Multigrid methods for pde optimization

Alfio Borzì

Università degli Studi del Sannio
Dipartimento e Facoltà di Ingegneria, Benevento, Italia
Multigrid methods and optimization with partial differential equations are two modern fields of research in applied mathematics and engineering, both starting in the seventies with the works of A. Brandt and W. Hackbusch, and J.L. Lions.

PDE-based optimization problems

- arise in application to
  - medicine
  - biology
  - fluid dynamics
  - mechanics
  - chemical processes
  - ...inverse problems, control problems, design problems

- are large-sized problems
- many have nonlinear nondifferentiable structure
- pose new theoretical & scientific computing challenges
\[
\frac{\partial y_1}{\partial t} = -ky_1(y_1 - a)(y_1 - 1) - y_1 y_2 + \sigma \Delta y_1 + u
\]
\[
\frac{\partial y_2}{\partial t} = [\epsilon_0 + \frac{\mu_1 y_2}{\mu_2 + y_1}][-y_2 - ky_1(y_1 - b - 1)]
\]

\(y_1\) - transmembrane potential and \(y_2\) - conductance

Austrian Science Fund SFB MOBIS ”Fast Multigrid Methods for Inverse Problems”
The BEC is described by the Gross-Pitaevskii equation

\[ i\dot{\psi}(x, t) = \left( -\frac{1}{2} \nabla^2 + V(x, u(t)) + g |\psi(x, t)|^2 \right) \psi(x, t) \]

Austrian Science Fund FWF Project "Quantum optimal control of semiconductor nanostructures"
Navier-Stokes equations, RSM turbulence models, multi-phase, spray and combustion
Outline of the talk


- Optimization with PDE constraints
- SQP multigrid
- Schur-complement-smoothing multigrid
- Collective-smoothing multigrid
- Multigrid for optimization
- Convergence theory
Optimization with PDE constraints
The formulation of PDE-based optimization problems

- A model of the governing system
- A description of the optimization mechanism
- A criterion that models the purpose of the optimization and its cost

We have an infinite-dimensional constrained minimization problem

\[
\begin{align*}
\text{minimize} & \quad J(y, u) \\
\text{under the constraint} & \quad c(y, u) = 0
\end{align*}
\]

\[L(y, u, p) = J(y, u) + (c(y, u), p)\]
Characterization of the optimizing solution

\[ \min_{u \in U_{ad}} J(y, u) := h(y) + \nu g(u) \quad J : Y \times U \to \mathbb{R} \]

s.t. \[ c(y, u) = 0 \]

The existence of \( c_y^{-1} \) enables a distinction between \( y \), the state variable, and \( u \in U_{ad} \subset U \), the optimization variable in the admissible set. So we have the mapping \( u \mapsto J(y(u), u) \) in the form

\[ u \mapsto y(u) \mapsto J(y(u), u) =: \hat{J}(u) \quad \text{reduced objective} \]

The solution of \( \min_{u \in U_{ad}} \hat{J}(u) \) is characterized by the following optimality system

\[ c(y, u) = 0 \]
\[ c_y(y, u) p^* = -h'(y) \]
\[ \nu g'(u) + c_u^* p , v - u ) \geq 0 \quad \text{for all } v \in U_{ad} \]

We have \( \nabla \hat{J}(u) = \nu g'(u) + c_u^* p(u) \), the reduced gradient.
Consider the simplest objective

\[ J(y, u) = \frac{1}{2} \|y - z\|_{L^2}^2 + \frac{\nu}{2} \|u\|_{L^2}^2 \]

where \( z \in L^2 \) is a target function.

**Linear elliptic optimal control problem**

\[ c(y, u) = -\Delta y - u - f \]

**Bilinear elliptic parameter identification problem**

\[ c(y, u) = -\Delta y - u y - f \]

we assume that \( c(y, u) = 0 \) provides the mapping \( u \mapsto y(u) \).

Elliptic optimality systems

Linear

\[-\Delta y - u = f \quad \text{in} \quad \Omega\]
\[-\Delta p + y = z \quad \text{in} \quad \Omega\]
\[\nu u - p = 0 \quad \text{in} \quad \Omega\]
\[y = 0 \quad \text{and} \quad p = 0 \quad \text{on} \quad \partial \Omega\]

\[\nabla \hat{J}(u) = \nu u - p\]
\[\nabla^2 \hat{J}(u) = \nu I + \Delta^{-2}\]

Bilinear

\[-\Delta y - uy = f \quad \text{in} \quad \Omega\]
\[-\Delta p + y - up = z \quad \text{in} \quad \Omega\]
\[\nu u - yp = 0 \quad \text{in} \quad \Omega\]
\[y = 0 \quad \text{and} \quad p = 0 \quad \text{on} \quad \partial \Omega\]

\[\nabla \hat{J}(u) = \nu u - yp\]
\[\nabla^2 \hat{J}(u) = \nu I + y(\Delta + u)^{-2}y + p(\Delta + u)^{-1}y + y(\Delta + u)^{-1}p\]
Quantum control problems

Control of Bose Einstein condensates in magnetic microtraps

\[
\min_{u \in U} J(\psi, u) := \frac{1}{2} \left( 1 - \left| \langle \psi_d | \psi(T) \rangle \right|^2 \right) + \frac{\gamma}{2} \int_0^T (\dot{u}(t))^2 \, dt
\]

\[
i \frac{\partial \psi}{\partial t}(x, t) = \left( -\frac{1}{2} \nabla^2 + V(x, u(t)) + g |\psi(x, t)|^2 \right) \psi(x, t)
\]

The optimal solution is characterized by the optimality system

\[ i \frac{\partial \psi}{\partial t} = \left( -\frac{1}{2} \nabla^2 + V_u + g |\psi|^2 \right) \psi \]

\[ i \frac{\partial p}{\partial t} = \left( -\frac{1}{2} \nabla^2 + V_u + 2g |\psi|^2 \right) p + g \psi^2 p^* \]

\[ \gamma \ddot{u} = -\Re \langle \psi | \frac{\partial V_u}{\partial u} | p \rangle , \]

with the initial and terminal conditions

\[ \psi(0) = \psi_0 \text{ and } ip(T) = -\langle \psi_d | \psi(T) \rangle \psi_d \]

\[ u(0) = 0, \quad u(T) = 1. \]
The optical flow \((u, v)\) determines the transformation from an image frame to the
next.

Find \(\vec{w}\) and \(I\) such that

\[
\begin{align*}
I_t + \vec{w} \cdot \nabla I &= 0, \quad \text{in } Q = \Omega \times (0, T), \\
I(\cdot, 0) &= Y_1,
\end{align*}
\]

and minimize

\[
J(I, \vec{w}) = \frac{1}{2} \sum_{k=1}^{N} \int_{\Omega} |I(x, y, t_k) - Y_k|^2 \, d\Omega + \frac{\alpha}{2} \int_{Q} \Phi(\frac{\partial \vec{w}}{\partial t} \, |^2) \, dq \\
+ \frac{\beta}{2} \int_{Q} \psi(\sqrt{\nabla^2 u^2 + \nabla^2 v^2}) \, dq + \frac{\gamma}{2} \int_{Q} |\nabla \cdot \vec{w}|^2 \, dq,
\]

where \(\alpha, \beta,\) and \(\gamma\) are given optimization weights.

A. B., K. Ito, and K. Kunisch, *Optimal control formulation for determining optical flow*,
Optimality system

The state equation evolving forward

\[ I_t + \vec{w} \cdot \nabla I = 0. \]

The adjoint equation evolving backwards

\[ p_t + \nabla \cdot (\vec{w} p) = 0, \text{ on } t \in (t_{k-1}, t_k), \]

\[ p(\cdot, t^+_k) - p(\cdot, t^-_k) = I(\cdot, t_k) - Y_k, \quad t = t_k, \]

for \( k = 2, \ldots N - 1. \)

Two (space-time) elliptic control equations

\[
\begin{align*}
\alpha \frac{\partial^2 u}{\partial t^2} + \beta \nabla \cdot [\Psi'(|\nabla \vec{w}|^2) \nabla u] + \gamma \frac{\partial}{\partial x} (\nabla \cdot \vec{w}) &= p \frac{\partial I}{\partial x}, \\
\alpha \frac{\partial^2 v}{\partial t^2} + \beta \nabla \cdot [\Psi'(|\nabla \vec{w}|^2) \nabla v] + \gamma \frac{\partial}{\partial y} (\nabla \cdot \vec{w}) &= p \frac{\partial I}{\partial y},
\end{align*}
\]

where \( |\nabla \vec{w}|^2 = |\nabla u|^2 + |\nabla v|^2. \)
Multigrid schemes
in the SQP approach
Multigrid SQP schemes

These schemes are recommended for \( u \in U = \mathbb{R}^{n_u} \) (and no further multigrid structure within \( U \)). An SQP method iterates over the following steps:

1. **Initialize** \( \ell = 0, y_0, u_0 \)
2. **Solve the adjoint problem**
   \[
   c_y^*(y_\ell, u_\ell) p_\ell = -J_y(y_\ell, u_\ell)
   \]
   and build the reduced gradient
   \[
   \nabla \hat{J}(u)_\ell = J_u^\top + c_u^*(y_\ell, u_\ell) p_\ell
   \]
3. **Build some approximation** \( B_\ell \approx H(y_\ell, u_\ell, p_\ell) \) of the reduced Hessian, e.g., by Quasi-Newton update formulae.
4. **Solve the linear problem**
   \[
   c_y(y_\ell, u_\ell) \Delta y = -(c_u(y_\ell, u_\ell) \Delta u + c(y_\ell, u_\ell))
   \]
5. **Update** \( (y_{\ell+1}, u_{\ell+1}) = (y_\ell, u_\ell) + \tau \cdot (\Delta y, \Delta u) \), where \( \tau \) is some line-search updating factor in the early iterations.

A multigrid scheme is applied in steps (1) and (4).
Consistency condition for multigrid SQP schemes

SQP convergence theory requires that the reduced gradient can be interpreted as a derivative, i.e. we need the consistency condition

$$\nabla \hat{J}(u)^T = \frac{\partial}{\partial u} J(y - Ac_u(y, u)u, u)$$

where $A \approx c_y(y, u)^{-1}$ is the approximation to $c_y(y, u)^{-1}$ defined by the multigrid algorithm for the forward problem.

This fact results in the following requirements for the construction of the multigrid components

$$A \mathcal{I}_{k-1} = (F \mathcal{I}_k)^*, \quad A \mathcal{S} = (FS)^*, \quad A \mathcal{I}_{k} = (F \mathcal{I}_{k-1})^*$$

where $A$ and $F$ label the operators for the adjoint and forward problems.

Schur-complement multigrid approach
The KKT-matrix

Many multigrid optimization approaches use a smoothing concept based on a Schur-complement splitting of the KKT-matrix.
Let \( w = (\Delta y, \Delta u, \Delta p) \) be the solution to the equation

\[
Aw = \begin{pmatrix}
-\nabla_y L(y, u, p) \\
-\nabla_u L(y, u, p) \\
-c(y, u)
\end{pmatrix} =: f. \tag{3}
\]

where \( L(y, u, p) \) is the Lagrangian of the optimization problem and the operator matrix

\[
A = \begin{bmatrix}
L_{yy} & L_{yu} & c_y^* \\
L_{uy} & L_{uu} & c_u^* \\
c_y & c_u & 0
\end{bmatrix} \tag{4}
\]

is the Karush-Kuhn-Tucker matrix, i.e. matrix of second order derivatives of the Lagrangian of the optimization problem.
All variants of SQP methods consider approximations of the matrix \( A \) above.
Use the iterative scheme

$$w^l = w^{l-1} + R(f - A w^{l-1})$$

For a linear elliptic optimal control problem

$$A = \begin{bmatrix}
    I & 0 & -\Delta \\
    0 & \nu I & -I \\
    -\Delta & -I & 0
\end{bmatrix}$$

where we can use a multigrid Poisson-solver for inverting $-\Delta$. This is the starting point of the early multigrid optimization methods.

Consider

\[ A = \begin{bmatrix} A & B^\top \\ B & D \end{bmatrix} \]

with symmetric blocks \( A \) and \( D \), and \( A \) invertible, is an explicit reformulation of a block-Gauss-decomposition, i.e.

\[ A \begin{bmatrix} I & -A^{-1}B^\top \\ 0 & I \end{bmatrix} = \begin{bmatrix} A & 0 \\ B & S \end{bmatrix} \]

where \( S = D - BA^{-1}B^\top \) is the Schur-complement.

Iterative Schur-complement solvers are based on the scheme

\[ w^l = w^{l-1} + \begin{bmatrix} I & -\tilde{A}^{-1}B^\top \\ 0 & I \end{bmatrix} \begin{bmatrix} \tilde{A} & 0 \\ B & \tilde{S} \end{bmatrix}^{-1} (f - A w^{l-1}). \]  

(5)

where \( \tilde{A} \) and \( \tilde{S} \) are approximations to \( A \) and \( S \).
Choice of the blocks, different strategies

First tentative choice (range space factorization)

\[ A = \begin{bmatrix} L_{yy} & L_{yu} \\ L_{uy} & L_{uu} \end{bmatrix} \]

The A-block may not be invertible. This arrangement is not well suited for PDE constrained optimization problems, unlike to variational problems like Stokes or Navier-Stokes [Braess-Sarazin '97, Wittum '88]

Interchanging the 2nd and 3rd row and column in the matrix \( A \) and identifying

\[ A = \begin{bmatrix} L_{yy} & c_y^* \\ c_y & 0 \end{bmatrix}, \quad B = \begin{bmatrix} L_{uy} & c_u^* \end{bmatrix}, \quad D = L_{uu} \]

leads to a nullspace decomposition.
Choice of the blocks, different strategies

With nullspace decomposition the Schur-complement reads

\[ S = L_{uu} - L_{uy} c_y^{-1} c_u - c_u^* c_y^{-1} L_{yu} + c_u^* c_y^{-1} L_{yy} c_y^{-1} c_u \]

that is the reduced Hessian that characterizes the optimization problem. Coercivity of the reduced Hessian guarantees well-posedness of the overall optimization problem.

In the case of a linear control problem one obtains

\[ S = \nu I - 0 \Delta^{-1} I - I \Delta^{-1} 0 + I \Delta^{-1} I \Delta^{-1} I = \nu I + (\Delta^{-1})^2 \]

With the choice

\[ \tilde{A} = \begin{bmatrix} L_{yy} & \tilde{c}_y^* \\ \tilde{c}_y & 0 \end{bmatrix} \quad \text{and} \quad \tilde{S} = \nu I \]

one obtains a transforming smoothing iteration.

Collective smoothing
multigrid approach
A collective smoothing multigrid (CSMG) approach means solving the optimality system for the state, the adjoint, and the control variables simultaneously in the multigrid process by using collective smoothers for the optimizations variables.

A CSMG based scheme aims at realizing the tight coupling in the optimality system along the hierarchy of grids in space and time.

The smoothing process: The basic idea is to solve the optimality system at grid or block-grid point level, within the admissible set and consistently with the differential structure of the problem.

The CSMG multigrid schemes result robust with respect to optimization parameters and provide mesh-independent convergence.
CSMG smoothing for a linear elliptic control problem

\[-\Delta y - u = f \]
\[-\Delta p + y = z \]
\[(\nu u - p, \nu - u) \geq 0 \quad \text{for all } \nu \in U_{ad}\]

where \(U_{ad} := \{u \in L^2(\Omega) \mid u_L(x) \leq u(x) \leq u_H(x) \text{ a.e. in } \Omega\}\)

Let \(A = -y_{i-1,j} - y_{i+1,j} - y_{i,j-1} - y_{i,j+1} - h^2 f_{i,j},\)
\(B = -p_{i-1,j} - p_{i+1,j} - p_{i,j-1} - p_{i,j+1} - h^2 z_{i,j}.\)

Then \(A + 4y_{i,j} - h^2 u_{i,j} = 0,\)
\(B + 4p_{i,j} + h^2 y_{i,j} = 0,\)
\(\nu u_{i,j} - p_{i,j} = 0.\)

\[\Rightarrow \quad y_{i,j}(u_{i,j}) = \frac{1}{4}(h^2 u_{i,j} - A) \]
\[\Rightarrow \quad p_{i,j}(u_{i,j}) = \frac{1}{16}(-h^2 u_{i,j} + h^2 A - 4B)\]

Setting \(\nabla \hat{J}(u) = \nu u - p(u) = 0,\) we obtain the auxiliary variable

\[\tilde{u}_{i,j} = \frac{1}{16\nu + h^4}(h^2 A - 4B).\]
The new value for $u_{i,j}$ resulting from the smoothing step is given by

$$u_{i,j} = \begin{cases} 
  u_{Hi,j} & \text{if } \tilde{u}_{i,j} \geq u_{Hi,j} \\
  \tilde{u}_{i,j} & \text{if } u_{Li,j} < \tilde{u}_{i,j} < u_{Hi,j} \\
  u_{Li,j} & \text{if } \tilde{u}_{i,j} \leq u_{Li,j}
\end{cases}.$$  

This collective Gauss-Seidel step satisfies the inequality constraint.

**Example:** $\tilde{u} > u_H$; then $u = u_H$.

Therefore $(\nu - u) \leq 0$ for any $\nu \in U_{ad}$, and

$$\nu u - p = \nu u - (4B - h^2 A - h^4 u_{ij})/16$$
$$= [(16\nu + h^4)u - (4B - h^2 A)]/16$$
$$< [(16\nu + h^4)\tilde{u} - (4B - h^2 A)]/16 = 0.$$  

Therefore $(\nu u - p) \cdot (\nu - u) \geq 0$ for all $\nu \in U_{ad}$. 
A discovery: Bang-bang control phenomenon

Consider \( u_L = -30 \) and \( u_H = 30 \), \( \nu = 0 \), \( z(x, y) = \sin(4\pi x) \sin(2\pi y) \).

| mesh       | \( \rho(y), \rho(p) \) | \( |y - z|_0 \) | \( |r(y)|_0, |r(p)|_0 \) |
|------------|-------------------------|----------------|------------------|
| 513 \times 513 | 0.12, 0.13             | 3.70 \times 10^{-1} | 2.9 \times 10^{-8}, 1.3 \times 10^{-13} |
| 1025 \times 1025 | 0.12, 0.13             | 3.70 \times 10^{-1} | 2.5 \times 10^{-8}, 4.2 \times 10^{-13} |
| 2049 \times 2049 | 0.12, 0.16             | 3.70 \times 10^{-1} | 1.9 \times 10^{-8}, 1.6 \times 10^{-12} |

The state (left) and the control (right).

Multigrid collective smoothing with FEM

Optimality system

\[ Q Y - M U = F, \]
\[ Q P + M Y = Z, \]
\[ (\nu U - P, \Phi - U) \geq 0, \quad \forall \Phi \in U. \]

Define

\[ C_1 = \sum_{j=1}^{n_k} q_{i,j} Y_j, \quad C_2 = \sum_{j=1}^{n_k} q_{i,j} P_j, \quad C_3 = \sum_{j=1}^{n_k} m_{i,j} Y_j, \quad \text{and} \quad C_4 = \sum_{j=1}^{n_k} m_{i,j} U_j. \]

\[ Y_i(U_i) = \frac{1}{q_{i,i}} (F_i - C_1 + C_4 + m_{i,i} U_i), \]
\[ P_i(U_i) = \frac{1}{q_{i,i}^2} [q_{i,i} (Z_i - C_2 - C_3) - m_{i,i} (F_i - C_1 + C_4 + m_{i,i} U_i)]. \]

Define the auxiliary variable as

\[ \widetilde{U}_i = \frac{1}{\nu q_{i,i}^2 + m_{i,i}^2} [q_{i,i} (Z_i - C_2 - C_3) - m_{i,i} (F_i - C_1 + C_4)]. \]
Numerical results - CSMG-FEM

Control-constrained linear problem

Figure: Numerical solutions $y$ (left) and $u$ (right) for the control-constrained linear elliptic optimal control problem with $\nu = 10^{-6}$. 
Let
\[ A = -y_{i-1,j} - y_{i+1,j} - y_{i,j-1} - y_{i,j+1} - h^2 f_{i,j}, \]
\[ B = -p_{i-1,j} - p_{i+1,j} - p_{i,j-1} - p_{i,j+1} - h^2 z_{i,j}. \]

Then
\[ A + 4y_{i,j} - h^2 u_{i,j} y_{i,j} = 0, \]
\[ B + 4p_{i,j} + h^2 y_{i,j} - h^2 u_{i,j} p_{i,j} = 0, \]
\[ \nu u_{i,j} - y_{i,j} p_{i,j} = 0. \]

\[ \Rightarrow y_{i,j}(u_{i,j}) = \frac{-1}{4 - h^2 u_{i,j}} A \]
\[ \Rightarrow p_{i,j}(u_{i,j}) = \frac{1}{(4 - h^2 u_{i,j})^2} (h^2 A + h^2 u_{i,j} - 4B) \]

Assume \((4 - h^2 u_{i,j}) \neq 0\) at any \(ij\). Setting \(\nabla \hat{J}(u) = \nu u - y(u)p(u) = 0\), we obtain \(u_{i,j}\) as solution of the quartic polynomial equation
\[ \nu h^6 u_{i,j}^4 - 12\nu h^4 u_{i,j}^3 + 48\nu h^2 u_{i,j}^2 - (64\nu + h^2 AB)u_{i,j} - (h^2 A^2 - 4AB) = 0. \]

Let the auxiliary variable \(\tilde{u}_{i,j}\) be the solution of the quartic polynomial
\[ \nu h^6 u_{i,j}^4 - 12\nu h^4 u_{i,j}^3 + 48\nu h^2 u_{i,j}^2 - (64\nu + h^2 AB)u_{i,j} - (h^2 A^2 - 4AB) = 0. \]
Bilinear control problem / parameter identification problem

\[ \min_{u \in U} J(y, u) := \frac{1}{2} \| y - z \|_{L^2}^2 + \frac{\nu}{2} \| u \|_{L^2}^2 \]

\[-\Delta y - u y = f \text{ in } \Omega \]

\[ y = 0 \text{ on } \partial \Omega \]

The (non attainable) target function \( z \) is given by

\[ z(x_1, x_2) = \begin{cases} 
2, & \text{on } (0.25, 0.75) \times (0.25, 0.75) \\
1, & \text{otherwise}
\end{cases} \]
Numerical results

Bilinear control problem / parameter identification problem

**Figure:** Numerical solutions $y$ (left) and $u$ (right) for the control-unconstrained bilinear elliptic optimal control problem with $\nu = 10^{-4}$. (CSMG, FDM)

<table>
<thead>
<tr>
<th>mesh</th>
<th>$\nu = 10^{-2}$</th>
<th>$\nu = 10^{-4}$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CSMG</td>
<td>NCG</td>
</tr>
<tr>
<td>$65 \times 65$</td>
<td>47.6</td>
<td>0.8</td>
</tr>
<tr>
<td>$129 \times 129$</td>
<td>221.9</td>
<td>4.0</td>
</tr>
<tr>
<td>$257 \times 257$</td>
<td>912.8</td>
<td>19.0</td>
</tr>
</tbody>
</table>
Other developments for elliptic optimization problems

- CSMG scheme for state-constrained elliptic control problems.
- Collective smoothing AMG for convection-diffusion optimality systems with jumping coefficients.
- CSMG schemes for optimality systems and higher-order discretization.
- Globalization techniques within CSMG schemes applied to non convex optimization problems.
- CSMG methods for singular control problems.
- CSMG methods for boundary control problems.
- CSMG convergence theory in the LFA and BPX frameworks.
CSMG schemes for time-dependent problems
Consider a reaction-diffusion model

\[
\min_{u \in L^2(Q)} J(y, u) \\
-\partial_t y + G(y) + \sigma \Delta y = u \quad \text{in } Q = \Omega \times (0, T) \\
y = y_0 \quad \text{in } \Omega \times \{t = 0\} \\
y = 0 \quad \text{on } \Sigma = \partial\Omega \times (0, T)
\]

Control required to
track a desired trajectory \(y_d(x, t)\)
reach a desired terminal state \(y_T(x)\)

For this purpose, the following objective can be considered

\[
J(y, u) = \frac{\alpha}{2} \|y - y_d\|_{L^2(Q)}^2 + \frac{\beta}{2} \|y(\cdot, T) - y_T\|_{L^2(\Omega)}^2 + \frac{\nu}{2} \|u\|_{L^2(Q)}^2
\]
Reaction diffusion process control through boundary terms

\[
\begin{aligned}
\min_{u \in L^2(\Sigma)} J(y, u) \\
-\partial_t y + G(y) + \sigma \Delta y &= 0 \quad \text{in } Q = \Omega \times (0, T) \\
y &= y_0 \quad \text{in } \Omega \times \{t = 0\} \\
-\frac{\partial y}{\partial n} &= u \quad \text{on } \Sigma = \partial\Omega \times (0, T)
\end{aligned}
\]

Control required to track a desired trajectory \( y_d(x, t) \)
reach a desired terminal state \( y_T(x) \)

For this purpose, the following objective can be considered

\[
J(y, u) = \frac{\alpha}{2} \| y - y_d \|^2_{L^2(Q)} + \frac{\beta}{2} \| y(\cdot, T) - y_T \|^2_{L^2(\Omega)} + \frac{\nu}{2} \| u \|^2_{L^2(\Sigma)}
\]
Optimality systems

Distributed source control: The solution is characterized by

\[-\partial_t y + G(y) + \sigma \Delta y = u \quad \text{in } Q\]
\[\partial_t p + G'(y)p + \sigma \Delta p + \alpha (y - y_d) = 0 \quad \text{in } Q\]
\[\nu u - p = 0 \quad \text{in } Q\]
\[y = 0, \quad p = 0 \quad \text{on } \Sigma\]

Neumann boundary control: The optimal solution satisfies

\[-\partial_t y + G(y) + \sigma \Delta y = 0 \quad \text{in } Q\]
\[\partial_t p + G'(y)p + \sigma \Delta p + \alpha (y - y_d) = 0 \quad \text{in } Q\]
\[\nu u - p = 0 \quad \text{on } \Sigma\]
\[-\frac{\partial y}{\partial n} = u, \quad -\frac{\partial p}{\partial n} = 0 \quad \text{on } \Sigma\]

With initial condition \(y(x,0) = y_0(x)\) for the state variable (evolving forward in time). And the terminal condition for the adjoint variable (evolving backward in time) \(p(x, T) = \beta(y(x, T) - y_T(x))\).
Discretization

The optimality systems are discretized by, e.g., finite differences and backward Euler scheme.

$\Omega_h$ defines the set of interior mesh-points, $(x_i, y_j), 2 \leq i, j \leq N_x$.

The space-time grid is defined by

$$Q_{h,\delta t} = \{(x, t_m) : x \in \Omega_h, t_m = (m - 1)\delta t, 1 \leq m \leq N_t + 1, \delta t = T / N_t\}$$

Time difference operators

$$\partial^- t y^m_h = \frac{y^m_h - y^{m-1}_h}{\delta t} \quad \text{and} \quad \partial^+ t p^m_h = \frac{p^{m+1}_h - p^m_h}{\delta t}$$

Example distributed control:

$$-\partial^- t y^m_h + G(y^m_h) + \sigma \Delta_h y^m_h = u^m_h$$

$$\partial^+ t p^m_h + G'(y^m_h)p^m_h + \sigma \Delta_h p^m_h + \alpha(y^m_h - y^m_{dh}) = 0$$

$$\nu u^m_h - p^m_h = 0$$

Boundary control: Consider the optimality system on the boundary and discretize the boundary derivative using a second-order centered scheme.
Coarsening strategy

Consider $L$ levels, $k = 1, \ldots, L$.

Coarsening in the space directions: $h = h_k = h_1/2^{k-1}$.
Coarsening in time direction: $\delta t = \delta t_k = \delta t_1/s^{k-1}$,

$s = 1$ semicoarsening in space;
$s = 2$ standard time coarsening;
$s = 4$ double time coarsening.

Transfer operators

$I_{k-1}^k$ denotes the injection or FW operator for restriction.
$I_k^{k-1}$ denotes bilinear interpolation in space and
If $s = 1$ no interpolation in time is needed,
if $s \in \{2, 4\}$ then $I_{k-1}^k$ corresponds to bilinear interpolation in space and in time.
Collective smoothing: Two essential requirements

- **Coupling** between state and control variables.
- Preserve opposite orientation of state and adjoint equations.

**Backward Euler discretization**

\[
\begin{align*}
  t &= 0 \\
  y_0 &\quad y \\
  t &= T
\end{align*}
\]

\[
\begin{align*}
  p &\quad p_0
\end{align*}
\]
Time-Splitted Collective Gauss-Seidel Iteration (TS-CGS)

\[
\begin{pmatrix}
  y^{(1)}_i \\
p^{(1)}_i
\end{pmatrix}_{jm} = \begin{pmatrix}
  y^{(0)}_i \\
p^{(0)}_i
\end{pmatrix}_{jm} + \left[ \begin{array}{cc}
  -(1 + 4\sigma\gamma + \delta t G') & -\delta t/\nu \\
  \delta t(\alpha + G'' p) & -(1 + 4\sigma\gamma + \delta t G')
\end{array} \right]^{-1} \begin{pmatrix}
  r_y \\
r_p
\end{pmatrix}_{jm}
\]

1. Set the starting approximation.

2. For \( m = 2, \ldots, N_t \) do

3. For \( ij \) in, e.g., lexicographic order do

   \[
y^{(1)}_{i,j,m} = y^{(0)}_{i,j,m} + \left[ \frac{-(1 + 4\sigma\gamma + \delta t G')}{[-(1 + 4\sigma\gamma + \delta t G')]^2 + \frac{\delta t^2}{\nu}(\alpha + G'' p)} \right]_{ij,m} r_y(w) + \frac{\delta t}{\nu} r_p(w) \]

4. end.
Consider the discrete optimality system at $i,j$ and for all time steps and solve the resulting block-tridiagonal system.

$$\begin{pmatrix} y \\ p \end{pmatrix}_{ij}^{(1)} = \begin{pmatrix} y \\ p \end{pmatrix}_{ij}^{(0)} + M^{-1} \begin{pmatrix} r_y \\ r_p \end{pmatrix}_{ij}$$

The block-tridiagonal system has the following form

$$M = \begin{bmatrix} A_2 & C_2 & C_3 & C_4 \\ B_3 & A_3 & C_3 & C_4 \\ B_4 & A_4 & C_4 & A_{N_{t+1}} \\ B_{N_{t+1}} & C_{N_{t+1}} & A_{N_{t+1}} & \end{bmatrix}, \ A_m = \begin{bmatrix} -(1 + 4\sigma \gamma) + \delta t G' & -\frac{\delta t}{\nu} \\ \delta t (\alpha + G'' p) & -(1 + 4\sigma \gamma) + \delta t G' \end{bmatrix}$$

$$B_m = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \text{ and } C_m = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$$

Centered at $t_m$, the entries $B_m$, $A_m$, $C_m$ refer to the variables $(y, p)$ at $t_{m-1}$, $t_m$, and $t_{m+1}$, respectively.
Convergence factors of TL-CGS scheme (left) and TS-CGS scheme

<table>
<thead>
<tr>
<th>$\eta$</th>
<th>$\nu$</th>
<th>TL-CGS</th>
<th>TS-CGS</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\gamma$</td>
<td>$10^{-8}$</td>
<td>$10^{-6}$</td>
<td>$10^{-4}$</td>
</tr>
<tr>
<td>32</td>
<td>0.04</td>
<td>0.11</td>
<td>0.11</td>
</tr>
<tr>
<td>48</td>
<td>0.07</td>
<td>0.11</td>
<td>0.11</td>
</tr>
<tr>
<td>64</td>
<td>0.09</td>
<td>0.11</td>
<td>0.11</td>
</tr>
</tbody>
</table>

The estimated convergence factors $\eta$ for TL-CGS and TS-CGS multigrid schemes.
Consider the optimal control problem of tracking $y_d$ for $t \geq 0$. Define time windows of size $\Delta t$. In each time window, an optimal control problem with tracking ($\alpha = 1$) and terminal observation ($\beta = 1$) is solved.

**Multigrid Receding Horizon Scheme (CSMG-RH)**

Applied to the control of a solid fuel ignition model:

$$-\partial_t y + \sigma \Delta y + \delta e^{y} = f$$
Other developments for parabolic optimization problems


MG OPT
multigrid for optimization
The MGOPT solution to the optimization problem $\min_u \hat{J}(u)$ requires to define a hierarchy of minimization problems

$$\min_{u_k} \hat{J}_k(u_k) \quad k = 1, 2, \ldots, L$$

where $u_k \in V_k$ and $\hat{J}_k(\cdot)$ is the reduced cost functional.

Among spaces $V_k$, restriction operators $I_{k-1}^k : V_k \to V_{k-1}$ and prolongation operators $I_{k-1}^k : V_{k-1} \to V_k$ are defined.

Require that $(I_{k-1}^k u, v)_{k-1} = (u, I_{k-1}^k v)_k$ for all $u \in V_k$ and $v \in V_{k-1}$.

We also choose an optimization scheme as 'smoother'

$$u_k^{(l)} = O_k(u_k^{(l-1)})$$

That provides sufficient reduction

$$\hat{J}_k(O_k(u_k^{(l)})) < \hat{J}_k(u_k^{(l)}) - \eta \| \nabla \hat{J}_k(u_k^{(l)}) \|^2$$

for some $\eta \in (0, 1)$. 
The MGOPT scheme is a FAS scheme applied to the gradient problem \( \nabla \hat{J}_k(u_k) = 0 \) where smoothing is replaced by any optimization scheme and a linesearch (damping) is applied to the coarse-grid correction step.

FAS Coarse Grid Problem

\[
\begin{align*}
A_k(u_k) &= f_k \\
A_{k-1}(u_{k-1}) &= I_{k-1}^k f_k + A_{k-1}(u_{k-1}) - I_{k-1}^k A_k(u_k)
\end{align*}
\]

MGLOPT Coarse Grid Problem

\[
\begin{align*}
\nabla \hat{J}_k(u_k) &= 0 \\
\nabla \hat{J}_{k-1}(u_{k-1}) &= \nabla \hat{J}_{k-1}(u_{k-1}) - I_{k-1}^k \nabla \hat{J}_k(u_k)
\end{align*}
\]

In the context of MGOPT, the ‘smoother’ is any standard one-grid optimization scheme.

- Steepest descent method
- Nonlinear conjugate gradient method
- Newton-type schemes

\[ u_{k+1} = u_k + \alpha_k p_k \quad \text{where} \quad p_k = -B_k^{-1} \nabla \hat{J}_k \]

MGOPT

- can be applied to any optimization problem given the reduced gradient
- independent of discretization
- independent of structure of the problem
MGOPT scheme

Initialize $u_0^k$. If $k = 1$, solve $\min_{u_k} \hat{J}_k(u_k) - (f_k, u_k)_k$ and return. Else,

1. **Pre-optimization**: $u_k^l = O_k(u_k^l, f_k)$, $l = 1, 2, \ldots, \gamma_1$

2. **Coarse grid problem**

   Restrict the solution of (1): $u_{k-1}^{\gamma_1} = I_k^{k-1}u_k^{\gamma_1}$

   Fine-to-coarse correction: $\tau_{k-1} = \nabla \hat{J}_{k-1}(u_{k-1}^{\gamma_1}) - I_k^{k-1}\nabla \hat{J}_k(u_k^{\gamma_1})$

   $f_{k-1} = I_k^{k-1}f_k + \tau_{k-1}$

   Apply MGOPT to the Coarse grid problem:

   $$\min_{u_{k-1}} \hat{J}_{k-1}(u_{k-1}) - (f_{k-1}, u_{k-1})_{k-1}$$

3. **Coarse grid correction**

   Prolongate the error: $d = I_k^{k-1}(u_{k-1} - u_{k-1}^{\gamma_1})$

   Perform a line search in the direction $d$ to obtain a step length $\alpha_k$.

   Coarse grid correction: $u_k^{\gamma_1+1} = u_k^{\gamma_1} + \alpha_k d$

4. **Post-optimization**: $u_k^l = O_k(u_k^l, f_k)$, $l = \gamma_1 + 2, \ldots, \gamma_1 + \gamma_2 + 1$
Control of Bose Einstein condensates in magnetic microtraps

We consider transport of Bose-Einstein condensates in magnetic microtraps, controllable by external parameters such as wire currents or radio-frequency fields.

The mean-field dynamics of the condensate is described by the Gross-Pitaevskii equation (GPE)

\[ i\dot{\psi}(x, t) = \left( -\frac{1}{2} \nabla^2 + V(x, u(t)) + g |\psi(x, t)|^2 \right) \psi(x, t) \]

\( V(x, u(t)) \) is a three-dimensional potential produced by a magnetic microtrap. \( u(t) \) is a control parameter that describes the variation of the confining potential.
Transport of Bose Einstein condensates in magnetic microtraps

Table: Computational performance of the CNCG and MGOPT schemes for different values of $g$; $T = 7.5$, $\gamma = 10^{-4}$, mesh $128 \times 1250$.

<table>
<thead>
<tr>
<th>$g$</th>
<th>CNCG</th>
<th>CPU</th>
<th>MGOPT</th>
<th>CPU</th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
<td>$3.89 \cdot 10^{-4}$</td>
<td>53</td>
<td>$7.08 \cdot 10^{-4}$</td>
<td>149</td>
</tr>
<tr>
<td>50</td>
<td>$2.35 \cdot 10^{-3}$</td>
<td>80</td>
<td>$9.84 \cdot 10^{-3}$</td>
<td>76</td>
</tr>
<tr>
<td>75</td>
<td>$5.54 \cdot 10^{-3}$</td>
<td>90</td>
<td>$1.85 \cdot 10^{-3}$</td>
<td>163</td>
</tr>
<tr>
<td>100</td>
<td>$4.94 \cdot 10^{-1}$</td>
<td>50</td>
<td>$5.44 \cdot 10^{-3}$</td>
<td>257</td>
</tr>
</tbody>
</table>
Multigrid convergence theory
1. Multigrid convergence theory for scalar elliptic equation

\[-\Delta y = f \text{ in } \Omega, \text{ and } y = 0 \text{ on } \partial \Omega.\]

discrete matrix form \[\Rightarrow A_k y_k = f_k\]

**Theorem 1:** Let \(M_k := I_k - B_k A_k\). Then there exists a positive constant \(\delta < 1\) such that

\[(A_k M_k y, y)_k \leq \delta (A_k y, y)_k, \text{ for all } y \in V_k.\]
2. Multigrid convergence theory for decoupled symmetric system

\[-\nu \Delta y = \nu f \text{ in } \Omega, \text{ and } y = 0 \text{ on } \partial \Omega,\]
\[-\Delta p = z \text{ in } \Omega, \text{ and } p = 0 \text{ on } \partial \Omega.\]

Discrete matrix form \( \Rightarrow \) \( \bar{A}_k w_k = g_k, \quad \bar{A}_k = \begin{pmatrix} \nu A_k & 0 \\ 0 & A_k \end{pmatrix} \)

This system is exactly two copies of Poisson problem. Hence the multigrid convergence theory for this system inherits the properties of the scalar case.

**Theorem 2:** Let \( \bar{M}_k := \bar{I}_k - \bar{B}_k \bar{A}_k \). Then there exists a positive constant \( \bar{\delta} < 1 \) such that

\[(\bar{A}_k \bar{M}_k w, w)_k \leq \bar{\delta}(\bar{A}_k w, w)_k, \quad \text{for all } w = (y, p) \in V_k \times V_k,\]

where \( \bar{\delta} = \delta \).
3. Multigrid convergence theory for optimality systems

\[-\nu \Delta y - p = \nu f \quad \text{in } \Omega, \quad \text{and } y = 0 \quad \text{on } \partial \Omega,\]
\[-\Delta p + y = z \quad \text{in } \Omega, \quad \text{and } p = 0 \quad \text{on } \partial \Omega.\]

Consider

\[A_k = \bar{A}_k + D_k, \quad D_k = \begin{pmatrix} 0 & -I_k \\ I_k & 0 \end{pmatrix}.\]

**Theorem 3:** Let \(M_k := I_k - B_k A_k\). Then there exist positive constants \(h_0\) and \(\hat{\delta} < 1\) such that \(\forall \ h_1 < h_0\)

\[(\bar{A}_k M_k w, w)_k \leq \hat{\delta}(\bar{A}_k w, w)_k, \quad \text{for all } w = (y, p) \in V_k \times V_k,\]

where \(\hat{\delta} = \delta + C h_1\).
Theoretical analysis of MGOPT

We assume that for each $k$, $\hat{J}_k$ is twice Fréchet differentiable and $\nabla^2 \hat{J}_k$ is (locally) positive definite and satisfies the conditions

$$(\nabla^2 \hat{J}_k(u) \nu, \nu)_k \geq \beta \|\nu\|^2_k$$

and

$$\|\nabla^2 \hat{J}_k(u) - \nabla^2 \hat{J}_k(\nu)\| \leq \lambda \|u - \nu\|_k$$

uniformly for some positive constants $\beta$ and $\lambda$.

We remark that

$$\nabla \left( \hat{J}_{k-1}(u_{k-1}) - (f_{k-1}, u_{k-1})_{k-1} \right)_{u_{k-1}=I^k_{k-1} u_k} = I_{k-1}^k \left( \nabla \hat{J}_k(u_k) - f_k \right),$$

We use the expansion

$$\hat{J}_k(u + z) = \hat{J}_k(u) + (\nabla \hat{J}_k(u), z)_k + \frac{1}{2} \int_0^1 (\nabla^2 \hat{J}_k(u + tz)z, z)_k dt.$$
Lemma 1: For $u, v \in V_k$ assume $(\nabla \hat{J}_k(u) - f_k, v)_k \leq 0$ and let $\gamma$ be such that

$$
0 \leq \gamma \leq -2\delta(\nabla \hat{J}_k(u) - f_k, v)_k \left[\int_0^1 (\nabla^2 \hat{J}_k(u + t\gamma v)v, v)_k dt\right]^{-1} \delta \in [0, 1].
$$

Then $-(1 - \delta)\gamma(\nabla \hat{J}_k(u) - f_k, v)_k \leq \hat{J}_k(u) - \hat{J}_k(u + \gamma v) + \gamma(f_k, v)_k \leq -\gamma(\nabla \hat{J}_k(u) - f_k, v)_k$.

We can find an explicit estimate for $\alpha$ such that a sufficient decrease is satisfied.

Lemma 2: For $u, v \in V_k$ assume $(\nabla \hat{J}_k(u) - f_k, v)_k \leq 0$ and let

$$
\alpha(u, v) = \min \left\{ 2, \frac{-(\nabla \hat{J}_k(u) - f_k, v)_k}{\lambda \|v\|_k^3 + (\nabla^2 \hat{J}_k(u)v, v)_k} \right\}, \quad \lambda > 0.
$$

Then

$$
\hat{J}_k(u + \alpha(u, v)v) \leq \hat{J}_k(u) + \alpha(u, v)(f_k, v) + \frac{1}{2}\alpha(u, v)(\nabla \hat{J}_k(u) - f_k, v)_k.
$$
The coarse-grid correction provides a descending direction

The following lemma states that the coarse-to-fine minimization step with step-length $\alpha$ is a minimizing step.

**Lemma 3:** Let $u \in V_k$ and define $\tilde{u} = I_{k-1}^k u$. Let $\tilde{v} \in V_{k-1}$ and define $v = I_{k-1}^k (\tilde{v} - \tilde{u})$. Assume

$$\hat{J}_{k-1}(\tilde{v}) - (f_{k-1}, \tilde{v})_{k-1} \leq \hat{J}_{k-1}(\tilde{u}) - (f_{k-1}, \tilde{u})_{k-1},$$

where

$$(f_{k-1}, \tilde{v})_{k-1} = \left( I_{k-1}^k f_k + \nabla \hat{J}_{k-1}(\tilde{v}) - I_{k-1}^k \nabla \hat{J}_k(v), \tilde{v} \right)_{k-1},$$

and

$$(f_{k-1}, \tilde{u})_{k-1} = \left( I_{k-1}^k f_k + \nabla \hat{J}_{k-1}(\tilde{u}) - I_{k-1}^k \nabla \hat{J}_k(u), \tilde{u} \right)_{k-1}.$$

Then

$$\hat{J}_k(u + \alpha(u, v)v) - \hat{J}_k(u) - \alpha(u, v)(f_k, v)_k \leq \frac{1}{2} \alpha(u, v)(\nabla \hat{J}_k(u) - f_k, v)_k.$$
Theorem 4: Let \( \hat{J}_k(u_k) - (f_k, u_k)_k \) be twice Fréchet differentiable and let \( \nabla^2 \hat{J}_k \) be locally Lipschitz continuous and satisfies the Hessian conditions in a neighborhood \( V_k^\epsilon \) of

\[
u_k^* = \arg\min_{u_k} (\hat{J}_k(u_k) - (f_k, u_k)_k).
\]

Then the MGOPT method provides a minimizing iteration.


Best regards and many thanks for your interest!!

See you in Benevento