On the convergence speed of iterative methods for linear inverse problems with sparsity constraints

Dirk A. Lorenz

joint work with Kristian Bredies

Applied Inverse Problems 2007 - Vanocuver

Yet another talk on sparsity constraints. . .

$$\min_{u \in H} \frac{\|Ku - f\|^2}{2} + \Phi(u)$$

- Tikhonov functionals for ill-posed problems
- Regularized optimal control problems

Φ models prior knowledge or a known constraint, e.g:

A sparsity constraint

For a given basis (ψ_k) and convex ϕ:

$$\Phi(u) = \sum_k w_k |\langle u, \psi_k \rangle|^p \quad 1 \leq p \leq 2$$

Yet another talk on minmization with l^1...

$$\min_{u\in\ell^2(\mathbf{N})} \frac{\|Ku-f\|^2_{\ell^2(\mathbf{N})}}{2} + \sum_{k\in\mathbf{N}} w_k|u_k|^p$$

- Ill-posed problems:
  The functional $\|Ku-f\|^2$ has no minimizer.
- $1 \leq p < 2$: The functional $\sum_k w_k|\langle u,\psi_k\rangle|^p$ is non-smooth.

Yet another talk on minmization with l^1...

$$\min_{u\in\ell^2(\mathbf{N})} \frac{\|Ku - f\|^2_{\ell^2(\mathbf{N})}}{2} + \sum_{k\in\mathbf{N}} w_k |u_k|^p$$

- Ill-posed problems:
  The functional $\|Ku - f\|^2$ has no minimizer.
- $1 \leq p < 2$: The functional $\sum_k w_k |\langle u, \psi_k\rangle|^p$ is non-smooth.

In this talk we will see...

- ...that iterated hard shrinkage can minimize.
- ...why iterated soft shrinkage is so slow, and how slow it is.

What is known?

- $1 \leq p < 2$ gives a regularization (Daubechies, Defrise, DeMol)
- iterated soft shrinkage converges to minimizer (DDD)
- iterated soft shrinkage combined with fixed-point iteration converges for non-linear inverse problems (Ramlau, Teschke)
- soft shrinkage is a gradient descent (Bredies, L., Maass)
- iterated soft shrinkage converges without fixed-point iteration (Bredies, L., Maass)
- geometry of sparsity in the finite dimensional setting (e.g. Tropp, Donoho and others)
- higher order minimization in the finite dimensional case (Candes)

Open: How fast is the iterative soft shrinkage? What about efficient algorithms in the infinite dimensional case?

Constraint minimization: Generalized gradient projection method (GGPM)

$$\min_{u \in H} F(u) + I_U(u)$$

Constraint minimization: Generalized gradient projection method (GGPM)

$$\min_{u \in H} F(u) + I_U(u)$$

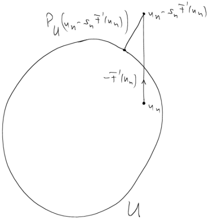
F' Lipschitz with constant L,
choose $u^0 \in U$ and stepsize $0 < s < \frac{2}{L}$.

Constraint minimization: Generalized gradient projection method (GGPM)

$$\min_{u \in H} F(u) + I_U(u)$$

F' Lipschitz with constant L,
choose $u^0 \in U$ and stepsize $0 < s < \frac{2}{L}$.
Iteration

$$u^{n+1} = P_U(u^n - sF'(u^n))$$



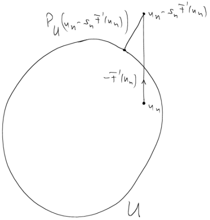
Projection:

$$P_U(u) = \operatorname{argmin} \tfrac{1}{2}\|v - u\|^2 + I_U(v)$$

Convergence analysis under construction...

Constraint minimization: Generalized gradient projection method (GGPM)

$$\min_{u \in H} F(u) + \Phi(u)$$



F' Lipschitz with constant L,
choose $u^0 \in U$ and stepsize $0 < s < \frac{2}{L}$.
Iteration

$$u^{n+1} = P_s(u^n - sF'(u^n))$$

Projection:

$$P_s(u) = \operatorname{argmin} \tfrac{1}{2}\|v - u\|^2 + s\Phi(v)$$

Convergence analysis under construction...

Reminder: Inverse Problems with sparsity

$$\min_{u \in H} F(u) + \Phi(u)$$

where

$$F(u) = \frac{\|Ku - f\|^2}{2} \quad \text{and} \quad \Phi(u) = \sum_k w_k |u_k|^p$$

Hence: $F'(u) = K^*(Ku - f)$ and $L = \|K\|^2$.

GGPM $\rightsquigarrow$ soft shrinkage

$$u^{n+1} = P_s(u^n - sK^*(Ku^n - f))$$

The generalized projection is

$$\begin{aligned} P_s(u) &= \operatorname{argmin} \tfrac{1}{2}\|v-u\|^2 + s\sum_k w_k |u_k|^p \\ &= \mathbf{S}_{p,sw}(u) \end{aligned}$$

GGPM $\rightsquigarrow$ soft shrinkage

$$u^{n+1} = P_s(u^n - sK^*(Ku^n - f))$$

The generalized projection is

$$\begin{aligned} P_s(u) &= \operatorname{argmin} \tfrac{1}{2}\|v - u\|^2 + s\sum_k w_k|u_k|^p \\ &= \mathbf{S}_{p,sw}(u) \end{aligned}$$

Iteration

$$u^{n+1} = \mathbf{S}_{p,sw}(u^n - sK^*(Ku^n - f))$$

Step-size $s = 1$ is OK for $\|K\| < \sqrt{2}$. (DDD: $\|K\| \leq 1$)

GGPM $\rightsquigarrow$ soft shrinkage

$$u^{n+1} = P_s(u^n - sK^*(Ku^n - f))$$

The generalized projection is

$$\begin{aligned} P_s(u) &= \operatorname{argmin} \tfrac{1}{2}\|v-u\|^2 + s\sum_k w_k |u_k|^p \\ &= \mathbf{S}_{p,sw}(u) \end{aligned}$$

Iteration

$$u^{n+1} = \mathbf{S}_{p,sw}(u^n - sK^*(Ku^n - f))$$

Step-size $s = 1$ is OK for $\|K\| < \sqrt{2}$. (DDD: $\|K\| \leq 1$)

Convergence and rates not proven yet...

Constraint minimization: Generalized conditional gradient method (GCGM)

$$\min_{u \in H} F(u) \text{ subject to } u \in U$$

1. Descent direction v^n by

$$\min_{v \in H} \langle F'(u^n), v \rangle \text{ subject to } v \in U$$

2. Choose a stepsize $s^n > 0$.
3. Update

$$u^{n+1} = u^n + s^n(v^n - u^n)$$

Constraint minimization: Generalized conditional gradient method (GCGM)

$$\min_{u\in H} F(u) + I_U(u)$$

1. Descent direction v^n by
$$\min_{v\in H}\langle F'(u^n), v\rangle + I_U(v)$$
2. Choose a stepsize $s^n > 0$.
3. Update
$$u^{n+1} = u^n + s^n(v^n - u^n)$$

Constraint minimization: Generalized conditional gradient method (GCGM)

$$\min_{u\in H} F(u) + \Phi(u)$$

1. Descent direction v^n by
$$\min_{v\in H}\langle F'(u^n), v\rangle + \Phi(v)$$
2. Choose a stepsize $s^n > 0$.
3. Update
$$u^{n+1} = u^n + s^n(v^n - u^n)$$

for Φ convex and lsc.
Note: $v_n = (\partial\Phi)^{-1}(-F'(u^n))$.

Convergence Analysis of GCGM

Theorem

Let $F(u) = \frac{1}{2}\|Ku - f\|^2$ and, Φ fulfill

- *Φ is proper, convex and lsc*
- *$\partial\Phi$ is onto and $(\partial\Phi)^{-1}$ bounded.*

Then the generalized conditional gradient method converges strongly to a minimizer of $F + \Phi$. *(Bredies, L. 2006)*

Remark

A local coercivity estimate around a minimizer u^* like

$$\langle K^*(Ku^* - f), v - u^* \rangle + \Phi(v) - \Phi(u^*) \geq C\|v - u^*\|^2$$

guarantees linear convergence.

A little workaround

Problem

$\Phi(u) = \sum_k w_k |u_k|^p$ is convex and lsc
but $\partial\Phi$ is not onto for $p = 1$.

$\rightsquigarrow \ v^n = (\partial\Phi)^{-1}(-K^*(Ku^n - f))$ must not exist!

A little workaround

Problem

$\Phi(u) = \sum_k w_k |u_k|^p$ is convex and lsc
but $\partial\Phi$ is not onto for $p = 1$.

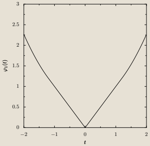
$\rightsquigarrow \; v^n = (\partial\Phi)^{-1}(-K^*(Ku^n - f))$ must not exist!

Workaround

Modify Φ for large values:

$$\tilde{\Phi}(u) = \sum_k w_k \phi(u_k)$$

(Quadratic extension in a C^1 way)



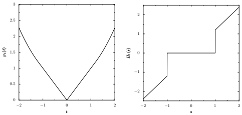
OK because for minimizer u^*: $|u_k^*| \leq \left(\frac{\|f\|^2}{2w_0}\right)^{1/p} =: S_0$.

Workaround $\rightsquigarrow$ hard shrinkage

The descent direction is

$$v^n = (\partial\Phi)^{-1}(-K^*(Ku^n - f))$$

and $(\partial\Phi)^{-1} =: \mathbf{H}_{p,w}$ is hard shrinkage !

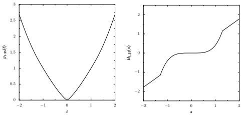


Workaround $\rightsquigarrow$ hard shrinkage

The descent direction is

$$v^n = (\partial\Phi)^{-1}(-K^*(Ku^n - f))$$

and $(\partial\Phi)^{-1} =: \mathbf{H}_{p,w}$ is hard shrinkage !

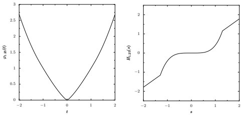


Workaround $\rightsquigarrow$ hard shrinkage

The descent direction is

$$v^n = (\partial\Phi)^{-1}(-K^*(Ku^n - f))$$

and $(\partial\Phi)^{-1} =: \mathbf{H}_{p,w}$ is hard shrinkage !



Iteration

$$u^{n+1} = (1 - s_n)u^n + s_n\mathbf{H}_{p,w}(-K^*(Ku^n - f))$$

(Step-size $s^n = \min(1, \frac{\Phi(u^n)-\Phi(v^n)+\langle Ku^n-f,K(u^n-v^n)\rangle}{\|K(v^n-u^n)\|^2})$ guarantees convergence.)

Properties of the iterated hard shrinkage

Theorem

Let u^* be a minimizer of

$$\Psi(u) = \tfrac{1}{2}\|Ku - f\|^2 + \textstyle\sum_k w_k |u_k|^p$$

and u^n, v^n be generated by the iterated hard shrinkage.

- *for $1 < p \leq 2$ it holds*
$$\|u^n - u^*\| \leq C\lambda^n$$
- *for $p = 1$ and K injective it holds*
$$\|u^n - u^*\| \leq Cn^{-1/2}.$$

Furthermore the estimate

$$D_n = \Phi(u^n) - \Phi(v^n) + \langle K^*(Ku^n - f), u^n - v^n\rangle \geq \tilde{\Psi}(u^n) - \tilde{\Psi}(u^*)$$

can serve as a stopping criterion.

Bredies, L. (2006)

Higher order methods for $p = 1$?

The equation

$$u = \mathbf{S}_w(u - K^*(Ku - f))$$

is almost smooth! $\rightsquigarrow$ Newton-methods?

Higher order methods for $p = 1$?

The equation

$$u = \mathbf{S}_w(u - K^*(Ku - f))$$

is almost smooth! $\rightsquigarrow$ Newton-methods?

Theorem

The function $F(u) = u - \mathbf{S}_w(u - K^*(Ku - f))$ is slantly differentiable and the Newton-method

$$u^{n+1} = u^n - F'(u^n)^{-1}F(u^n)$$

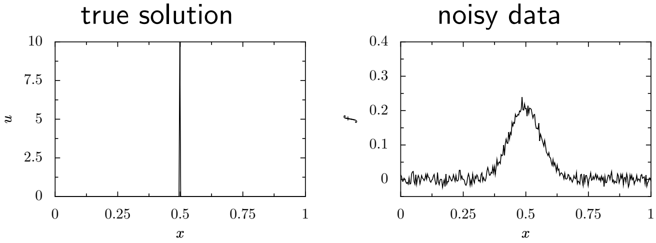
converges locally and superlinearly. *(Griesse, L., 2007)*

Per step: solve small (esp. finite) linear system.

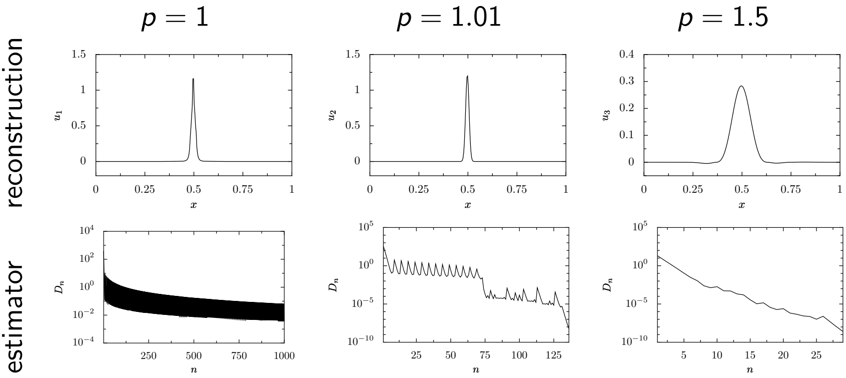
Backwards heat conduction

$$u_t = u_{xx} \quad \text{for } (t,x) \in [0,T] \times [0,1]$$
$$u(0,x) = u^0(x)$$
$$u(t,0) = u(t,1) = 0.$$

$$Ku^0 = u(T)$$



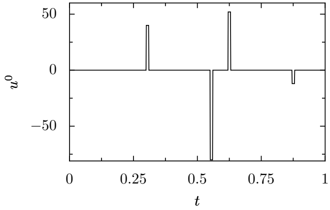
Backwards heat conduction



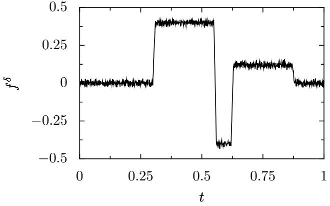
Inverse integration

$$Ku(t) = \int_0^t u(s)\mathrm{d}s, \quad s, t \in [0,1], \quad u \in L^2[0,1].$$

true solution



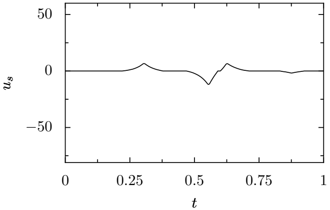
noisy data



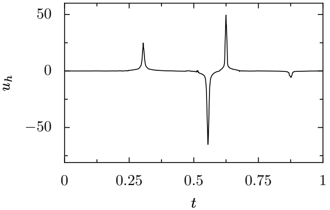
Inverse integration

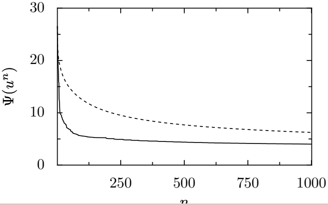
After 1000 iterations of…

…soft shrinkage



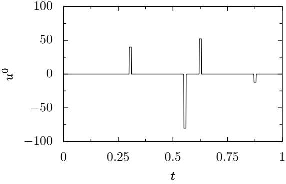
…hard shrinkage



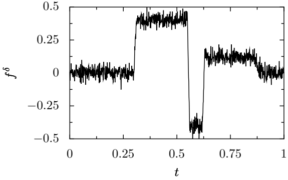


Experiments with Newton

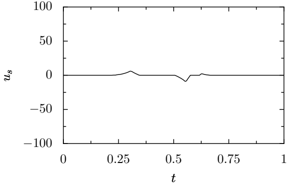
true solution



noisy data

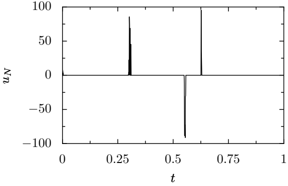


1000 soft shrinkage iterations



$\Psi(u_s) = 15.22$

. . . and 11 Newton iterations



$\Psi(u_N) = 13.90$

Conclusion

- Tikhonov functionals with sparsity constraints can be minimized by iterated hard shrinkage as well as soft shrinkage.
- Both are generalized gradient methods.
- The iterated hard shrinkage is very easy to implement.
- The convergence speed changes drastically from $p > 1$ to $p = 1$.
- The distance to the minimizer can be estimated easily.

Conclusion

- Tikhonov functionals with sparsity constraints can be minimized by iterated hard shrinkage as well as soft shrinkage.
- Both are generalized gradient methods.
- The iterated hard shrinkage is very easy to implement.
- The convergence speed changes drastically from $p > 1$ to $p = 1$.
- The distance to the minimizer can be estimated easily.

Outlook:

- Globalize convergence of Newton methods?
- Use small p to get good initial value for $p = 1$?