Dynamical low-rank approximation

Christian Lubich
Univ. Tübingen

Genève, Swiss Numerical Analysis Day, 17 April 2015

Coauthors

- Othmar Koch 2007, 2010
- Achim Nonnenmacher 2008
- Dajana Conte 2010
- Thorsten Rohwedder & Reinhold Schneider 2013
- Bart Vandereycken 2013, 2015
- Ivan Oseledets 2014, 2015
- Jutho Haegeman & Frank Verstraete 2014^{pre}
- Emil Kieri & Hanna Walach 2015^{pre}

Motivation: Data reduction / model reduction

- Encyclopedia: term-document matrix
  latent semantic indexing: low-rank approximation by a truncated SVD: $A \approx \sum_{j=1}^{r} \sigma_j u_j v_j^T$

  time-dependent encyclopedia? (à la Wikipedia)

- Multi-particle quantum dynamics
  time-dependent Schrödinger equation $i \frac{\partial \psi}{\partial t} = H\psi$
  for the wavefunction $\psi = \psi(x_1, \ldots, x_N, t)$

  MCTDH: reduced model via low-rank tensor approximation

Outline

Dynamical low-rank approximation

Differential equations

Splitting integrator

Robustness to small singular values

Extensions to tensors

Outline

Dynamical low-rank approximation

Differential equations

Splitting integrator

Robustness to small singular values

Extensions to tensors

Best low-rank approximation

given a matrix $A \in \mathbb{R}^{m \times n}$ (huge m, n)

best approximation to A of rank r: matrix X in the manifold $\mathcal{M}_r$ of rank-r matrices with

$$X \in \mathcal{M}_r \quad \text{such that} \quad \|X - A\| = \min!$$

Frobenius norm: problem solved by truncated SVD

$$X = U\Sigma_r V^T = \sum_{i=1}^{r} \sigma_i u_i v_i^T$$

Best low-rank approximation

given matrices $A(t) \in \mathbb{R}^{m \times n}$ for $0 \le t \le T$ (huge m, n)

best approximation to $A(t)$ of rank r: matrix $X(t)$ in the manifold $\mathcal{M}_r$ of rank-r matrices with

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

Frobenius norm: problem solved by truncated SVD, but expensive!

Need an updating technique

Dynamical low-rank approximation

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

replaced by initial value problem of differential equations on $\mathcal{M}_r$:

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

Motivation

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

- use sparse increments $\dot{A}(t)$ instead of the complete data matrix $A(t)$: time-continuous updating
- linear projection onto the tangent space instead of a nonlinear, nonconvex minimization problem
- extends to matrix differential equations $\dot{A} = F(A)$: minimum defect approximation

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

- ...

Outline

Non-unique factorisation of rank-r matrices

$$Y = USV^T$$

with invertible $S \in \mathbb{R}^{r \times r}$

$U \in \mathbb{R}^{m \times r}$ and $V \in \mathbb{R}^{n \times r}$ with orthonormal columns

---

SVD has diagonal S, is *not* assumed here!

Unique decomposition in the tangent space

non-unique decomposition $Y = USV^T \in \mathcal{M}_r$

tangent matrix $\dot{Y} \in T_Y\mathcal{M}_r$:

$$\dot{Y} = \dot{U}SV^T + U\dot{S}V^T + US\dot{V}^T$$

with $\dot{S} \in \mathbb{R}^{r\times r}$ and skew-symmetric $U^T\dot{U}$, $V^T\dot{V}$

$\dot{S}, \dot{U}, \dot{V}$ are uniquely determined by $\dot{Y}$ and S, U, V
under the orthogonality constraints

$$U^T\dot{U} = 0, \qquad V^T\dot{V} = 0$$

---

gauge conditions enforce uniqueness

Equivalent formulations of dynamical low-rank approximation

- $\dot{Y} \in T_Y\mathcal{M}_r$ such that $\|\dot{Y} - \dot{A}\| = \min!$
- $\langle \dot{Y} - \dot{A}, Z \rangle = 0$ for all $Z \in T_Y\mathcal{M}_r$
- $\dot{Y} = P(Y)\dot{A}$ with $P(Y) =$ orth. projection onto $T_Y\mathcal{M}_r$

ODEs for dynamical low-rank approximation

$$Y = USV^T$$

with

$$\dot{U} = (I_m - UU^T)\dot{A}VS^{-1}$$
$$\dot{V} = (I_n - VV^T)\dot{A}^T US^{-T}$$
$$\dot{S} = U^T\dot{A}V$$

Koch & L. 2007

cf. ODEs for SVD (*Wright* 1992 and *Dieci & Eirola* 1999)
but here, no singularities arise for coalescing singular values

Low-rank approximation of matrix ODEs

$$\dot{A} = F(A)$$

solution A approximated by $Y = USV^T$ with

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

---

minimum defect: $\dot{Y} \in T_Y\mathcal{M}_r$ with $\|\dot{Y} - F(Y)\| = \min!$

Outline

Equivalent formulations of dynamical low-rank approximation

- $\dot{Y} \in T_Y\mathcal{M}_r$ such that $\|\dot{Y} - \dot{A}\| = \min!$
- $\langle \dot{Y} - \dot{A}, Z \rangle = 0$ for all $Z \in T_Y\mathcal{M}_r$
- $\dot{Y} = P(Y)\dot{A}$ with $P(Y)$ = orth. projection onto $T_Y\mathcal{M}_r$:

$$P(Y)\dot{A} = \dot{A}P_{\mathcal{R}(Y^T)} - P_{\mathcal{R}(Y)}\dot{A}P_{\mathcal{R}(Y^T)} + P_{\mathcal{R}(Y)}\dot{A}$$

Idea: split the projection $\rightarrow$ <u>the</u> integrator (wait and see!)

L. & Oseledets 2014

Splitting integrator, abstract form

1. Solve the differential equation

$$\dot{Y}_I = \dot{A} P_{\mathcal{R}(Y_I^T)}$$

   with initial value $Y_I(t_0) = Y_0$ for $t_0 \leq t \leq t_1$.

2. Solve

$$\dot{Y}_{II} = -P_{\mathcal{R}(Y_{II})} \dot{A} P_{\mathcal{R}(Y_{II}^T)}$$

   with initial value $Y_{II}(t_0) = Y_I(t_1)$ for $t_0 \leq t \leq t_1$.

3. Solve

$$\dot{Y}_{III} = P_{\mathcal{R}(Y_{III})} \dot{A}$$

   with initial value $Y_{III}(t_0) = Y_{II}(t_1)$ for $t_0 \leq t \leq t_1$.

Finally, take $Y_1 = Y_{III}(t_1)$ as an approximation to $Y(t_1)$.

Solving the split differential equations

The solution of 1. is given by

$$Y_I = U_I S_I V_I^T \quad \text{with} \quad (U_I S_I)^{\cdot} = \dot{A} V_I, \quad \dot{V}_I = 0:$$

$$U_I(t) S_I(t) = (A(t) - A(t_0)) V_I(t_0), \qquad V_I(t) = V_I(t_0).$$

and similarly for 2. and 3.

Splitting integrator, practical form

Start from $Y_0 = U_0 S_0 V_0^T \in \mathcal{M}_r$.

1. With the increment $\Delta A = A(t_1) - A(t_0)$, set
$$K_1 = U_0 S_0 + \Delta A\, V_0$$
and orthogonalize:
$$K_1 = U_1 \widetilde{S}_1,$$
where $U_1 \in \mathbb{R}^{m\times r}$ has orthonormal columns, and $\widetilde{S}_1 \in \mathbb{R}^{r\times r}$.
2. Set $\widetilde{S}_0 = \widetilde{S}_1 - U_1^T\, \Delta A\, V_0$.
3. Set $L_1 = V_0 \widetilde{S}_0^T + \Delta A^T U_1$ and orthogonalize:
$$L_1 = V_1 S_1^T,$$
where $V_1 \in \mathbb{R}^{n\times r}$ has orthonormal columns, and $S_1 \in \mathbb{R}^{r\times r}$.

The algorithm computes a factorization of the rank-r matrix
$$Y_1 = U_1 S_1 V_1^T \approx Y(t_1).$$

Splitting integrator, cont.

- use symmetrized variant (Strang splitting)
- for a matrix differential equation $\dot{A} = F(A)$:
  in substep 1. solve

$$\dot{K} = F(KV_0^T)V_0, \qquad K(t_0) = U_0 S_0$$

  by a step of a numerical method (Runge-Kutta etc.),
  and similarly in substeps 2. and 3.

Outline

ODEs for dynamical low-rank approximation

$$Y = USV^T$$

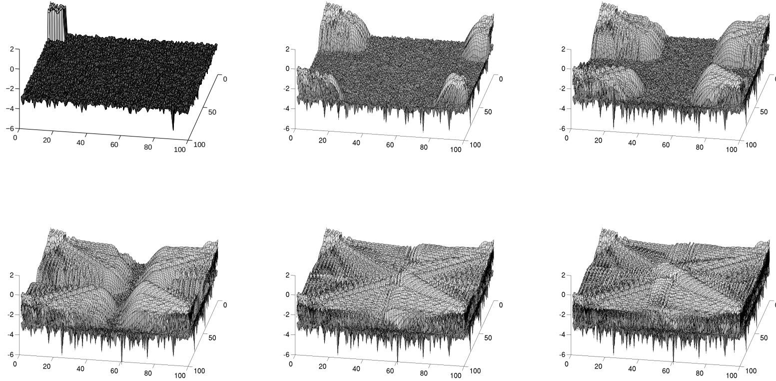
with

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

What if S is ill-conditioned? (effective rank smaller than r)

Numerical experiment

ε-perturbed rank-10 matrices for $t = 0, 0.2, \ldots, 1$, ε=1e$-$3.



Numerical experiment: errors at $t = 1$

Table: Approximation rank $r = 10$.

ε	$\|X - A\|$	$\|Y - A\|$	$\|S^{-1}\|$
1e−1	7.3762e+00	1.3478e+01	7.9878e−01
1e−2	9.3381e−01	5.2203e+00	1.4487e+00
1e−3	1.8293e−01	2.1549e−01	2.6232e+00
1e−4	1.8310e−02	2.1550e−02	2.6232e+00

Table: Approximation rank $r = 20$.

ε	$\|X - A\|$	$\|Y - A\|$	$\|S^{-1}\|$
1e−1	6.0335e+00	1.3094e+01	1.5749e+00
1e−2	6.1246e−01	1.0885e+00	1.3569e+01
1e−3	6.1280e−02	1.0354e−01	1.3474e+02
1e−4	6.1282e−03	1.0298e−02	1.2940e+03

Robustness under over-approximation

Theorem

- $A(t) = X_q(t) + E(t)$ *with* $\|E(t)\| \leq \varepsilon_0$, $\|\dot{E}(t)\| \leq \varepsilon_1$ *for* $X_q(t) \in \mathcal{M}_q$ *with* $q \leq r$
- $\sigma_q(A(t)) \geq \rho > 0$
- $\|\dot{A}(t)\| \leq \mu$
- $Y(0) = X_r(0) \in \mathcal{M}_r$

Then,

$$\|Y(t) - X_q(t)\| \leq \varepsilon_0 + 6t\varepsilon_1 \quad \textit{for} \;\; t \leq \frac{\rho}{16\mu}$$

---

good rank-r approximation even in case of effective rank $q < r$

Numerical experiment: integrator errors ($h = 10^{-3}$)

	p	Appr. err.
Midpoint	2.0023	0.2200
KLS	1.0307	1.8133
KLS(symm)	1.8226	0.2215
KSL	1.0089	0.2188
KSL(symm)	2.005	0.2195

Table: $\varepsilon = 10^{-3}, r = 10$

	p	Appr. err.
Midpoint	2.0024	0.0188
KLS	1.0309	1.8030
KLS(symm)	1.8231	0.0324
KSL	1.0082	0.0002
KSL(symm)	2.0049	0.0002

Table: $\varepsilon = 10^{-6}, r = 10$

	p	Appr. err.
Midpoint	0.0001	0.1006
KLS	0.8154	1.4224
KLS(symm)	1.4911	0.3142
KSL	1.0354	0.0913
KSL(symm)	1.9929	0.0913

Table: $\varepsilon = 10^{-3}, r = 20$

	p	Appr. err.
Midpoint	-	failed
KLS	0.9633	1.3435
KLS(symm)	0.3127	1.5479
KSL	1.0362	9.1316e-05
KSL(symm)	1.993	9.1283e-05

Table: $\varepsilon = 10^{-6}, r = 20$

An exactness result for the splitting method

Theorem

If $A(t)$ has rank at most r, then the splitting integrator is exact:

$$Y_1 = A(t_1)$$

Ordering of the splitting is essential! (KSL, not KLS)

Approximation is robust to small singular values

$$\dot{A} = F(t, A), \qquad A(t_0) = Y_0 \in \mathcal{M}_r$$

- F is locally Lipschitz-continuous
- $\|(I - P(Y))F(t, Y)\| \leq \varepsilon$ for all $Y \in \mathcal{M}_r$.

$Y_n \in \mathcal{M}_r$ result of the projector-splitting integrator after n steps

Theorem

$$\|Y_n - A(t_n)\| \leq c_1\varepsilon + c_2 h \qquad \textit{for} \;\; t_n \leq T,$$

where c_1, c_2 depend only on the local Lipschitz constant and bound of F, and on T.

Kieri, L. & Walach 2015[pre]

Remarks on the proof

The method splits $P(Y) = P_I(Y) - P_{II}(Y) + P_{III}(Y)$ in

$$\dot{Y} = P(Y)F(t, Y).$$

Difficulty: cannot use the Lipschitz continuity of the tangent space projection $P(\cdot)$ and its subprojections, because the Lipschitz constants become large for small singular values

Rescue:

- use the previous exactness result
- use the conservation of the subprojections in the substeps

Outline

Tensors in Tucker format

Approximation of time-dependent tensors

$$A(t) \in \mathbb{R}^{n_1 \times \cdots \times n_d} \quad \text{with entries} \quad A(k_1, \ldots, k_d; t)$$

by tensors $Y(t) \in \mathbb{R}^{n_1 \times \cdots \times n_d}$ of the form, with $r_i \ll n_i$,

$$Y(k_1, \ldots, k_d; t) = \sum_{j_1=1}^{r_1} \cdots \sum_{j_d=1}^{r_d} c_{j_1,\ldots,j_d}(t)\, u_{j_1}^{(1)}(k_1; t) \ldots u_{j_d}^{(d)}(k_d; t),$$

Tucker format of multilinear rank $r = (r_1, \ldots, r_d)$:

- Each 1-mode matrix unfolding of the core tensor has full rank.
- The vectors $u_1^{(i)}(t), \ldots, u_{r_i}^{(i)}(t) \in \mathbb{R}^{n_i}$ are orthonormal.

Dynamical tensor approximation: Koch & L. 2010
Projector-splitting integrator: L. 2015[pre]

Tensor trains / matrix product states

Non-commutative separation of variables:

$$Y(k_1, \dots, k_d) = G_1(k_1) \dots G_d(k_d)$$

where the $G(k_i)$ are $r_{i-1} \times r_i$ matrices, with $r_0 = r_d = 1$.

Attractive because of low data requirement: dKr^2

Matrix product states:	e.g., Verstraete, Murg & Cirac 2008
Tensor trains:	Oseledets 2011
Manifold structure:	Holtz, Rohwedder & Schneider 2012
	Uschmajew & Vandereycken 2013
Dynamical approximation in the tensor train format:	
	L., Rohwedder, Schneider, Vandereycken 2013
Projector-splitting integrator:	L., Oseledets & Vandereycken 2015
in physics context:	Haegeman, L., O., V. & Verstraete 2014[pre]
Error analysis:	Kieri, L. & Walach 2015[pre]

Outline

Dynamical low-rank approximation

Differential equations

Splitting integrator

Robustness to small singular values

Extensions to tensors

Review

C.L., Low-rank dynamics

in Extraction of Quantifiable Information from Complex Systems (S. Dahlke et al., eds.), Springer Lecture Notes Comput. Sci. Eng. 102, 2014.

na.uni-tuebingen.de