Mesh-Independent Algorithms for Optimal Control of Partial Differential Equations

Roland Herzog

Interdisciplinary Center for Scientific Computing

Heidelberg University

Summer School on Hardware-Aware Scientific Computing

October 04–15, 2021

Purpose of This Talk

- Give a brief introduction to basic PDE-constrained optimization problems . . .
- as a class of (hopefully interesting) target problems for HPC architectures.
- Emphasize the importance to view these problems from the continuous level . . .
- in order to devise overall efficient solution algorithms.

Table of Contents

Hilbert Spaces

Suppose that H is a vector space endowed with an inner product $(\cdot,\cdot)$:

- $(u,v)_H = (v,u)_H$
- $(\alpha_1\, u_1 + \alpha_2\, u_2, v)_H = \alpha_1 (u_1, v)_H + \alpha_2 (u_2, v)_H$
- $(u,u)_H \ge 0$
- $(u,u)_H = 0$ if and only if $u = 0$

The inner product generates a norm in H:

$$\|u\|_H = \sqrt{(u,u)_H}.$$

H is called a Hilbert space provided that it is complete with respect to this norm.

Examples of Hilbert Spaces

Suppose that $\Omega \subset \mathbb{R}^n$ is an open set.

1. $$L^2(\Omega) = \left\{ u \colon \Omega \to \mathbb{R} \text{ measurable} \,\middle|\, \int_\Omega |u|^2 \, \mathrm{d}x \text{ is finite} \right\}$$

   is a Hilbert space w.r.t. the inner product

   $$(u, v)_{L^2(\Omega)} := \int_\Omega u \, v \, \mathrm{d}x.$$

2. $$H^1(\Omega) = \left\{ u \colon \Omega \to \mathbb{R} \text{ measurable} \,\middle|\, \int_\Omega |\nabla u|_2^2 + |u|^2 \, \mathrm{d}x \text{ is finite} \right\}$$

   is a Hilbert space w.r.t. the inner product

   $$(u, v)_{H^1(\Omega)} := \int_\Omega \nabla u \cdot \nabla v + u \, v \, \mathrm{d}x.$$

Unconstrained Optimization in Hilbert Space

$$\text{Minimize} \quad f(u) \quad \text{where } u \in H$$

$f : H \to \mathbb{R}$ is called the objective functional.

$f'(u) \in \mathcal{L}(H, \mathbb{R})$ is called the (Fréchet) derivative of f at $u \in H$ provided that

$$\frac{1}{\|\delta u\|_H} \big[f(u + \delta u) - f(u) - f'(u)\, \delta u \big] \longrightarrow 0 \quad \text{as } \delta u \to 0.$$

- The space $\mathcal{L}(H, \mathbb{R}) = H^*$ of bounded linear functionals $H \to \mathbb{R}$ is called the dual space of H.
- The dual space H^* carries the operator norm:

$$\|\xi\|_{H^*} := \sup\big\{ \langle \xi \, , u \rangle \;\big|\; \|u\|_H = 1 \big\}.$$

Implications for Iterative Optimization Methods

Minimize $f(u)$ where $u \in H$

At a point (iterate) $u \in H$, we require a search direction $d \in H$ to come to the next iterate

$$u + \alpha\, d \in H.$$

How do we obtain a direction $d \in H$ in which to proceed using the derivative $f'(u) \in H^*$?

The Riesz Map

Theorem

1. *In a Hilbert space* H*, the elements* ξ *of its dual space* H^* *are precisely of the form* $(u_\xi, \cdot)_H$.
2. *The relation* $R\colon \xi \mapsto u_\xi$ *is linear (the* Riesz map*).*
3. *Using the Riesz map, we can define an inner product on* H^**:*

$$(\xi, \varphi)_{H^*} \coloneqq (R\xi, R\varphi)_H = \langle \xi \, , R\varphi \rangle = \langle \varphi \, , R\xi \rangle.$$

4. $(\cdot, \cdot)_{H^*}$ *defines a norm, which agrees with the operator norm.*
5. H^* *is itself a* Hilbert space *and* $R\colon H^* \to H$ *is an isometry.*

This statement can be summarized as

A Hilbert space H “can be identified with” its dual H^*.

but it often leads to confusion.

Implications for Iterative Optimization Methods

Minimize $f(u)$ where $u \in H$

At a point (iterate) $u \in H$, we require a search direction $d \in H$ to come to the next iterate

$$u + \alpha\, d \in H.$$

How do we obtain a direction $d \in H$ in which to proceed using the derivative $f'(u) \in H^*$?

Implications for Iterative Optimization Methods

Minimize $\quad f(u) \quad$ where $u \in H$

At a point (iterate) $u \in H$, we require a search direction $d \in H$ to come to the next iterate

$$u + \alpha\, d \in H.$$

How do we obtain a direction $d \in H$ in which to proceed using the derivative $f'(u) \in H^*$?

The Riesz representer of $f'(u) \in H^*$ is called the gradient of f at u:

$$\nabla f(u) = R\, f'(u).$$

Steepest Descent Method

Minimize $f(u)$ where $u \in H$

In a steepest descent (gradient descent) method, we use $d := -\nabla f(u)$ as search direction and update an iterate u according to $u + \alpha\, d$.

The negative gradient direction d is obtained by solving the linear system

$$(d, v)_H = -f'(u)\, v \quad \text{for all } v \in H.$$

But What About $H = \mathbb{R}^n$?

Isn't it true that in $\mathbb{R}^n$, the gradient $\nabla f(u) \in \mathbb{R}^n$ is simply the transpose of the derivative $f'(u) \in \mathbb{R}_n$?

But What About $H = \mathbb{R}^n$?

Isn't it true that in $\mathbb{R}^n$, the gradient $\nabla f(u) \in \mathbb{R}^n$ is simply the transpose of the derivative $f'(u) \in \mathbb{R}_n$?

That is the case only when the Euclidean inner product is used:

$$\begin{aligned} & (\nabla f(u), v)_{\mathsf{id}} = f'(u)\, v \quad \text{for all } v \in \mathbb{R}^n \\ \Leftrightarrow \quad & \nabla f(u) = f'(u)^{\mathsf{T}}. \end{aligned}$$

But What About $H = \mathbb{R}^n$?

Isn't it true that in $\mathbb{R}^n$, the gradient $\nabla f(u) \in \mathbb{R}^n$ is simply the transpose of the derivative $f'(u) \in \mathbb{R}_n$?

That is the case only when the Euclidean inner product is used:

$$
\begin{aligned}
& (\nabla f(u), v)_{\mathsf{id}} = f'(u)\, v \quad \text{for all } v \in \mathbb{R}^n \\
\Leftrightarrow \quad & \nabla f(u) = f'(u)^{\mathsf{T}}.
\end{aligned}
$$

When the inner product is represented by an s. p. d. matrix M:

$$
\begin{aligned}
& (\nabla f(u), v)_M = f'(u)\, v \quad \text{for all } v \in \mathbb{R}^n \\
\Leftrightarrow \quad & M\, \nabla f(u) = f'(u)^{\mathsf{T}}.
\end{aligned}
$$

Does the Inner Product Really Matter?

For a quadratic polynomial in $\mathbb{R}^n$

$$f(u) = \frac{1}{2} u^\mathsf{T} A\, u - b^\mathsf{T} u + c \quad \text{with an s. p. d. matrix } A,$$

the speed of convergence of the steepest descent method (with exact line search) is determined by its condition number

$$\kappa := \frac{\lambda_{\max}(A)}{\lambda_{\min}(A)}.$$

Often one refers to the gradient method in a non-Euclidean inner product as a preconditioned gradient method.

Does the Inner Product Really Matter?

For a quadratic polynomial in $\mathbb{R}^n$

$$f(u) = \frac{1}{2} u^\mathsf{T} A\, u - b^\mathsf{T} u + c \quad \text{with an s. p. d. matrix } A,$$

the speed of convergence of the steepest descent method (with exact line search) is determined by its generalized condition number

$$\kappa := \frac{\lambda_{\max}(A; M)}{\lambda_{\min}(A; M)}.$$

Often one refers to the gradient method in a non-Euclidean inner product as a preconditioned gradient method.

Intermediate Summary

1. One needs to distinguish between derivative $f'(u) \in H^*$ and gradient $\nabla f(u) \in H$.
2. When talking about gradients and the steepest descent method, always mention which inner product is being used.
3. Even in $\mathbb{R}^n$:
$$M \, \nabla f(u) = f'(u)^\mathsf{T}$$
4. The specification of the inner product is in the user's responsibility.

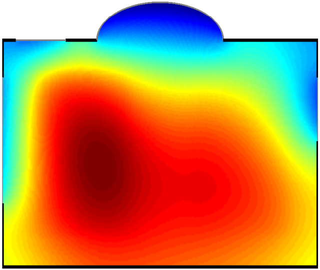
When solving discretized version of infinite-dimensional optimization problems, we can expect to achieve mesh-independent algorithms only when respecting the appropriate inner product of the underlying undiscretized problem.

Table of Contents

A Model Problem

As an introduction, we are going to consider (variations on) the following model problem for the stationary heat equation:

$$\begin{cases} -\kappa\, \Delta y(x) = \chi_{\text{ctrl}}\, u(x) & \text{in } \Omega \\ \kappa\, \frac{\partial}{\partial n} y(x) + \alpha\, y(x) = 0 & \text{on } \Gamma = \partial\Omega \end{cases}$$

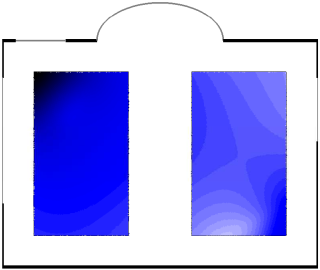


state temperature $y \in \mathcal{Y} := H^1(\Omega)$

A Model Problem

As an introduction, we are going to consider (variations on) the following model problem for the stationary heat equation:

$$\begin{cases} -\kappa\,\Delta y(x) = \chi_{\text{ctrl}}\,u(x) & \text{in } \Omega \\ \kappa\,\frac{\partial}{\partial n} y(x) + \alpha\,y(x) = 0 & \text{on } \Gamma = \partial\Omega \end{cases}$$



state: temperature $y \in \mathcal{Y} := H^1(\Omega)$

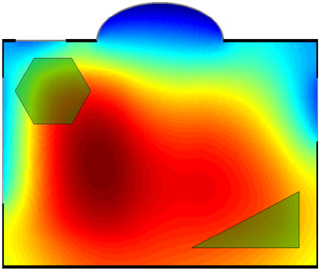
control: heating power $u \in \mathcal{U} := L^2(\Omega_{\text{ctrl}})$

A Model Problem

As an introduction, we are going to consider (variations on) the following model problem for the stationary heat equation:

$$\text{Minimize} \quad \frac{1}{2}\int_{\Omega_{\text{obs}}} \big(y(x) - y_d(x)\big)^2 \, \mathrm{d}x$$

$$\text{s.t.} \quad \begin{cases} -\kappa\, \Delta y(x) = \chi_{\text{ctrl}}\, u(x) & \text{in } \Omega \\ \kappa\, \frac{\partial}{\partial n} y(x) + \alpha\, y(x) = 0 & \text{on } \Gamma = \partial\Omega \end{cases}$$



state: temperature $y \in \mathcal{Y} := H^1(\Omega)$

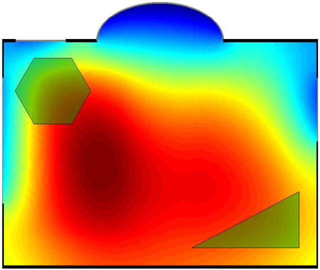
control: heating power $u \in \mathcal{U} := L^2(\Omega_{\text{ctrl}})$

A Model Problem

As an introduction, we are going to consider (variations on) the following model problem for the stationary heat equation:

$$\text{Minimize} \quad \frac{1}{2}\int_{\Omega_{\text{obs}}} \big(y(x) - y_d(x)\big)^2 \, \mathrm{d}x + \frac{\gamma}{2}\int_{\Omega_{\text{ctrl}}} u(x)^2 \, \mathrm{d}x, \qquad \gamma > 0$$

$$\text{s.t.} \quad \begin{cases} -\kappa \, \Delta y(x) = \chi_{\text{ctrl}} \, u(x) & \text{in } \Omega \\ \kappa \, \frac{\partial}{\partial n} y(x) + \alpha \, y(x) = 0 & \text{on } \Gamma = \partial\Omega \end{cases}$$



state — temperature $y \in \mathcal{Y} := H^1(\Omega)$

control — heating power $u \in \mathcal{U} := L^2(\Omega_{\text{ctrl}})$

Optimality Conditions

Lagrangian:

$$\mathcal{L}(y,u,p) = \frac{1}{2}\int_{\Omega_{\text{obs}}} (y - y_d)^2 \, \mathrm{d}x + \frac{\gamma}{2}\int_{\Omega_{\text{ctrl}}} u^2 \, \mathrm{d}x \qquad \leftarrow \text{objective}$$

$$+ \int_{\Omega} \kappa \nabla y \cdot \nabla p \, \mathrm{d}x + \int_{\Gamma} \alpha \, y \, p \, \mathrm{d}s - \int_{\Omega_{\text{ctrl}}} u \, p \, \mathrm{d}x \qquad \leftarrow \text{PDE}$$

Optimality Conditions

Lagrangian:

$$\mathcal{L}(y,u,p) = \frac{1}{2}\int_{\Omega_{\text{obs}}} (y - y_d)^2 \, \mathrm{d}x + \frac{\gamma}{2}\int_{\Omega_{\text{ctrl}}} u^2 \, \mathrm{d}x \qquad \leftarrow \text{objective}$$

$$+ \int_{\Omega} \kappa \nabla y \cdot \nabla p \, \mathrm{d}x + \int_{\Gamma} \alpha \, y \, p \, \mathrm{d}s - \int_{\Omega_{\text{ctrl}}} u \, p \, \mathrm{d}x \qquad \leftarrow \text{PDE}$$

$$\mathcal{L}_y(y,u,p)\,\delta y = \int_{\Omega_{\text{obs}}} (y - y_d)\,\delta y \, \mathrm{d}x + \int_{\Omega} \kappa \nabla \delta y \cdot \nabla p \, \mathrm{d}x + \int_{\Gamma} \alpha \, \delta y \, p \, \mathrm{d}s$$

$$= 0 \quad \text{for all } \delta y \in \mathcal{Y} = H^1(\Omega)$$

Optimality Conditions

Lagrangian:

$$\mathcal{L}(y,u,p) = \frac{1}{2}\int_{\Omega_{\text{obs}}} (y - y_d)^2 \, \mathrm{d}x + \frac{\gamma}{2}\int_{\Omega_{\text{ctrl}}} u^2 \, \mathrm{d}x \qquad \leftarrow \text{objective}$$

$$+ \int_{\Omega} \kappa \, \nabla y \cdot \nabla p \, \mathrm{d}x + \int_{\Gamma} \alpha \, y \, p \, \mathrm{d}s - \int_{\Omega_{\text{ctrl}}} u \, p \, \mathrm{d}x \qquad \leftarrow \text{PDE}$$

adjoint PDE $\begin{cases} -\kappa \, \Delta p = -\chi_{\text{obs}} \, (y - y_d) & \text{in } \Omega \\ \kappa \, \frac{\partial}{\partial n} p + \alpha \, p = 0 & \text{on } \Gamma \end{cases}$

Optimality Conditions

Lagrangian:

$$\mathcal{L}(y,u,p) = \frac{1}{2}\int_{\Omega_{\text{obs}}} (y-y_d)^2 \, \mathrm{d}x + \frac{\gamma}{2}\int_{\Omega_{\text{ctrl}}} u^2 \, \mathrm{d}x \qquad \leftarrow \text{objective}$$

$$+ \int_\Omega \kappa \, \nabla y \cdot \nabla p \, \mathrm{d}x + \int_\Gamma \alpha \, y \, p \, \mathrm{d}s - \int_{\Omega_{\text{ctrl}}} u \, p \, \mathrm{d}x \qquad \leftarrow \text{PDE}$$

adjoint PDE
$$\begin{cases} -\kappa \, \Delta p = -\chi_{\text{obs}} \, (y-y_d) & \text{in } \Omega \\ \kappa \frac{\partial}{\partial n} p + \alpha \, p = 0 & \text{on } \Gamma \end{cases}$$

$$\mathcal{L}_u(y,u,p) \, \delta u = \gamma \int_{\Omega_{\text{ctrl}}} u \, \delta u \, \mathrm{d}x - \int_{\Omega_{\text{ctrl}}} \delta u \, p \, \mathrm{d}x = 0 \quad \text{for all } \delta u \in \mathcal{U} := L^2(\Omega_{\text{ctrl}}$$

Optimality Conditions

Lagrangian:

$$\mathcal{L}(y,u,p) = \frac{1}{2}\int_{\Omega_{\text{obs}}} (y - y_d)^2 \, \mathrm{d}x + \frac{\gamma}{2}\int_{\Omega_{\text{ctrl}}} u^2 \, \mathrm{d}x \qquad \leftarrow \text{objective}$$

$$+ \int_{\Omega} \kappa \, \nabla y \cdot \nabla p \, \mathrm{d}x + \int_{\Gamma} \alpha \, y \, p \, \mathrm{d}s - \int_{\Omega_{\text{ctrl}}} u \, p \, \mathrm{d}x \qquad \leftarrow \text{PDE}$$

adjoint PDE

$$\begin{cases} -\kappa \, \Delta p = -\chi_{\text{obs}} \, (y - y_d) & \text{in } \Omega \\ \kappa \, \frac{\partial}{\partial n} p + \alpha \, p = 0 & \text{on } \Gamma \end{cases}$$

gradient equation

$$\{ \gamma \, u - p = 0 \quad \text{in } \Omega_{\text{ctrl}}$$

Optimality Conditions

Lagrangian:

$$\mathcal{L}(y,u,p) = \frac{1}{2}\int_{\Omega_{\text{obs}}} (y - y_d)^2 \, \mathrm{d}x + \frac{\gamma}{2}\int_{\Omega_{\text{ctrl}}} u^2 \, \mathrm{d}x \qquad \leftarrow \text{objective}$$

$$+ \int_\Omega \kappa \, \nabla y \cdot \nabla p \, \mathrm{d}x + \int_\Gamma \alpha \, y \, p \, \mathrm{d}s - \int_{\Omega_{\text{ctrl}}} u \, p \, \mathrm{d}x \qquad \leftarrow \text{PDE}$$

adjoint PDE

$$\begin{cases} -\kappa \, \Delta p = -\chi_{\text{obs}} \, (y - y_d) & \text{in } \Omega \\ \kappa \, \frac{\partial}{\partial n} p + \alpha \, p = 0 & \text{on } \Gamma \end{cases}$$

gradient equation

$$\{\gamma \, u - p = 0 \quad \text{in } \Omega_{\text{ctrl}}$$

$$\mathcal{L}_p(y,u,p) \, \delta p = \int_\Omega \kappa \, \nabla y \cdot \nabla \delta p \, \mathrm{d}x + \int_\Gamma \alpha \, y \, \delta p \, \mathrm{d}s - \int_{\Omega_{\text{ctrl}}} u \, \delta p \, \mathrm{d}x$$

$$= 0 \quad \text{for all } \delta p \in \mathcal{Y} := H^1(\Omega)$$

Optimality Conditions

Lagrangian:

$$\mathcal{L}(y,u,p) = \frac{1}{2}\int_{\Omega_{\text{obs}}} (y - y_d)^2 \, \mathrm{d}x + \frac{\gamma}{2}\int_{\Omega_{\text{ctrl}}} u^2 \, \mathrm{d}x \qquad \leftarrow \text{objective}$$

$$+ \int_\Omega \kappa \, \nabla y \cdot \nabla p \, \mathrm{d}x + \int_\Gamma \alpha \, y \, p \, \mathrm{d}s - \int_{\Omega_{\text{ctrl}}} u \, p \, \mathrm{d}x \qquad \leftarrow \text{PDE}$$

adjoint PDE
$$\begin{cases} -\kappa \, \Delta p = -\chi_{\text{obs}} \, (y - y_d) & \text{in } \Omega \\ \kappa \, \frac{\partial}{\partial n} p + \alpha \, p = 0 & \text{on } \Gamma \end{cases}$$

gradient equation
$$\{ \gamma \, u - p = 0 \quad \text{in } \Omega_{\text{ctrl}}$$

state PDE
$$\begin{cases} -\kappa \, \Delta y = \chi_{\text{ctrl}} \, u & \text{in } \Omega \\ \kappa \, \frac{\partial}{\partial n} y + \alpha \, y = 0 & \text{on } \Gamma \end{cases}$$

Optimality System as Saddle-Point Problem

self-adjoint saddle-point system:

$$\begin{bmatrix} \chi_{\text{obs}} & \cdot & -\Delta \\ \cdot & \gamma & -\chi_{\text{ctrl}} \\ -\Delta & -\chi_{\text{ctrl}} & \cdot \end{bmatrix} \begin{pmatrix} {\color{red}y} \\ {\color{blue}u} \\ {\color{green}p} \end{pmatrix} = \begin{pmatrix} y_d \\ 0 \\ 0 \end{pmatrix}$$

Optimality System as Saddle-Point Problem

self-adjoint saddle-point system:

$$\begin{bmatrix} \chi_{\text{obs}} & \cdot & -\Delta \\ \cdot & \gamma & -\chi_{\text{ctrl}} \\ -\Delta & -\chi_{\text{ctrl}} & \cdot \end{bmatrix} \begin{pmatrix} y \\ u \\ p \end{pmatrix} = \begin{pmatrix} y_d \\ 0 \\ 0 \end{pmatrix}$$

$$\langle \mathcal{A}(y,u), (z,v) \rangle = a\big((y,u),(z,v)\big)$$

$$\mathcal{A} \colon \mathcal{Y} \times \mathcal{U} \longrightarrow \mathcal{Y}^* \times \mathcal{U}^*$$

$$a\big((y,u),(z,v)\big) = \int_{\Omega_{\text{obs}}} y\, z \,\mathrm{d}x + \gamma \int_{\Omega_{\text{ctrl}}} u\, v \,\mathrm{d}x$$

Optimality System as Saddle-Point Problem

self-adjoint saddle-point system:

$$\begin{bmatrix} \chi_{\text{obs}} & \cdot & -\Delta \\ \cdot & \gamma & -\chi_{\text{ctrl}} \\ -\Delta & -\chi_{\text{ctrl}} & \cdot \end{bmatrix} \begin{pmatrix} y \\ u \\ p \end{pmatrix} = \begin{pmatrix} y_d \\ 0 \\ 0 \end{pmatrix}$$

$$\langle \mathcal{A}\,(y,u)\,,(z,v)\rangle = a\big((y,u),(z,v)\big) \qquad \mathcal{A}\colon \mathcal{Y}\times\mathcal{U} \longrightarrow \mathcal{Y}^* \times \mathcal{U}^*$$

$$a\big((y,u),(z,v)\big) = \int_{\Omega_{\text{obs}}} y\,z \;\mathrm{d}x + \gamma \int_{\Omega_{\text{ctrl}}} u\,v \;\mathrm{d}x$$

$$\langle \mathcal{B}\,(y,u)\,,q\rangle = b\big((y,u),q\big) \qquad \mathcal{B}\colon \mathcal{Y}\times\mathcal{U} \longrightarrow \mathcal{Y}^*$$

$$b\big((y,u),q\big) = \int_{\Omega} \kappa\,\nabla y \cdot \nabla q \;\mathrm{d}x + \int_{\Gamma} \alpha\,y\,q \;\mathrm{d}s - \int_{\Omega_{\text{ctrl}}} u\,q \;\mathrm{d}x$$

How Does This Relate to Section 1?

We have here a problem with a PDE constraint.

$$\begin{aligned} -\kappa\,\Delta y &= \chi_{\text{ctrl}}\,u && \text{in } \Omega \\ \kappa\,\frac{\partial}{\partial n} y + \alpha\, y &= 0 && \text{on } \Gamma = \partial\Omega \end{aligned}$$

For every control $u \in L^2(\Omega_{\text{ctrl}})$, there is a unique state $y \in H^1(\Omega)$ solving the PDE.

When we use the control-to-state map $S\colon L^2(\Omega_{\text{ctrl}}) \to H^1(\Omega) \hookrightarrow L^2(\Omega_{\text{obs}})$, we can eliminate the state variable y:

Minimize $J(y, u)$ s. t. (y, u) satisfies PDE	Minimize $f(u) := J(Su, u)$

Gradient of the Reduced Problem

The reduced objective is of the form

$$\text{Minimize} \quad f(u) := \frac{1}{2}\|Su - y_d\|_H^2 + \frac{\gamma}{2}\|u\|_U^2.$$

The derivative and gradient are given by

$$\begin{aligned} f'(u)\,\delta u &= \big(Su - y_d, S\delta u\big)_H + \gamma\,\big(u, \delta u\big)_U \\ &= \big(S^\circ(Su - y_d), \delta u\big)_U + \gamma\,\big(u, \delta u\big)_U \\ &= \big(S^\circ(Su - y_d) + \gamma\,u, \delta u\big)_U \end{aligned}$$

$$\Rightarrow \quad \nabla f(u) = \; S^\circ(Su - y_d) + \gamma\,u.$$

S° is the Hilbert-space adjoint of S.

Evaluation of the Gradient

$$\nabla f(u) = S^\circ(Su - y_d) + \gamma\, u$$

$y = Su$ means

$$\begin{aligned} -\kappa\,\Delta y &= \chi_{\text{ctrl}}\, u && \text{in } \Omega \\ \kappa\,\frac{\partial}{\partial n} y + \alpha\, y &= 0 && \text{on } \Gamma \end{aligned}$$

$p = -S^\circ(y - y_d)$ means

$$\begin{aligned} -\kappa\,\Delta p &= -\chi_{\text{obs}}\,(y - y_d) && \text{in } \Omega \\ \kappa\,\frac{\partial}{\partial n} p + \alpha\, p &= 0 && \text{on } \Gamma \end{aligned}$$

$$\nabla f(u) = -p + \gamma\, u$$

Each evaluation of $\nabla f(u)$ requires

- one solve of the state (forward) PDE
- one solve of the adjoint (backward) PDE

Discretization by Finite Elements

Suppose that $\{\varphi_j\}$ are $\mathcal{P}_1$ (piecewise linear, globally continuous) finite element basis functions.

objective:

$$\frac{1}{2}\int_{\Omega_{\text{obs}}} (y - y_d)^2 \, \mathrm{d}x + \frac{\gamma}{2}\int_{\Omega_{\text{ctrl}}} u^2 \, \mathrm{d}x = \frac{1}{2}(\mathbf{y} - \mathbf{y}_d)^\mathsf{T} \mathbf{M}_{\text{obs}} (\mathbf{y} - \mathbf{y}_d) + \frac{\gamma}{2}\mathbf{u}^\mathsf{T} \mathbf{M}_{\text{ctrl}} \, \mathbf{u}$$

PDE:

$$\int_\Omega \kappa \nabla y \cdot \nabla \varphi_i \, \mathrm{d}x + \int_\Gamma \alpha \, y \, \varphi_i \, \mathrm{d}s - \int_{\Omega_{\text{ctrl}}} u \, \varphi_i \, \mathrm{d}x = 0 \quad \text{for all } i$$

$$\Leftrightarrow \quad \mathbf{K}\,\mathbf{y} - \mathbf{M}_{\text{ctrl}} \, \mathbf{u} = \mathbf{0} \quad \Leftrightarrow \quad \mathbf{S}\,\mathbf{u} = \mathbf{K}^{-1}\mathbf{M}_{\text{ctrl}} \, \mathbf{u}$$

adjoint PDE:

$$\int_\Omega \kappa \nabla p \cdot \nabla \varphi_i \, \mathrm{d}x + \int_\Gamma \alpha \, p \, \varphi_i \, \mathrm{d}s + \int_{\Omega_{\text{obs}}} (y - y_d) \, \varphi_i \, \mathrm{d}x = 0 \quad \text{for all } i$$

$$\Leftrightarrow \quad \mathbf{K}^\mathsf{T}\mathbf{p} + \mathbf{M}_{\text{obs}} (\mathbf{y} - \mathbf{y}_d) = \mathbf{0} \quad \Leftrightarrow \quad \mathbf{S}^\circ \, \mathbf{z} = \mathbf{K}^{-\mathsf{T}}\mathbf{M}_{\text{obs}} \, \mathbf{z}$$

Representation of the Derivative and Gradient

$$\mathbf{f}(\mathbf{u}) = \frac{1}{2}(\mathbf{K}^{-1}\mathbf{M}_{\mathsf{ctrl}}\,\mathbf{u} - \mathbf{y}_d)^{\mathsf{T}}\mathbf{M}_{\mathsf{obs}}\,(\mathbf{K}^{-1}\mathbf{M}_{\mathsf{ctrl}}\,\mathbf{u} - \mathbf{y}_d) + \frac{\gamma}{2}\mathbf{u}^{\mathsf{T}}\mathbf{M}_{\mathsf{ctrl}}\,\mathbf{u}$$

$$\begin{aligned} \mathbf{f}'(\mathbf{u})\,\delta\mathbf{u} &= (\mathbf{K}^{-1}\mathbf{M}_{\mathsf{ctrl}}\,\mathbf{u} - \mathbf{y}_d)^{\mathsf{T}}\mathbf{M}_{\mathsf{obs}}\,\mathbf{K}^{-1}\mathbf{M}_{\mathsf{ctrl}}\,\delta\mathbf{u} + \gamma\,\mathbf{u}^{\mathsf{T}}\mathbf{M}_{\mathsf{ctrl}}\,\delta\mathbf{u} \\ &= (\nabla\mathbf{f}(\mathbf{u}), \delta\mathbf{u})_{\mathbf{M}_{\mathsf{ctrl}}} \end{aligned}$$

$$\begin{aligned} \nabla\mathbf{f}(\mathbf{u}) &= \underbrace{\mathbf{K}^{-\mathsf{T}}\mathbf{M}_{\mathsf{obs}}}_{=\mathbf{S}^\circ}(\underbrace{\mathbf{K}^{-1}\mathbf{M}_{\mathsf{ctrl}}}_{=\mathbf{S}}\,\mathbf{u} - \mathbf{y}_d) + \gamma\,\mathbf{u} \\ &= \mathbf{S}^\circ(\mathbf{S}\,\mathbf{u} - \mathbf{y}_d) + \gamma\,\mathbf{u} \end{aligned}$$

- This agrees with the formulation of $\nabla f(u)$ in the continuous setting.
- Interestingly, the evaluation of the gradient $\nabla\mathbf{f}(\mathbf{u})$ from the derivative $\mathbf{f}'(\mathbf{u})$ does not require the solution of a linear system with $\mathbf{M}_{\mathsf{ctrl}}$ here.

Mesh-Independence of Gradient Descent

- The objective is quadratic:

$$f(u) = \frac{1}{2}\|Su - y_d\|_H^2 + \frac{\gamma}{2}\|u\|_U^2 = \frac{1}{2}(Au, u)_U - (b, u)_U + c$$

where $A = S^\circ S + \gamma\,\mathrm{id}$ is U-selfadjoint and positive definite

- The Q-linear contraction rate can be estimated by

$$\kappa := \underbrace{\frac{\sup\big\{(Au, u)_U \mid \|u\|_U = 1\big\}}{\inf\big\{(Au, u)_U \mid \|u\|_U = 1\big\}}}_{\text{on } U} \ge \underbrace{\frac{\sup\big\{(\mathbf{A}\mathbf{u}, \mathbf{u})_U \mid \|\mathbf{u}\|_U = 1\big\}}{\inf\big\{(\mathbf{A}\mathbf{u}, \mathbf{u})_U \mid \|\mathbf{u}\|_U = 1\big\}}}_{\text{on any subspace } \mathbf{U} \subset U}$$

- An implementation of $\mathbf{f}$, $\nabla \mathbf{f}$ etc. must be matrix-free.
- To obtain an optimal implementation, we need to use optimal solvers of the state and adjoint equations.

Conjugate Gradient Method

A much better algorithm for the minimization of

$$\frac{1}{2}(Au, u)_U - (b, u)_U + c$$

in a Hilbert space U with self-adjoint, positive definite operator A is the conjugate gradient method.

- Every new search direction is A-orthogonalized w.r.t. all previous ones yet with short recursion.
- The dominant cost per iteration is the same as in gradient descent.
- The inner product in U takes the same role as preconditioner as in gradient descent.

[Günnel, Herzog, Sachs, 2014]

Conjugate Gradient Method

Input: initial guess $u_0 \in U$, tolerance ε
Output: approximate minimizer of $\frac{1}{2}(Au, u)_U - (b, u)_U + c$

Set $r_0 := \nabla f(u_0) = A\, u_0 - b \in U$
Set $d_0 := -r_0$ // initial search direction
Set $\delta_0 := \|r_0\|_U^2$
Set $n := 0$
while $n < n_{\text{max}}$ und $\delta_n > \varepsilon^2 \delta_0$ **do**
  Set $q_n := A\, d_n$
  Set $\alpha_n := \frac{\delta_n}{(d_n, q_n)_U}$ // exact line search
  Set $u_{n+1} := u_n + \alpha_n\, d_n$ // update solution
  Set $r_{n+1} := r_n + \alpha_n\, q_n$ // update gradient/residual
  Set $\delta_{n+1} := \|r_{n+1}\|_U^2$
  Set $\beta_{n+1} := \frac{\delta_{n+1}}{\delta_n}$
  Set $d_{n+1} := -r_{n+1} + \beta_{n+1}\, d_n$ // A-orthogonalize gradient
  Set $n := n + 1$
end while
return u_n

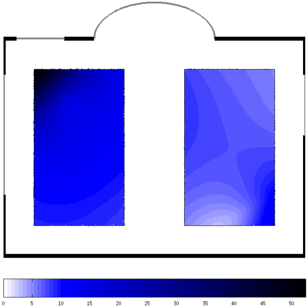
Table of Contents

Model Problem with Control Constraints

$$\text{Minimize} \quad \frac{1}{2}\int_{\Omega_{\text{obs}}} \big(y(x) - y_d(x)\big)^2 \,\mathrm{d}x + \frac{\gamma}{2}\int_{\Omega_{\text{ctrl}}} u(x)^2 \,\mathrm{d}x, \qquad \gamma > 0$$

$$\text{s.t.} \quad \begin{cases} -\kappa\,\Delta y(x) = \chi_{\text{ctrl}}\, u(x) & \text{in } \Omega \\ \kappa\,\frac{\partial}{\partial n} y(x) + \alpha(x)\, y(x) = 0 & \text{on } \Gamma = \partial\Omega \end{cases}$$

control



Model Problem with Control Constraints

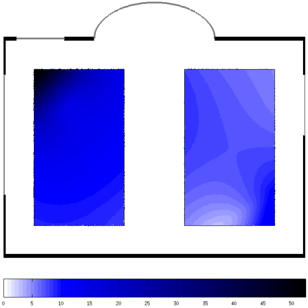
$$\text{Minimize} \quad \frac{1}{2}\int_{\Omega_{\text{obs}}} \big(y(x) - y_d(x)\big)^2 \, \mathrm{d}x + \frac{\gamma}{2}\int_{\Omega_{\text{ctrl}}} u(x)^2 \, \mathrm{d}x, \quad \gamma > 0$$

$$\text{s.t.} \quad \begin{cases} -\kappa\, \Delta y(x) = \chi_{\text{ctrl}}\, u(x) & \text{in } \Omega \\ \kappa \frac{\partial}{\partial n} y(x) + \alpha(x)\, y(x) = 0 & \text{on } \Gamma = \partial\Omega \end{cases}$$

$$\text{and} \quad u_a(x) \le u(x) \le u_b(x) \quad \text{in } \Omega_{\text{ctrl}}$$

Such constraints are usually motivated by technological limitations.

control



- upper bound u_b on the heating power
- lower bound, e.g., $u_a = 0$ (no cooling)

Optimality Conditions

$$\mathcal{L}(y, u, p, \mu) = \frac{1}{2} \int_{\Omega_{\text{obs}}} (y - y_d)^2 \, \mathrm{d}x + \frac{\gamma}{2} \int_{\Omega_{\text{ctrl}}} u^2 \, \mathrm{d}x \qquad \leftarrow \text{objective}$$

$$+ \int_{\Omega} \kappa \, \nabla y \cdot \nabla p \, \mathrm{d}x + \int_{\Gamma} \alpha \, y \, p \, \mathrm{d}s - \int_{\Omega_{\text{ctrl}}} u \, p \, \mathrm{d}x \qquad \leftarrow \text{PDE}$$

adjoint PDE
$$\begin{cases} -\kappa \, \Delta p = -\chi_{\text{obs}} \, (y - y_d) & \text{in } \Omega \\ \kappa \frac{\partial}{\partial n} p + \alpha \, p = 0 & \text{on } \Gamma \end{cases}$$

gradient eq.
$$\{ \, \gamma \, u - p = 0 \quad \text{in } \Omega_{\text{ctrl}}$$

state PDE
$$\begin{cases} -\kappa \, \Delta y = \chi_{\text{ctrl}} \, u & \text{in } \Omega \\ \kappa \frac{\partial}{\partial n} y + \alpha \, y = 0 & \text{on } \Gamma \end{cases}$$

Optimality Conditions

$$\mathcal{L}(y,u,p,\mu) = \frac{1}{2}\int_{\Omega_{\text{obs}}} (y - y_d)^2 \, \mathrm{d}x + \frac{\gamma}{2}\int_{\Omega_{\text{ctrl}}} u^2 \, \mathrm{d}x \qquad \leftarrow \text{objective}$$

$$+ \int_\Omega \kappa \, \nabla y \cdot \nabla p \, \mathrm{d}x + \int_\Gamma \alpha \, y \, p \, \mathrm{d}s - \int_{\Omega_{\text{ctrl}}} u \, p \, \mathrm{d}x \qquad \leftarrow \text{PDE}$$

$$+ \int_{\Omega_{\text{ctrl}}} \mu^+ \, (u - u_b) \, \mathrm{d}x + \int_{\Omega_{\text{ctrl}}} \mu^- \, (u_a - u) \, \mathrm{d}x \qquad \leftarrow \text{constraints}$$

adjoint PDE
$$\begin{cases} -\kappa \, \Delta p = -\chi_{\text{obs}} \, (y - y_d) & \text{in } \Omega \\ \kappa \, \frac{\partial}{\partial n} p + \alpha \, p = 0 & \text{on } \Gamma \end{cases}$$

gradient eq.
$$\{ \; \gamma \, u - p + \mu^+ - \mu^- = 0 \quad \text{in } \Omega_{\text{ctrl}}$$

state PDE
$$\begin{cases} -\kappa \, \Delta y = \chi_{\text{ctrl}} \, u & \text{in } \Omega \\ \kappa \, \frac{\partial}{\partial n} y + \alpha \, y = 0 & \text{on } \Gamma \end{cases}$$

complementarity
$$\begin{cases} \mu^+ \geq 0, \quad u - u_b \leq 0, \quad \mu^+ \, (u - u_b) = 0 \text{ in } \Omega_{\text{ctrl}} \\ \mu^- \geq 0, \quad u_a - u \leq 0, \quad \mu^- \, (u_a - u) = 0 \text{ in } \Omega_{\text{ctrl}} \end{cases}$$

Optimality Conditions

$$
\begin{aligned}
\mathcal{L}(y,u,p,\mu) = {} & \frac{1}{2}\int_{\Omega_{\text{obs}}} (y - y_d)^2 \, \mathrm{d}x + \frac{\gamma}{2}\int_{\Omega_{\text{ctrl}}} u^2 \, \mathrm{d}x && \leftarrow \text{objective} \\
& + \int_{\Omega} \kappa \, \nabla y \cdot \nabla p \, \mathrm{d}x + \int_{\Gamma} \alpha \, y \, p \, \mathrm{d}s - \int_{\Omega_{\text{ctrl}}} u \, p \, \mathrm{d}x && \leftarrow \text{PDE} \\
& + \int_{\Omega_{\text{ctrl}}} \mu^+ \, (u - u_b) \, \mathrm{d}x + \int_{\Omega_{\text{ctrl}}} \mu^- \, (u_a - u) \, \mathrm{d}x && \leftarrow \text{constraints}
\end{aligned}
$$

adjoint PDE
$$
\begin{cases} -\kappa \, \Delta p = -\chi_{\text{obs}} \, (y - y_d) & \text{in } \Omega \\ \kappa \, \frac{\partial}{\partial n} p + \alpha \, p = 0 & \text{on } \Gamma \end{cases}
$$

gradient eq.
$$
\{ \, \gamma \, u - p + \mu = 0 \quad \text{in } \Omega_{\text{ctrl}}
$$

state PDE
$$
\begin{cases} -\kappa \, \Delta y = \chi_{\text{ctrl}} \, u & \text{in } \Omega \\ \kappa \, \frac{\partial}{\partial n} y + \alpha \, y = 0 & \text{on } \Gamma \end{cases}
$$

complementarity
$$
\{ \, \mu = \max\{0, \mu + \quad (u - u_b)\} + \min\{0, \mu + \quad (u - u_a)\}
$$

Optimality Conditions

$$\mathcal{L}(y,u,p,\mu) = \frac{1}{2}\int_{\Omega_{\text{obs}}} (y - y_d)^2 \,\mathrm{d}x + \frac{\gamma}{2}\int_{\Omega_{\text{ctrl}}} u^2 \,\mathrm{d}x \qquad \leftarrow \text{objective}$$

$$+ \int_{\Omega} \kappa \nabla y \cdot \nabla p \,\mathrm{d}x + \int_{\Gamma} \alpha\, y\, p \,\mathrm{d}s - \int_{\Omega_{\text{ctrl}}} u\, p \,\mathrm{d}x \qquad \leftarrow \text{PDE}$$

$$+ \int_{\Omega_{\text{ctrl}}} \mu^+ (u - u_b) \,\mathrm{d}x + \int_{\Omega_{\text{ctrl}}} \mu^- (u_a - u) \,\mathrm{d}x \qquad \leftarrow \text{constraints}$$

adjoint PDE
$$\begin{cases} -\kappa\,\Delta p = -\chi_{\text{obs}}\,(y - y_d) & \text{in } \Omega \\ \kappa \frac{\partial}{\partial n} p + \alpha\, p = 0 & \text{on } \Gamma \end{cases}$$

gradient eq.
$$\left\{ \gamma\, u - p + \mu = 0 \quad \text{in } \Omega_{\text{ctrl}} \right.$$

state PDE
$$\begin{cases} -\kappa\,\Delta y = \chi_{\text{ctrl}}\, u & \text{in } \Omega \\ \kappa \frac{\partial}{\partial n} y + \alpha\, y = 0 & \text{on } \Gamma \end{cases}$$

complementarity
$$\left\{ \mu = \max\{0, \mu + c\,(u - u_b)\} + \min\{0, \mu + c\,(u - u_a)\} \right.$$

Solution by Primal-Dual Active Set Method

The pointwise linearization of the complementarity function

$$\mu = \max\{0, \mu + c\,(u - u_b)\} + \min\{0, \mu + c\,(u - u_a)\},$$

$$\begin{aligned} c\,(u + \delta u - u_b) &= 0 \quad \text{on } \mathcal{A}^+ := \{x \in \Omega_{\text{ctrl}} : \mu + c\,(u - u_b) > 0\}, \\ c\,(u + \delta u - u_a) &= 0 \quad \text{on } \mathcal{A}^- := \{x \in \Omega_{\text{ctrl}} : \mu + c\,(u - u_a) < 0\}, \\ \mu + \delta\mu &= 0 \quad \text{on } \mathcal{I} := \Omega_{\text{ctrl}} \setminus \big(\mathcal{A}^+ \cup \mathcal{A}^-\big) \end{aligned}$$

is a generalized differentiation concept (even in function space).

- It leads to a locally superlinearly convergent active set method which generalizes Newton's method.
- In each step, we need to solve a problem with equality constraints only.

[Chen, Nashed, Qi, 2000; Hintermüller, Ito, Kunisch, 2002; Ulbrich, 2003]

UNIVERSITÄT HEIDELBERG ZUKUNFT SEIT 1386

Algorithmic Considerations

Full-Space Approach

$$\begin{bmatrix} \chi_{\text{obs}} & \cdot & -\Delta & \cdot \\ \cdot & \gamma & -\chi_{\text{ctrl}} & \chi_{\mathcal{A}} \\ -\Delta & -\chi_{\text{ctrl}} & \cdot & \cdot \\ \cdot & \chi_{\mathcal{A}} & \cdot & -\frac{1}{c}\chi_{\mathcal{I}} \end{bmatrix} \begin{pmatrix} \delta y \\ \delta u \\ \delta p \\ \delta \mu \end{pmatrix} = \cdots$$

- large-scale, self-adjoint, indefinite saddle-point system
- solution by preconditioned Minres, a Krylov subspace method

Reduced Approach

$$\chi_{\mathcal{I}} \big(S^\circ S + \gamma \, \text{id} \big) \, \chi_{\mathcal{I}} \, \delta u = \cdots$$

- S = forward PDE solver, S° = adjoint PDE solver
- smaller scale, self-adjoint, positive definite system (id + compact)
- solution by subspace CG, superlinear in function space

[Herzog, Kunisch (2010); Herzog, Sachs (SIMAX, 2010 & SINUM, 2015)]

Numerical Experiment

Setup of the eperiment:

- reduced primal-dual active set method
- matrix-free implementation of $S^\circ S + \gamma\,\mathrm{id}$
- inexact solves by subspace CG method
- discretization by $\mathcal{P}_1$ finite elements
- multi-level (nested) approach on a sequence of refined meshes

Caution:

$$\mu = \max\{0, \qquad \mu + c\,(u - u_b)\} + \min\{0, \qquad \mu + c\,(u - u_a)\}$$

Numerical Experiment

Setup of the eperiment:

- reduced primal-dual active set method
- matrix-free implementation of $S^\circ S + \gamma\,\mathrm{id}$
- inexact solves by subspace CG method
- discretization by $\mathcal{P}_1$ finite elements
- multi-level (nested) approach on a sequence of refined meshes

Caution:

$$\mathbf{M}_{\mathsf{ctrl}}\boldsymbol{\mu} = \max\{0, \mathbf{M}_{\mathsf{ctrl}}\boldsymbol{\mu} + c\,(\mathbf{u} - \mathbf{u}_b)\} + \min\{0, \mathbf{M}_{\mathsf{ctrl}}\boldsymbol{\mu} + c\,(\mathbf{u} - \mathbf{u}_a)\}$$

- lumped control space mass matrix: pointwise $\leftrightarrow$ componentwise

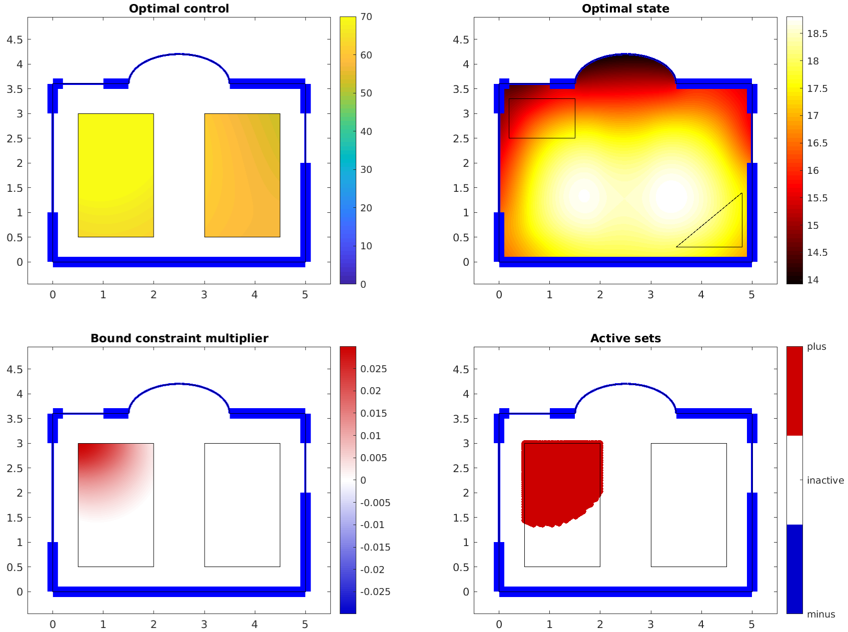
Numerical Experiment

Setup of the eperiment:

- reduced primal-dual active set method
- matrix-free implementation of $S^\circ S + \gamma\,\mathrm{id}$
- inexact solves by subspace CG method
- discretization by $\mathcal{P}_1$ finite elements

dofs state	dofs ctrl	dofs total	active	PDAS/CG iter		CPU
226	95	642	29.47%	7	1.7	0.1 s
849	324	2346	29.63%	4	1.8	0.1 s
3289	1190	8958	30.00%	3	1.7	0.2 s
12 945	4554	34 998	30.15%	2	1.5	0.4 s
51 361	17 810	138 342	30.20%	2	1.5	1.3 s
204 609	70 434	550 086	30.23%	2	1.5	5.9 s

Numerical Results

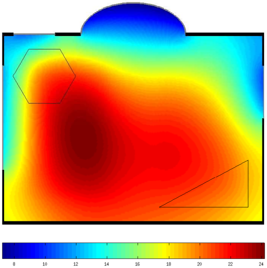


Model Problem with State Constraints

Minimize $\quad \displaystyle \frac{1}{2}\int_{\Omega_{\text{obs}}} \big(y(x) - y_d(x)\big)^2 \,\mathrm{d}x + \frac{\gamma}{2}\int_{\Omega_{\text{ctrl}}} u(x)^2 \,\mathrm{d}x, \quad \gamma > 0$

s.t. $\quad \begin{cases} -\kappa\,\Delta y(x) = \chi_{\text{ctrl}}\,u(x) & \text{in } \Omega \\ \kappa\,\frac{\partial}{\partial n} y(x) + \alpha(x)\,y(x) = 0 & \text{on } \Gamma = \partial\Omega \end{cases}$

state



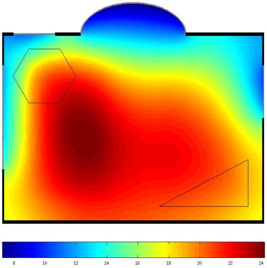
Model Problem with State Constraints

Minimize $\displaystyle \frac{1}{2}\int_{\Omega_{\text{obs}}} \big(y(x) - y_d(x)\big)^2 \, \mathrm{d}x + \frac{\gamma}{2}\int_{\Omega_{\text{ctrl}}} u(x)^2 \, \mathrm{d}x, \quad \gamma > 0$

s.t. $$\begin{cases} -\kappa\, \Delta y(x) = \chi_{\text{ctrl}}\, u(x) & \text{in } \Omega \\ \kappa \frac{\partial}{\partial n} y(x) + \alpha(x)\, y(x) = 0 & \text{on } \Gamma = \partial\Omega \end{cases}$$

and $y_a(x) \le y(x) \le y_b(x)$ in $\overline{\Omega}$

state



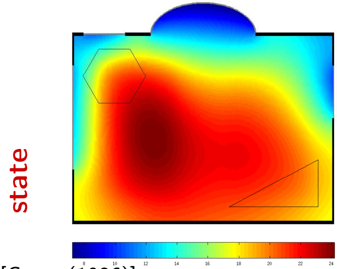
- Such constraints are motivated by, e.g., safety limitations.

Model Problem with State Constraints

$$\text{Minimize} \quad \frac{1}{2}\int_{\Omega_{\text{obs}}} \big(y(x) - y_d(x)\big)^2 \, \mathrm{d}x + \frac{\gamma}{2}\int_{\Omega_{\text{ctrl}}} u(x)^2 \, \mathrm{d}x, \quad \gamma > 0$$

$$\text{s.t.} \quad \begin{cases} -\kappa\,\Delta y(x) = \chi_{\text{ctrl}}\, u(x) & \text{in } \Omega \\ \kappa \frac{\partial}{\partial n} y(x) + \alpha(x)\, y(x) = 0 & \text{on } \Gamma = \partial\Omega \end{cases}$$

$$\text{and} \quad y_a(x) \le y(x) \le y_b(x) \quad \text{in } \overline{\Omega}$$



[Casas (1986)]

- Such constraints are motivated by, e.g., safety limitations.
- Analysis and numerical solution way more challenging than for control / sparsity constraints.
- Lagrange multiplier is a measure $\mu \in C(\overline{\Omega})^*$, not a function.

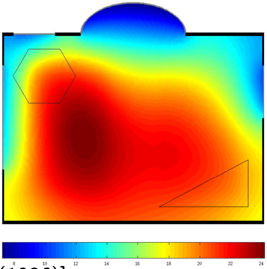
Model Problem with State Constraints

$$\text{Minimize} \quad \frac{1}{2}\int_{\Omega_{\text{obs}}} \big(y(x) - y_d(x)\big)^2 \, \mathrm{d}x + \frac{\gamma}{2}\int_{\Omega_{\text{ctrl}}} u(x)^2 \, \mathrm{d}x, \quad \gamma > 0$$

$$+ \frac{1}{2\,\varepsilon}\int_{\overline{\Omega}} \max\left\{0, y(x) - y_b(x)\right\}^2 + \min\left\{0, y(x) - y_a(x)\right\}^2 \, \mathrm{d}x$$

$$\text{s.t.} \quad \begin{cases} -\kappa\,\Delta y(x) = \chi_{\text{ctrl}}\, u(x) & \text{in } \Omega \\ \kappa\,\frac{\partial}{\partial n} y(x) + \alpha(x)\, y(x) = 0 & \text{on } \Gamma = \partial\Omega \end{cases}$$

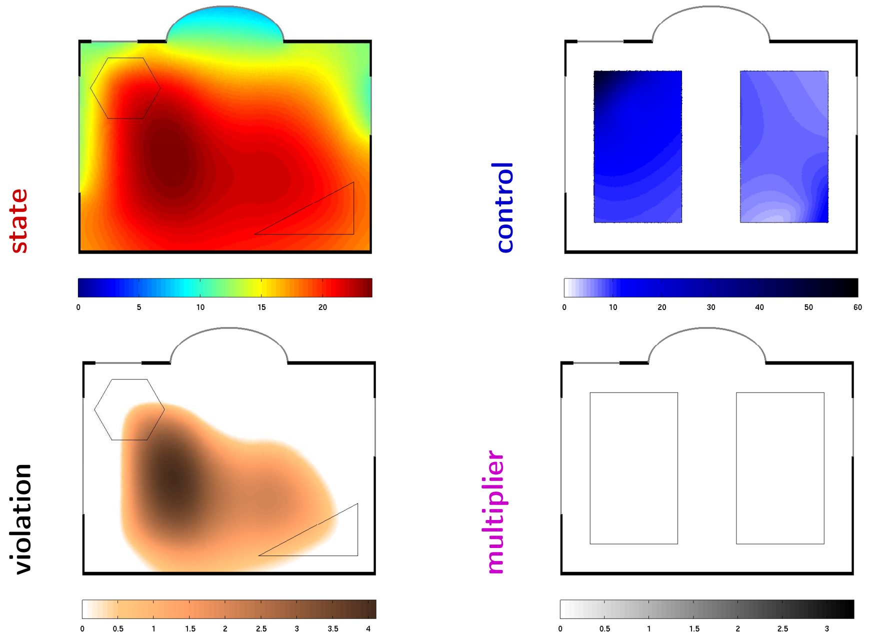
state



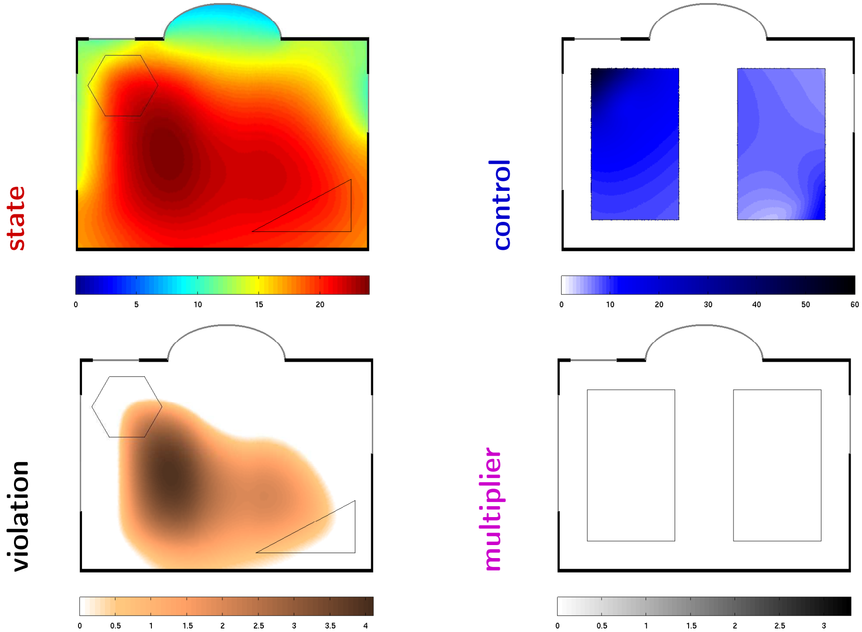
[Casas (1986)]

- Such constraints are motivated by, e.g., safety limitations.
- Analysis and numerical solution way more challenging than for control / sparsity constraints.
- Lagrange multiplier is a measure $\mu \in C(\overline{\Omega})^*$, not a function.

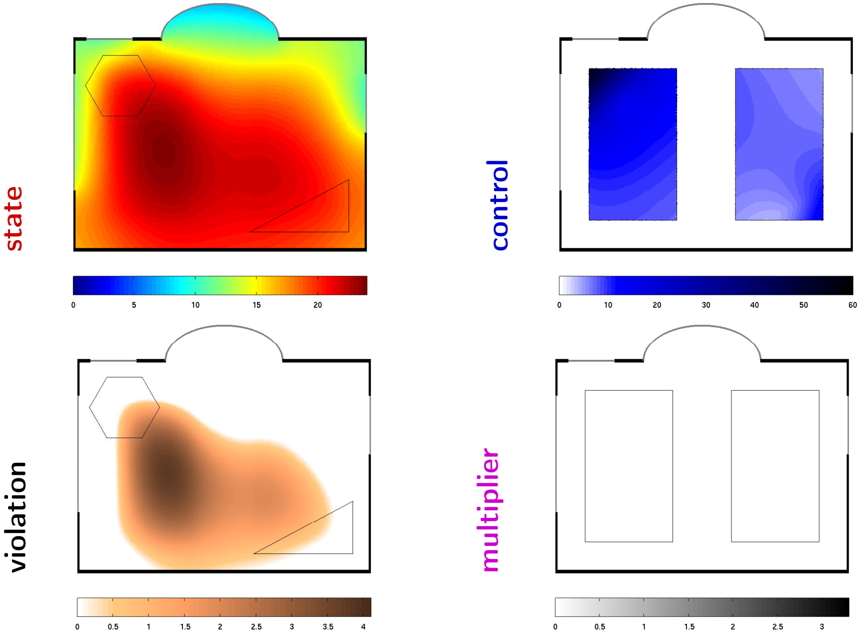
Numerical Example with $\varepsilon = 1000$



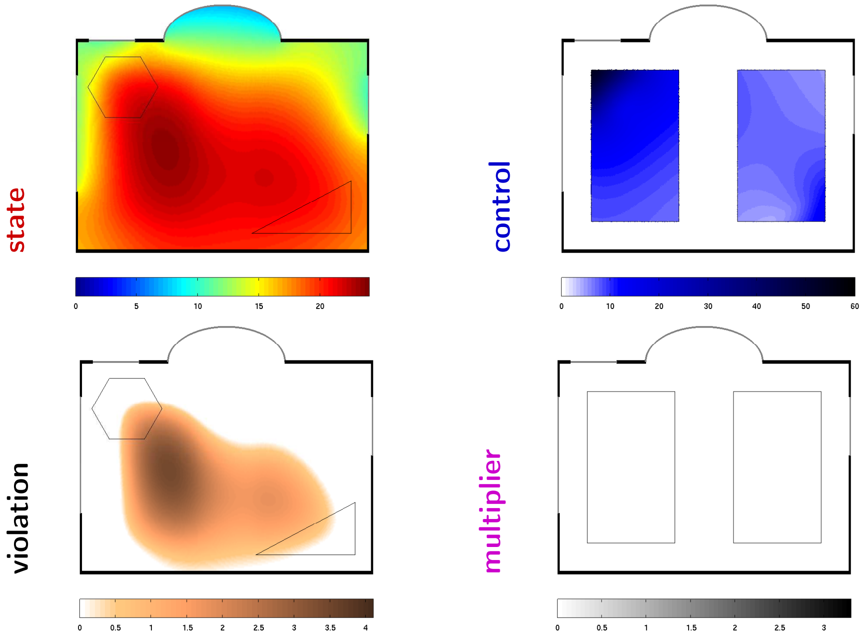
Numerical Example with $\varepsilon = 464.2$



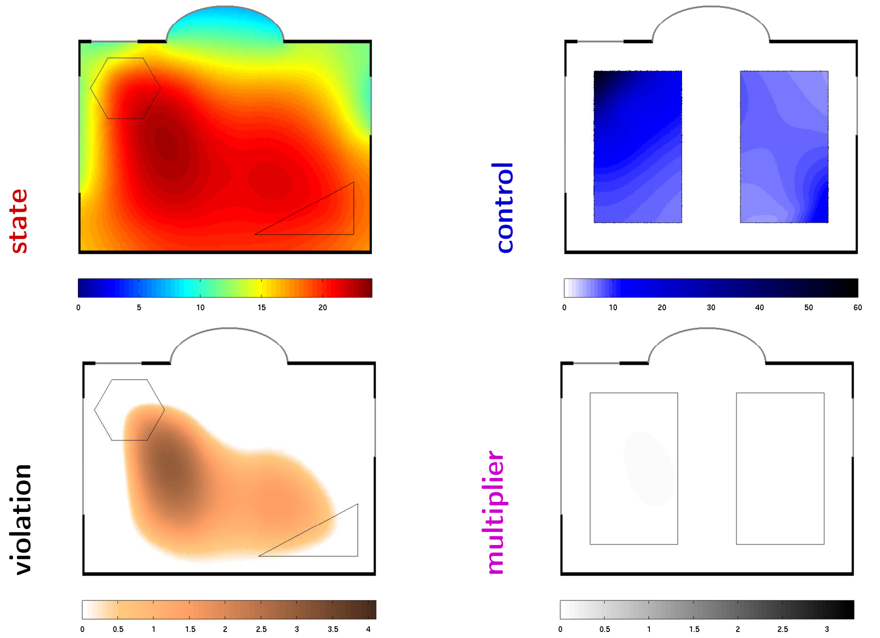
Numerical Example with $\varepsilon = 215.4$



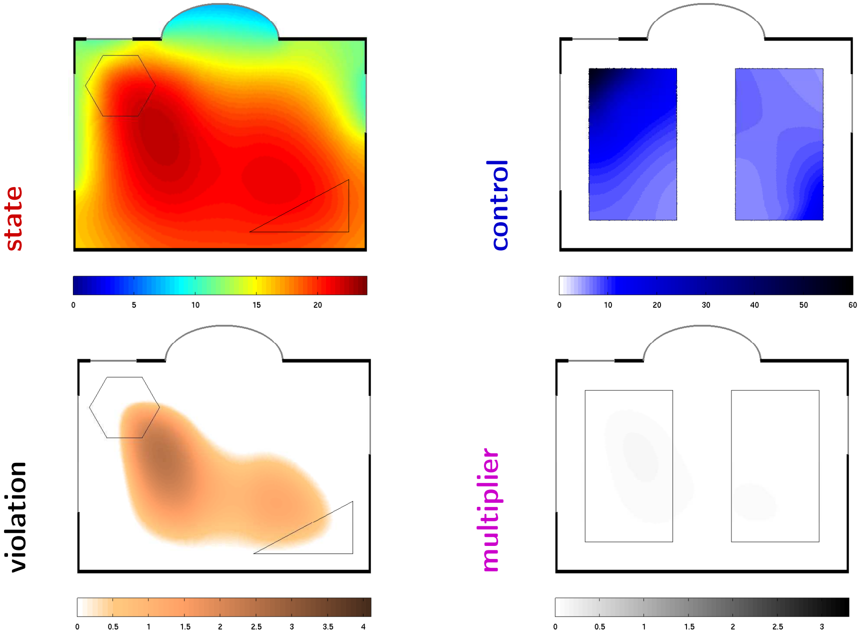
Numerical Example with $\varepsilon = 100.0$



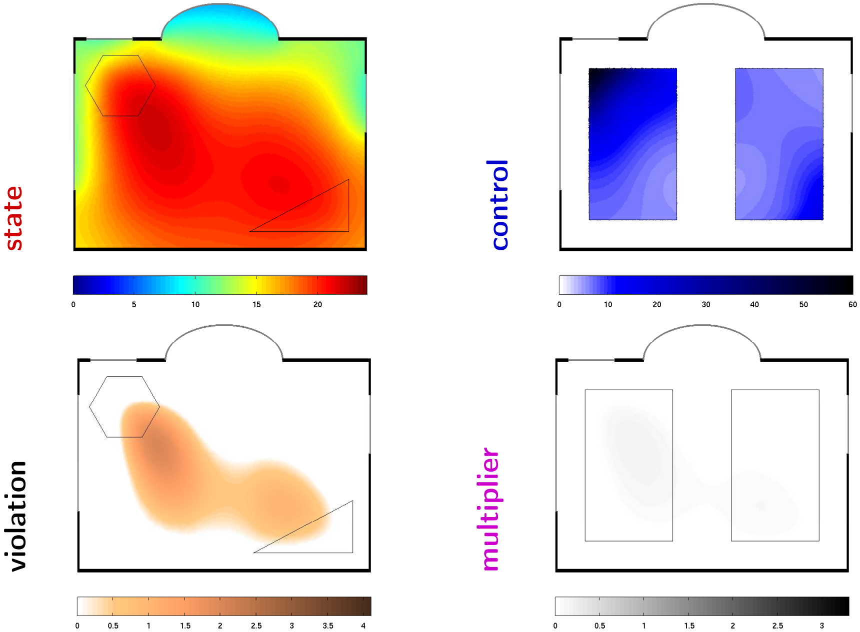
Numerical Example with $\varepsilon = 46.42$



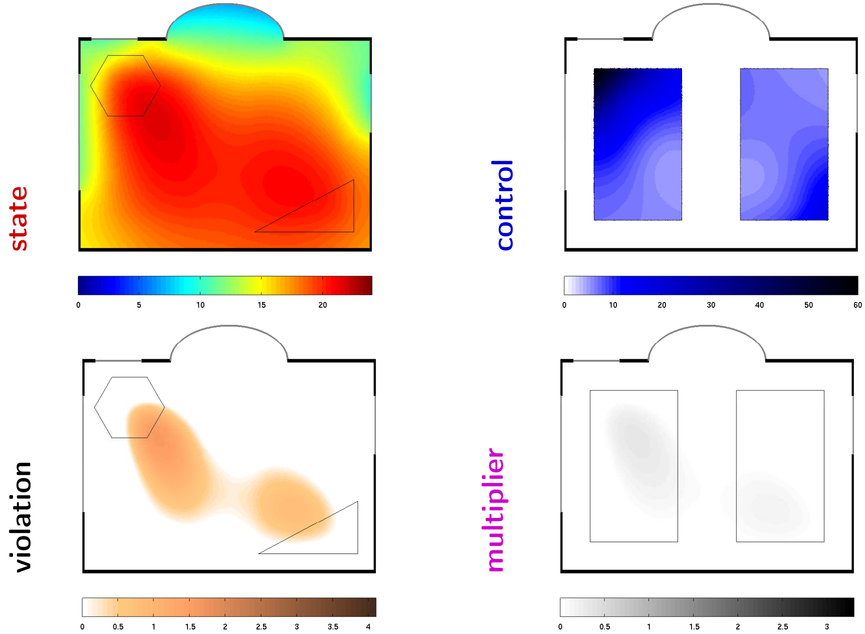
Numerical Example with $\varepsilon = 21.54$



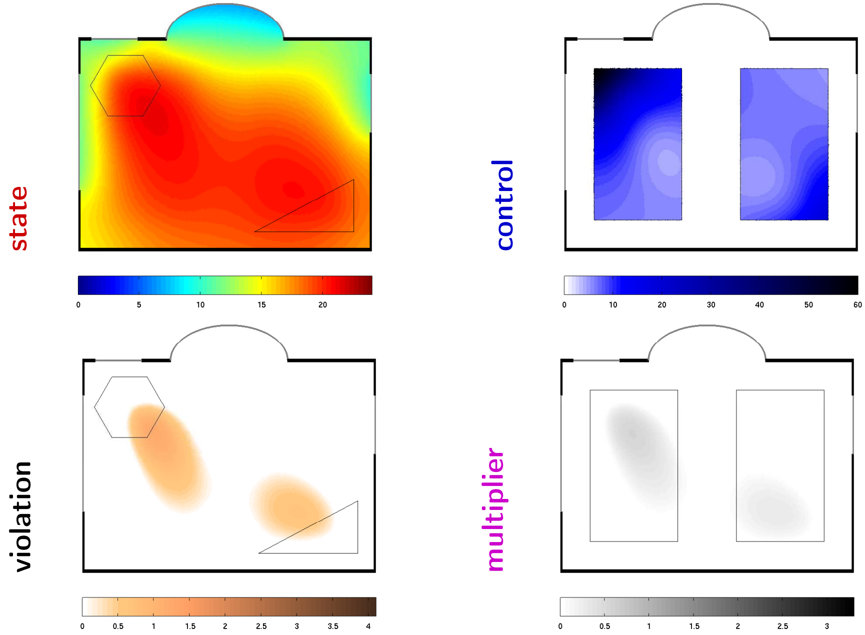
Numerical Example with $\varepsilon = 10.00$



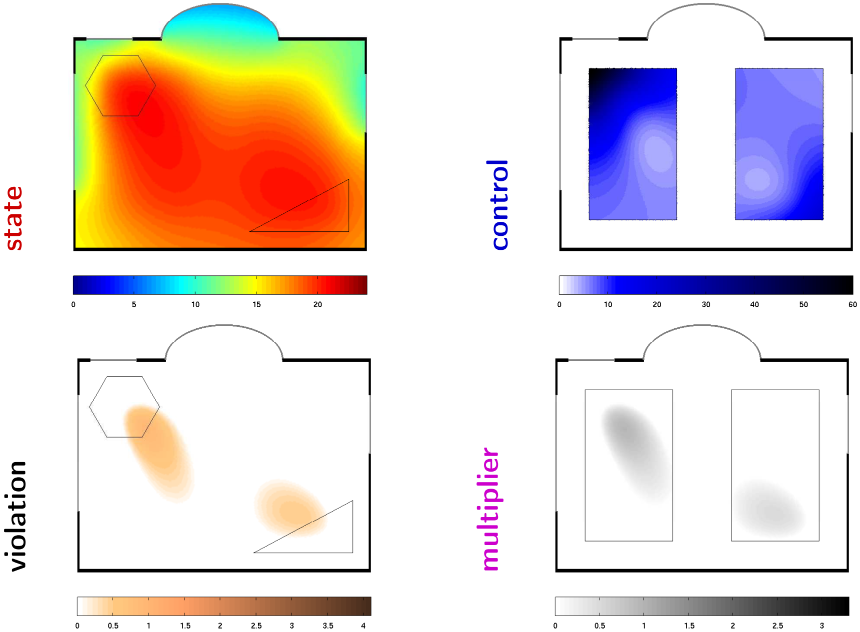
Numerical Example with $\varepsilon = 4.642$



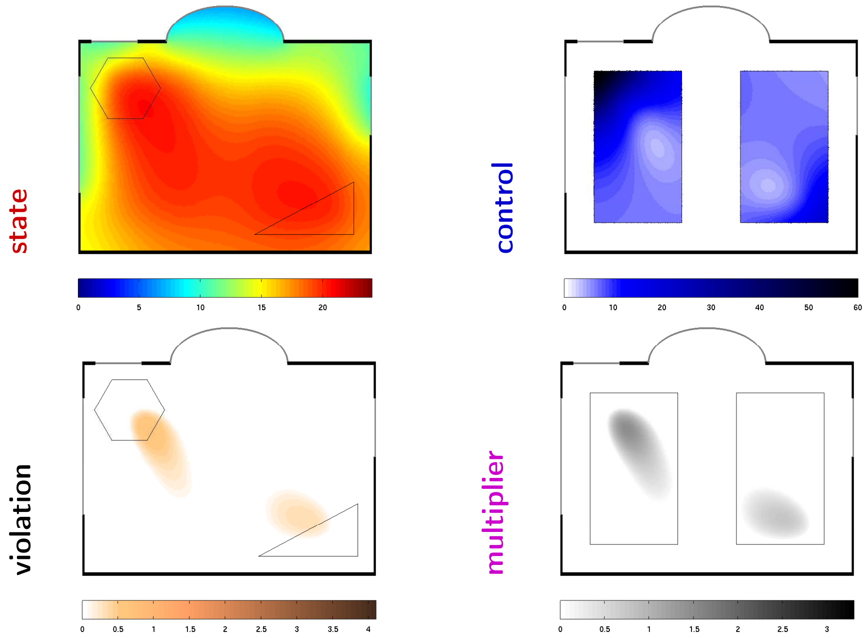
Numerical Example with $\varepsilon = 2.154$



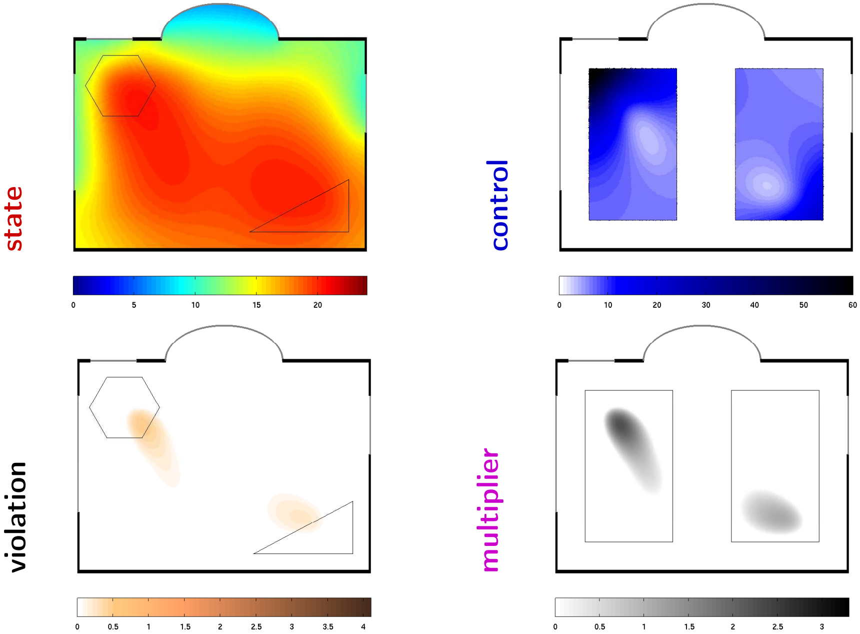
Numerical Example with $\varepsilon = 1.000$



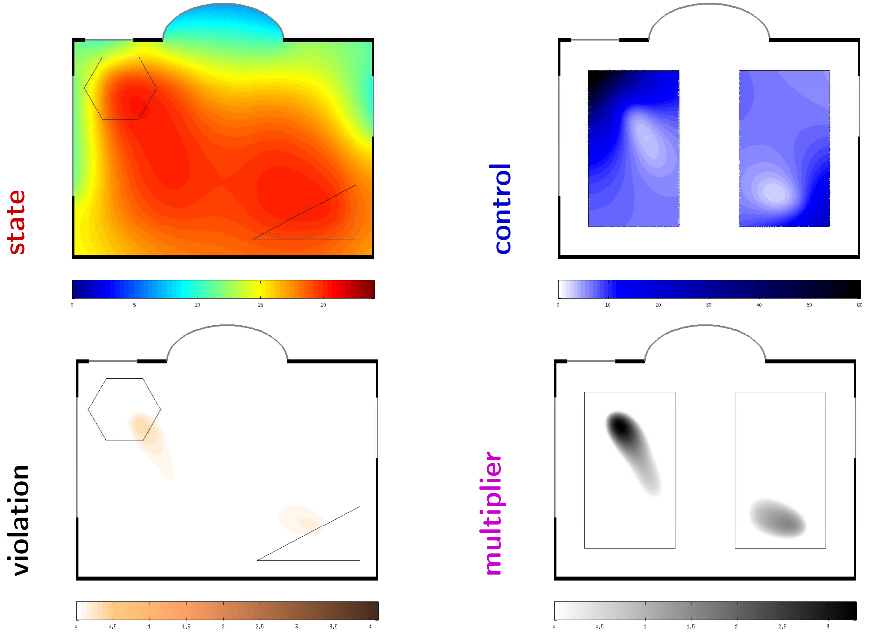
Numerical Example with $\varepsilon = 0.464$



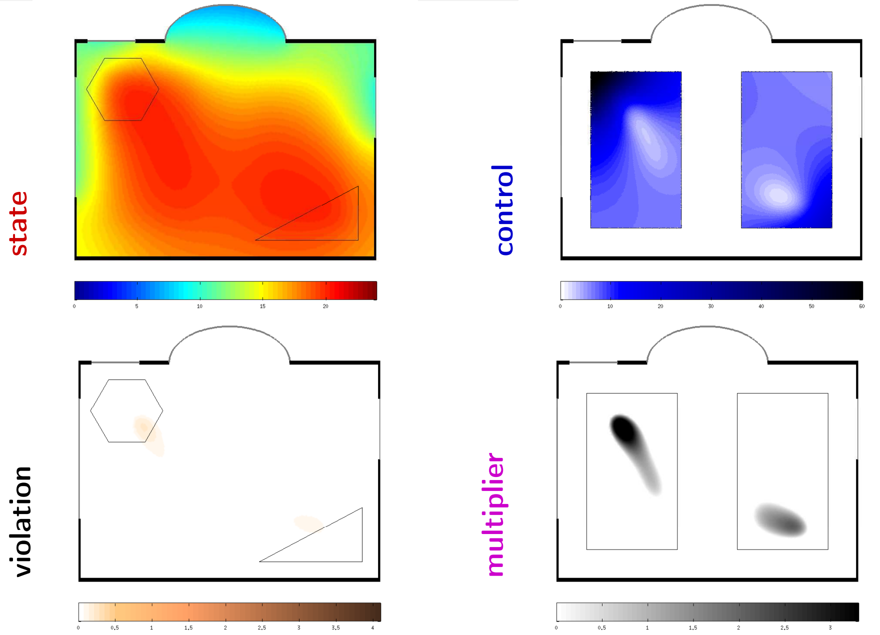
Numerical Example with $\varepsilon = 0.215$



Numerical Example with $\varepsilon = 0.100$

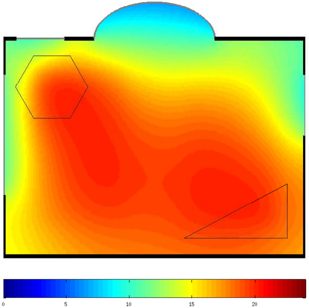


Numerical Example with $\varepsilon = 0.046$

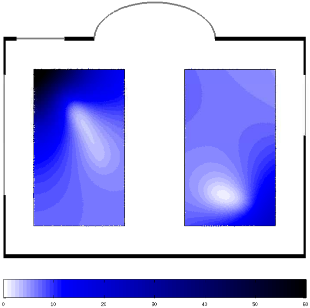


Numerical Example with $\varepsilon = 0.022$

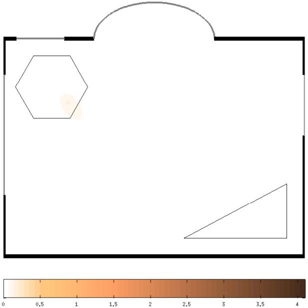
state



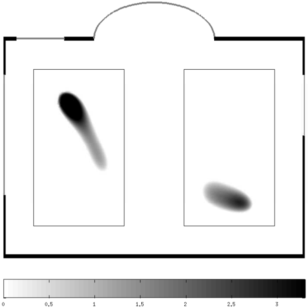
control



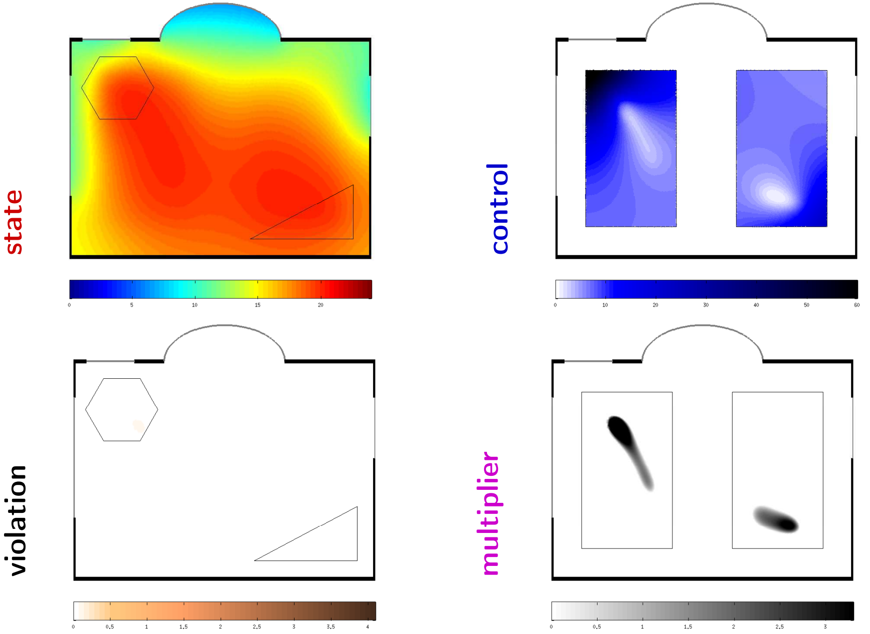
violation



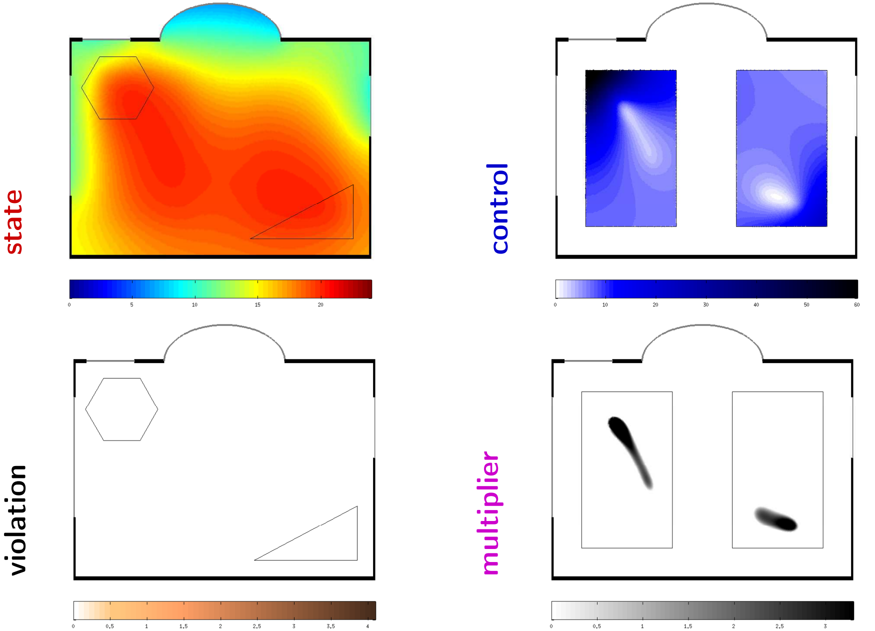
multiplier



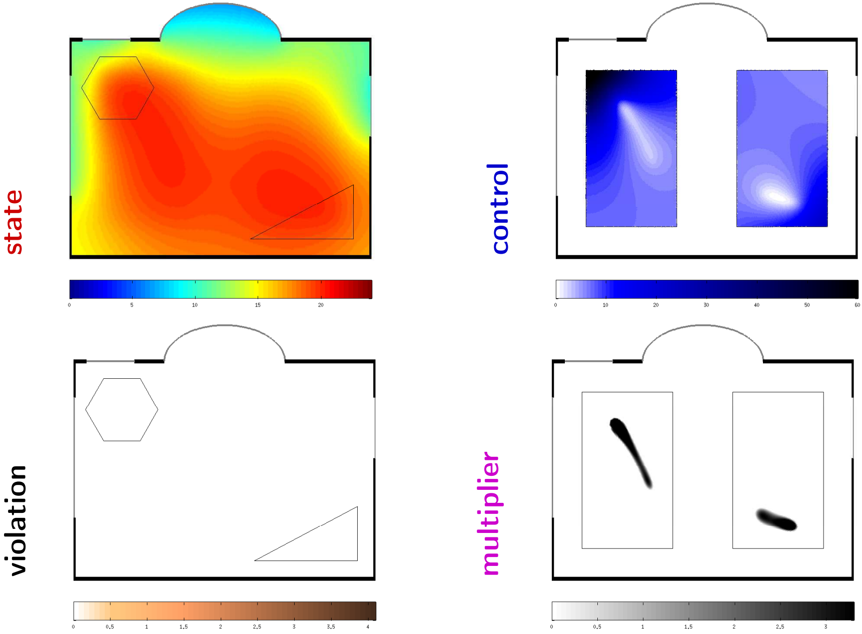
Numerical Example with $\varepsilon = 0.010$



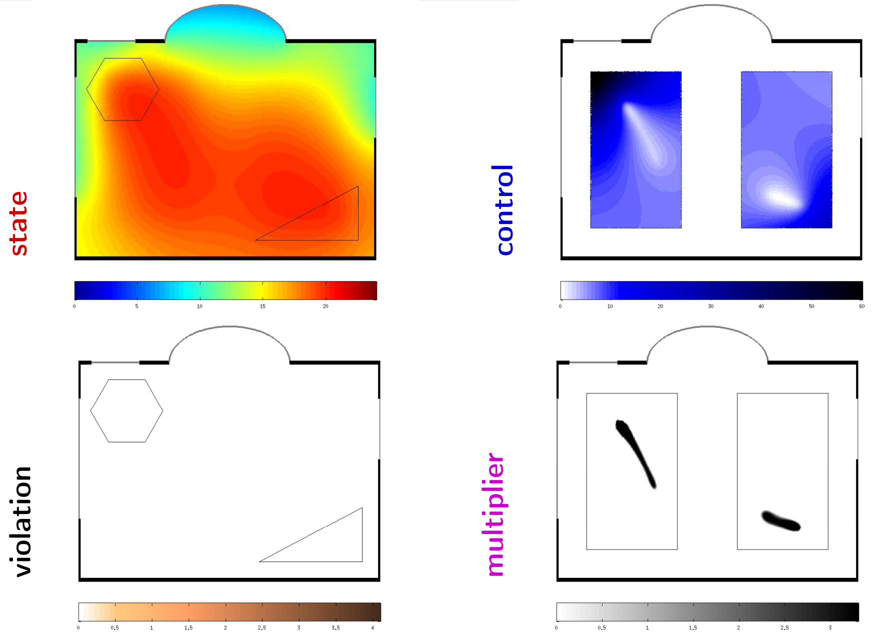
Numerical Example with $\varepsilon = 0.005$



Numerical Example with $\varepsilon = 0.002$



Numerical Example with $\varepsilon = 0.001$



Considerations for HPC Implementations

reduced formulation

- solve the PDE repeatedly
- converge only the control variable
- system solves often symmetric, positive definite (depends on PDE)

full-space formulation

- converge state, control and adjoint state simultaneously
- never solve the PDE
- system solves often symmetric, but indefinite

in between: solutions on subdomains (as in domain decomposition)

Questions for HPC implementations:

- How to handle varying effective problem dimensions due to changing active constraints?
- GenEO approach for optimal control problems?

Summary

- introduced optimal control problems with PDEs
- lead to large-scale, highly structured discrete problems
- optimization often adds an additional iteration layer (gradient descent, semi-smooth Newton, . . .)
- function-space awareness deals with mesh independence of the outer iteration layer
- complements efficient hardware-aware implementations on lower iteration layers

References I

Chen, X.; Z. Nashed; L. Qi (2000). "Smoothing methods and semismooth methods for nondifferentiable operator equations". *SIAM Journal on Numerical Analysis* 38, pp. 1200–1216. doi: 10.1137/S0036142999356719.

Günnel, A.; R. Herzog; E. Sachs (2014). "A note on preconditioners and scalar products in Krylov subspace methods for self-adjoint problems in Hilbert space". *Electronic Transactions on Numerical Analysis* 41, pp. 13–20. url: https://etna.mcs.kent.edu/volumes/2011-2020/vol41/.

Herzog, R.; K. Kunisch (2010). "Algorithms for PDE-constrained optimization". *GAMM Reports* 33.2, pp. 163–176. doi: 10.1002/gamm.201010013.

Herzog, R.; E. Sachs (2009). *Preconditioned conjugate gradient method for optimal control problems with control and state constraints*. Tech. rep. SPP1253-088. Priority Program 1253, German Research Foundation.

Hintermüller, M.; K. Ito; K. Kunisch (2002). "The primal-dual active set strategy as a semismooth Newton method". *SIAM Journal on Optimization* 13.3, pp. 865–888. doi: 10.1137/s1052623401383558.

Ulbrich, M. (2003). "Semismooth Newton methods for operator equations in function spaces". *SIAM Journal on Control and Optimization* 13.3, pp. 805–842. doi: 10.1137/s1052623400371569.