Time-stepping of low-rank approximations with small singular values

Emil Kieri

(joint work with Christian Lubich, Hanna Walach)

We consider low-rank approximations to matrix differential equations of the form

$$\dot{A}(t) = F(t, A(t)), \qquad A(0) \in \mathbb{C}^{m\times n}.$$

Such problems appear, e.g., after spatial discretisation of time-dependent partial differential equations (PDEs) in two spatial dimensions. Higher-dimensional PDEs give rise to higher order tensor differential equations of similar form. Tensor differential equations in tensor-train format [5] are also covered by the present analysis, but for the sake of simplicity we will restrict this presentation to the matrix case. Low-rank approximations have been used with success for high-dimensional PDEs, most predominantly in the multi-configurational time-dependent Hartree method [4]. There are, however, several open questions regarding the convergence of low-rank approximations. In this work, we answer one of these questions positively. We prove that if F has a modest Lipschitz constant and maps almost onto the tangent space of the low-rank manifold, then a recently proposed time-stepping scheme for low-rank approximations [2, 3] will give an accurate result, also in the presence of small singular values [1]. The scheme was first proposed for matrices in [2], and then extended to tensors in tensor-train format in [3].

In the vicinity of small nonzero singular values, the manifold $\mathcal{M}_r \subset \mathbb{C}^{m\times n}$ of rank-r matrices has strong curvature. That is, if σ_r is the rth singular value of $X \in \mathcal{M}_r$, then the Lipschitz constant of the projection $P(X)$ onto the tangent space $T_X\mathcal{M}_r$ of $\mathcal{M}_r$ at X is proportional to σ_r^{-1}. This means that standard time-stepping schemes for the dynamical low-rank approximation

$$\dot{Y}(t) = P(Y)F(t, Y(t)), \qquad Y(0) = Y_0 \approx A(0)$$

will break down in the presence of small singular values. The scheme of [2, 3], on the other hand, is robust in this case.

With the decomposition $Y = USV^*$, where $S \in \mathbb{C}^{r\times r}$ and U, V have orthonormal columns, the projection onto the tangent space has the representation

$$P(Y)Z = ZVV^* - UU^*ZVV^* + UU^*Z \qquad \text{for } Z \in \mathbb{C}^{m\times n}.$$

The method of [2, 3] uses this decomposition of the projection to construct a splitting scheme. For matrices, given $Y_0 = U_0S_0V_0^*$, the scheme reads:

- Solve $\dot{K}(t) = F(t, K(t)V_0^*)$, $\quad K_0 = U_0S_0$.
- Make a QR-decomposition $[U_1, \widehat{S}_1] = \mathrm{qr}\,(K(h))$.
- Solve $\dot{S}(t) = -U_1^*F(t, U_1S(t)V_0^*)V_0$, $\quad S(0) = \widehat{S}_1$, and let $\widetilde{S}_0 = S(h)$.
- Solve $\dot{L}(t) = F(t, UL(t)^*)^*U_1$, $\quad L(0) = V_0\widetilde{S}_0^*$.
- Make a QR-decomposition $[V_1, S_1^*] = \mathrm{qr}\,(L(h))$.

Then, $Y_1 = U_1S_1V_1^*$ is a consistent approximation of $Y(h)$. For many problems F can be evaluated without forming the full $m \times n$ matrix. In such cases the scheme is very efficient, as it only works with much smaller matrices.

In [1] we prove that if $F(t, Y)$ is Lipschitz continuous in its second argument, and if it up to an ε-perturbation maps onto $T_Y\mathcal{M}_r$, then the error of the splitting scheme can be bounded in terms of ε and the time step h, independently of the singular values. The proof is based on two important properties of the scheme:

- If $F(t, A) = F(t)$, i.e., the right-hand side does not explicitly depend on A, and if $A(t) \in \mathcal{M}_r$ for all t, then the splitting scheme is exact for any time step h. This result was proven in [2].
- The matrices U and V stay constant during some of the substeps. Furthermore, the perturbation from the exact flow of FVV^* arising in the first substep is invariant to projection with VV^*, and similarly in substeps two and three with the relevant projections.

To prove the result, we construct a path $X(t)$ on $\mathcal{M}_r$ close to $A(t)$. By the assumptions on F, such a path exists. By the exactness result, the splitting schemed applied to $\dot{A}(t) = \dot{X}(t)$ would give $A(h) = X(h)$. Using the Gröbner–Alekseev lemma we estimate how much we, in each substep, deviate from the scheme applied to $\dot{X}(t)$. As $X(t)$ and $Y(t)$ are different paths on the manifold, the Lipschitz constant of F will here enter in the estimate. Using the preservation of U and V in the relevant substeps we can bound the effect of these deviations independently of the singular values. For the details we refer to [1].

As the error estimate depends on the Lipschitz constant of F, it is not valid for (spatial discretisations of) PDEs. Extending the error analysis to cover also this case remains an open problem. Numerically we get much better results for PDEs than what is explained by the present analysis.

References

[1] E. Kieri, Ch. Lubich and H. Walach, *Discretized dynamical low-rank approximation in the presence of small singular values*, SIAM J. Numer. Anal. **54** (2016), 1020–1038.

[2] Ch. Lubich and I. V. Oseledets, *A projector-splitting integrator for dynamical low-rank approximation*, BIT **54** (2014), 171–188.

[3] Ch. Lubich, I. V. Oseledets and B. Vandereycken, *Time integration of tensor trains*, SIAM J. Numer. Anal. **53** (2015), 917–941.

[4] H.-D. Meyer, U. Manthe and L. S. Cederbaum, *The multi-configurational time-dependent Hartree approach*, Chem. Phys. Lett. **165** (1990), 73–78.

[5] I. V. Oseledets, *Tensor-train decomposition*, SIAM J. Sci. Comput. **33** (2011), 2295–2317.