Seminar on Numerical Analysis

Peter Gangl

—

Semismooth Newton methods

12-01-2010

Recall:

- We can reformulate optimization problems as (possibly nonsmooth) operator equation

$$G(x) = 0, \qquad (1)$$

where $G : X \to Y, \quad X, Y$ Banach Spaces.

Recall:

- We can reformulate optimization problems as (possibly nonsmooth) operator equation

$$G(x) = 0, \tag{1}$$

where $G : X \to Y, \quad X, Y$ Banach Spaces.

- Equation (1) can be solved using some generalized Newton methods:
  1. Choose $x^0 \in X$
     For $k = 0, 1, 2, \ldots$
  2. Choose an invertible operator $M_k \in \mathcal{L}(X, Y)$.
  3. Obtain s^k by solving
     $$M_k s^k = -G(x^k)$$
     and set $x^{k+1} = x^k + s^k$.

Convergence results

- Let (x^k) be the sequence generated by the generalized Newton method where x^0 is sufficiently close to the solution $\bar{x} \in X$. Then we have:

Convergence results

- Let (x^k) be the sequence generated by the generalized Newton method where x^0 is sufficiently close to the solution $\bar{x} \in X$. Then we have:
  1. (x^k) converges q-linearly to $\bar{x}$ with rate $\gamma \in (0,1)$ iff

$$\|M_k^{-1}(G(\bar{x}+d^k) - G(\bar{x}) - M_k d^k)\|_X \leq \gamma \|d^k\|_X \qquad (2)$$

     $\forall k$ with $\|d^k\|_X$ sufficiently small.

Convergence results

- Let (x^k) be the sequence generated by the generalized Newton method where x^0 is sufficiently close to the solution $\bar{x} \in X$. Then we have:
  1. (x^k) converges q-linearly to $\bar{x}$ with rate $\gamma \in (0,1)$ iff
     $$\|M_k^{-1}(G(\bar{x}+d^k) - G(\bar{x}) - M_k d^k)\|_X \leq \gamma \|d^k\|_X \qquad (2)$$
     $\forall k$ with $\|d^k\|_X$ sufficiently small.
  2. (x^k) converges q-superlinearly to $\bar{x}$ iff
     $$\|M_k^{-1}(G(\bar{x}+d^k) - G(\bar{x}) - M_k d^k)\|_X = o(\|d^k\|_X) \qquad (3)$$
     for $\|d^k\|_X \to 0$

Convergence results

- Let (x^k) be the sequence generated by the generalized Newton method where x^0 is sufficiently close to the solution $\bar{x} \in X$. Then we have:
  1. (x^k) converges q-linearly to $\bar{x}$ with rate $\gamma \in (0,1)$ iff
  $$\|M_k^{-1}(G(\bar{x}+d^k) - G(\bar{x}) - M_k d^k)\|_X \leq \gamma \|d^k\|_X \tag{2}$$
  $\forall k$ with $\|d^k\|_X$ sufficiently small.
  2. (x^k) converges q-superlinearly to $\bar{x}$ iff
  $$\|M_k^{-1}(G(\bar{x}+d^k) - G(\bar{x}) - M_k d^k)\|_X = o(\|d^k\|_X) \tag{3}$$
  for $\|d^k\|_X \to 0$
  3. (x^k) converges q-order $1+\alpha > 1$ to $\bar{x}$ iff
  $$\|M_k^{-1}(G(\bar{x}+d^k) - G(\bar{x}) - M_k d^k)\|_X = O(\|d^k\|_X^{1+\alpha}) \tag{4}$$
  for $\|d^k\|_X \to 0$

Convergence results

The estimates (2), (3) are meant uniformly in k.

Convergence results

The estimates (2), (3) are meant uniformly in k.

It is convenient to split the smallness assumptions on

$$\|M_k^{-1}(G(\bar{x}+d^k)-G(\bar{x})-M_k d^k)\|_X$$

in two parts:

Convergence results

The estimates (2), (3) are meant uniformly in k.

It is convenient to split the smallness assumptions on

$$\|M_k^{-1}(G(\bar{x}+d^k)-G(\bar{x})-M_k d^k)\|_X$$

in two parts:

1. Regularity condition:

$$\|M_k^{-1}\|_{Y\to X} \leq C, \quad \forall k \geq 0. \tag{5}$$

Convergence results

The estimates (2), (3) are meant uniformly in k.

It is convenient to split the smallness assumptions on

$$\|M_k^{-1}(G(\bar{x}+d^k)-G(\bar{x})-M_k d^k)\|_X$$

in two parts:

1. Regularity condition:

$$\|M_k^{-1}\|_{Y\to X} \leq C, \quad \forall k \geq 0. \tag{5}$$

2. Approximation condition:

$$\|G(\bar{x}+d^k)-G(\bar{x})-M_k d^k\|_Y = o(\|d^k\|_X) \text{ for } \|d\|_X \to 0 \tag{6}$$

or

$$\|G(\bar{x}+d^k)-G(\bar{x})-M_k d^k\|_Y = O(\|d^k\|_X^{1+\alpha}) \text{ for } \|d\|_X \to 0 \tag{7}$$

Convergence results

Theorem 1. Consider the operator equation (1) with $G : X \to Y$, where X and Y are Banach spaces. Let (x^k) be generated by the generalized Newton method. Then:

Convergence results

Theorem 1. Consider the operator equation (1) with $G : X \to Y$, where X and Y are Banach spaces. Let (x^k) be generated by the generalized Newton method. Then:

1. If x^0 is sufficiently close to $\bar{x}$ and (2) holds then $x^k \to \bar{x}$ q-linearly with rate γ.

Convergence results

Theorem 1. Consider the operator equation (1) with $G : X \to Y$, where X and Y are Banach spaces. Let (x^k) be generated by the generalized Newton method. Then:

1. If x^0 is sufficiently close to $\bar{x}$ and (2) holds then $x^k \to \bar{x}$ q-linearly with rate γ.
2. If x^0 is sufficiently close to $\bar{x}$ and (3) (or (5) and (6)) holds then $x^k \to \bar{x}$ q-superlinearly.

Convergence results

Theorem 1. Consider the operator equation (1) with $G : X \to Y$, where X and Y are Banach spaces. Let (x^k) be generated by the generalized Newton method. Then:

1. If x^0 is sufficiently close to $\bar{x}$ and (2) holds then $x^k \to \bar{x}$ q-linearly with rate γ.
2. If x^0 is sufficiently close to $\bar{x}$ and (3) (or (5) and (6)) holds then $x^k \to \bar{x}$ q-superlinearly.
3. If x^0 is sufficiently close to $\bar{x}$ and (4) (or (5) and (7)) holds then $x^k \to \bar{x}$ q-superlinearly with order $1 + \alpha$.

Question: How to choose $M_k \in \mathcal{L}(X, Y)$ in the algorithm?

Question: How to choose $M_k \in \mathcal{L}(X, Y)$ in the algorithm?

- If G is continously F-differentiable and G' is continously invertible at the solution $\bar{x}$:

  Choose $M_k = G'(x^k)$ for all k ($\rightarrow$Classical Newton's Method)

Question: How to choose $M_k \in \mathcal{L}(X, Y)$ in the algorithm?

- If G is continously F-differentiable and G' is continously invertible at the solution $\bar{x}$:

  Choose $M_k = G'(x^k)$ for all k ($\rightarrow$Classical Newton's Method)
- If G is nonsmooth:

  Find a suitable substitute for G'($\rightarrow$ **Semismooth Newton's Method**)

Generalized Differential and Semismoothness

Considering set-valued generalized differentials $\partial G : X \rightrightarrows \mathcal{L}(X, Y)$, we will choose M_k point-based in each step, i.e.

$$M_k \in \partial G(x^k).$$

Generalized Differential and Semismoothness

Considering set-valued generalized differentials $\partial G : X \rightrightarrows \mathcal{L}(X, Y)$, we will choose M_k point-based in each step, i.e.

$$M_k \in \partial G(x^k).$$

Since we want M_k to satisfy the superlinear approximation condition (6), we have to require

$$\sup_{M \in \partial G(\bar{x}+d)} \|G(\bar{x}+d) - G(\bar{x}) - Md\|_Y = o(\|d\|_X) \text{ for } \|d\|_X \to 0.$$

This leads us to the definition of semismoothness:

Definition 1. (Semismoothness) Let X, Y be Banach spaces, $G : X \to Y$ a continuous operator and let be given the set-valued mapping $\partial G : X \rightrightarrows \mathcal{L}(X, Y)$ with nonempty images. Then

Definition 1. (Semismoothness) Let X, Y be Banach spaces, $G : X \to Y$ a continuous operator and let be given the set-valued mapping $\partial G : X \rightrightarrows \mathcal{L}(X, Y)$ with nonempty images. Then

1. G is called ∂G-semismooth at $x \in X$ if

$$\sup_{M \in \partial G(x+d)} \|G(x+d) - G(x) - Md\|_Y = o(\|d\|_X) \text{ for } \|d\|_X \to 0. \tag{8}$$

Definition 1. (Semismoothness) Let X, Y be Banach spaces, $G : X \to Y$ a continuous operator and let be given the set-valued mapping $\partial G : X \rightrightarrows \mathcal{L}(X, Y)$ with nonempty images. Then

1. G is called ∂G-semismooth at $x \in X$ if

$$\sup_{M \in \partial G(x+d)} \|G(x+d) - G(x) - Md\|_Y = o(\|d\|_X) \text{ for } \|d\|_X \to 0. \tag{8}$$

2. G is called ∂G-semismooth of order $\alpha > 0$ at $x \in X$ if

$$\sup_{M \in \partial G(x+d)} \|G(x+d) - G(x) - Md\|_Y = O(\|d\|_X^{1+\alpha}) \text{ for } \|d\|_X \to 0. \tag{9}$$

Lemma 1 Let $G : X \to Y$ be continuously F-differentiable near $x \in X$ and $\{G'\}$ the set-valued operator $\{G'\} : X \rightrightarrows \mathcal{L}(X, Y)$ mapping x to the one element set $\{G'(x)\}$.

Lemma 1 Let $G : X \to Y$ be continuously F-differentiable near $x \in X$ and $\{G'\}$ the set-valued operator $\{G'\} : X \rightrightarrows \mathcal{L}(X, Y)$ mapping x to the one element set $\{G'(x)\}$.

- Then G is $\{G'\}$-semismooth at x.

Lemma 1 Let $G : X \to Y$ be continuously F-differentiable near $x \in X$ and $\{G'\}$ the set-valued operator $\{G'\} : X \rightrightarrows \mathcal{L}(X, Y)$ mapping x to the one element set $\{G'(x)\}$.

- Then G is $\{G'\}$-semismooth at x.
- Furthermore, if G' is α-order Hoelder continuous near x, then G is $\{G'\}$-semismooth at x of order α.

Lemma 1 Let $G : X \to Y$ be continuously F-differentiable near $x \in X$ and $\{G'\}$ the set-valued operator $\{G'\} : X \rightrightarrows \mathcal{L}(X, Y)$ mapping x to the one element set $\{G'(x)\}$.

- Then G is $\{G'\}$-semismooth at x.
- Furthermore, if G' is α-order Hoelder continuous near x, then G is $\{G'\}$-semismooth at x of order α.

Proof: Blackboard

Theorem 2 Let X, Y, Z, X_i, Y_i be Banach spaces.

Theorem 2 Let X, Y, Z, X_i, Y_i be Banach spaces.

1. If the operators $G_i : X \to Y_i$ are ∂G_i-semismooth at x then (G_1, G_2) is $(\partial G_1, \partial G_2)$-semismooth at x.

Theorem 2 Let X, Y, Z, X_i, Y_i be Banach spaces.

1. If the operators $G_i : X \to Y_i$ are ∂G_i-semismooth at x then (G_1, G_2) is $(\partial G_1, \partial G_2)$-semismooth at x.
2. If $G_i : X \to Y$, $i = 1, 2$, are ∂G_i-semismooth at x then $G_1 + G_2$ is $(\partial G_1 + \partial G_2)$-semismooth at x.

Theorem 2 Let X, Y, Z, X_i, Y_i be Banach spaces.

1. If the operators $G_i : X \to Y_i$ are ∂G_i-semismooth at x then (G_1, G_2) is $(\partial G_1, \partial G_2)$-semismooth at x.
2. If $G_i : X \to Y$, $i = 1, 2$, are ∂G_i-semismooth at x then $G_1 + G_2$ is $(\partial G_1 + \partial G_2)$-semismooth at x.
3. Let $G_1 : Y \to Z$ and $G_2 : X \to Y$ be ∂G_i-semismooth at $G_2(x)$ and x, respectively. Assume that ∂G_1 is bounded near $y = G_2(x)$ and that G_2 is Lipschitz continuous near x. Then $G = G_1 \circ G_2$ is ∂G-semismooth with

$$\partial G(x) = \{M_1 M_2 : M_1 \in \partial G_1(G_2(x)), M_2 \in \partial G_2(x)\}.$$

Theorem 2 Let X, Y, Z, X_i, Y_i be Banach spaces.

1. If the operators $G_i : X \to Y_i$ are ∂G_i-semismooth at x then (G_1, G_2) is $(\partial G_1, \partial G_2)$-semismooth at x.
2. If $G_i : X \to Y$, $i = 1, 2$, are ∂G_i-semismooth at x then $G_1 + G_2$ is $(\partial G_1 + \partial G_2)$-semismooth at x.
3. Let $G_1 : Y \to Z$ and $G_2 : X \to Y$ be ∂G_i-semismooth at $G_2(x)$ and x, respectively. Assume that ∂G_1 is bounded near $y = G_2(x)$ and that G_2 is Lipschitz continuous near x. Then $G = G_1 \circ G_2$ is ∂G-semismooth with

$$\partial G(x) = \{M_1 M_2 : M_1 \in \partial G_1(G_2(x)), M_2 \in \partial G_2(x)\}.$$

Proof: Blackboard

Semismooth Newton methods

In addition to the superlinear approximation condition (6) (or (7)), the operator M_k of the generalized Newton method should also satisfy the regularity condition (5). Therefore we require the following:

$$\exists C > 0 \exists \delta > 0 : \|M^{-1}\|_{Y \to X} \leq C \quad \forall M \in \partial G(x) \quad \forall x \in X, \|x - \bar{x}\|_X < \delta. \tag{10}$$

Semismooth Newton methods

In addition to the superlinear approximation condition (6) (or (7)), the operator M_k of the generalized Newton method should also satisfy the regularity condition (5). Therefore we require the following:

$$\exists C > 0 \exists \delta > 0 : \|M^{-1}\|_{Y \to X} \leq C \quad \forall M \in \partial G(x) \quad \forall x \in X, \|x - \bar{x}\|_X < \delta. \tag{10}$$

Algorithm 1 (Semismooth Newton's method)

1. Choose $x^0 \in X$ (sufficiently close to the solution $\bar{x}$).
   For $k = 0, 1, 2, \ldots$:
2. Choose $M_k \in \partial G(x^k)$.
3. Obtain s^k by solving

$$M_k s^k = -G(x^k),$$

and set $x^{k+1} = x^k + s^k$.

Corollary 1 Let $G : X \to Y$ be continuous and ∂G-semismooth at a solution $\bar{x}$ of (1). Furthermore, assume that the regularity condition (10) holds. Then there exists $\delta > 0$ such that for all $x^0 \in X$, $\|x^0 - \bar{x}\|_X < \delta$, the semismooth Newton method (Algorithm 1) converges q-superlinearly to $\bar{x}$.

If G is ∂G-semismooth of order $\alpha > 0$ at $\bar{x}$, then the convergence is of order $1 + \alpha$.

Finite Dimensional Example

For locally Lipschitz-continuous functions $G : \mathbf{R}^n \to \mathbf{R}^m$, the standard choice for ∂G is Clarke's generalized Jacobian

$$\begin{aligned} \partial^{cl} G : \mathbf{R}^n &\rightrightarrows \mathbf{R}^{m \times n} \\ \partial^{cl} G(x) &= \mathrm{conv}\{M \in \mathbf{R}^{m \times n} : \exists (x^k) \to x, G \text{ differentiable at } x^k : \\ &\quad G'(x^k) \to M\}. \end{aligned} \tag{11}$$

Finite Dimensional Example

For locally Lipschitz-continuous functions $G : \mathbf{R}^n \to \mathbf{R}^m$, the standard choice for ∂G is Clarke's generalized Jacobian

$$\begin{aligned} \partial^{cl} G : \mathbf{R}^n &\rightrightarrows \mathbf{R}^{m \times n} \\ \partial^{cl} G(x) &= \mathrm{conv}\{M \in \mathbf{R}^{m \times n} : \exists (x^k) \to x, G \text{ differentiable at } x^k : \\ &\quad G'(x^k) \to M\}. \end{aligned} \tag{11}$$

Example 1 Consider $\psi : \mathbf{R} \to \mathbf{R}$, $\psi(x) = P_{[a,b]}(x)$, $a < b$, then Clarke's generalized derivative is

$$\partial^{cl} \psi(x) = \begin{cases} \{0\} & x < a \text{ or } x > b, \\ \{1\} & a < x < b, \\ \mathrm{conv}\{0, 1\} = [0, 1] & x = a \text{ or } x = b. \end{cases}$$

Example in Function Spaces (1)

Let $\Omega \subset \mathbf{R}^n$ be measurable, $f : L^2(\Omega) \to \mathbf{R}$ twice continuously F-differentiable, a, $b \in L^\infty(\Omega)$ with $b(x) > a(x)\ \forall x \in \Omega$.
We consider the problem

$$\min_{u \in L^2(\Omega)} f(u), \quad a \leq u \leq b \quad \text{a.e. on } \Omega$$

Example in Function Spaces (1)

Let $\Omega \subset \mathbf{R}^n$ be measurable, $f : L^2(\Omega) \to \mathbf{R}$ twice continuously F-differentiable, $a, b \in L^\infty(\Omega)$ with $b(x) > a(x)\ \forall x \in \Omega$.
We consider the problem

$$\min_{u \in L^2(\Omega)} f(u), \quad a \leq u \leq b \quad \text{a.e. on } \Omega$$

We can transform the bounds to constant bounds using $u \mapsto \frac{u-a}{b-a}$, so that we consider the problem

$$\min_{u \in L^2(\Omega)} f(u), \quad \beta_l \leq u \leq \beta_r \quad \text{a.e. on } \Omega \tag{12}$$

Example in Function Spaces (1)

Let $\Omega \subset \mathbf{R}^n$ be measurable, $f : L^2(\Omega) \to \mathbf{R}$ twice continuously F-differentiable, $a, b \in L^\infty(\Omega)$ with $b(x) > a(x) \; \forall x \in \Omega$.
We consider the problem

$$\min_{u \in L^2(\Omega)} f(u), \quad a \leq u \leq b \quad \text{a.e. on } \Omega$$

We can transform the bounds to constant bounds using $u \mapsto \frac{u-a}{b-a}$, so that we consider the problem

$$\min_{u \in L^2(\Omega)} f(u), \quad \beta_l \leq u \leq \beta_r \quad \text{a.e. on } \Omega \tag{12}$$

Note: Many PDE-constrained optimal control problems lead to this setting (see last talk).

Example in Function Spaces (2)

If we set $S = \{u \in L^2(\Omega) : \beta_l \leq u \leq \beta_r\}$, problem (12) is equivalent to the variational inequality

$$u \in S, \quad (\nabla f(u), v - u)_{L^2(\Omega)} \geq 0 \ \ \forall v \in S. \tag{13}$$

Example in Function Spaces (2)

If we set $S = \{u \in L^2(\Omega) : \beta_l \leq u \leq \beta_r\}$, problem (12) is equivalent to the variational inequality

$$u \in S, \quad (\nabla f(u), v - u)_{L^2(\Omega)} \geq 0 \;\; \forall v \in S. \tag{13}$$

Using the L^2-projection P_S given by

$$P_S(v)(x) = P_{[\beta_l, \beta_r]}(v(x)), \;\; x \in \Omega,$$

we can reformulate (13) as the nonsmooth operator equation

$$\Phi(u) := u - P_S(u - \theta \nabla f(u)) = 0, \tag{14}$$

where $\theta > 0$ is arbitrary, but fixed.

Example in Function Spaces (3)

Aim: Define a generalized differential $\partial\Phi$ for Φ in such a way that Φ is $\partial\Phi$-semismooth.

Example in Function Spaces (3)

Aim: Define a generalized differential $\partial\Phi$ for Φ in such a way that Φ is $\partial\Phi$-semismooth.

By the sum rule we developed in *Theorem 2*, this reduces to the question how a suitable differential for the superposition $P_{[\beta_l,\beta_r]}(v(\cdot))$ can be defined.

Example in Function Spaces (3)

Aim: Define a generalized differential $\partial\Phi$ for Φ in such a way that Φ is $\partial\Phi$-semismooth.

By the sum rule we developed in *Theorem 2*, this reduces to the question how a suitable differential for the superposition $P_{[\beta_l,\beta_r]}(v(\cdot))$ can be defined.

Therefore we can use a result proved by M. Ulbrich, for the formulation of which we need a definition of generalized differentials of superposition operators of the form $\psi(G(\cdot))$, where G is a continuously F-differentiable operator.

Example in Function Spaces (4)

Definition 2 Let $\psi : \mathbf{R}^m \to \mathbf{R}$ be Lipschitz continuous and $(\partial^{cl}\psi)$-semismooth. Furthermore, let $1 \leq q \leq p \leq \infty$ be given, consider

$$\Psi_G : U \to L^q(\Omega), \;\; \Psi_G(u)(x) = \psi(G(u)(x)),$$

where $G : U \to L^p(\Omega)^m$ is continuously F-differentiable and U is a Banach space. We define the differential

$$\begin{aligned} \partial\Psi_G : U &\rightrightarrows \mathcal{L}(U, L^q(\Omega)), \\ \partial\Psi_G(u) &= \{M : Mv = g^T(G'(u)v), g \in L^\infty(\Omega)^m, \\ &\qquad g(x) \in \partial^{cl}\psi(G(u)(x)) \;\; \dot{\forall} x \in \Omega\}. \end{aligned} \quad (15)$$

Example in Function Spaces (4)

Definition 2 Let $\psi : \mathbf{R}^m \to \mathbf{R}$ be Lipschitz continuous and $(\partial^{cl}\psi)$-semismooth. Furthermore, let $1 \leq q \leq p \leq \infty$ be given, consider

$$\Psi_G : U \to L^q(\Omega), \quad \Psi_G(u)(x) = \psi(G(u)(x)),$$

where $G : U \to L^p(\Omega)^m$ is continuously F-differentiable and U is a Banach space. We define the differential

$$\begin{aligned} \partial\Psi_G : U &\rightrightarrows \mathcal{L}(U, L^q(\Omega)), \\ \partial\Psi_G(u) &= \{M : Mv = g^T(G'(u)v), g \in L^\infty(\Omega)^m, \\ &\quad g(x) \in \partial^{cl}\psi(G(u)(x)) \ \dot{\forall} x \in \Omega\}. \end{aligned} \tag{15}$$

Note: This is just the differential that we would obtain by the construction in part (3) of *Theorem 2*.

Example in Function Spaces (5)

Now we can state the following semismoothness result:

Example in Function Spaces (5)

Now we can state the following semismoothness result:

Theorem 3 Let $\Omega \subset \mathbf{R}^n$ be measurable with $0 < |\Omega| < \infty$. Furthermore, let $\psi : \mathbf{R}^m \to \mathbf{R}$ be Lipschitz continuous and $\partial^{cl}\psi$-semismooth. Let U be a Banach space, $1 \leq q < p \leq \infty$, and assume that the operator $G : U \to L^q(\Omega)^m$ is continuously F-differentiable and that G maps U locally Lipschitz continuously to $L^p(\Omega)^m$. Then, the operator

$$\Psi_G : U \to L^q(\Omega), \ \ \Psi_G(u)(x) = \psi(G(u)(x)),$$

is $\partial\Psi_G$-semismooth, where $\partial\Psi_G$ is defined in (15).

Example in Function Spaces (5)

Now we can state the following semismoothness result:

Theorem 3 Let $\Omega \subset \mathbf{R}^n$ be measurable with $0 < |\Omega| < \infty$. Furthermore, let $\psi : \mathbf{R}^m \to \mathbf{R}$ be Lipschitz continuous and $\partial^{cl}\psi$-semismooth. Let U be a Banach space, $1 \leq q < p \leq \infty$, and assume that the operator $G : U \to L^q(\Omega)^m$ is continuously F-differentiable and that G maps U locally Lipschitz continuously to $L^p(\Omega)^m$. Then, the operator

$$\Psi_G : U \to L^q(\Omega), \;\; \Psi_G(u)(x) = \psi(G(u)(x)),$$

is $\partial\Psi_G$-semismooth, where $\partial\Psi_G$ is defined in (15).

Proof M. Ulbrich: "Nonsmooth Newton-like methods for variational inequalities and constrained optimization problems in function spaces" (2001)

Example in Function Spaces (6)

In order to be able to apply this result to the second summand of our problem

$$\begin{aligned} \Psi_G : U = L^2(\Omega) &\rightarrow L^2(\Omega), \\ \Psi_G(u)(x) &= \psi(G(u)(x)) = P_{[\beta_l, \beta_r]}((u - \theta \nabla f(u))(x)), \end{aligned}$$

we have to make some assumptions on the structure of ∇f (which are fulfilled by many optimal control problems):

Example in Function Spaces (6)

In order to be able to apply this result to the second summand of our problem

$$\begin{aligned} \Psi_G : U = L^2(\Omega) &\rightarrow L^2(\Omega), \\ \Psi_G(u)(x) &= \psi(G(u)(x)) = P_{[\beta_l,\beta_r]}((u - \theta\nabla f(u))(x)), \end{aligned}$$

we have to make some assumptions on the structure of ∇f (which are fulfilled by many optimal control problems):

There exist $\alpha > 0$ and $p > 2$ sucht that

- $\nabla f(u) = \alpha u + H(u)$,
- $H : L^2(\Omega) \rightarrow L^2(\Omega)$ continuously F-differentiable,
- $H : L^2(\Omega) \rightarrow L^p(\Omega)$ locally Lipschitz continuous.

Example in Function Spaces (7)

Under these assumptions and by setting $\theta = \frac{1}{\alpha}$, Ψ_G reduces to

$$\Psi_G(u)(x) = \psi(G(u)(x)) = P_{[\beta_l,\beta_r]}(-\frac{1}{\alpha}H(u)(x)),$$

so setting $q = 2, \psi = P_{[\beta_l,\beta_r]}$ and $G = -\frac{1}{\alpha}H$, we can apply *Theorem 3* and obtain that the operator Ψ_G is $\partial\Psi_G$-semismooth with $\partial\Psi_G$ defined in (15).

Example in Function Spaces (7)

Under these assumptions and by setting $\theta = \frac{1}{\alpha}$, Ψ_G reduces to

$$\Psi_G(u)(x) = \psi(G(u)(x)) = P_{[\beta_l,\beta_r]}(-\frac{1}{\alpha}H(u)(x)),$$

so setting $q = 2, \psi = P_{[\beta_l,\beta_r]}$ and $G = -\frac{1}{\alpha}H$, we can apply *Theorem 3* and obtain that the operator Ψ_G is $\partial\Psi_G$-semismooth with $\partial\Psi_G$ defined in (15).

Therefore, by *Theorem 2*, the operator $\Phi = I - \Psi_G$ in our problem (14) is semismooth w.r.t $\partial\Phi = I - \partial\Psi_G$ defined by

$$\begin{aligned} \partial\Phi : L^2(\Omega) &\rightrightarrows \mathcal{L}(L^2(\Omega), L^2(\Omega)), \\ \partial\Phi(u) &= \{M : M = I + \frac{g}{\alpha} \cdot H'(u), g \in L^\infty(\Omega), \\ &\quad g(x) \in \partial^{cl} P_{[\beta_l,\beta_r]}(-(1/\alpha)H(u)(x)) \dot\forall x \in \Omega\} \end{aligned} \tag{16}$$

and (under some additional regularity condition) we can apply the semismooth Newton method (*Algorithm 1*).

Application to Optimal Control (1)

Now we can apply the theory we just developed to the following elliptic optimal control problem:

$$\begin{aligned} \min_{y \in H_0^1(\Omega), u \in L^2(\Omega)} J(y,u) &\stackrel{def}{=} \frac{1}{2}\|y - y_d\|^2_{L^2(\Omega)} + \frac{\alpha}{2}\|u\|^2_{L^2(\Omega)} \\ s.t. \;\; Ay &= r + Bu, \;\; \beta_l \le u \le \beta_r. \end{aligned}$$

with

Application to Optimal Control (1)

Now we can apply the theory we just developed to the following elliptic optimal control problem:

$$\min_{y \in H_0^1(\Omega), u \in L^2(\Omega)} J(y,u) \stackrel{def}{=} \frac{1}{2}\|y - y_d\|^2_{L^2(\Omega)} + \frac{\alpha}{2}\|u\|^2_{L^2(\Omega)}$$
$$s.t. \quad Ay = r + Bu, \quad \beta_l \le u \le \beta_r.$$

with

- $y \in H_0^1(\Omega)$ the *state*

Application to Optimal Control (1)

Now we can apply the theory we just developed to the following elliptic optimal control problem:

$$\min_{y\in H_0^1(\Omega), u\in L^2(\Omega)} J(y,u) \overset{def}{=} \frac{1}{2}\|y-y_d\|^2_{L^2(\Omega)} + \frac{\alpha}{2}\|u\|^2_{L^2(\Omega)}$$

$$s.t. \quad Ay = r + Bu, \quad \beta_l \le u \le \beta_r .$$

with

- $y \in H_0^1(\Omega)$ the *state*
- $u \in L^2(\Omega)$ the *control*

Application to Optimal Control (1)

Now we can apply the theory we just developed to the following elliptic optimal control problem:

$$\min_{y\in H_0^1(\Omega),u\in L^2(\Omega)} J(y,u) \stackrel{def}{=} \frac{1}{2}\|y-y_d\|^2_{L^2(\Omega)} + \frac{\alpha}{2}\|u\|^2_{L^2(\Omega)}$$
$$s.t. \quad Ay = r + Bu, \quad \beta_l \le u \le \beta_r.$$

with

- $y \in H_0^1(\Omega)$ the *state*
- $u \in L^2(\Omega)$ the *control*
- $A : H_0^1(\Omega) \to H^{-1}(\Omega) = H_0^1(\Omega)^*$ a (for simplicity) linear elliptic partial differential operator (e.g. $A = -\triangle$)

Application to Optimal Control (1)

Now we can apply the theory we just developed to the following elliptic optimal control problem:

$$\min_{y\in H_0^1(\Omega),u\in L^2(\Omega)} J(y,u) \stackrel{def}{=} \frac{1}{2}\|y-y_d\|^2_{L^2(\Omega)} + \frac{\alpha}{2}\|u\|^2_{L^2(\Omega)}$$
$$s.t. \quad Ay = r + Bu, \quad \beta_l \le u \le \beta_r.$$

with

- $y \in H_0^1(\Omega)$ the *state*
- $u \in L^2(\Omega)$ the *control*
- $A : H_0^1(\Omega) \to H^{-1}(\Omega) = H_0^1(\Omega)^*$ a (for simplicity) linear elliptic partial differential operator (e.g. $A = -\triangle$)
- $r \in H^{-1}(\Omega)$ given

Application to Optimal Control (1)

Now we can apply the theory we just developed to the following elliptic optimal control problem:

$$\min_{y \in H_0^1(\Omega), u \in L^2(\Omega)} J(y,u) \overset{def}{=} \frac{1}{2}\|y - y_d\|^2_{L^2(\Omega)} + \frac{\alpha}{2}\|u\|^2_{L^2(\Omega)}$$
$$s.t. \quad Ay = r + Bu, \quad \beta_l \leq u \leq \beta_r.$$

with

- $y \in H_0^1(\Omega)$ the *state*
- $u \in L^2(\Omega)$ the *control*
- $A : H_0^1(\Omega) \to H^{-1}(\Omega) = H_0^1(\Omega)^*$ a (for simplicity) linear elliptic partial differential operator (e.g. $A = -\triangle$)
- $r \in H^{-1}(\Omega)$ given
- $B : L^{p'}(\Omega) \to H^{-1}(\Omega)$ the continuous and linear control operator with $p' \in [1, 2)$

Application to Optimal Control (1)

Now we can apply the theory we just developed to the following elliptic optimal control problem:

$$\begin{aligned} \min_{y \in H_0^1(\Omega), u \in L^2(\Omega)} J(y,u) &\overset{def}{=} \frac{1}{2}\|y - y_d\|^2_{L^2(\Omega)} + \frac{\alpha}{2}\|u\|^2_{L^2(\Omega)} \\ s.t. \;\; Ay &= r + Bu, \;\; \beta_l \le u \le \beta_r. \end{aligned}$$

with

- $y \in H_0^1(\Omega)$ the *state*
- $u \in L^2(\Omega)$ the *control*
- $A : H_0^1(\Omega) \to H^{-1}(\Omega) = H_0^1(\Omega)^*$ a (for simplicity) linear elliptic partial differential operator (e.g. $A = -\triangle$)
- $r \in H^{-1}(\Omega)$ given
- $B : L^{p'}(\Omega) \to H^{-1}(\Omega)$ the continuous and linear control operator with $p' \in [1, 2)$
- the constant bounds $\beta_l < \beta_r$

Application to Optimal Control (2)

By eliminating the state y via the state equation $y = y(u) = A^{-1}(r + Bu)$, we obtain a reduced problem, which is of the form (12):

$$\min_{u \in L^2(\Omega)} \quad \hat{J}(u) \stackrel{def}{=} J(y(u), u) \quad \stackrel{def}{=} \frac{1}{2}\|y(u) - y_d\|^2_{L^2(\Omega)} + \frac{\alpha}{2}\|u\|^2_{L^2(\Omega)}$$

$$s.t. \;\; \beta_l \leq u \leq \beta_r.$$

Application to Optimal Control (2)

By eliminating the state y via the state equation $y = y(u) = A^{-1}(r + Bu)$, we obtain a reduced problem, which is of the form (12):

$$\min_{u \in L^2(\Omega)} \quad \hat{J}(u) \overset{def}{=} J(y(u), u) \quad \overset{def}{=} \frac{1}{2}\|y(u) - y_d\|^2_{L^2(\Omega)} + \frac{\alpha}{2}\|u\|^2_{L^2(\Omega)}$$
$$s.t. \;\; \beta_l \leq u \leq \beta_r.$$

For the F-derivative $\nabla \hat{J}$ we obtain

$$\begin{aligned} \nabla \hat{J}(u) &= \alpha u + y'(u)^*(y(u) - y_d) \qquad (17) \\ &= \alpha u + B^*(A^{-1})^*(A^{-1}(r + Bu) - y_d) \overset{def}{=} \alpha u + H(u). \end{aligned}$$

Application to Optimal Control (3)

Since $B \in \mathcal{L}(L^{p'}(\Omega), H^{-1}(\Omega))$, we have $B^* \in \mathcal{L}(H_0^1(\Omega), L^p(\Omega))$ with $p = p'/(p-1) > 2$. Hence the affine linear operator $H(u)$ defined in (17) is a continuous affine linear mapping $L^2(\Omega) \to L^p(\Omega)$ and we can apply *Theorem 3* to rewrite the optimality conditions as a semismooth operator equation

$$\Phi(u) \overset{def}{=} u - P_{[\beta_l, \beta_r]}(-(1/\alpha)H(u)) = 0.$$

Application to Optimal Control (4)

Considering the generalized differential $\partial\Phi$ we developed in (16), the Newton system in *Algorithm 1* now reads

$$\left(I + \frac{1}{\alpha} g^k \cdot H'(u^k)\right) s^k = -\Phi(u^k), \tag{18}$$

and $g \cdot H'(u)$ stands for $v \mapsto g \cdot (H'(u)v)$ and $g^k \in L^\infty(\Omega)$ is chosen such that

$$g^k(x) = \begin{cases} = 0 & -(1/\alpha)H(u^k)(x) \notin [\beta_l, \beta_r], \\ = 1 & -(1/\alpha)H(u^k)(x) \in (\beta_l, \beta_r), \\ = [0,1] & -(1/\alpha)H(u^k)(x) \in \{\beta_l, \beta_r\}. \end{cases}$$

The operator on the left side has the form

$$M_k \stackrel{def}{=} I + \frac{1}{\alpha} g^k \cdot H'(u^k) = I + \frac{1}{\alpha} g^k \cdot B^*(a^{-1})^* A^{-1} B.$$

Application to Optimal Control (5)

For solving (18), it can be advantageous to note that s^k solves (18) if and only if $s^k = d_u^k$ and $(d_y^k, d_u^k, d_\mu^k)^T$ solves

$$\begin{pmatrix} I & 0 & A^* \\ 0 & I & -\frac{1}{\alpha} g^k \cdot B^* \\ A & -B & 0 \end{pmatrix} \begin{pmatrix} d_y^k \\ d_u^k \\ d_\mu^k \end{pmatrix} = \begin{pmatrix} 0 \\ -\Phi(u^k) \\ 0 \end{pmatrix}$$

to which multigrid methods can be applied.

References

1. M.Hinze, R.Pinnau, M.Ulbrich, S.Ulbrich. Optimization with PDE Constraints, Springer, New York (2009)
2. M. Ulbrich. Nonsmooth Newton-like Methods for Variational Inequalities and Constrained Optimization Problems in Function Spaces (Habilitation; 2001)
Available at:
http://www-m1.ma.tum.de/foswiki/pub/Lehrstuhl /PublikationenUlbrich/Habil.pdf