Proximal gradient methods applied to optimization problems with L^p-cost, $p \in [0, 1)$

Daniel Wachsmuth
joint work with Carolin Natemeyer

University of Würzburg

A workshop on Optimization and Operator Theory
Technion - 2021

Outline

The sparse control problem

Optimal control problem = minimization problem

$$\min f(u)$$

with controls

$$u : \Omega \to \mathbb{R}, \quad u \in L^2(\Omega)$$

on bounded set $\Omega \in \mathbb{R}^d$.

Sparse controls: Penalize measure of support of controls

$$\|u\|_0 := \text{meas}\{x \in \Omega : \; u(x) \neq 0\}$$

or some p-'norm' $\|u\|_p$ with $p \in (0, 1)$

$$\|u\|_p := \int_\Omega |u(x)|^p \, \mathrm{d}x.$$

Optimization problem

Let $p \in [0, 1)$.

Minimize

$$J(u) := f(u) + \frac{\alpha}{2}\|u\|^2_{L^2(\Omega)} + \beta\|u\|_p$$

over all $u \in L^2(\Omega)$ satisfying

$$u \in \mathcal{U}_{\text{ad}} := \{v \in L^2(\Omega) : |v| \le b \text{ a.e. on } \Omega\}.$$

Reference: [Ito-Kunisch '14][DW '19]

Example

$f(u) = \frac{1}{2}\|y(u) - y_d\|^2_{L^2(\Omega)}$ with $y(u)$ being the solution of a pde to control u.

Lower semicontinuity

- $x \mapsto |x|_p$ is lower semicontinuous on $\mathbb{R}$,
- $u \mapsto \|u\|_p$ is lower semicontinuous on $L^q(\Omega)$,
- $u \mapsto \|u\|_p$ is seq. weakly lower semicontinuous on ℓ^q,
- $u \mapsto \|u\|_p$ is *NOT* seq. weakly lower semicontinuous on $L^q(\Omega)$.

Example

Define $\Omega = (0,1)$ and

$$f_n(x) := 1 + \operatorname{sign}\sin(2n\pi x).$$

Then $f_n \rightharpoonup 1 =: f$ in $L^q(\Omega)$, $1 \leq q < \infty$, but

$$\|f_n\|_p = 2^{p-1} < 1, \quad \|f\|_p = 1.$$

Existence of solutions cannot be proven. No regularizing effect.

Outline

Finite-dimensional case

Let $f : \mathbb{R}^n \to \mathbb{R}$ be continuously differentiable. Consider

$$\min f(x) + \|x\|_p$$

with $\|x\|_p := \sum_i |x_i|_p$.

(1) The point $x = 0$ is always a local minimum.

(2) If $\bar{x}$ is locally optimal, then

$$\nabla f(\bar{x})_i + p\,\mathrm{sign}(\bar{x}_i)|\bar{x}_i|^{p-1} = 0 \quad \forall i:\ \bar{x}_i \neq 0.$$

No information on $\nabla f(\bar{x})_i$ if $\bar{x}_i = 0$.

(3) Analogous results hold for optimization in sequence spaces.

Pontryagin maximum principle

Theorem

Let $\bar{u}$ be a local solution of the L^p-problem. Then it holds

$$\bar{u}(x) = \arg\min_{|u|\le b} \nabla f(\bar{u})(x) \cdot u + \frac{\alpha}{2}|u|^2 + \beta|u|_p$$

for almost all $x \in \Omega$.

- Stronger optimality condition than in the finite-dimensional case.
- No derivative of $u \mapsto |u|_p$ appears.

Proof uses needle perturbations

$$d_n(x) = \chi_{B_{r_n}(x_0)}(x) \cdot (v - \bar{u}(x)) : \quad r_n \to 0,\ |v| \le b$$

and Lebesgue's differentiation theorem.

[Casas '94][Ito, Kunisch '14]

Optimality conditions - ideas of proof

Convex differentiable optimization:
Fix direction $u - \bar{u}$, compute

$$\lim_{t \searrow 0} \frac{f(\bar{u} + t(u - \bar{u})) - f(\bar{u})}{t}.$$

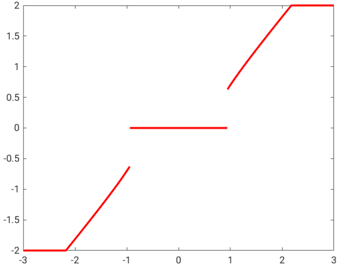
Pontryagin maximum principle:
Fix amplitude $v \in \mathbb{R}$, compute

$$\lim_{r \searrow 0} \frac{f(\bar{u} + \chi_{B_r(x)}(v - \bar{u})) - f(\bar{u})}{|B_r(x)|}.$$

Pontryagin maximum principle

The control $\bar{u}$ satisfies PMP if and only if

$$(-\nabla f(\bar{u})(x),\ \bar{u}(x)) \in P \quad \text{for a.a. } x$$



$P : \mathbb{R} \rightrightarrows \mathbb{R}$ is monotone with closed graph, but not maximal monotone.

Outline

Motivation - steepest descent

Consider

$$\min f(x).$$

Gradient descent step

$$x_{k+1} = x_k - t_k \nabla f(x_k)$$

is a solution of

$$\min f(x_k) + \nabla f(x_k)(x - x_k) + \frac{1}{2t_k}\|x - x_k\|^2.$$

Motivation - proximal gradient

Consider

$$\min f(x) + g(x).$$

Compute x_{k+1} as one solution of

$$\min f(x_k) + \nabla f(x_k)(x - x_k) + \frac{1}{2t_k}\|x - x_k\|^2 + g(x).$$

or equivalently as

$$x_{k+1} \in \mathrm{prox}_{t_k g}(x_k - t_k \nabla f(x_k))$$

Still involves the non-smooth function g.

Example

For $g = \|\cdot\|_1$ the resulting method is soft-thresholding applied to a gradient step.

Proximal gradient step - pointwise

Set $L := \frac{1}{t}$.

Determine u_{k+1} as a solution of

$$\min f(u_k) + \nabla f(u_k)(u - u_k) + \frac{L}{2}\|u - u_k\|^2_{L^2(\Omega)} + \frac{\alpha}{2}\|u\|^2_{L^2(\Omega)} + \beta\|u\|_p + I_{\mathcal{U}_{\text{ad}}}(u).$$

Proximal gradient step - pointwise

Set $L := \frac{1}{t}$.

Determine u_{k+1} as a solution of

$$\min f(u_k) + \nabla f(u_k)(u - u_k) + \frac{L}{2}\|u - u_k\|^2_{L^2(\Omega)} + \frac{\alpha}{2}\|u\|^2_{L^2(\Omega)} + \beta\|u\|_p + I_{\mathcal{U}_{\text{ad}}}(u).$$

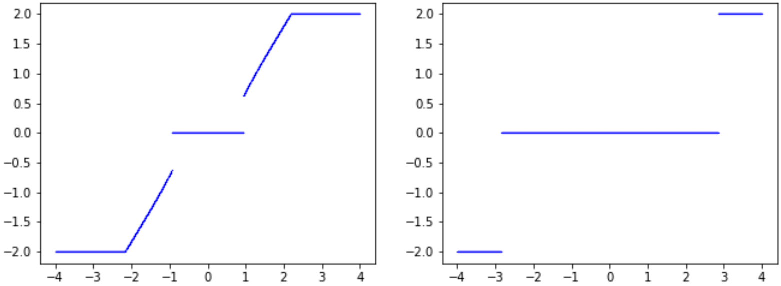
Pointwise minimization:

$$u_{k+1}(x) \in \operatorname{prox}_{L^{-1}g}\left(u_k(x) - L^{-1}\nabla f(u_k)(x)\right) \quad \text{for a.a. } x \in \Omega$$

where

$$g(u) := \frac{\alpha}{2}|u|^2 + \beta|u|_p + I_{[-b,b]}(u).$$

Proximal operator of g



Important property for $u \in \text{prox}_{L^{-1}g}(v)$:

$$\text{Either } u = 0 \text{ or } |u| \geq u_0$$

with $u_0 > 0$ depending on L.

Iteration methods based on PMP

PMP is equivalent to minimizing some Hamiltonian:

$$\bar{u}(x) = \arg\min_{|u|\leq b} H(\bar{y}(x), u, \bar{p}(x)).$$

Fixed-point method [Krylov, Černous′ko '62]

$$u_{k+1} := \arg\min_{|u|\leq b} H(y_k(x), u, p_k(x))$$

.. and with prox-term

$$u_{k+1} := \arg\min_{|u|\leq b} H(y_k(x), u, p_k(x)) + \frac{L_k}{2}(u - u_k(x))^2$$

[Sakawa, Shindo '80, Bonnans '86, Breitenbach, Borzi '19]

If control appears linearly in the state equation, this method is equivalent to the proximal gradient method.

Outline

Setting

(A1) $g(u) := \frac{\alpha}{2}|u|^2 + \beta|u|_p + I_{[-b,b]}(u)$

(A2) $f : L^2(\Omega) \to \mathbb{R}$ is weakly lower semicontinuous and bounded from below,

(A3) $f : L^2(\Omega) \to \mathbb{R}$ is Fréchet differentiable,

(A4) ∇f is Lipschitz continuous from $L^2(\Omega)$ to $L^2(\Omega)$ with modulus L_f:

$$\|\nabla f(u_1) - \nabla f(u_2)\|_{L^2(\Omega)} \leq L_f \|u_1 - u_2\|_{L^2(\Omega)}$$

for all $u_1, u_2 \in L^2(\Omega)$.

Convergence analysis

Theorem

Suppose $L > L_f$.

(1) The sequence $(f(u_k) + g(u_k))$ is monotonically decreasing and converging.

(2) $\sum_{k=1}^{\infty} \|u_{k+1} - u_k\|_{L^2(\Omega)}^2 \to 0$.

(3) The sequences (u_k) and $(\nabla f(u_k))$ are bounded in $L^2(\Omega)$ if $\alpha > 0$ or $b < +\infty$.

Characteristic functions (χ_k) converge: no oscillation.

Convergence analysis

Theorem

Suppose $L > L_f$.

(1) The sequence $(f(u_k) + g(u_k))$ is monotonically decreasing and converging.

(2) $\sum_{k=1}^{\infty} \|u_{k+1} - u_k\|_{L^2(\Omega)}^2 \to 0$.

(3) The sequences (u_k) and $(\nabla f(u_k))$ are bounded in $L^2(\Omega)$ if $\alpha > 0$ or $b < +\infty$.

Define

$$\chi_k(x) := |u_k(x)|_0.$$

(4) $\sum_{k=1}^{\infty} \|\chi_k - \chi_{k+1}\|_{L^1(\Omega)} < +\infty$.

(5) $\chi_k \to \chi$ in $L^1(\Omega)$ for some characteristic function χ.

Characteristic functions (χ_k) converge: no oscillation.

Weak limit points for L^0-problems

Theorem

Let g be induced by $\|\cdot\|_0$.

Let $u^* \in U_{ad}$ be a weak sequential limit point of the iterates (u_k) in $L^2(\Omega)$. Then it holds

$$f(u^*) + g(u^*) \le \liminf_{k\to\infty}(f(u_k) + g(u_k))$$

and

$$(1-\chi)u^* = 0$$

with χ as above.

Lower semicontinuity along the iterates for L^0-problems.

[DW '19]

Strong convergence for integer-valued problems

Theorem

Suppose $g = \delta_{\mathbb{Z}}$.

Then $u_k \to u^*$ in $L^1(\Omega)$.

Proof: $\|u_{k+1} - u_k\|^2_{L^2(\Omega)} \geq \|u_{k+1} - u_k\|^1_{L^1(\Omega)}$

$$\sum_{k=0}^{\infty} \|u_{k+1} - u_k\|_{L^1(\Omega)} \leq \sum_{k=0}^{\infty} \|u_{k+1} - u_k\|^2_{L^2(\Omega)} < +\infty$$

Passing to the limit in PMP

Two problems:

(1) $\frac{L}{2}(u-u_k)^2$ is there to stay: u_{k+1} satisfies

$$\nabla f(u_k)(u_{k+1}-u_k) + \frac{L}{2}(u_{k+1}-u_k)^2 + g(u_{k+1})$$
$$\leq$$
$$\nabla f(u_k)(u-u_k) + \frac{L}{2}(u-u_k)^2 + g(u)$$

for all $u \in \mathbb{R}$ for a.a. x.

(2) weak convergence of (u_k) in $L^2(\Omega)$, no pointwise convergence.

Stationarity of limit points

Goal: Rewrite the inclusion

$$u_{k+1} \in \text{prox}_{L^{-1}g}\left(u_k - L^{-1}\nabla f(u_k)\right)$$

Stationarity of limit points

Goal: Rewrite the inclusion

$$u_{k+1} \in \text{prox}_{L^{-1}g}\left(u_k - L^{-1}\nabla f(u_k)\right)$$

Suppose $\nabla f : L^2(\Omega) \to L^2(\Omega)$ is completely continuous, e.g.,

$$v_k \rightharpoonup v \;\Rightarrow\; \nabla f(v_k) \to \nabla f(v).$$

Then $u_{k_n} \rightharpoonup u^*$ in $L^2(\Omega)$ implies

$$\nabla f(u_{k_n}) + L(u_{k_n+1} - u_{k_n}) \to \nabla f(u^*).$$

Stationarity of limit points

Goal: Rewrite the inclusion

$$u_{k+1} \in \operatorname{prox}_{L^{-1}g}\left(u_k - L^{-1}\nabla f(u_k)\right)$$

Suppose $\nabla f : L^2(\Omega) \to L^2(\Omega)$ is completely continuous, e.g.,

$$v_k \rightharpoonup v \;\Rightarrow\; \nabla f(v_k) \to \nabla f(v).$$

Then $u_{k_n} \rightharpoonup u^*$ in $L^2(\Omega)$ implies

$$\nabla f(u_{k_n}) + L(u_{k_n+1} - u_{k_n}) \to \nabla f(u^*).$$

Define G (depending on L) by

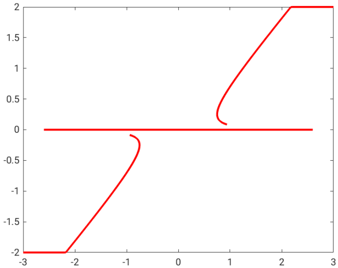
$$u_{k+1} \in G\Big(-\big(\nabla f(u_k) + L(u_{k+1} - u_k)\big)\Big).$$

(Fixed points of $G \circ (-\nabla f)$ are L-stationary [Beck, Eldar '13])

Stationarity of limit points

Given

$$u_{k+1}(x) \in G\Big(-\big(\nabla f(u_k) + L(u_{k+1} - u_k)\big)(x)\Big)$$



$G, \tilde{G} : \mathbb{R} \rightrightarrows \mathbb{R}$ have closed graphs, are not single-valued, not monotone, $G \supset \tilde{G} \supset P$.

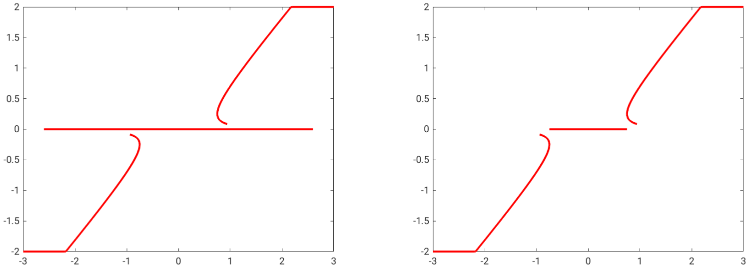
Stationarity of limit points

Given

$$u_{k+1}(x) \in G\Big(-\big(\nabla f(u_k) + L(u_{k+1} - u_k)\big)(x)\Big)$$

construct $\tilde{u}_{k+1}$ such that $\chi_{k+1}\tilde{u}_{k+1} = \chi_{k+1}u_{k+1}$ and

$$\tilde{u}_{k+1}(x) \in \chi_{k+1}(x)\tilde{G}\Big(-\big(\nabla f(u_k) + L(u_{k+1} - u_k)\big)(x)\Big).$$



$G, \tilde{G} : \mathbb{R} \rightrightarrows \mathbb{R}$ have closed graphs, are not single-valued, not monotone, $G \supset \tilde{G} \supset P$.

Stationarity of limit points

Theorem

Suppose $\nabla f : L^2(\Omega) \to L^2(\Omega)$ *is completely continuous.*
Let u^* *be a weak limit point of* (u_k)*. Then it holds*

$$u^*(x) \in \Big[G \cup \overline{\mathrm{conv}}\, \tilde{G} \Big] \Big(-\nabla f(u^*)(x) \Big)$$

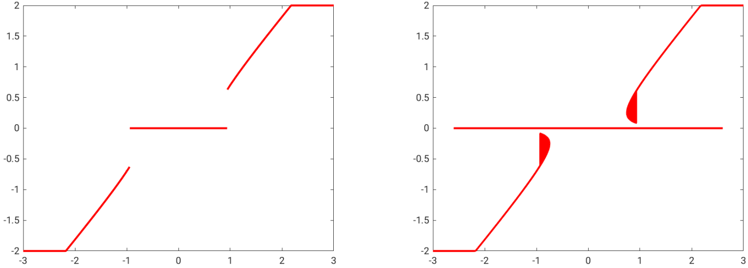
and

$$(1-\chi)u^* = 0.$$

- Weaker than PMP (even without taking $\overline{\mathrm{conv}}$). Maps $G, \tilde{G}$ depend monotonically on L. For $L = 0$, PMP is recovered.
- For convex g, weak sequential limit points would be stationary.
- If $(\nabla f(u_k))$ converges pointwise a.e., then result holds without $\overline{\mathrm{conv}}$.

PMP versus stationarity of limit points

Comparison of P and $G \cup \overline{\text{conv}}\tilde{G}$:



Strong convergence

$$\Omega_{k,\mathrm{bad}} := \Big\{x : \ u_k(x) \neq 0 \text{ and } u_k(x) \text{ is on bad branch of } G(\cdot)\Big\}$$

Theorem

Suppose $\nabla f : L^2(\Omega) \to L^2(\Omega)$ is completely continuous. Let g be induced by $\|\cdot\|_p$, $p \in (0,1)$. Let $\alpha > 0$.
Assume

$$\mathrm{meas}(\Omega_{k,\mathrm{bad}}) \to 0.$$

Suppose $u_{k_n} \rightharpoonup u^*$. Then $u_{k_n} \to u^*$ and

$$\chi u^*(x) \in \chi G\Big(-\nabla f(u^*)(x)\Big).$$

Outline

Example

Define $f(u)$ by

$$f(u) := \frac{1}{2}\|y_u - y_d\|^2_{L^2(\Omega)},$$

where y_u denotes the weak solution of the elliptic partial differential equation

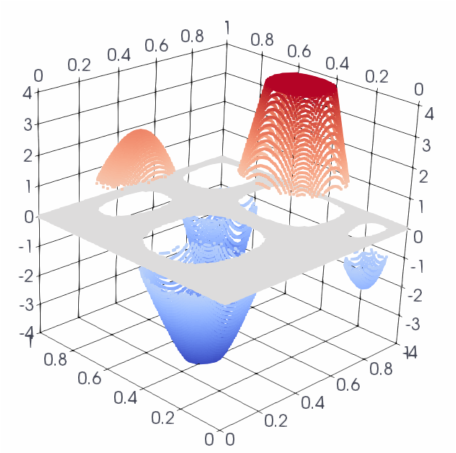
$$-\Delta y = u \quad \text{in } \Omega, \quad y = 0 \quad \text{on } \partial\Omega.$$

Here, we chose $\Omega = (0,1)^2$, [Ito, Kunisch '14]

$$y_d(x_1, x_2) = 10x_1 \sin(5x_1)\cos(7x_2), \quad \alpha = 0.01, \quad \beta = 0.01, \quad b = 4$$

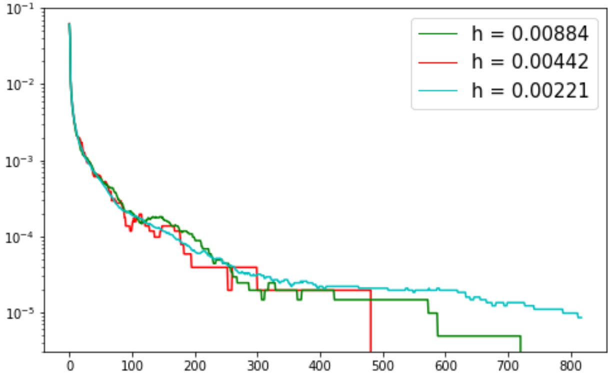
Solution

Optimal control u for $p = 0.8$



The bad set

Measure of $\Omega_{k,\mathrm{bad}}$ for three different discretizations



Outlook

- results extend to sparsity-promoting g
- L^0-constraints $\|u\|_0 \leq \tau$,
- group sparsity: L^0 in space, L^2 in time,
- second-order optimality conditions in cooperation with Eduardo Casas [SICON'20] and with Gerd Wachsmuth,
- H^1-regularization (existence of solutions, but no maximum principle),
- H^s-regularization.