Algorithms for dynamical low-rank approximation

Bart Vandereycken

Section of mathematics
University of Geneva

IPAM Workshop
Mathematical Foundations and Algorithms for Tensor Computations
May 2021

A matrix/tensor-valued ODE

Consider the initial value problem

$$\dot{X}(t) = F(t, X(t)), \quad X(0) = X_0,$$

with smooth $F\colon \mathbb{R} \times \mathcal{H} \to \mathcal{H}$ and $X(t) \in \mathcal{H}$ is a time-dependent

- **matrix** in $\mathbb{R}^{n_1 \times n_2}$;
- d-dimensional tensor (array) in $\mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d}$.

Suppose $X(t)$ can be well approximated by a *low rank matrix/tensor*. Applications: Schrödinger eq., stochastic/parametric eq.'s, kinetic eq.'s, ...

Topic of this talk: Provably compute this low rank approximation efficiently.

Overview:

- Approximation: DLRA and TDVP
- Algorithms: intrinsic and extrinsic
- Analysis: robustness to small singular values

Low-rank approximation

- Take $\mathcal{M} \subset \mathcal{H}$ the approximation space of matrices (tensors) of (TT) rank r.

  $\rightsquigarrow$ $\mathcal{M}$ is a smooth embedded manifold of $\mathcal{H}$.

- Given $\dot{X} = F(t, X)$, we would like to compute

  $$Y(t) \in \mathcal{M} \quad \text{s.t.} \quad \|X(t) - Y(t)\| \to \min.$$

  $\rightsquigarrow$ The SVD gives a (quasi) best approximation *if* $X(t)$ *is known in all* t.

- Instead, we compute the dynamical low-rank approximation (DLRA)

  $$\dot{Y}(t) \in T_{Y(t)}\mathcal{M} \qquad \text{s.t.} \quad \|\dot{X}(t) - \dot{Y}(t)\| \to \min.$$

  $\rightsquigarrow$ Dirac–Frenkel time-dependent variational principle (TDVP) [Dirac 30].

- Many names: MCDTH [Worth/Beck/Jäckle/Meyer 07], DLRA [Koch/Lubich 07], Dynamically orthogonal field eqs. [Sapsis/Lermusiaux 09], TDVP [Haegeman *et.al.* 11], Bi-orthogonal decomposition [Cheng/Hou/Zhang 13], . . .

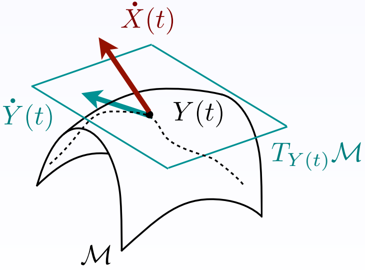
Projected vector field

Recall $\dot{X} = F(t, X)$ and the dynamical low-rank approximation

$$\dot{Y}(t) \in T_{Y(t)}\mathcal{M} \qquad \text{s.t.} \quad \|\dot{X}(t) - \dot{Y}(t)\| \to \min.$$

Hence, by construction $Y(t) \in \mathcal{M}$ if $Y(0) \in \mathcal{M}$.

In a picture:



Locally equivalent to:

$$\dot{Y} = \mathrm{P}(Y)(F(t, Y)) \qquad \text{(DLRA)}$$

with the orthogonal projection

$$\mathrm{P}(Y)\colon \mathcal{H} \to T_Y\mathcal{M}.$$

How do we integrate (DLRA) in practice / numerically?

Geometry of matrices of fixed rank

A smooth submanifold embedded in $\mathcal{H} = \mathbb{R}^{n_1 \times n_2}$:

$$\mathcal{M} = \{X \in \mathbb{R}^{n_1 \times n_2} : \ \mathrm{rank}(X) = r\}.$$

- Compact representation $X = USV^\intercal$ where

$$U \in \mathbb{R}^{n_1 \times r},\ U^\intercal U = I, \quad S \in \mathbb{R}^{r \times r}, \quad V \in \mathbb{R}^{n_2 \times r},\ V^\intercal V = I.$$

- Tangent space of all first-order variations to $\mathcal{M}$ at X:

$$T_X\mathcal{M} = \{\delta U\, S\, V^\intercal + U\, \delta S\, V^\intercal + U\, S\, \delta V^\intercal \ : \ \delta U, \delta S, \delta V\}$$

  can be **uniquely parametrized** using the normalization (horizontal space)

$$\mathcal{H}_X = \{\delta U \in \mathbb{R}^{n_1 \times k}, \quad \delta S \in \mathbb{R}^{k \times k}, \quad \delta V \in \mathbb{R}^{n_2 \times k} \ : \ U^\intercal \delta U = 0,\ V^\intercal \delta V = 0\}.$$

- Orthogonal projection $\mathrm{P}(X) \colon \mathbb{R}^{n_1 \times n_2} \to T_X\mathcal{M} \simeq \mathcal{H}_X$:

$$Z \mapsto \begin{cases} \delta U = \mathrm{P}_U^{\perp}\ Z\, V S^{-1} \\ \delta S = U^\intercal\, Z\, V \\ \delta V = \mathrm{P}_V^{\perp}\ Z^\intercal\, U S^{-\intercal} \end{cases}$$

Efficient implementation of DLRA for matrices

DLRA in (normalized) parameter space:

$$\begin{cases} \dot{U} = \mathrm{P}_U^{\perp} F(t,Y) V S^{-1}, & U(0) = U_0 \in \mathbb{R}^{n_1 \times r} \\ \dot{S} = U^{\intercal} F(t,Y) V, & S(0) = S_0 \in \mathbb{R}^{r \times r} \\ \dot{V} = \mathrm{P}_V^{\perp} F(t,Y)^{\intercal} U S^{-\intercal}, & V(0) = V_0 \in \mathbb{R}^{n_2 \times r} \end{cases} \tag{$*$}$$

- When $r \ll n$, the matrices U, S, V are small.

  $\rightsquigarrow$ Can integrate $(*)$ using standard methods (like Runge–Kutta).

- Important structure: F has low operator rank (MPO)

  $$\text{if } \operatorname{rank}(Y) = k \text{ then } \operatorname{rank}(F(t,Y)) = \text{poly}(k).$$

  $\rightsquigarrow$ $F(t,Y)V$ and $F(t,Y)^{\intercal}U$ is reasonably fast.

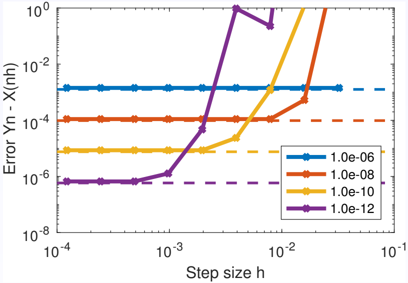
- Example: $F(t,Y) = HY + YH + Y \cdot |Y|^2$ has rank $\leq 2k + k^3$

Numerical experiment

Integrate DLRA in (normalized) parameter space.

Synthetic example $X(t) = e^{t\Omega_1} e^{tD} e^{-t\Omega_2}$

Check influence of smallest singular val.: $\sigma_{\min}(X_0) \in \{10^{-6}, 10^{-8}, 10^{-10}, 10^{-12}\}$.



- **Step size restriction**: Asymptotic decay $h \leq h_0$ but $h_0 \sim \sigma_{\min}(X_0) \to 0$.
- Error $\|Y_n - Y(nh)\| \leq Ch^p$ satisfies

$$C \propto e^L \qquad \text{with} \qquad L \propto \|S^{-1}\|_2 = 1/\sigma_r(Y)$$

$\rightsquigarrow$ related to curvature of $\mathcal{M}$:

$$\| \mathrm{P}(X)(Z) - \mathrm{P}(Y)(Z)\| \lesssim \frac{1}{\sigma_r(X)} \|X - Y\| \|Z\|.$$

DLRA for tensors

⊕ DLRA can be formulated for SVD-based tensor networks in trees: Tucker [Koch/Lubich 10], MPS [Haegeman/Cirac/Osborne/Pižorn/Verschelde/Verstraete 10], TT [Holtz/Rohwedder/Schneider 12], HT [Uschmajew/V. 13], general tree networks [Ceruti/Lubich/Wallach 21]

- Smooth embedded submanifolds in $\mathcal{H}$ with efficient projections $\mathrm{P}(Y)$
- Normalized parameter space of small dimension.
- Quasi-optimal approximation properties (for small time) [Koch/Lubich 07,10], [Lubich/Rohwedder/Schneider/V. 13], [Musharbash/Nobile/Zhou 15]

⊖ The DLRA differential equations in parameter space use the inverse of (reduced) Gram matrices of $Y(t)$.

- Suffer from step size restriction when $\sigma_r \to 0$.
- Conclusion: DLRA in parameter space is very expensive to integrate numerically when $\sigma_r \to 0$. Better integrators are needed!

Better integrators for DLRA

Robust DLRA integrators suitable for SVD-based tensor formats.

Intrinsic (never leaves $\mathcal{M}$)

- Exploits that $\mathcal{M}$ is like a ruled surface.
- Preserves many properties of F, like energy and norm conservation.
- [Lubich/Oseledets 14], [Lubich/Oseledets/V. 15], [Haegeman *et.al* 16]

Extrinsic (leaves $\mathcal{M}$)

- Exploits that we can quasi-project onto $\mathcal{M}$ using HOSVD.
- Naturally rank adaptive.
- [Kieri/V. 18], [Feppon/Lermusiaux 18], [Rodgers/Dektor/Venturi 21], ...

Both: provably robust error for $\sigma_r \to 0$ when F is close to the tangent bundle:

$$\|F(Y) - \mathrm{P}(Y)(F(Y))\| \leq \varepsilon \quad \text{for all } Y \in \mathcal{M} \text{ near } X(t)$$

Intrinsic: projector splitting [Lubich/Oseledets 14]

Let $X = USV^\intercal$ and $\mathrm{P}_U = UU^\intercal$. Rewrite orthogonal projection onto $T_X\mathcal{M}$ as

$$\begin{aligned} \mathrm{P}(X)(Z) &= \mathrm{P}_U\, Z\, \mathrm{P}_V + \mathrm{P}_U^\perp\, Z\, \mathrm{P}_V + \mathrm{P}_U\, Z\, \mathrm{P}_V^\perp \\ &= Z\, \mathrm{P}_V - \mathrm{P}_U\, Z\, \mathrm{P}_V + \mathrm{P}_U\, Z \end{aligned}$$

First-order scheme: Lie–Trotter splitting for $\dot{Y} = \mathrm{P}(Y)F(Y)$

1) integrate for $t \in [t_0, t_1]$: $\dot{Y}_1 = +F(Y)\,\mathrm{P}_{V_1(t)}$, $Y_1(t_0) = Y(t_0)$
2) integrate for $t \in [t_0, t_1]$: $\dot{Y}_2 = -\mathrm{P}_{U_2(t)}\,F(Y)\,\mathrm{P}_{V_2(t)}$, $Y_2(t_0) = Y_1(t_1)$
3) integrate for $t \in [t_0, t_1]$: $\dot{Y}_3 = +\mathrm{P}_{U_3(t)}\,F(Y)$, $Y_3(t_0) = Y_2(t_1)$

where $Y_i(t) = U_i(t)S_i(t)V_i(t) \in \mathcal{M}$. Result is $Y_3(t_1) \simeq Y(t_1)$.

Strang splitting: Composing a forward half step with a backward (adjoint) half step is **second-order accurate** at nearly the same cost.

Important observation: The substeps can be solved much more easily . . .

Intrinsic: Practical computation

Let $Y = USV^\intercal$ and $K = US$. The first substep for $Y(t) = K(t)V(t)^\intercal$ satisfies

$$F(KV^\intercal)VV^\intercal = \dot{Y} = \dot{K}\,V^\intercal + K\,\dot{V}^\intercal,\ Y(0) = Y_0.$$

We can choose $\dot{V} = 0$ since

$$\dot{K}\,V_0^\intercal = F(KV_0^\intercal)V_0V_0^\intercal \quad \Rightarrow \quad \dot{K} = F(KV_0^\intercal)V_0,\ \ K(0) = U_0\,S_0.$$

- Substep evolves in a **flat** subspace $\rightsquigarrow$ curvature will not restrict step size!
- If F is low-rank operator, $F(KV_0^\intercal)V_0$ can be evaluated efficiently.
- Rest of the scheme is derived similarly:

$$\begin{array}{lll}
\text{1a) integrate for } t \in [t_0, t_1]: & \dot{K} = +F(KV_0^\intercal)V_0, & K(t_0) = U_0S_0 \\
\quad \text{1b) update with QR:} & K(t_1) = U_1R_1, \quad S_1 = R_1S_0 & \\
\text{2a) integrate for } t \in [t_0, t_1]: & \dot{S} = -U_1F(U_1SV_0^\intercal)V_0, & S(t_0) = S_1 \\
\text{3a) integrate for } t \in [t_0, t_1]: & \dot{L} = +F(U_1L^\intercal)^\intercal U_1, & L(t_0) = V_1S(t_1)^\intercal
\end{array}$$

Result is $Y_3(t_1) = U_1L(t_1)^\intercal$.

Properties

Theorem ([Kieri/Lubich/Wallach '16]) For $0 \leq nh \leq T$, the DLRA error for projector-splitting with $Y_0 = X_0$ satisfies

$$\|Y_n - X(nh)\| \leq C\,(\varepsilon + \boldsymbol{h}),$$

where C does not depend on h or $\sigma_r(Y_n)$. In addition, if $X(nh) \in \mathcal{M}$, then there is no error: $Y_n = X(nh)$.

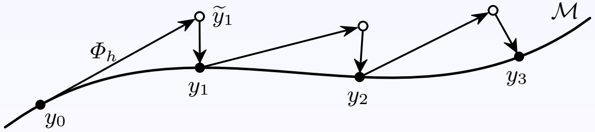
Can be extended to tensor networks based on trees by a clever ordering and splitting of the tangent space projector

$$\mathrm{P}(X) = \sum_{i=1}^{d-1} (\mathrm{P}_{\leq i-1}\,\mathrm{P}_{\geq i+1} - \mathrm{P}_{\leq i}\,\mathrm{P}_{\geq i+1}) + \mathrm{P}_{\leq d-1}\,\mathrm{P}_{\geq d+1},$$

Details: [Lubich/Oseledets/V. 15], [Haegeman *et.al* 16], [Ceruti/Lubich/Wallach 21].

Extrinsic: projection method

Simple and standard technique: **projection method**.



(1) Calculate / approximate flow over on $[t_i, t_i + h]$: $\widetilde{Y}_{i+1} = \Phi_h(Y_i)$.

(2) Project back to $\mathcal{M}$: $Y_{i+1} = \min\{X - \widetilde{Y}_{i+1} : X \in \mathcal{M}\}$.

Advantage:

1. Can use any standard integrator for F.
2. Quasi-optimal projection $\mathcal{R} : \mathcal{H} \to \mathcal{M}$ from HOSVD is sufficient.
3. These projections do not destroy high order local error

Questions: Robust when $\sigma_r(Y) \to 0$? Efficient projection?

Geometric numerical integration, [Hairer/Lubich/Wanner 06]

Higher-order Runge–Kutta: algorithm [Kieri/V. 18]

Faster convergence using high-order RK methods.

- Applied to $\dot{X} = F(X)$ on $\mathcal{H}$:

$$\widetilde{Z}_j = X_i + h \sum_{l=1}^{j-1} a_{jl} F(\widetilde{Z}_l), \quad j = 1, \ldots, s,$$
$$X_{i+1} = X_i + h \sum_{j=1}^{s} b_j F(\widetilde{Z}_j).$$

- Extend to $\dot{Y} = \mathrm{P}(Y)(F(Y))$ on $\mathcal{M}$ by internal projection

$$Z_j = Y_i + h \sum_{l=1}^{j-1} a_{jl} \mathrm{P}(\mathcal{R}(Z_l)) F(\mathcal{R}(Z_l)), \quad j = 1, \ldots, s,$$
$$Y_{i+1} = \mathcal{R}(Y_i + h \sum_{j=1}^{s} b_j \mathrm{P}(\mathcal{R}(Z_j)) F(\mathcal{R}(Z_j))).$$

$\rightsquigarrow$ Lucky linear algebra: computing $\mathcal{R}(Z_j)$ is $O(dj^2r^2(n+r^2))$.

Higher-order Runge–Kutta: analysis

Theorem ([Kieri/V. '18]) Given an explicit Runge–Kutta method of order p and stage orders $q_1 \leq q_2 \leq \cdots \leq q_s$. Denote

$$q = \begin{cases} \min(p, q_2 + 1), & \text{if } b_2 \neq 0, \\ \min(p, q_3 + 1, q_2 + 2), & \text{if } b_2 = 0. \end{cases}$$

Then, for $0 \leq nh \leq T$, the DLRA error for projected RK with $Y_0 = X_0$ satisfies

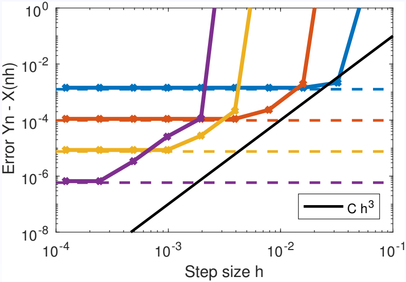
$$\|Y_n - X(nh)\| \leq C\,(\varepsilon + \boldsymbol{h^q}),$$

where C does not depend on h or $\sigma_r(Y_n)$.

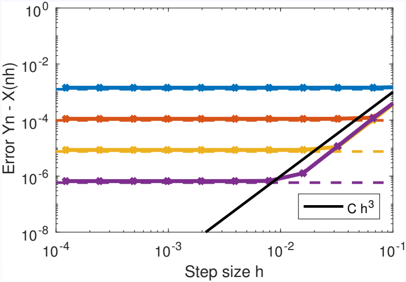
- When $p \leq 3$, we obtain the same order as original method: $q = p$.
- For high order methods, internal stage orders limit final order.
- Nevertheless, improves on KSL where robustness only for $p = 1$.

Numerical experiment

Synthetic example from before: $X(t) = e^{t\Omega_1} e^{tD} e^{-t\Omega_2}$.



RK3 on U, S, V



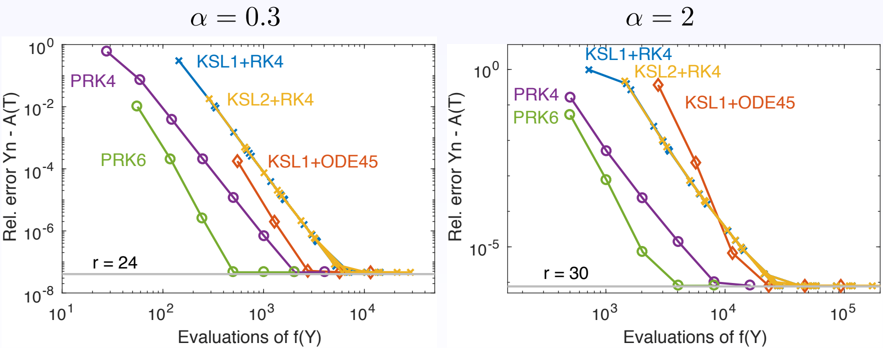
projected RK3

- Robust error behavior, higher order.
- Slightly higher cost ($O(s^2 r^2)$ vs. $O(sr^2)$) for s stages.

Numerical results

Schrödinger equation with cubic non-linearity:

$$\mathrm{i}\dot{X} = -\tfrac{1}{2}(BX + XB) - \alpha|X|^2 X.$$



- Higher order is more efficient.
- Integration of subproblems in projector splitting need to be done carefully. Here we did not exploit the structure of the original ODE (e.g., Krylov, exponential, Strang splitting).

Summary and outlook

To **summarize**, dynamical low-rank approximation:

- avoid integrating in parameter space;
- extrinsic and intrinsic approaches are preferable;
- generalizable to tensor networks on trees.

Outlook and not discussed:

- Stiff problems ($\ell \ll L$): [Ostermann/Piazzola/Walach 19], [Conte 20]
- Rank adaptivity: [Rodgers/Dektor/Venturi 21], [Dektor/Rodger/Venturi 21], [Schrammer 21]
- Parallelism in space [Secular *et.al* 20] and in time [Carrel/V. 2X]
- Hybrid schemes: enrich subspace of projector splitting [Yang/White 20], [Ceruti/Kusch/Lubich 21]
- Kinetic equations: [Einkemmer/Lubich 19], [Einkemmer/Yu/Ying 21]

Thank you for listening