Sparse Regularization with Non-convex Regularization Terms

Markus Grasmair

Computational Science Center,
University of Vienna,
Austria

April 13, 2010

Outline

Setting

Let U, V be Hilbert spaces and $F\colon U \to V$ bounded linear. Consider the approximative solution of the operator equation

$$F(u) = v$$

for given $v \in V$.

If the operator equation is ill-posed (e.g. if $F\colon U \to V$ is compact), regularization is necessary: search for $u_\alpha \in U$ minimizing

$$\mathcal{T}(u; \alpha, v) := \|Fu - v\|^2 + \alpha\mathcal{R}(u)$$

with *regularization functional* $\mathcal{R}\colon U \to \mathbb{R}_{\geq 0}$ and *regularization term* $\alpha > 0$.

Quadratic Regularization

Consider

$$\mathcal{T}(u;\alpha,v) := \|Fu - v\|_V^2 + \alpha\|u\|_U^2 .$$

Original setting by Tikhonov: $U = W^{k,2}(\Omega)$.
Regularization term enforces *smoothness* (differentiability) of the approximate solutions.

In some applications (signal / image processing), smoothness no realistic assumption / no important property.

Recent idea: instead of smoothness, try to enforce sparsity with respect to a given basis / frame / subset.

Sparsity

Let $(\phi_i)_{i\in I} \subset U$ be some bounded family. Find

$$u = \sum_{i\in I} u_i \phi_i \in U$$

with

$$Fu \approx v$$

such that (weak sense)

$$\#\{i \in I : |u_i| \gg 0\} \text{ is small}$$

or (strong sense)

$$\#\{i \in I : u_i \neq 0\} \text{ is finite .}$$

Sub-quadratic Regularization

Let U be a Hilbert space and $(\phi_i)_{i\in I}$ an orthonormal basis. Then

$$\|u\|_U^2 = \sum_i |u_i|^2 \,.$$

Idea: Obtain (weak) sparsity by replacing quadratic regularization with sub-quadratic regularization:
Consider

$$\mathcal{R}(u) := \sum_i |u_i|^p \qquad \text{with } 1 \le p < 2 \,.$$

Stronger penalization of small coefficients, weaker penalization of large coefficients.
Minimizer of Tikhonov functional will be in ℓ^p.

Sparsity by Differentiability

Consider $(\phi_k)_{k\in\mathbb{N}}$ Fourier-basis of $L^2(0,2\pi)$. Then

$$\|u'\|^2 = \sum_k k^2 |u_k|^2 \,.$$

Reverse Hölder inequality:

$$\sum_k k^2 |u_k|^2 \geq \Bigl(\sum_k k^{-2q}\Bigr)^{-1/q} \Bigl(\sum_k |u_k|^{2p}\Bigr)^{1/p}$$

with $0 < p < 1$, $q > 0$, and $1/p - 1/q = 1$. Therefore:

$$\|u'\|^2 < +\infty \implies (u_k)_{k\in\mathbb{N}} \in \ell^{2/3+\varepsilon} \,.$$

But: “Scale dependent sparsity pattern.”

Non-convex Regularization

Achieve sparsity in the strong sense by choosing

$$\mathcal{R}(u) = \sum_i |u_i|^p \qquad \text{with } 0 < p < 1\,.$$

More general setting:

$$\mathcal{R}(u) := \sum_i w_i \phi(u_i)$$

with

$$w_i > w_{\min} \geq 0 \qquad \text{and} \qquad \phi\colon \mathbb{R} \to [0, +\infty]\,.$$

Conditions for this to make sense?

Outline

Important Properties

Existence:

$\mathcal{T}(\cdot;\alpha,v)$ attains a minimizer for every $\alpha > 0$ and $v \in V$.

Stability:

If $y^{(k)} \to y$ and $u_\alpha^{(k)} \in \arg\min \mathcal{T}(\cdot;\alpha,v^{(k)})$ then

$$u_\alpha^{(k)} \to u_\alpha \in \arg\min \mathcal{T}(\cdot;\alpha,v)\,.$$

Convergence:

If $\|v^\delta - v^\dagger\| \le \delta \to 0$, $\alpha \to 0$ and $\delta^2/\alpha \to 0$, then

$$\arg\min \mathcal{T}(\cdot;\alpha,v^\delta) = u_\alpha^\delta \;\to\; u^\dagger = \arg\min\bigl\{\mathcal{R}(u) : Fu = v^\dagger\bigr\}\,.$$

Lower Semi-continuity

Direct method of the calculus of variations:

If $\mathcal{T}$ is weakly lower semi-continuous and coercive, then a minimizer exists.

Recall:

$$\mathcal{R}(u) = \sum_i w_i \phi(u_i) \ .$$

Lemma

The following are equivalent:

- *The functional $\mathcal{R}$ is lower semi-continuous.*
- *The functional $\mathcal{R}$ is weakly lower semi-continuous.*
- *The function ϕ is lower semi-continuous.*

Coercivity

Recall: $\mathcal{R}$ is coercive, if $\|u\| \to \infty$ implies $\mathcal{R}(u) \to \infty$.

Lemma

Assume that there exists $C > 0$ with

$$\phi(t) \geq \frac{Ct^2}{1+t^2}$$

and

$$\lim_{t \to \pm\infty} \phi(t) = +\infty\ .$$

Then $\mathcal{R}$ is coercive.

If $\sup_i w_i < +\infty$, then also the converse holds.

If $\sup_i w_i = +\infty$, weaker growth of ϕ is possible.

Radon–Riesz Property

Under the natural assumptions introduced above, the functional $\mathcal{R}$ satisfies the *Radon–Riesz* property:

Lemma

Let the assumptions for lower semi-continuity and coercivity be satisfied.
If $(u^{(k)})_{k\in\mathbb{N}} \subset \ell^2$ *converges weakly to* $u \in \ell^2$ *and* $\mathcal{R}(u^{(k)}) \to \mathcal{R}(u) < +\infty$, *then*

$$\|u^{(k)} - u\| \to 0 \,.$$

Important property for deducing stability and convergence in norm topology.

Well-posedness

Theorem

Assume that the following hold:

- *The function ϕ is lower semi-continuous and $\phi(0) = 0$.*
- $\lim_{t\to\pm\infty} \phi(t) = +\infty$.
- *There exists $C > 0$ with*

$$\phi(t) \geq \frac{Ct^2}{1+t^2} .$$

Then Tikhonov regularization with $\mathcal{R}(u) = \sum_i w_i \phi(u_i)$ is well-posed, stable, and convergent.

Here: stability and convergence in sub-sequential sense.

Necessary Conditions for Sparsity

Lemma

Assume that there exists $C > 0$ such that

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

Then every local minimizer of $\mathcal{T}(\cdot;\alpha,v)$ is sparse in the strong sense.

In a sense:

$$\text{Sparsity of solution} \iff \text{Non-differentiability of } \phi \text{ at } 0.$$

Outline

Convergence Rates for Quadratic Regularization

Quantitative estimate of the quality of the regularized solutions:

Let

$$\mathcal{R}(u) = \|u\|^2 \qquad \text{and} \qquad u^\dagger \in \operatorname{Ran} F^* \,.$$

If

$$\delta \sim \alpha\,, \qquad \|v^\delta - v^\dagger\| \le \delta \to 0\,,$$

and

$$u_\alpha^\delta \in \arg\min_u \mathcal{T}(u; \alpha, v^\delta)\,,$$

then

$$\|u_\alpha^\delta - u^\dagger\| = O(\sqrt{\delta}) \qquad \text{as } \delta \to 0\,.$$

Infinite Growth at Zero

Recall: lower Dini derivatives of $\phi\colon \mathbb{R} \to [0, +\infty]$ defined as

$$D_+\phi(t) = \liminf_{\varepsilon \to 0^+} \frac{\phi(t+\varepsilon) - \phi(t)}{\varepsilon} ,$$
$$D_-\phi(t) = \liminf_{\varepsilon \to 0^-} \frac{\phi(t+\varepsilon) - \phi(t)}{\varepsilon} .$$

In particular:

$$D_\pm\phi(0) = \pm\infty \qquad \Longleftrightarrow \qquad \phi \text{ has infinite growth near zero} .$$

Convergence Rates for Sparse Regularization

Define
$$\operatorname{supp}(u^\dagger) := \left\{ i \in I : u_i^\dagger \neq 0 \right\} .$$

Theorem

Assume that the following hold:

- $D_\pm \phi(0) = \pm\infty$.
- $u^\dagger$ *is the unique* $\mathcal{R}$*-minimizing solution of* $Fu = y^\dagger$.
- $u^\dagger$ *is sparse and* F *is injective on* $\ell^2(\operatorname{supp}(u^\dagger))$.

If $\alpha \sim \delta$ *we have*
$$\| u_\alpha^\delta - u^\dagger \| = O(\delta) .$$

Restricted Injectivity

Uniqueness of $u^\dagger$ necessary for obtaining any convergence rates (else: convergence to different minimizers).

In non-convex case, uniqueness of $u^\dagger$ stronger condition than restricted injectivity:

If $\phi \in C^2(\mathbb{R} \setminus \{0\})$ and $\phi''(t) \leq 0$ for $t \neq 0$, then:

$$\text{Uniqueness of } u^\dagger \implies \text{Injectivity of } F \text{ on } \operatorname{supp}(u^\dagger).$$

More general: same implication holds, if ϕ is locally concave on $\mathbb{R} \setminus \{0\}$.

Summary

- General setting for (componentwise) Tikhonov regularization on ℓ^2:
$$\mathcal{R}(u) = \sum_i w_i \phi(u_i) \,.$$
- Necessary and sufficient conditions for well-posedness (existence, stability, convergence) only in terms of ϕ.
- Sparsity for sufficiently fast growth at zero.
- Linear convergence rates for infinite growth at zero — if minimizer is unique.