Dynamical Low Rank Approximation in Optimal Control and Molecular Dynamics

Martin Eigel
Reinhold Schneider
David Sommer

RG4

June 6, 2022

Motivation: Automated heating of buildings

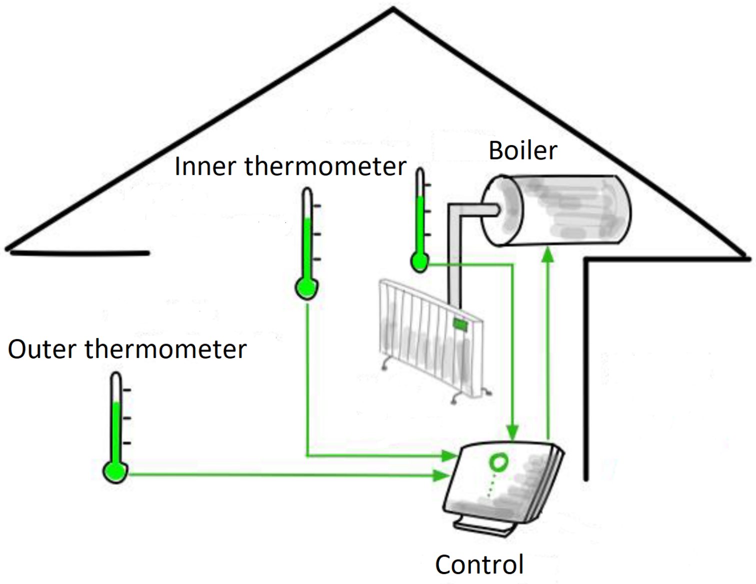


Figure 1: Control Problem: Heating of a building

- heating/cooling power $u(t) \in \mathbb{R}^m$.
- room temperatures $x(t) \in \mathbb{R}^d$.
- target temperatures $x_T(t) \in \mathbb{R}^d$.
- Running cost

$$\ell(t, x, u) = \|x - x_T(t)\|_Q^2 + \lambda \|u\|_R^2.$$

- Dynamics modelled by ODE $\dot{x} = f(t, x, u)$

Problem (Infinite Horizon)

$$\min_u \int_0^\infty \ell(t, x(t), u(t))dt,$$
$$\dot{x}(t) = f(t, x(t), u(t)), \quad x(0) = x_0.$$

$\hookrightarrow$ Often infeasible.

Model Predictive Control

Problem (Infinite Horizon)

$$
\begin{aligned}
&\min_u \int_0^\infty \ell(t, x(t), u(t))dt, \\
&\dot{x}(t) = f(t, x(t), u(t)), \quad x(0) = x_0.
\end{aligned}
$$

Model Predictive Control

Problem (Finite Horizon)

Find $\bar{u}(t)$ as a solution to

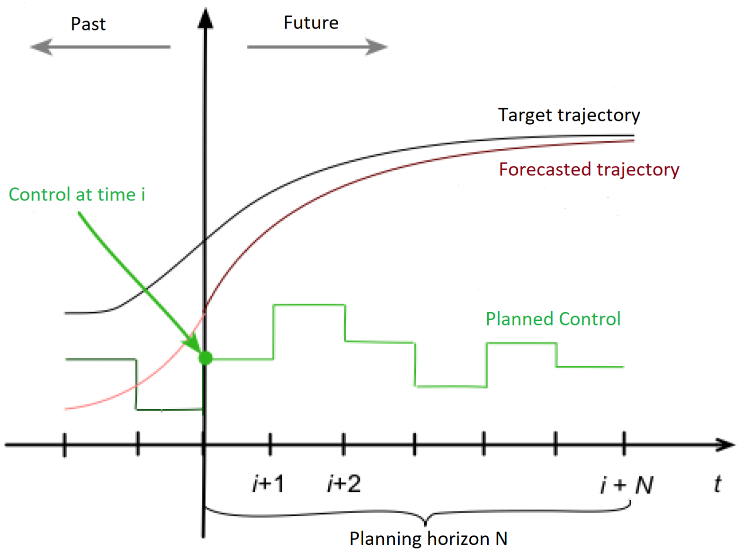
$$\min_{\bar{u}} \int_{t_i}^{t_i+T} \ell(\bar{x}(t), \bar{u}(t))dt + E(\bar{x}(t_i + T)),$$
$$\dot{\bar{x}}(t) = f(\bar{x}(t), \bar{u}(t)), \quad \bar{x}(t_i) = x(t_i).$$

Model Predictive Control

Problem (Finite Horizon)

Find $\bar{u}(t)$ as a solution to

$$\min_{\bar{u}} \int_{t_i}^{t_i+T} \ell(\bar{x}(t), \bar{u}(t))dt + E(\bar{x}(t_i + T)),$$
$$\dot{\bar{x}}(t) = f(\bar{x}(t), \bar{u}(t)), \quad \bar{x}(t_i) = x(t_i).$$



At each time step $t_i = i\delta$, $i \in \mathbb{N}_0$,

- measure the state $x(t_i)$
- solve problem (FH) for $\bar{u}$
- Apply the input

$$u_{MPC}(t) = \bar{u}(t; t_i, x(t_i))$$

for $t \in [t_i, t_i + \delta)$ to the system.

Downside of standard MPC:

- State feedback only at discrete time points $t_i = i\delta$
  $\hookrightarrow$ δ needs to be small for robustness
- Repeated online computation necessary
  $\hookrightarrow$ Each optimization needs to be performed in time δ.
  $\hookrightarrow$ Need to approximate, i.e. piecewise constant controls.

Possibility to alleviate this problem:

MPC with **feedback laws**.

Feedback MPC

Problem (Feedback Finite Horizon)

Find $\alpha(t,x)$ *as a solution to*

$$\min_{\alpha\in\mathcal{A}} \int_0^T \ell(\bar{x}(t),\alpha(t,\bar{x}(t)))dt + E(\bar{x}(T)),$$
$$\dot{\bar{x}}(t) = f(\bar{x}(t),\alpha(t,\bar{x}(t))), \quad \bar{x}(0) = x_0,$$

for all $x_0 \in \Omega \subset \mathbb{R}^d$.

Feedback MPC

Problem (Feedback Finite Horizon)

Find $\alpha(t,x)$ *as a solution to*

$$\min_{\alpha\in\mathcal{A}} \int_0^T \ell(\bar{x}(t),\alpha(t,\bar{x}(t)))dt + E(\bar{x}(T)),$$
$$\dot{\bar{x}}(t) = f(\bar{x}(t),\alpha(t,\bar{x}(t))), \quad \bar{x}(0) = x_0,$$

for all $x_0 \in \Omega \subset \mathbb{R}^d$.

Solve (FFH) for $\alpha(t,x)$ **offline**.
At each time interval $t_i = i\delta$, $i \in \mathbb{N}_0$,

- apply input

$$u_{MPC}(t) = \alpha(t,x(t)) \quad \text{for } t \in [t_i, t_i+\delta)$$

  to the system.

The Hamilton Jacobi Bellman equation

Assumptions:

- Dynamics linear in u: $f(x,u) = f(x) + g(x)u$
- Cost quadratic in u: $\ell(x,u) = c(x) + \|u\|_R^2$

Problem (HJB Equation)

Find the value function V as the solution to

$$\frac{\partial}{\partial t}V(t,x) + \nabla_x V(t,x)^\intercal(f(x) + g(x)\alpha(t,x)) + c(x) + \|\alpha(t,x)\|_R^2 = 0, \quad (1)$$

$$V(T,\cdot) = E, \quad (2)$$

where the optimal feedback control α satisfies

$$\alpha(t,x) = -\frac{1}{2}R^{-1}g(x)^\intercal\nabla_x V(t,x). \quad (3)$$

The Hamilton Jacobi Bellman equation

Assumptions:

- Dynamics linear in u: $f(x,u) = f(x) + g(x)u$
- Cost quadratic in u: $\ell(x,u) = c(x) + \|u\|_R^2$

Problem (HJB Equation)

Find the value function V as the solution to

$$\frac{\partial}{\partial t}V(t,x) + \nabla_x V(t,x)^\intercal (f(x) + g(x)\alpha(t,x)) + c(x) + \|\alpha(t,x)\|_R^2 = 0, \quad (1)$$

$$V(T,\cdot) = E, \quad (2)$$

where the optimal feedback control α satisfies

$$\alpha(t,x) = -\frac{1}{2}R^{-1}g(x)^\intercal \nabla_x V(t,x). \quad (3)$$

- Choose approximation class $\longrightarrow$ Tensor Trains with polynomial basis
- Choose approximation method $\longrightarrow$ Dynamical low rank approximation

The Tensor Train format

Consider one dimensional basis $\{\phi_i : \mathbb{R} \to \mathbb{R}\}_{i=1}^n$ and a fully tensorized function

$$v(x) = A\Phi(x) = \sum_{i_1,\dots,i_d=1}^{n} A(i_1,\dots,i_d)\phi_{i_1}(x_1)\cdot\ldots\cdot\phi_{i_d}(x_d)$$

The Tensor Train format

Consider one dimensional basis $\{\phi_i : \mathbb{R} \to \mathbb{R}\}_{i=1}^n$ and a fully tensorized function

$$v(x) = A\Phi(x) = \sum_{i_1,\ldots,i_d=1}^{n} A(i_1,\ldots,i_d)\phi_{i_1}(x_1)\cdot\ldots\cdot\phi_{i_d}(x_d)$$

$\hookrightarrow$ storage of $A \in \mathbb{R}^{n\times n}$ is $\mathcal{O}(n^d)$, curse of dimensionality.

The Tensor Train format

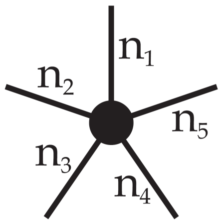
Consider one dimensional basis $\{\phi_i : \mathbb{R} \to \mathbb{R}\}_{i=1}^{n}$ and a fully tensorized function

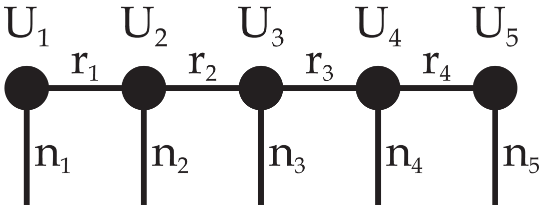
$$v(x) = A\Phi(x) = \sum_{i_1,\dots,i_d=1}^{n} A(i_1,\dots,i_d)\phi_{i_1}(x_1)\cdot\ldots\cdot\phi_{i_d}(x_d)$$

$\hookrightarrow$ storage of $A \in \mathbb{R}^{n\times n}$ is $\mathcal{O}(n^d)$, curse of dimensionality.

Instead, learn low rank Tensor Train (TT) approximation of A [Ose11, OSS19]:

$$A(i_1,\dots,i_d) \approx \sum_{k_1,\dots,k_{d-1}}^{r_1,\dots,r_{d-1}} U_1(i_1,k_1)U_2(k_1,i_2,k_2)\cdot\ldots\cdot U_d(k_{d-1},i_d) \tag{4}$$





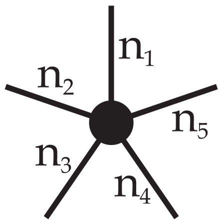
The Tensor Train format

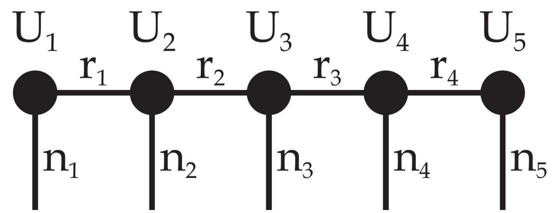
Consider one dimensional basis $\{\phi_i : \mathbb{R} \to \mathbb{R}\}_{i=1}^{n}$ and a fully tensorized function

$$v(x) = A\Phi(x) = \sum_{i_1,\dots,i_d=1}^{n} A(i_1,\dots,i_d)\phi_{i_1}(x_1)\cdot\ldots\cdot\phi_{i_d}(x_d)$$

$\hookrightarrow$ storage of $A \in \mathbb{R}^{n\times n}$ is $\mathcal{O}(n^d)$, curse of dimensionality.
Instead, learn low rank Tensor Train (TT) approximation of A [Ose11, OSS19]:

$$A(i_1,\dots,i_d) \approx \sum_{k_1,\dots,k_{d-1}}^{r_1,\dots,r_{d-1}} U_1(i_1,k_1)U_2(k_1,i_2,k_2)\cdot\ldots\cdot U_d(k_{d-1},i_d) \qquad (4)$$





$\hookrightarrow$ $\mathcal{O}(ndr^2)$, curse lifted! (provided the ranks r_i are bounded)
If equality holds in (4), $\mathbf{r} = (r_1,\dots,r_{d-1})$ is called the TT-rank of A.

Dynamical Low Rank Approximation ([KL10, CKL22])

Consider a tensor valued ODEs of the form

$$\dot{A}(t) = F(t, A(t)), \tag{5}$$
$$A(0) = A_0, \tag{6}$$

where $A(t) \in \mathbb{R}^{n \times \ldots \times n}$.

Dynamical Low Rank Approximation ([KL10, CKL22])

Consider a tensor valued ODEs of the form

$$\dot{A}(t) = F(t, A(t)), \tag{5}$$
$$A(0) = A_0, \tag{6}$$

where $A(t) \in \mathbb{R}^{n\times\ldots\times n}$.

The set

$$\mathcal{M}_{\mathbf{r}} = \{A \in \mathbb{R}^{n\times\ldots\times n} : A \text{ has TT-rank } \mathbf{r}\}$$

defines a smooth manifold in the full tensor space $\mathbb{R}^{n\times\ldots\times n}$.

The tangent space at a point $A \in \mathcal{M}_{\mathbf{r}}$ is denoted $\mathcal{T}_A(\mathcal{M}_{\mathbf{r}})$.

Dynamical Low Rank Approximation ([KL10, CKL22])

Consider a tensor valued ODEs of the form

$$\dot{A}(t) = F(t, A(t)), \tag{5}$$
$$A(0) = A_0, \tag{6}$$

where $A(t) \in \mathbb{R}^{n\times\ldots\times n}$.

The set

$$\mathcal{M}_{\mathbf{r}} = \{A \in \mathbb{R}^{n\times\ldots\times n} : A \text{ has TT-rank } \mathbf{r}\}$$

defines a smooth manifold in the full tensor space $\mathbb{R}^{n\times\ldots\times n}$.

The tangent space at a point $A \in \mathcal{M}_{\mathbf{r}}$ is denoted $\mathcal{T}_A(\mathcal{M}_{\mathbf{r}})$.

A DLRA approximation $Y(t) \in \mathcal{M}_{\mathbf{r}}$ of $A(t)$ is defined by

$$\dot{Y}(t) = \underset{\vartheta\in\mathcal{T}_{Y(t)}(\mathcal{M}_{\mathbf{r}})}{\arg\min} \|\vartheta - F(t, Y(t))\|_F, \tag{7}$$
$$Y(0) = Y_0, \tag{8}$$

where $Y_0 \in \mathcal{M}_{\mathbf{r}}$ is an approximation of the initial condition A_0.

DLRA applied to the HJB equation

Formally, we can write the HJB equation as

$$\partial_t V(t,x) = [\mathcal{L}V](t,x), \tag{9}$$

where $\mathcal{L}$ is a nonlinear differential operator.

Given: basis Φ, TT-rank $\mathbf{r}$

Goal: find approximate solution $\hat{V}(t,x) = A(t)\Phi(x)$, where $A(t) \in \mathcal{M}_{\mathbf{r}}$.

Question: How to obtain $A(t)$?

DLRA applied to the HJB equation

Formally, we can write the HJB equation as

$$\partial_t V(t,x) = [\mathcal{L}V](t,x), \tag{9}$$

where $\mathcal{L}$ is a nonlinear differential operator.

Given: basis Φ, TT-rank $\mathbf{r}$

Goal: find approximate solution $\hat{V}(t,x) = A(t)\Phi(x)$, where $A(t) \in \mathcal{M}_{\mathbf{r}}$.

Question: How to obtain $A(t)$?

Idea of DLRA: Project r.h.s. of (9) onto the current tangent space

$$\dot{A}(t) = \underset{B \in \mathcal{T}_{A(t)}(\mathcal{M}_{\mathbf{r}})}{\arg\min} \|B\Phi - [\mathcal{L}\hat{V}](t,\cdot)\|_{L^2(\Omega)}. \tag{10}$$

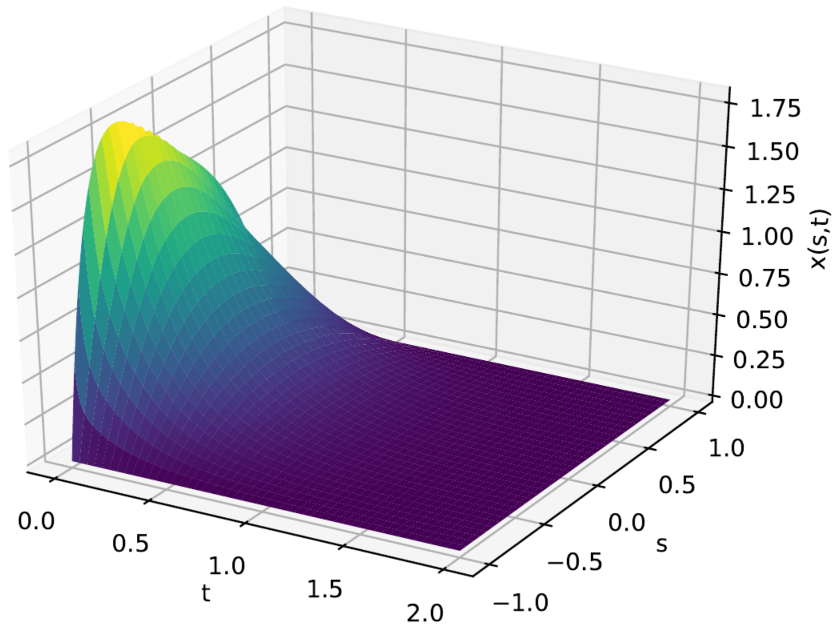
Results [ESS21]

Consider a semi-discretised 1D heat Eq. with unstable reaction term

$$\dot{x} = Ax + x^3 + gu, \tag{11}$$
$$x(0) = x_0, \tag{12}$$

with

- $x(t) \in \Omega = (-2,2)^d, d = 12$
- Scalar control $u \in \mathbb{R}$ and $g \in \mathbb{R}^d$
- Basis $\{\phi_i\}_i^n$ of $H^2_{\text{mix}}(\Omega)$-orthonormal polynomials up to degree $n = 8$
- TT-rank $\mathbf{r} = (3, 5, 5, \dots, 5, 5, 3)$



Dimension reduction from the full tensor space

$$n^d = 9^{12} > 282 \text{ billion} \qquad \text{to less than} \qquad nd \max_{i=1,\dots,d} r_i^2 = 2700$$

Results [ESS21]

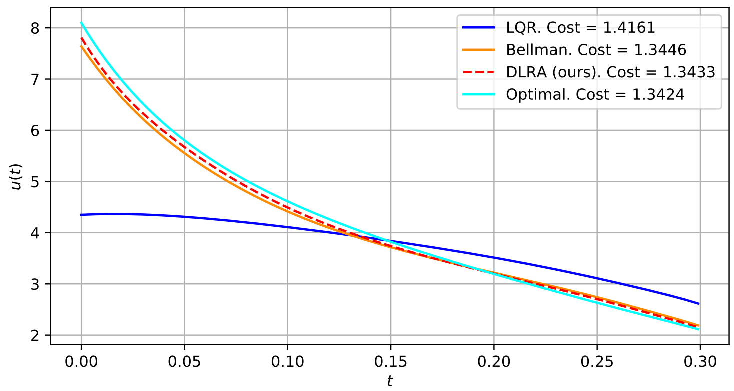


Figure 2: Control trajectories for a polynomial initial condition

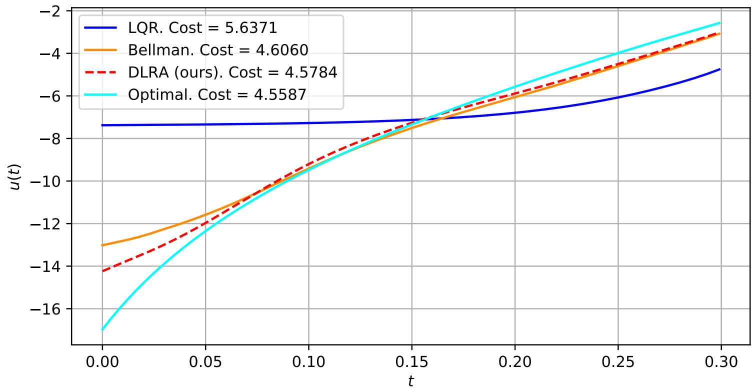


Figure 3: Control trajectories for a constant initial condition

	Bellman		DLRA		Hybrid	
	comp. time	mean cost	comp. time	mean cost	comp. time	mean cost
pol. deg. 4	3078.44	1.8822	333.29	2.6147	909.65	1.8804
pol. deg. 6	4270.33	1.8801	421.52	1.8802	1851.93	1.8798
pol. deg. 8	5967.91	1.8800	499.96	1.8799	–	–

Table 1: Computation time of the methods in seconds as well as mean costs of polynomial initial conditions for the heat Eq. with unstable reaction term.
The mean cost of the optimal control is 1.8793.

Outlook: DLRA Molecular Dynamics

Task: Determine development of observable means

$$g(t,x) = \mathbb{E}[f(X_t)|X_0 = x]$$

of high dimensional molecular dynamics such as overdamped Langevin dynamics

$$dX_t = -\frac{1}{\gamma}\nabla V(X_t)dt + \sqrt{2\beta^{-1}\gamma^{-1}}dW_t, \; X_t \in \mathbb{R}^d$$

The expectation g satisfies the Kolmogorov Backward Equation (KBE)

$$\partial_t g = \mathcal{L}g, \tag{13}$$
$$g(0,x) = f(x) \tag{14}$$

with

$$[\mathcal{L}g](t,x) = -\frac{1}{\gamma}\nabla V(x)\cdot\nabla g(t,x) + \frac{1}{\beta\gamma}\Delta g(t,x).$$

Outlook: DLRA in Molecular Dynamics

Problem: $d = 3N \gg 1$ (spatial coordinates of every atom)

Tools:

- Physics informed coordinate transform (*coarse graining*) $\xi : \mathbb{R}^d \to \mathbb{R}^m$, $m \leq d$, and projected dynamics driven by $\mathcal{L}^\xi$.
- Tensor Train Ansatz with 1D basis functions ϕ_i on reduced variables:

$$g(t, x) \approx A(t)(\phi \circ \xi)(x) = \sum_{i_1,\dots,i_d} A_{i_1,\dots,i_d}(t)\phi_{i_1}(\xi_1(x)) \cdot \ldots \cdot \phi_{i_d}(\xi_d(x)).,$$

- DLRA with empirical norm via samples from the invariant measure μ.

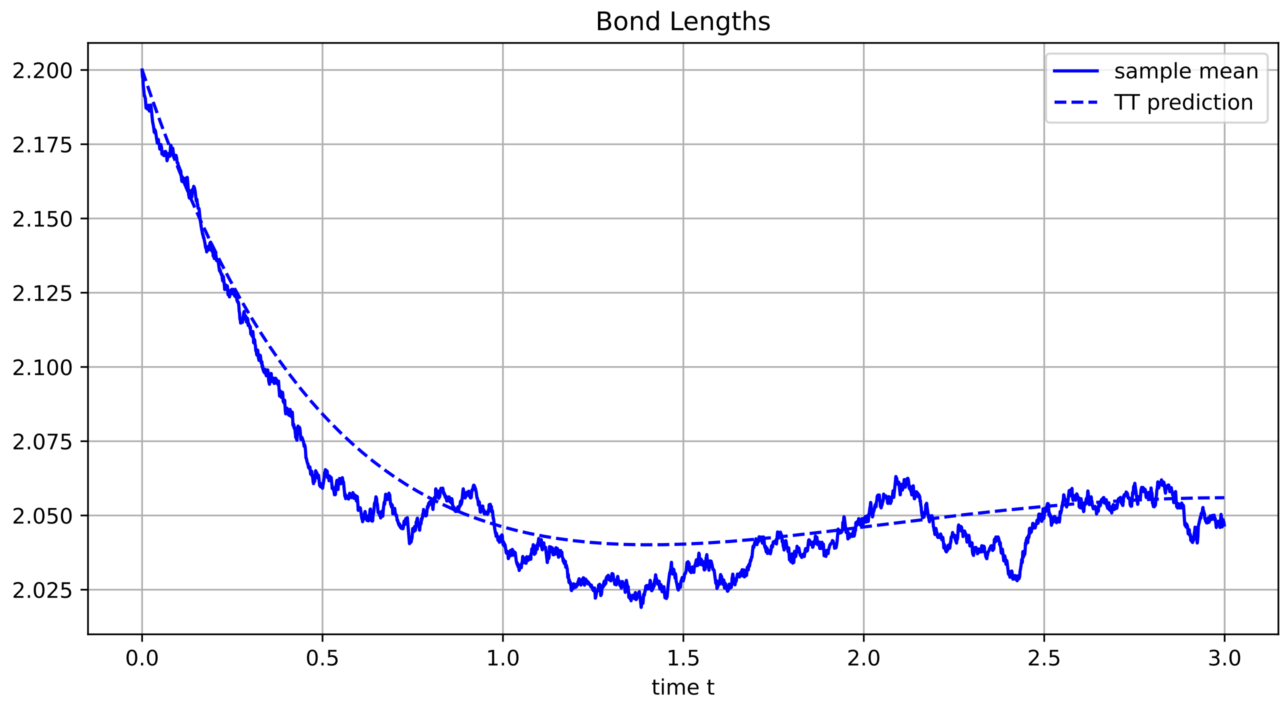


Figure 4: TT prediction (dashed) and empirical mean of a bond length between two atoms of a toy molecule developing over time.

Thank you very much for your attention!

[BPK16] Steven L. Brunton, Joshua L. Proctor, and J. Nathan Kutz, *Discovering governing equations from data by sparse identification of nonlinear dynamical systems*, Proceedings of the National Academy of Sciences **113** (2016), no. 15, 3932–3937.

[CKL22] Gianluca Ceruti, Jonas Kusch, and Christian Lubich, *A rank-adaptive robust integrator for dynamical low-rank approximation*, BIT Numerical Mathematics (2022), 1–26.

[ESS21] Martin Eigel, Reinhold Schneider, and David Sommer, *Dynamical low-rank approximations of solutions to the hamilton-jacobi-bellman equation*, 2021.

[FPM+20] Carl Folkestad, Daniel Pastor, Igor Mezic, Ryan Mohr, Maria Fonoberova, and Joel Burdick, *Extended dynamic mode decomposition with learned koopman eigenfunctions for prediction and control*, 2020.

[KK18] Dante Kalise and Karl Kunisch, *Polynomial approximation of high-dimensional hamilton–jacobi–bellman equations and applications to feedback control of semilinear parabolic pdes*, SIAM Journal on Scientific Computing **40** (2018), no. 2, A629–A652.

[KKB21] Eurika Kaiser, J. Nathan Kutz, and Steven L. Brunton, *Data-driven discovery of koopman eigenfunctions for control*, 2021.

[KL10] Othmar Koch and Christian Lubich, *Dynamical tensor approximation*, SIAM Journal on Matrix Analysis and Applications **31** (2010), no. 5, 2360–2375.

[LRSV13] Christian Lubich, Thorsten Rohwedder, Reinhold Schneider, and Bart Vandereycken, *Dynamical approximation by hierarchical tucker and tensor-train tensors*, SIAM Journal on Matrix Analysis and Applications **34** (2013), 470–494.

[Ose11] Ivan Oseledets, *Tensor-train decomposition*, SIAM J. Scientific Computing **33** (2011), 2295–2317.

[OSS19] Mathias Oster, Leon Sallandt, and Reinhold Schneider, *Approximating the stationary hamilton-jacobi-bellman equation by hierarchical tensor products*, 2019.

Dynamics of the system

General dynamics

$$\dot{x} = f(x, u)$$

can be transformed by defining $\hat{x} = (x, u)$ and $\hat{u} = \dot{u}$ to:

$$\dot{\hat{x}} = \begin{pmatrix} \dot{x} \\ \dot{u} \end{pmatrix} = \begin{pmatrix} f(x, u) \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ I \end{pmatrix} \hat{u} = g(\hat{x}) + B\hat{u}.$$

Linear in the control!

Nonlinear function g can be learned with

- Neural networks, Gaussian processes, tensor networks
- Koopman based methods like EDMD (e.g. [FPM$^+$20, KKB21]
- Sparse methods, i.e. SINDy ([BPK16])

or combinations of the above.

Solving the FFH problem with Policy Iteration

Problem (FFH)

Find $\bar{u}(t,x)$ *as a solution to*

$$\min_{\bar{u}\in\mathcal{A}} \int_0^T \ell(\bar{x},\bar{u})dt + E(\bar{x}(T)),$$
$$\dot{\bar{x}} = g(\bar{x}) + B\bar{u}, \quad \bar{x}(0) = x_0$$

for all $x_0 \in \mathcal{X}_0$.

Assumption:

$$F(x,u) = l(x) + 2\sum_{i=1}^{m} \lambda_i \int_0^{u_i} \mathcal{P}_i^{-1}(\mu)d\mu,$$

- $\mathcal{P}_i \in \mathcal{C}^1(\mathbb{R}, (u_i^{min}, u_i^{max}))$ odd, integrable, strictly increasing and bijective, $\lambda_i > 0$
- l continuous, bounded below by a class $\mathcal{K}_\infty$ function, $l(0) = 0$.

Iterate until convergence:

(1) Approximate $V_{\bar{u}}(t,x) = \int_t^T \ell(\bar{x},\bar{u})dt + E(\bar{x}(T))$ in suitable function class

(2) update $\bar{u}(t,x) \leftarrow -\mathcal{P}\left(\frac{1}{2}\Lambda^{-1}g(x)^T\nabla_x V_{\bar{u}}(t,x)\right)$

To approximate

$$V(t,x) = \int_t^T \ell(\bar{x},\bar{u})dt + E(\bar{x}(T))$$

(a) Propagate samples $\{x_k\}_{k=1}^M$ through the dynamics and add up costs (MC) to get snapshots

$$\{y_{jk}\}_{j,k} = \{V(t_j, x_k)\}_{j,k}$$

at discrete time points t_j.

(b) Fit $V^j(x) \approx V(t_j, x)$ by solving

$$\min_{V^j \in \mathcal{M}} \sum_{k=1}^M |V^j(x_k) - y_k^j|^2 + \mu \|V^j\|_{H^1}^2$$

(c) interpolate (e.g.) linearly between $V^j(x)$ to obtain $\tilde{V}(t,x) \approx V(t,x)$.

Often suited for function class $\mathcal{M}$ in (b): **polynomials** (see, eg. [KK18])

Error bounds in the abstract setting

In the abstract setting, error bounds can be derived, which we quote for the sake of completeness.

Theorem ([LRSV13])

Suppose that $\dot{A}(t) \leq \mu$ *and that a continuously differentiable best approximation* $X(t) \in \mathcal{M}_{\mathbf{r}}$ *to* $A(t)$ *exists for* $t \in [0, T]$. *Let* $\delta > 0$ *be such that the smallest nonzero singular value of every matrix unfolding of* $X(t)$ *is greater or equal to* ρ, *and assume that the best-approximation error is bounded by* $\|X(t) - A(t)\| \leq c\rho$ *for* $t \in [0, T]$ *with a constant* c *depending only on the dimension* d. *Then, the approximation error of the dynamical low-rank approximation defined by* (7) *with initial value* $Y(0) = X(0)$ *is bounded by*

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

with $\beta = C\mu\rho - 1$ *for* $t \in [0, T]$, *as long as the right-hand side remains bounded by* $c\rho$. *The constant* C *is only dependent on* d *and is given in [LRSV13].*

In practice we set

$$\Delta A = \operatorname*{arg\,min}_{B \in \mathcal{T}_{A(t)}(\mathcal{M}_{\mathsf{r}})} \|B\Phi - [\mathcal{L}\hat{V}](t,\cdot)\|_{L^2(\Omega,{\color{red}M})}$$

$$= \min_{B \in \mathcal{T}_{A(t)}(\mathcal{M}_{\mathsf{r}})} \sum_{k=1}^{M} |B\Phi(x_k) - [\mathcal{L}\hat{V}](t,x_k)|^2, \quad x_k \text{ uniform},$$

In practice we set

$$\Delta A = \operatorname*{arg\,min}_{B\in\mathcal{T}_{A(t)}(\mathcal{M}_{\mathbf{r}})} \|B\Phi - [\mathcal{L}\hat{V}](t,\cdot)\|_{L^2(\Omega,{\color{red}M})}$$
$$= \min_{B\in\mathcal{T}_{A(t)}(\mathcal{M}_{\mathbf{r}})} \sum_{k=1}^{M} |B\Phi(x_k) - [\mathcal{L}\hat{V}](t,x_k)|^2, \quad x_k \text{ uniform},$$

and make an Euler steps

$$\hat{A}_{i+1} = A_i + \tau\Delta A_i \qquad \text{(addition leaves the manifold)},$$
$$A_{i+1} = \mathcal{R}(\hat{A}_{i+1}) \qquad \text{(retraction back to the manifold)}.$$

In the case of the TT-manifold, the retraction $\mathcal{R}$ can be done by simple rank truncation [Ose11].

Example: Heating of a building

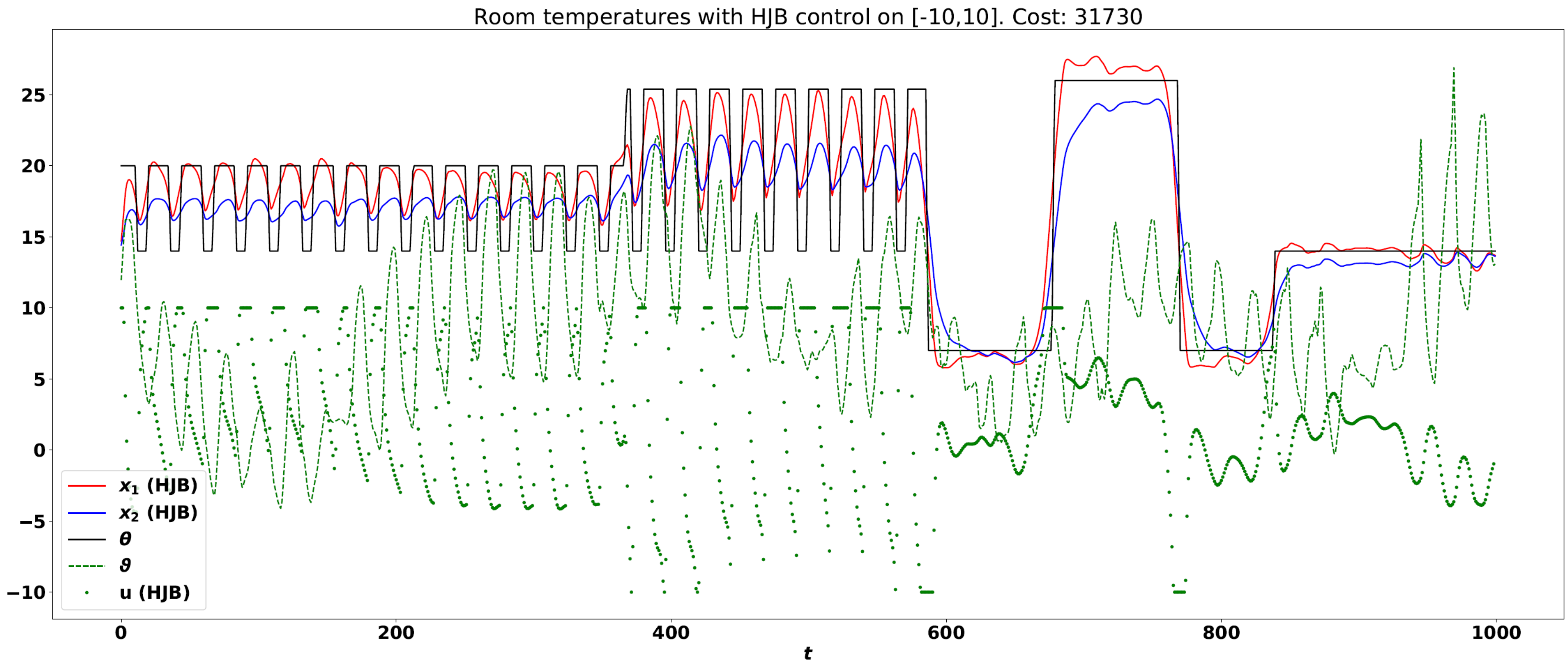


Figure 5: Controlling the temperature of two rooms (simulation).
red, blue: room temperatures
black: target temperature
green - - -: outer temperature
green · · ·: control values