields $\lim_{k\to\infty}\gamma_k = 0$. Furthermore, it holds $\lim_{k\to\infty}\|y^{k+1}-y^k\| = 0$ by (A8). Then, by applying the triangular inequality, we conclude that $\lim_{k\to\infty}\|\hat{y}^{k+1}-y^k\| = 0$ and consequently $\lim_{k\to\infty}\rho_k = 0$. Hence, the thesis of item (iii) follows.

Item (iv). Since the subdifferential of the sum of two functions, one of which is continuously differentiable, is given by the sum of the two subdifferentials [27, Exercise 8.8], it is possible to write the following equivalences

$$\begin{aligned}x^{k+1}\in\operatorname{prox}_{\lambda_k L(\cdot,y^k)}(x^k) &\Leftrightarrow 0\in(x^{k+1}-y^k)+\partial\iota_A(x^{k+1})+\frac{1}{\lambda_k}(x^{k+1}-x^k),\\ \hat{y}^{k+1}\in\operatorname{prox}_{\mu_k L(x^{k+1},\cdot)} &\Leftrightarrow 0\in-(x^{k+1}-\hat{y}^{k+1})+\partial\iota_B(\hat{y}^{k+1})+\frac{1}{\mu_k}(\hat{y}^{k+1}-y^k).\end{aligned}$$

Hence there exist $w_1^{k+1}\in\partial\iota_A(x^{k+1})$ and $w_2^{k+1}\in\partial\iota_B(\hat{y}^{k+1})$ such that

$$w_1^{k+1}=-(x^{k+1}-y^k)-\frac{1}{\lambda_k}(x^{k+1}-x^k),\quad w_2^{k+1}=(x^{k+1}-\hat{y}^{k+1})-\frac{1}{\mu_k}(\hat{y}^{k+1}-y^k).$$

The subdifferential of L at any point (x,y) writes as $\partial L(x,y) = \{x-y+\partial\iota_A(x)\}\times\{y-x+\partial\iota_B(y)\}$, see e.g. [12, Proposition 2.1]. Then, if we define the vectors

$$\begin{aligned}v_1^{k+1}&=x^{k+1}-\hat{y}^{k+1}+w_1^{k+1}=-\frac{1}{\lambda_k}(x^{k+1}-x^k)-(\hat{y}^{k+1}-y^k),\\ v_2^{k+1}&=\hat{y}^{k+1}-x^{k+1}+w_2^{k+1}=-\frac{1}{\mu_k}(\hat{y}^{k+1}-y^k),\end{aligned}$$

the vector $v^{k+1} = (v_1^{k+1}, v_2^{k+1})$ belongs to the subdifferential of L at $(x^{k+1}, \hat{y}^{k+1})$. By setting $\kappa_- = \min\{\lambda_-, \mu_-\}$, the norm of v^{k+1} can be upper bounded by proceeding as follows

$$\begin{aligned}
\|v^{k+1}\| &= \left\| \left(-\frac{1}{\lambda_k}(x^{k+1} - x^k) - (\hat{y}^{k+1} - y^k), -\frac{1}{\mu_k}(\hat{y}^{k+1} - y^k) \right) \right\| \\
&= \left\| \left(-\frac{1}{\lambda_k}(x^{k+1} - x^k), -\frac{1}{\mu_k}(y^{k+1} - y^k) \right) + \left(-(\hat{y}^{k+1} - y^k), -\frac{1}{\mu_k}(\hat{y}^{k+1} - y^{k+1}) \right) \right\| \\
&\leq \frac{1}{\kappa_-} \left\| (x^{k+1} - x^k, y^{k+1} - y^k) \right\| + \left\| \left(-(\hat{y}^{k+1} - y^{k+1}), -\frac{1}{\mu_k}(\hat{y}^{k+1} - y^{k+1}) \right) \right\| \\
&+ \left\| -\left(y^{k+1} - y^k, 0\right) \right\| \leq \left(1 + \frac{1}{\kappa_-}\right) \|(x^{k+1} - x^k, y^{k+1} - y^k)\| + \sqrt{1 + \frac{1}{\kappa_-^2}} \|\hat{y}^{k+1} - y^{k+1}\| \\
&\leq \left(1 + \frac{1}{\kappa_-}\right) \|(x^{k+1} - x^k, y^{k+1} - y^k)\| + \sqrt{-\left(1 + \frac{1}{\kappa_-^2}\right)\left(\frac{1-\zeta}{\zeta}\right) Q^{k+1}(y^{k+1})}, \qquad \text{(A9)}
\end{aligned}$$

where the last inequality follows from condition (3.5). Finally, we observe that the subdifferential of the function F defined in (3.13) can be written as $\partial F(x, y, \rho) = \partial L(x, y) \times \{\rho\}$. Hence, since the vector $v^{k+1} \in \mathbb{R}^n$ belongs to the subdifferential $\partial L(x^{k+1}, \hat{y}^{k+1})$, we have that $(v^{k+1}, \rho^{k+1}) \in \partial F(x^{k+1}, \hat{y}^{k+1}, \rho^{k+1})$, and the norm of such subgradient can be upper bounded as

$$\begin{aligned}
&\|(v^{k+1}, \rho^{k+1})\| \leq \|v^{k+1}\| + |\rho^{k+1}| \leq \left(1 + \frac{1}{\kappa_-}\right) \|(x^{k+1} - x^k, y^{k+1} - y^k)\| \\
&+ \sqrt{-\left(1 + \frac{1}{\kappa_-^2}\right)\left(\frac{1-\zeta}{\zeta}\right) Q^{k+1}(y^{k+1})} + \sqrt{\frac{1}{\mu_k}} \|\hat{y}^{k+1} - y^k\| + \sqrt{-2\left(\frac{1-\zeta}{\zeta}\right) Q^{k+1}(y^{k+1})} \\
&\leq \left(1 + \frac{1}{\kappa_-} + \frac{1}{\sqrt{\kappa_-}}\right) \|(x^{k+1} - x^k, y^{k+1} - y^k)\| \\
&+ \sqrt{\frac{1-\zeta}{\zeta}} \left(\sqrt{\frac{1}{\kappa_-}} + \sqrt{1 + \frac{1}{\kappa_-^2}} + \sqrt{2} \right) \sqrt{-Q^{k+1}(y^{k+1})},
\end{aligned}$$

where the second inequality is due to (A9) and the definition of $\{\rho_k\}_{k\in\mathbb{N}}$ in (3.16), whereas the third one follows by employing the triangular inequality and the inexactness condition (3.5). Then, the thesis of item (iv) is obtained by applying (3.17) to the previous inequality.

References

[1] Neumann, J.: Functional Operators. Annals of Mathematics Studies, vol. 22. Princeton University Press, Princeton, NJ (2010)

[2] Bauschke, H.H., Borwein, J.M.: On projection algorithms for solving convex feasibility problems. SIAM Rev. **38**(3), 367–426 (1996)

[3] Bauschke, H.H., Borwein, J.M.: On the convergence of von Neumann's alternating projection algorithm for two sets. Set-Valued Anal. **1**, 185–212 (1993)

[4] Bauschke, H.H., Borwein, J.M., Lewis, A.S.: The method of cyclic projections for closed convex sets in Hilbert space. Contemp. Math. **204**, 1–38 (1997)

[5] Escalante, R., Raydan, M.: Alternating Projection Methods. Society for Industrial and Applied Mathematics, Philadelphia, PA (2011). https://doi.org/10.1137/9781611971941 . https://epubs.siam.org/doi/abs/10.1137/9781611971941

[6] Behling, R., Bello-Cruz, Y., Santos, L.R.: Infeasibility and error bound imply finite convergence of alternating projections. SIAM J. Optim. **31**(4), 2863–2892 (2021) https://doi.org/10.1137/20M1358669 https://doi.org/10.1137/20M1358669

[7] Drusvyatskiy, D.: Slope and Geometry in Variational Mathematics. Ph.D. Thesis **47**, 669–681 (2013)

[8] Drusvyatskiy, D., Ioffe, A.D., Lewis, A.S.: Transversality and alternating projections for nonconvex sets. Found. Comput. Math. **15**(6), 1637–1651 (2015)

[9] Lewis, A.S., Malick, J.: Alternating projections on manifolds. Math. Oper. Res. **33**(1), 216–234 (2008)

[10] Lewis, A.S., Luke, D.R., Malick, J.: Local linear convergence of alternating and averaged nonconvex projections. Found. Comput. Math. **9**(4), 485–513 (2009)

[11] Noll, D., Rondepierre, A.: On Local Convergence of the Method of Alternating Projections. Found. Comput. Math. **16**, 425–455 (2016)

[12] Attouch, H., Bolte, J., Redont, P., Soubeyran, A.: Proximal alternating minimization and projection methods for nonconvex problems: An approach based on the Kurdyka-Łojasiewicz inequality. Math. Oper. Res. **35**(2), 438–457 (2010)

[13] Birgin, E.G., Martínez, J.M., Raydan, M.: Inexact spectral projected gradient methods on convex sets. IMA J. Numer. Anal. **23**(4), 539–559 (2003)

[14] Ferreira, O.P., Lemes, M., Prudente, L.F.: On the inexact scaled gradient projection method. Comput. Optim. Appl. **81**, 91–125 (2022)

[15] Bonettini, S., Ochs, P., Prato, M., Rebegoldi, S.: An abstract convergence framework with application to inertial inexact forward–backward methods. Comput. Optim. Appl. **84**, 319–362 (2023)

[16] Recht, B., Fazel, M., Parrilo, P.A.: Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization. SIAM Rev. **52**(3), 471–501 (2010) https://doi.org/10.1137/070697835 https://doi.org/10.1137/070697835

[17] Saad, Y.: Numerical Methods for Large Eigenvalue Problems: Revised Edition. SIAM, ??? (2011)

[18] Kruger, A.Y., Thao, N.H.: Regularity of collections of sets and convergence of inexact alternating projections. J. Convex Anal. **23**, 823–847 (2016)

[19] Díaz Millán, R., Ferreira, O.P., Prudente, L.F.: Alternating conditional gradient method for convex feasibility problems. Comput. Optim. Appl. **80**, 245–269 (2021)

[20] Drusvyatskiy, D., Lewis, A.S.: Local linear convergence for inexact alternating projections on nonconvex sets. Vietnam J. Math. **47**, 669–681 (2019)

[21] Luke, D.R.: Prox-regularity of rank constraint sets and implications for algorithms. J. Math. Imaging Vis. **47**, 231–238 (2013)

[22] Bonettini, S., Prato, M., Rebegoldi, S.: Convergence of inexact forward-backward algorithms using the forward-backward envelope. SIAM J. Optim. **30**(4), 3069–3097 (2020)

[23] Bonettini, S., Prato, M., Rebegoldi, S.: New convergence results for the inexact variable metric forward-backward method. Appl. Math. Comput. **392**, 125719 (2021)

[24] Guerreiro, R.F.C., Aguiar, P.M.Q.: Estimation of rank deficient matrices from partial observations: two-step iterative algorithms. In: Int. Workshop on Energy Minimization Methods in Comp. Vision and Pattern Recognition, pp. 450–466 (2003). Springer

[25] Wang, H., Cai, J.F., Wang, T., Wei, K.: Fast Cadzow's algorithm and a gradient variant. J. Sci. Comput. **88**(2), 41 (2021)

[26] Bolte, J., Daniilidis, A., Lewis, A.: The Łojasiewicz inequality for nonsmooth subanalytic functions

with applications to subgradient dynamical systems. SIAM J. Optim. **17**(4), 1205–1223 (2007)

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

[28] Bauschke, H.H., Noll, D.: On cluster points of alternating projections. Serdica Math. J. **39**, 355–364 (2013)

[29] Attouch, H., Redont, P., Soubeyran, A.: A new class of alternating proximal minimization algorithms with costs-to-move. SIAM J. Optim. **18**(3), 1061–1081 (2007) https://doi.org/10.1137/060657248

[30] Bolte, J., Sabach, S., Teboulle, M.: Proximal alternating linearized minimization for nonconvex and nonsmooth problems. Math. Program. **146**(1–2), 459–494 (2014)

[31] Frankel, P., Garrigos, G., Peypouquet, J.: Splitting methods with variable metric for Kurdyka–Łojasiewicz functions and general convergence rates. J. Opt. Theory Appl. **165**, 874–900 (2015)

[32] Attouch, H., Bolte, J.: On the convergence of the proximal algorithm for nonsmooth functions involving analytic features. Math. Program. **116**, 5–16 (2009)

[33] Candes, E., Recht, B.: Exact matrix completion via convex optimization. Found. Comput. Math. **9**, 717–772 (2009)

[34] Golub, G.H., Van Loan, C.F.: Matrix Computations. JHU press, ??? (2013)

[35] Yu, P., Li, G., Pong, T.K.: Kurdyka–łojasiewicz Exponent via Inf-projection. Found. Comput. Math. **22**, 1171–1217 (2022)

[36] Li, G., Pong, T.K.: Calculus of the Exponent of Kurdyka–łojasiewicz Inequality and Its Application to Linear Convergence of First-Order Methods. Found. Comput. Math. **18**, 1199–1232 (2018)

[37] Larsen, R.M.: Lanczos bidiagonalization with partial reorthogonalization. Dept. Computer Science, Aarhus University, Technical report, DAIMI **PB**(537) (1998)

[38] Paige, C.C., Saunders, M.A.: Lsqr: An algorithm for sparse linear equations and sparse least squares. ACM Trans. Math. Softw. **8**(1), 43–71 (1982)

[39] Simon, H.D., Zha, H.: Low-rank matrix approximation using the Lanczos bidiagonalization process with applications. SIAM J. Sci. Comput. **21**(6), 2257–2274 (2000) https://doi.org/10.1137/s1064827597327309

[40] Ramlatchan, A., Yang, M., Liu, Q., Li, M., Wang, J., Li, Y.: A survey of matrix completion methods for recommendation systems. Big Data Min. Anal. **1**(4), 308–323 (2018)

[41] Nguyen, L.T., Kim, J., Shim, B.: Low-rank matrix completion: A contemporary survey. IEEE Access **7**, 94215–94237 (2019)

[42] Agarwal, A., Amjad, M.J., Shah, D., Shen, D.: Model agnostic time series analysis via matrix estimation. In: Proceedings of the ACM on Measurement and Analysis of Computing Systems, vol. 2, pp. 1–39. ACM New York, NY, USA, ??? (2018)

[43] Silei, M., Bellavia, S., Superchi, F., Bianchini, A.: Recovering corrupted data in wind farm measurements: A matrix completion approach. Energies **16**(4) (2023) https://doi.org/10.3390/en16041674

[44] Bellavia, S., Gondzio, J., Porcelli, M.: A relaxed interior point method for low-rank semidefinite programming problems with applications to matrix completion. J. Sci. Comput. **89**(2), 46 (2021)