FRIEDRICH-ALEXANDER UNIVERSITÄT ERLANGEN-NÜRNBERG
FACULTY OF SCIENCES

CHAIR IN APPLIED ANALYSIS FAU

STOCHASTIC SIMULATION AND OPTIMIZATION FOR DYNAMICAL SYSTEMS

Daniël Veldman Applied Analysis - AvH Professorship, FAU Erlangen-Nürnberg

Enrique Zuazua Applied Analysis - AvH Professorship, FAU Erlangen-Nürnberg

Introduction

Stochastic optimization techniques such as the Random Batch Method (RBM) are well-established and widely used in modern intelligent systems such as search engines, recommendation software, and speech and image recognition platforms [1].
The RBM has also been applied to the simulation and optimization of interacting particle systems where it leads to a significant reduction in computational cost [2,3]. Inspired by these results, we consider here a stochastic method to speed up the simulation and optimization of large-scale linear dynamical systems.

Optimal control

In a classical control problem, the aim is to find the optimal control $u^*(t)$ that minimizes

$$J = \int_0^T \Big((x(t) - x_d(t))^\top Q (x(t) - x_d(t)) + u(t)^\top R u(t) \Big) \ \mathrm{d}t, \tag{1}$$

on a finite time interval $[0, T]$ subject to the dynamics

$$\dot{x}(t) = Ax(t) + Bu(t), \qquad x(0) = x_0, \tag{2}$$

where $x(t) \in \mathbb{R}^N$ is the state, $u(t) \in \mathbb{R}^q$ is the control, $x_d(t)$ is a given desired trajectory, and $Q \succeq 0$, $R \succ 0$, A, and B are constant matrices. Without loss of generality, we can assume that $q \leq N$.
It is well known that the problem (1)–(2) has a unique solution $u^* \in L^2(0, T; \mathbb{R}^q)$.
However, computing u^* can be challenging because (2) and the corresponding adjoint equation need to be solved several times and each time step typically has a computational complexity of $O(N^3)$.

Especially when N is large, finding $u^*(t)$ is computationally demanding.

Stochastic simulation and optimization method

Step 1 Decompose the matrix A into submatrices A_m as

$$A = \sum_{m=1}^{M} A_m. \tag{3}$$

The submatrices A_m are chosen such that replacing A by A_m in (2) reduces the computational cost per time step. Typically, the submatrices A_m will be more sparse than A.

Step 2 Let $\{\mathcal{S}_\ell\}_{1 \leq \ell \leq 2^M}$ denote the collection of 2^M subsets of $\{1, 2, \ldots, M\}$.
Assign to each subset a probability $p_\ell \in [0, 1]$ with which the subset $\mathcal{S}_\ell$ will be selected such that

- $\sum_{\ell=1}^{2^M} p_\ell = 1$,
- $\pi_m = \sum_{\{\ell \mid m \in \mathcal{S}_\ell\}} p_\ell > 0$ for each $m \in \{1, 2, \ldots, M\}$.

Note that π_m is the probability that the index m is an element of the selected subset.

Step 3 Partition the time interval $I = [0, T]$ into K subintervals $I_k = [t_{k-1}, t_k]$ of length $\leq h$. In each time interval I_k, choose a subset $\mathcal{S}_{\ell(k)}$ according to the probabilities p_ℓ from Step 2 and define

$$A_h(t) = \sum_{m \in \mathcal{S}_{\ell(k)}} \frac{A_m}{\pi_m}, \qquad t \in I_k. \tag{4}$$

This definition ensures that $\mathbb{E}[A_h(t)] = A$ for all $t \in [0, T]$.

Step 4 Find the optimal control $u_h^*(t)$ that minimizes

$$J_h = \int_0^T \Big((x_h(t) - x_d(t))^\top Q (x_h(t) - x_d(t)) + u_h(t)^\top R u_h(t) \Big) \ \mathrm{d}t, \tag{5}$$

subject to the dynamics

$$\dot{x}_h(t) = A_h(t) x_h(t) + B u_h(t), \qquad x(0) = x_0. \tag{6}$$

When the $N \times N$-matrix A is decomposed into M blocks of size $N/\sqrt{M} \times N/\sqrt{M}$, the computational complexity for each time step of (6) and the corresponding adjoint equation is reduced to $O(N^3/M^{3/2})$.

Typically, $u_h^*(t)$ can be computed faster than $u^*(t)$.

Convergence results

Fix any control $u \in L^2(0, T; \mathbb{R}^q)$ in (2). When the control $u_h(t)$ in (6) is equal to the control $u(t)$ in (2), we can prove similarly as in [3] that

$$\lim_{h \to 0} \mathbb{E}\left[|x_h(t) - x(t)|^2 \right] = 0, \tag{7}$$

for all $t \in [0, T]$. Using this result and the strong convexity of the cost functional J_h, we proved that

$$\lim_{h \to 0} \mathbb{E}\left[\|u_h^* - u^*\|_{L^2(0,T)}^2 \right] = 0. \tag{8}$$

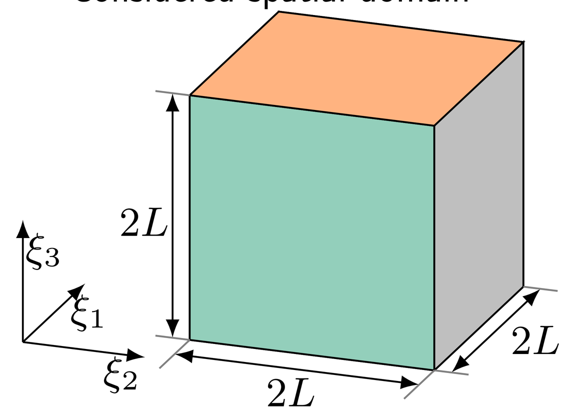
These results imply that for any $\varepsilon > 0$ and $\delta > 0$, there exists an $h > 0$ such that $\mathbb{P}[\max_{t \in [0,T]} |x_h(t) - x(t)|^2 > \delta] < \varepsilon$ and $\mathbb{P}[\|u_h^* - u^*\|_{L^2(0,T)}^2 > \delta] < \varepsilon$.

Numerical example

The temperature in the space $\boldsymbol{\xi} = (\xi_1, \xi_2, \xi_3) \in [-L, L]^3$ is modeled by the heat equation $y_t = \Delta y$. Our aim is to keep the temperature on $S_{\text{top}} = \{\xi_3 = L\}$ (the orange surface in the figure) close to zero by applying a uniform heat load $-\partial y(t, \boldsymbol{\xi})/\partial \xi_1 = u(t)$ on $S_1 = \{\xi_1 = -L\}$ (the green surface in the figure). Zero Neumann boundary conditions are applied except on S_1. For simplicity, the control $u(t)$ is taken independent of space. The initial condition is $y(0, \boldsymbol{\xi}) = \exp(-|\boldsymbol{\xi}|^2/(8L^2))$ and the cost functional is taken as

$$\mathcal{J} = \int_0^T \iint_{S_{\text{top}}} y(t, \boldsymbol{\xi})^2 \ \mathrm{d}\xi_1 \ \mathrm{d}\xi_2 \ \mathrm{d}t + 10^{-4} \int_0^T u(t)^2 \ \mathrm{d}t. \tag{9}$$

Considered spatial domain



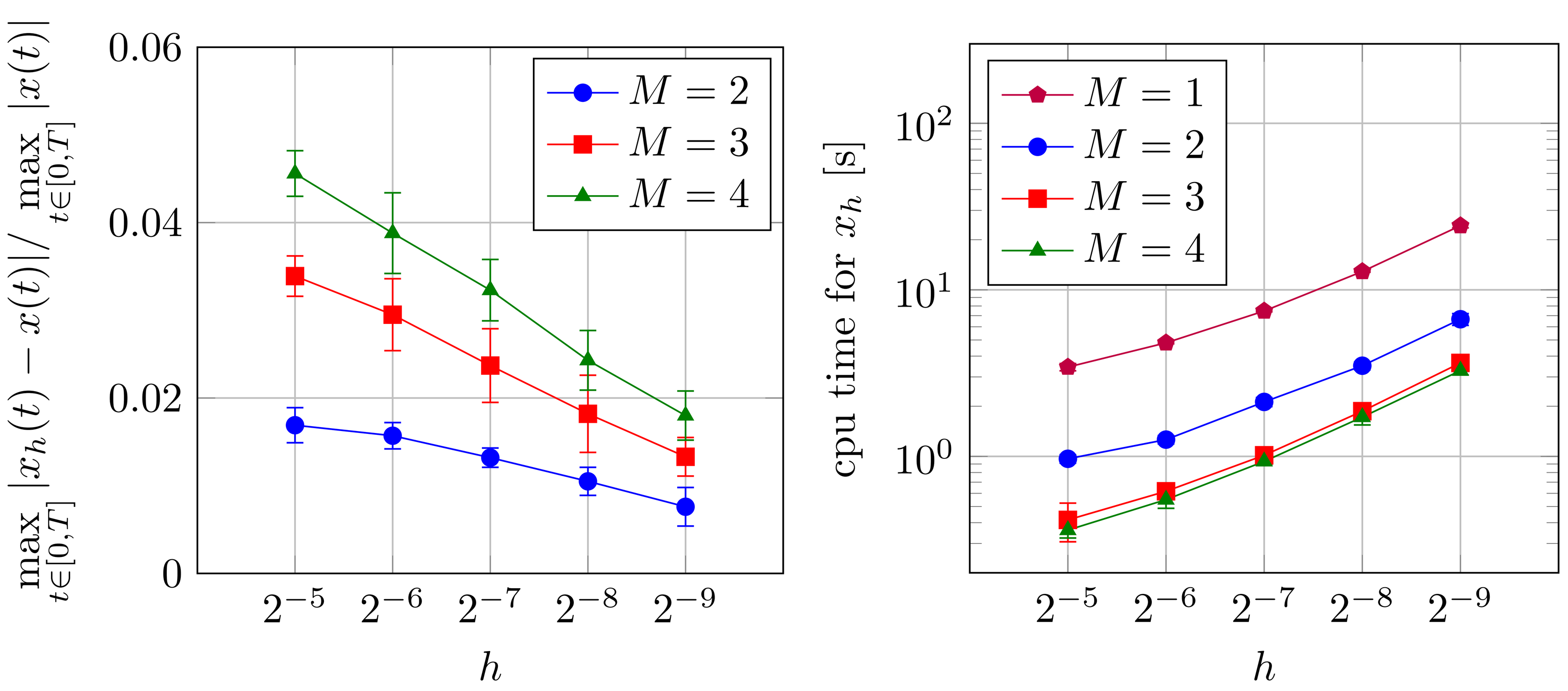
The time horizon is $T = 2$ and $L = 1.5$. A finite difference discretization on a uniform rectangular grid with $N = 31^3 = 29,791$ nodes results in a system of the form (1)–(2). The time grid is uniform with stepsize h. Observe that the matrix A can be written as the sum of interaction matrices A_{ij} of the form

$$A_{ij}[i, i] = A_{ij}[j, j] = -\tfrac{1}{\Delta\xi^2}, \qquad A_{ij}[i, j] = A_{ij}[j, i] = \tfrac{1}{\Delta\xi^2}, \tag{10}$$

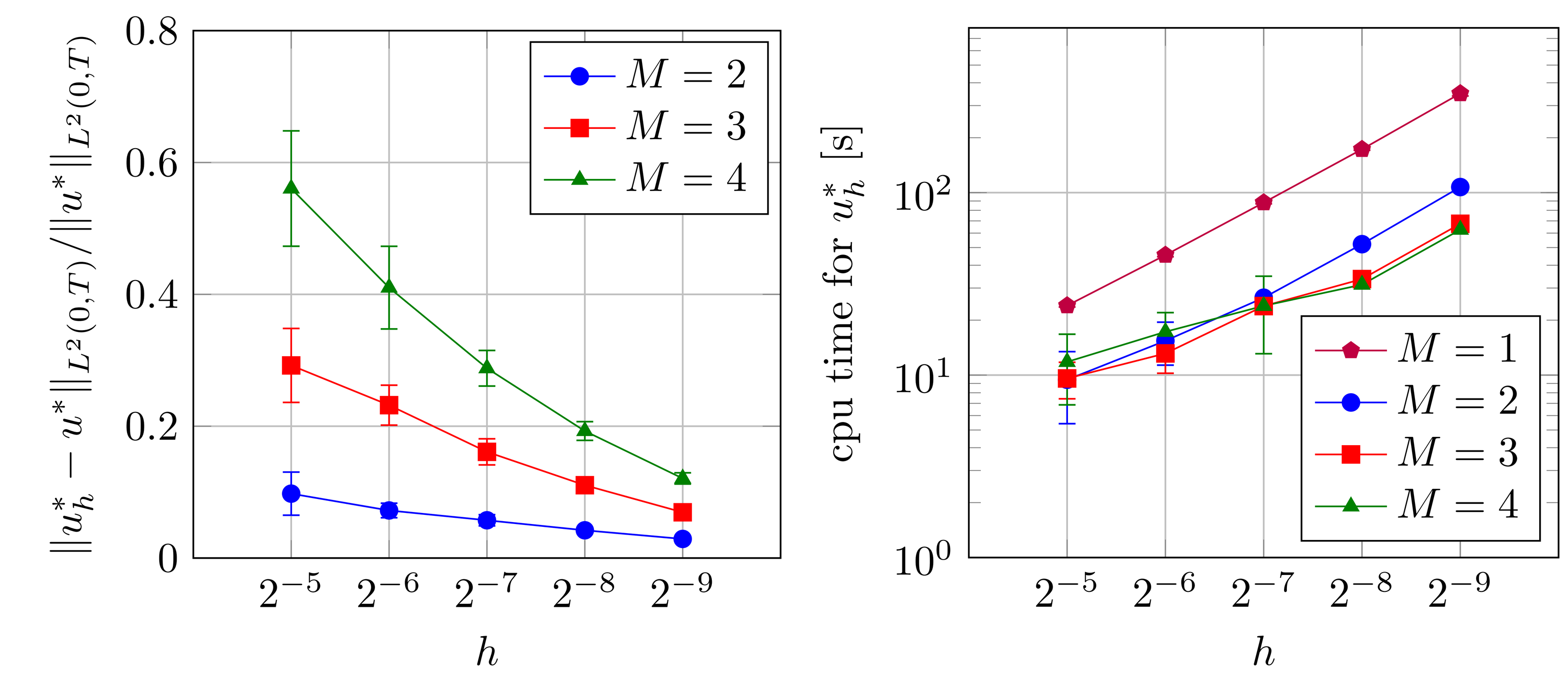
where i and j are adjacent nodes in the spatial grid with spacing $\Delta\xi$. The set of $86,490$ interaction matrices A_{ij} is randomly partitioned into M subsets of approximately equal size. Each submatrix A_m is the sum of the interaction matrices A_{ij} in one subset. Note that this construction assures that (3) holds. The probabilities p_ℓ are chosen as $p_\ell = 1/M$ when $\mathcal{S}_\ell = \{m\}$ for some $m \in \{1, 2, \ldots, M\}$ and $p_\ell = 0$ otherwise. It follows that $\pi_m = 1/M$. Note that $A_h(t) = A$ when $M = 1$.

We present numerical results for two situations. Situation I illustrates the convergence result (7) for the solution $x_h(t)$ of the forward dynamics (6) with $u_h(t) = 0$. Situation II illustrates the convergence result (8) for the control $u_h^*(t)$ that minimizes the cost J_h in (5). In both cases, $x_h(t)$ and $u_h^*(t)$ are computed for 10 different realizations of the sets $\mathcal{S}_{\ell(k)}$ (but for the same decomposition of A into submatrices A_m). The figures below show the mean and the (estimated) 2σ-confidence interval of the error and the computational time (based on these 10 realizations). Observe that the solutions $x_h(t)$ and $u_h^*(t)$ for $M = 1$ are equal to $x(t)$ and $u^*(t)$.

Situation I: Relative error and computational time for the response $x_h(t)$ with $u_h(t) = 0$



Situation II: Relative error and computational time for the optimal control $u_h^*(t)$



The figures show that the errors in $x_h(t)$ and $u_h^*(t)$ decrease with h.
Increasing M increases the error but speeds up the computations.

Conclusions, discussions, and further research

The accuracy of the proposed stochastic simulation and optimization method for large-scale linear dynamical systems increases when the time step h is decreased. Our analysis shows that $\mathbb{E}[\max_{t\in[0,T]} |x_h(t) - x(t)|]$ and $\mathbb{E}[\|u_h^* - u^*\|_{L^2(0,T)}]$ converge to zero as $\sqrt{h}$. This rate can also be observed in the numerical example.

In the considered numerical example, the reduction in computational time varies between a factor 2 and 10 and generally increases when h decreases. With $M = 2$ and $h = 2^{-9}$, optimal controls with an expected 4%-error are computed 3 times faster. We expect an even larger reduction in computational time when the state dimension N is increased further.

Note that x_h depends nonlinearly on A_h so that $\mathbb{E}[x_h] \neq x$ and $\mathbb{E}[u_h^*] \neq u^*$ for $h > 0$. Because of this bias, the variation in the x_h and u_h^* obtained for different realizations of the sets $\mathcal{S}_{\ell(k)}$ on the same time grid cannot be used to estimate the expected errors $\mathbb{E}[|x_h(t) - x(t)|^2]$ and $\mathbb{E}[\|u_h^* - u^*\|_{L^2(0,T)}^2]$.

Topics for further research:

- What is the best choice for the probabilities p_ℓ?
- What is the best way to decompose A into submatrices A_m?
- Extensions to infinite-dimensional and nonlinear problems.

Selected publications

[1] Bottou, L., Curtis, F. E., Nocedal, J. (2018). **Optimization methods for large-scale machine learning.** Siam Review, 60(2), 223-311.

[2] Ko, D., Zuazua, E. (2020). **Model predictive control with random batch methods for a guiding problem.** arXiv:2004.14834.

[3] Jin, S., Li, L., Liu, J. G. (2020). **Random Batch Methods (RBM) for interacting particle systems.** J. Comput. Phys., 400, 108877.

[4] Veldman, D.W.M., Zuazua, E. **Stochastic time-splitting methods in optimal control.** In preparation.

caa-avh.nat.fau.eu
04/2021

Department of DATA SCIENCE

Unterstützt von / Supported by
Alexander von Humboldt Stiftung/Foundation