Received: 15 November 2021 | Revised: 13 April 2022 | Accepted: 2 June 2022

DOI: 10.1002/nla.2459

SPECIAL ISSUE PAPER

WILEY

Local convergence of alternating low-rank optimization methods with overrelaxation

Ivan V. Oseledets[1] | **Maxim V. Rakhuba**[2] | **André Uschmajew**[3]

[1]Skolkovo Institute of Science and Technology, Moscow, Russia

[2]HSE University, Moscow, Russia

[3]Max Planck Institute for Mathematics in the Sciences, Leipzig, Germany

Correspondence
André Uschmajew, Max Planck Institute for Mathematics in the Sciences, 04103 Leipzig, Germany.
Email: uschmajew@mis.mpg.de

Funding information
Russian Science Foundation, Grant/Award Number: 21-71-00119; Ministry of Science and Higher Education of the Russian Federation, Grant/Award Number: 075-10-2021-068

Abstract

The local convergence of alternating optimization methods with overrelaxation for low-rank matrix and tensor problems is established. The analysis is based on the linearization of the method which takes the form of an SOR iteration for a positive semidefinite Hessian and can be studied in the corresponding quotient geometry of equivalent low-rank representations. In the matrix case, the optimal relaxation parameter for accelerating the local convergence can be determined from the convergence rate of the standard method. This result relies on a version of Young's SOR theorem for positive semidefinite 2×2 block systems.

KEYWORDS

ALS, low-rank optimization, overrelaxation, SOR method

1 | INTRODUCTION

We consider a low-rank matrix optimization problem of the form

$$\min_{\operatorname{rank}(X)\leq k} f(X), \tag{1}$$

where $f : \mathbb{R}^{m\times n} \to \mathbb{R}$ is a smooth function on the space of real $m \times n$ matrices. It will be mostly assumed that f is strongly convex. This generic problem appears in a large number of applications, where low-rank matrices serve as nonlinear model classes, such as in matrix recovery, or are employed for reducing numerical complexity when dealing with large-scale matrices.

Since the constraint set admits the explicit parameterization $X = UV^\top$, the problem can be rewritten as

$$\min_{U\in\mathbb{R}^{m\times k},\ V\in\mathbb{R}^{n\times k}} F(U,V) = f(UV^\top). \tag{2}$$

One of the basic methods for solving (2) is the alternating optimization (AO) method, which optimizes the factor matrices U and V in an alternating manner. Conceptually, ignoring the question of unique solvability of subproblems, the method looks as follows:

$$\begin{aligned} U_{\ell+1} &= \operatorname{argmin}_U F(U, V_\ell), \\ V_{\ell+1} &= \operatorname{argmin}_V F(U_{\ell+1}, V). \end{aligned} \tag{3}$$

While this is certainly a standard approach from the viewpoint of nonlinear optimization, where such a scheme is also known as nonlinear Gauss–Seidel method, it is worth emphasizing that the special structure of low-rank problems is

particularly amenable to it. This is due to the bilinearity of the parameterization $UV^\top$, which turns the subproblems of (3) into optimization problems for the same initial function f, but on lower dimensional linear subspaces. Therefore, when f is a quadratic function, this method is called the alternating least squares (ALS) method.

While the study of global convergence of the AO method (3) is usually difficult, its local convergence properties are well-understood.[1-3] The local analysis is based on the fact that the linearized version of the method at a critical point (U_*, V_*) takes the form of a block Gauss–Seidel method for the Hessian $\nabla^2 F(U_*, V_*)$. Due to the intrinsic overparameterization of rank-k matrices by the representation $UV^\top$, the Hessian is at best positive semidefinite, but never positive definite. The Gauss–Seidel error iteration matrix is not contractive on the null space of the Hessian, but it turns out that this problem can be overcome by passing to the corresponding quotient geometry of equivalent low-rank representations $X = UV^\top$. This is possible thanks to an invariance of the AO method under changes of the representation. In fact, this invariance allows one to regard the method (3) as a well-defined iteration on the set of rank-k matrices.

In this work, we consider the acceleration of the local convergence of the AO method (3) by means of overrelaxation. This is a classic idea in nonlinear optimization; see, for example, References 4-7 to mention some early works. Several variants of such acceleration have been proposed for low-rank matrix problems, for example, for matrix completion.[8,9] The basic overrelaxation method that we consider has already been proposed in Reference 10 for the more general low-rank tensor train (TT) format and in the matrix case reads as follows:

$$\begin{aligned} U_{\ell+1} &= (1-\omega)U_\ell + \omega \operatorname{argmin}_U F(U, V_\ell), \\ V_{\ell+1} &= (1-\omega)V_\ell + \omega \operatorname{argmin}_V F(U_{\ell+1}, V). \end{aligned} \tag{4}$$

Here $\omega > 0$ is a relaxation parameter, which sometimes is also called a shift. It can be observed numerically that a suitable choice of the shift significantly improves the convergence speed.

Our goal is to study the local convergence of this iteration for low-rank optimization in a similar spirit as for the plain AO method (3), which corresponds to the case $\omega = 1$. This will be done in Section 2. The linearization of (4) at a critical point (U_*, V_*) leads to a 2×2 block SOR method for the Hessian $\nabla^2 F(U_*, V_*)$. Using the fact that for $0 < \omega < 2$ and a positive semidefinite Hessian with positive definite block diagonal the SOR error iteration is contractive on any subspace complementary to the null space of the Hessian, we obtain local convergence results for this range of ω. This result is stated in Theorem 1.

It is then natural to ask for the optimal shift ω achieving the fastest local convergence rate, which requires to minimize the spectral radius of the SOR error iteration matrix. For positive definite 2×2 block systems this can be achieved using a well-known theorem of Young. It is however possible to adjust the arguments to the positive semidefinite case, as will be done in Lemma 1. This yields the expected, yet not entirely trivial, formula for the asymptotically optimal shift in terms of the convergence rate of the standard AO method with $\omega = 1$. The result is stated in Theorem 2. In practice, this means that the optimal shift can be estimated adaptively and at practically zero cost from the observed convergence rate of the standard method.

Of course, overrelaxation can also be applied to AO methods for low-rank tensor optimization. In Section 3, we focus on the low-rank TT format as in Reference 10. Like low-rank matrix factorization, the TT decomposition is subject to an intrinsic overparameterization which can be described by a simple group action in parameter space, but leads to formally semidefinite Hessians in critical points. By passing to suitable quotient spaces, the local convergence of the method can be established in essentially the same way as for low-rank matrices (Theorem 3). However, a main difference to the matrix case is that the formula for the optimal shift cannot be made rigorous under reasonable assumptions, although it can still serve as a useful heuristic.

In Section 4, we report on some numerical experiments that illustrate the advantage of using shifts in low-rank AO methods, and validate our theoretical findings regarding the optimal shift in the matrix case. We also demonstrate the adaptive procedure for choosing an almost optimal shift based on the observed convergence rate of the standard method.

2 | ALTERNATING OPTIMIZATION WITH RELAXATION FOR LOW-RANK MATRICES

In this section, we first formalize the basic AO iteration (3) for low-rank matrix problems and recall some of its basic properties. We then proceed to the method with overrelaxation, establish its local convergence and determine the optimal shift parameter.

2.1 | Standard AO method

Consider the scheme (3) and assume f to be strongly convex. Then the first argmin is uniquely defined if $\operatorname{rank}(V_\ell) = k$, since it corresponds to minimizing the strongly convex function f on a linear subspace of $\mathbb{R}^{m\times n}$ which is the image of the injective linear map $U \mapsto UV_\ell^\top$. Likewise, the second argmin is well-defined if $\operatorname{rank}(U_{\ell+1}) = k$ and returns the unique minimum of f on the linear subspace $V \mapsto U_{\ell+1}V^\top$. Therefore, in some open and dense subsets both argmins define smooth maps $\hat{S}_1$ and $\hat{S}_2$, respectively, such that

$$U_{\ell+1} = \hat{S}_1(V_\ell), \quad V_{\ell+1} = \hat{S}_2(U_{\ell+1}). \tag{5}$$

One full update of the method then takes the form of a fixed point iteration

$$\begin{pmatrix} U_{\ell+1} \\ V_{\ell+1} \end{pmatrix} = S\begin{pmatrix} U_\ell \\ V_\ell \end{pmatrix} := \begin{pmatrix} \hat{S}_1(V_\ell) \\ \hat{S}_2(\hat{S}_1(V_\ell)) \end{pmatrix}. \tag{6}$$

The map S is well-defined and smooth on any open subset of

$$\mathcal{D} = \{(U, V) \in \mathbb{R}^{m\times k} \times \mathbb{R}^{n\times k} : V \text{ and } \hat{S}_1(V) \text{ have full column rank } k\}.$$

In particular, any critical point (U_*, V_*) of F in (2) for which U_* and V_* have full column rank belongs to $\mathcal{D}$ and is a fixed point of S. To see this, note that $U \mapsto F(U, V_*) = f(UV_*^\top)$ is strongly convex since $U \mapsto UV_*^\top$ is injective. Since U_* is a critical point of that function, it is the global minimum and hence $\hat{S}_1(V_*) = U_*$. The argument for V_* is the same. Such a critical point of F possesses an open neighborhood in $\mathcal{D}$ in which S is well-defined and smooth. Conversely, any fixed point $(U_*, V_*) \in \mathcal{D}$ of S must be a critical point of F since it implies that U_* is the global minimum of $U \mapsto F(U, V_*)$ and V_* is the global minimum of $V \mapsto F(U_*, V)$. Hence the partial gradients $\nabla_U F(U_*, V_*)$ and $\nabla_V F(U_*, V_*)$ are both zero.

By passing from the initial constrained problem (1) to the factorized problem (2), we formally introduced an ambiguity arising from the fact that the factorization $X = UV^\top$ of a rank-k matrix is not unique. In particular, $X = UAA^{-1}V^\top$ for any invertible $k \times k$ matrix A so that the function F has level sets of at least dimension k^2 (when U, V have full column rank). Therefore, a fixed point $(U_*, V_*) \in \mathcal{D}$ of S is never locally unique. However, this issue of nonuniqueness is only a formal one since one is ultimately interested in the sequence of generated matrices $X_\ell = U_\ell V_\ell^\top$. Assuming $\operatorname{rank}(X_\ell) = k$ for all ℓ, this sequence is not affected by any reparameterization $(U_\ell, V_\ell) \to (U_\ell A_\ell, V_\ell A_\ell^{-T})$ with invertible matrices A_ℓ during the iteration. This is due to the invariance properties

$$\begin{aligned} \hat{S}_1(VA^{-T}) &= \hat{S}_1(V)A, \\ \hat{S}_2(UA) &= \hat{S}_2(U)A^{-T} \end{aligned} \tag{7}$$

of the maps $\hat{S}_1$ and $\hat{S}_2$, which hold whenever U and V have full column rank. To see this, let $U_+ = \hat{S}_1(V)$ and $\hat{U}_+ = \hat{S}_1(VA^{-T})$. Then by construction $U_+V^\top$ and $\hat{U}_+A^{-1}V^\top$ are the unique minimizers of f on the linear subspaces $\{UV^\top : U \in \mathbb{R}^{m\times k}\}$ and $\{UA^{-1}V^\top : U \in \mathbb{R}^{m\times k}\}$, respectively. Obviously both spaces are equal, hence $U_+V^\top = \hat{U}_+A^{-1}V^\top$. Since V has full column rank, we obtain $U_+ = \hat{U}_+A^{-1}$, the first identity in (7). The argument for $\hat{S}_2$ is analogous.

The above invariance of the AO method allows us to interpret it as a method

$$X_{\ell+1} = \mathbf{S}(X_\ell)$$

in the full matrix space $\mathbb{R}^{m\times n}$, or more precisely on the subset of matrices of rank at most k. This viewpoint has been taken in Reference 3 and will be helpful in this work, too. From an algorithmic perspective, the AO viewpoint (6) is more useful since it operates on the smaller matrices U and V instead of the full matrix X. Furthermore, the invariance with respect to the described change of parameterization allows for a robust implementation of the AO method by orthogonalizing the columns of U_ℓ and V_ℓ after every partial update, without affecting the generated sequence X_ℓ of matrices. This method is a special case of Algorithm 1 with $\omega = 1$.

2.2 | Overrelaxation

Instead of (5), we now consider the more general update rule with a shift,

$$\begin{aligned} U_{\ell+1} &= (1-\omega)U_\ell + \omega \hat{S}_1(V_\ell), \\ V_{\ell+1} &= (1-\omega)V_\ell + \omega \hat{S}_2(U_{\ell+1}), \end{aligned} \tag{8}$$

which corresponds to (4). For $\omega = 1$, this iteration equals the standard AO method (5). By defining the map

$$S_\omega \begin{pmatrix} U \\ V \end{pmatrix} = (1-\omega)\begin{pmatrix} U \\ V \end{pmatrix} + \omega \begin{pmatrix} \hat{S}_1(V) \\ \hat{S}_2((1-\omega)U + \omega \hat{S}_1(V)) \end{pmatrix}, \tag{9}$$

we can write (8) as a nonlinear fixed point iteration

$$\begin{pmatrix} U_{\ell+1} \\ V_{\ell+1} \end{pmatrix} = S_\omega \begin{pmatrix} U_\ell \\ V_\ell \end{pmatrix}. \tag{10}$$

The map S_ω is well-defined and smooth on any open subset of

$$\mathcal{D}_\omega = \{(U,V) \in \mathbb{R}^{m\times k} \times \mathbb{R}^{n\times k} \,:\, V \text{ and } (1-\omega)U + \omega \hat{S}_1(V) \text{ have full column rank}\}.$$

Note that $(U_*, V_*) \in \mathcal{D}_\omega$ is a fixed point of S_ω if and only if $(U_*, V_*) \in \mathcal{D}$ and (U_*, V_*) is a fixed point of S. In particular, any critical point (U_*, V_*) of F in (2) such that U_* and V_* have full column rank belongs to $\mathcal{D}_\omega$ and is a fixed point of S_ω. Moreover, any such critical point possesses an open neighborhood in $\mathcal{D}_\omega$ such that S_ω is well-defined and smooth on this neighborhood. Again, the converse is also true, that is, a fixed point $(U_*, V_*) \in \mathcal{D}_\omega$ of S_ω is a critical point of F.

The iteration (10) exhibits the same invariance under changes of representation as the standard AO method. Let $(U,V) \in \mathcal{D}_\omega$ and $(U_+, V_+) = S_\omega(U,V)$, then from (9) and (7) one verifies

$$S_\omega \begin{pmatrix} UA \\ VA^{-T} \end{pmatrix} = \begin{pmatrix} U_+A \\ V_+A^{-T} \end{pmatrix}. \tag{11}$$

Therefore, the generated sequence $X_\ell = U_\ell V_\ell^\top$ is essentially (if $\operatorname{rank}(X_\ell) = k$ for all ℓ) invariant under changes of representation during the iteration. In particular, QR decomposition can be used in numerical implementation for keeping the argmin problems well-conditioned. The resulting method is denoted in Algorithm 1.

Algorithm 1. Low-rank AO with overrelaxation and QR

Data: $V_0 \in \mathbb{R}^{n\times k}$, relaxation parameter ω
for $\ell = 0, 1, 2, \ldots$ **do**
  $U \leftarrow \operatorname{argmin}_{\hat{U}\in\mathbb{R}^{m\times k}} F(\hat{U}, V_\ell)$ $U \leftarrow (1-\omega)U_\ell + \omega U$, $\quad U = Q_1 R_1$ $V \leftarrow \operatorname{argmin}_{\hat{V}\in\mathbb{R}^{n\times k}} F(Q_1, \hat{V})$ $V \leftarrow (1-\omega)V_\ell R_1^\top + \omega V$,
  $V = Q_2 R_2$ $U_{\ell+1} := Q_1 R_2^\top$, $\quad V_{\ell+1} := Q_2$
end

Since the goal of this work is a local convergence analysis of the fixed point iteration (10), it is important to observe that the invariance under the group action also carries over to the asymptotic linear convergence rate of the method, which depends on the eigenvalues of the derivative $S'_\omega(U_*, V_*)$ at a fixed point $(U_*, V_*) \in \mathcal{D}_\omega$. To see this invariance, it is convenient to introduce the corresponding group action θ_A of $\mathrm{GL}(k)$ acting on $\mathbb{R}^{m\times k} \times \mathbb{R}^{n\times k}$ via

$$A \mapsto \theta_A \cdot \begin{pmatrix} U \\ V \end{pmatrix} = \begin{pmatrix} UA \\ VA^{-T} \end{pmatrix}. \tag{12}$$

In this notation (11) reads

$$S_\omega\left(\theta_A \cdot \begin{pmatrix} U \\ V \end{pmatrix}\right) = \theta_A \cdot S_\omega \begin{pmatrix} U \\ V \end{pmatrix}. \tag{13}$$

For fixed A, (12) defines an invertible linear map θ_A on $\mathbb{R}^{m\times k} \times \mathbb{R}^{n\times k}$ with $\theta_A^{-1} = \theta_{A^{-1}}$. Differentiating both sides of (13) it then follows that

$$S'_\omega \begin{pmatrix} U_* A \\ V_* A^{-T} \end{pmatrix} = \theta_A \cdot S'_\omega \begin{pmatrix} U_* \\ V_* \end{pmatrix} \cdot \theta_A^{-1}. \tag{14}$$

This shows that $S'_\omega(U_*A, V_*A^{-T})$ has the same eigenvalues as $S'_\omega(U_*, V_*)$ and allows us to study the local convergence rate of the iteration (10) at any particular fixed point (U_*, V_*).

As for the standard AO method, the invariance property allows for an interpretation of the method (10) as an iteration

$$X_{\ell+1} = \mathbf{S}_\omega(X_\ell) \tag{15}$$

on the manifold

$$\mathcal{M}_k = \{X \in \mathbb{R}^{m\times n} : \operatorname{rank}(X) = k\},$$

where $\mathbf{S}_\omega : \mathcal{O} \subseteq \mathcal{M}_k \to \mathbb{R}^{m\times n}$ is defined through

$$\mathbf{S}_\omega(X) = \tau(S_\omega(U, V)), \quad X = UV^\top, \tag{16}$$

with the map

$$\tau(U, V) = UV^\top.$$

Here the domain of definition $\mathcal{O}$ of $\mathbf{S}_\omega$ should be contained in the image of $\mathcal{D} \cap \mathcal{D}_\omega$ under τ. In particular, let $(U_*, V_*) \in \mathcal{D}$ be a fixed point of the map S (the standard AO method), that is, a critical point of F. Then $\mathbf{S}_\omega$ is well-defined and smooth in some open neighborhood $\mathcal{O} \subseteq \mathcal{M}_k$ of $X_* = U_*V_*^\top$ and X_* is a fixed point of $\mathbf{S}_\omega$. This manifold viewpoint will be useful in the local convergence analysis conducted in the next section.

2.3 | Local convergence

Let $X_* = U_*V_*^\top \in \mathcal{M}_k$ be a fixed point of $\mathbf{S}_\omega$. Then $\mathbf{S}_\omega$ locally maps to $\mathcal{M}_k$ and thus the derivative $\mathbf{S}'(X_*)$ maps the tangent space $T_{X_*}\mathcal{M}_k$ to itself. This provides the following local convergence criterion.

Proposition 1. *Let f be strongly convex and (U_*, V_*) be a critical point of F in (2) with U_*, V_* having full column rank k. Then (U_*, V_*) is a fixed point of S_ω and $X_* = U_*V_*^\top \in \mathcal{M}_k$ is a fixed point of $\mathbf{S}_\omega$. Let $\mathbf{S}'_\omega(X^*) : T_{X_*}\mathcal{M}_k \to \mathbb{R}^{m\times n}$ denote the derivative of $\mathbf{S}_\omega$ at X_*, and $\mathbf{P}_{X_*} : \mathbb{R}^{m\times n} \to T_{X_*}\mathcal{M}_k$ the tangent space projection. If for the spectral radius*

$$\rho_\omega = \rho(\mathbf{P}_{X_*}\mathbf{S}'_\omega(X_*)) < 1, \tag{17}$$

then for $X_0 = U_0V_0^\top$ close enough to X_* the iterates $X_\ell = U_\ell V_\ell^\top$ generated by Algorithm 1 converge to X_* at an asymptotic linear rate ρ_ω.

To study the convergence criterion in more detail we investigate $\mathbf{S}'_\omega(X_*)$. For this we repeat some well-known computations. We first consider the map S_ω in parameter space. From (9), its derivative at (U_*, V_*) takes the form of a block matrix

$$S'_\omega \begin{pmatrix} U_* \\ V_* \end{pmatrix} = (1-\omega)\begin{pmatrix} I & 0 \\ 0 & I \end{pmatrix} + \omega \begin{pmatrix} 0 & \hat{S}'_1(V_*) \\ (1-\omega)\hat{S}'_2(U_*) & \omega\hat{S}'_2(U_*)\hat{S}'_1(V_*) \end{pmatrix},$$

where we have used that $(1-\omega)U_* + \omega\hat{S}_1(V_*) = U_*$. Setting

$$L = \begin{pmatrix} 0 & 0 \\ \hat{S}_2'(U_*) & 0 \end{pmatrix}, \quad R = \begin{pmatrix} 0 & \hat{S}_1'(V_*) \\ 0 & 0 \end{pmatrix}$$

one then verifies the identity

$$S_\omega' \begin{pmatrix} U_* \\ V_* \end{pmatrix} = (I - \omega L)^{-1}[(1-\omega)I + \omega R]. \tag{18}$$

This linear operator can be interpreted as a block SOR error iteration matrix for the Hessian

$$H = \begin{pmatrix} \nabla_{UU}F(U_*, V_*) & \nabla_{UV}F(U_*, V_*) \\ \nabla_{VU}F(U_*, V_*) & \nabla_{VV}F(U_*, V_*) \end{pmatrix}$$

of F at (U_*, V_*), written in 2×2 block form. To see this, consider the usual decomposition

$$H = D + E + E^\top$$

with

$$D = \begin{pmatrix} \nabla_{UU}F(U_*, V_*) & 0 \\ 0 & \nabla_{VV}F(U_*, V_*) \end{pmatrix}, \quad E = \begin{pmatrix} 0 & 0 \\ \nabla_{VU}F(U_*, V_*) & 0 \end{pmatrix}.$$

Assuming that the block diagonal part D is invertible and differentiating the equations

$$\nabla_U F(\hat{S}_1(V), V) = 0, \quad \nabla_V F(U, \hat{S}_2(U)) = 0 \tag{19}$$

(which implicitly define $\hat{S}_1$ and $\hat{S}_2$), one finds that

$$\hat{S}_1'(V_*) = -[\nabla_{UU}F(U_*, V_*)]^{-1}\nabla_{UV}F(U_*, V_*)$$

and

$$\hat{S}_2'(U_*) = -[\nabla_{VV}F(U_*, V_*)]^{-1}\nabla_{VU}F(U_*, V_*).$$

In other words, $L = -D^{-1}E$, $R = -D^{-1}E^\top$, and

$$L + R = I - D^{-1}H.$$

Using the expressions for L and R in (18), one obtains the alternative formula

$$S_\omega' \begin{pmatrix} U_* \\ V_* \end{pmatrix} = T_\omega := I - N_\omega^{-1}H, \quad N_\omega = \frac{1}{\omega}D + E. \tag{20}$$

We see from (20) that $S_\omega'(U_*, V_*)$ equals the error iteration matrix T_ω for the two-block SOR method for H. It is well-known that T_ω has spectral radius less than one if $0 < \omega < 2$ and H is positive definite. However, the latter is never the case here. Since ∇F is constantly zero on the orbit $\theta_A \cdot (U_*, V_*)$ (this follows from the chain rule by differentiating $F = F \circ \theta_A$ for fixed A), the Hessian at critical points $(U_*, V_*) \in \mathcal{D}$ has at least a k^2-dimensional kernel $\ker H$ containing the tangent space to the orbit. On $\ker H$ the matrix T_ω acts as identity. However, if $0 < \omega < 2$, D is positive definite and H is at least positive semidefinite, then by classic results it still holds that T_ω is a contraction on any invariant subspace

complementary to $\ker H$; see, for example, Reference 11, Section 3 or Reference 12, Corollary 2.1. Specifically, as follows from Reference 11, the space

$$\mathcal{W}_\omega = N_\omega^{-1}(\ker H)^\perp \tag{21}$$

is an invariant subspace of T_ω, which splits the parameter space into a direct sum*

$$\mathbb{R}^{m\times k} \times \mathbb{R}^{n\times k} = \ker H \oplus \mathcal{W}_\omega, \tag{22}$$

and T_ω is a contraction on $\mathcal{W}_\omega$.

At this point, we can exploit that we are actually interested in the convergence of the products $X_\ell = U_\ell V_\ell^\top$. Under the assumption that $\ker H$ equals the tangent space to the orbit $\theta_A \cdot (U_*, V_*)$, any complementary subspace, such as $\mathcal{W}_\omega$, satisfies the properties of a so-called horizontal space for the quotient manifold structure of $\mathcal{M}_k$. For us, this means the following.

Proposition 2. *Assume $X_* = U_* V_*^\top$ has rank k, H is positive semidefinite,* $\dim(\ker H) = k^2$ *and a decomposition (22) holds. Then the map $\tau(U, V) = UV^\top$ is a local diffeomorphism between a (relative) neighborhood of (U_*, V_*) in $(U_*, V_*) + \mathcal{W}_\omega$ and a neighborhood of X_* in the embedded submanifold $\mathcal{M}_k \subseteq \mathbb{R}^{m\times n}$.*

Proof. The proof can be given without particular reference to quotient manifolds, but assuming knowledge that $\mathcal{M}_k$ is a smooth embedded submanifold of dimension $mk + nk - k^2$,[13(Example8.14)] and τ is a local submersion on $\mathcal{M}_k$ in a neighborhood of (U_*, V_*) (it is not difficult to verify that $\tau'(U_*, V_*)$ has rank $mk + nk - k^2$). Then since τ is constant on the θ_A-orbit of (U_*, V_*), its derivative vanishes on the tangent space to that orbit at (U_*, V_*), which is of dimension k^2. We already noted that $\ker H$ contains that tangent space, so if $\dim(\ker H) = k^2$, then $\tau'(U_*, V_*)$ vanishes on $\ker H$. Hence, due to (22), $\tau'(U_*, V_*)$ must be a bijection between $\mathcal{W}_\omega$ and the tangent space $T_{X_*}\mathcal{M}_k$. The assertion follows by the inverse function theorem. ∎

We are now in the position to formulate a local convergence result for the iteration (15).

Theorem 1. *Let f be strongly convex and (U_*, V_*) be a critical point of F in (2) with U_*, V_* having full column rank k. Assume that the Hessian $H = \nabla^2 F(U_*, V_*)$ is positive semidefinite and* $\dim(\ker H) = k^2$. *Fix $0 < \omega < 2$. Then for $X_0 = U_0 V_0^\top$ close enough (this may depend on ω) to $X_* = U_* V_*^\top$ Algorithm 1 is well-defined and the iterates $X_\ell = U_\ell V_\ell^\top$ converge to X_* at an asymptotic linear rate*

$$\rho_\omega = \limsup_{\ell\to\infty} \|X_\ell - X_*\|^{1/\ell} < 1,$$

where ρ_ω is the spectral radius of T_ω on $\mathcal{W}_\omega$.

The convergence rate ρ_ω is determined in the next section. We stated the above result separately because its proof can be easily generalized to alternating optimization methods for low-rank tensor formats that admit a similar invariance under a group action. This will be outlined for the TT format in Section 3.

Proof. Since V_* has full column rank, the linear map $U \mapsto UV_*^\top$ is injective and hence the restricted map $U \mapsto F(U, V_*) = f(UV_*^T)$ is strongly convex. Therefore, $\nabla_{UU} F(U_*, V_*)$ is positive definite. Likewise, $\nabla_{VV} F(U_*, V_*)$ is positive definite, so that the block diagonal part D of H is positive definite. As a result, the decomposition (22) of the parameter space applies and T_ω is a contraction on its invariant subspace $\mathcal{W}_\omega$. In a neighborhood of X_* in $\mathcal{M}_k$ the map $\mathbf{S}_\omega$ in (16) can be written as

$$\mathbf{S}_\omega = \tau \circ S_\omega \circ \tau^{-1},$$

where we have restricted τ to the affine subspace $(U_*, V_*) + \mathcal{W}_\omega$. Therefore, by chain rule,

$$\mathbf{S}'_\omega(X_*) = [\tau'(U_*, V_*)] \circ T_\omega \circ [\tau'(U_*, V_*)]^{-1}. \tag{23}$$

By Proposition 2, the derivative $\tau'(U_*, V_*)$ is an isomorphism between $\mathcal{W}_\omega$ and $T_{X_*}\mathcal{M}_k$. Due to (23), this implies that the convergence criterion (17) in Proposition 1 is satisfied. ∎

Remark 1. The assumptions that H is positive semidefinite and $\dim(\ker H) = k^2$ already imply by themselves that D is positive definite. Indeed, as noted in the proof of Proposition 2, $\dim(\ker H) = k^2$ means that $\ker H$ equals the tangent space to the θ_A-orbit of (U_*, V_*), which however does not contain elements of the form $(U, 0)$ or $(0, V)$ (this can be seen from (12)). This allows to define a nonlinear SOR process in a neighborhood of such a critical point based on the implicit definitions (19) of $\hat{S}_1$ and $\hat{S}_2$ even when f is not strongly convex; compare Reference 6, Theorem 10.3.5.

To get a better intuition for the assumptions in the theorem, it is useful to write the Hessian as a bilinear form

$$\nabla^2 F(U_*, V_*)[h, h] = \langle \tau'(U_*, V_*)[h], \nabla^2 f(X_*) \cdot \tau'(U_*, V_*)[h] \rangle + \langle \nabla f(X_*), \tau''(U_*, V_*)[h, h] \rangle,$$

where $h = (\delta U, \delta V)$. If f is strictly convex, then the first term is nonnegative and equal to zero if and only if $\tau'(U_*, V_*)[h] = 0$, that is, if h is in the tangent space to the θ_A-orbit at (U_*, V_*). Thus, an important situation in which the assumptions of the theorem are satisfied is when $\nabla f(X_*) = 0$, that is, when $X_* = U_* V_*^\top$ is a global minimum of f. In Section 4.1, we conduct some numerical experiments for a matrix completion problem (32) admitting such a global minimum $X_* = U_* V_*^\top$ with $\nabla f(X_*) = 0$. In this application, however, f is only convex, but not strictly convex. Then in order to satisfy the assumptions of the theorem at X_* one would need that the tangent space $T_{X_*}\mathcal{M}_k$ (the image of $\tau'(U_*, V_*)$) does not intersect the null space of $\nabla^2 f(X_*) = P_\Omega$, but we will not investigate this condition in detail.

2.4 | Asymptotically optimal relaxation

It is well-known that under certain assumptions the relaxation parameter ω in the linear SOR method can be optimized using a theorem of Young; see, for example, Reference 14, Section 6.2 or Reference 15, Section 4.6.2. This theory is usually presented for positive definite systems. However, for 2×2 block systems it is possible to adjust the arguments to the positive semidefinite case.

Lemma 1. *Let $H = D + E + E^\top \in \mathbb{R}^{p \times p}$ be a positive semidefinite 2×2 block matrix with positive definite block diagonal D and such that $\frac{1}{\omega}D + E$ is invertible for any $0 < \omega < 2$. Assume $q = \dim(\ker H) < p/2$. Let $\sigma(I - D^{-1}H)$ denote the spectrum of $I - D^{-1}H$, then*

$$\beta := \max\{|\mu| : \mu \in \sigma(I - D^{-1}H) \setminus \{\pm 1\}\} < 1.$$

The matrix $T_\omega = I - N_\omega^{-1}H$, where $N_\omega = \frac{1}{\omega}D + E$, induces a decomposition (22) into two invariant subspaces and the spectral radius ρ_ω of T_ω on $\mathcal{W}_\omega$ equals

$$\rho_\omega = \begin{cases} 1 - \omega + \frac{1}{2}\omega^2\beta^2 + \omega\beta\sqrt{1 - \omega + \frac{1}{4}\omega^2\beta^2}, & \text{if } 0 < \omega \le \omega_{\text{opt}}, \\ \omega - 1, & \text{if } \omega_{\text{opt}} \le \omega < 2, \end{cases}$$

where

$$\omega_{\text{opt}} = \frac{2}{1 + \sqrt{1 - \beta^2}} > 1. \tag{24}$$

The value of ρ_ω is minimal for $\omega = \omega_{\text{opt}}$. It holds that $\beta^2 = \rho_1$ is the spectral radius for the standard AO method with $\omega = 1$ (on its invariant subspace $\mathcal{W}_1$).

Proof. The decomposition (22) into invariant subspaces has already been verified (see Endnote *). We follow the arguments in the proof of Theorem 4.27 in Reference 15. There it is shown that the eigenvalues μ of $I - D^{-1}H$ and λ of T_ω are related via

$$\lambda = 1 - \omega + \frac{1}{2}\omega^2\mu^2 \pm \omega\mu\sqrt{1 - \omega + \frac{1}{4}\omega^2\mu^2}. \tag{25}$$

Indeed, note that under the given assumptions $I - D^{-1}H = -D^{-1}(E + E^\top)$ is a two-cyclic matrix and has only real eigenvalues, of which the nonzero ones come in pairs $\pm\mu$. By (25), both μ and $-\mu$ create a same pair of eigenvalues λ.

Eigenvalues $\mu = 1$ of $I - D^{-1}H$ must belong to eigenvectors in $\ker H$. Therefore, $\mu = 1$ and $\mu = -1$ both have multiplicity q. They yield eigenvalues $\lambda = 1$ and $\lambda = (1-\omega)^2$ of T_ω. Since eigenvectors of T_ω with $\lambda = 1$ must belong to $\ker H$, we conclude that the restriction of T_ω to the invariant subspace $\mathcal{W}_\omega$ has an eigenvalue $(1-\omega)^2$ and its other eigenvalues are generated from formula (25) with $|\mu| \neq 1$. Since $2q < p$ such μ must exist. Rewriting the eigenvalue equation $(I - D^{-1}H)x = \mu x$ in the two ways

$$Hx = (1-\mu)Dx, \quad (2D - H)x = (1+\mu)Dx,$$

and using a special property of 2×2 block matrices that H and $2D - H$ have the same eigenvalues, we obtain $|\mu| \leq 1$ since both H and $2D - H$ are positive semidefinite and D is positive definite. This shows $\beta < 1$.

Consider eigenvalues μ of $I - D^{-1}H$ with $|\mu| < 1$. If $\omega \geq \omega_{\text{opt}}$, then for such μ the expression under the square root in formula (25) is always negative. Hence they generate pairs of conjugate complex eigenvalues λ, but one verifies that they all have the same modulus $|\lambda| = |1-\omega|$, independent from $|\mu|$. Clearly $|1-\omega| > (1-\omega)^2$ so that the asserted formula for ρ_ω is proven for $\omega \geq \omega_{\text{opt}}$. When $0 < \omega < \omega_{\text{opt}}$ the expression under the square root in (25) may be negative or not. If it is negative, we have already seen that $|\lambda| = |1-\omega|$ is generated. If it is nonnegative, which in particular is the case for $\mu = \pm\beta$, the corresponding λ with the larger absolute value is

$$\lambda = 1 - \omega + \frac{1}{2}\omega^2\mu^2 + \omega\,|\mu|\,\sqrt{1 - \omega + \frac{1}{4}\omega^2\mu^2}$$

(since the sum before $\pm$ in (25) then is nonnegative, too). This expression is maximized for $\mu = \pm\beta$ and also is then larger than $|1-\omega|$ on the interval $0 < \omega < \omega_{\text{opt}}$. The statements of the lemma follow. ∎

Remark 2. In the setting of the lemma one always has $q = \dim(\ker H) \leq p/2$, but (if p is even) equality $q = 2p$ could in principle hold. It is then interesting to note that in this case $\rho_\omega = (1-\omega)^2$, which is minimized for $\omega = 1$, yielding a superlinear convergence rate. However, this case is not relevant in the context of our work, where $p = km + kn$ with a rank $k < \min(m, n)$. In the following theorem, we assume $q = k^2$ so that $q < p/2$ is satisfied.

Applying Lemma 1 in the context of Theorem 1 immediately provides our main result on the asymptotically optimal choice of the shift ω for Algorithm 1.

Theorem 2. *Let f be strongly convex and (U_*, V_*) be a critical point of F in (2) with U_*, V_* having full column rank $k < \min(m, n)$. Assume that the Hessian $H = \nabla^2 F(U_*, V_*)$ is positive semidefinite and $\dim(\ker H) = k^2$. Let $\rho_1 < 1$ be the asymptotic linear convergence rate of the standard AO method with $\omega = 1$. Fix $0 < \omega < 2$. Then for $X_0 = U_0 V_0^\top$ close enough (this may depend on ω) to $X_* = U_* V_*^\top$ Algorithm 1 is well-defined and the iterates $X_\ell = U_\ell V_\ell^\top$ converge to X_* at an asymptotic linear rate $\rho_\omega < 1$ given in Lemma 1 with $\beta^2 = \rho_1$. The optimal asymptotic rate is achieved for*

$$\omega_{\text{opt}} = \frac{2}{1 + \sqrt{1 - \beta^2}}.$$

In practice, the simplest approach for approximating ω_{opt} adaptively is by running the standard method with $\omega = 1$ and estimating $\beta^2 \approx \rho_1$ based on its numerically observed convergence rate. The efficiency of this approach will be illustrated in Section 4.

3 | LOW-RANK TENSOR PROBLEMS

Clearly, the nonlinear SOR method can be applied to functions with more than two block variables. In low-rank tensor optimization one frequently considers problems of the form

$$\min F(U^1, \ldots, U^D) = f(\tau(U^1, \ldots, U^D)), \tag{26}$$

where now f is a smooth function on a tensor space $\mathbb{R}^{n_1 \times \cdots \times n_d}$, and $\tau : \mathcal{V}_1 \times \cdots \times \mathcal{V}_D \to \mathbb{R}^{n_1 \times \cdots \times n_d}$ is a multilinear map parameterizing a low-rank tensor format. Such a problem is amenable to alternating optimization since the update for a single block variable U^μ is just an optimization problem for the function f, but on a linear subspace of $\mathbb{R}^{n_1 \times \cdots \times n_d}$.

As an important example, we mention optimization in the TT format.[16] Here $D = d$ and

$$\tau : \mathcal{V} := \mathbb{R}^{n_1 \times k_1} \times \mathbb{R}^{k_1 \times n_2 \times k_2} \times \cdots \times \mathbb{R}^{k_{d-2} \times n_{d-1} \times k_{d-1}} \times \mathbb{R}^{k_{d-1} \times n_d} \to \mathbb{R}^{n_1 \times \cdots \times n_d}$$

is defined via

$$X = \tau(U^1, \dots, U^d) \quad \Leftrightarrow \quad X(i_1, \dots, i_d) = U^1(i_1, :)U^2(:, i_2, :) \cdots U^d(:, i_d), \tag{27}$$

which are matrix products of corresponding slices in the so called TT cores $U^\mu \in \mathbb{R}^{k_{\mu-1} \times n_\mu \times k_\mu}$ (one fixes $k_0 = k_1 = 1$). The minimal possible values $(k_1, \dots, k_{d-1})$, which determine the sizes of the TT cores, such that such a decomposition is possible are called the TT-ranks of tensor X. Alternating optimization methods form the basis for the majority of computational methods in the TT format.[17]

An AO method with relaxation for (26) takes the form

$$\begin{aligned} U^1_{\ell+1} &= (1-\omega)U^1_\ell + \omega \hat{S}_1(U^2_\ell, \dots, U^d_\ell), \\ &\vdots \\ U^\mu_{\ell+1} &= (1-\omega)U^\mu_\ell + \omega \hat{S}_\mu(U^1_{\ell+1}, \dots, U^{\mu-1}_{\ell+1}, U^{\mu+1}_\ell, \dots, U^d_\ell), \\ &\vdots \\ U^d_{\ell+1} &= (1-\omega)U^d_\ell + \omega \hat{S}_d(U^1_{\ell+1}, \dots, U^{d-1}_{\ell+1}), \end{aligned} \tag{28}$$

where the $\hat{S}_\mu$ return minimizers (or critical points) of the restricted functions $U^\mu \mapsto F(\dots, U^\mu, \dots)$ with the other block variables being fixed. In this form, the method has been proposed for the TT format in Reference 10. Under suitable assumptions such an iteration defines a smooth map S_ω from an open subset of $\mathcal{V}_1 \times \cdots \times \mathcal{V}_D$ to $\mathbb{R}^{n_1 \times \cdots \times n_d}$ for which a similar fixed point analysis as in the matrix case can be conducted. In the following, we sketch this for the TT format, but the ideas can be applied to general tree tensor network formats such as the Tucker or hierarchical Tucker format. We will make use of several well-known properties of the TT format, in particular the quotient manifold structure of tensors of fixed TT-rank and the orbital invariance of AO methods. Most of the related details can be found in References 2 and 18.

For the TT format (27), we assume that $\mathbf{k} = (k_1, \dots, k_{d-1})$ is chosen such that tensors of TT-rank $\mathbf{k}$ exist. Then in fact on a dense and open subset $\mathcal{V}'$ of $\mathcal{V}$ the map τ maps to tensors of fixed TT-rank $\mathbf{k}$. For convenience we will use the notation $\mathbf{U} = (U^1, \dots, U^d)$ for the elements in $\mathcal{V}$. As in the matrix case, τ in (27) is invariant under a group action, namely,

$$\mathcal{G} = \mathrm{GL}(k_1) \times \cdots \times \mathrm{GL}(k_{d-1}) \ni \mathbf{A} = (A_1, \dots, A_{d-1}) \mapsto \theta_{\mathbf{A}} \cdot \mathbf{U},$$

which inserts the product $A_\mu A_\mu^{-1}$ between the matrices $U^\mu(:, i_\mu, :)$ and $U^{\mu+1}(:, i_{\mu+1}, :)$ in (27), that is, the slices of the TT cores are transformed according to

$$U^\mu(:, i_\mu, :) \to A^{-1}_{\mu-1} U^\mu(:, i_\mu, :) A_\mu \tag{29}$$

(here $A_0 = A_d = 1$). The corresponding restriction of τ to the quotient manifold $\mathcal{V}'/\mathcal{G}$ is a diffeomorphism onto the set $\mathcal{M}_{\mathbf{k}}$ of tensors of fixed TT-rank $\mathbf{k}$, which is an embedded submanifold of $\mathbb{R}^{n_1 \times \cdots \times n_d}$ of dimension $\dim(\mathcal{M}_{\mathbf{k}}) = \dim(\mathcal{V}) - \dim(\mathcal{G})$. Notably, let $\mathbf{U} \in \mathcal{V}'$, then $\tau'(\mathbf{U}) = 0$ on the tangent space of the orbit $\theta_{\mathbf{A}} \cdot \mathbf{U}$ at $\mathbf{U}$. On any complementary subspace $\mathcal{W}$ to that tangent space, $\tau'(\mathbf{U})$ is a bijection from $\mathcal{W}$ to the tangent space of $\mathcal{M}_{\mathbf{k}}$ at $\tau(\mathbf{U})$.

Assume again that f is smooth and strongly convex. Then any critical point $\mathbf{U}_*$ of the function $F = f \circ \tau$ that lies in $\mathcal{V}'$ is a fixed point of the iteration (28) since the restricted linear maps $U^\mu \mapsto \tau(U^1_*, \dots, U^\mu, \dots, U^d_*)$ are injective so that the corresponding restriction $U^\mu \mapsto F(U^1_*, \dots, U^\mu, \dots, U^d_*)$ is strongly convex. Moreover, the whole process is well-defined in some neighborhood of (the orbit of) $\mathbf{U}_*$ where it can be written as

$$\mathbf{U}_{\ell+1} = S_\omega(\mathbf{U}_\ell)$$

with a smooth map S_ω. A key observation to make is that the maps $\hat{S}_1, \dots, \hat{S}_d$ in (28) that realize the updates of single TT cores exhibit an analogous compatibility with the group action as in (7) for the matrix case, namely

$$\hat{S}_\mu(\theta_{\mathbf{A}} \cdot \mathbf{U}) = A^{-1}_{\mu-1} \hat{S}_\mu(\mathbf{U}) A_\mu,$$

where the matrix product is understood slice-wise as in (29) (and we slightly abused notation since $\hat{S}_\mu$ does not depend on U^μ). It entails a corresponding invariance

$$S_\omega(\theta_{\mathbf{A}} \cdot \mathbf{U}) = \theta_{\mathbf{A}} \cdot S_\omega(\mathbf{U}) \tag{30}$$

of a full update loop, in analogy to (13). This allows us to regard (28) as a well-defined iteration

$$X_{\ell+1} = \mathbf{S}_\omega(X_\ell)$$

on the manifold $\mathcal{M}_{\mathbf{k}}$, at least locally in a neighborhood of $\tau(\mathbf{U}_*)$. From a practical viewpoint, the invariance (30) admits to change the TT representation in every substep of (28) in order to make the restricted linear maps $U^\mu \mapsto \tau(\dots, U^\mu, \dots)$ orthogonal and improve numerical stability. We refer to References 2 and 17 for details on orthogonalization of substeps.

Based on these similarities to the matrix case, one can proceed in almost the same way as in Section 2. Let $\mathbf{U}_* \in \mathcal{V}'$ be a critical point of F, that is, $\nabla F(\mathbf{U}_*) = 0$. Due to the orbital invariance of F, the Hessian $H = \nabla^2 F(\mathbf{U}_*)$ has a kernel of dimension at least $\dim(\mathcal{G})$ since it contains the tangent space to the orbit at $\mathbf{U}_*$. However, in the block decomposition

$$H = D + E + E^\mathsf{T} \tag{31}$$

into a block diagonal part D (corresponding to the block variables $U^1, \dots, U^d$) and lower block triangular part E, the block matrix D is positive definite since f is strongly convex.[†] The derivative of S_ω then again takes the form of an SOR error iteration matrix

$$T_\omega = I - N_\omega^{-1} H, \quad N_\omega = \frac{1}{\omega} D + E,$$

similar to (20); see, for example, Reference 6, Theorems 10.3.4 and 10.3.5 for the derivation. For $0 < \omega < 2$, a decomposition $\mathcal{V} = \ker H \oplus \mathcal{W}_\omega$ as in (22) applies and T_ω is a contraction on the invariant subspace $\mathcal{W}_\omega$ if H is positive semidefinite. Using the same proof as for Theorem 1, we obtain the analogous local convergence result for TT optimization. Recall that we assume that $\mathbf{k}$ is properly chosen so that τ maps the open and dense subset $\mathcal{V}'$ to the manifold $\mathcal{M}_{\mathbf{k}}$.

Theorem 3. *Let $\mathbf{U}_* \in \mathcal{V}'$ be a critical point of function F in (26) where f is strongly convex. Assume that the Hessian $H = \nabla^2 F(\mathbf{U}_*)$ is positive semidefinite and $\dim(\ker H) = \dim(\mathcal{G}) = k_1^2 + \dots + k_{d-1}^2$. Fix $0 < \omega < 2$. Then for $X_0 = \tau(\mathbf{U}_0)$ close enough (this may depend on ω) to $X_* = \tau(\mathbf{U}_*)$ the iteration (28) is well-defined and the iterates $X_\ell = \tau(\mathbf{U}_\ell)$ converge to X_* at an asymptotic linear rate*

$$\rho_\omega = \limsup_{\ell \to \infty} \|X_\ell - X_*\|^{1/\ell} < 1,$$

where ρ_ω is the spectral radius of T_ω on $\mathcal{W}_\omega$.

While so far everything looks conceptually almost identical to the matrix case, a major difference arises when proceeding to determine the optimal shift parameter ω. It is not clear whether Lemma 1 can be generalized. This is in fact already an issue with the linear SOR method with more than two blocks for positive definite systems, since certain conditions on the decomposition (31) of H are required in order to derive a formula like (25) for the eigenvalues of T_ω; compare Reference 15, Section 4.6. In the matrix case, the fact that $E + E^\mathsf{T}$ is two-cyclic makes this possible but for more than two block variables assuming such conditions on E does not appear very reasonable, especially when taking into account that the critical point (U_*, V_*) and hence its Hessian are not given a priori. Moreover, even if the formula (25) would apply, one would need that the eigenvalues of the Jacobi error iteration matrix $I - D^{-1}H$ have absolute value at most one, but for a block decomposition (31) with more than two blocks this does not follow from the positive definiteness of D alone. Thus, for the TT format the estimation of an optimal parameter ω_{opt} from formula (24) remains a heuristic.

4 | NUMERICAL EXPERIMENTS

In this section, we present some numerical experiments to illustrate the benefit of overrelaxation in low-rank optimization.

4.1 | Matrix completion problem

First, we apply the proposed AO overrelaxation scheme in Algorithm 1 to the following nonconvex formulation of a low-rank matrix completion problem:

$$\min_{U\in\mathbb{R}^{m\times k},V\in\mathbb{R}^{n\times k}} F(U,V)=\frac{1}{2}\left\|P_\Omega(A-UV^\top)\right\|_F^2. \tag{32}$$

Here Ω is a given set of index pairs, and the linear operator $P_\Omega:\mathbb{R}^{m\times n}\to\mathbb{R}^{m\times n}$ is defined as

$$(P_\Omega(X))_{ij}=\begin{cases} x_{ij}, & (i,j)\in\Omega,\\ 0, & \text{otherwise.}\end{cases}$$

In our experiments, the set Ω consists of randomly generated index pairs. We set $m=n$ and choose A to be a random rank-k matrix, that is, $A=U_*V_*^\top$, where $U_*,V_*\in\mathbb{R}^{n\times r}$ are random matrices with each element sampled from a standard Gaussian distribution. The number of sampled entries $|\Omega|$ is defined by an oversampling parameter $\mathtt{OS}\geq 1$,

$$|\Omega|=\mathtt{OS}\cdot(2nk-k^2),$$

since we want $|\Omega|$ to be larger than $2nk-k^2=\dim(\mathcal{M}_k)$, which is the number of essential degrees of freedom for an $n\times n$ matrix of rank k. In the experiments $\mathtt{OS}=3$.

In Figure 1, we present the convergence plots for an experiment with $n=2000$ for several choices of ω and different ranks k. We report the relative residuals

$$\mathtt{err}_\ell=\frac{\left\|P_\Omega(A-U_\ell V_\ell^\top)\right\|_F}{\|P_\Omega(A)\|_F}, \tag{33}$$

where the sequence (U_ℓ,V_ℓ) is generated by Algorithm 1 with a shift parameter ω. The only difference with Algorithm 1 is that we always start with $\omega=1$ (standard ALS) and only turn on the shift after the convergence has stabilized, in this experiment usually after 12 iterations. The optimal shift ω_{opt} from (24) depends on the convergence rate $\beta^2=\rho_1$ of the standard AO method, which is estimated while running the iteration with $\omega=1$ using

$$\beta^2\approx\sqrt{\frac{\mathtt{err}_{\ell+2}}{\mathtt{err}_\ell}}. \tag{34}$$

As expected, using overrelaxation accelerates the convergence of the ALS method if ω is chosen properly. We note that the additional computations arising from the overrelaxation scheme come in asymptotically negligible cost as compared

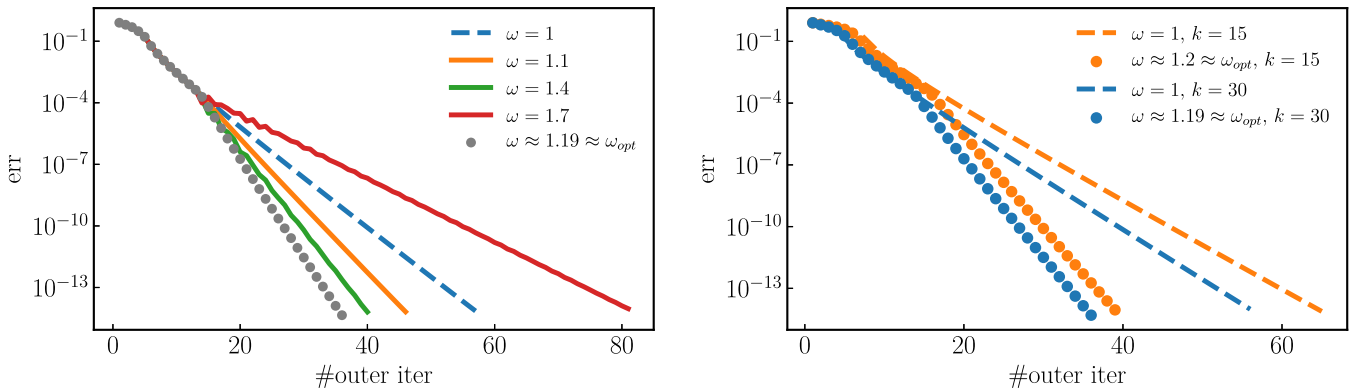

FIGURE 1 Relative residuals (33) of Algorithm 1 for the completion problem (32) with respect to the number of outer iterations using various shift parameters ω for rank $k=30$ (left) and for rank values $k=15,30$ (right). The parameter $\omega=1$ corresponds to the standard ALS method, while parameters $\omega\in(1,2)$ represent the version of the iteration with overrelaxation. The $\omega\approx\omega_{\text{opt}}$ case corresponds to the choice (24) with β^2 estimated using (34).

with the basic ALS. In turn, the proposed approach leads to a significant reduction of the total number of iterations for achieving a high accuracy.

4.2 | Low-rank solution of the Lyapunov equation

Consider the Lyapunov equation

$$AX + XA^\top = B, \quad A, B \in \mathbb{R}^{n\times n}, \tag{35}$$

where $X \in \mathbb{R}^{n\times n}$ is a matrix to be found. In case of a symmetric positive definite matrix A, Equation (35) represents the optimality condition for the strongly convex optimization problem

$$\min_{X\in\mathbb{R}^{n\times n}} f(X) = \frac{1}{2}\langle AX + XA^\top, X\rangle_F - \langle B, X\rangle_F.$$

A rank-k approximation to the solution is, therefore, obtained by solving the problem

$$\min_{U,V\in\mathbb{R}^{n\times k}} F(U,V) = f\left(UV^\top\right)$$

instead, which is of the form (2). For this we employ the proposed overrelaxation algorithm.

In the experiment, we choose $A = (n+1)^2\,\mathrm{tridiag}(-1, 2, -1)$, set $n = 256$ and generate the right-hand side $B = AX_* + X_*A^\top$ from a specified solution X_*. Specifically, we choose $k = 2$ and generate the second and the third singular values of X_* such that their ratio equals to 0.99, which is similar to experiments conducted in Reference 3. There it has been numerically observed that such a large ratio at the target singular value can lead to slow convergence of the standard ALS method.

Due to the fact that X_* cannot be approximated with high accuracy using $k = 2$, the function values $f(X)$ will not converge to zero and hence cannot be taken as an appropriate error measure. Instead, we compute the values

$$\texttt{proj\_err}_\ell = \frac{\|\mathbf{P}_{X_\ell}\nabla f(X_\ell)\|_F}{\|\mathbf{P}_{X_\ell}(B)\|_F} = \frac{\|\mathbf{P}_{X_\ell}\left(AX_\ell + X_\ell A^\top - B\right)\|_F}{\|\mathbf{P}_{X_\ell}(B)\|_F}, \quad X_\ell = U_\ell V_\ell^\top, \tag{36}$$

where $\mathbf{P}_{X_\ell}$ denotes the orthogonal projection operator to the tangent space of the manifold $\mathcal{M}_k$ of fixed rank-k matrices at X_ℓ; see, for example, Reference 19. This reflects the fact that the method can be regarded as a minimization method on that manifold. Similar to (34), we can then use these values for approximating the optimal shift parameter ω_{opt} using (24) with β^2 estimated from

$$\beta^2 \approx \sqrt{\frac{\texttt{proj\_err}_{\ell+2}}{\texttt{proj\_err}_\ell}}. \tag{37}$$

In Figure 2, we plot the values of $\texttt{proj\_err}_\ell$ against the number of outer iterations ℓ for several values of shifts ω, including the basic ALS and the approximated optimal shift. In all cases, the shift is activated after 50 iterations. All considered shifts lead to convergence improvement with the shift that approximates the optimal one being the best.

4.3 | Linear systems in the quantized TT format

Finally, we test our approach for solving linear systems in the TT format. In particular, we apply the so-called quantized TT (QTT) format[20,21] to solve Equation (35) by fixing $n = 2^d$, $d = 12$, and by representing $X \in \mathbb{R}^{2^d\times 2^d}$ as order-$2d$ tensors in $\mathbb{R}^{2\times\cdots\times 2}$ using reshape in the lexicographical order. These tensors are then further restricted to the TT format with the TT-rank equal to $(4, 4, \ldots, 4)$ (this choice of ranks led to a much slower convergence of the ALS method as compared to other rank values). The right-hand side B was selected to be a matrix of all ones, which trivially admits a QTT

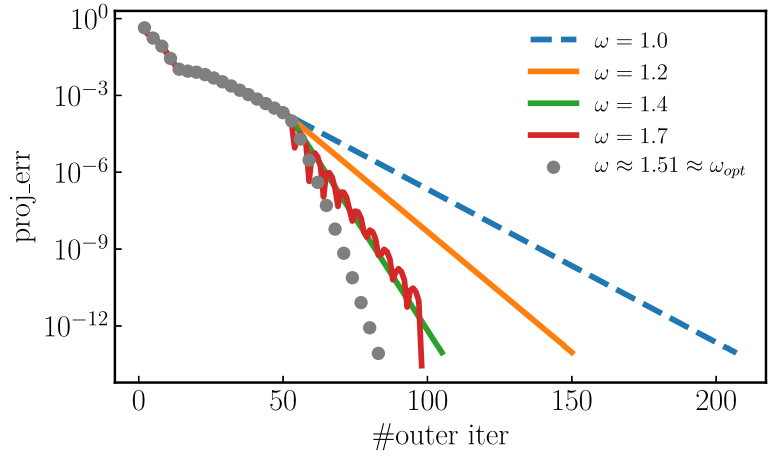

FIGURE 2 Relative residuals (36) of Algorithm 1 for the low-rank solution of a Lyapunov equation (35) with respect to the number of outer iterations using various shift parameters ω. Here $k = 2$. The parameter $\omega = 1$ corresponds to the standard ALS method, while parameters $\omega \in (1, 2)$ represent the version of the iteration with overrelaxation. The $\omega \approx \omega_{\text{opt}}$ case corresponds to the choice (24) with β^2 estimated using (37).

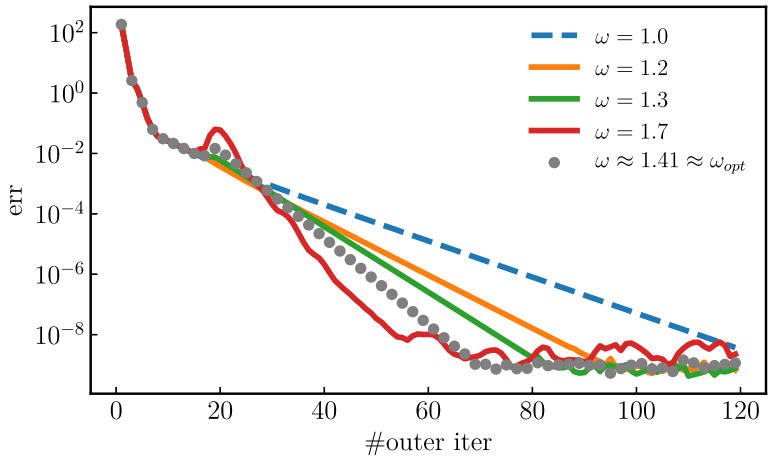

FIGURE 3 Maximal (within one ALS sweep) relative local residual with respect to the number of outer iterations of the QTT ALS method for solving a Lyapunov equation (35) using various shifts ω. The QTT ranks are $(4, \ldots, 4)$. The parameter $\omega = 1$ corresponds to the standard QTT ALS method, while parameters $\omega \in (1, 2)$ represent the iteration with overrelaxation. The $\omega \approx \omega_{\text{opt}}$ case corresponds to (24) with β^2 estimated from the standard ALS method, even though for tensor problems there is no theoretical guarantee that this choice is actually close to optimal. Shifts are activated after 15 iterations.

representation with all TT-ranks equal to one. Note that all computations were performed directly in the TT format, that is, no full tensors were formed.

As an error measure $\texttt{err}_\ell$ we take the maximum relative norm of all local residuals within one sweep of the standard ALS.[22] Based on this error we estimate β^2 and use the same formula (34) for ω_{opt}, but as noted in Section 3 there is no theoretical guarantee that this formula provides the optimal shift parameter in the tensor case. Nevertheless, the results from Figure 3 suggest that this choice leads to nearly the fastest convergence among the considered choices of shifts.

ACKNOWLEDGMENTS

The work of Ivan V. Oseledets was supported by the Ministry of Science and Higher Education of the Russian Federation under Grant No. 075-10-2021-068. Results obtained in Section 4 (performed by Maxim V. Rakhuba) were obtained within

the Russian Science Foundation under Grant No. 21-71-00119. Open Access funding enabled and organized by Projekt DEAL.

CONFLICT OF INTEREST

The authors declare no potential conflict of interests.

DATA AVAILABILITY STATEMENT

Data sharing is not applicable to this article as no datasets were generated or analyzed during the current study.

ENDNOTES

*Obviously, $\mathcal{W}_\omega$ is an invariant subspace of T_ω and has the correct dimension. To see that it is complementary to $\ker H$, note that any $x \in \mathcal{W}_\omega$ satisfies $N_\omega x \in (\ker H)^\perp$. If $x \in \ker H$, one verifies $0 = \langle x, N_\omega x\rangle = \left(\frac{1}{\omega} - \frac{1}{2}\right)\langle x, Dx\rangle$, which under the given assumptions implies $x = 0$.

†As in Remark 1, this also follows from the assumptions that H is positive semidefinite and $\dim(\ker H) = \dim(\mathcal{G})$, since the tangent space to the orbit at $\mathbf{U}_* \in \mathcal{V}'$ does not contain elements of the form $(0, \ldots, 0, U^\mu, 0, \ldots, 0)$.

REFERENCES

1. Uschmajew A. Local convergence of the alternating least squares algorithm for canonical tensor approximation. SIAM J Matrix Anal Appl. 2012;33(2):639–52.
2. Rohwedder T, Uschmajew A. On local convergence of alternating schemes for optimization of convex problems in the tensor train format. SIAM J Numer Anal. 2013;51(2):1134–62.
3. Oseledets IV, Rakhuba MV, Uschmajew A. Alternating least squares as moving subspace correction. SIAM J Numer Anal. 2018;56(6):3459–79.
4. Schechter S. Iteration methods for nonlinear problems. Trans Am Math Soc. 1962;104:179–89.
5. Ortega JM, Rockoff ML. Nonlinear difference equations and Gauss-Seidel type iterative methods. SIAM J Numer Anal. 1966;3:497–513.
6. Ortega JM, Rheinboldt WC. Iterative solution of nonlinear equations in several variables. New York-London: Academic Press; 1970.
7. Hageman LA, Porsching TA. Aspects of nonlinear block successive overrelaxation. SIAM J Numer Anal. 1975;12:316–35.
8. Wen Z, Yin W, Zhang Y. Solving a low-rank factorization model for matrix completion by a nonlinear successive over-relaxation algorithm. Math Program Comput. 2012;4(4):333–61.
9. Wang J, Wang YP, Xu Z, Wang CL. Accelerated low rank matrix approximate algorithms for matrix completion. Comput Math Appl. 2019;77(2):334–41.
10. Grasedyck L, Kluge M, Krämer S. Variants of alternating least squares tensor completion in the tensor train format. SIAM J Sci Comput. 2015;37(5):A2424–50.
11. Weissinger J. Verallgemeinerungen des Seidelschen Iterationsverfahrens. Z Angew Math Mech. 1953;33:155–63.
12. Keller HB. On the solution of singular and semidefinite linear systems by iteration. J Soc Ind Appl Math Ser B Numer Anal. 1965;2:281–90.
13. Lee JM. Introduction to smooth manifolds. New York: Springer-Verlag; 2003.
14. Young DM. Iterative solution of large linear systems. New York-London: Academic Press; 1971.
15. Hackbusch W. Iterative solution of large sparse systems of equations. 2nd ed. Cham: Springer; 2016.
16. Oseledets IV. Tensor-train decomposition. SIAM J Sci Comput. 2011;33(5):2295–317.
17. Holtz S, Rohwedder T, Schneider R. The alternating linear scheme for tensor optimization in the tensor train format. SIAM J Sci Comput. 2012;34(2):A683–713.
18. Uschmajew A, Vandereycken B. The geometry of algorithms using hierarchical tensors. Linear Algebra Appl. 2013;439(1):133–66.
19. Vandereycken B. Low-rank matrix completion by Riemannian optimization. SIAM J Optim. 2013;23(2):1214–36.
20. Khoromskij BN. $O(d \log N)$-quantics approximation of N-d tensors in high-dimensional numerical modeling. Constr Approx. 2011;34(2):257–80.
21. Oseledets IV. Approximation of $2^d \times 2^d$ matrices using tensor decomposition. SIAM J Matrix Anal Appl. 2009;10;31(4):2130–45.
22. Oseledets IV, Dolgov SV. Solution of linear systems and matrix inversion in the TT-format. SIAM J Sci Comput. 2012;34(5):A2718–39.

How to cite this article: Oseledets IV, Rakhuba MV, Uschmajew A. Local convergence of alternating low-rank optimization methods with overrelaxation. Numer Linear Algebra Appl. 2023;30(3):e2459. https://doi.org/10.1002/nla.2459