Low Rank Approximation
Lecture 9

Daniel Kressner
Chair for Numerical Algorithms and HPC
Institute of Mathematics, EPFL
daniel.kressner@epfl.ch

Time-dependent low-rank approximation

Goal: Approximate given matrix $A(t)$ at every time $t \in [0, T]$ by rank-r matrix.

Motivation for taking time-dependence into account:

- If $\sigma_r(t) > \sigma_{r+1}(t)$ for all $t \in [0, T]$ then $\mathcal{T}_r(A(t))$ inherits smoothness of A wrt t [Baumann/Helmke'2003; Mehrmann/Rath'1993].
- Low-rank approximation at t_i may yield valuable information for low-rank approximation at nearby t_{i+1}.
- Allows us to work with tangent space (linear) instead of manifold (nonlinear). Much easier to impose additional structure.

More general setting: Approximate solution of ODE

$$\dot{A}(t) = F(A(t)), \quad A(0) = A_0.$$

by trajectory $X(t)$ in $\mathcal{M}_r$.

Main application for low-rank tensors: Discretized time-dependent high-dimensional PDEs.

Dynamical low-rank approximation

Recall definition of tangent vectors. If $X : [0, T] \to \mathbb{R}^{m\times n}$ is a smooth curve in $\mathcal{M}_r$ then

$$\dot{X}(t) \in T_{X(t)}\mathcal{M}_r, \qquad \forall t \in [0, T].$$

Dynamical low-rank approximation: Given $A : [0, T] \to \mathbb{R}^{m\times n}$, for every t construct $X(t) \in \mathcal{M}_r$ satisfying

$$\dot{X}(t) \in \mathcal{T}_{X(t)}\mathcal{M}_r \quad \text{such that} \quad \|\dot{X}(t) - \dot{A}(t)\| = \min!$$

- Differential equation needs to be supplemented by initial value, say $X(0) = \mathcal{T}_r(A(0))$, which is in $\mathcal{M}_r$ unless $A(0)$ has rank less than r.
- Hope: X stays close to A if A admits good rank-r approximation for all t.
- For efficiency, aim at finding equivalent system of $\approx \dim \mathcal{M}_r$ differential equations.

Dynamical low-rank approximation

Counterexample to hope:

$$A(t) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 10^{-5}e^{t-1} & 0 \\ 0 & 0 & 10^{-5}e^{1-t} \end{pmatrix}, \quad t \in [0, 10].$$

For $r = 2$, dynamical low-rank approximation yields

$$X(t) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 10^{-5}e^{1-t} \end{pmatrix}.$$

$$\|A(T) - X(T)\|_F = 10^4 \quad \text{but} \quad \|A(T) - \mathcal{T}_r(A(T))\|_F = 10^{-14}.$$

Dynamical low-rank approximation is “blind” to evolution of $10^{-5}e^{t-1}$.

Can only hope for good approximation on short time intervals and if $\sigma_r(t) > \sigma_{r+1}(t)$.

Dynamical low-rank approximation

$$\dot{X}(t) \in \mathcal{T}_{X(t)}\mathcal{M}_r \quad \text{such that} \quad \|\dot{X}(t) - \dot{A}(t)\| = \min!$$

is equivalent to differential equation

$$\dot{X}(t) = P_{X(t)}(\dot{A}(t)), \tag{1}$$

where $P_{X(t)}$ is the orthogonal projection onto $\mathcal{T}_{X(t)}\mathcal{M}_r$.

Will now omit dependence on time. Given $X = USV^T$ with $S \in \mathbb{R}^{r\times r}$ (not necessarily diagonal), $U \in \mathrm{St}(m,r)$ and $V \in \mathrm{St}(n,r)$, we have

$$\begin{aligned} P_X(\dot{A}) &= P_U \dot{A} P_V + P_U^{\perp} \dot{A} P_V + P_V^{\perp} \dot{A} P_U \\ &= U\dot{S}V^T + \dot{U}SV^T + US\dot{V}^T, \end{aligned}$$

where

$$\begin{aligned} \dot{S} &= U^T \dot{A} V \\ \dot{U} &= P_U^{\perp} \dot{A} V S^{-1} \\ \dot{V} &= P_V^{\perp} \dot{A}^T U S^{-T}. \end{aligned}$$

Dynamical low-rank approximation

Dynamical low-rank approximation is equivalent to system of differential equations

$$
\begin{aligned}
\dot{S} &= U^T \dot{A} V \\
\dot{U} &= P_U^{\perp} \dot{A} V S^{-1} \\
\dot{V} &= P_V^{\perp} \dot{A}^T U S^{-T}
\end{aligned}
$$

with initial values $S(0) = S_0 \in \mathbb{R}^{r\times r}$, $U(0) = U_0 \in \mathrm{St}(m, r)$, $V(0) = V_0 \in \mathrm{St}(n, r)$.

Remarks:

- $\dot{U} \in T_U \mathrm{St}(m, r)$ and thus U stays in $\mathrm{St}(m, r)$; $\dot{V} \in T_V \mathrm{St}(n, r)$ and thus V stays in $\mathrm{St}(n, r)$. This can be preserved using geometric integration [Hairer/Lubich/Wanner'2006] or combining standard integrators with retractions.
- Step size control should aim at controlling error for USV^T and not for individual factors
- Presence of S^{-1} makes differential equations very stiff for ill-conditioned S (σ_r close to zero); results in small step sizes of explicit integrators.

Dynamical low-rank approximation

Error bound from Theorem 5.1 in [Koch/Lubich'2007].

Theorem

Suppose that for $t \in [0, T]$ we have

- $\sigma_r(A(t)) > \sigma_{r+1}(A(t))$;
- $\|A(t) - \mathcal{T}_r(A(t))\|_F \leq \rho/16$ *with* $\rho = \min_{t \in [0,T]} \sigma_r(A(t))$.

Then, with $X(0) = \mathcal{T}_r(A(0))$, we have

$$\|X(t) - \mathcal{T}_r(A(t))\|_F \leq 2\beta e^{\beta t} \int_0^t \|A(s) - \mathcal{T}_r(A(s))\|_F \, \mathrm{d}s$$

where $\beta = 8\mu/\rho$.

Dynamical low-rank approximation

Extension to dynamical system $\dot{A} = F(A)$ with $F : \mathbb{R}^{m\times n} \to \mathbb{R}^{m\times n}$.

Dynamical low-rank approximation: Construct $X(t) \in \mathcal{M}_r$ satisfying

$$\dot{X}(t) \in \mathcal{T}_{X(t)}\mathcal{M}_r \quad \text{such that} \quad \|\dot{X}(t) - F(X(t))\| = \min!$$

Equivalent to dynamical system

$$\dot{X}(t) = P_{X(t)}(F(X(t))).$$

Equivalent to

$$\begin{aligned} \dot{S} &= U^T F(X(t)) V \\ \dot{U} &= P_U^{\perp} F(X(t)) V S^{-1} \\ \dot{V} &= P_V^{\perp} F(X(t))^T U S^{-T} \end{aligned}$$

EFY. Consider the linear matrix differential equation

$$\dot{A} = LA + AR$$

for $L(t) \in \mathbb{R}^{m\times m}$ and $R(t) \in \mathbb{R}^{n\times n}$. Show that $A(t) \in \mathcal{M}_r$ for all $t > 0$ if $A(0) \in \mathcal{M}_r$. Write down the differential equations for U, S, V.

EFY. Consider the nonlinear matrix differential equation

$$\dot{A} = F(A) := LA + AR + A \circ A,$$

where $\circ$ denotes the elementwise product. Develop an efficient method for evaluating $F(A(t))$ for $A(t) \in \mathcal{M}_r$ when $r \ll m, n$.

Dynamical low-rank approximation

For error analysis assume for $Y(t) := \mathcal{T}_r(A(t))$ that

- $\|F(Y(t))\|_F \leq \mu$, $\|F(X(t))\|_F \leq \mu$ for $t \in [0, T]$;
- $\langle F(Z_1) - F(Z_2), Z_1 - Z_2 \rangle \leq \lambda \|Z_1 - Z_2\|_F^2$ for all $Z_1, Z_2 \in \mathcal{M}_r$ and some fixed $\lambda \in \mathbb{R}$;
- $\|F(Y(t)) - F(A(t))\|_F \leq L \|Y(t) - A(t)\|_F$ for $t \in [0, T]$.

Theorem 6.1 in [Koch/Lubich'2007]:

Theorem

Under assumptions above, suppose that for $t \in [0, T]$ we have

- $\sigma_r(A(t)) > \sigma_{r+1}(A(t))$;
- $\|A(t) - \mathcal{T}_r(A(t))\|_F \leq \rho/16$ *with* $\rho = \min_{t \in [0,T]} \sigma_r(A(t))$.

Then, with $X(0) = \mathcal{T}_r(A(0))$, we have

$$\|X(t) - \mathcal{T}_r(A(t))\|_F \leq (2\beta + L) e^{(2\beta+\lambda)t} \int_0^t \|A(s) - \mathcal{T}_r(A(s))\|_F \, \mathrm{d}s$$

Dynamical low-rank approximation

Numerical integrators face severe problems as $\sigma_r \to 0$.

Unfortunately, $\sigma_r \approx 0$ is a very likely situation, as problems of interest typically feature quick singular value decay.

Regularization ($=$ increasing small singular values of S by $\epsilon > 0$) introduces additional errors whose effect is poorly understood.

Fundamental underlying problem: U, V ill-conditioned in the presence of small σ_r.

Idea by [Lubich/Oseledets'2014]: Splitting integrator that is insensitive to $\sigma_r \to 0$.

Projector-splitting integrator

Recall that

$$\dot{X}(t) = P_{X(t)}(\dot{A}(t)),$$

with

$$\begin{aligned} P_X(Z) &= P_U Z P_V + P_U^{\perp} Z P_V + P_V^{\perp} Z P_U \\ &= ZVV^T - UU^T ZVV^T + UU^T Z \\ &= ZP_{\text{range}(X^T)} - P_{\text{range}(X)} ZP_{\text{range}(X^T)} + P_{\text{range}(X)} Z. \end{aligned}$$

One step of Lie-Trotter splitting integrator from t_0 to $t_1 = t_0 + h$ applied to this decomposition:

1. Solve $\dot{X}_I = \dot{A}P_{\text{range}(X_I^T)}$ on $[t_0, t_1]$ with i.v. $X_I(t_0) = X_0 \in \mathcal{M}_r$.
2. Solve $\dot{X}_{II} = -P_{\text{range}(X_{II})}\dot{A}P_{\text{range}(X_{II}^T)}$ on $[t_0, t_1]$ with i.v. $X_{II}(t_0) = X_I(t_1)$.
3. Solve $\dot{X}_{III} = P_{\text{range}(X_{III})}\dot{A}$ on $[t_0, t_1]$ with i.v. $X_{III}(t_0) = X_{II}(t_1)$.

Approximation returned: $X_1 := X_{III}(t_1)$.

Standard theory [Hairer/Lubich/Wanner'2006]:
Lie-Trotter splitting integrator has convergence order one.

Projector-splitting integrator

Consider first step:
Rhs $\dot{A}P_{\mathrm{range}(X_I^T)} \in \mathcal{T}_{X_I}\mathcal{M}_r$ and hence $X_I \in \mathcal{M}_r$ for $t \in [t_0, t_1]$.
$\rightsquigarrow$ factorization

$$X_I = U_I S_I V_I^T.$$

Differentiation $\rightsquigarrow$

$$\frac{1}{\partial t}(U_I S_I) V_I^T + U_I S_I \dot{V}_I^T = \dot{X}_I \overset{!}{=} \dot{A} V_I V_I^T.$$

Satisfied if

$$\frac{1}{\partial t}(U_I S_I) = \dot{A} V_I, \quad \dot{V}_I = 0.$$

In particular, primitive of $\dot{A}V_I$ is AV_I. In turn, this equation is solved by

$$U_I(t) S_I(t) = U_I(t_0) S_I(t_0) + \big(A(t) - A(t_0)\big) V_I(t_0).$$

Similar considerations for second and third step.

Projector-splitting integrator

Solution of three split equations given by

1. $U_I(t)S_I(t) = U_I(t_0)S_I(t_0) + \big(A(t) - A(t_0)\big) V_I(t_0)$,
2. $S_{II}(t) = S_{II}(t_0) - U_{II}(t_0)^T \big(A(t) - A(t_0)\big)^T V_{II}(t_0)$,
3. $V_{III}(t)S_{III}(t)^T = V_{III}(t_0)S_{III}(t_0)^T + \big(A(t) - A(t_0)\big)^T U_{III}(t_0)$,

Unassigned Quantities ($V_I, U_{II}, V_{II}, U_{III}$) remain constant.

Practical algorithm:

Given $X_0 = U_0 S_0 V_0^T$ and with $\Delta_A := A(t_1) - A(t_0)$, compute

1. QR factorization $K_I = U_I S_I$ for $K_I = U_0 S_0 + \Delta_A V_0$.
2. $S_{II} = S_I - U_I^T \Delta_A V_0$
3. QR factorization $L_{III} = V_{III} S_{III}^T$ for $L_{III} = V_0 S_{II}^T + \Delta_A^T U_I$.

Returns $U_1 S_1 V_1^T = U_I S_{III} V_{III}^T$

Projector-splitting integrator

Remarks:

- KSL splitting is exact if $A(t)$ has rank at *most* r (Theorem 4.1 by Lubich/Oseledets) $\rightsquigarrow$ robustness in presence of small singular value σ_r [Kieri/Lubich/Wallach'2016].
- Symmetrization (KSLSK) leads to second order.
- Splitting extends to $\dot{A} = F(A)$ by setting $\triangle_A := hF(Y_0)$ (but exactness result does not hold).

Further literature

- Extension to Tucker [Koch/Lubich'2010, Lubich/Vandereycken/Wallach'2018], tensor train [Lubich/Oseledets/Vandereycken'2015]
- Application to 3D nonlinear PDEs [Nonnenmacher/Lubich'2009], matrix differential equations [Mena et al.'2018], Vlasov-Poisson equation [Einkemmer/Lubich'2018], time-dependent PDEs with uncertainties [Musharbash/Nobile/Zhou'2014], quantum chemistry [Kloss/Burghardt/Lubich'2017] and physics [Haegeman et al.'2016].
- [Kieri/Vandereycken'2017]: Conceptually simpler approach by combining standard integrators with low-rank truncation.