LINEAR CONVERGENCE OF A PROXIMAL ALTERNATING MINIMIZATION METHOD WITH EXTRAPOLATION FOR $\ell_1$-NORM PRINCIPAL COMPONENT ANALYSIS*

PENG WANG†, HUIKANG LIU‡, AND ANTHONY MAN-CHO SO§

Abstract. A popular robust alternative of the classic principal component analysis (PCA) is the $\ell_1$-norm PCA (L1-PCA), which aims to find a subspace that captures the most variation in a dataset as measured by the $\ell_1$-norm. L1-PCA has shown great promise in alleviating the effect of outliers in data analytic applications. However, it gives rise to a challenging non-smooth non-convex optimization problem, for which existing algorithms are either not scalable or lack strong theoretical guarantees on their convergence behavior. In this paper, we propose a proximal alternating minimization method with extrapolation (PAMe) for solving a two-block reformulation of the L1-PCA problem. We then show that for both the L1-PCA problem and its two-block reformulation, the Kurdyka-Lojasiewicz exponent at any of the limiting critical points is $1/2$. This allows us to establish the linear convergence of the sequence of iterates generated by PAMe and to determine the criticality of the limit of the sequence with respect to both the L1-PCA problem and its two-block reformulation. To complement our theoretical development, we show via numerical experiments on both synthetic and real-world datasets that PAMe is competitive with a host of existing methods. Our results not only significantly advance the convergence theory of iterative methods for L1-PCA but also demonstrate the potential of our proposed method in applications.

Key words. L1-PCA, Kurdyka-Lojasiewicz exponent, proximal alternating minimization, extrapolation, linear convergence

AMS subject classifications. 49J52, 58C05, 58C20, 90C30

1. Introduction. Dimension reduction is a powerful paradigm for facilitating information extraction from large datasets. Among the many existing dimension reduction techniques, perhaps the most classic and widely used one is principal component analysis (PCA), which aims to identify a low-dimensional subspace that captures the most variation in the dataset. Concretely, let $X = [x_1, \ldots, x_n] \in \mathbb{R}^{d \times n}$ be the data matrix, where $n$ and $d$ denote the number of samples and dimension of the data points, respectively. Suppose that the sample mean of the dataset $\{x_1, \ldots, x_n\}$ is zero. Then, a common formulation for finding the desired low-dimensional subspace is given by

$$\max_{Q \in \text{St}(d,K)} \|X^TQ\|_F,$$

where $K$ is the dimension of the subspace with $K \leq \min\{n, d\}$ and $\text{St}(d, K) = \{Q \in \mathbb{R}^{d \times K} : Q^TQ = I_K\}$ is the compact Stiefel manifold with $I_K$ being the $K \times K$ identity matrix.


Funding: The first author was supported in part by the ARO YIP award W911NF1910027 and in part by the NSF CAREER award CCF-1845076. The third author was supported in part by the Hong Kong RGC GRF project CUHK 14205421.

†Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, MI 48109 USA (pengwa@umich.edu).

‡School of Information Management and Engineering, Shanghai University of Finance and Economics, Shanghai, China (liuhuikang@shufe.edu.cn).

§Department of Systems Engineering and Engineering Management, The Chinese University of Hong Kong, Shatin NT, Hong Kong SAR, China (manchoso@se.cuhk.edu.hk, http://www.se.cuhk.edu.hk/~manchoso/).
matrix [19]. Despite its non-convexity, Problem (1.1), which we shall refer to as L2-PCA, can be solved efficiently by computing the singular value decomposition (SVD) of the data matrix $X$. Moreover, the subspace spanned by the columns of an optimal solution to Problem (1.1) possesses many nice properties [18]. Nevertheless, it has long been known that L2-PCA is sensitive to corruptions in the dataset (generically referred to as outliers); see, e.g., [11]. This makes L2-PCA ill-suited for many contemporary applications, as the datasets are often collected by automated devices and can be prone to outliers. Over the years, there has been much effort in developing alternatives to L2-PCA that are robust against outliers; see, e.g., [22, 30] and the references therein. One approach is to replace the $\ell_2$- (Frobenius) norm in (1.1) with a suitable measure of dispersion in the dataset called scale function. Various scale functions and their statistical properties have been studied in the literature; see, e.g., [31] and the references therein. In particular, by taking the $\ell_1$-norm as the scale function, we obtain the following robust alternative to L2-PCA, which we shall refer to as L1-PCA:

$$(1.2) \max_{Q \in \mathbb{S}(d,K)} \|X^TQ\|_1.$$ 

Here, $\|A\|_1 = \sum_{i,j} |A_{ij}|$ denotes the $\ell_1$-norm of the matrix $A$. Besides being of interest in its own right, L1-PCA is also related to other data analytic tools, such as independent component analysis [32] and linear discriminant analysis [33]. However, unlike L2-PCA, which can essentially be solved in polynomial time, L1-PCA gives rise to a challenging computational problem. Indeed, it is shown in [34] that Problem (1.2) is NP-hard even when $K = 1$. This motivates the development of numerically efficient algorithms for solving the L1-PCA problem.

Many of the earlier algorithms for L1-PCA, such as [5, 9, 10], are heuristic in nature. In particular, there is no guarantee that the outputs of these algorithms satisfy any optimality condition of Problem (1.2). Among the first algorithms for L1-PCA that come with theoretical guarantees are those proposed by Kwak [21] and Nie et al. [35], which are based on fixed-point (FP) iterations. The former applies to Problem (1.2) with $K = 1$, while the latter can handle general $K \geq 1$. The per-iteration computational costs of these two algorithms are bounded by $\mathcal{O}(ndK + dK^2)$, which is cheap in the practically relevant case where $K \ll \min\{n,d\}$. Moreover, it is shown that for both algorithms, the iterates generated have a limit point (i.e., subsequential convergence of the iterates) and every limit point satisfies certain first-order optimality condition of the problem. However, the convergence rates of the two algorithms remain unknown. Around the same time, McCoy and Tropp [34] studied a semidefinite relaxation (SDR) approach (see [27] for an overview) to solving Problem (1.2) when $K = 1$. It is shown that with high probability, the solution obtained via this approach will have an objective value that is at least $c \cdot \sqrt{2/\pi}$ times the optimal value for any fixed $c \in (0,1)$. However, standard interior-point method-based implementations of the SDR approach have a computational complexity of roughly $\mathcal{O}(n^{3.5})$, which renders the approach impractical when the dataset is large. Later, Markopoulos et al. [28] proposed an exact algorithm for solving Problem (1.2) that runs in $\mathcal{O}(n^{dK-K+1})$ time. Although this algorithm is impractical due to its high computational cost, it shows that Problem (1.2) is actually polynomial-time solvable when both $d$ and $K$ are fixed. Moreover, it can be used to benchmark the solution quality of different L1-PCA algorithms. In a follow-up work, Markopoulos et al. [29] developed an algorithm based on bit-flipping (BF) iterations for tackling Problem (1.2). On one hand, the computational cost of each BF iteration is $\mathcal{O}(ndK + nK^3)$, which is inferior to that of the FP iteration developed in [35] when $n \geq d$. On
the other hand, the algorithm based on BF iterations is guaranteed to converge in a finite number of steps, while that based on FP iterations is not known to possess such a property. Nevertheless, the number of BF iterations needed can be exponential in \( n \) and \( K \) in the worst case. Moreover, it is not clear whether the solution obtained from the BF iterations satisfies any optimality condition of Problem (1.2). Recently, Kim and Klabjan [20] revisited Problem (1.2) under the setting where \( K = 1 \) and proposed an algorithm similar to those in [21, 35] for tackling it. It is shown that the sequence of iterates generated by the algorithm will converge in a finite number of steps. This qualitatively improves upon the subsequential convergence results in [21, 35]. Moreover, by pretending that the objective function of (1.2) is smooth, it is claimed that the limit of the sequence is a local maximum of the problem. However, a rigorous proof of this claim is still missing.

1.1. Our Contributions. In view of the above discussion, our goal in this paper is to develop an iterative method for solving Problem (1.2) that is numerically efficient and has strong theoretical guarantees on its convergence behavior. To begin, observe that since \( |x| = \max\{x, -x\} \) for any \( x \in \mathbb{R} \), we can reformulate Problem (1.2) as

\[
(1.3) \min_{P \in B(n, K), Q \in \text{St}(d, K)} -\langle P, X^TQ \rangle,
\]

where \( \langle A, B \rangle = \text{tr}(A^TB) \) denotes the Euclidean inner product of two matrices \( A, B \) of the same dimensions and \( B(n, K) = \{ P \in \mathbb{R}^{n \times K} : P_{ij} \in \{\pm 1\}; \ i = 1, \ldots, n; \ j = 1, \ldots, K \} \) is the set of \( n \times K \pm 1 \) matrices. Noting that Problem (1.3) has two separate blocks of variables \( P \) and \( Q \), we can tackle it using the proximal alternating minimization (PAM) method [4, 2]. To achieve further speed-up, we equip the method with an extrapolation scheme, in which a point extrapolated from previous iterates of the block variable \( Q \) is used in the update of the block variable \( P \). It is worth noting that such a scheme differs from those developed for accelerating various proximal block coordinate descent-type methods (see, e.g., [37, 15, 16] and the references therein) and seems to be new. The resulting method, which we call proximal alternating minimization with extrapolation (PAMe), admits an efficient implementation, as the update of each block can essentially be given in closed form. In particular, it has a per-iteration computational cost of \( \mathcal{O}(ndK + dK^2) \), which is competitive with the methods based on FP iterations in [21, 35, 20]. Thus, PAMe is well suited to tackle large instances of Problem (1.2). Our numerical experiments on both synthetic and real-world datasets show that PAMe can be significantly faster than PAM and is competitive, in terms of both computational efficiency and solution quality, with a host of existing methods.

To shed light on the numerical performance of and obtain strong theoretical convergence guarantees for our proposed method PAMe, a key step is to characterize the growth behavior of the objective function of (1.3) around the limiting critical points (see Subsection 1.2 for the definition) of the problem. Towards that end, we first show that the Kurdyka-Lojasiewicz (KL) exponent at any limiting critical point of certain orthogonality constrained linear optimization (LO-OC) problem is 1/2. This result is new and complements that in [24] for (homogeneous) quadratic optimization with orthogonality constraint. Moreover, it implies, through a calculus rule established in [23], that the KL exponent at any limiting critical point of the original L1-PCA formulation (1.2) is 1/2. Then, we relate the limiting critical points of (1.3) to those of a particular instance of the LO-OC problem and show that the KL exponent at any of the former is also 1/2. With this characterization, we can utilize the analysis
framework in [2, 3] to establish the linear convergence of PAMe to a limit \((P^*, Q^*)\), which is a limiting critical point of Problem (1.3). Moreover, we show that the limit \(Q^*\) is a critical point (see Subsection 1.2 for the definition) of Problem (1.2) under certain conditions on the step sizes of PAMe. To the best of our knowledge, our work is the first to determine the KL exponent at the limiting critical points of both (1.2) and (1.3) and to present a first-order method that provably converges to a limiting critical point of (1.3) at a linear rate.

The rest of this paper is organized as follows. In Section 2, we introduce our proposed method PAMe and present the main results of this paper. We then prove the main results in Section 3 (concerning the KL exponent at the limiting critical points of Problems (1.2) and (1.3)) and Section 4 (concerning the convergence behavior of PAMe). In Section 5, we report the numerical performance of PAMe and other existing methods on both synthetic and real-world datasets. We end with some closing remarks in Section 6.

1.2. Notation and Definitions. In addition to the notation introduced earlier, we will use the following throughout the paper. Let \(O^n = \text{St}(n,n)\) denote the set of \(n \times n\) orthogonal matrices and \(S^n\) denote the set of \(n \times n\) symmetric matrices. For any \(a \in \mathbb{R}\), let

\[
\text{sgn}(a) = \begin{cases} 
\{a/|a|\}, & a \neq 0, \\
\{-1, 1\}, & a = 0 
\end{cases}
\]

denote (a variant of) the sign function that will be used to express the subdifferential of \(x \mapsto -|x|\). Given a matrix \(A\), let \(\text{sgn}(A)\) denote the matrix obtained by applying \(\text{sgn}()\) to each entry of \(A\); \(\|A\|_F\) and \(\|A\|\) denote the Frobenius norm and spectral norm of \(A\), respectively; \(\lambda_k(A)\) denote the \(k\)-th largest eigenvalue of \(A\) if \(A\) is symmetric. Given a vector \(x\), let \(\|x\|_2\) denotes the Euclidean norm of \(x\) and \(\text{Diag}(x)\) denote the diagonal matrix with \(x\) as its diagonal. Given square matrices \(Y_1, \ldots, Y_n\), let \(\text{BlkDiag}(Y_1, \ldots, Y_n)\) denote the block diagonal matrix with \(Y_1, \ldots, Y_n\) as its diagonal blocks.

Next, we introduce some concepts in non-smooth analysis that will be needed in our subsequent development. The details can be found in, e.g., [38]. For a non-empty closed set \(\mathcal{S} \subseteq \mathbb{R}^p\), the indicator function \(\delta_\mathcal{S} : \mathbb{R}^p \to \{0, +\infty\}\) associated with \(\mathcal{S}\) is defined as

\[
\delta_\mathcal{S}(x) = \begin{cases} 
0, & x \in \mathcal{S}, \\
+\infty, & \text{otherwise}; 
\end{cases}
\]

the projection onto \(\mathcal{S}\) is the set-valued mapping \(\text{Proj}_\mathcal{S} : \mathbb{R}^p \rightrightarrows \mathbb{R}^p\) given by \(\text{Proj}_\mathcal{S}(x) = \arg\min_{y \in \mathcal{S}} \|y-x\|_2\); the distance between \(\mathcal{S}\) and another non-empty closed set \(\mathcal{T} \subseteq \mathbb{R}^p\) is defined as \(\text{dist}(\mathcal{S}, \mathcal{T}) = \inf_{x \in \mathcal{S}, y \in \mathcal{T}} \|x-y\|_2\).

Let \(f : \mathbb{R}^p \to (-\infty, +\infty]\) be a given function. The domain of \(f\) is defined as \(\text{dom}(f) = \{x \in \mathbb{R}^p : f(x) < +\infty\}\). The function \(f\) is said to be proper if \(\text{dom}(f) \neq \emptyset\). A vector \(v \in \mathbb{R}^p\) is said to be a Fréchet subgradient of \(f\) at \(x \in \text{dom}(f)\) if

\[
(1.4) \quad \liminf_{y \to x} \frac{f(y) - f(x) - \langle v, y - x \rangle}{\|y - x\|_2} \geq 0.
\]

The set of vectors \(v \in \mathbb{R}^p\) satisfying (1.4) is called the Fréchet subdifferential of \(f\) at \(x \in \text{dom}(f)\) and denoted by \(\partial f(x)\). The limiting subdifferential, or simply the subdifferential, of \(f\) at \(x \in \text{dom}(f)\) is defined as

\[
\partial f(x) = \left\{ v \in \mathbb{R}^p : \exists x^k \to x, v^k \to v \text{ with } f(x^k) \to f(x), v^k \in \partial f(x^k) \right\}.
\]
By convention, if \( x \not\in \text{dom}(f) \), then \( \partial f(x) = \emptyset \). The domain of \( \partial f \) is defined as 
\[
\text{dom}(\partial f) = \{ x \in \mathbb{R}^p : \partial f(x) \neq \emptyset \}. 
\]
For the indicator function \( \delta_S : \mathbb{R}^p \rightarrow \{0, +\infty\} \) associated with the non-empty closed set \( S \subseteq \mathbb{R}^p \), we have
\[
\hat{\partial} \delta_S(x) = \left\{ v \in \mathbb{R}^p : \limsup_{y \to x, y \in S} \frac{\langle v, y - x \rangle}{\|y - x\|_2} \leq 0 \right\} \quad \text{and} \quad \partial \delta_S(x) = \mathcal{N}_S(x)
\]
for any \( x \in S \), where \( \mathcal{N}_S(x) \) is the normal cone to \( S \) at \( x \).

Now, suppose that the function \( f : \mathbb{R}^p \rightarrow (-\infty, +\infty] \) is proper and lower semi-continuous. A point \( x \in \mathbb{R}^p \) satisfying \( \partial f(x) = \emptyset \) is called a limiting critical point of \( f \). By the generalized Fermat rule (see, e.g., [38, Theorem 10.1]), a local minimizer of \( f \) is a limiting critical point of \( f \). The function \( f \) is said to have a KL exponent of \( \theta \in [0, 1) \) at the point \( \bar{x} \in \text{dom}(\partial f) \) if there exist constants \( \epsilon, \eta > 0, \nu \in (0, +\infty) \) such that
\[
dist(0, \partial f(x)) \geq \eta |f(x) - f(\bar{x})|^{\theta}
\]
whenever \( \|x - \bar{x}\| \leq \epsilon \) and \( f(\bar{x}) < f(x) < f(\bar{x}) + \nu \); see [23] and cf. [2, 3].

Upon writing Problem (1.2) as
\[
(1.5) \quad \min_{Q \in \mathbb{R}^{d \times K}} \{ \ell(Q) = -\|X^T Q\|_1 + \delta_{\text{St}(d,K)}(Q) \}
\]
and invoking the subdifferential calculus rules in [38, Chapter 10B], we see that every locally optimal solution \( Q \in \text{St}(d,K) \) to Problem (1.2) satisfies
\[
(1.6) \quad 0 \in -X \text{sgn}(X^T Q) + \mathcal{N}_{\text{St}(d,K)}(Q).
\]
A point \( Q \in \text{St}(d,K) \) satisfying (1.6) is called a critical point of \( \ell \). It should be noted that every limiting critical point of \( \ell \) is a critical point of \( \ell \), but the converse is not known to hold.

2. Main Results. As mentioned in Subsection 1.1, our strategy for tackling Problem (1.2) is to apply a proximal alternating minimization scheme to its two-block reformulation (1.3). Let us now formalize this strategy and introduce our proposed method PAME.

To begin, observe that Problem (1.3) can be written as
\[
(2.1) \quad \min_{P \in \mathbb{R}^{n \times K}, Q \in \mathbb{R}^{d \times K}} \left\{ h(P, Q) = -\langle P, X^T Q \rangle + \delta_{\text{St}(n,K)}(P) + \delta_{\text{St}(d,K)}(Q) \right\},
\]
which is in a form that is amenable to the PAM method developed in [4] (see also [2]). Given the current iterate \( (P^k, Q^k) \in \mathcal{B}(n,K) \times \text{St}(d,K) \), the method generates the next iterate \( (P^{k+1}, Q^{k+1}) \in \mathcal{B}(n,K) \times \text{St}(d,K) \) via
\[
(2.2) \quad P^{k+1} \in \arg \min_{P \in \mathcal{B}(n,K)} \left\{ -\langle P, X^T Q^k \rangle + \frac{\alpha_k}{2} \|P - P^k\|_F^2 \right\},
\]
\[
(2.3) \quad Q^{k+1} \in \arg \min_{Q \in \text{St}(d,K)} \left\{ -\langle P^{k+1}, X^T Q \rangle + \frac{\beta_k}{2} \|Q - Q^k\|_F^2 \right\},
\]
where \( \alpha_k, \beta_k > 0 \) are the step sizes. Motivated by the desire to accelerate the PAM iterations, we incorporate an extrapolation step when updating the block variable \( P \). Specifically, we replace (2.2) by
\[
E^k = Q^k + \gamma_k (Q^k - Q^{k-1}),
\]
\[
(2.4) \quad \min_{P \in \mathcal{B}(n,K)} \left\{ -\langle P, X^T E^k \rangle + \frac{\alpha_k}{2} \|P - P^k\|_F^2 \right\},
\]
where $E^k \in \mathbb{R}^{d \times K}$ is the point extrapolated from $Q^{k-1}$ and $Q^k$ and $\gamma_k \in [0, 1)$ is the extrapolation parameter. Now, note that both $P^{k+1}$ in (2.4) and $Q^{k+1}$ in (2.3) admit closed-form expressions. On one hand, it is easy to verify that

$$P^{k+1} = \text{sgn}(P^k + X^TE^k/\alpha_k).$$

On the other hand, the update (2.3) is an instance of the orthogonal Procrustes problem [40], whose solution is given by

$$Q^{k+1} = U^{k+1}V^{k+1T}.$$  

Here, $U^{k+1} \in \text{St}(d, K)$ and $V^{k+1} \in \mathbb{O}^K$ are obtained from a thin SVD $U^{k+1}\Sigma^{k+1}V^{k+1T}$ of $Q^k + XPK^{k+1}/\beta_k$. The above development leads to our proposed method PAMe, whose complete description can be found in Algorithm 2.1. Since the costs of implementing (2.5) and (2.6) are $O(ndK)$ and $O(dK^2)$, respectively, the per-iteration cost of Algorithm 2.1 is $O(ndK + dK^2)$, which is cheap when $K \ll \min\{n, d\}$.

---

**Algorithm 2.1** Proximal Alternating Minimization with Extrapolation (PAMe) for L1-PCA

1: **Input:** $X \in \mathbb{R}^{d \times n}$, $P^0 \in \mathbb{B}(n, K)$, $Q^{-1} = Q^0 \in \text{St}(d, K)$
2: for $k = 0, 1, 2, \ldots$ do
3: choose step sizes $\alpha_k, \beta_k > 0$ and extrapolation parameter $\gamma_k \in [0, 1)$;
4: set $E^k \leftarrow Q^k + \gamma_k(Q^k - Q^{k-1})$;
5: pick $P^{k+1} \in \text{sgn}(P^k + X^TE^k/\alpha_k)$;
6: compute a thin SVD $Q^k + XPK^{k+1}/\beta_k = U^{k+1}\Sigma^{k+1}V^{k+1T}$;
7: set $Q^{k+1} \leftarrow U^{k+1}V^{k+1T}$;
8: terminate if stopping criteria are met;
9: end for

Although at first sight the extrapolation scheme introduced above is similar to those used in various inertial proximal block coordinate descent-type methods (see, e.g., [45, 14, 37, 7, 15, 16, 36] and the references therein), there are two crucial differences. First, instead of performing an extrapolation step in each of block updates, PAMe performs such a step in only one of the block updates. Second, in most existing extrapolation schemes, each block update involves an extrapolation point that is obtained from previous iterates of that same block. By contrast, PAMe uses previous iterates of the block variable $Q$ to generate an extrapolation point for the update of the block variable $P$; see (2.4). As we shall see in Section 5, our proposed extrapolation scheme has better numerical performance than existing ones when tackling Problem (1.3). It is also interesting to note that the updates in PAMe are similar to those obtained when applying, in a formal manner, proximal difference-of-convex algorithms with extrapolation (see, e.g., [44, 26]) to Problem (1.5). For instance, the update of the block variable $Q$ in the method pDCAe developed in [44] amounts to projecting $E^k + X\xi^k/\beta_k$ with $\xi^k \in \text{sgn}(X^TQ^k)$ onto $\text{St}(d, K)$, while that in our proposed method PAMe amounts to projecting $Q^k + X\xi^k/\beta_k$ with $\xi^k \in \text{sgn}(P^k + X^TE^k/\alpha_k)$ onto $\text{St}(d, K)$ (see (2.3) and (2.5)). In fact, for the update in PAMe, we can take $\xi^k \in \text{sgn}(X^TE^k)$ when $\alpha_k > 0$ is sufficiently small. This further brings out the resemblance between the updates in pDCAe and PAMe. Nevertheless, since the objective function $\ell$ of Problem (1.5) is not of the difference-of-convex type (it involves the indicator function $\delta_{\text{St}(d, K)}$ of the non-convex set $\text{St}(d, K)$), existing analyses (such
as those in [26]) do not yield any guarantee on the convergence behavior of proximal difference-of-convex algorithms when applied to Problem (1.5). Moreover, we observe that PAMe outperforms pDCAe in our numerical experiments; see Section 5 for details.

Next, we present the main theoretical contributions of this paper. Our first result states that the objective function $\ell$ of Problem (1.5) (resp. $h$ of Problem (2.1)) has a KL exponent of $1/2$ at any of its limiting critical points. Combining this with the result in [23, Lemma 2.1], we conclude that $\ell$ (resp. $h$) has a KL exponent of $1/2$ at any $Q \in \text{dom}(\partial \ell)$ (resp. $(P, Q) \in \text{dom}(\partial h)$). This opens the possibility of determining the convergence rates of a host of iterative methods for solving Problems (1.5) and (2.1); see, e.g., [2, 3].

**Theorem 2.1.** Let $Q^* \in \text{St}(d, K)$ be a limiting critical point of Problem (1.5). Then, there exist $\epsilon_\ell \in (0, 1)$ and $\eta_\ell > 0$ such that for all $Q \in \text{St}(d, K)$ with $\|Q - Q^*\|_F \leq \epsilon_\ell$,

$$\text{dist}(0, \partial \ell(Q)) \geq \eta_\ell |\ell(Q) - \ell(Q^*)|^{1/2}. \quad (2.7)$$

Moreover, let $Z^* \in \mathcal{B}(n, K) \times \text{St}(d, K)$ be a limiting critical point of Problem (2.1). Then, there exist $\epsilon_h \in (0, 1)$ and $\eta_h > 0$ such that for all $Z \in \mathcal{B}(n, K) \times \text{St}(d, K)$ with $\|Z - Z^*\|_F \leq \epsilon_h$,

$$\text{dist}(0, \partial h(Z)) \geq \eta_h |h(Z) - h(Z^*)|^{1/2}. \quad (2.8)$$

Theorem 2.1 implies that we can take $\epsilon = \epsilon_\ell$, $\eta = \eta_\ell$, and $\nu = +\infty$ (resp. $\epsilon = \epsilon_h$, $\eta = \eta_h$, and $\nu = +\infty$) in the definition of the KL exponent of $\ell$ (resp. $h$) at $Q^*$ (resp. $Z^*$); see Subsection 1.2. We remark that the constants $\epsilon_\ell$, $\eta_\ell$ and $\epsilon_h$, $\eta_h$ can be determined explicitly; see Subsection 3.3.

With Theorem 2.1 at our disposal, we can study the convergence behavior of PAMe (Algorithm 2.1). Our second result has two parts. The first part states that with suitable choices of the parameters in PAMe, the iterates $\{(P^k, Q^k)\}_{k \geq 0}$ generated by the method will converge linearly to a limiting critical point $(P^*, Q^*)$ of Problem (2.1). Now, recall that our original problem of interest is (1.5). Although there is a correspondence between the optimal solutions to Problems (1.5) and (2.1), it is not clear whether such a correspondence carries over to the (limiting) critical points of the two problems. Thus, it is natural to ask whether $Q^*$ is a (limiting) critical point of Problem (1.5). Unfortunately, we do not yet know the answer to this question. The second part of our result, which provides a partial answer, gives a sufficient condition for $Q^*$ to be a critical point of Problem (1.5) (i.e., $Q^*$ satisfies (1.6)).

**Theorem 2.2.** Let $\{(P^k, Q^k)\}_{k \geq 0}$ be the sequence of iterates generated by Algorithm 2.1, where the step sizes $\{\alpha_k\}_{k \geq 0}$, $\{\beta_k\}_{k \geq 0}$ and extrapolation parameters $\{\gamma_k\}_{k \geq 0}$ satisfy (i) $\alpha_k \leq \alpha_k \leq \alpha^*$ for some $\alpha^* \in (0, +\infty)$, (ii) $3\beta_*/2 \leq \beta_k \leq \beta^*$ for some $\beta_*, \beta^* \in (0, +\infty)$, and (iii) $0 \leq \gamma_k < \gamma^* = \max\{1, \alpha_\nu \beta_*/2\|X\|^2\}$. Then, the sequence $\{(P^k, Q^k)\}_{k \geq 0}$ converges at least linearly to a limiting critical point $(P^*, Q^*)$ of Problem (2.1); i.e., there exist constants $c > 0$ and $\tau \in [0, 1)$ such that

$$\|\alpha_k \|_F \leq cr^k \quad \text{for } k = 0, 1, \ldots$$

Moreover, if $\alpha_k = \alpha_\nu$ for $k \geq 0$, then $Q^*$ is a solution to the following generalized equation:

$$0 \in -X \text{sgn}(P^* + X^T Q/\alpha_\nu) + \mathcal{N}_{\text{St}(d, K)}(Q). \quad (2.9)$$
In particular, if $\alpha_*$ satisfies
\begin{equation}
0 < \alpha_* < \min\{ |(X^TQ^*)_{ij}| : (X^TQ^*)_{ij} \neq 0; i = 1, \ldots, n; j = 1, \ldots, K\},
\end{equation}
then $Q^*$ is a critical point of Problem (1.5). Conversely, every critical point $Q$ of Problem (1.5) satisfying $0 \in -XP^* + N_{St(d,K)}(Q)$ and $P^* \in \text{sgn}(X^TQ)$ is a solution to the generalized equation (2.9), regardless of whether (2.10) holds.

Condition (2.10) suggests that PAMe is more likely to return a critical point of Problem (1.5) if we choose a smaller step size $\alpha_*$. In fact, we observe from our numerical experiments that a small $\alpha_*$ often leads to favorable performance of PAMe on the L1-PCA problem; see Section 5 for details.

3. Characterizing the KL exponent for Problems (1.5) and (2.1). Our goal in this section is to prove Theorem 2.1. This is achieved in three steps. First, we invoke a calculus rule established in [23] to show that the task of estimating the KL exponent at a limiting critical point of Problem (1.5) reduces to that of estimating the KL exponent at a limiting critical point of an LO-OC problem. Then, we establish a local error bound for the LO-OC problem and use it to characterize the KL exponent for that problem. Lastly, we utilize the result obtained for the LO-OC problem and the structures of Problems (1.5) and (2.1) to complete the proof.

3.1. Relation with Linear Optimization over the Stiefel Manifold. Let $P_1, \ldots, P_{2^nK}$ be an enumeration of the elements in $B(n,K)$. By definition of the $\ell_1$-norm, we can express the objective function $\ell$ of Problem (1.5) as the pointwise minimum of finitely many proper, lower semicontinuous functions:
\begin{equation}
\ell(Q) = \min_{i \in \{1, \ldots, 2^nK\}} \left\{ \langle XP_i, Q \rangle + \delta_{St(d,K)}(Q) \right\}.
\end{equation}
Since $\text{dom}(\partial \ell) \subseteq \text{dom}(\ell) = St(d,K)$ by definition and $\ell(Q) = -\|X^TQ\|_1$ for any $Q \in St(d,K)$, it is immediate that $\ell$ is continuous on $\text{dom}(\partial \ell)$. Moreover, using the result in [38, Exercise 8.8], we have $\partial \ell_i(Q) = XP_i + N_{St(d,K)}(Q)$ for any $Q \in St(d,K)$, which implies that $\text{dom}(\ell_i) = \text{dom}(\partial \ell_i) = St(d,K)$ for $i = 1, \ldots, 2^nK$. Thus, by [23, Corollary 3.1], in order to determine the KL exponent of the function $\ell$ at a point $x \in \text{dom}(\partial \ell)$, it suffices to determine the KL exponents of the functions $\ell_1, \ldots, \ell_{2^nK}$ at the point $x$. Noting that $\ell_i$ ($i = 1, \ldots, 2^nK$) is the sum of a linear function and the indicator function associated with $St(d,K)$, a natural approach is to study the following general LO-OC problem, where $A \in \mathbb{R}^{d \times K}$ is any given matrix:
\begin{equation}
(\text{LO-OC}) \quad \min_{Q \in \mathbb{R}^{d \times K}} \left\{ g(Q) = \langle A, Q \rangle + \delta_{St(d,K)}(Q) \right\}.
\end{equation}
By [23, Lemma 2.1], for any $\theta \in [0,1)$, the function $g$ has a KL exponent of $\theta$ at any of its non-limiting critical point. Thus, we shall focus on determining the KL exponents of $g$ at its limiting critical points.

3.2. Estimating the KL Exponent for Problem (LO-OC). Let
\begin{equation}
Q = \{ Q \in St(d,K) : 0 \in \partial g(Q) \}
\end{equation}
declare the set of limiting critical points of Problem (LO-OC). Based on the development in the previous subsection, our next step is to prove the following result, which can be of independent interest.
Theorem 3.1. There exist \( \epsilon_g \in (0, 1), \eta_g > 0 \) such that for all \( Q \in \text{St}(d, K) \) and \( Q^* \in \mathcal{Q} \) with \( \|Q - Q^*\|_F \leq \epsilon_g \),
\[
\text{dist}(0, \partial g(Q)) \geq \eta_g |g(Q) - g(Q^*)|^{1/2}.
\]

Theorem 3.1 implies that the KL exponent at any limiting critical point of Problem \((\text{LO-OC})\) is \( 1/2 \) with \( \epsilon = \epsilon_g \), \( \eta = \eta_g \), and \( \nu = +\infty \). Moreover, the constants \( \epsilon, \eta, \nu \) are uniform over all limiting critical points in \( \mathcal{Q} \). We remark that in the recent work [17], Jiang and Li studied the KL exponents at the optimal solutions to the so-called trust region sub-problem
\[
(\text{TRS}) \quad \min_{q \in \mathbb{R}^d : \|q\|_2 \leq 1} \left\{ \langle q, Wq \rangle + \langle a, q \rangle \right\},
\]
where \( W \in \mathbb{S}^d \) is a given symmetric matrix and \( a \in \mathbb{R}^d \) is a given vector. The result in [17, Theorem 4.8] implies that when \( W = 0 \), the KL exponent at any optimal solution to Problem \((\text{TRS})\) is \( 1/2 \). This in turn implies that when \( K = 1 \), the KL exponent at any optimal solution to Problem \((\text{LO-OC})\) is \( 1/2 \), which is consistent with the result in Theorem 3.1. We should point out that [17, Theorem 4.8] also applies to instances of Problem \((\text{TRS})\) with \( W \neq 0 \), though it only gives the KL exponents at the optimal solutions to the problem. It would be interesting to extend the results in [24, Theorem 1], [17, Theorem 4.8], and Theorem 3.1 and determine the KL exponents at the limiting critical points of the problem of minimizing an inhomogeneous quadratic function over the Stiefel manifold. We leave this as a future work.

The proof of Theorem 3.1 follows the same three-step strategy in [24]—i.e., we first characterize the limiting critical points of Problem \((\text{LO-OC})\) (Subsection 3.2.1), then establish a local error bound for the problem (Subsection 3.2.2), and finally determine the KL exponent at the said limiting critical points (Subsection 3.2.3). However, since Problem \((\text{LO-OC})\) involves minimizing a linear rather than homogeneous quadratic function over the Stiefel manifold, the structure of its limiting critical point set is quite different from that of the problem studied in [24]. Thus, we need to develop some new techniques to carry out the above strategy. We should mention that the proof of Theorem 3.1 is quite long and technical. Readers who are interested in how Theorem 3.1 is used to complete the proof of Theorem 2.1 can skip ahead to Subsection 3.3.

3.2.1. Structure of the Limiting Critical Point Set. We begin with the following result, which provides, among other things, a characterization of \( \mathcal{Q} \).

Proposition 3.2. Consider the map \( R : \text{St}(d, K) \to \mathbb{R}^{d \times K} \) given by
\[
R(Q) = A - QA^T Q.
\]
We have
\[
\text{dist}(0, \partial g(Q)) = \left\| \left( I_d - \frac{1}{2} QQ^T \right) R(Q) \right\|_F
\]
and
\[
\frac{1}{2} \|R(Q)\|_F \leq \text{dist}(0, \partial g(Q)) \leq \|R(Q)\|_F.
\]
In particular, we have \( Q \in \mathcal{Q} \) if and only if
\[
Q \in \text{St}(d, K) \quad \text{and} \quad R(Q) = 0.
\]
where $\theta = s \delta$ implies that $\overline{A}$ is a limiting critical point of the function $g_A$ at $\theta$. Thus, we can assume without loss of generality that $\theta = s \delta$.

Let $F = \{1, \ldots, K\}$  and $\overline{A} = \overline{\Sigma}_A = \text{Diag}(\sigma_1, \ldots, \sigma_r)$ with $\sigma_1 \geq \cdots \geq \sigma_r > 0$ being the positive singular values of $A$: $U_A \in \mathcal{O}_d$ with $U_{A,1} \in \mathbb{R}^{d \times r}$, $U_{A,2} \in \mathbb{R}^{d \times (d-r)}$; $V_A \in \mathcal{O}_K$ with $V_{A,1} \in \mathbb{R}^{K \times r}$, $V_{A,2} \in \mathbb{R}^{K \times (K-r)}$. Then, for any $Q \in \text{St}(d, K)$, we have $\langle A, Q \rangle = \langle \overline{\Sigma}_A, U_A^T Q V_A \rangle$ with $U_A^T Q V_A \in \text{St}(d, K)$. Moreover, Proposition 3.2 implies that $Q \in \text{St}(d, K)$ is a limiting critical point of the function $\langle Q \rangle \mapsto \langle A, Q \rangle + \delta_{\text{St}(d,K)}(Q)$ if and only if $U_A^T Q V_A \in \text{St}(d, K)$ is a limiting critical point of the function $Q \mapsto (\overline{\Sigma}_A, Q) + \delta_{\text{St}(d,K)}(Q)$. Thus, we can assume without loss of generality that

$$A = \begin{bmatrix} \overline{A} & 0 \\ 0 & 0 \end{bmatrix},$$

where $\overline{A} = \text{Diag}(a_1, \ldots, a_r)$ with $a_1 \geq \cdots \geq a_r > 0$. Suppose that $A$ has $p \geq 1$ distinct positive singular values. In other words, there exist indices $s_0, s_1, \ldots, s_p$ such that $0 = s_0 < s_1 < \cdots < s_p = r$ and

$$a_{s_0 + 1} = \cdots = a_{s_1} > a_{s_1 + 1} = \cdots = a_{s_2} > \cdots > a_{s_{p-1} + 1} = \cdots = a_{s_p}.$$
Let \( h_i = s_i - s_{i-1} \) be the multiplicity of the \( i \)-th largest positive singular value, where \( i = 1, \ldots, p \). Then, we clearly have \( \sum_{i=1}^{p} h_i = r \) and

\[
\tilde{A} = \text{BlkDiag}(a_1 I_{h_1}, \ldots, a_p I_{h_p}).
\]

Based on the block structures of \( A \) in (3.6) and \( \tilde{A} \) in (3.8), let us partition \( Q \in \mathbb{R}^{d \times K} \) as

\[
Q = \begin{bmatrix} Q_1 & Q_2 \\ Q_3 & Q_4 \end{bmatrix} \quad \text{and} \quad Q_1 = \begin{bmatrix} Q_{h_1 h_1} & \cdots & Q_{h_1 h_p} \\ \vdots & \ddots & \vdots \\ Q_{h_p h_1} & \cdots & Q_{h_p h_p} \end{bmatrix},
\]

where \( Q_1 \in \mathbb{R}^{r \times r}, \ Q_2 \in \mathbb{R}^{r \times (K-r)}, \ Q_3 \in \mathbb{R}^{(d-r) \times r}, \ Q_4 \in \mathbb{R}^{(d-r) \times (K-r)} \), and \( Q_{h_i h_j} \in \mathbb{R}^{h_i \times h_j} \) for \( i, j \in \{1, \ldots, p\} \). With the above partition, we can further elucidate the structure of a limiting critical point of Problem (LO-OC). Specifically, we establish the following result:

**Proposition 3.3.** Suppose that \( A \in \mathbb{R}^{d \times K} \) has the form given in (3.6). Then, we have \( Q \in \mathcal{Q} \) if and only if

\[
Q = \begin{bmatrix} U \text{Diag}(q) U^T & 0 \\ 0 & V \end{bmatrix},
\]

where \( U = \text{BlkDiag}(U_1, \ldots, U_p) \) with \( U_i \in \mathcal{O}^{h_i} \) for \( i = 1, \ldots, p \), \( q \in \{\pm 1\}^r \), and \( V \in \text{St}(d-r, K-r) \).

**Proof.** If \( Q \) is of the form given in (3.10), then using the block structure of \( A \) in (3.6), it is straightforward to verify that \( Q \in \text{St}(d, K) \) and \( A - QA^T Q = 0 \). By Proposition 3.2, we conclude that \( Q \in \mathcal{Q} \).

Conversely, suppose that \( Q \in \mathcal{Q} \). By Proposition 3.2, we have \( A - QA^T Q = 0 \). Since \( Q \in \text{St}(d, K) \), this implies that

\[
Q^T A - A^T Q = 0,
\]

which, together with \( A - QA^T Q = 0 \), yields

\[
A - QQ^T A = 0.
\]

Using the block structures of \( A \) in (3.6) and \( Q \) in (3.9), we have \( Q^T A = \begin{bmatrix} Q_1^T \tilde{A} & 0 \\ Q_2^T \tilde{A} & 0 \end{bmatrix} \).

It then follows from (3.11) that \( Q_1^T \tilde{A} = \tilde{A}^T Q_1 \) and \( Q_2^T \tilde{A} = 0 \). Since \( \tilde{A} \) has full rank, the latter implies that \( Q_2 = 0 \), which in turn implies that \( Q_1^T Q_4 = I_{K-r} \) because we have \( Q \in \text{St}(d, K) \). Using \( Q_2 = 0 \) and (3.12), we obtain

\[
A - QQ^T A = \begin{bmatrix} \tilde{A} - Q_1 Q_1^T \tilde{A} & 0 \\ -Q_3 Q_4^T \tilde{A} & 0 \end{bmatrix} = 0;
\]

i.e., \( (I_r - Q_1 Q_1^T) \tilde{A} = 0 \) and \( Q_3 Q_4^T \tilde{A} = 0 \). These, together with the fact that \( \tilde{A} \) has full rank, imply that \( Q_1 \in \mathcal{O}^r \) and \( Q_3 = 0 \).

Now, using the block structures of \( \tilde{A} \) in (3.8) and \( Q_1 \) in (3.9), we get

\[
Q_1^T \tilde{A} = \begin{bmatrix} a_{s_1} Q_{h_1 h_1}^T & \cdots & a_{s_1} Q_{h_1 h_p}^T \\ \vdots & \ddots & \vdots \\ a_{s_p} Q_{h_p h_1}^T & \cdots & a_{s_p} Q_{h_p h_p}^T \end{bmatrix}, \quad \tilde{A}^T Q_1 = \begin{bmatrix} a_{s_1} Q_{h_1 h_1} & \cdots & a_{s_1} Q_{h_1 h_p} \\ \vdots & \ddots & \vdots \\ a_{s_p} Q_{h_p h_1} & \cdots & a_{s_p} Q_{h_p h_p} \end{bmatrix}.
\]
Since \( Q_1^T \tilde{A} = \tilde{A}^T Q_1 \), we have
\[
Q_{h_i h_j}^T = \frac{a_{s_i}}{a_{s_j}} Q_{h_j h_i}, \quad \text{for } i, j \in \{1, \ldots, p\},
\]
which implies that
\[
\sum_{i=1}^{p} Q_{h_i h_j}^T Q_{h_j h_i} = \sum_{i=1}^{p} \frac{a_{s_i}^2}{a_{s_j}^2} Q_{h_j h_i} Q_{h_j h_i}^T, \quad \text{for } j \in \{1, \ldots, p\}.
\]
Moreover, the fact that \( Q_1 \in \mathcal{O}^r \) implies
\[
\sum_{i=1}^{p} Q_{h_i h_j}^T Q_{h_i h_j} = I_{h_j}, \quad \sum_{i=1}^{p} Q_{h_i h_j} Q_{h_i h_j}^T = I_{h_j}, \quad \text{for } j \in \{1, \ldots, p\}.
\]
It then follows from (3.15) and (3.16) that
\[
\sum_{i \neq j} \left(1 - \frac{a_{s_i}^2}{a_{s_j}^2}\right) \|Q_{h_j h_i}\|_F^2 = 0 \quad \text{for } j \in \{1, \ldots, p\}.
\]
By rewriting (3.14) as \( Q_{h_i h_j}^T = \frac{a_{s_j}}{a_{s_i}} Q_{h_i h_j} \) for \( i, j \in \{1, \ldots, p\} \) and repeating the above argument, we get
\[
\sum_{i \neq j} \left(1 - \frac{a_{s_i}^2}{a_{s_j}^2}\right) \|Q_{h_i h_j}\|_F^2 = 0 \quad \text{for } j \in \{1, \ldots, p\}.
\]
Since \( a_{s_i} > \cdots > a_{s_p} > 0 \) by (3.7), the identities in (3.17) and (3.18) imply that \( Q_{h_i h_j} = 0 \) for \( i, j \in \{1, \ldots, p\}; i \neq j \). This, together with (3.14) and (3.16), yields
\[
Q_{h_i h_i} = Q_{h_i h_i}^T, \quad \text{and} \quad Q_{h_i h_i}^T Q_{h_i h_i} = I_{h_j}, \quad \text{for } i \in \{1, \ldots, p\}.
\]
Let \( Q_{h_i h_i} = U_i \Lambda_i U_i^T \) (\( i = 1, \ldots, p \)) be an eigen-decomposition of \( Q_{h_i h_i} \), where \( U_i \in \mathcal{O}^{h_i} \), and \( \Lambda_i = \text{Diag}(\lambda_{s_i-1}, \ldots, \lambda_{s_i}) \). Then, we have \( Q_{h_i h_i}^T Q_{h_i h_i} = U_i \Lambda_i^2 U_i^T = I_{h_i} \) from (3.19), which implies that \( \Lambda_i^2 = I_{h_i} \). It follows that \( \lambda_{s_i-1}, \ldots, \lambda_{s_i} \in \{\pm 1\} \).

Putting all the pieces together, we see that \( Q_1 \) takes the form
\[
Q_1 = \text{BlkDiag}(U_1, \ldots, U_p) \cdot \text{Diag}(q) \cdot \text{BlkDiag}(U_1^T, \ldots, U_p^T)
\]
with \( U_i \in \mathcal{O}^{h_i} \) for \( i = 1, \ldots, p \) and \( q \in \{\pm 1\}^r \), \( Q_2 = 0 \), \( Q_3 = 0 \), and \( Q_4 \in \text{St}(d - r, K - r) \). This completes the proof. \( \Box \)

Proposition 3.3 suggests that when \( A \in \mathbb{R}^{d \times K} \) has the form given in (3.6), the set \( Q \) of limiting critical points of Problem (LO-OC) can be expressed as
\[
Q = \bigcup_{q \in \{\pm 1\}^r} Q_q,
\]
where
\[
Q_q = \left\{ \begin{bmatrix} U & \text{Diag}(q) U^T \\ 0 & V \end{bmatrix} : U = \text{BlkDiag}(U_1, \ldots, U_p), U_i \in \mathcal{O}^{h_i} \text{ for } i \in \{1, \ldots, p\}, V \in \text{St}(d - r, K - r) \right\}.
\]
The following result shows that the collection \( \{Q_q\}_{q \in \{\pm 1\}^r} \) essentially forms a well-separated partition of the set \( Q \).
PROPOSITION 3.4. Suppose that $A \in \mathbb{R}^{d \times K}$ has the form given in (3.6). Let $q, q' \in \{\pm 1\}^{r}$ be arbitrary. Then, we either have $Q_q = Q_{q'}$ or $Q_q \cap Q_{q'} = \emptyset$. Moreover, if the latter holds, then $\text{dist}(Q_q, Q_{q'}) \geq 2$.

Proof. Let $q = (q_1, \ldots, q_p)$ and $q' = (q'_1, \ldots, q'_p)$, where $q_i, q'_i \in \{\pm 1\}^{b_i}$ for $i = 1, \ldots, p$. Suppose that $Q = \begin{bmatrix} Q_1 & 0 \\ 0 & V \end{bmatrix} \in Q_q$. By definition of $Q_q$ in (3.20), for $i = 1, \ldots, p$, the eigenvalues of the $i$-th diagonal block of $Q_1$ are given by the entries of $q_i$. Thus, if $Q \in Q_q \cap Q_{q'}$, then both $q_i$ and $q'_i$ are vectors of eigenvalues of the $i$-th diagonal block of $Q_1$, which implies that $q_i$ and $q'_i$ are equal up to a permutation for $i = 1, \ldots, p$. It follows that whenever $Q_q \cap Q_{q'} \neq \emptyset$, we have $Q_q = Q_{q'}$.

Now, suppose that $Q_q \cap Q_{q'} = \emptyset$. Let

\[
Q = \begin{bmatrix} U \text{Diag}(q) U^T & 0 \\ 0 & V \end{bmatrix} \in Q_q, \quad Q' = \begin{bmatrix} U' \text{Diag}(q') U'^T & 0 \\ 0 & V' \end{bmatrix} \in Q_{q'}
\]

be arbitrary, where $U = \text{BlkDiag}(U_1, \ldots, U_p)$, $U' = \text{BlkDiag}(U'_1, \ldots, U'_p)$ with $U_i, U'_i \in \mathcal{O}^{b_i}$ for $i = 1, \ldots, p$ and $V, V' \in \text{St}(d-r, K-r)$. Then, we have

\[
\|Q - Q'\|_F^2 = \sum_{i=1}^{p} \|U_i \text{Diag}(q_i) U_i^T - U'_i \text{Diag}(q'_i) U'_i^T\|_F^2 + \|V - V'\|_F^2
\]

\[
\geq \sum_{i=1}^{p} \min_{U_i \in \mathcal{O}^{b_i}} \|U_i \text{Diag}(q_i) U_i^T - \text{Diag}(q'_i)\|_F^2 + \min_{V \in \text{St}(d-r, K-r)} \|V - V'\|_F^2
\]

\[= \sum_{i=1}^{p} \min_{U_i \in \mathcal{O}^{b_i}} \|U_i \text{Diag}(q_i) U_i^T - \text{Diag}(q'_i)\|_F^2, \tag{3.21}\]

where the last equality follows from the fact that $V' \in \text{St}(d-r, K-r)$. For $i = 1, \ldots, p$, let $t_i$ and $t'_i$ denote the number of 1's in $q_i$ and $q'_i$, respectively. If there exists a $j \in \{1, \ldots, p\}$ such that $t_j \neq t'_j$, then we can find a $k \in \{1, \ldots, h_j\}$ such that for any $U_j \in \mathcal{O}^{b_j}$,

\[
\|U_j \text{Diag}(q_j) U_j^T - \text{Diag}(q'_j)\|_F^2 \geq \|U_j \text{Diag}(q_j) U_j^T - \text{Diag}(q'_j)\|_F^2
\]

\[\geq \|\lambda_k (U_j \text{Diag}(q_j) U_j^T - \text{Diag}(q'_j)\) - \lambda_k (\text{Diag}(q'_j))\|^2 = 4, \tag{3.22}\]

where the second inequality follows from classic perturbation results for eigenvalues of symmetric matrices (see, e.g., [41, Corollary 4.10]) and the last equality is due to the fact that $q_j, q'_j \in \{\pm 1\}^{b_j}$ and $t_j \neq t'_j$. Since $Q \in Q_q, Q' \in Q_{q'}$ are arbitrary, we conclude from (3.21) and (3.22) that $\text{dist}(Q_q, Q_{q'}) \geq 2$. Otherwise, we have $t_i = t'_i$ for $i = 1, \ldots, p$, which implies that $q_i$ and $q'_i$ are equal up to a permutation for $i = 1, \ldots, p$. In this case, we have $Q_q = Q_{q'}$, which contradicts our assumption that $Q_q \cap Q_{q'} = \emptyset$. This completes the proof. \[\square\]

3.2.2. Local Error Bound. Equipped with the results in the previous section, our next task is to establish the following local error bound for Problem (LO-OC), which provides an estimate of the distance between any point from a certain subset of $\text{St}(d, K)$ to the set of limiting critical points of Problem (LO-OC) using the map $R$ introduced in (3.1). As we shall see, such an error bound plays a crucial role in determining the KL exponent at the limiting critical points of Problem (LO-OC).
Theorem 3.5. Let $A \in \mathbb{R}^{d \times K}$ be an arbitrary rank-$r$ matrix whose SVD is given by (3.5) and whose positive singular values are given by (3.7). Then, for any $q \in \{\pm 1\}^r$, we have

\begin{equation}
\text{dist}(Q, \bar{Q}_q) \leq \kappa \|R(Q)\|_F \quad \text{for all } Q \in \text{St}(d, K) \text{ with } \text{dist}(Q, \bar{Q}_q) < 1,
\end{equation}

where $\bar{Q}_q = U_A Q_q V_A^T = \{A W V_A^T : W \in Q_q\}$ and

\begin{align*}
\kappa &= \frac{1}{a_r} \left( 13 + 6(6p - 5) \left( \min_{i,j \in \{1, \ldots, p\}} \delta_{ij}^2 \right) \right)^{-1/2} \\
\delta_{ij} &= \frac{a_{si}}{a_{sj}} - \frac{a_{si}}{a_{sj}} \quad \text{for } i, j \in \{1, \ldots, p\}; \ i \neq j.
\end{align*}

It is worth noting that error bounds of similar nature have been extensively used to study the convergence behavior of various iterative methods; see, e.g., \[6, 25, 46, 24\] for some recent developments. Thus, Theorem 3.5 can be of independent interest.

To prove Theorem 3.5, observe that since

\begin{align*}
\text{dist}(Q, \bar{Q}_q) = \text{dist}(U_A^T Q V_A, Q_q), \\
\|A - QA^T Q\|_F = \|\Sigma_A - (U_A^T Q V_A) \Sigma_A^T (U_A^T Q V_A)\|_F, \\
U_A^T Q V_A \in \text{St}(d, K),
\end{align*}

it suffices to establish (3.23) for the case where $A$ has the block structure given in (3.6) (in particular, we have $U_A = I_d$, $V_A = I_K$, and $Q_q = Q_q$). In view of the structure of $Q_q$ given in (3.20), a natural idea is to first consider the partition $Q = \begin{bmatrix} Q_1 & Q_2 \\ Q_3 & Q_4 \end{bmatrix} \in \text{St}(d, K)$ as in (3.9) and observe that

\begin{equation}
\text{dist}^2(Q, Q_q) = \text{dist}^2(Q_1, Q_q^1) + \|Q_2\|_F^2 + \|Q_3\|_F^2 + \min_{V \in \text{St}(d-r, K-r)} \|Q_4 - V\|_F^2,
\end{equation}

where

\begin{equation}
Q_q^1 = \{U \text{ Diag}(q) U^T : U = \text{BlkDiag}(U_1, \ldots, U_p), \ U_i \in \mathcal{O}^{h_i} \text{ for } i \in \{1, \ldots, p\}\}.
\end{equation}

Then, it suffices to bound each of the terms on the right-hand side of (3.24) separately. Let us begin by dispensing with the easy cases.

Proposition 3.6. Suppose that $A \in \mathbb{R}^{d \times K}$ has the form given in (3.6) and $Q \in \text{St}(d, K)$ is partitioned according to (3.9). Then, the following hold:

\begin{align}
\text{dist}^2(Q, Q_q) &= \text{dist}^2(Q_1, Q_q^1) + \|Q_2\|_F^2 + \|Q_3\|_F^2 + \min_{V \in \text{St}(d-r, K-r)} \|Q_4 - V\|_F^2, \\
\min_{V \in \text{St}(d-r, K-r)} \|Q_4 - V\|_F^2 &\leq \|Q_2\|_F^2.
\end{align}

Proof. We first prove (3.26). Using the block structures of $A$ in (3.6) and $Q$ in (3.9) and the fact that $Q_1^T Q_1 + Q_3^T Q_3 = I_r$, we compute

\begin{align}
\|R(Q)\|_F^2 &= \|\bar{A} - Q_1 \bar{A}^T Q_1\|_F^2 + \|Q_1 \bar{A}^T Q_2\|_F^2 + \|Q_3 \bar{A}^T Q_1\|_F^2 + \|Q_3 \bar{A}^T Q_2\|_F^2 \\
&= \|\bar{A} - Q_1 \bar{A}^T Q_1\|_F^2 + \|Q_3 \bar{A}^T Q_1\|_F^2 + \|\bar{A}^T Q_2\|_F^2.
\end{align}
This, together with the definition of $\tilde{A}$, implies that $\|R(Q)\|^2_F \geq \|\tilde{A}^T Q_2\|^2_F \geq a_2^2\|Q_2\|^2_F$. Moreover, since
\begin{equation}
0 \leq \langle B - B^T, B - B^T \rangle = 2(\langle B, B \rangle - \langle B^T, B \rangle)
\end{equation}
for any $B \in \mathbb{R}^{r \times r}$, we obtain from (3.28) that
\begin{align*}
\|R(Q)\|^2_F &\geq \|\tilde{A} - Q_1, \tilde{A}^T Q_1\|^2_F + \|Q_3, \tilde{A}^T Q_1\|^2_F \\
&= \|\tilde{A}\|^2_F - 2\langle \tilde{A}, Q_1, \tilde{A}^T Q_1 \rangle + \|\tilde{A}^T Q_1\|^2_F \\
&= \|\tilde{A}\|^2_F - \langle Q_1^T \tilde{A}, \tilde{A}^T Q_1 \rangle - \langle Q_1^T \tilde{A}, \tilde{A}^T Q_1 \rangle + \langle \tilde{A}^T Q_1, \tilde{A}^T Q_1 \rangle \\
&\geq \|\tilde{A}\|^2_F - \langle Q_1^T \tilde{A}, \tilde{A}^T Q_1 \rangle.
\end{align*}

Using the fact that $Q_1^T Q_1 + Q_3^T Q_3 = I_r$ and invoking (3.29), (3.30), we obtain
\begin{align*}
a_2^2\|Q_1\|^2_F &\leq \|\tilde{A} Q_1\|^2_F = \|\tilde{A}\|^2_F - \|\tilde{A} Q_1\|^2_F \\
&\leq \|\tilde{A}\|^2_F - \langle Q_1^T \tilde{A}, A Q_1^T \rangle = \|\tilde{A}\|^2_F - \langle Q_1^T \tilde{A}, \tilde{A}^T Q_1 \rangle \leq \|R(Q)\|^2_F.
\end{align*}

Next, we prove (3.27). Let $Q_4 = U_4 \Sigma_4 V_4^T$ be a thin SVD of $Q_4$, where $\Sigma_4 = \text{Diag}(\sigma_1, \ldots, \sigma_{K-r})$ with $\sigma_1 \geq \cdots \geq \sigma_{K-r} \geq 0$ being the singular values of $Q_4$, $U_4 \in \text{St}(d-r, K-r)$, and $V_4 \in \mathcal{O}^{K-r}$. Noting that the left-hand side of (3.27) is an instance of the orthogonal Procrustes problem [40], we have
\begin{equation}
\min_{V \in \text{St}(d-r, K-r)} \|Q_4 - V\|^2_F = \|Q_4 - U_4 V_4^T\|^2_F = \sum_{i=1}^{K-r} (1 - \sigma_i)^2.
\end{equation}

Using the facts that (i) $(1-x)^2 \leq (1-x)^2(1+x)^2$ for any $x \geq 0$, (ii) $Q_2^T Q_2 + Q_4^T Q_4 = I_{K-r}$, and (iii) $\|Q_2\| \leq \|Q\| \leq 1$, we obtain
\begin{equation}
\sum_{i=1}^{K-r} (1 - \sigma_i)^2 \leq \sum_{i=1}^{K-r} (1 - \sigma_i^2)^2 = \|I_{K-r} - Q_4^T Q_4\|^2_F = \|Q_2^T Q_2\|^2_F \leq \|Q_2\|^2_F.
\end{equation}

This completes the proof. \qed

Now, it remains to bound $\text{dist}^2(Q_1, Q_2)$. The following technical lemma, whose proof can be found in Appendix A, will be useful for that purpose. Recall that $\delta_{ij} = \frac{a_{ij}}{a_{ij}} - \frac{a_{ji}}{a_{ji}}$ for $i, j \in \{1, \ldots, p\}; i \neq j$.

**Lemma 3.7.** Suppose that $A \in \mathbb{R}^{d \times K}$ has the form given in (3.6) and $Q \in \text{St}(d, K)$ is partitioned according to (3.9). Then, the following hold:
\begin{align}
\sum_{i=1}^{p} \|Q_{h_i} - Q_{h_i}^T\|^2_F &\leq \|R(Q)\|^2_F, \\
\sum_{i,j \in \{1, \ldots, p\}} \delta_{ij}^2 \sum_{i=1}^{p} \|Q_{h_i} - Q_{h_i}^T\|^2_F &\leq 6\|R(Q)\|^2_F.
\end{align}

**Proposition 3.8.** Suppose that $A \in \mathbb{R}^{d \times K}$ has the form given in (3.6) and let $q \in \{\pm 1\}^r$ be given. Furthermore, suppose that $Q \in \text{St}(d, K)$ satisfies $\text{dist}(Q, Q_d) < 1$ and is partitioned according to (3.9). Then,
\begin{equation}
\text{dist}^2(Q_1, Q_2^T) \leq \frac{1}{a_2^2} \left(10 + 6p - 5 \left(\min_{i, j \in \{1, \ldots, p\}} \delta_{ij}^2 \right)^{-1}\right) \|R(Q)\|^2_F.
\end{equation}
Proof. Based on the block structure of $Q_1$ in (3.9) and the definition of $Q_q^1$ in (3.25), we have

\begin{equation}
\text{dist}^2(Q_1, Q_q^1) = \sum_{i=1}^{p} \sum_{j \neq i} \|Q_{h_i, h_j}\|_F^2 + \sum_{i=1}^{p} \min_{U_i \in O^{h_i}} \|Q_{h_i, h_i} - U_i \text{Diag}(q_i) U_i^T\|_F^2.
\end{equation}

The first term on the right-hand side of (3.34) can be bounded using (3.32). Thus, it suffices to bound the second term. Towards that end, let $\tilde{Q}_{h_i, h_i} = (Q_{h_i, h_i} + Q_{h_i, h_i}^T)/2 \in S^{h_i}$ for $i = 1, \ldots, p$. Since $Q_{h_i, h_i} - \tilde{Q}_{h_i, h_i}$ is skew-symmetric and $Q_{h_i, h_i} - U_i \text{Diag}(q_i) U_i^T$ is symmetric, we have $(Q_{h_i, h_i} - \tilde{Q}_{h_i, h_i}, \tilde{Q}_{h_i, h_i} - U_i \text{Diag}(q_i) U_i^T) = 0$. This, together with (3.31), implies that

\begin{equation}
\sum_{i=1}^{p} \min_{U_i \in O^{h_i}} \|Q_{h_i, h_i} - U_i \text{Diag}(q_i) U_i^T\|_F^2 \\
= \sum_{i=1}^{p} \|Q_{h_i, h_i} - \tilde{Q}_{h_i, h_i}\|_F^2 + \sum_{i=1}^{p} \min_{U_i \in O^{h_i}} \|\tilde{Q}_{h_i, h_i} - U_i \text{Diag}(q_i) U_i^T\|_F^2 \\
= \frac{1}{4} \sum_{i=1}^{p} \|Q_{h_i, h_i} - \tilde{Q}_{h_i, h_i}^T\|_F^2 + \sum_{i=1}^{p} \min_{U_i \in O^{h_i}} \|\tilde{Q}_{h_i, h_i} - U_i \text{Diag}(q_i) U_i^T\|_F^2 \\
\leq \frac{1}{4a^2} \|R(Q)\|_F^2 + \sum_{i=1}^{p} \min_{U_i \in O^{h_i}} \|\tilde{Q}_{h_i, h_i} - U_i \text{Diag}(q_i) U_i^T\|_F^2.
\end{equation}

Now, for $i = 1, \ldots, p$, let $\tilde{Q}_{h_i, h_i} = U_i A_{h_i} U_i^T$ be an eigen-decomposition of $\tilde{Q}_{h_i, h_i} \in S^{h_i}$, where $A_{h_i} = \text{Diag}(\lambda_1^{(h_i)}, \ldots, \lambda_{h_i}^{(h_i)})$ with $\lambda_1^{(h_i)} \geq \cdots \geq \lambda_{h_i}^{(h_i)}$ being the eigenvalues of $\tilde{Q}_{h_i, h_i}$. Since $\|Q_{h_i, h_i}\| \leq \|Q_1\| \leq \|Q\| \leq 1$ for $i = 1, \ldots, p$, we have $|\lambda_k^{(h_i)}| \leq 1$ for $i = 1, \ldots, p$ and $k = 1, \ldots, h_i$. We claim that for $i = 1, \ldots, p$ and $k = 1, \ldots, h_i$, the $k$-th largest eigenvalue of $Q_{h_i, h_i}$ has the same sign as the $k$-th largest eigenvalue of $U_i \text{Diag}(q_i) U_i^T$. Indeed, if the claim is not true for some $i \in \{1, \ldots, p\}$ and $k \in \{1, \ldots, h_i\}$, then for any $U_i \in O^{h_i}$, we have

\begin{align*}
\|\tilde{Q}_{h_i, h_i} - U_i \text{Diag}(q_i) U_i^T\|_F^2 &\geq \|\tilde{Q}_{h_i, h_i} - U_i \text{Diag}(q_i) U_i^T\|^2 \\
&\geq |\lambda_k^{(h_i)} - \lambda_k(U_i \text{Diag}(q_i) U_i^T)| \geq 1,
\end{align*}

where the second inequality follows from classic perturbation results for eigenvalues of symmetric matrices (see, e.g., [41, Corollary 4.10]) and the third follows the fact that the sign of $\lambda_k(U_i \text{Diag}(q_i) U_i^T)$ with $q_i \in \{\pm 1\}^{h_i}$ is different from that of $\lambda_k^{(h_i)}$. This implies that $\text{dist}(Q, Q_q) \geq 1$, which contradicts our assumption that $\text{dist}(Q, Q_q) < 1$.

Using the above claim, we can bound

\begin{equation}
\min_{U_i \in O^{h_i}} \|\tilde{Q}_{h_i, h_i} - U_i \text{Diag}(q_i) U_i^T\|_F^2 \leq \sum_{k=1}^{h_i} \left(1 - |\lambda_k^{(h_i)}|\right)^2 \sum_{k=1}^{h_i} \left(1 - |\lambda_k^{(h_i)}|\right)^2 \\
\leq \sum_{k=1}^{h_i} \left(1 - |\lambda_k^{(h_i)}|\right)^2 \left(1 + |\lambda_k^{(h_i)}|\right)^2 = \|I_{h_i} - Q_{h_i, h_i}^T Q_{h_i, h_i}\|_F^2.
\end{equation}
for \( i = 1, \ldots, p \). Let us turn to bound \( \sum_{i=1}^{p} \| I_{h_{i}} - Q_{h_{h_{i}}}^{T} Q_{h_{h_{i}}}, \|_{F}^{2} \). Observe that with \( \Delta_{h_{i}} = Q_{h_{h_{i}}} - Q_{h_{h_{i}}}^{T} \) for \( i = 1, \ldots, p \), we have

\[
\sum_{i=1}^{p} \| I_{h_{i}} - Q_{h_{h_{i}}}^{T} Q_{h_{h_{i}}}, \|_{F}^{2} = \sum_{i=1}^{p} \left\| I_{h_{i}} - \left( Q_{h_{h_{i}}}^{T} + \frac{1}{2} \Delta_{h_{i}} h_{i} \right) \left( Q_{h_{h_{i}}} - \frac{1}{2} \Delta_{h_{i}} h_{i} \right) \right\|_{F}^{2} \leq 3 \sum_{i=1}^{p} \left( \| I_{h_{i}} - Q_{h_{h_{i}}}^{T} Q_{h_{h_{i}}}, \|_{F}^{2} + \| Q_{h_{h_{i}}} \Delta_{h_{i}} h_{i}, \|_{F}^{2} + \frac{1}{16} \| \Delta_{h_{i}} h_{i}, \|_{F}^{2} \right) \leq 3 \sum_{i=1}^{p} \| I_{h_{i}} - Q_{h_{h_{i}}}^{T} Q_{h_{h_{i}}}, \|_{F}^{2} + \frac{15}{4} \sum_{i=1}^{p} \| \Delta_{h_{i}} h_{i}, \|_{F}^{2}. \]

(3.37)

where the second-to-last inequality follows from the fact that \( \| Q_{h_{h_{i}}} \| \leq 1 \) and \( \| \Delta_{h_{i}} h_{i} \| \leq 2 \| Q_{h_{h_{i}}} \| \leq 2 \), and the last inequality follows from (3.31). Continuing, we bound

\[
\sum_{j=1}^{p} \| I_{h_{j}} - Q_{h_{h_{j}}}^{T} Q_{h_{h_{j}}}, \|_{F}^{2} = \sum_{j=1}^{p} \left\| I_{h_{j}} - \sum_{i=1}^{p} Q_{h_{h_{i}}}^{T} Q_{h_{h_{i}}} + \sum_{i \neq j} Q_{h_{h_{i}}}^{T} Q_{h_{h_{i}}}, \right\|_{F}^{2} \leq 2 \sum_{j=1}^{p} \left\| I_{h_{j}} - \sum_{i=1}^{p} Q_{h_{h_{i}}}^{T} Q_{h_{h_{i}}}, \right\|_{F}^{2} + \frac{15}{4} \sum_{j=1}^{p} \| \Delta_{h_{j}} h_{j}, \|_{F}^{2}. \]

(3.38)

where the second inequality follows from the fact that \( \{ I_{h_{j}} - \sum_{i=1}^{p} Q_{h_{h_{i}}}^{T} Q_{h_{h_{i}}} \}_{j=1}^{p} \) are the diagonal blocks of \( I_{r} - Q_{i}^{T} Q_{i} \) and the last is due to \( Q_{i}^{T} Q_{i} + Q_{i}^{T} Q_{3} = I_{r} \), \( \| Q_{3} \| \leq 1 \), and \( \| Q_{h_{j}} \| \leq 1 \) for \( i, j \in \{1, \ldots, p\} \).

Upon putting (3.34)–(3.38) together and invoking (3.26) and (3.32), we obtain (3.33). This completes the proof.

We now have all the ingredients to finish the proof of Theorem 3.5.

Proof of Theorem 3.5. Let \( q \in \{ \pm 1 \}^{r} \) be given. Using (3.24) and the results in Propositions 3.6 and 3.8, we get

\[
\text{dist}^{2}(Q, Q_{q}) \leq \frac{1}{a_{r}^{2}} \left( 13 + 6(6p - 5) \left( \min_{i,j \in \{1, \ldots, p\}} \delta_{ij}^{2} \right)^{-1} \right) \| R(Q) \|_{F}^{2}
\]

for any \( Q \in \text{St}(d, K) \) satisfying \( \text{dist}(Q, Q_{q}) < 1 \). This implies (3.23), as desired. □
3.2.3. From Error Bound to KL Exponent. Once we have the local error bound \((3.23)\), it is rather straightforward to determine the KL exponent at the limiting critical points of Problem (\textsc{LO-OC}). We remark that although there are works showing how various error bounds can be used to determine the KL exponent for a host of optimization problems (see, e.g., \([6, 23, 24]\)), they focus either on settings that only allow convex feasible sets or on homogeneous quadratic objective functions. As such, the results in these works cannot be applied directly to our problem setting.

**Proof of Theorem 3.1.** Let \(Q \in \text{St}(d, K)\) and \(Q^* \in Q\) be such that \(\|Q - Q^*\|_F < 1\). Furthermore, let \(Q^* \in Q\) be such that \(\text{dist}(Q, Q^*) = \|Q - Q^*\|_F\). Clearly, we have \(\|Q - Q^*\|_F < 1\). We claim that \(Q^*, \tilde{Q}^* \in \tilde{Q}_q\) for some \(q \in \{\pm 1\}^\tau\). Indeed, if this is not the case, then we have \(Q^* \in \tilde{Q}_q\) and \(Q^* \in \tilde{Q}_q\) for some \(q, q^* \in \{\pm 1\}^\tau\) with \(q \neq q^*\). Since \(\text{dist}(\tilde{Q}_q, \tilde{Q}_{q^*}) = \text{dist}(Q_q, Q_{q^*})\) (recall the definition of \(Q_q\) in Theorem 3.5), Proposition 3.4 implies that \(\|Q^* - \tilde{Q}^*\|_F \geq 2\). However, our assumption gives \(\|Q^* - \tilde{Q}^*\|_F \leq \|Q - \tilde{Q}^*\|_F + \|Q - Q^*\|_F < 2\), which is a contradiction. This establishes the claim.

Next, we claim that \(g\) is constant on \(\tilde{Q}_q\). Indeed, for any \(W \in \tilde{Q}_q\), we have

\[
g(W) = \langle A, W \rangle = \langle \Sigma_A, U_A^T W V_A \rangle = \sum_{i=1}^p \sum_{j=1}^h a_{ij} q_{ij},
\]

where the first equality is due to the fact that \(W \in \text{St}(d, K)\); the second inequality uses the SVD of \(A\) (\(3.5\)); the third inequality follows from the definition of \(\Sigma_A\) (cf. (3.6) and (3.8)), the fact that \(U_A^T W V_A \in \tilde{Q}_q\), and the definition of \(Q_q\) in (3.20). Upon noting that the rightmost expression does not depend on \(W\), the claim is established. In particular, we obtain \(g(Q^*) = g(\tilde{Q}^*)\).

Since \(\tilde{Q}^* \in Q\), we have \(A = \tilde{Q}^* A^T \tilde{Q}^*\) by Proposition 3.2, which implies that \(A = Q^* (Q^*)^T A\) (see (3.12)). It follows that

\[
g(Q) - g(\tilde{Q}^*) = \langle A, Q - \tilde{Q}^* \rangle = \langle \tilde{Q}^* A^T \tilde{Q}^* Q - A, \tilde{Q}^* \rangle
\]

\[
= \langle A^T \tilde{Q}^*, (\tilde{Q}^*)^T Q - I_K \rangle,
\]

\[
g(Q) - g(\tilde{Q}^*) = \langle A, Q - \tilde{Q}^* \rangle = \langle \tilde{Q}^* (\tilde{Q}^*)^T A, Q \rangle - \langle A, \tilde{Q}^* \rangle
\]

\[
= \langle A^T \tilde{Q}^*, Q^T \tilde{Q}^* - I_K \rangle.
\]

Summing the above two equalities yields

\[
|g(Q) - g(\tilde{Q}^*)| = \frac{1}{2} \left| \langle A^T \tilde{Q}^*, (\tilde{Q}^*)^T Q + Q^T \tilde{Q}^* - 2I_K \rangle \right|
\]

\[
= \frac{1}{2} \left| \langle A^T \tilde{Q}^*, (Q - \tilde{Q}^*)^T (Q - \tilde{Q}^*) \rangle \right| \leq \frac{1}{2} \|A\| \cdot \|Q - \tilde{Q}^*\|_F^2
\]

\[
\leq \frac{1}{2} \kappa^2 \|A\| \cdot \|R(Q)\|_F^2 \leq 2\kappa^2 \|A\| \cdot \text{dist}^2(0, \partial g(Q)),
\]

where the first inequality follows from the Cauchy-Schwarz inequality and the fact that \(\|Q^*\| \leq 1\); the second inequality follows from the assumption that \(\text{dist}(Q, \tilde{Q}_q) = \|Q - \tilde{Q}^*\|_F\) and Theorem 3.5; the last inequality follows from Proposition 3.2. Recalling that \(g(Q^*) = g(\tilde{Q}^*)\), we establish Theorem 3.1 with \(\eta_g = (2\kappa^2 \|A\|)^{-1/2}\) and \(\epsilon_g < 1\).

3.3. Completing the Proof. We are now ready to achieve our original goal of characterizing the KL exponent for Problems (1.5) and (2.1).
Proof of Theorem 2.1. Recall that the objective function \( \ell \) of Problem (1.5) takes the form \( \ell(Q) = \min_{i \in \{1, \ldots, 2^m \}} \ell_i(Q) \), where \( \ell_i(Q) = \langle XP_i, Q \rangle + \delta_{St(d,K)}(Q) \). For any \( Q \in St(d,K) \), let \( \mathcal{I}(Q) = \{ i \in \{1, \ldots, 2^m \} : \ell(Q) = \ell_i(Q) \} \) denote the set of active indices of \( \ell \) at \( Q \). Furthermore, let \( Q^* \in St(d,K) \) be a limiting critical point of \( \ell \). By definition, we have \( \min_{i \in \mathcal{I}(Q^*)} \ell_i(Q^*) > \ell(Q^*) \). Thus, there exists an \( \epsilon > 0 \) such that for all \( Q \in St(d,K) \) with \( \|Q - Q^*\|_F \leq \epsilon \), we have \( \min_{i \in \mathcal{I}(Q^*)} \ell_i(Q) > \ell(Q) \); i.e., \( \mathcal{I}(Q) \subseteq \mathcal{I}(Q^*) \). By adapting the proof of [23, Theorem 3.1] and invoking Theorem 3.1, we conclude that \( \text{dist}(0, \partial \ell(Q)) \geq \eta \ell(Q) - \ell(Q^*) \) for all \( Q \in St(d,K) \) with \( \|Q - Q^*\|_F \leq \epsilon \), where \( \epsilon \) is the constant obtained from Theorem 3.1 by taking \( g = \ell \). This establishes (2.7).

Next, recall that the objective function \( h \) of Problem (2.1) takes the form \( h(P, Q) = -\langle P, X^T Q \rangle + \delta_{B(n,K)}(P) + \delta_{St(d,K)}(Q) \). Let \( Z^* = (P^*, Q^*) \in B(n,K) \times St(d,K) \) be a critical point of \( h \). Furthermore, let \( Z = (P, Q) \in B(n,K) \times St(d,K) \) be such that \( \|Z - Z^*\|_F \leq \epsilon_h \) with \( \epsilon_h < 1 \). Since \( \|P - P^*\|_F \leq \|Z - Z^*\|_F < 1 \) and \( P, P^* \in B(n,K) \), we have \( P = P^* \). Moreover, we have

\[
(3.39) \quad \partial h(P, Q) = \{-XP + N_{St(d,K)}(Q)\} \times \{-XP + N_{B(n,K)}(P)\}
\]

by [2, Proposition 2.1]. This, together with the fact that \( (0, 0) \in \partial h(P^*, Q^*) \), implies that \( 0 \in -XP + N_{B(n,K)}(P^*) \); i.e., \( Q^* \) is a limiting critical point of the function \( Q \mapsto h(P^*, Q) \). Hence, by Theorem 3.1, there exists an \( \eta_h > 0 \) such that

\[
\eta_h|h(Z) - h(Q^*)|^{1/2} = \eta_h |\langle XP^*, Q \rangle + \langle XP^*, Q^* \rangle|^{1/2} \leq \text{dist}(0, \partial h(P^*, Q)) = \text{dist}(0, \partial h(Z)).
\]

This establishes (2.8).

4. Convergence Analysis of PAMe. Our goal in this section is to prove Theorem 2.2, which concerns the convergence behavior of our proposed method PAMe (Algorithm 2.1). Towards that end, we first combine the characterization of the KL exponent for Problem (2.1) in Theorem 2.1 with the abstract convergence results for descent methods in [2, 3] to establish the linear convergence of PAMe to a limiting critical point \( (P^*, Q^*) \) of Problem (2.1). Then, by noting that \( (P^*, Q^*) \) is a solution to certain generalized equations, we obtain a sufficient condition for \( Q^* \) to be a critical point of Problem (1.5).

4.1. Basic Properties of PAMe. To study the convergence behavior of PAMe using the analysis framework developed in [2, 3], a key first step is to show that the iterates \( \{P_k, Q_k\}_{k \geq 0} \) generated by PAMe achieve sufficient decrease and satisfy a relative error (also referred to as safeguard in [39, 25, 24]) condition with respect to some potential function. One immediate choice of the potential function is the objective function \( h \) of Problem (2.1) itself. However, due to the extrapolation step in line 4 of Algorithm 2.1, it is not clear whether the sequence \( \{h(P_k, Q_k)\}_{k \geq 0} \) satisfies the sufficient decrease and relative error conditions. To circumvent this difficulty, let \( \beta \geq 0 \) be a parameter and consider the potential function \( \Psi_{\beta} : \mathbb{R}^{n \times K} \times \mathbb{R}^{d \times K} \rightarrow (-\infty, +\infty) \) given by

\[
\Psi_{\beta}(P, Q, Q') = h(P, Q) + \frac{\beta}{2} ||Q - Q'||_F^2.
\]

We note that similar potential functions have previously been used in the convergence analysis of iterative methods with inertial terms/extrapolation steps; see, e.g., [37, 44,
26, 16]. The following result shows that if the step sizes and extrapolation parameters in PAMe are suitably chosen, then there exists a \( \beta > 0 \) such that the sequence \( \{\Psi_\beta(P^k, Q^k, Q^{k-1})\}_{k \geq 0} \) satisfies the two conditions mentioned earlier.

**Proposition 4.1.** Consider the setting of Theorem 2.2. Let \( C^k = (P^k, Q^k, Q^{k-1}) \) for \( k \geq 0 \). Then, the following hold (recall that \( \beta_* \) is given in Theorem 2.2):

1. The sequence \( \{C^k\}_{k \geq 0} \) is bounded.
2. There exists a constant \( \kappa_1 > 0 \) such that for \( k \geq 0 \),
   \[
   \Psi_\beta(C^{k+1}) - \Psi_\beta(C^k) \leq -\kappa_1 \|C^{k+1} - C^k\|^2_F.
   \]
3. There exists a constant \( \kappa_2 > 0 \) such that for \( k \geq 0 \),
   \[
   \text{dist}(0, \partial \Psi_\beta(C^{k+1})) \leq \kappa_2 \|C^{k+1} - C^k\|_F.
   \]

**Proof.** The proof of (a) is immediate, as \( C^k \in B(n, K) \times St(d, K) \times St(d, K) \) for \( k \geq 0 \) and both \( B(n, K) \) and \( St(d, K) \) are bounded.

To prove (b), we first observe from the updates (2.4) and (2.3) that
\[
\begin{align*}
-\langle P^{k+1}, X^T E^k \rangle + \langle P^k, X^T E^k \rangle &\leq -\frac{\alpha_k}{2} \|P^{k+1} - P^k\|^2_F, \\
-\langle P^{k+1}, X^T Q^{k+1} \rangle + \langle P^k, X^T Q^k \rangle &\leq -\frac{\beta_k}{2} \|Q^{k+1} - Q^k\|^2_F.
\end{align*}
\]

Let \( \Delta_{P}^{k+1} = P^{k+1} - P^k \) and \( \Delta_{Q}^{k+1} = Q^{k+1} - Q^k \) for \( k \geq 0 \). Since \( E^k = Q^k + \gamma_k(Q^k - Q^{k-1}) \), it follows that
\[
\begin{align*}
-\langle P^{k+1}, X^T Q^{k+1} \rangle + \langle P^k, X^T Q^k \rangle &\leq -\frac{\alpha_k}{2} \|P^{k+1} - P^k\|^2_F - \frac{\beta_k}{2} \|Q^{k+1} - Q^k\|^2_F + \gamma_k \langle P^{k+1} - P^k, X^T(Q^k - Q^{k-1}) \rangle \\
&\leq -\frac{\alpha_k}{2} \|\Delta_{P}^{k+1}\|^2_F - \frac{\beta_k}{2} \|\Delta_{Q}^{k+1}\|^2_F + \frac{\gamma_k}{2} \|X\|^2 \|\Delta_{P}^{k+1}\|_F^2 \\
&\leq -\frac{\alpha_k}{2} \|\Delta_{P}^{k+1}\|^2_F - \frac{\beta_k}{2} \|\Delta_{Q}^{k+1}\|^2_F + \frac{\gamma_k}{2\alpha_k} \|X\|^2 \|\Delta_{Q}^{k+1}\|_F^2,
\end{align*}
\]
where the second inequality uses the fact that \( 2\langle A, B \rangle \leq \rho\|A\|^2_F + \|B\|^2_F / \rho \) for any \( \rho > 0 \). Now, by letting \( \kappa_1 = \min \{\alpha_*(1 - \gamma^*)/2, \beta_*/4\} > 0 \), we obtain
\[
\begin{align*}
\Psi_\beta(C^{k+1}) - \Psi_\beta(C^k) &\leq h(P^{k+1}, Q^{k+1}) + \frac{\beta_*}{2} \|Q^{k+1} - Q^k\|^2_F - h(P^k, Q^k) - \frac{\beta_*}{2} \|Q^k - Q^{k-1}\|^2_F \\
&\leq -\frac{\alpha_k}{2} \|\Delta_{P}^{k+1}\|^2_F - \frac{1}{2}(\beta_k - \beta_*) \|\Delta_{Q}^{k+1}\|_F^2 - \frac{1}{2} (\beta_* - \gamma_k \|X\|^2 / \alpha_k) \|\Delta_{Q}^{k+1}\|_F^2 \\
&\leq -\kappa_1 \|\Delta_{P}^{k+1}\|_F^2 + \|\Delta_{Q}^{k+1}\|_F^2 + \|\Delta_{Q}^{k+1}\|_F^2 = -\kappa_1 \|C^{k+1} - C^k\|^2_F,
\end{align*}
\]

as desired.

Lastly, let us prove (c). Again, using the updates (2.4) and (2.3), we have
\[
\begin{align*}
0 &\in -X^T E^k + \alpha_k (P^{k+1} - P^k) + N_{B(n,K)}(P^{k+1}), \\
0 &\in -X^T P^{k+1} + \beta_k (Q^{k+1} - Q^k) + N_{St(d,K)}(Q^{k+1}).
\end{align*}
\]
By adapting the proof of [2, Proposition 2.1], we obtain
\[
\partial \Psi_\beta(P^{k+1}, Q^{k+1}, Q^k) = \{-X^TQ^{k+1} + N_{B(n,K)}(P^{k+1})\} \\
\times \{-XP^{k+1} + \beta_s(Q^{k+1} - Q^k) + N_{St(d,K)}(Q^{k+1})\} \\
\times \{\beta_s(Q^k - Q^{k+1})\}.
\]
This, together with (4.2) and (4.3), implies that
\[
dist^2(0, \partial \Psi_\beta(P^{k+1}, Q^{k+1}, Q^k)) \\
\leq \|X^T(E^k - Q^{k+1}) - \alpha_k \Delta^k_{P^*}\|_F^2 + \|\beta_s - \beta_k\|_F^2 + \|\beta_s\Delta^k_{Q^*}\|_F^2 \\
= \|\beta_s - \beta_k\|_F^2 + \|\beta_s\Delta^k_{Q^*}\|_F^2 + \|\beta_s\Delta^k_{Q^*}\|_F^2 \\
\leq 3\alpha_s^2 \|\Delta^k_{P^*}\|_F^2 + 3\|X\|^2 \|\Delta^k_{Q^*}\|_F^2.
\]
By taking \(\kappa_2 = \left(\max\{3\alpha_s^2, (\beta_s - \beta_k)^2 + 2\alpha_s^2 + 3\|X\|^2\}\right)^{1/2}\), we obtain
\[
dist^2(0, \partial \Psi_\beta(P^{k+1}, Q^{k+1}, Q^k)) \leq \kappa_2^2 \left(\|\Delta^k_{P^*}\|_F^2 + \|\Delta^k_{Q^*}\|_F^2\right) \\
= \kappa_2^2 \|C^{k+1} - C^k\|^2_F,
\]
which implies the desired result.

**4.2. Linear Convergence of PAMe and Properties of Limit Points.** Proposition 4.1 shows that the sequence \(\{C^k\}_{k \geq 0}\) is bounded and satisfies both the sufficient decrease and relative error conditions with respect to the potential function \(\Psi_\beta\). Thus, a natural next step is to study the convergence behavior of the sequence \(\{C^k\}_{k \geq 0}\) with respect to the potential function \(\Psi_\beta\) and then use the result to deduce the convergence behavior of the sequence \(\{(P^k, Q^k)\}_{k \geq 0}\) with respect to the objective function \(h\) of Problem (2.1). To begin, let us prove two technical lemmas. The first establishes a relationship between the limiting critical points of \(h\) and \(\Psi_\beta\).

**Lemma 4.2.** Let \(\beta > 0\) be given. Suppose that \((P, Q, Q') \in B(n, K) \times St(d, K) \times R^{d \times K}\) is a limiting critical point of \(\Psi_\beta\). Then, we have \(Q = Q'\). Moreover, \((P, Q, Q)\) is a limiting critical point of \(\Psi_\beta\) if and only if \((P, Q)\) is a limiting critical point of \(h\).

**Proof.** Recall that
\[
\partial \Psi_\beta(P, Q, Q') = \{-X^TQ + N_{B(n,K)}(P)\} \times \{-XP + \beta(Q - Q') + N_{St(d,K)}(Q)\} \\
\times \{\beta(Q' - Q)\}.
\]
Thus, if \(\beta > 0\) and \((0, 0, 0) \in \partial \Psi_\beta(P, Q, Q')\), then we must have \(Q = Q'\). Moreover, we have \((0, 0, 0) \in \partial \Psi_\beta(P, Q, Q)\) if and only if
\[
(0, 0) \in \{-X^TQ + N_{B(n,K)}(P)\} \times \{-XP + N_{St(d,K)}(Q)\}.
\]
By (3.39), the latter condition holds if and only if \((0, 0) \in \partial h(P, Q)\). 

The second is motivated by the update of the block variable \(P\) in Algorithm 2.1 and shows that a limit point of the sequence \(\{P^k\}_{k \geq 0}\) satisfies certain fixed-point inclusion.

**Lemma 4.3.** Let \(\alpha > 0\) be given. Suppose that the sequences \(\{P^k\}_{k \geq 0}\) and \(\{Y^k\}_{k \geq 0}\) satisfy
\[
P^k \in B(n, K), \ Y^k \in R^{n \times K}, \ P^{k+1} \in sgn(P^k + Y^k/\alpha) \ for \ k = 0, 1, \ldots, \\
P^k \to P^*, \ Y^k \to Y^*.
\]
Then, we have $P^* \in \text{sgn}(P^* + Y^*/\alpha)$. Moreover, for any $Y \in \mathbb{R}^{n \times K}$ satisfying $P^* \in \text{sgn}(Y)$, we have $P^* \in \text{sgn}(P^* + Y/\alpha)$.

Proof. Let $i \in \{1, \ldots, n\}$ and $j \in \{1, \ldots, K\}$ be arbitrary. If $P^*_{ij} + Y^*_{ij}/\alpha = 0$, then $\text{sgn}(P^*_{ij} + Y^*_{ij}/\alpha) = \{-1, 1\}$ by definition. Since $P^* \in B(n, K)$, we have $P^*_{ij} \in \text{sgn}(P^*_{ij} + Y^*_{ij}/\alpha)$. On the other hand, if $P^*_{ij} + Y^*_{ij}/\alpha \neq 0$, then the assumption that $P_k \to P^*$, $Y^k \to Y^*$ implies $\text{sgn}(P^*_{ij} + Y^*_{ij}/\alpha) = \text{sgn}(P^*_{ij} + Y^*_{ij}/\alpha)$ for all sufficiently large $k \geq 0$. As $P^{k+1} \in \text{sgn}(P^k + Y^k/\alpha)$ for $k \geq 0$, we conclude that $P^*_{ij} = \text{sgn}(P^*_{ij} + Y^*_{ij}/\alpha)$. This establishes the first claim.

Now, let $Y \in \mathbb{R}^{n \times K}$ be such that $P^* \in \text{sgn}(Y)$. If $Y_{ij} = 0$, then $P^*_{ij} = \text{sgn}(P^*_{ij})$ trivially. On the other hand, if $Y_{ij} \neq 0$, then $P^*_{ij} = \text{sgn}(Y_{ij}/\alpha) = \text{sgn}(P^*_{ij} + Y_{ij}/\alpha)$. This establishes the second claim.

We are now ready to establish the main convergence result for our proposed method PAMe.

Proof of Theorem 2.2. Recall from Theorem 2.1 that the objective function $h$ of Problem (2.1) has a KL exponent of 1/2 at any of its limiting critical points. Hence, by [23, Theorem 3.6] and Lemma 4.2, for any $\beta > 0$, the potential function $\Psi_{\beta}$ has a KL exponent of 1/2 at any of its limiting critical points. It then follows from [23, Lemma 2.1] that $\Psi_{\beta}$ has a KL exponent of 1/2 at any $(P, Q, Q^*) \in \text{dom}(\partial \Psi_{\beta})$. This, together with the results in Proposition 4.1, allows us to invoke [3, Theorem 2.9] to conclude that under the setting of Theorem 2.2, the sequence $\{C^k\}_{k \geq 0}$ converges to a limiting critical point $(P^*, Q^*, Q^*)$ of the potential function $\Psi_{\beta}$. Moreover, by [2, Theorem 3.4], there exist constants $c > 0$ and $\tau \in (0, 1)$ such that

$$\|P^{k+1} - (P^*, Q^*)\|_F \leq c \tau^k$$

for $k = 0, 1, \ldots$.

It follows from Lemma 4.2 that the sequence $\{(P^k, Q^k)\}_{k \geq 0}$ converges at least linearly to the limiting critical point $(P^*, Q^*)$ of Problem (2.1).

Now, suppose that $\alpha_k = \alpha_*$ for $k \geq 0$ in Algorithm 2.1. According to the update (2.5), the sequence $\{(P^k, Q^k)\}_{k \geq 0}$ satisfies $P^{k+1} = \text{sgn}(P^k + X^T E^k/\alpha_*)$, where $E^k = Q^k + \gamma_k (Q^* - Q^k)$. Since $(P^k, Q^k) \to (P^*, Q^*)$ and $(P^*, Q^*)$ is a limiting critical point of $h$, we have $0 \in -X \text{sgn}(P^* + X^T Q^*/\alpha_*) + N_{\text{St}(d,K)}(Q^*)$ by (3.39) and the result in Lemma 4.3; i.e., $Q^*$ is a solution to the generalized equation (2.9). In particular, noting that $P^* \in B(n, K)$, if $\alpha_*$ satisfies (2.10), then $\text{sgn}(P^* + X^T Q^*/\alpha_*) \subseteq \text{sgn}(X^T Q^*)$. Consequently, we obtain $0 \in -X \text{sgn}(X^T Q^*) + N_{\text{St}(d,K)}(Q^*)$, which, in view of (1.6), shows that $Q^*$ is a critical point of Problem (1.5). Conversely, let $Q$ be a critical point of Problem (1.5) that satisfies $0 \in -X P^* + N_{\text{St}(d,K)}(Q^*)$ and $P^* \in \text{sgn}(X^T Q)$. By Lemma 4.3, we have $P^* \in \text{sgn}(P^* + X^T Q/\alpha_*)$. It then follows that $Q$ is a solution to the generalized equation (2.9).

5. Numerical Results. In this section, we report the numerical performance of different L1-PCA algorithms—including our proposed method PAMe, the standard PAM method (see (2.2) and (2.3)), the method based on FP iterations (FPM) in [35], the method pDCAe in [44], the inertial proximal alternating linearized minimization (iPALM) method in [37], and the Gauss-Seidel-type iPAM (GiPAM) method in [15]—on both synthetic and real-world datasets. We remark that pDCAe is applied to Problem (1.5) only in a formal manner, as the objective function $\ell$ is not of the difference-of-convex type. We do not include the inertial proximal block coordinate descent-type algorithm in [16] in our experiments, as it has essentially the same updates as those of iPAM when applied to Problem (2.1). We also do not include
the exact algorithm in [28] or the algorithm based on BF iterations in [29] in our experiments, as the datasets we used are too large for them to tackle. All the numerical experiments were conducted on a PC running Windows 10 with an Intel® Core™ i5-8600 3.10GHz CPU and 16GB memory. Our code runs in MATLAB R2020a and is available at https://github.com/peng8wang/L1-PCA-PAMe.

5.1. Convergence Performance and Solution Quality. We begin by studying the convergence performance and solution quality of the different algorithms when applied to both synthetic and real-world instances of the L1-PCA problem. The data matrix \( X \in \mathbb{R}^{d \times n} \) in a synthetic instance of the L1-PCA problem is generated according to the fixed effect model in [5]. Specifically, for \( i = 1, \ldots, n \), the \( i \)-th column of \( X \) is given by \( x_i = z_i + e_i \), where \( z_i \in \mathbb{R}^d \) is called a fixed effect and \( e_i \in \mathbb{R}^d \) is a random noise. The model assumes that the fixed effects \( z_1, \ldots, z_n \) lie on a \( K \)-dimensional subspace and satisfy \( \sum_{i=1}^n z_i = 0 \), and that the noise vectors \( e_1, \ldots, e_n \) have entries that are independent and identically distributed (i.i.d.) according to the Laplace distribution with mean 0 and variance \( \sigma^2 \). In our experiments, we generate the fixed effects \( z_1, \ldots, z_n \) in two steps. First, we generate a basis \( U \in \text{St}(d, K) \) of the target \( K \)-dimensional subspace by \( U = Y(Y^T Y)^{-1/2} \), where the entries of \( Y \in \mathbb{R}^{d \times K} \) are i.i.d. according to the standard normal distribution. Then, we set \( z_i = U(a_i - \bar{a}) \) for \( i = 1, \ldots, n \), where each entry of \( a_i \in \mathbb{R}^K \) is i.i.d. according to the standard uniform distribution and \( \bar{a} = \frac{1}{n} \sum_{i=1}^n a_i \). With the above setup, we set \( \sigma = 0.5 \), \( K = 50 \) and generate two synthetic instances whose data matrices have dimensions \( (n, d) = (4000, 2000) \) and \( (n, d) = (2000, 4000) \), respectively. For the real-world instance, we set \( K = 20 \) and extract a data matrix of dimensions \( (n, d) = (15935, 62061) \) from the dataset news20 in LIBSVM [8].

The parameters of the various algorithms are set as follows. To be fair, we employ the same step sizes when updating the block variables \( P \) and \( Q \) in all the PA(L)M-type methods. Specifically, for the two synthetic instances, we set \( (\alpha_k, \beta_k) = (10^{-4}, 10^3) \) and \( (\alpha_k, \beta_k) = (10^{-4}, 10^2) \) for \( k \geq 0 \), respectively; for the real-world instance, we set \( (\alpha_k, \beta_k) = (10^{-3}, 40) \) for all \( k \geq 0 \). The step size for updating the block variable \( Q \) in pDCAe is set as \( \beta_k = 1 \) for \( k \geq 0 \). There is no need to choose any step size for FPM. Next, we specify the extrapolation parameters in the PA(L)M-type methods. For PAMe, we explore three different settings of the extrapolation parameter. In the first one, we simply set the extrapolation parameter as \( \gamma_k = 1 \) for \( k \geq 0 \). The resulting method is denoted by PAMe-1. In the second, we set the extrapolation parameter as \( \gamma_k = \min\{1, \alpha_k \beta_k / 2 \|X\|^2\} \) for all \( k \geq 0 \) according to Theorem 2.2. The resulting method is denoted by PAMe-2. In the last one, we use the fixed restart scheme suggested in [44] with the fixed restart interval \( \bar{T} = 10 \). The resulting method is denoted by PAMe-R. Although the settings for PAMe-1 and PAMe-R may violate the condition in Theorem 2.2, they both work effectively in our experiments. For iPALM, we set the extrapolation parameters when updating the block variables \( P \) and \( Q \) both as \( \frac{k-1}{k+2} \) for \( k \geq 1 \). Such a choice is motivated by the numerical results in [37]. For GiPALM, we set the extrapolation parameters when updating the block variables \( P \) and \( Q \) as \( 1/2 \) and \( 1/4 \) for \( k \geq 0 \), respectively. For pDCAe, we set the extrapolation parameter when updating the block variable \( Q \) using the fixed restart scheme as suggested in [44] with the fixed restart interval \( \bar{T} = 10 \). In each test, we adopt the same starting point for all the algorithms and terminate them when the Frobenius norm of the difference of two consecutive iterates is less than \( 10^{-8} \).

https://www.csie.ntu.edu.tw/~cjlin/libsvmtools/datasets/
To demonstrate the convergence performance of the tested methods, we plot the function value gap $h(P^k, Q^k) - h(P^*, Q^*)$ and the iterate gap $\|Q^k - Q^*\|_F$ against the iteration number for each method in Figures 1 and 2, respectively, where $(P^*, Q^*)$ is the last iterate of the corresponding method. It can be observed that for all the tested methods, both the sequence of function value gaps and the sequence of iterate gaps converge linearly. In particular, the convergence performance of PAMe supports our linear convergence result in Theorem 2.2. Moreover, our numerical results demonstrate that PAMe converges substantially faster than the standard PAM method and also faster than FPM, pDCAe, iPALM, and GiPALM.

To compare the quality of the solution returned by each method, we use the total explained variation (TEV) measure as in [20], which in our setting is given by

$$\text{TEV} = \frac{\sum_{i=1}^{K} q_i^T X^T X q_i}{\sum_{i=1}^{K} \bar{q}_i^T X^T X \bar{q}_i}.$$ 

Here, $q_i$ is the $i$-th column of the solution returned by the tested method and $\bar{q}_i$ is the $i$-th leading eigenvector of $X^T X$. Table 1 summarizes the TEV of the tested methods. It can be observed that the performance of PAMe is comparable to those of the other methods.

5.2. Application to Clustering on a Subspace. As suggested in [13], another way of evaluating the performance of an L1-PCA algorithm is to study the clustering accuracy of a dataset on the subspace found by the algorithm. The procedure is as follows. First, we apply the L1-PCA algorithm to the given dataset to compute a subspace. Then, we project the data points onto the subspace and perform $k$-means...
clustering on the projected points. Finally, we record the fraction of data points that are correctly clustered. In our experiments, we use the real-world datasets a9a, colon-cancer, gisette, rcv1.binary, real-sim, and w8a in LIBSVM [8], whose dimensions can be found in Table 2. In each of these datasets, the data points are given one of two possible labels. These labels serve as the ground truth and naturally divide the data points into two clusters. The dimension $K$ of the subspace used by L1-PCA to capture the variation in the data matrix $X \in \mathbb{R}^{d \times n}$ is chosen such that the fraction of total variance explained by the leading $K$ singular values of $X$ is not less than 0.8; i.e., $K$ satisfies $\sum_{k=1}^{K} \sigma_k^2 \geq 0.8 \sum_{k=1}^{p} \sigma_k^2$, where $p = \min\{n, d\}$ and $\sigma_1 \geq \cdots \geq \sigma_p \geq 0$ are the singular values of $X$. If $p$ is so large (say, $p \geq 10000$) that it becomes too expensive to compute all the singular values of $X$, we simply set $K = 50$.

The step sizes used by the PA(L)M-type methods are listed in Table 2. The step size for updating the block variable $Q$ in pDCAe is given by $\beta_k$ in Table 2. We use the same extrapolation parameters for PAMe-1, PAMe-2, PAMe-R, pDCAe, iPALM, and GiPALM as those in Subsection 5.1.

We terminate the tested methods when either the number of iterations reaches 1000 or the Frobenius norm of the difference of two consecutive iterates is less than $10^{-6}$. To compare the computational efficiency and clustering accuracy of the tested methods, we record their CPU times and ratios of correctly clustered points, averaged over 10 randomly chosen initial points, in Tables 3 and 4, respectively. It can be observed that the CPU times consumed by PAMe-1 and PAMe-R are generally less than those consumed by the other methods on the tested data sets, especially on rcv1.binary, real-sim, and w8a. Moreover, the clustering accuracy of PAMe is comparable to those of the other methods. These demonstrate the efficiency and efficacy of PAMe when performing clustering on a subspace.

### Table 1

<table>
<thead>
<tr>
<th>Method</th>
<th>PAMe-1</th>
<th>PAMe-2</th>
<th>PAMe-R</th>
<th>PAM</th>
<th>FPM</th>
<th>pDCAe</th>
<th>iPALM</th>
<th>GiPALM</th>
</tr>
</thead>
<tbody>
<tr>
<td>syn 1</td>
<td>0.8392</td>
<td>0.8214</td>
<td>0.8271</td>
<td>0.8213</td>
<td>0.8232</td>
<td>0.8230</td>
<td>0.8231</td>
<td>0.8287</td>
</tr>
<tr>
<td>syn 2</td>
<td>0.7746</td>
<td>0.7209</td>
<td>0.7321</td>
<td>0.7209</td>
<td>0.7227</td>
<td>0.7231</td>
<td>0.7215</td>
<td>0.7823</td>
</tr>
<tr>
<td>new20</td>
<td>0.9655</td>
<td>0.9632</td>
<td>0.9629</td>
<td>0.9629</td>
<td>0.9625</td>
<td>0.9625</td>
<td>0.9635</td>
<td>0.9623</td>
</tr>
</tbody>
</table>

### Table 2

<table>
<thead>
<tr>
<th>Dataset</th>
<th>$(n, d)$</th>
<th>$K$</th>
<th>$\alpha_k$</th>
<th>$\beta_k$</th>
</tr>
</thead>
<tbody>
<tr>
<td>a9a</td>
<td>(32561, 123)</td>
<td>6</td>
<td>$10^{-8}$</td>
<td>1</td>
</tr>
<tr>
<td>colon-cancer</td>
<td>(62, 2000)</td>
<td>9</td>
<td>$10^{-6}$</td>
<td>1</td>
</tr>
<tr>
<td>gisette</td>
<td>(6000, 5000)</td>
<td>1</td>
<td>$10^{-6}$</td>
<td>1</td>
</tr>
<tr>
<td>rcv1.binary</td>
<td>(20242, 47236)</td>
<td>50</td>
<td>$10^{-10}$</td>
<td>10</td>
</tr>
<tr>
<td>real-sim</td>
<td>(72309, 20958)</td>
<td>50</td>
<td>$10^{-10}$</td>
<td>1</td>
</tr>
<tr>
<td>w8a</td>
<td>(49749, 300)</td>
<td>30</td>
<td>$10^{-10}$</td>
<td>1</td>
</tr>
</tbody>
</table>

### 6. Concluding Remarks.

In this paper, we proposed a fast iterative method called PAMe to tackle the two-block reformulation (1.3) of the L1-PCA problem (1.2). We proved that the sequence of iterates generated by PAMe converges linearly to a limiting critical point of Problem (1.3) and gave a sufficient condition under which the said limiting critical point yields a critical point of the original problem (1.2). We also demonstrated the efficiency and efficacy of PAMe via numerical experiments.
on both synthetic and real-world datasets. As a key step in establishing the linear convergence of PAMe, we showed that the KL exponent at any limiting critical point of Problems (1.2) and (1.3) is $1/2$. This result not only is significant in its own right but also opens the possibility of establishing strong theoretical guarantees on the convergence behavior of other iterative methods (see, e.g., [12]) for solving (1.2) or (1.3). One possible future direction is to consider the design and analysis of fast iterative methods for other $\ell_1$-norm-based variants of PCA (see, e.g., [22, 42]).

REFERENCES


Recalling that \((A.3)\) In a similar fashion, we get \((A.2)\) using the fact that \((A.1)\) in (3.7) imply that \((A.7)\) in (3.13) and the ordering of the singular values of \(A\) in (3.7) imply that

\[
\|\tilde{A}^T Q_1 - Q_1^T \tilde{A}\|_F^2 \geq \sum_{i=1}^p \|a_{si} (Q_{h_i h_i} - Q_{h_i h_i}^T)\|_F^2 \geq a_{sp}^2 \sum_{i=1}^p \|Q_{h_i h_i} - Q_{h_i h_i}^T\|_F^2.
\]

This, together with (A.1), yields (3.31).

Using (3.7) and (3.13) again, we have

\[
\|\tilde{A}^T Q_1 - Q_1^T \tilde{A}\|_F^2 \geq \sum_{i=1}^p \sum_{j \neq i} \|a_{si} Q_{h_i h_j} - a_{sj} Q_{h_j h_i}\|_F^2 \geq a_{sp}^2 \sum_{i=1}^p \sum_{j \neq i} \|a_{si} Q_{h_i h_j} - a_{sj} Q_{h_j h_i}\|_F^2.
\]

In a similar fashion, we get

\[
\|\tilde{A} Q_1^T - Q_1^{T \tilde{A}}\|_F^2 \geq a_{sp}^2 \sum_{i=1}^p \sum_{j \neq i} \|a_{si} Q_{h_i h_j} - a_{sj} Q_{h_j h_i}\|_F^2.
\]

Recalling that \(\delta_{ij} = \frac{a_{si}}{a_{sj}} - \frac{a_{sj}}{a_{si}}\) for \(i, j \in \{1, \ldots, p\}; i \neq j\) and using (A.2) and (A.3),
we bound

\[
\left( \min_{i,j \in \{1, \ldots, p\}} \delta_{ij}^2 \right) \sum_{i=1}^{p} \sum_{j \neq i} \| Q_{h_i h_j} \|_F^2 \leq \sum_{i=1}^{p} \sum_{j \neq i} \| \delta_{ij} Q_{h_i h_j} \|_F^2
\]

\[
\leq 2 \sum_{i=1}^{p} \sum_{j \neq i} \left( \left\| \frac{a_{s_i}}{a_{s_j}} Q_{h_i h_j} - Q_{h_j h_i}^T \right\|_F^2 + \left\| \frac{a_{s_j}}{a_{s_i}} Q_{h_i h_j} - Q_{h_j h_i}^T \right\|_F^2 \right)
\]

\[
(A.4) \quad \leq \frac{2}{a_{s_p}^2} \left( \| \tilde{A}^T Q_1 - Q_1^T \tilde{A} \|_F^2 + \| \tilde{A} Q_1^T - Q_1 \tilde{A}^T \|_F^2 \right).
\]

Similar to the derivation of (A.1), we have

\[
(A.5) \quad \| \tilde{A} Q_1^T - Q_1 \tilde{A}^T \|_F^2 \leq 2 \left( \| \tilde{A} \|_F^2 - \langle \tilde{A} Q_1^T, Q_1 \tilde{A}^T \rangle \right) \leq 2 \| R(Q) \|_F^2,
\]

where the second inequality follows from (3.30) and the fact that \( \langle \tilde{A} Q_1^T, Q_1 \tilde{A}^T \rangle = \langle Q_1^T \tilde{A}, \tilde{A}^T Q_1 \rangle \). Putting (A.1), (A.4), and (A.5) together, we obtain (3.32).