Variational Inequalities and Convergence Rates for Non-convex Regularization

Markus Grasmair

Computational Science Center,
University of Vienna,
Austria

June 23, 2010

Outline

Outline

Inverse Problems

Let X, Y be topological spaces and $F\colon X \to Y$ and solve, for given data $y \in Y$, the equation

$$F(x) = y\,. \tag{1}$$

If (1) is ill-posed, regularization is necessary:
Search for $x_\alpha \in X$ minimizing

$$\mathcal{T}(x;\alpha,y) := \mathcal{S}\big(F(x),y\big) + \alpha\mathcal{R}(x)\,.$$

Here,

$\mathcal{S}\colon Y \times Y \to \mathbb{R}_{\geq 0}$	...	non-negative *distance measure*,
$\mathcal{R}\colon X \to \mathbb{R}_{\geq 0}$	...	non-negative *regularization functional*,
$\alpha > 0$	...	*regularization parameter*.

Well-Posedness

Existence:
$\mathcal{T}(\cdot;\alpha,y)$ attains a minimizer for every $\alpha > 0$ and $y \in Y$.

Stability:
If $\mathcal{S}(y^{(k)},y) \to 0$ and $x_\alpha^{(k)} \in \arg\min_x \mathcal{T}(x;\alpha,y^{(k)})$, then

$$x_\alpha^{(k)} \to x_\alpha \in \arg\min_x \mathcal{T}(x;\alpha,y) .$$

Convergence:
If $\mathcal{S}(y^\delta,y^\dagger) \le \delta \to 0$ and $\alpha \to 0$ sufficiently slowly ($\delta/\alpha \to 0$), then

$$\arg\min_x \mathcal{T}(x;\alpha,y^\delta) \ni x_\alpha^\delta \;\to\; x^\dagger \in \arg\min\bigl\{\mathcal{R}(x) : F(x) = y^\dagger\bigr\} .$$

Conditions for well-posedness in: Hofmann et al. 2007, Pöschl 2008, Scherzer et al. 2009.

Convergence Rates

Measure speed of convergence: Let

$$\Sigma(x^\dagger;\alpha,\delta) := \Big\{x_\alpha^\delta \in \arg\min_x \mathcal{T}(x;\alpha,y^\delta) : \mathcal{S}(y^\delta,y^\dagger) \le \delta\Big\}$$

and define for some *distance measure*

$$D \colon X \times X \to [0,+\infty]$$

the function

$$H(x^\dagger;\alpha,\delta) := \sup\Big\{D(x^\dagger,x_\alpha^\delta) : x_\alpha^\delta \in \Sigma(x^\dagger;\alpha,\delta)\Big\}\,.$$

Convergence rate: behaviour of H as α and δ tend to zero $\sim$ accuracy of the regularization method (for small noise level).

Classical Convergence Rates in Hilbert Spaces

Setting: X, Y Hilbert spaces, $F\colon X \to Y$ bounded linear.
Let

$$\mathcal{S}(y_1, y_2) = \|y_1 - y_2\|_Y^2 , \qquad \mathcal{R}(x) = \|x\|_X^2 ,$$
$$D(x_1, x_2) := \|x_1 - x_2\|_X^2 .$$

If $x^\dagger$ satisfies the *range condition*

$$x^\dagger \in \operatorname{Ran} F^*$$

then there exists a constant $\gamma > 0$ such that

$$H(x^\dagger; \alpha, \delta) \le \frac{\delta}{\alpha} + \gamma\sqrt{\delta} + \frac{\gamma^2}{2}\alpha .$$

Note that $\delta \simeq \mathcal{S}(y^\dagger, y^\delta) = \|y^\dagger - y^\delta\|^2$.

Banach Spaces: Bregman Distances

Let X, Y be Banach spaces, $F\colon X \to Y$ bounded linear, $\mathcal{S}(y_1, y_2) = \|y_1 - y_2\|_Y^2$. Let

$$\mathcal{R}\colon X \to [0, +\infty] \text{ convex and lower semi-continuous,}$$
$$D(x_1, x_2) := \mathcal{R}(x_1) - \mathcal{R}(x_2) - \langle \partial\mathcal{R}(x_2), x_1 - x_2 \rangle .$$

$D \ldots$ *Bregman distance.*
If $x^\dagger$ satisfies the range condition

$$\operatorname{Ran} F^* \cap \partial\mathcal{R}(x^\dagger) \neq \emptyset$$

then there exists a constant $\gamma > 0$ such that

$$H(x^\dagger; \alpha, \delta) \leq \frac{\delta}{\alpha} + \gamma\sqrt{\delta} + \frac{\gamma^2}{2}\alpha .$$

Range Condition and Variational Inequalities

The range condition

$$\operatorname{Ran} F^* \cap \partial\mathcal{R}(x^\dagger) \neq \emptyset$$

is equivalent to the inequality

$$\langle \partial\mathcal{R}(x^\dagger), x^\dagger - x \rangle \leq \gamma \|F(x^\dagger - x)\| . \tag{2}$$

Proofs of rates rely on (2) rather than on the range condition. Slight modification of proofs yields similar rates under the weaker condition

$$\langle \partial\mathcal{R}(x^\dagger), x^\dagger - x \rangle \leq \eta D(x^\dagger, x) + \gamma \|F(x^\dagger) - F(x)\| .$$

for some $0 < \eta < 1$. No linearity of F is required.

Variational Inequalities

Let X, Y be Banach spaces and $F\colon X \to Y$ sufficiently regular. Assume that, for some $\beta > 0$, $\gamma > 0$,

$$\beta D(x, x^\dagger) \le \mathcal{R}(x) - \mathcal{R}(x^\dagger) + \gamma \|F(x) - F(x^\dagger)\|$$

whenever x sufficiently close to $x^\dagger$ and $|\mathcal{R}(x) - \mathcal{R}(x^\dagger)|$ small enough. Then

$$\beta H(x^\dagger; \alpha, \delta) \le \frac{\delta}{\alpha} + \gamma\sqrt{\delta} + \frac{\gamma^2}{2}\alpha$$

whenever δ, α, and δ/α are small enough.

Outline

Abstract Convexity

Definition

Let X be a set and let W be a family of functions

$$w\colon X \to \bar{\mathbb{R}} = \mathbb{R} \cup \{+\infty\} .$$

A function $\mathcal{R}\colon X \to \bar{\mathbb{R}}$ is

W-convex at $x \in X$,

if for every $\varepsilon > 0$ there exists $w \in W$ such that

$$\mathcal{R}(\tilde{x}) \geq \mathcal{R}(x) + \big(w(\tilde{x}) - w(x)\big) - \varepsilon$$

for all $\tilde{x} \in X$.

Abstract Bregman Distance

Definition

Let $\mathcal{R}$ be a W-convex function.
The W-sub-differential of $\mathcal{R}$ at $x \in X$ is defined as

$$\partial_W \mathcal{R}(X) := \left\{ w \in W : \mathcal{R}(\tilde{x}) \geq \mathcal{R}(x) + w(\tilde{x}) - w(x) \right\} .$$

We define, for $w \in \partial_W \mathcal{R}(x)$, the *$W$-Bregman distance* with respect to w as

$$D^w(x, \tilde{x}) = \mathcal{R}(\tilde{x}) - \mathcal{R}(x) - \big(w(\tilde{x}) - w(x) \big) \geq 0 .$$

Example — Classical Convexity

Let X be a Banach space with dual X^*.

A function $\mathcal{R}\colon X \to \bar{\mathbb{R}}$ is X^*-convex, if and only it is lower semi-continuous and convex in the classical sense.

We have

$$\partial_{X^*}\mathcal{R}(x) = \bigl\{\xi \in X^* : \mathcal{R}(\tilde{x}) \geq \mathcal{R}(x) + \langle \xi, \tilde{x}\rangle - \langle \xi, x\rangle\bigr\} = \partial\mathcal{R}(x)\,.$$

Moreover,

$$D^\xi(x,\tilde{x}) = \mathcal{R}(\tilde{x}) - \mathcal{R}(x) - \langle \xi, \tilde{x} - x\rangle$$

is the usual Bregman distance.

Example — Clarke Sub-differential

Let X be a Hilbert space. The

$$\textit{proximal sub-differential } \partial_P \mathcal{R}(x)$$

is defined as the set of all $\xi \in X$ such that

$$\mathcal{R}(\tilde{x}) \geq \mathcal{R}(x) + \langle \xi, \tilde{x} - x \rangle - \sigma \|\tilde{x} - x\|^2$$

for some $\sigma \geq 0$ and all $\tilde{x}$ near x.

Define W by

$$w \in W \iff w(\tilde{x}) = \langle \xi, \tilde{x} - x \rangle - \sigma \|\tilde{x} - x\|^2$$

for some $\xi \in X$, $\sigma \geq 0$, and $\tilde{x}$ close to x. Then

$$\partial_P \mathcal{R}(x) \neq \emptyset \iff \partial_W \mathcal{R}(x) \neq \emptyset\ .$$

Example — Generalized Sub-differential

Define W as the set of all functions of the form

$$w(\tilde{x}) = \langle \xi, \tilde{x} - x \rangle - A(\tilde{x} - x, \tilde{x} - x)$$

for $\tilde{x}$ close to x, with $\xi \in X$ and A a positive semi-definite, symmetric, bounded quadratic form.

Define the *generalized sub-differential* of $\mathcal{R}$ at x as $\partial_W \mathcal{R}(x)$. Again,

$$\partial_P \mathcal{R}(x) \neq \emptyset \iff \partial_W \mathcal{R}(x) \neq \emptyset\,.$$

We have the Bregman distance

$$D^w(x, \tilde{x}) = \mathcal{R}(\tilde{x}) - \mathcal{R}(x) - \langle \xi, \tilde{x} - x \rangle + A(\tilde{x} - x, \tilde{x} - x)$$

for $\tilde{x}$ close to x.

Generalized Variational Inequalities

Let

$$x^\dagger \in \arg\min\{\mathcal{R}(x) : Ax = y^\dagger\} .$$

and let $\Phi\colon \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ concave and strictly increasing with $\Phi(0) = 0$.

Definition

We say that a *variational inequality* at $x^\dagger$ holds with $\beta > 0$ and Φ, if

$$\beta D^w(x^\dagger, x) \leq \mathcal{R}(x) - \mathcal{R}(x^\dagger) + \Phi\big(\mathcal{S}(F(x), F(x^\dagger))\big)$$

for all x in a neighbourhood of $x^\dagger$ with $\mathcal{R}(x)$ close to $\mathcal{R}(x^\dagger)$.

Convergence Rates

Theorem

Assume that a variational inequality at $x^\dagger$ holds with $\beta > 0$ and Φ. Then for α and δ small enough we have the following estimates:

- *If* $\lim_{t\to 0^+} \Phi(t)/t < +\infty$, *then*

$$\beta H(x^\dagger; \alpha, \delta) \le \frac{\delta}{\alpha} + \gamma \Phi(\delta) \, .$$

- *If* $\lim_{t\to 0^+} \Phi(t)/t = +\infty$, *then*

$$\beta H(x^\dagger; \alpha, \delta) \le \frac{\delta}{\alpha} + \gamma_1 \Phi(\delta) + \gamma_2 \frac{\Psi(\alpha)}{\alpha}$$

with Ψ *denoting the convex conjugate of* Φ^{-1}.

Convergence Rates — Asymptotics

Let now

$$x_\alpha^\delta \in \arg\min_x \mathcal{T}(x;\alpha,y^\delta) \qquad \text{with} \qquad \mathcal{S}(y^\delta,y^\dagger) \le \delta\,.$$

Corollary

Assume that a variational inequality at $x^\dagger$ holds with $\beta > 0$ and Φ.

- *If $\lim_{t\to 0^+} \Phi(t)/t < +\infty$, then we have for $\alpha = \text{const}$ small enough*
$$D^w(x^\dagger, x_\alpha^\delta) = O(\delta)\,.$$
- *If $\lim_{t\to 0^+} \Phi(t)/t = +\infty$ and $\alpha \sim \delta/\Phi(\delta)$ then*
$$D^w(x^\dagger, x_\alpha^\delta) = O\big(\Phi(\delta)\big)\,.$$

Outline

Metric Regularization

Let Y be a metric space and

$$\mathcal{S}(y_1, y_2) = d(y_1, y_2)^p \qquad \text{with} \qquad p > 1\,.$$

If the variational inequality

$$\beta D^w(x^\dagger, x) \le \mathcal{R}(x) - \mathcal{R}(x^\dagger) + \gamma d\big(F(x), F(x^\dagger)\big)$$

holds, then we have for a parameter choice

$$\alpha \sim d(y^\dagger, y^\delta)^{p-1}$$

the rate

$$D^w(x^\dagger, x_\alpha^\delta) \le O\big(d(y^\dagger, y^\delta)\big)\,.$$

Setting

Let X and Y be Hilbert spaces and $F\colon X \to Y$ bounded linear. Let moreover

$$\mathcal{S}(y_1, y_2) = \|y_1 - y_2\|^p \qquad \text{with} \qquad p > 1\,.$$

Assume that $\mathcal{R}$ has a proximal sub-differential w at $x^\dagger$, that is,

$$\mathcal{R}(x) \geq \mathcal{R}(x^\dagger) + \langle \xi, x - x^\dagger \rangle - A(x - x^\dagger, x - x^\dagger)$$

with $\xi \in X$ and $A\colon X \to X$ positive semi-definite, symmetric, bounded, bilinear.
Then there exists $L\colon X \to X$ bounded linear and self-adjoint such that

$$A(x_1, x_2) = \big\langle L x_1, x_2 \big\rangle\,.$$

Convergence Rates

Lemma

Assume that for some $\mu > 0$ *the mapping* $\mu^2 F^* F - L$ *is positive semi-definite and that*

$$\xi \in \operatorname{Ran}\left(\sqrt{\mu^2 F^* F - L}\right) .$$

Then the variational inequality

$$D^w(x^\dagger, x) \le \mathcal{R}(x) - \mathcal{R}(x^\dagger) + \gamma \|F(x - x^\dagger)\|$$

holds for some $\gamma > 0$*. In particular, with a parameter choice* $\alpha \sim \|y^\dagger - y^\delta\|^{p-1}$,

$$D^w(x^\dagger, x_\alpha^\delta) = O(\|y^\dagger - y^\delta\|) .$$

Setting

Let Y be a Hilbert space, $X = \ell^2$, and $F\colon \ell^2 \to Y$ bounded linear. Let $\mathcal{S}(y_1, y_2) = \|y_1 - y_2\|^2$ and define

$$\mathcal{R}(x) = \sum_\lambda \phi(x_\lambda) \qquad \text{for some} \qquad \phi\colon \mathbb{R} \to [0, +\infty] \,.$$

Let $1 < p < 2$ and consider the set W of functions of the form

$$w(x) = \langle \xi, x - x^\dagger \rangle - \sum_\lambda c_\lambda |x_\lambda - x^\dagger_\lambda|^p$$

with $\xi \in \ell^2$ and $c_\lambda > 0$. Assume that, for some $p > q > 0$ and $C > 0$,

$$\phi(t) \geq \frac{C|t|^q}{1 + |t|^q}$$

Convergence Rates

Lemma

Assume that the following hold:

- $x^\dagger$ *is the unique* $\mathcal{R}$*-minimizing solution of* $Fx^\dagger = y^\dagger$.
- $\operatorname{supp}(x^\dagger)$ *is finite (*$x^\dagger$ *is sparse).*
- $F|_{\ell^2(\operatorname{supp}(x^\dagger))}$ *is injective.*

Assume that

$$\tilde{w} = x \mapsto \langle \xi, x - x^\dagger \rangle - \sum_\lambda c_\lambda |x_\lambda - x_\lambda^\dagger|^p \in \partial_W \mathcal{R}(x^\dagger) .$$

If $\xi \in \operatorname{Ran}(F^*)$ *and* $\operatorname{supp}(\xi) = \operatorname{supp}(x^\dagger)$, *then, for some* $w \in \partial_W \mathcal{R}(x^\dagger)$ *and* $\gamma > 0$,

$$\gamma \|x_\alpha^\delta - x^\dagger\|^p \le D^w(x^\dagger, x_\alpha^\delta) = O(\|y^\delta - y^\dagger\|) .$$

Summary

- Derivation of convergence rates for non-convex Tikhonov regularization.
- Variational inequalities allow generalization by means of abstract concepts of convexity.
- Connection to standard range condition for linear operators on Hilbert spaces.
- Convergence rates for sparse regularization with non-convex regularization term.