The Lepskiĭ balancing principle for conjugate gradient regularization

Peter Mathé

Weierstrass Institute Berlin

Warszawa, 09 lutego 2010

Outline

Krylov minimization

Conjugate gradient regularization (*cg*) is an iterative solver for linear equations in Hilbert space. Here we shall apply this to noisy equations

$$y^\delta = Tx + \delta\xi,$$

where

- $T\colon X \to Y$ has non-closed range (ill-posedness),
- ξ is bounded deterministic noise,
- $\delta > 0$ is the noise level.

Krylov minimization

Conjugate gradient regularization (*cg*) is an iterative solver for linear equations in Hilbert space. Here we shall apply this to noisy equations

$$y^\delta = Tx + \delta\xi,$$

where

- $T\colon X \to Y$ has non-closed range (ill-posedness),
- ξ is bounded deterministic noise,
- $\delta > 0$ is the noise level.

cg determines at each step k the element x_k^δ as minimizer of

$$\delta_k := \|y^\delta - Tx_k^\delta\| = \min\left\{\|y^\delta - Tx\|, \quad x \in \mathcal{K}_k(T^* y^\delta, T^* T)\right\},$$

where

$$\mathcal{K}_k(T^* y^\delta, T^* T) = \operatorname{span}\left\{(T^* T)^{k-1}\, T^* y^\delta, \quad k = 1, \ldots, k\right\},$$

denotes the kth *Krylov-subspace*.

Background

cg has a clear geometric interpretation, as it determines the descent directions such that the consecutive residuals are *TT^*-orthogonal*, i.e.,

$$\langle T^*(y^\delta - Tx_k^\delta), T^*(y^\delta - Tx_{k+1}^\delta)\rangle = 0.$$

Therefore, it should be called *conjugate residuals*!
The resulting algorithmic description is short, and 'easy to compute'.

Remark

- *M. R. Hestenes and E. Stiefel [3], 1952, original study,*
- *J. R. Shewchuk [5], 1994, "An Introduction to the Conjugate Gradient Method Without the Agonizing Pain", the most cited introduction to the subject.*

Using *cg* for regularization

In regularization *cg* is used for the *normal equations*

$$T^* y^\delta = T^* T x + \delta T^* \xi,$$

Since x_k^δ solves a least-squares problem:
If $y^\delta \notin \mathcal{R}(T)$ then $\|x_k^\delta\|$ must explode as $k \to \infty$. Therefore regularization (stopping) $k = k(\delta, y^\delta)$ is necessary.

Remark

- *A. Nemirowskiĭ [4], 1986, "Regularizing properties of the conjugate gradient method in ill-posed problems", original study,*
- *M. Hanke [2], 1995, monograph, polishing the first study,*
- *Engl, Hanke and Neubauer [1], 1996, chapter of* cg*, that's where we refer to!*

Challenges

We recall that x_k^δ solves a minimization problem in a Krylov subspace

$$x_k^\delta = g_k(T^*T)T^*y^\delta, \quad \deg(g_k) \leq k - 1.$$

The polynomial $g_k = g_k(y^\delta)$ depends on the data y^δ, therefore *cg* is a *nonlinear iterative method*!

Consequently,

- there is no *immediate bias-variance decomposition*,
- no control of growth of the noise propagation ,
- no *a priori parameter choice*, and
- it is hard to check the *qualification* of *cg*.

Challenges

We recall that x_k^δ solves a minimization problem in a Krylov subspace

$$x_k^\delta = g_k(T^*T)T^*y^\delta, \quad \deg(g_k) \leq k-1.$$

The polynomial $g_k = g_k(y^\delta)$ depends on the data y^δ, therefore *cg* is a *nonlinear iterative method*!

Consequently,

- there is no *immediate bias-variance decomposition*,
- no control of growth of the noise propagation ,
- no *a priori parameter choice*, and
- it is hard to check the *qualification* of *cg*.

Theorem (Nemirovskiĭ)

If $x \in \mathcal{R}\,(T^*T)^\mu$, and if K_{DP} then is according to the discrepancy principle *then*

$$\|x - x_{K_{DP}}^\delta\| \leq C\delta^{\mu/(\mu+1/2)}.$$

Goals, achievements

- We want to extend *cg* to statistical ill-posed problems. Therefore, the discrepancy principle causes problems!
- Need other parameter choice.

In our study we achieved

- application of the *Lepskiĭ balancing principle* for parameter choice,
- to obtain an *oracle inequality*, and
- to extend *cg* to (a class of) *general smoothness assumptions*.

Goals, achievements

- We want to extend *cg* to statistical ill-posed problems. Therefore, the discrepancy principle causes problems!
- Need other parameter choice.

In our study we achieved

- application of the *Lepskiĭ balancing principle* for parameter choice,
- to obtain an *oracle inequality*, and
- to extend *cg* to (a class of) *general smoothness assumptions*.

Definition

An index function ψ is said to be majorized by the power μ *if the function $t \mapsto t^{\mu}/\psi(t)$ is an index function. An index function is said to be* majorized by a power *if it is majorized by the power μ for some $\mu > 0$.*

Goals, achievements

- We want to extend *cg* to statistical ill-posed problems. Therefore, the discrepancy principle causes problems!
- Need other parameter choice.

In our study we achieved

- application of the *Lepskiĭ balancing principle* for parameter choice,
- to obtain an *oracle inequality*, and
- to extend *cg* to (a class of) *general smoothness assumptions*.

Definition

An index function ψ is said to be majorized by the power μ *if the function $t \mapsto t^{\mu}/\psi(t)$ is an index function. An index function is said to be* majorized by a power *if it is majorized by the power μ for some $\mu > 0$.*

Remark

G. Blanchard, P. Mathé, Conjugate gradient regularization under general smoothness assumptions, 2010.

Outline

Error decomposition

Theorem

Suppose that ψ is majorized by a power and that $x \in H_\psi$. There is a constant C such that

$$\|x - x_k^\delta\| \leq C\psi\left(\Theta_\psi^{-1}(d_k)\right) + 3\frac{\delta}{\sqrt{\alpha_k}},$$

where $\Theta_\psi(t) = \sqrt{t}\psi(t)$, and

- $d_k := \max\left\{\delta, \|y^\delta - Tx_k^\delta\|\right\}$, *and* $\alpha_k := \left|r_k'(0)\right|^{-1}$, $k = 1, 2, \ldots$.

Error decomposition

Theorem

Suppose that ψ is majorized by a power and that $x \in H_\psi$. There is a constant C such that

$$\|x - x_k^\delta\| \leq C\psi\left(\Theta_\psi^{-1}(d_k)\right) + 3\frac{\delta}{\sqrt{\alpha_k}},$$

where $\Theta_\psi(t) = \sqrt{t}\psi(t)$, and

- $d_k := \max\left\{\delta, \|y^\delta - Tx_k^\delta\|\right\}$, *and* $\alpha_k := \left|r_k'(0)\right|^{-1}$, $k = 1, 2, \ldots$.

Remark

Notice that

- $\alpha_k \searrow 0$ *as* $k \to \infty$, *and* $\|y^\delta - Tx_k^\delta\| \searrow 0$ *as* $k \to \infty$.
- $d_k \not\to 0$ *as* $k \to \infty$. *This reflects that the first summand subsumes* non-linearity *terms.*
- *There is* no control of increase of $\left|r_k'(0)\right|$*!*

What is $|r_k'(0)|$?

- Recall that $x_k^\delta := g_k(T^*T)T^*y^\delta$ with polynomial g_k.
- We assign $r_k(\lambda) := 1 - \lambda g_k(\lambda)$.
- This is a *polynomial of* $\deg(r_k) = k,\ r_k(0) = 1$.
- The polynomials $r_k, k = 1, \dots$ are *orthogonal* with respect to $d\mu(\lambda) := \lambda\, d\|F_\lambda y^\delta\|^2$.
- The quantity $\left|r_k'(0)\right|$ is the (abs. value of) *first derivative*.
- It can be shown that $\left|r_k'(0)\right| \nearrow \infty$ as $k \to \infty$.
- The polynomials are easily calculated along with the steps of *cg*.

cg under the discrepancy principle

Theorem

Suppose that ψ is majorized by a power, *and that* K_{DP} is according to the discrepancy principle. *Then there is a constant* $C < \infty$ *such that for* $x \in H_\psi$ *we have that*

$$\|x - x^\delta_{K_{DP}}\| \leq C\psi\left(\Theta_\psi^{-1}(\delta)\right), \quad \delta \to 0,$$

provided that $K_{DP} > 1$.

cg under the discrepancy principle

Theorem

Suppose that ψ is majorized by a power, *and that* K_{DP} is according to the discrepancy principle. *Then there is a constant* $C < \infty$ *such that for* $x \in H_\psi$ *we have that*

$$\|x - x^\delta_{K_{DP}}\| \leq C\psi\left(\Theta_\psi^{-1}(\delta)\right), \quad \delta \to 0,$$

provided that $K_{DP} > 1$.

- This extends the Nemirovskiĭ result.
- Since *qualification* can only be checked for *classical source conditions*, we use the *method of distance functions* as in B. Hofmann, P. Mathé, *Analysis of profile functions for general linear regularization methods, SINUM, 2007.*

Oracle inequality

Assume $\{x_1, \dots, x_m\}$ is a finite set in a metric space (M, d), and $x \in M$ ("the truth") satisfies $d(x, x_i) \leq \frac{1}{2}(\Phi(i) + \Psi(i))$, $i = 1, \dots, m$, where $\Phi : \{1, \dots, m\} \to \mathbb{R}^+$ *is non-decreasing*, and $\Psi : \{1, \dots, m\} \to \mathbb{R}^+$ *is non-increasing*.

Fix any $K > 1$, and define the set

$$\Delta = \left\{1 \leq j \leq m : d(x_i, x_j) \leq K\Psi(i), \text{ for all } i \leq j\right\}.$$

Oracle inequality

Assume $\{x_1, \dots, x_m\}$ is a finite set in a metric space (M, d), and $x \in M$ ("the truth") satisfies $d(x, x_i) \leq \frac{1}{2}\left(\Phi(i) + \Psi(i)\right), \; i = 1, \dots, m,$ where $\Phi : \{1, \dots, m\} \to \mathbb{R}^+$ *is non-decreasing*, and $\Psi : \{1, \dots, m\} \to \mathbb{R}^+$ *is non-increasing*.

Fix any $K > 1$, and define the set

$$\Delta = \left\{1 \leq j \leq m : d(x_i, x_j) \leq K\Psi(i), \; \text{for all } i \leq j\right\}.$$

Definition

Any integer satisfying $\bar{j} \in \Delta$ and $(\bar{j} + 1) \notin \Delta$ is called Lepskiĭ parameter.

Oracle inequality

Assume $\{x_1, \dots, x_m\}$ is a finite set in a metric space (M, d), and $x \in M$ ("the truth") satisfies $d(x, x_i) \leq \frac{1}{2}(\Phi(i) + \Psi(i))$, $i = 1, \dots, m$, where $\Phi : \{1, \dots, m\} \to \mathbb{R}^+$ *is non-decreasing*, and $\Psi : \{1, \dots, m\} \to \mathbb{R}^+$ *is non-increasing*.

Fix any $K > 1$, and define the set

$$\Delta = \left\{1 \leq j \leq m : d(x_i, x_j) \leq K\Psi(i), \text{ for all } i \leq j\right\}.$$

Definition

Any integer satisfying $\bar{j} \in \Delta$ and $(\bar{j} + 1) \notin \Delta$ is called Lepskiĭ parameter.

Proposition (Abstract oracle inequality)

Any Lepskiĭ parameter $\bar{j}$ satisfies

$$d(x, x_{\bar{j}}) \leq \frac{1}{2} \min_{1 \leq i \leq m} \left\{\frac{K}{K-1}\Phi(i) + (1 + 2K)\Psi(i)\right\}.$$

cg under Lepskiĭ parameter choice

Theorem

Suppose that ψ is majorized by a power, *and that* K_L is the Lepskiĭ parameter choice. *Then there is a constant* $C < \infty$ *such that for* $x \in H_\psi$ *we have that*

$$\|x - x^\delta_{K_L}\| \le C\psi\left(\Theta_\psi^{-1}(\delta)\right), \quad \delta \to 0,$$

provided that $1 < K_L < k_{\max}$.

cg under Lepskiĭ parameter choice

Theorem

Suppose that ψ is majorized by a power, *and that* K_L is the Lepskiĭ parameter choice. *Then there is a constant* $C < \infty$ *such that for* $x \in H_\psi$ *we have that*

$$\|x - x^\delta_{K_L}\| \leq C\psi\left(\Theta_\psi^{-1}(\delta)\right), \quad \delta \to 0,$$

provided that $1 < K_L < k_{\max}$.

Remark

Although the oracle bound holds without any assumptions, the optimal order can be guaranteed only if K_L is internal.

- Heinz W. Engl, Martin Hanke, and Andreas Neubauer.
  Regularization of inverse problems, volume 375 of *Mathematics and its Applications*.
  Kluwer Academic Publishers Group, Dordrecht, 1996.
- Martin Hanke.
  Conjugate gradient type methods for ill-posed problems, volume 327 of *Pitman Research Notes in Mathematics Series*.
  Longman Scientific & Technical, Harlow, 1995.
- Magnus R. Hestenes and Eduard Stiefel.
  Methods of conjugate gradients for solving linear systems.
  J. Research Nat. Bur. Standards, 49:409–436 (1953), 1952.
- A. S. Nemirovskiĭ.
  Regularizing properties of the conjugate gradient method in ill-posed problems.
  Zh. Vychisl. Mat. i Mat. Fiz., 26(3):332–347, 477, 1986.
- Jonathan R Shewchuk.

An introduction to the conjugate gradient method without the agonizing pain.
Technical report, Carnegie Mellon University, Pittsburgh, PA, USA, 1994.