Dynamical Low Rank Approximation of Matrices and Tensors

Othmar Koch

joint work with Christian Lubich

Supported by the Austrian Academy of Sciences, APART program.

Low Rank Approximation of Matrices

$A(t) \in \mathbb{R}^{m \times n}, \quad A \in C^1, \quad A(t) = (a_{i,j}(t))_{i,j}$

Best approximation of rank r (wrt the Frobenius norm)

$$\|X(t) - A(t)\|_F^2 = \sum_{i,j} (x_{i,j}(t) - a_{i,j}(t))^2 \to \min!$$

Singular value decomposition (pointwise)

$$X(t) = \hat{U}(t)\hat{S}(t)\hat{V}^T(t),$$

$\hat{U}(t) \in \mathbb{R}^{m \times r}$, $\hat{V}(t) \in \mathbb{R}^{n \times r}$ ... orthogonal,
$\hat{S}(t) \in \mathbb{R}^{r \times r}, \quad \hat{S} = \mathrm{diag}(\sigma_1(t), \ldots, \sigma_r(t)), \quad r \ll \min(m,n).$

Remark: $\hat{U}, \hat{V}$ discontinuous in general.

Dynamical Low Rank Approximation

Koch, Lubich 2007: Replace best approximation by

$$Y(t) = U(t)S(t)V^T(t) = \sum_{i,j} s_{i,j}(t)u_i(t)v_j^T(t) \in \mathcal{M}_r,$$

$U(t) \in \mathbb{R}^{m\times r},\ V(t) \in \mathbb{R}^{n\times r}$ ... orthogonal,
$S(t) \in \mathbb{R}^{r\times r}$ ... general nonsingular matrix,

$$\left\langle \delta Y \,\middle|\, \dot{Y} - \dot{A} \right\rangle_F = 0 \quad \forall \text{ variations } \delta Y \in \mathcal{T}_Y\mathcal{M}_r.$$

Orthogonality conditions $U^T(t)\dot{U}(t) = 0,\ V^T(t)\dot{V}(t) = 0.$
Yields equations of motion

$$\begin{aligned}
\dot{S} &= U^T \dot{A} V, \\
\dot{U} &= (I_m - UU^T)\dot{A}VS^{-1}, \\
\dot{V} &= (I_n - VV^T)\dot{A}^T U S^{-T}.
\end{aligned}$$

Benefits from Variational Approach

Advantages of this approximation:

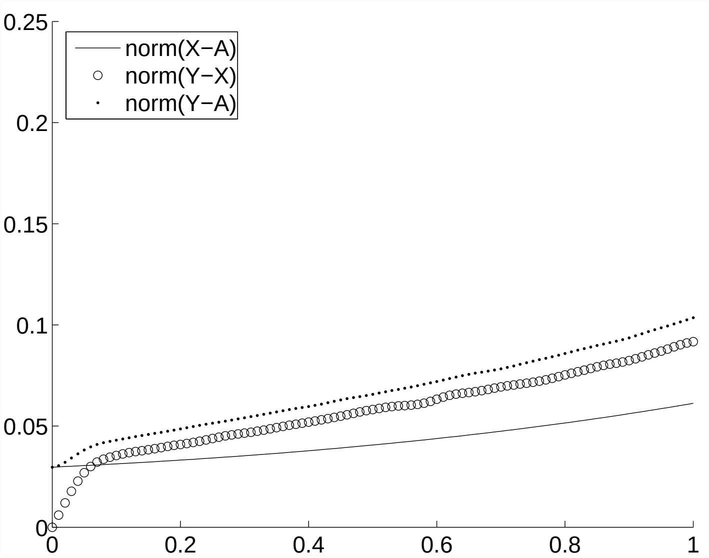
- Factors in the approximation are smooth. The equations of motion do not break down when singular values cross.
- If $\dot{A}$ is sparser than A, computations are cheap.
- Approach extends easily to A defined by matrix differential equation $\dot{A} = f(A)$ — minimum defect approximation: Find Y such that

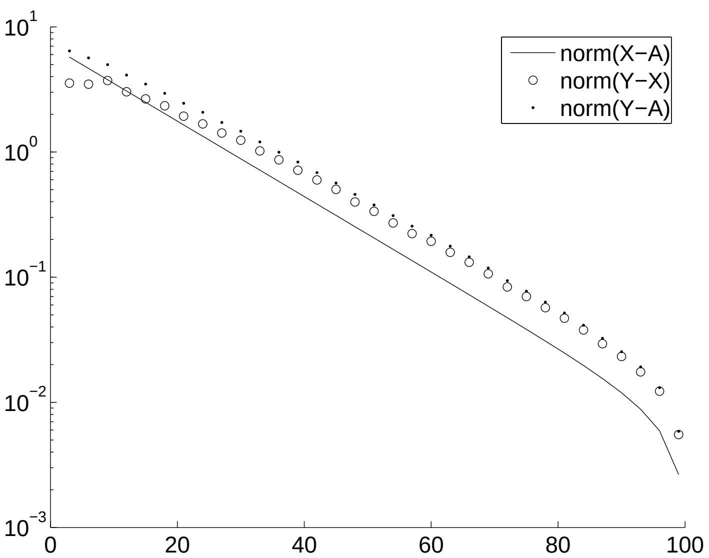
$$\left\langle \delta Y \,\middle|\, \dot{Y} - f(Y) \right\rangle_F = 0 \quad \forall \text{ variations } \delta Y \in \mathcal{T}_Y \mathcal{M}_r.$$

Approximation Result

Theorem (Koch, Lubich 2007):
Let $X(t)$ ... best approximation

$$\|Y(t) - X(t)\|_F = O(t) \qquad \text{for } t \text{ not too large.}$$





Error as a function of t (left) and of r (right)

Outline of Analysis

Rewrite DLRA:

$$\dot{Y} = \mathcal{P}_r(Y)\dot{A},$$

with

$$\mathcal{P}_r(Y)B = B - (I_m - UU^T)B(I_n - VV^T) \quad \text{for } Y = USV^T.$$

Lemma: $X \in \mathcal{M}$ with $\sigma_r(X) \geq \rho > 0$, $Y \in \mathcal{M}$ with $\|Y - X\|_F \leq \frac{1}{8}\rho$: Then we have the *curvature bounds*

$$\|\mathcal{P}_r(Y) - \mathcal{P}_r(X)\|_F \leq \frac{8}{\rho}\|Y - X\|_F,$$

$$\|Y - X - \mathcal{P}_r(Y)(Y - X)\|_F \leq \frac{4}{\rho}\|Y - X\|_F^2.$$

Local Quasi-Optimality

Theorem: Let

- $\|\dot{A}(t)\|_2 \le \mu$,
- $\sigma_r(A(t)) \ge \rho > 0$,
- Best approximation $X(t) \in \mathcal{M}_r$ is C^1 and $\|X(t) - A(t)\|_F \le \frac{1}{16}\rho$.

Then

$$\|Y(t) - X(t)\|_F \le 2\beta e^{\beta t} \int_0^t \|X(s) - A(s)\|_F \, ds$$

with $\beta = 8\mu\rho^{-1}$ as long as the right-hand side is $\le \frac{1}{8}\rho$.

Linear Error Growth

Theorem: Let

$$\begin{aligned} A(t) &= X(t) + E(t), \\ & \quad X(t) \in \mathcal{M}_r \text{ not necessarily best approx.} \end{aligned}$$

with

$$\sigma_r(X) \geq \rho > 0, \quad \|\dot{X}(t)\|_2 \leq \mu, \quad \|\dot{E}(t)\|_F \leq \varepsilon, \quad \varepsilon \leq \frac{1}{8}\mu.$$

Then $Y(0) = X(0)$ implies

$$\|Y(t) - X(t)\|_F \leq 2t\varepsilon, \quad \text{for } t \leq \frac{\rho}{4\sqrt{2\mu\varepsilon}}.$$

Example: Image Compression (1)

A series of images (defining a *pseudo time*):

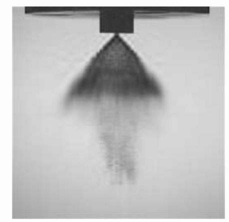
(courtesy Robert Bosch GmbH Stuttgart)

Example: Image Compression (2)

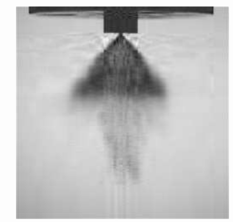
Comparison of approximation quality of last image:

Approx. of rank 16 (top row) and rank 32 (bottom row)

(a) orig.



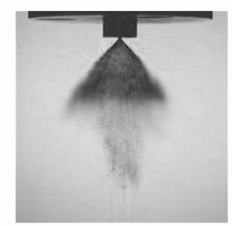
(b) best



(c) dlra

(a) orig.

(b) best



(c) dlra

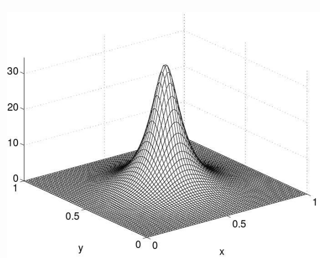
Considerable speed-up by DLRA!

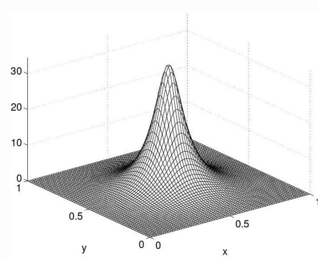
From: Lubich, Nonnenmacher 2008.

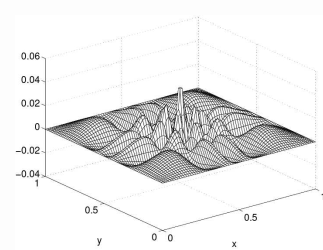
Example: 2D Blowup PDE

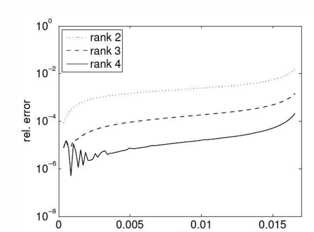
$$\frac{\partial u}{\partial t} = \Delta u + u^3, \quad (x, y) \in [-1, 1]^2,\ u(0) = u_0.$$

- standard FD space discretisation on a tensor grid
- DLRA of the matrix of grid values
- effective rank of the solution near blow-up ≤ 10
- full FD / rank-3 DLRA / difference / error vs. t









From: Lubich, Nonnenmacher 2008.

Tensor Approximation

Extension to higher order tensors $\mathcal{A}(t) \in \mathbb{R}^{N_1 \times \cdots \times N_f}$:

$$
\begin{aligned}
\mathcal{A}(t) &\approx \sum_{j_1=1}^{n_1} \cdots \sum_{j_f=1}^{n_f} s_{j_1,\ldots,j_f}(t) \phi_{j_1}^{(1)}(t) \otimes_1 \cdots \otimes_{f-1} \phi_{j_f}^{(f)}(t) \\
&= \mathcal{S} \times_1 \phi^{(1)} \times_2 \cdots \times_f \phi^{(f)} \\
&=: \mathcal{U}(t) \in \mathcal{M},
\end{aligned}
$$

where

$$
\begin{aligned}
\mathcal{S} &= (s_{j_1,\ldots,j_f})_{j_k=1,\ldots,n_k,\ k=1,\ldots,f} \in \mathbb{R}^{n_1 \times \cdots \times n_f}, \\
\phi^{(k)} &= \left(\phi_{j_k}^{(k)}\right)_{j_k=1,\ldots,n_k} \in \mathbb{R}^{N_k \times n_k}, \qquad k = 1, \ldots, f.
\end{aligned}
$$

Equations of Motion

- For a tensor $\mathcal{A} \in \mathbb{R}^{N_1 \times \cdots \times N_f}$, the *matrix unfolding* $A_{(k)}$ is defined as the matrix resulting from linearly aligning matrix slices.
- The *density matrices* are given by

$$\rho^{(k)}(\mathcal{A}) = A_{(k)} A_{(k)}^T, \quad k = 1, \ldots, f.$$

- The *inner product* and associated norm of tensors $\mathcal{A},\ \mathcal{B} \in \mathbb{R}^{N_1 \times \cdots \times N_f}$ are given by

$$\langle \mathcal{A} | \mathcal{B} \rangle := \sum_{j_1=1}^{N_1} \cdots \sum_{j_f=1}^{N_f} a_{j_1,\ldots,j_f} b_{j_1,\ldots,j_f},$$

$$\|\mathcal{A}\| := \sqrt{\langle \mathcal{A} | \mathcal{A} \rangle}.$$

Equations of Motion (2)

Variational principle

$$\left\langle \delta\mathfrak{U} \middle| \dot{\mathfrak{U}} - \dot{\mathcal{A}} \right\rangle = 0 \quad \forall\, \delta\mathfrak{U} \in \mathcal{T}_{\mathfrak{U}}\mathcal{M},$$

together with orthogonality conditions yields

$$\begin{aligned} \dot{a}_J &= \langle \Phi_J | \mathfrak{B} \rangle, \quad \forall J, \\ \dot{\phi}^{(k)}_{l_k} &= \left(I - P^{(k)} \right) \sum_{j_k=1}^{n_k} \left(\rho^{(k)}(\mathfrak{U})^{-1} \right)_{l_k, j_k} \left\langle \psi^{(k)}_{j_k} \middle| \mathfrak{B} \right\rangle, \quad \forall l_k, k, \end{aligned}$$

$$\Phi_J := \phi^{(1)}_{j_1} \otimes_1 \cdots \otimes_{f-1} \phi^{(f)}_{j_f}, \quad P^{(k)} = \sum_{j_k=1}^{n_k} \phi^{(k)}_{j_k} \left(\phi^{(k)}_{j_k} \right)^T,$$

$$\psi^{(k)}_{j_k} = \left(\mathfrak{S} \times_1 \phi^{(1)} \times_2 \cdots \phi^{(k-1)} \times_{k+1} \phi^{(k+1)} \ldots \times_f \phi^{(f)} \right).$$

Approximation Properties

Approximation results carry over from matrix case through matrix unfoldings!

For a tensor $\mathcal{A} \in \mathbb{R}^{N_1 \times \cdots \times N_f}$, the *matrix unfolding* $A_{(n)} \in \mathbb{R}^{N_n \times (N_{n+1} N_{n+2} \cdots N_f N_1 \cdots N_{n-1})}$ is defined as the matrix containing the element $a_{j_1,\ldots,j_f}$ at the position with row number j_n and column number resulting from linearly aligning matrix slices.

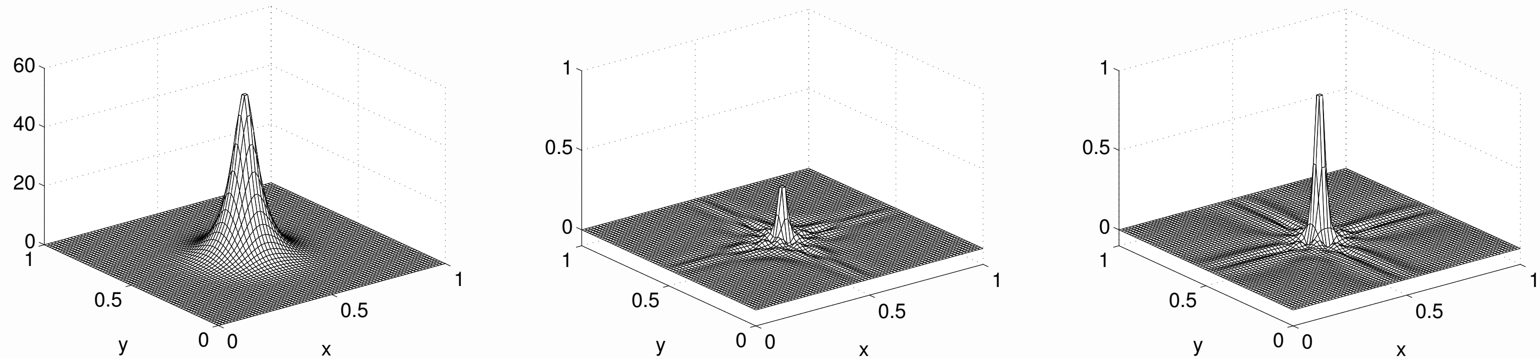
- Curvature bounds of the manifold. ✓
- Local quasi-optimality. ✓
- Linear error growth. ✓

Proof: Koch, Kolda, Lubich (2008).

Numerical Illustration

Parabolic blowup PDE in 3D

$$\frac{\partial u}{\partial t} = \Delta u + u^3, \qquad (x, y, z) \in [0, 1]^3, \quad t > 0.$$



Approximation of rank $(5, 5, 5)$ and error as compared to the rank-$(4, 4, 4)$ and rank-$(3, 3, 3)$ approximations.

From: Lubich, Nonnenmacher (2008).