Algorithm for Overcoming the Curse of Dimensionality For Time-Dependent Non-convex Hamilton–Jacobi Equations Arising From Optimal Control and Differential Games Problems

Yat Tin Chow, Jerome Darbon, Stanley Osher, Wotao Yin

Presentation by Bumsu Kim and Wonjun Lee

University of California - Los Angeles

March 11, 2019

Contents

Introduction

Differential games with time dependent control sets

Given:

- A Fixed terminal time $T \in \mathbb{R}$.
- An initial time $t < T$.
- Two balanced compact convex sets $C, D \subset \mathbb{R}^d$ (a set C is balanced if $C = -C$)

Consider two families of balanced compact convex sets

$$\{C(s)\}_{t<s<T}, \{D(s)\}_{t<s<T}$$
$$C(s) \subset C, D(s) \subset D \text{ for all } s$$

Introduction

Denote

$$\mathcal{A} := \{a : [t, T] \to C : a \text{ is measurable, } a(s) \in C(s)\}$$
$$\mathcal{B} := \{b : [t, T] \to C : a \text{ is measurable, } b(s) \in D(s)\}$$

a: control, b: possible errors in the system

Initial Value Problem

Given an initial state $x \in \mathbb{R}^d$, consider (Lipschitz) solution $x : [t, T] \to \mathbb{R}^d$ such that

$$\frac{dx}{ds}(s) = Mx(s) + N_C(s)a(s) + N_D(s)b(s) \text{ in } (t, T) \tag{1}$$

$$x(t) = x \tag{2}$$

$M, \{N_C(s)\}_{t<s<T}, \{N_D(s)\}_{t<s<T}$: $d \times d$ matrices with real entries.

Initial Value Problem

Using change of variable $z(s) = e^{-sM}x(s)$,

$$\frac{dz}{ds}(s) = e^{-sM}N_C(s)a(s) + e^{-sM}N_D(s)b(s) \text{ in } (t, T) \tag{3}$$

$$z(t) = e^{-tM}x \tag{4}$$

$M, \{N_C(s)\}_{t<s<T}, \{N_D(s)\}_{t<s<T}$: $d \times d$ matrices with real entries.

Cost Functional

Consider a running-cost Lagrangian functional

- $L : [t, T] \times C \times D \to \mathbb{R} \cup \{+\infty\}$
- L is proper lower-semicontinuous 1-coercive convex in a.
- L is proper upper-semicontinuous 1-coercive concave in b.

and $J : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$: proper lower-semicontinuous 1-coercive convex terminal cost functional.

Cost Functional

Consider the cost functional for a given intial time $t < T, z \in \mathbb{R}^d, a, b,$

$$K(t, z, a(\cdot), b(\cdot)) = \int_t^T L(s, a(s), b(s))ds + J(z(T))$$

Cost Functional

Consider the cost functional for a given intial time $t < T, z \in \mathbb{R}^d, a, b,$

$$K(t, z, a(\cdot), b(\cdot)) = \int_t^T L(s, a(s), b(s))ds + J(z(T))$$

Goal

a: minimize K

b: maximize K

Change of Variables

Let

$$\tilde{z}(\tau) = z(T - \tau) = z(s)$$

Then

$$\frac{d\tilde{z}(\tau)}{d\tau} = \frac{dz(s)}{ds}\frac{ds}{d\tau} = -\frac{dz(s)}{ds}$$

Change of Variables

$$\begin{aligned}\frac{dz}{ds}(s) &= e^{-sM}N_C(s)a(s) + e^{-sM}N_D(s)b(s) \text{ in } (t,T)\\ &= f(s,a,b)\\ z(t) &= e^{-tM}x\end{aligned}$$

$$\begin{aligned}\frac{d\tilde{z}}{d\tau}(\tau) &= -f(T-\tau,a,b)\\ \tilde{z}(T-t) &= e^{-(T-t)M}x\end{aligned}$$

Change of Variables

$$\begin{aligned}
\tilde{K}(T-t,\tilde{z};a,b) &= \int_{T-t}^{0} \tilde{L}(\tau,a,b)d\tau + J(\tilde{z}(0)) \\
&= \int_{T-t}^{0} L(T-\tau,a,b)d\tau + J(\tilde{z}(0)) \\
&= -\int_{t}^{T} L(s,a,b)ds + J(z(T)) \\
&= -K(t,z;a,b) + 2J(z(T))
\end{aligned}$$

Change of Variables

$$\begin{aligned}
\tilde{K}(T-t,\tilde{z};a,b) &= \int_{T-t}^{0} \tilde{L}(\tau,a,b)d\tau + J(\tilde{z}(0)) \\
&= \int_{T-t}^{0} L(T-\tau,a,b)d\tau + J(\tilde{z}(0)) \\
&= -\int_{t}^{T} L(s,a,b)ds + J(z(T)) \\
&= -K(t,z;a,b) + 2J(z(T))
\end{aligned}$$

Goal

a: minimize K $\iff$ maximize $\tilde{K}$

b: maximize K $\iff$ minimize $\tilde{K}$

Lower Value and Upper Value

Define the lower value as

$$V(T-t,\tilde{z}) := \inf_{\beta\in\Delta(T-t)} \sup_{a\in\mathcal{A}(T-t)} \tilde{K}(T-t,\tilde{z};a(\cdot),\beta[a](\cdot)) \tag{5}$$

and upper value as

$$U(T-t,\tilde{z}) := \sup_{\alpha\in\Gamma(T-t)} \inf_{b\in\mathcal{B}(T-t)} \tilde{K}(T-t,\tilde{z};\alpha[b](\cdot),b(\cdot)) \tag{6}$$

$\Gamma(t)$: the set of all strategies for the **control** beginning at t.
$\Delta(t)$: the set of all strategies for the **error** beginning at t.

Hamilton-Jacobi PDE

$V(T-t,\tilde{z})$ and $U(T-t,\tilde{z})$ are the viscosity solution to the following HJ PDE:

$$\frac{\partial}{\partial t}V(T-t,\tilde{z})+H^{-}(t,\nabla_z V(T-t,\tilde{z}))=0 \qquad \text{in } \mathbb{R}^d\times(0,T) \tag{7}$$

$$V(T,\tilde{z})=J(\tilde{z}) \qquad \text{in } \mathbb{R}^d \tag{8}$$

$$\frac{\partial}{\partial t}U(T-t,\tilde{z})+H^{+}(t,\nabla_z(T-t,\tilde{z}))=0 \qquad \text{in } \mathbb{R}^d\times(0,T) \tag{9}$$

$$U(T,\tilde{z})=J(\tilde{z}) \qquad \text{in } \mathbb{R}^d \tag{10}$$

Hamilton-Jacobi PDE

$$H^-(t,p) = \min_{b\in D(T-t)} \max_{a\in C(T-t)} \{\langle -f(T-t,a,b), p\rangle + L(T-t,a,b)\}$$
$$H^+(t,p) = \max_{a\in C(T-t)} \min_{b\in D(T-t)} \{\langle -f(T-t,a,b), p\rangle + L(T-t,a,b)\}$$

To see the proof: Evans, L.C. Souganidis, P.E.: Differential games and representation formulas for solutions of Hamilton-Jacobi-Isaacs equations. Indiana Univ. Math. J. **38,** 773-797 (1984)

Problem Setup

The class of problems the paper is focusing on:

- $\mathrm{L} = 0$
- $C(t) = \{\langle x - a_C(t), [Q_C(t)]^{-1}(x - a_C(t))\rangle \leq 1\}$
- $D(t) = \{\langle x - a_D(t), [Q_D(t)]^{-1}(x - a_D(t))\rangle \leq 1\}$
- Q_C, Q_D: some positive definite matrices depending on t.

Problem Setup

$$
\begin{aligned}
&H^-(t,p)\\
&= \min_{b\in D(T-t)} \max_{a\in C(T-t)} \{\langle -f(T-t,a,b),p\rangle\}\\
&= \min_{b\in D(T-t)} \max_{a\in C(T-t)} \{\langle -e^{-(T-t)M}N_C(T-t)a - e^{-(T-t)M}N_D(T-t)b,p\rangle\}\\
&= \max_{a\in C(T-t)} \min_{b\in D(T-t)} \{\langle -e^{-(T-t)M}N_C(T-t)a - e^{-(T-t)M}N_D(T-t)b,p\rangle\}\\
&= H^+(t,p)
\end{aligned}
$$

Let $H(t,p) := H^-(t,p) = H^+(t,p)$

HJE PDE Solution

Thus, we can check that $\phi(t,z) := V(T-t,z) = U(T-t,z)$ is the viscosity solution to the following:

$$\frac{\partial}{\partial t}\phi(t,\tilde{z}) + H(t,\nabla_z \phi(t,\tilde{z})) = 0 \qquad \text{in } \mathbb{R}^d \times (0,T) \tag{11}$$

$$\phi(0,\tilde{z}) = J(\tilde{z}) \qquad \text{in } \mathbb{R}^d \tag{12}$$

HJE PDE Solution

Thus, we can check that $\phi(t,z) := V(T-t,z) = U(T-t,z)$ is the viscosity solution to the following:

$$\frac{\partial}{\partial t}\phi(t,\tilde{z}) + H(t, \nabla_z \phi(t,\tilde{z})) = 0 \qquad \text{in } \mathbb{R}^d \times (0,T) \tag{11}$$

$$\phi(0,\tilde{z}) = J(\tilde{z}) \qquad \text{in } \mathbb{R}^d \tag{12}$$

The viscosity solution to the above PDE given by the generalized Hopf formula is

$$\phi(t,z) = -\min_{p\in\mathbb{R}^d}\left\{ J^*(p) + \int_0^t H(s,p)ds - \langle z,p\rangle \right\}$$

Optimization Method

Generalized Hopf Formula

The solution of the Hamilton-Jacobi equation is given by

$$\phi(t,z) = -\min_{p\in\mathbb{R}^d}\left\{J^*(p) + \int_0^t H(s,p)ds - \langle z,p\rangle\right\}$$

We calculate the solution of the PDE via solving minimization problem. This is nice because

- All the values ϕ at different points (t,z) are *decoupled* and thus they can be implemented in *parallel*!
- Unlike FDM/FEM, where the error $\to 0$ as the mesh size $\to 0$ (and the number of mesh points/elements grows exponentially with the dimension), it gives an *exact solution*!

Optimization Method

Generalized Hopf Formula

The solution of the Hamilton-Jacobi equation is given by

$$\phi(t,z) = -\min_{p\in\mathbb{R}^d}\left\{J^*(p) + \int_0^t H(s,p)ds - \langle z,p\rangle\right\}$$

Our objective function has a special form $F_1(p) + F_2(p) - F_3(p)$, where F_1 is smooth and convex, F_2 and F_3 are convex homogeneous functions. Recall that the Hamiltonian H is homogeneous of degree one:

$$H(s,p) = \sqrt{\langle Ap, QAp\rangle} - \sqrt{\langle Bp, Q'Bp\rangle} + \langle Ap, a\rangle + \langle Bp, b\rangle$$

Optimization Method

Two methods to perform minimization:

- Coordinate descent
- ADMM/split Bregman

Coordinate Descent

First, take an initial guess of the Lipschitz constant L, and set $count = 0$. Initialize $j_1 := 1$ and a parameter $\alpha := 1/L$.

Algorithm 1 Coordinate Descent

1: **for** $k = 1, 2, \cdots, M$ **do**
2: For $i = j_k$, $p_i^{k+1} = p_i^k - \alpha(\partial_i J^*(p^k) + \int_0^t \partial_i H(s, p^k)ds - z_i)$
3: Set $j_{k+1} = j_k + 1$. If $j_{k+1} = d + 1$, set $j_k + 1 = 1$.
4: If $|p_i^{k+1} - p_i^k| > \epsilon$, set $count = 0$.
If $k = M$, reset $k = 0$ and set $\alpha = \alpha/2$.
5: If $|p_i^{k+1} - p_i^k| < \epsilon$, set $count = count + 1$.
6: If $count = d$, stop.
7: **end for**

Coordinate Descent

Advantages:

- Easy implementation
- May choose larger step size for gradient descent

Possible concern:

- May not converge in non-convex case

$\Rightarrow$ Use multiple initial guess to avoid non-optimal stationary points
(The only non-smooth point for our Hamiltonian is the origin)

ADMM/split-Bregman

Splitting schemes might compute faster in some cases, especially when the Hamiltonian is of special form: homogeneous of degree 1.

Moreover, in our case, ADMM is guaranteed to converge due to convexity of the functional.

ADMM/split-Bregman

Consider two homogeneous of degree 1 convex functionals:

$$\Psi_1(t,p) = \sqrt{\langle p, Q_C(T-t)p\rangle} + \langle p, a_C(T-t)\rangle$$
$$\Psi_1(t,p) = \sqrt{\langle p, Q_C(T-t)p\rangle} + \langle p, a_C(T-t)\rangle$$
$$R_1(t) = [-\exp(-(T-t)M)N_C(T-t)]^T$$
$$R_2(t) = [-\exp(-(T-t)M)N_D(T-t)]^T$$

Then $H(t,p) = \Psi_1(t, R_1(t)p) - \Psi_2(t, R_2(t)p)$, and our objective functinoal is

$$\mathcal{J}(p) = J^*(p) + \int_0^t \Psi_1(s, R_1(s)p) - \Psi_2(s, R_2(s)p)ds - \langle z, p\rangle$$

Minimizing $\mathcal{J}$ is equivalent to

ADMM/split-Bregman

Minimizing

$$J^*(p) + \int_0^t \Psi_1(s, \gamma_1(s)) - \Psi_2(s, \gamma_2(s))ds - \langle z, p \rangle$$
$$\text{subject to } R_1(s)p = \gamma_1(s), R_2(s)p = \gamma_2(s) \text{ for } 0 < s < t.$$

So we introduce the augmented Lagrangian functional:

$$\mathcal{L}(v, \gamma) = J^*(p) + \int_0^t \Psi_1(s, \gamma_1(s)) - \Psi_2(s, \gamma_2(s))ds - \langle z, p \rangle$$
$$+ \sum_{i=1,2} \frac{\rho_i}{2} \int_0^t \|\lambda_i(s) - \gamma_i(s) + R_i(s)v\|^2 ds$$

ADMM/split-Bregman

Algorithm 2 ADMM/split-Bregman

1: **for** $n = 1, 2, \cdots$ **do**

2: For each $s \in (0, t)$, compute

$$\gamma_i^{k+1}(s) = \text{argmin} \left\{ (-1)^{i+1} \Psi_i(s, I) + \frac{\rho_i}{2} \| \lambda_i^k(s) - I + R_i(s) v^k \|^2 \right\}$$

3: Compute $v^{k+1} = \text{argmin}\ \mathcal{L}(v, \gamma^{k+1})$

4: For each $s \in (0, t)$, set

$$\lambda_i^{k+1}(s) = \lambda_i^k(s) - \gamma_i^{k+1}(s) + R_i(s) v^{k+1}$$

5: If $|v^{k+1} - v^k| < \epsilon$, stop.

6: **end for**

ADMM/split-Bregman

Advantages:

- Only involves only explicit and easily computable proximal maps*
- May choose larger step size for gradient descent

Disadvantages:

- Computing variables for each $s \in (0, t)$ costs a lot for large T

It is difficult to say which algorithm performs better in a particular problem.

We present results from the coordinate descent algorithm here

Numerical Experiments - $d = 6$

Here we present an example with dimension 6 on a 2-dimensional cross section Ω of the form $[-3, 3]^2 \times \{0\}^4$. We computed $\phi(x)$ for each $x = (-3 + 0.1p, -3 + 0.1q),\ p, q = 0, \cdots, 60$.

$$J(x) = \frac{1}{2}\langle x, Ax\rangle - \frac{1}{2}, \quad A = \mathrm{diag}(1, \frac{25}{4}, 0.5, \cdots, 0.5)$$

$$M = 0.1 \times \begin{pmatrix} -20 & 0 & 0 & 0 & 0 & 0 \\ 0 & -25 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ -0.744 & -0.032 & 0 & -0.154 & -0.0052 & 1.54 \\ 0.337 & -1.12 & 0 & 0.249 & -0.1 & -5.2 \\ -0.02 & 0 & 0.0386 & -0.996 & -0.000295 & -0.117 \end{pmatrix}$$

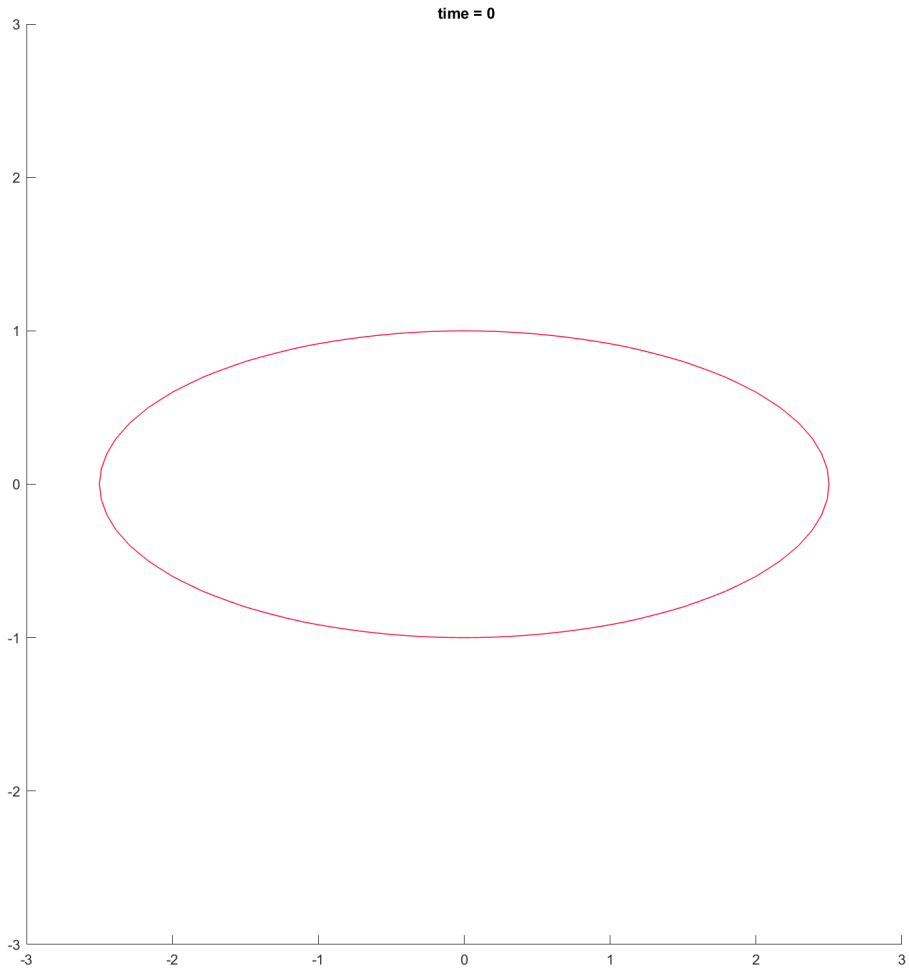
$$N_C = 0.1 \times \mathrm{diag}(20, 25, 0, \cdots, 0), \quad N_D = 0$$

$$C(t) = \{\langle x - a_C, Q_C^{-1}(x - a_C)\rangle \le 1\}, \quad D(t) = \{0\}.$$

where Q_C is tridiagonal with 3's on the diagonal and 1's above and below, $a_C = (-1, -0.5, 0, \cdots, 0)^T$ and $T = 0.5$.

We will show the zero contours $\{z \in \Omega : \phi(t, z) = 0\}$ for $t \in [0, 0.5]$.

Numerical Experiments - $d = 6$

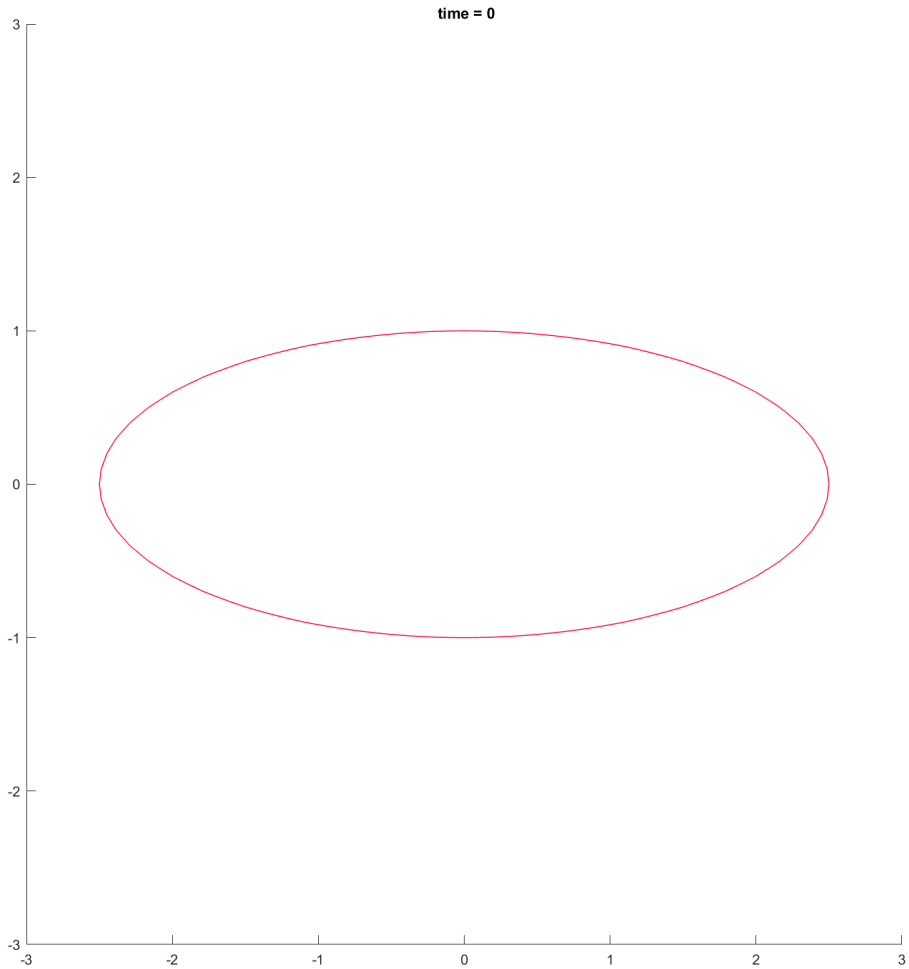


Numerical Experiments - $d = 6$, Non-convex Hamiltonian

Recall that the implementation is higly parallelizable. In fact, in *our* implementation, parallelization makes it roughly 2 times faster, so that the computational time per point is only about 3ms per point, even with a mobile CPU (1.8GHz intel i7-8550U).

The next example is non-convex Hamiltonian case: we have the same M and N_C, but now $N_D = 0.1 \times I_6$, Q_C and Q_D has exponential factors, and a_D equals a_C with the same initial value as previous example, but changes by $0.1 \times (\cos(\frac{\pi t}{T}), \sin(\frac{\pi t}{T}))$ with t.
In this example, due to non-convexity, for optimization problem we used 20 random initial points and took the best one.

Numerical Experiments - $d = 6$, Non-convex Hamiltonian



Numerical Experiments - $d = 2, 3. \cdots, 20$

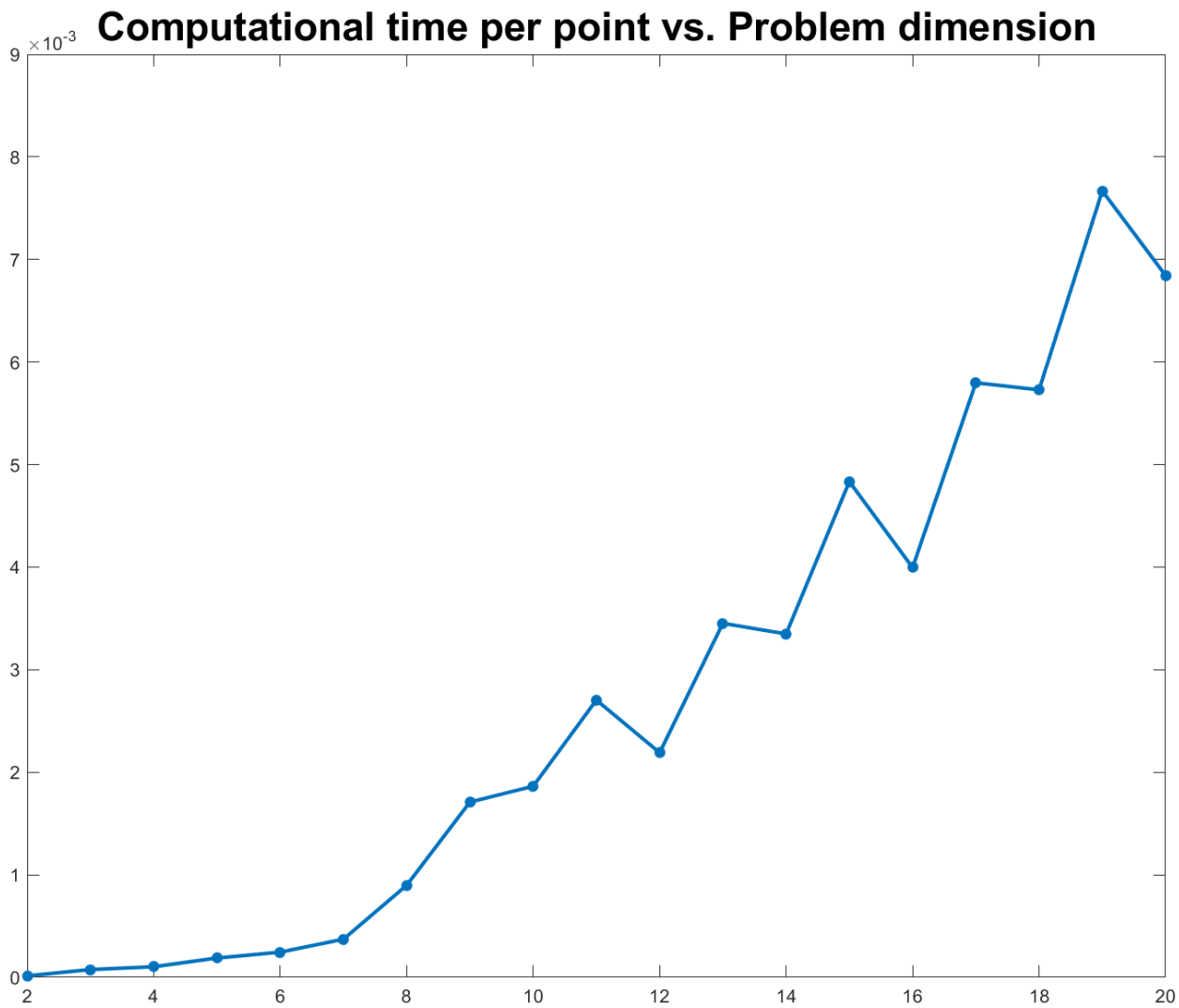
Finally, we now show the computational time per dimension d. Remember that, for FDM/FEM, the number of grid points or mesh elements grows exponentially with respect to d!

In this example,

- M : tridiagonal with only 1's
- N_C : tridiagonal with 1's (main diagonal) and 0.5's (above/below)
- Q_C: tridiagonal with 0.3's (main) and 0.1's (above/below)
- $N_D = 0.1 \times I_d$, $Q_D = 0.01 \times I_d$, $a_C = a_D = 0$
- $d = 2, 3, 4, \cdots, 20$.

Average time of seven experiments

Numerical Experiments - $d = 2, 3. \cdots, 20$



Conclusion

- Generalized Hopf formula is applicable for Hamilton-Jacobi equations arising from certain types of optimal control and differential game problems
- Using Coordinate descent or ADMM/split-Bregman method, we can effectively solve the HJ equation, which is possibly non-convex time-dependent.
- Unlike FDM/FEM, it gives exact solutions and does not suffer from exponentially growing grid points