STOCHASTIC OPTIMIZATION METHODS FOR THE SIMULTANEOUS CONTROL OF PARAMETER-DEPENDENT SYSTEMS

Umberto Biccari
Fundación Deusto and Universidad de Deusto, Bilbao, Spain
umberto.biccari@deusto.es cmc.deusto.es/umberto-biccari

joint work with:
Ana Navarro - Universitat de València
Enrique Zuazua - FAU, Fundación Deusto and Universidad Autónoma de Madrid

June 12, 2020
Introduction
Key concepts of the presentation:

- parameter-depending models
- simultaneous controllability
- stochastic optimization
Parameter-depending models appear in many real-life applications, to describe physical phenomena which may have different realizations:

\[
\begin{cases}
    x_\nu'(t) = A_\nu x_\nu(t) + B u(t), & 0 < t < T, \\
    x_\nu(0) = x_0, & \nu \in \mathcal{K}
\end{cases}
\]

**Example 1:** linearized cart-inverted pendulum system

\[
\begin{pmatrix}
    \dot{x}_\nu \\
    \dot{v}_\nu \\
    \dot{\theta}_\nu \\
    \dot{\omega}_\nu
\end{pmatrix}
= 
\begin{pmatrix}
    0 & 0 & 1 & 0 \\
    0 & -\frac{\nu}{M} & 0 & 0 \\
    0 & 0 & 0 & 1 \\
    0 & \frac{\nu + M}{M\ell} & 0 & 0
\end{pmatrix}
\begin{pmatrix}
    x_\nu \\
    v_\nu \\
    \theta_\nu \\
    \omega_\nu
\end{pmatrix}
+ 
\begin{pmatrix}
    0 \\
    1 \\
    0 \\
    -1
\end{pmatrix}
\cdot u.
\]
Parameter-depending models appear in many real-life applications, to describe physical phenomena which may have different realizations

\[ \begin{cases} x'_\nu(t) = A_\nu x_\nu(t) + B u(t), & 0 < t < T, \\ x_\nu(0) = x^0, & \nu \in \mathcal{K} \end{cases} \]

**Example 2:** system of thermoelasticity

\[
\begin{align*}
w_{tt} - \mu \Delta w - (\lambda + \mu) \nabla \text{div}(w) + \alpha \nabla \theta &= u 1_\omega \\
\theta_t - \Delta \theta + \beta \text{div}(w_t) &= 0
\end{align*}
\]

Lebeau and Zuazua, Null controllability of a system of linear thermoelasticity, 2002
Simultaneous controllability

We look for a unique **parameter-independent** control $u$ such that, at time $T > 0$, the corresponding solution $x_\nu$ satisfies

$$x_\nu(T) = x^T, \quad \text{for all} \quad \nu \in \mathcal{K}$$

In the ODE setting, simultaneous controllability is equivalent to the classical controllability of the augmented system

$$\dot{x} = \mathcal{A}x + \mathcal{B}u$$

with $x = (x_\nu_1, \ldots, x_\nu_{|\mathcal{K}|})^T \in \mathbb{R}^{N|\mathcal{K}|}$, $u = (u, \ldots, u)^T \in L^2(0, T; \mathbb{R}^{N|\mathcal{K}|})$, and where the matrices $\mathcal{A}$ and $\mathcal{B}$ are given by

$$\mathcal{A} = \begin{pmatrix} A_\nu_1 & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & A_\nu_{|\mathcal{K}|} \end{pmatrix} \in \mathbb{R}^{N|\mathcal{K}| \times N|\mathcal{K}|} \quad \text{and} \quad \mathcal{B} = \begin{pmatrix} B \\ \vdots \\ B \end{pmatrix} \in \mathbb{R}^{N|\mathcal{K}| \times 1}$$

Lohéac and Zuazua, *From averaged to simultaneous controllability*, 2016
Simultaneous controllability

We look for a unique parameter-independent control $u$ such that, at time $T > 0$, the corresponding solution $x_\nu$ satisfies

$$x_\nu(T) = x^T, \quad \text{for all } \nu \in \mathcal{K}$$

In the ODE setting, simultaneous controllability is equivalent to the classical controllability of the augmented system

$$\dot{x} = Ax + Bu$$

with $x = (x_{\nu_1}, \ldots, x_{\nu|\mathcal{K}|})^T \in \mathbb{R}^{N|\mathcal{K}|}$, $u = (u, \ldots, u)^T \in L^2(0, T; \mathbb{R}^{N|\mathcal{K}|})$, and where the matrices $A$ and $B$ are given by

$$A = \begin{pmatrix} A_{\nu_1} & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & A_{\nu|\mathcal{K}|} \end{pmatrix} \in \mathbb{R}^{N|\mathcal{K}| \times N|\mathcal{K}|} \quad \text{and} \quad B = \begin{pmatrix} B \\ \vdots \\ B \end{pmatrix} \in \mathbb{R}^{N|\mathcal{K}| \times 1}$$

Lohéac and Zuazua, From averaged to simultaneous controllability, 2016
Computation of simultaneous controls

\[ \hat{u} = \min_{u \in L^2(0,T;\mathbb{R}^M)} F_\nu(u) \]

\[ F_\nu(u) := \frac{1}{2} E \left[ \left\| x_\nu(T) - x^T \right\|_{\mathbb{R}^N}^2 \right] + \frac{\beta}{2} \|u\|_{L^2(0,T;\mathbb{R}^M)}^2 \]

\[ F_\nu(u) := \frac{1}{|\mathcal{K}|} \sum_{\nu \in \mathcal{K}} f_{\nu_k} + \frac{\beta}{2} \|u\|_{L^2(0,T;\mathbb{R}^M)}^2 \]

Typical approaches:

- Gradient Descent (GD): \( u^{k+1} = u^k - \eta_k \nabla F_\nu(u^k) \)
- Conjugate Gradient (CG)

Nocedal and Wright, Numerical optimization, 1999

Ciarlet, Introduction à l’analyse numérique matricielle et à l’optimisation, 1988

Both approaches have a high computational cost when dealing with large parameter sets.
Computation of simultaneous controls

\[
\hat{u} = \min_{u \in L^2(0,T;\mathbb{R}^M)} F_\nu(u)
\]

\[
F_\nu(u) := \frac{1}{2} \mathbb{E} \left[ \left\| x_\nu(T) - x^T \right\|_{\mathbb{R}^N}^2 \right] + \frac{\beta}{2} \|u\|_{L^2(0,T;\mathbb{R}^M)}^2
\]

\[
F_\nu(u) := \frac{1}{|\mathcal{K}|} \sum_{\nu_k \in \mathcal{K}} f_{\nu_k} + \frac{\beta}{2} \|u\|_{L^2(0,T;\mathbb{R}^M)}^2
\]

Typical approaches:

- Gradient Descent (GD):
  \[
u^{k+1} = \nu^k - \eta_k \nabla F_\nu(\nu^k)\]

- Conjugate Gradient (CG)

Nocedal and Wright, Numerical optimization, 1999

Ciarlet, Introduction à l’analyse numérique matricielle et à l’optimisation, 1988

Both approaches have a high computational cost when dealing with large parameter sets.
Computation of simultaneous controls

\[ \hat{u} = \min_{u \in L^2(0,T;\mathbb{R}^M)} F_{\nu}(u) \]

\[
F_{\nu}(u) := \frac{1}{2} \mathbb{E} \left[ \| x_\nu(T) - x^T \|_{\mathbb{R}^N}^2 \right] + \frac{\beta}{2} \| u \|_{L^2(0,T;\mathbb{R}^M)}^2
\]

\[
F_{\nu}(u) := \frac{1}{|\mathcal{K}|} \sum_{\nu_k \in \mathcal{K}} f_{\nu_k} + \frac{\beta}{2} \| u \|_{L^2(0,T;\mathbb{R}^M)}^2
\]

Typical approaches:

- Gradient Descent (GD): \( u^{k+1} = u^k - \eta_k \nabla F_{\nu}(u^k) \)
- Conjugate Gradient (CG)

Nocedal and Wright, Numerical optimization, 1999

Ciarlet, Introduction à l’analyse numérique matricielle et à l’optimisation, 1988

Both approaches have a high computational cost when dealing with large parameter sets.
**STOCHASTIC GRADIENT DESCENT (SGD)**

This is a simplification of the classical GD in which, instead of computing $\nabla F_\nu$ for all parameters $\nu \in \mathcal{K}$, in each iteration this gradient is estimated on the basis of a single randomly picked configuration

$$u^{k+1} = u^k - \eta_k \nabla f_{\nu_k}(u^k)$$


**CONTINUOUS STOCHASTIC GRADIENT (CSG)**

This is a variant of SGD, based on the idea of reusing previously obtained information to improve the efficiency of the algorithm

$$u^{k+1} = u^k - \eta_k G^k, \quad G^k = \sum_{\ell=1}^{k} \alpha_\ell \nabla f_{\nu_\ell}(u^\ell)$$

Pflug, Bernhardt, Grieshammer and Stingl, *A new stochastic gradient method for the efficient solution of structural optimization problems with infinitely many state problems*, 2020
Optimization algorithms
Gradient Descent

\[ u^{k+1} = u^k - \eta_k \nabla F_\nu(u^k) \]

Convergence

Since \( F_\nu \) is convex, if we take \( \eta_k \) constant small enough, we have

\[ \| u^k - \hat{u} \|_{\mathbb{R}^N}^2 \leq \| u^0 - \hat{u} \|_{\mathbb{R}^N}^2 e^{-2C_{GD}k}, \quad C_{GD} = \ln \left( \frac{\rho + 1}{\rho - 1} \right) \]

\[ \| u^k - \hat{u} \|_{\mathbb{R}^N}^2 < \varepsilon \quad \rightarrow \quad k = O \left( \frac{\ln(\varepsilon^{-1})}{C_{GD}} \right) \rightarrow \quad \text{cost}_{GD} = O \left( \frac{|K| \ln(\varepsilon^{-1})}{C_{GD}} \right) \]
\[ u^{k+1} = u^k - \eta_k \left( \beta u^k - \frac{1}{|\mathcal{K}|} \sum_{\nu \in \mathcal{K}} B^T p^k_\nu \right) \]

\[
\begin{cases}
    x'_\nu(t) = A_\nu x_\nu(t) + B u, & 0 < t < T \\
    p'_\nu(t) = -A^T_\nu p_\nu(t), & 0 < t < T \\
    x_\nu(0) = x^0, & p_\nu(T) = -(x_\nu(T) - x^T)
\end{cases}
\]

**Convergence**

Since \( F_\nu \) is convex, if we take \( \eta_k \) constant small enough, we have

\[
\left\| u^k - \hat{u} \right\|_{\mathbb{R}^N}^2 \leq \left\| u^0 - \hat{u} \right\|_{\mathbb{R}^N}^2 e^{-2C_{GD} k}, \quad C_{GD} = \ln \left( \frac{\rho + 1}{\rho - 1} \right)
\]

\[
\left\| u^k - \hat{u} \right\|_{\mathbb{R}^N}^2 < \varepsilon \quad \Rightarrow \quad k = \mathcal{O} \left( \frac{\ln(\varepsilon^{-1})}{C_{GD}} \right) \quad \Rightarrow \quad \text{cost}_{GD} = \mathcal{O} \left( \frac{|\mathcal{K}| \ln(\varepsilon^{-1})}{C_{GD}} \right)
\]
The expected exponential convergence of GD may be violated in practice. The convergence rate is given in terms of the constant $C_{GD}(\rho)$ which is positive decreasing converge to zero as $\rho \to +\infty$.

A bad conditioning in a minimization problem affects the actual convergence of GD.

Example

$$\min_{x \in \mathbb{R}} \left( \frac{1}{2} x^T Q \tau x - b^T x \right)$$

$$Q \tau = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \tau & 0 \\ 0 & 0 & \tau^2 \end{pmatrix} \quad b = -\begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

$$\rho = \frac{\lambda_{\text{max}}}{\lambda_{\text{min}}} = \tau^2$$

<table>
<thead>
<tr>
<th>$\tau$</th>
<th>iterations</th>
<th>$\rho$</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>27</td>
<td>4</td>
</tr>
<tr>
<td>5</td>
<td>161</td>
<td>25</td>
</tr>
<tr>
<td>10</td>
<td>633</td>
<td>100</td>
</tr>
<tr>
<td>20</td>
<td>2511</td>
<td>400</td>
</tr>
<tr>
<td>50</td>
<td>15619</td>
<td>2500</td>
</tr>
</tbody>
</table>

Meza, Steepest descent, 2010
The expected exponential convergence of GD may be violated in practice.
The convergence rate is given in terms of the constant $C_{GD}(\rho)$ which is positive decreasing converge to zero as $\rho \to +\infty$.

A bad conditioning in a minimization problem affects the actual convergence of GD.

**Example**

$$\min_{x \in \mathbb{R}} \left( \frac{1}{2} x^T Q_{\tau} x - b^T x \right)$$

$$Q_{\tau} = \begin{pmatrix}
1 & 0 & 0 \\
0 & \tau & 0 \\
0 & 0 & \tau^2 \\
\end{pmatrix} \quad b = - \begin{pmatrix}
1 \\
1 \\
\end{pmatrix}$$

$$\rho = \frac{\lambda_{\text{max}}}{\lambda_{\text{min}}} = \tau^2$$

<table>
<thead>
<tr>
<th>$\tau$</th>
<th>iterations</th>
<th>$\rho$</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>27</td>
<td>4</td>
</tr>
<tr>
<td>5</td>
<td>161</td>
<td>25</td>
</tr>
<tr>
<td>10</td>
<td>633</td>
<td>100</td>
</tr>
<tr>
<td>20</td>
<td>2511</td>
<td>400</td>
</tr>
<tr>
<td>50</td>
<td>15619</td>
<td>2500</td>
</tr>
</tbody>
</table>

Meza, Steepest descent, 2010
Conjugate Gradient

$$\nabla F_\nu(u) = \beta u - \frac{1}{|K|} \sum_{\nu \in K} B^T p_\nu$$

**Convergence**

$$\|u^k - \hat{u}\|_{\mathbb{R}^N}^2 \leq 4 \|u^0 - \hat{u}\|_{\mathbb{R}^N}^2 e^{-2c_{CG}k}, \quad c_{CG} = \ln \left( \frac{\sqrt{\rho} + 1}{\sqrt{\rho} - 1} \right)$$

$$\|u^k - \hat{u}\|_{\mathbb{R}^N}^2 < \varepsilon \quad \Rightarrow \quad k = \mathcal{O} \left( \frac{\ln(\varepsilon^{-1})}{c_{CG}} \right) \quad \Rightarrow \quad cost_{CG} = \mathcal{O} \left( \frac{|K| \ln(\varepsilon^{-1})}{c_{CG}} \right)$$
Conjugate Gradient

\[ \nabla F_\nu(u) = (\beta I + \mathbb{E}[\mathcal{L}_{T,\nu}^* \mathcal{L}_{T,\nu}]) u + \mathbb{E}[\mathcal{L}_{T,\nu}^* (y_\nu(T) - x^T)] \rightarrow Au = b \]

\[ \mathcal{L}_{T,\nu} : L^2(0, T; \mathbb{R}^M) \rightarrow \mathbb{R}^N \quad u \mapsto z_\nu(T) \]

\[ \mathcal{L}_{T,\nu}^* : \mathbb{R}^N \rightarrow L^2(0, T; \mathbb{R}^M) \quad p_{T,\nu} \mapsto B^T p_\nu \]

\[ \begin{cases} \ y'_\nu(t) = A_\nu y_\nu(t), & 0 < t < T \\ \ y_\nu(0) = x_0 \end{cases} \]

\[ \begin{cases} \ z'_\nu(t) = A_\nu z_\nu(t) + Bu(t), & 0 < t < T \\ \ z_\nu(0) = 0 \end{cases} \]

Convergence

\[ \left\| u^k - \hat{u} \right\|_{\mathbb{R}^N}^2 \leq 4 \left\| u^0 - \hat{u} \right\|_{\mathbb{R}^N}^2 e^{-2C_{CG}k}, \quad C_{CG} = \ln \left( \frac{\sqrt{\rho} + 1}{\sqrt{\rho} - 1} \right) \]

\[ \left\| u^k - \hat{u} \right\|_{\mathbb{R}^N}^2 < \varepsilon \rightarrow k = \mathcal{O} \left( \frac{\ln(\varepsilon^{-1})}{C_{CG}} \right) \rightarrow \text{cost}_{CG} = \mathcal{O} \left( \frac{|\mathcal{K}| \ln(\varepsilon^{-1})}{C_{CG}} \right) \]
The expected exponential convergence of CG may be violated in practical experiments, although the situation is less critical than in GD.

- The constant $C_{CG}(\rho)$ depends on the square root of $\rho$, hence CG is less sensible to the conditioning of the problem.

- CG enjoys the **finite termination property**. This means that, if we apply CG to solve a $N$-dimensional problem, the algorithm will converge in at most $N$-iterations.
Applying SGD for minimizing $F_{\nu}(u)$ requires, at each iteration $k$, only one resolution of the dynamics.
Stochastic Gradient Descent

\[ u^{k+1} = u^k - \eta_k \left( \beta u^k - B^T p^k_{\nu_k} \right) \]

\[
\begin{cases}
    x'_{\nu_k}(t) = A_{\nu_k} x_{\nu_k}(t) + Bu, & 0 < t < T \\
p'_{\nu_k}(t) = -A_{\nu_k}^T p_{\nu_k}(t), & 0 < t < T \\
x_{\nu_k}(0) = x^0, & p_{\nu_k}(T) = -(x_{\nu_k}(T) - x^T)
\end{cases}
\]

Applying SGD for minimizing \( F_{\nu}(u) \) requires, at each iteration \( k \), only one resolution of the dynamics.
In SGD the iterate sequence \((u^k)_{k \geq 1}\) is a stochastic process determined by the random sequence \((\nu_k)_{k \geq 1} \subset \mathcal{K}\). Hence, the convergence properties are defined in **expectation** \(\mathbb{E} \left[ \|u^k + 1 - \hat{u}\|_{\mathbb{R}^N}^2 \right]\) or in the context of **almost sure convergence**.

Bach and Moulines, Non-asymptotic analysis of stochastic approximation algorithms for machine learning, 2011

Bottou, Online learning and stochastic approximations, 1998

In SGD, convergence is guaranteed if the step-sizes are chosen such that \(\mathbb{E} \left[ \|\nabla F_\nu(u^k)\|^2 \right]\) is bounded above by a deterministic quantity. In particular, a fixed step-size \(\eta_k = \bar{\eta}\), even if small, does not allow to converge. A standard approach is to use as a decreasing sequence such that

\[
\sum_{k=1}^{\infty} \eta_k = +\infty \quad \text{and} \quad \sum_{k=1}^{\infty} \eta_k^2 < +\infty
\]

Robbins and Monro, A stochastic approximation method, 1951

Bottou, Curtis and Nocedal, Optimization methods for large-scale machine learning, 2018
If $\eta_k$ is properly chosen, by means of standard martingale techniques we can show that the SGD converges almost surely

$$ u^k \overset{a.s.}{\to} \hat{u}, \quad \text{as} \; k \to +\infty $$

**Convergence rate**

Because of the noise introduced by the random selection of the descent direction the convergence of SGD is linear

$$ \mathbb{E} \left[ \left\| u^k - \hat{u} \right\|_{\mathbb{R}^N}^2 \right] = O \left( k^{-1} \right) $$

$$ \mathbb{E} \left[ \left\| u^k - \hat{u} \right\|_{\mathbb{R}^N}^2 \right] < \varepsilon \quad \to \quad k = O \left( \varepsilon^{-1} \right) \quad \to \quad \text{cost}_{SGD} = O \left( \varepsilon^{-1} \right) $$
Stochastic Gradient Descent - convergence

If $\eta_k$ is properly chosen, by means of standard martingale techniques we can show that the SGD converges almost surely

$$u^k \overset{a.s.}{\to} \hat{u}, \quad \text{as } k \to +\infty$$

Convergence rate

Because of the noise introduced by the random selection of the descent direction the convergence of SGD is linear

$$\mathbb{E} \left[ \left\| u^k - \hat{u} \right\|_{\mathbb{R}^N}^2 \right] = \mathcal{O} \left( k^{-1} \right)$$

$$\mathbb{E} \left[ \left\| u^k - \hat{u} \right\|_{\mathbb{R}^N}^2 \right] < \varepsilon \quad \rightarrow \quad k = \mathcal{O} \left( \varepsilon^{-1} \right) \quad \rightarrow \quad \text{cost}_{SGD} = \mathcal{O} \left( \varepsilon^{-1} \right)$$
Continuous Stochastic Gradient

\[ u^{k+1} = u^k - \eta_k G^k, \quad G^k = \sum_{\ell=1}^{k} \alpha_\ell \nabla f_{\nu_\ell}(u^\ell) \]

**Convergence properties**

As the optimization process evolves, the approximated gradient \( G^k \) converges almost surely to the full gradient of the objective functional

\[ G^k \xrightarrow{a.s} \nabla F_\nu, \quad \text{as } k \to +\infty \]

In particular, CSG is a less noisy algorithm and has a better convergence behavior. In particular, convergence may be guaranteed also choosing a fixed learning rate sequence \( \eta_k = \bar{\eta} \).

Pflug, Bernhardt, Grieshammer and Stingl, A new stochastic gradient method for the efficient solution of structural optimization problems with infinitely many state problems, 2020
Continuous Stochastic Gradient

\[ u^{k+1} = u^k - \eta_k \sum_{\ell=1}^{k} \alpha_{\ell} \left( \beta u_\ell - B^\top p_{\nu \ell} \right) \]

**Convergence properties**

As the optimization process evolves, the approximated gradient \( G^k \) converges almost surely to the full gradient of the objective functional

\[ G^k \overset{a.s.}{\to} \nabla F_{\nu}, \quad \text{as } k \to +\infty \]

In particular, CSG is a less noisy algorithm and has a better convergence behavior. In particular, convergence may be guaranteed also choosing a fixed learning rate sequence \( \eta_k = \bar{\eta} \).

Pflug, Bernhardt, Grieshammer and Stingl, A new stochastic gradient method for the efficient solution of structural optimization problems with infinitely many state problems, 2020
Numerical simulations
The system includes a cart of mass $M$ and a rigid pendulum of length $\ell$.

The pendulum is anchored to the cart and at the free extremity it is placed a variable mass described by the parameter $\nu$.

The cart moves on a horizontal plane. The states $x_\nu(t)$ and $v_\nu(t)$ describe its position and velocity, respectively.

During the motion of the cart the pendulum deviates from the initial vertical position by an angle $\theta_\nu(t)$, with an angular velocity $\omega_\nu(t)$.

Starting from an initial state $(x^i, v^i, 0, 0)$, we want to compute a parameter-independent control function $u$ steering all the realizations of the system in time $T$ to the final state $(x^f, 0, 0, 0)$. 
The system includes a cart of mass $M$ and a rigid pendulum of length $\ell$.

The pendulum is anchored to the cart and at the free extremity it is placed a variable mass described by the parameter $\nu$.

The cart moves on a horizontal plane. The states $x_\nu(t)$ and $v_\nu(t)$ describe its position and velocity, respectively.

During the motion of the cart the pendulum deviates from the initial vertical position by an angle $\theta_\nu(t)$, with an angular velocity $\omega_\nu(t)$.

Starting from an initial state $(x^i, v^i, 0, 0)$, we want to compute a parameter-independent control function $u$ steering all the realizations of the system in time $T$ to the final state $(x^f, 0, 0, 0)$. 
Numerical simulations

Input data

- \( x^0 = (-1, 1, 0, 0)^T \)
- \( x^T = (0, 0, 0, 0)^T \)
- \( T = 1s \)
- \( \varepsilon = 10^{-4} \)
- \( M = 10 \)
- \( \ell = 1 \)
- \( \nu \in \mathcal{K} = \{\nu_1, \ldots, \nu_{|\mathcal{K}|}\} \) with \( \nu_1 = 0.1 \) and \( \nu_{|\mathcal{K}|} = 1 \)
Numerical simulations

Input data

- $x^0 = (-1,1,0,0)^T$
- $x^T = (0,0,0,0)^T$
- $T = 1s$
- $\varepsilon = 10^{-4}$

- $M = 10$
- $\ell = 1$
- $\nu \in \mathcal{K} = \{\nu_1, \ldots, \nu_{|\mathcal{K}|}\}$ with $\nu_1 = 0.1$ and $\nu_{|\mathcal{K}|} = 1$

Control

![Control Plot](image)
Numerical simulations

| $|\mathcal{K}|$ | Iter. | Time    | Iter. | Time    | Iter. | Time    | Iter. | Time    |
|-------|-------|---------|-------|---------|-------|---------|-------|---------|
| 2     | 1868  | 45.1s   | 12    | 1.1s    | 2195  | 33.1s   | 930   | 18.6s   |
| 10    | 1869  | 150.1s  | 13    | 2.6s    | 2106  | 31.4s   | 923   | 17.4s   |
| 100   | 1870  | 1799.5s | 12    | 17.7s   | 2102  | 28.9s   | 929   | 17.4s   |
| 250   |       |         | 13    | 50.3s   | 2080  | 28.2s   | 928   | 17.9s   |
| 500   |       |         | 13    | 101.3s  | 2099  | 32.9s   | 927   | 21.5s   |

![Graph showing computational time and number of parameters for different algorithms](image-url)
Numerical simulations

CSG outperforms SGD in terms of the number of iterations it requires to converge and, consequently, of the total computational time. This because the optimization process is less noisy than SGD, yielding to a better convergence behavior.
Conclusions

We compared the GD, CG, SGD and CSG algorithms for the minimization of a quadratic functional associated to the simultaneous controllability of linear parameter-dependent models.

We observed the following:

1. The GD approach is the worst one in terms of the computational complexity, as a consequence of the bad conditioning of the simultaneous controllability problem.

2. The choice of SGD and CSG instead of CG is preferable only when dealing with parameter sets of large cardinality $|\mathcal{K}|$. 
In the PDE setting, simultaneous controllability is a quite delicate issue because of the appearance of peculiar phenomena which are not detected at the ODE level.

For some PDE systems, simultaneous controllability may be understood by looking at the spectral properties of the model. Roughly speaking, one needs all the eigenvalues to have multiplicity one in order to be able to observe every eigenmode independently. This fact generally yields restrictions to the validity of simultaneous controllability, which may be difficult to tackle at the numerical level.

Dáger and Zuazua, Controllability of star-shaped networks of strings, 2001
Open problems

COMPARISON WITH THE GREEDY METHODOLOGY

- The greedy approach aims to approximate the dynamics and controls of linear parameter-dependent by identifying the most meaningful realizations of the parameters.
  Lazar and Zuazua, Greedy controllability of finite dimensional linear systems, 2016
  Hernández-Santamaría, Lazar and Zuazua, Greedy controllability of finite dimensional linear systems, 2019
- A comparison of the greedy and stochastic would be an interesting issue.
THANK YOU FOR YOUR ATTENTION!

This project has received funding from the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation program (grant agreement No 694126-DYCON).