Forschungsschwerpunkt S92

Industrial Geometry

http://www.industrial-geometry.at

FSP Report No. 74

Well-posedness and Convergence Rates for Sparse Regularization with Sublinear l^q Penalty Term

M. Grasmair

August 2008

Der Wissenschaftsfonds.

Well-posedness and Convergence Rates for Sparse Regularization with Sublinear l^q Penalty Term

Markus Grasmair

Department of Mathematics
University of Innsbruck
Technikerstr. 21a
6020 Innsbruck, Austria

August 11, 2008

Abstract

This paper deals with the application of non-convex, sublinear penalty terms to the regularization of possibly non-linear inverse problems the solutions of which are assumed to have a sparse expansion with respect to some given basis or frame. It is shown that this type of regularization is well-posed and yields sparse results. Moreover, linear convergence rates are derived under the additional assumption of a certain range condition.

MSC: 65J20; 65J22, 49N45.

1 Introduction and Statement of the Results

In the recent years, the notion of sparsity has received much attention both for compression and denoising, and the solution of inverse and ill-posed problems (see for instance [1, 2, 3, 4]). In the inverse problems case, one is given an equation $F(u) = v$ to solve, where $F\colon U \to V$ is a possibly non-linear operator between the Banach spaces U and V. The goal is to find an approximate solution u that is sparse with respect to a sequence $(\phi_i)_{i\in\mathbb{N}} \subset U$ in the sense that it can be written as a finite linear combination of some ϕ_i. This sparsity can be enforced by minimizing the regularization functional

$$\mathcal{S}_{\alpha,v}(u) := \|F(u) - v\|^p + \alpha \inf\Big\{\sum_{i\in\mathbb{N}} |u_i|^q : \sum_{i\in\mathbb{N}} u_i\phi_i = u\Big\}$$

over U, where $\alpha > 0$, $p \geq 1$, and $q < 2$.

The functional $\mathcal{S}_{\alpha,v}$ with $p = 2$ and $1 \leq q < 2$ has been proposed in [3], where convergence and stability of an iterative thresholding algorithm for the minimization of $\mathcal{S}_{\alpha,v}$ have been derived. In [6], the functional $\mathcal{S}_{\alpha,v}$ has been considered from an inverse problems point of view. Under the assumption that only noisy data v^δ satisfying $\|v - v^\delta\| \le \delta$ are available, the properties of minimizers u_α^δ of $\mathcal{S}_{\alpha,v^\delta}$ in dependence of α and δ have been analysed. In particular, convergence rates of order $O(\sqrt{\delta})$ were shown to hold under a certain range condition and parameter choice. These rates have been improved to $O(\delta^{1/q})$ in [5].

In contrast to the superlinear case $1 \le q < 2$, which by now is well-understood, there exist hardly any results concerning the non-convex regularization functional that arises for $0 < q < 1$. Though in [6] already a special case of this regularization method has been considered, no attempt for a systematic study has been made. This short note wants to draw the attention to this case and prove the theoretical suitability of sublinear penalty terms for the regularization of ill-posed problems with sparse solutions.

Using the results of [5], existence of a minimizer, stability of the set of minimizers with respect to the input data, and convergence as the noise approaches zero easily follow (see Thm. 2). In addition, it is shown that the choice of the exponent $0 < q < 1$ forces every (local) minimizer of $\mathcal{S}_{\alpha,v}$ to be sparse, as long as the operator F is locally Lipschitz (see Thm. 4). Finally, convergence rates of order $O(\delta)$ are derived under the assumption that a certain range condition is satisfied (see Thm. 5).

For proving the above mentioned properties, it is convenient to work directly with the coefficients of $u \in U$ with respect to $(\phi_i)_{i\in\mathbb{N}}$. To that end one regards the operator equation on the sequence space l^1 via the embedding $L\colon l^1 \to U$ that maps a sequence $u = (u_i)_{i\in\mathbb{N}}$ to $Lu := \sum_{i\in\mathbb{N}} u_i\phi_i$. Then one minimizes

$$\mathcal{T}_{\alpha,v}(u) := \|G(u) - v\|^p + \alpha\mathcal{R}(u) := \|G(u) - v\|^p + \alpha\sum_{i\in\mathbb{N}}|u_i|^q$$

over l^1 instead, where $G := F \circ L$. Note that the embedding L is well-defined and bounded, provided the sequence $(\phi_i)_{i\in\mathbb{N}} \subset U$ is bounded. Also, under this assumption the functionals $\mathcal{S}_{\alpha,v}$ and $\mathcal{T}_{\alpha,v}$ are equivalent in the sense that $\hat{u} \in U$ minimizes $\mathcal{S}_{\alpha,v}$ if and only if there exists $u \in l^1$ minimizing $\mathcal{T}_{\alpha,v}$ such that $Lu = \hat{u}$.

Throughout this paper the following assumptions hold:

- The space V is a reflexive Banach space.
- The parameters α, p, and q satisfy $\alpha > 0$, $p \ge 1$, and $0 < q < 1$.
- The operator $G\colon l^1 \to V$ is sequentially closed in the following sense: Whenever a bounded sequence $(u^{(k)})_{k\in\mathbb{N}} \subset l^1$ satisfies $u_i^{(k)} \to u_i$ for every $i \in \mathbb{N}$ and $G(u^{(k)}) \rightharpoonup v \in V$, then $G(u) = v$.

Remark 1. The notion of sequential convergence introduced above is in fact defined by the weak* topology on l^1. This topology is induced by the pre-dual c_0 of l^1, which consists of all sequences $(w_i)_{i\in\mathbb{N}}$ satisfying $w_i \to 0$ and is endowed with the norm $\|w\|_\infty := \max_{i\in\mathbb{N}}|w_i|$. ∎

Theorem 2 (Well-Posedness). *The following hold:*

1. *The functional* $\mathcal{T}_{\alpha,v}$ *has a minimizer for every* $\alpha > 0$ *and* $v \in V$.
2. *Assume that* $(v^{(k)})_{k\in\mathbb{N}} \subset V$ *converges to* v. *Then every sequence* $u^{(k)} \in \arg\min\{\mathcal{T}_{\alpha,v^{(k)}}(u) : u \in l^1\}$ *has a subsequence converging to a minimizer of* $\mathcal{T}_{\alpha,v}$.
3. *Assume that the equation* $G(u) = v$ *has a solution in* l^q. *Let* $(v^{(k)})_{k\in\mathbb{N}} \subset V$ *converge to* v *and let* $\alpha^{(k)} > 0$ *satisfy* $\alpha^{(k)} \to 0$ *and* $\|v^{(k)} - v\|^p/\alpha^{(k)} \to 0$ *as* $k \to \infty$. *Then every sequence* $u^{(k)} \in \arg\min\{\mathcal{T}_{\alpha^{(k)},v^{(k)}}(u) : u \in l^1\}$ *has a subsequence converging to an* $\mathcal{R}$*-minimizing solution of* $G(u) = v$.

Proof. The proof is along the lines of [5, Propositions 5–7]. □

Remark 3. Note that, in fact, the results in [5] imply that the stability and convergence assertions hold with respect to the functional $\mathcal{R}$. That is, the subsequence $(u^{(k_l)})_{l\in\mathbb{N}}$ in Item 2 satisfies $\mathcal{R}(u^{(k_l)} - u_\alpha) \to 0$ for some minimizer u_α of $\mathcal{T}_{\alpha,v}$, and the subsequence $(u^{(k_l)})_{l\in\mathbb{N}}$ in Item 3 satisfies $\mathcal{R}(u^{(k_l)} - u^\dagger) \to 0$ for some $\mathcal{R}$-minimizing solution $u^\dagger$ of $G(u) = v$. ■

Theorem 4 (Sparsity). *Assume that* $G\colon l^1 \to V$ *is locally Lipschitz. Then, every local minimizer of* $\mathcal{T}_{\alpha,v}$ *is sparse.*

Proof. See Section 2. □

Theorem 5 (Convergence Rates). *Let* $u^\dagger$ *be an* $\mathcal{R}$*-minimizing solution of* $G(u) = v$ *that is sparse. Assume that* $G\colon U \to V$ *is Gâteaux differentiable in* $u^\dagger$ *and that*

$$e_i \in \mathrm{Range}\big(G'(u^\dagger)^\#\big) \qquad \text{for every} \qquad i \in \mathbb{N}\,,$$

where $e_i \in l^\infty$ *denotes the* i*-th unit vector. Moreover, assume that there exist* γ_1, $\gamma_2 > 0$, $\sigma > 0$, *and* $\rho > \mathcal{R}(u^\dagger)$ *such that*

$$\mathcal{R}(u) - \mathcal{R}(u^\dagger) \geq \gamma_1 \big\|G'(u^\dagger)(u - u^\dagger)\big\| - \gamma_2 \big\|G(u) - G(u^\dagger)\big\| \tag{1}$$

for every $u \in l^1$ *satisfying* $\mathcal{R}(u) < \rho$ *and* $\|G(u) - G(u^\dagger)\| < \sigma$.

For $v^\delta \in U$ *satisfying* $\|v^\delta - v\| \leq \delta$ *and* $\alpha > 0$ *let* $u_\alpha^\delta \in \arg\min\{\mathcal{T}_{\alpha,v^\delta}(u) : u \in l^1\}$. *Assume that* $\alpha = \alpha(\delta)$ *satisfies* $C_1\alpha(\delta) \leq \delta^{p-1} \leq C_2\alpha(\delta)$ *for some* $C_2 \geq C_1 > 0$ *(or* $C_1 \leq \alpha(\delta) \leq C_2$ *for* C_2 *small enough in the case* $p = 1$*). Then*

$$\big\|u_\alpha^\delta - u^\dagger\big\|_1 = O(\delta) \qquad \text{as} \qquad \delta \to 0\,.$$

Proof. See Section 2. □

2 Proofs

Sparsity

We now prove Theorem 4, which asserts that local Lipschitz continuity of G implies that every local minimizer of $\mathcal{T}_{\alpha,v}$ is sparse.

Proof (of Theorem 4). Let u be a local minimizer of $\mathcal{T}_{\alpha,v}$. Denote by $I := \{i \in \mathbb{N} : u_i \neq 0\}$ the set of non-zero components of u. We have to show that I is a finite set.

The local minimality of u implies the existence of some $\varepsilon > 0$ such that $\mathcal{T}_{\alpha,v}(u) \leq \mathcal{T}_{\alpha,v}(w)$ whenever $\|u - w\|_1 < \varepsilon$. Since G is locally Lipschitz, there exists $C > 0$ such that

$$\big| \|G(u + te_i) - v\|^p - \|G(u) - v\|^p \big| \leq C|t|$$

whenever $i \in \mathbb{N}$ and $|t| < \varepsilon$.

Now let $i \in I$, let $|t| < \varepsilon$, and $w = u + te_i$. Then $\|u - w\|_1 = t < \varepsilon$ and therefore

$$\|G(u) - v\|^p + \alpha\mathcal{R}(u) \leq \|G(w) - v\|^q + \alpha\mathcal{R}(w) .$$

Since $\mathcal{R}(u) - \mathcal{R}(w) = |u_i|^q - |u_i + t|^q$, we obtain that

$$\alpha\big(|u_i|^q - |u_i + t|^q\big) \leq \|G(w) - v\|^p + \|G(u) - v\|^p \leq C|t| . \tag{2}$$

Now by assumption $u_i \neq 0$, which implies that the function $t \mapsto |u_i + t|^q$ is differentiable at $t = 0$ with derivative $q \operatorname{sgn}(u_i)|u_i|^{q-1}$. Thus (2) shows that

$$q|u_i|^{q-1} \leq C$$

whenever $i \in I$. Since $0 < q < 1$, this in turn implies that the absolute value of the non-zero coefficients u_i is bounded from below. On the other hand, $\mathcal{R}(u) = \sum_{i \in I} |u_i|^q < \infty$. This shows that the set I is finite, and therefore u is sparse. □

Convergence Rates

Now we derive the convergence rates asserted in Theorem 5.

Proof (of Theorem 5). Define the set $I := \{i \in \mathbb{N} : u_i^\dagger \neq 0\}$. Since by assumption $e_i \in \operatorname{Range}\big(G'(u^\dagger)^\#\big)$ for every $i \in I$ and I is finite, there exists $C_1 > 0$ such that

$$|w_i| = |\langle e_i, u\rangle| \leq C_1 \big\| G'(u^\dagger) w \big\|$$

for every $w \in l^1$ and $i \in I$. Moreover, there exists $C_2 > 0$ such that

$$|u_i^\dagger|^q - |t|^q \leq C_2 |u_i^\dagger - t|$$

for every $t \in \mathbb{R}$. Finally, there exists $C_3 > 0$ such that

$$|t| \leq C_3 |t|^q$$

for every $t \in \mathbb{R}$ satisfying $|t|^q \le \rho$.

Consequently,

$$\begin{aligned}\|u - u^\dagger\|_1 &= \sum_{i \notin I} |u_i| + \sum_{i \in I} |u_i - u_i^\dagger| \\ &\le C_3 \sum_{i \notin I} |u_i|^q + \sum_{i \in I} |u_i - u_i^\dagger| \\ &= C_3 \big(\mathcal{R}(u) - \mathcal{R}(u^\dagger)\big) + \sum_{i \in I} \big(|u_i - u_i^\dagger| - C_3 |u_i|^q + C_3 |u_i^\dagger|^q\big) \\ &\le C_3 \big(\mathcal{R}(u) - \mathcal{R}(u^\dagger)\big) + (1 + C_2 C_3) \sum_{i \in I} |u_i - u_i^\dagger| \\ &\le C_3 \big(\mathcal{R}(u) - \mathcal{R}(u^\dagger)\big) + (1 + C_2 C_3) C_1 |I| \|G'(u^\dagger)(u - u^\dagger)\|\end{aligned}$$

for every $u \in l^q$ satisfying $\mathcal{R}(u) \le \rho$. Here, $|I|$ denotes the number of elements contained in I. Defining $C_4 := (1 + C_2 C_3) C_1 |I| / \gamma_1$, we obtain from (1) that

$$\|u - u^\dagger\|_1 \le (C_3 + C_4) \big(\mathcal{R}(u) - \mathcal{R}(u^\dagger)\big) + C_4 \gamma_2 \|G(u) - G(u^\dagger)\|$$

for every $u \in l^q$ satisfying $\mathcal{R}(u) \le \rho$ and $\|G(u) - G(u^\dagger)\| < \sigma$. Now the assertion follows from [5, Prop. 8]. □

3 Summary

This paper shows that regularization with non-convex, sublinear regularization terms is well-posed and yields sparse results. Additionally, linear convergence rates have been derived for operators F satisfying a range condition that basically postulates that F acts 'nicely' on the subset $(\phi_i)_{i \in \mathbb{N}} \subset U$ with respect to which sparsity is required. In particular, the derived results imply that several properties of convex regularization methods carry over to certain non-convex ones, which therefore may also be applied to the regularization of inverse and ill-posed problems.

Acknowledgements

This work has been supported by the Austrian Science Fund (FWF) within the national research network Industrial Geometry, project 9203-N12.

References

[1] E. J. Candès, J. Romberg, and T. Tao. Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. *IEEE Trans. Inf. Theory*, 52(2):489–509, 2006.

[2] C. Chaux, P. L. Combettes, J.-C. Pesquet, and V. R. Wajs. A variational formulation for frame-based inverse problems. *Inverse Probl.*, 23(4):1495–1518, 2007.

[3] I. Daubechies, M. Defrise, and C. De Mol. An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. *Comm. Pure Appl. Math.*, 57(11):1413–1457, 2004.

[4] D. L. Donoho. Compressed sensing. *IEEE Trans. Inf. Theory*, 52(4):1289–1306, 2006.

[5] M. Grasmair, M. Haltmeier, and O. Scherzer. Sparse regularization with l^q penalty term. Reports of FSP S105 - "Photoacoustic Imaging" 1, University of Innsbruck, Austria, 2008. submitted.

[6] D. Lorenz. Convergence rates and source conditions for Tikhonov regularization with sparsity constraints. *arXiv:0801.1774v1, submitted*, 2008.