On Generalized Gradient Methods for Non-differentiable Functionals

Application to Tikhonov Functionals with Sparsity Constraints

Dirk A. Lorenz – joint work with Kristian Bredies

RICAM Linz, 2007-02-27

1 The Problem: Sparsity Constraints

2 Gradient Methods
- Classical Gradient Methods
- Generalizations

3 Application to Tikhonov Functionals
- Soft shrinkage
- Hard shrinkage

4 Examples
- Backwards heat condution
- Inverse integration

1 The Problem: Sparsity Constraints

2 Gradient Methods
- Classical Gradient Methods
- Generalizations

3 Application to Tikhonov Functionals
- Soft shrinkage
- Hard shrinkage

4 Examples
- Backwards heat condution
- Inverse integration

What this talk is about

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

- 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:

$$\Phi(u) = \sum_k \phi(\langle u, \psi_k \rangle),$$

What this talk is about

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

- 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:

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

What this talk is about

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

- 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:

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

What this talk is about

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

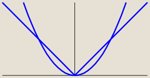
- Tikhonov functionals for ill-posed problems
- Regularized optimal control problems

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

A sparsity constraint

For a given basis (ψ_k) and convex ϕ:

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



Problems with minimization

$$\min_{u\in\ell^2} \frac{\|Ku - f\|_{\ell^2}^2}{2} + \sum_k w_k |u_k|^p$$

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

Goal of this talk: Develop generalized gradient methods for the minimization.

1 The Problem: Sparsity Constraints

2 Gradient Methods
- Classical Gradient Methods
- Generalizations

3 Application to Tikhonov Functionals
- Soft shrinkage
- Hard shrinkage

4 Examples
- Backwards heat condution
- Inverse integration

Constraint minimization: Conditional gradient method

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

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: Conditional gradient method

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

1. Descent direction v^n by

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

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

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

Constraint minimization: Conditional gradient method

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

1. Descent direction v^n by

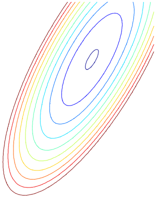
$$\min_{v\in H} \langle F'(u^n), v\rangle + I_U(v)$$

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

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

Generalization of the conditional gradient method

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



1. Descent direction v^n by

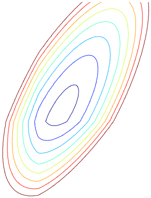
$$\min_{v \in H} \langle F'(u^n), v \rangle + \Phi(v)$$

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

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

Generalization of the conditional gradient method

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



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

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

Convergence Analysis

Condition

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

Theorem

Let F be Lipschitz and convex, Φ like above. Then the generalized conditional gradient method converges to a minimizer of $F + \Phi$.

(Descent direction is $v^n = (\partial\Phi)^{-1}(-F'(u^n))$.)

1 The Problem: Sparsity Constraints

2 Gradient Methods
- Classical Gradient Methods
- Generalizations

3 Application to Tikhonov Functionals
- Soft shrinkage
- Hard shrinkage

4 Examples
- Backwards heat condution
- Inverse integration

Inverse Problems with sparsity

$$\min_{u \in H} \frac{\|Ku - f\|^2}{2} + \underbrace{\sum_k w_k |u_k|^p}_{\Phi(u)}$$

For $p = 1$:

- Φ is convex and lsc,
- but $\partial\Phi(u) = \sum_k w_k \operatorname{sgn}(u_k)\psi_k$ is not onto.

Workaround 1 $\rightsquigarrow$ soft shrinkage

$$\min_{u \in H} \frac{\|Ku - f\|^2}{2} - \frac{\lambda\|u\|^2}{2} + \underbrace{\frac{\lambda\|u\|^2}{2} + \sum_k w_k |u_k|^p}_{\Phi(u)}$$

$\partial\Phi(u) = \lambda u + \sum_k w_k \operatorname{sgn}(u_k)\psi_k$
and hence the descent direction is

$$\begin{aligned} v^n &= (\partial\Phi)^{-1}(\lambda u^n - K^*(Ku^n - f)) \\ &= S_{w/\lambda}(u^n - K^*(Ku^n - f)/\lambda) \end{aligned}$$

(For $\|K\| \leq 1, \lambda = 1$ stepsizes$^n = 1$ is ok,
$\rightsquigarrow u^{n+1} = v^n = S_w(u^n - K^*(Ku^n - f))$.)

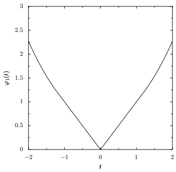
Workaround 2

Estimate the norm of the minimizer u^*:

$$\|u^*\|^p \leq \frac{(F+\Phi)(u^*)}{w_0} \leq \frac{(F+\Phi)(0)}{w_0} = \frac{\|f\|^2}{2w_0}$$

Hence,

$$|u_k| \leq \left(\frac{\|f\|^2}{2w_0}\right)^p =: S_0.$$



Modify Φ for large values:

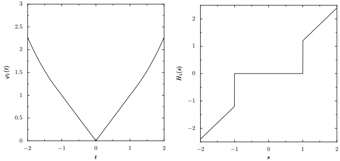
$$\tilde{\Phi}(u) = \sum_k w_k \phi(u_k)$$

(Quadratic extension in a C^1 way)

Workaround 2 $\rightsquigarrow$ hard shrinkage!

The descent direction is

$$\begin{aligned} v^n &= (\partial\Phi)^{-1}(-K^*(Ku^n - f)) \\ v_k^n &= (\partial\phi)^{-1}(-K^*(Ku^n - f)_k / w_k) \end{aligned}$$

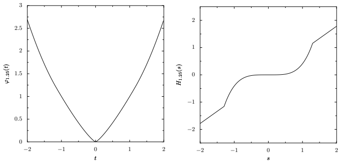


(The special step-size $s^n = \min(1, \frac{\Phi(u^n) - \Phi(v^n) + \langle Ku^n - f, K(u^n - v^n)\rangle}{\|K(v^n - u^n)\|^2})$ guarantees convergence.)

Workaround 2 $\rightsquigarrow$ hard shrinkage!

The descent direction is

$$
\begin{aligned}
v^n &= (\partial\Phi)^{-1}(-K^*(Ku^n - f)) \\
v^n_k &= (\partial\phi)^{-1}(-K^*(Ku^n - f)_k / w_k)
\end{aligned}
$$



(The special 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

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

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)

1 The Problem: Sparsity Constraints

2 Gradient Methods
- Classical Gradient Methods
- Generalizations

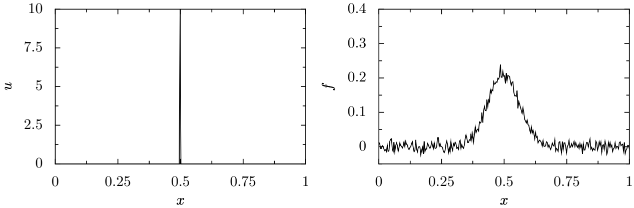
3 Application to Tikhonov Functionals
- Soft shrinkage
- Hard shrinkage

4 Examples
- Backwards heat condution
- Inverse integration

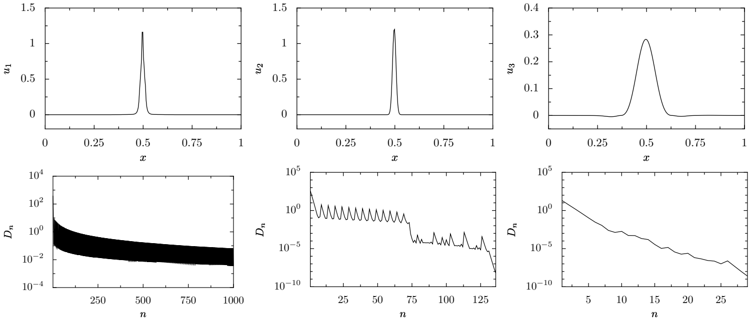
Backwards heat conduction

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

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

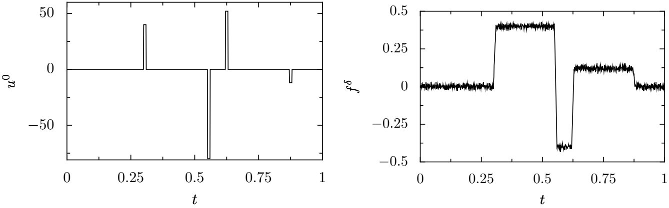


Backwards heat conduction

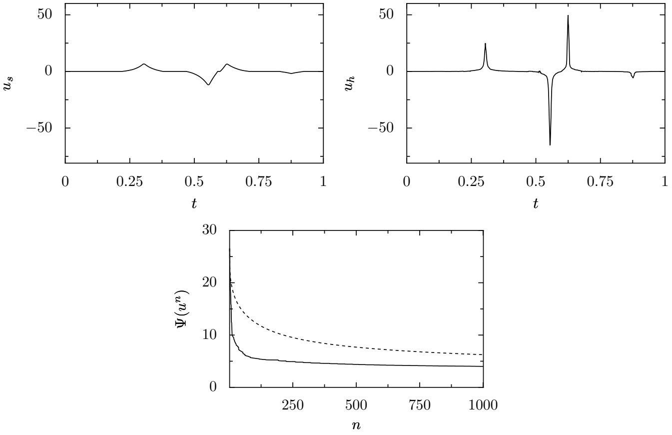


Inverse integration

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



Inverse integration



Conclusion

- Tikhonov functionals with sparsity constraints can be minimized by iterated hard shrinkage.
- 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.