A semismooth Newton method for a class of semilinear optimal control problems with box and volume constraints

Antoine Laurain (joint work with Samuel Amstutz)

Technical University of Berlin

PICOF'12

Optimal control problem

General formulation

- $(\mathcal{P})$: $\min_{(u,y)\in U_{ad}\times Y} J(u,y) \qquad$ subject to $E(u,y)=0$,

$$U := \{u \in L^2(D),\ 0 \leq u \leq 1 \text{ a.e. in } D\},$$
$$U_{ad} := \{u \in U,\ \int_D u = m\}, \qquad 0 < m < |D|,$$

- D is a bounded domain of $\mathbb{R}^N$, $N \in \{2,3\}$.
- $J : L^2(D) \times Y \to \mathbb{R}$ and $E : U \times Y \to Z$.
- Y, Z are Banach spaces.
- E denotes a class of semilinear equations.

- Ito & Kunisch (2004): L^2-control cost, primal-dual active set method for nonlinear problems and bilateral constraints.
- Stadler (2009), Wachsmuth (2011): L^1-control cost, linear case.

Motivation

Shape/topology optimization

Shape optimization: the control is a set $\Omega \subset \mathbb{R}^N$, or its indicator function χ_Ω which takes values in $\{0, 1\}$.
Topology optimization: if the topology is unknown.

- Allaire, Bendsøe...: Relaxed formulation.
- Delfour-Zolésio, Murat-Simon..: Smooth boundary perturbations.
- Masmoudi, Sokolowski..: Topological derivative.

Optimal control approach

- With an L^1-control cost and a linear elliptic state equation, the control u eventually takes $0-1$ values.
- Nonsmooth Newton methods available.
- L^1-control cost $\|u\|_{L^1}$ corresponds to a volume constraint.
- Total Variation $\|Du\|_{L^1}$ corresponds to a perimeter constraint.

Nonsmooth Newton Method

Main idea

- Reformulation of the optimality conditions for $(\mathcal{P})$:

$$\Phi(u, y, p, \lambda) = 0$$

  where (p, λ) are Lagrange multipliers.
- Φ is a nonsmooth, nonlinear vector function.
- Generalized differentiability of Φ: Newton derivative.
- Use a semismooth Newton Method (Hintermüller,Ito,Kunisch ...)

Binary and sparse solutions

- In certain cases we show that the solution is binary.
- Numerical solutions exhibits in general a piecewise constant nature for the semilinear problem.
- The volume constraint allows to exactly control the level of sparsity of u.

Problem statement and optimality conditions

Definition

For every $u \in U_{ad}$ we define the cone $K(u) \subset L^2(D)$ by

$$\forall v \in L^2(D), \qquad v \in K(u) \Longleftrightarrow \begin{cases} v = 0 \text{ a.e. in } [0 < u < 1], \\ v \geq 0 \text{ a.e. in } [u = 0], \\ v \leq 0 \text{ a.e. in } [u = 1]. \end{cases}$$

Theorem (optimality conditions)

Let $(\bar{u}, \bar{y})$ be an optimal solution of $(\mathcal{P})$. With appropriate minimal assumptions on E, J, there exists $(\bar{\lambda}, \bar{p}) \in \mathbb{R} \times Z'$ such that

$$\begin{aligned} L_u(\bar{u}, \bar{y}, \bar{p}) + \bar{\lambda} &\in K(\bar{u}), \\ L_y(\bar{u}, \bar{y}, \bar{p}) &= 0, \\ L_p(\bar{u}, \bar{y}, \bar{p}) &= 0, \\ \int_D \bar{u} &= m. \end{aligned}$$

Problem statement and optimality conditions

For all $(u, y, p, \lambda, g) \in L^2(D) \times Y \times Z' \times \mathbb{R} \times L^2(D)$ we set

$$T(u, g) := u \max(0, g) + (1 - u) \min(0, g),$$

and

$$\Phi(u, y, p, \lambda) := \begin{pmatrix} T(u, L_u(u, y, p) + \lambda) \\ L_y(u, y, p) \\ L_p(u, y, p) \\ \int_D u - m \end{pmatrix}.$$

Theorem

Let $(\bar{u}, \bar{y}, \bar{p}, \bar{\lambda}) \in L^2(D) \times Y \times Z' \times \mathbb{R}$. The optimality conditions are equivalent to

$$\Phi(\bar{u}, \bar{y}, \bar{p}, \bar{\lambda}) = 0.$$

Semismooth Newton method

Newton derivative

Let $\mathcal{X}, \mathcal{Y}$ be Banach spaces and $\mathcal{U} \subset \mathcal{X}$ open. If there exists $G : \mathcal{U} \to \mathcal{L}(\mathcal{X}, \mathcal{Y})$ such that for all $\mathcal{U} \in V$

$$\lim_{h \to 0} \frac{1}{\|h\|_{\mathcal{X}}} \|F(u+h) - F(u) - G(u+h)h\|_{\mathcal{Y}} = 0$$

then $F : \mathcal{U} \to \mathcal{Y}$ is Newton differentiable, G is the Newton derivative.

Semismooth Newton method

Suppose $F(u^*) = 0$ and $F : \mathcal{X} \to \mathcal{Y}$ is Newton differentiable in $\mathcal{U}$ containing u^*, with Newton derivative G. If $G(u)$ is nonsingular for all $u \in \mathcal{U}$ and $\{\|G(u)^{-1}\|_{\mathcal{L}(\mathcal{Y},\mathcal{X})}, u \in \mathcal{U}\}$ is bounded, then

$$u_{n+1} = u_n - G(u_n)^{-1} F(u_n)$$

converges superlinearly to u^*, if $\|u_0 - u^*\|_{\mathcal{X}}$ is sufficiently small.

Regularization

- Problem: the generalized Jacobian of Φ is not invertible.
- Remedy: regularization of Φ by means of

$$\Phi^\varepsilon(u,y,p,\lambda) := \begin{pmatrix} T^\varepsilon(u, L_u(u,y,p)+\lambda) \\ L_y(u,y,p) \\ L_p(u,y,p) \\ \langle 1,u\rangle - m \end{pmatrix}.$$

- Examples of regularization:

$$T^\varepsilon(u,g) = u\max(0, g+\varepsilon) + (1-u)\min(0, g-\varepsilon),$$
$$T^\varepsilon(u,g) = \sqrt{\varepsilon^2+g^2} + \min(0,g).$$

- $D\Phi^\varepsilon(u,y,p,\lambda)$ is invertible and $\|D\Phi^\varepsilon(u,y,p,\lambda)^{-1}\|$ is uniformly bounded for (u,y,p,λ) close to the solution.

Semilinear problem

- $(\mathcal{P})$: $\min_{(u,y)\in U_{ad}\times Y} J(u,y) \qquad \text{subject to } E(u,y)=0,$

$$J(u,y) = \frac{1}{2}\int_D (y - y^\dagger)^2,$$
$$E(u,y) = Ay + \psi(y) - u,$$

  where $y^\dagger \in L^2(D)$, $Y = H_0^1(D)$, $Z = H^{-1}(D)$, $A = -\Delta$, and $\psi \in \mathcal{C}^3(\mathbb{R})$, non-decreasing, and such that

$$\|\psi^{(k)}\|_{L^\infty} \leq M_\psi^k, \qquad k = 1, 2, 3,$$

  for some positive constants M_ψ^k.
- D is of class $\mathcal{C}^2$ or convex.
- For all $u \in L^2(D)$, there exists a unique solution $y(u) \in H^2(D) \cap H_0^1(D)$ to the equation $E(u, y(u)) = 0$.

Existence of Regularized solutions

(A): Assumption on p

$\exists\gamma > 0$ such that, for all $(u, y, p) \in U \times H_0^1(D) \times H_0^1(D)$ satisfying

$$Ay + \psi(y) = u,$$
$$[A + \psi'(y)]p = -(y - y^\dagger),$$

there holds

$$1 + \psi''(y)p \geq \gamma.$$

This assumption is fulfilled if M_ψ^2 is small enough.

Theorem (Existence of solutions)

Let Assumption (A) hold. For each $\varepsilon > 0$ there exists

$$(u_\varepsilon, y_\varepsilon, p_\varepsilon, \lambda_\varepsilon) \in L^2(D, [0, 1]) \times (H^2 \cap H_0^1)(D) \times (H^2 \cap H_0^1)(D) \times \mathbb{R}$$

such that $\Phi^\varepsilon(u_\varepsilon, y_\varepsilon, p_\varepsilon, \lambda_\varepsilon) = 0$.

Convergence of the Newton algorithm

Theorem

Assume Assumption (A) holds and $\Phi^\varepsilon(\zeta^\varepsilon) = 0$ *with* $\zeta^\varepsilon = (u_\varepsilon, y_\varepsilon, p_\varepsilon, \lambda_\varepsilon)$. *Then*

$$\zeta_{n+1} = \zeta_n - D\Phi^\varepsilon(\zeta_n)^{-1}\Phi^\varepsilon(\zeta_n)$$

is well-defined and converges superlinearly to ζ^ε *as long as* $\|\zeta_0 - \zeta^\varepsilon\|$ *is sufficiently small.*

Proof

The main tasks are:

- the invertibility of the generalized gradient $D\Phi^\varepsilon(\zeta)$,
- a uniform bound on $\|D\Phi^\varepsilon(\zeta)^{-1}\|$ in an appropriate norm.

Convergence of the regularized solution

Theorem

Let $\{\varepsilon_k\}_{k\in\mathbb{N}}$, ε_k positive, $\varepsilon_k \to 0$ and $\Phi^{\varepsilon_k}(\zeta^{\varepsilon_k}) = 0$.

- *For any $s < 2$ there exists a subsequence $\{\varepsilon_{k_l}\}_{l\in\mathbb{N}}$ and $(u^*, \lambda^*) \in L^2(D, [0,1]) \times \mathbb{R}$ such that*

$$u^{\varepsilon_{k_l}} \rightharpoonup u^* \text{ weakly in } L^2(D), \qquad y^{\varepsilon_{k_l}} \to y^* \text{ strongly in } H^s(D),$$
$$p^{\varepsilon_{k_l}} \to p^* \text{ strongly in } H^s(D), \qquad \lambda^{\varepsilon_{k_l}} \to \lambda^* \text{ in } \mathbb{R},$$

  where y^*, p^* are given by

$$Ay^* + \psi(y^*) = u^*,$$
$$B(y^*)p^* = -(y^* - y^\dagger).$$

- *Every cluster point ζ^* (for $s < 2$ large enough) satisfies*

$$\Phi(\zeta^*) = 0.$$

Numerical method

- Ensure a constant rate of convergence of the merit function

$$R(\varepsilon) = \frac{1}{2}\|\Phi(\zeta^{\varepsilon})\|^2.$$

- Look for a sequence $\{\varepsilon_k\}$ such that

$$\frac{R(\varepsilon_{k+1})}{R(\varepsilon_k)} \approx \tau \quad \text{with } 0 < \tau < 1.$$

- Define $\Psi(\varepsilon, \zeta) = \Phi^{\varepsilon}(\zeta)$. Update

$$\varepsilon_{k+1} = \varepsilon_k \tau^{\beta_k},$$
$$\beta_k = \frac{-R(\varepsilon_k)}{\varepsilon_k \langle D\Phi(\zeta^{\varepsilon_k}) D\Phi^{\varepsilon_k}(\zeta^{\varepsilon_k})^{-1} D_{\varepsilon}\Psi(\varepsilon_k, \zeta^{\varepsilon_k}), \Phi(\zeta^{\varepsilon_k})\rangle}.$$

- Define $\rho(\ln \varepsilon) = \ln R(\varepsilon)$. Stopping criterions

$$\frac{\|\zeta_k - \zeta_{k-1}\|}{\|\zeta_{k-1}\|} < \kappa_N \text{ and } \rho'(\ln \varepsilon) < \kappa_E.$$

Numerical experiments: linear case

- domain $D =]0,1[^2$, volume constraint $m = 0.5$, $n = 39601$ nodes,
- convergence rate for $R(\varepsilon)$: $\tau = 0.1$
- stopping criterions: $\kappa_N = 10^{-8}$, $\kappa_E = 10^{-3}$
- $y_1^\dagger \equiv 0.01$ and $y_2^\dagger(x_1, x_2) = \sin(2\pi x_1)\sin(2\pi x_2)$
- CPU time: 5 minutes (Matlab-Desktop PC)

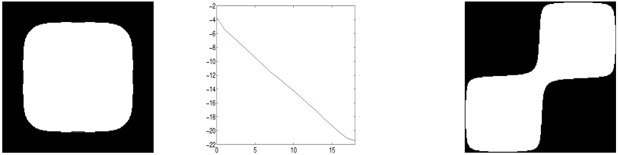

Figure: optimal control for $y^\dagger = y_1^\dagger$ (left), convergence history of $\log_{10} R(\varepsilon)$ for $y^\dagger = y_1^\dagger$ (middle), and optimal control for $y^\dagger = y_2^\dagger$ (right).

Numerical experiments: nonlinear case

- Fix $y^\dagger = 0.01$, and consider two functions ψ:

$$\psi_1(t) = e^{at} - 1, \qquad a = 10^3,$$
$$\psi_2(t) = \arctan(at), \qquad a = 10^2.$$

- ψ_2 satisfies our assumptions, but not ψ_1.
- appearance of intermediate regions.

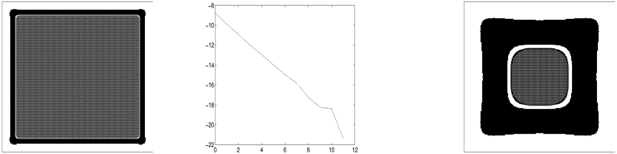

Figure: optimal control for $\psi = \psi_1$ (left), convergence history of $\log_{10} R(\varepsilon)$ for $\psi = \psi_1$ (middle), and optimal control for $\psi = \psi_2$ (right).

Thanks for your attention!