Variational Inequalities and Convergence Rates for Non-convex Regularization

Markus Grasmair

Computational Science Center,
University of Vienna,
Austria

June 23, 2010
Outline

1 Variational Inequalities and Convergence Rates
   - Convergence Rates
   - Variational Methods

2 Abstraction
   - Abstract Convexity
   - Variational Inequalities

3 Examples
   - Metric Regularization
   - Non-convex Regularization on Hilbert Spaces
   - Sparse Regularization
Outline

1. Variational Inequalities and Convergence Rates
   - Convergence Rates
   - Variational Methods

2. Abstraction
   - Abstract Convexity
   - Variational Inequalities

3. Examples
   - Metric Regularization
   - Non-convex Regularization on Hilbert Spaces
   - Sparse Regularization
Let $X$, $Y$ be topological spaces and $F: X \to Y$ and solve, for given data $y \in Y$, the equation

$$F(x) = y.$$ \quad (1)

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

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

Here,

- $S: Y \times Y \to \mathbb{R}_{\geq 0}$ \ldots non-negative distance measure,
- $R: X \to \mathbb{R}_{\geq 0}$ \ldots non-negative regularization functional,
- $\alpha > 0$ \ldots regularization parameter.
Well-Posedness

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

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

Convergence:
If $S(y^\delta, y^\dagger) \leq \delta \to 0$ and $\alpha \to 0$ sufficiently slowly ($\delta/\alpha \to 0$), then
$$\text{arg min}_x \mathcal{T}(x; \alpha, y^\delta) \ni x^\delta_{\alpha} \to x^\dagger \in \text{arg min}\{\mathcal{R}(x) : F(x) = y^\dagger\}.$$  

Measure speed of convergence: Let

\[ \Sigma(x^†; \alpha, \delta) := \left\{ x^{\delta}_{\alpha} \in \arg \min_{x} T(x; \alpha, y^{\delta}) : S(y^{\delta}, y^†) \leq \delta \right\} \]

and define for some distance measure

\[ D : X \times X \to [0, +\infty] \]

the function

\[ H(x^†; \alpha, \delta) := \sup \left\{ D(x^†, x^{\delta}_{\alpha}) : x^{\delta}_{\alpha} \in \Sigma(x^†; \alpha, \delta) \right\} . \]

Convergence rate: behaviour of \( H \) as \( \alpha \) and \( \delta \) tend to zero \( \sim \) accuracy of the regularization method (for small noise level).
Classical Convergence Rates in Hilbert Spaces

Setting: $X$, $Y$ Hilbert spaces, $F: X \rightarrow Y$ bounded linear.

Let

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

If $x^\dagger$ satisfies the range condition

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

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

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

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

Markus Grasmair
Let $X$, $Y$ be Banach spaces, $F : X \to Y$ bounded linear, $S(y_1, y_2) = \|y_1 - y_2\|_Y^2$. Let

$$R : X \to [0, +\infty] \text{ convex and lower semi-continuous,}$$

$$D(x_1, x_2) := R(x_1) - R(x_2) - \langle \partial R(x_2), x_1 - x_2 \rangle.$$

$D \ldots$ Bregman distance.

If $x^\dagger$ satisfies the range condition

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

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

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

is equivalent to the inequality

$$\langle \partial \mathcal{R}(x^\dagger), x^\dagger - x \rangle \leq \gamma \| F(x^\dagger - x) \| .$$  \hspace{1cm} (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.
Let $X, Y$ be Banach spaces and $F : X \to Y$ sufficiently regular.
Assume that, for some $\beta > 0$, $\gamma > 0$,

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

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

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

whenever $\delta$, $\alpha$, and $\delta/\alpha$ are small enough.
Outline

1 Variational Inequalities and Convergence Rates
   - Convergence Rates
   - Variational Methods

2 Abstraction
   - Abstract Convexity
   - Variational Inequalities

3 Examples
   - Metric Regularization
   - Non-convex Regularization on Hilbert Spaces
   - Sparse Regularization
Abstract Convexity

Definition

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

$$w : X \rightarrow \bar{\mathbb{R}} = \mathbb{R} \cup \{+\infty\}.$$ 

A function $\mathcal{R} : X \rightarrow \bar{\mathbb{R}}$ is

$W$-convex at $x \in X,$

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

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

for all $\tilde{x} \in X.$
Abstract Bregman Distance

**Definition**

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

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

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

$$D^w(x, \tilde{x}) = \mathcal{R}(\tilde{x}) - \mathcal{R}(x) - (w(\tilde{x}) - w(x)) \geq 0 .$$
Let $X$ be a Banach space with dual $X^*$.

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

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

Moreover,

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

is the usual Bregman distance.
Example — Clarke Sub-differential

Let $X$ be a Hilbert space. The

\[ \text{proximal sub-differential } \partial_P \mathcal{R}(x) \]

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

\[ \partial_P \mathcal{R}(x) \neq \emptyset \iff \partial_W \mathcal{R}(x) \neq \emptyset . \]
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 $R$ at $x$ as $\partial_W R(x)$. Again,

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

We have the Bregman distance

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

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

Markus Grasmair
Let
\[ x^\dagger \in \arg \min \{ \mathcal{R}(x) : Ax = y^\dagger \} . \]
and let \( \Phi : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0} \) concave and strictly increasing with \( \Phi(0) = 0 \).

**Definition**

We say that a variational inequality at \( x^\dagger \) holds with \( \beta > 0 \) and \( \Phi \), if

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

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

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

- If $\lim_{t \to 0^+} \Phi(t)/t < +\infty$, then
  \[
  \beta H(x^\dagger; \alpha, \delta) \leq \frac{\delta}{\alpha} + \gamma \Phi(\delta).
  \]

- If $\lim_{t \to 0^+} \Phi(t)/t = +\infty$, then
  \[
  \beta H(x^\dagger; \alpha, \delta) \leq \frac{\delta}{\alpha} + \gamma_1 \Phi(\delta) + \gamma_2 \frac{\psi(\alpha)}{\alpha}
  \]

with $\Psi$ denoting the convex conjugate of $\Phi^{-1}$. 
Let now
\[ x^\delta_\alpha \in \arg \min_x \mathcal{T}(x; \alpha, y^\delta) \quad \text{with} \quad S(y^\delta, y^\dagger) \leq \delta . \]

**Corollary**

Assume that a variational inequality at \( x^\dagger \) holds with \( \beta > 0 \) and \( \Phi \).

- If \( \lim_{t \to 0^+} \Phi(t)/t < +\infty \), then we have for \( \alpha = \text{const small enough} \)
  \[ D^w(x^\dagger, x^\delta_\alpha) = O(\delta) . \]

- If \( \lim_{t \to 0^+} \Phi(t)/t = +\infty \) and \( \alpha \sim \delta/\Phi(\delta) \) then
  \[ D^w(x^\dagger, x^\delta_\alpha) = O(\Phi(\delta)) . \]
Outline

1. Variational Inequalities and Convergence Rates
   - Convergence Rates
   - Variational Methods

2. Abstraction
   - Abstract Convexity
   - Variational Inequalities

3. Examples
   - Metric Regularization
   - Non-convex Regularization on Hilbert Spaces
   - Sparse Regularization
Let $Y$ be a metric space and
\[ S(y_1, y_2) = d(y_1, y_2)^p \quad \text{with} \quad p > 1. \]

If the variational inequality
\[ \beta D^w(x^\dagger, x) \leq \mathcal{R}(x) - \mathcal{R}(x^\dagger) + \gamma d(F(x), F(x^\dagger)) \]
holds, then we have for a parameter choice
\[ \alpha \sim d(y^\dagger, y^\delta)^p - 1 \]
the rate
\[ D^w(x^\dagger, x_\alpha^\delta) \leq O(d(y^\dagger, y^\delta)). \]
Setting

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

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

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 $A: X \to X$ positive semi-definite, symmetric, bounded, bilinear.

Then there exists $L: X \to X$ bounded linear and self-adjoint such that

$$A(x_1, x_2) = \langle Lx_1, x_2 \rangle.$$
Lemma

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

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

Then the variational inequality

$$D^w(x^\dagger, x) \leq R(x) - R(x^\dagger) + \gamma \|F(x - x^\dagger)\|$$

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

$$D^w(x^\dagger, x_\alpha^\delta) = O(\|y^\dagger - y^{\delta}\|) .$$
Let $Y$ be a Hilbert space, $X = \ell^2$, and $F : \ell^2 \to Y$ bounded linear. Let $S(y_1, y_2) = \|y_1 - y_2\|^2$ and define

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

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 $C > 0$,

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

Assume that the following hold:

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

Assume that

\[
\tilde{w} = x \mapsto \langle \xi, x - x^\dagger \rangle - \sum_{\lambda} c_\lambda |x_\lambda - x^\dagger_\lambda|^p \in \partial W R(x^\dagger).
\]

If \( \xi \in \text{Ran}(F^*) \) and \( \text{supp}(\xi) = \text{supp}(x^\dagger) \), then, for some \( w \in \partial W R(x^\dagger) \) and \( \gamma > 0 \),

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