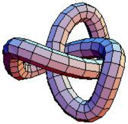
Accelerated Landweber iteration in Banach spaces

TORSTEN HEIN

Chemnitz University of Technology
Department of Mathematics
09107 Chemnitz, Germany



Warsaw, February 2010

torsten.hein@mathematik.tu-chemnitz.de

Outline

The situation

- $\mathcal{X}$ and $\mathcal{Y}$ denote real Banach spaces with dual spaces $\mathcal{X}^*$ and $\mathcal{Y}^*$, corresponding norms $\|\cdot\|$ and duality products $\langle\cdot,\cdot\rangle$.
- $\mathcal{X}$ is supposed to be reflexive.
- The linear and bounded operator $A:\mathcal{X}\longrightarrow\mathcal{Y}$ has non-closed range, i.e. $\overline{\mathcal{R}(A)}\neq\mathcal{R}(A)$.
- We consider the ill-posed operator equation

$$Ax=y, \qquad x\in\mathcal{X},\ y\in\mathcal{Y}$$

- given: only noisy data y^δ with $\|y-y^\delta\|\leq\delta$
- There exists a solution $x^\dagger$ for given exact data, i.e. $Ax^\dagger=y$ holds.

Tikhonov regularization

Tikhonov-like regularization approach:

- choose $1 < p, s < \infty$,
- define Tikhonov functional with norm penalty term

$$T_{\alpha,y^\delta}(x) := \frac{1}{p}\|Ax - y^\delta\|^p + \frac{\alpha}{s}\|x\|^s$$

  for given regularization parameter $\alpha > 0$ and
- calculate

$$x_\alpha^\delta := \text{argmin } \left\{ T_{\alpha,y^\delta}(x) \; : \; x \in \mathcal{X} \right\}$$

  as regularized approximate solution of equation $Ax = y$.

For 'optimal' choice of parameter α:

- have to solve the minimization problem several times for different regularization parameter α

$\Rightarrow$ numerically ineffective

Iterative regularization

- instead of Tikhonov regularization we deal with iterative approaches for solving the minimization problem

$$\Phi(x) := \frac{1}{p}\|Ax - y^\delta\|^p \to \min$$

- regularization: early termination of the iteration (stopping criterion)
- Minimization: gradient type method
- classical Landweber iteration (in Hilbert spaces): constant step size
- acceleration: control of the step size in each iteration
- still slow convergence expected, but:

The stopping criterion often terminates the iteration before convergence becomes too slow!

Outline

Duality mappings

In order to iterate in the correct space we need duality mappings:

Definition

Let $1 < p < \infty$.

- The *duality mapping* $J_p : \mathcal{X} \longrightarrow 2^{\mathcal{X}^*}$ with gauge function $t \mapsto t^{p-1}$ is defined as

$$J_p(x) := \{x^* \in \mathcal{X}^* \ : \ \langle x^*, x\rangle = \|x^*\|\|x\|, \|x^*\| = \|x\|^{p-1}\}.$$

- The Banach space $\mathcal{X}$ is called *smooth* if duality mappings are always single valued.

Example

Assume $\mathcal{X} = L^r(0,1)$, $1 < r < \infty$. Then $J_p(x) = \|x\|^{p-r}|x|^{r-1}\mathrm{sgn}(x)$, $x \neq 0$.

The parameter p is just a scaling factor. It is more important to choose the space $\mathcal{X}$ (and $\mathcal{Y}$) suitably!

Duality mappings

Properties:

- If $J_p(x)$ is single-valued (i.e. if $\mathcal{X}$ is smooth) then

$$J_p(x) = \left(\frac{1}{p}\|x\|^p\right)' \in \mathcal{X}^*.$$

- J_2 is linear if and only if $\mathcal{X}$ is a Hilbert space.
- $\mathcal{X}$ reflexive and $\mathcal{X}, \mathcal{X}^*$ are smooth then $J_p^{-1} = J_{p^*}^* : \mathcal{X}^* \longrightarrow \mathcal{X}$ with $(p^*)^{-1} + p^{-1} = 1$.

In the following we assume $\mathcal{X}, \mathcal{X}^*$ and $\mathcal{Y}$ to be smooth.

Notation: duality mappings $J_p : \mathcal{Y} \longrightarrow \mathcal{Y}^*$ and $J^*_{s^*} : \mathcal{X}^* \longrightarrow \mathcal{X}$ with $(s^*)^{-1} + s^{-1} = 1$, $A^\star : \mathcal{Y}^* \longrightarrow \mathcal{X}^*$ adjoint of A.

Algorithm – general scheme

(S0) Init. Choose $1 < p, s < \infty$, $x_0^* \in \mathcal{X}^*$, $x_0^\delta := J^*_{s^*}(x_0^*)$. Choose upper bound $\overline{\mu} \in (0, \infty]$ for the step size and define parameter $\tau > 1$. Set $n := 0$.

(S1) STOP, if for $\delta > 0$ the discrepancy criterion $\|A x_n^\delta - y^\delta\| \le \tau\delta$ is fulfilled or we have $A x_n^\delta = y$ for $\delta = 0$.

(S2) Calculate $\psi_n^* := \Phi'(x_n^\delta) = A^\star J_p(A x_n^\delta - y^\delta)$ and determine the step size $\mu_n > 0$.

(S3) Calculate the new iterate

$$x_{n+1}^* := x_n^* - \mu_n \psi_n^*, \quad x_{n+1}^\delta := J^*_{s^*}(x_{n+1}^*).$$

Set $n := n + 1$ and go to step (S1).

Outline

Bregman distances

Bregman distance of the functional $x \mapsto \frac{1}{s}\|x\|^s$:

$$\begin{aligned}\Delta_s(\tilde{x}, x) &:= \frac{1}{s}\|\tilde{x}\|^s - \frac{1}{s}\|x\|^s - \langle J_s(x), \tilde{x} - x\rangle \\ &= \frac{1}{s}\|\tilde{x}\|^s + \frac{1}{s^*}\|x\|^s - \langle J_s(x), \tilde{x}\rangle \\ &= \frac{1}{s}\|\tilde{x}\|^s + \frac{1}{s^*}\|J_s(x)\|^{s^*} - \langle J_s(x), \tilde{x}\rangle\end{aligned}$$

From convexity: $\Delta_s(\tilde{x}, x) \geq 0$ for all $\tilde{x}, x \in \mathcal{X}$, but

- no symmetry and
- no transitivity.

On the other hand: if $\mathcal{X}$ is uniformly convex (i.e. $\mathcal{X} = L^r$, $1 < r < \infty$) then

$$\Delta_s(\tilde{x}, x) \to 0 \;\Leftrightarrow\; \|\tilde{x} - x\| \to 0.$$

Convergence analysis

For proving convergence of the algorithm in Banach spaces we observe the following:

- $\mathcal{X}$ Hilbert space:

$$\frac{1}{2}\|\tilde{x}-x\|^2 = \frac{1}{2}\|\tilde{x}\|^2 - \frac{1}{2}\|x\|^2 - \langle x, \tilde{x}-x\rangle_{\mathcal{X},\mathcal{X}}$$

- $\mathcal{X}$ Banach space:

$$\Delta_s(\tilde{x},x) = \frac{1}{s}\|\tilde{x}\|^s - \frac{1}{s}\|x\|^s - \langle J_s(x), \tilde{x}-x\rangle_{\mathcal{X}^*,\mathcal{X}}$$

Therefore: replace $\|x_n^\delta - x^\dagger\|^2$ by $\Delta_s(x^\dagger, x_n^\delta)$ (or $\Delta_s(x_n^\delta, x^\dagger)$)!

Choice of the step size

Remember

$$x_{n+1}^\delta := J_{s^*}^*(x_n^\delta - \mu_n \psi_n^*).$$

Keep $\mu > 0$ variable we derive

$$\begin{aligned}
&\Delta_s(x^\dagger, J_{s^*}^*(x_n^* - \mu\psi_n^*)) - \Delta_s(x^\dagger, x_n^\delta) \\
&\quad = \frac{1}{s^*}\|x_n^* - \mu\psi_n^*\|^{s^*} - \frac{1}{s^*}\|x_n^*\|^{s^*} + \mu\langle\psi_n^*, x^\dagger\rangle \\
&\quad = \frac{1}{s^*}\|x_n^* - \mu\psi_n^*\|^{s^*} - \frac{1}{s^*}\|x_n^*\|^{s^*} + \mu\langle\psi_n^*, x_n^\delta\rangle \\
&\qquad + \mu\langle J_p(Ax_n^\delta - y^\delta), Ax^\dagger - y^\delta + y^\delta - Ax_n^\delta\rangle \\
&\quad \le \frac{1}{s^*}\|x_n^* - \mu\psi_n^*\|^{s^*} - \frac{1}{s^*}\|x_n^*\|^{s^*} + \mu\langle\psi_n^*, x_n^\delta\rangle \\
&\qquad - \mu\left(\|Ax_n^\delta - y^\delta\|^p - \delta\|Ax_n^\delta - y^\delta\|^{p-1}\right)
\end{aligned}$$

- $\delta = 0$, $\mathcal{X}, \mathcal{Y}$ are Hilbert spaces: This is the method of minimal error!

Choice of the step size

Set

$$c_n^\delta := \|A x_n^\delta - y^\delta\|^p - \delta \|A x_n^\delta - y^\delta\|^{p-1}$$

and

$$f(\mu) := \frac{1}{s^*} \|x_n^* - \mu \psi_n^*\|^{s^*} + \mu \langle \psi_n^*, x_n^\delta \rangle - \mu c_n^\delta.$$

Lemma

Assume $\psi_n^* \neq 0$. Then the minimization problem

$$f(\mu) \to \min \quad \textit{subject to} \quad \mu > 0$$

has a unique solution $\mu^* > 0$ as long as $c_n^\delta > 0$.

We have

$$f'(\mu) = -\langle \psi_n^*, J_{s^*}^*(x_n^* - \mu \psi_n^*) + x_n^\delta \rangle - c_n^\delta$$

which is continuous and strictly increasing with $f'(0) = -c_n^\delta < 0$.

Choice of the step size

We now can specify step (S2) in the general algorithm

Algorithm I

(S2) Calculate $\psi_n^* = A^\star J_p(A x_n^\delta - y^\delta)$. Find the solution μ^* of the equation

$$f'(\mu) = 0, \qquad \mu \geq 0.$$

Set $\mu_n := \min\{\mu^*, \overline{\mu}\|A x_n^\delta - y^\delta\|^{s-p}\}$.

- we observe

$$c_n^\delta > 0 \quad \Leftrightarrow \quad \|A x_n^\delta - y^\delta\| > \delta$$

- consequence: can choose $\tau > 1$ arbitrary close to 1
- The estimate ensures $\Delta_s(x^\dagger, x_{n+1}^\delta) < \Delta_s(x^\dagger, x_n^\delta)$

Choice of the step size - second approach

We need some further condition on the space $\mathcal{X}$.

Assumption (Xu/Roach inequality I)

The space $\mathcal{X}$ is supposed to be *s*-convex for some $2 \leq s < \infty$, i.e. there exists a constant C_s such that

$$\frac{1}{s}\|\tilde{x}\|^s - \frac{1}{s}\|x\|^s - \langle J_s(x), \tilde{x} - x\rangle \geq \frac{C_s}{s}\|\tilde{x} - x\|^s$$

holds for all $\tilde{x}, x \in \mathcal{X}$.

- Choice of s in the algorithm is determined by the space $\mathcal{X}$.
- s-convexity implies reflexivity and smoothness of $\mathcal{X}^*$.

Choice of the step size - second approach

What we really need is the following:

Corollary (Xu/Roach inequality II)

Assume $\mathcal{X}$ to be s-convex for some $2 \leq s < \infty$. Then the dual space $\mathcal{X}^*$ is s^*-smooth, $(s^*)^{-1} + s^{-1} = 1$, i.e. there exists a constant $G^*_{s^*}$ such that

$$\frac{1}{s^*}\|\tilde{x}^*\|^{s^*} - \frac{1}{s^*}\|x^*\|^{s^*} - \langle J^*_{s^*}(x^*), \tilde{x}^* - x^*\rangle \leq \frac{G^*_{s^*}}{s^*}\|\tilde{x}^* - x^*\|^{s^*}$$

holds for all $\tilde{x}^*, x^* \in \mathcal{X}^*$.

Example

Assume $\mathcal{X}^* = L^r$, $1 < r < \infty$. Then $\mathcal{X}^*$ is $\min\{r,2\}$-smooth and $\max\{r,2\}$-convex. Moreover we have

$$G^*_{\min\{r,2\}} := \begin{cases} 2^{2-r} & r \leq 2 \\ r-1 & r \geq 2. \end{cases}$$

Choice of the step size - second approach

We recall

$$\begin{aligned}&\Delta_s(x^\dagger, J_{s^*}^*(x_n^* - \mu\psi_n^*)) - \Delta_s(x^\dagger, x_n^\delta)\\&= \frac{1}{s^*}\|x_n^* - \mu\psi_n^*\|^{s^*} - \frac{1}{s^*}\|x_n^*\|^{s^*} + \mu\langle\psi_n^*, x^\dagger\rangle\end{aligned}$$

Using s^*-smoothness of $\mathcal{X}^*$ we derive

$$\begin{aligned}&\tfrac{1}{s^*}\|x_n^* - \mu\psi_n^*\|^{s^*} - \tfrac{1}{s^*}\|x_n^*\|^{s^*}\\&\leq \langle J_{s^*}^*(x_n^*), -\mu\psi_n^*\rangle + \frac{G_{s^*}^*}{s^*}\|\mu\psi_n^*\|^{s^*}\\&= -\mu\langle x_n^\delta, \psi_n^*\rangle + \frac{G_{s^*}^*}{s^*}\mu^{s^*}\|\psi_n^*\|^{s^*}\end{aligned}$$

and hence

$$\begin{aligned}&\Delta_s(x^\dagger, J_{s^*}^*(x_n^* - \mu\psi_n^*)) - \Delta_s(x^\dagger, x_n^\delta)\\&\leq -\mu\left(\|Ax_n^\delta - y^\delta\|^p - \delta\|Ax_n^\delta - y^\delta\|^{p-1}\right) + \frac{G_{s^*}^*}{s^*}\|\psi_n^*\|^{s^*}\mu^{s^*}\\&= -c_n^\delta\mu + \frac{G_{s^*}^*}{s^*}\|\psi_n^*\|^{s^*}\mu^{s^*}.\end{aligned}$$

Choice of the step size - second approach

We observe:

- minimizer μ^* of the right hand side can be calculated explicitly:

$$\mu^* := \left(\frac{c_n^\delta}{G_{s^*}^* \|\psi_n^*\|^{s^*}} \right)^{\frac{1}{s^*-1}} > 0 \quad \Leftrightarrow \quad c_n^\delta > 0$$

  (and hence $\psi_n^* \neq 0$).
- price: less optimality than the first approach
- but: use this variant as auxiliary problem for proving convergence of the first algorithm

Choice of the step size - second approach

We now can specify step (S2) in the general algorithm again

Algorithm II

(S2) Calculate $\psi_n^* = A^\star J_p(A x_n^\delta - y^\delta)$ and

$$\mu^* := \left(\frac{\|A x_n^\delta - y^\delta\|^p - \delta \|A x_n^\delta - y^\delta\|^{p-1}}{G_{s^*}^* \|\psi_n^*\|^{s^*}} \right)^{\frac{1}{s^*-1}} .$$

Set $\mu_n := \min\{\mu^*, \overline{\mu}\|A x_n^\delta - y^\delta\|^{s-p}\}$.

Choice of μ_n: techniqual reason by observing

$$\mu^* \geq \underline{\mu}_\tau \|A x_n^\delta - y^\delta\|^{s-p}$$

for some constant $\underline{\mu}_\tau$ which depends neither on x_n^δ nor on y^δ.

Choice of the step size - second approach

Assume $\delta = 0$ and suppose $\|A x_n^0 - y\| \to 0$ as $n \to \infty$.
Consequence:

- $\mu_n \to \infty$ for $s < p$ and
- $\mu_n \to 0$ for $s > p$ and $\overline{\mu} < \infty$.

This might be one reason why a constant step size $\mu_n \equiv \mu$ does not provide linear convergence ($s < p$) or we even cannot prove convergence ($s > p$).

We summarize the assumptions:

Assumptions

(A1) The space $\mathcal{X}$ is smooth and s-convex for some $2 \leq s < \infty$.

(A2) The space $\mathcal{Y}$ is smooth.

(A3) There exists an element $x^\dagger \in \mathcal{X}$ satisfying $A\,x^\dagger = y$.

Convergence - Algorithm II

Theorem I (noiseless data)

Assume (A1)-(A3), $\delta = 0$ and let $\{x_n^0\}$ be generated by Algorithm II. Then the algorithm stops either after a finite number N of iterations with $x_N^0 = \tilde{x}$ or we have convergence $x_n^0 \to \tilde{x}$ as $n \to \infty$. In both cases $\tilde{x}$ satisfies $A\tilde{x} = y$.

Theorem II (noisy data)

Assume (A1)-(A3), $\delta > 0$ and let $\{x_n^\delta\}$ be generated by Algorithm II. Then the algorithm stops after a finite number $N(\delta, y^\delta)$ of iterations. If, additionally, J_p is continuous then we have convergence $x^\delta_{N(\delta,y^\delta)} \to \tilde{x}$ with $A\tilde{x} = y$ as $\delta \to 0$.

Corollary

If $N(\delta, y^\delta) \to \infty$ as $\delta \to 0$ we have $x^\delta_{N(\delta,y^\delta)} \to x^\dagger$ with

$$\Delta_s(x^\dagger, x_0^\delta) := argmin\left\{\Delta_s(x, x_0^\delta) \,:\, Ax = y\right\}$$

Convergence - Algorithm I

Theorem III

Assume (A1)-(A3) and let the sequence $\{x_n^\delta\}$ be generated by Algorithm I. Then the results of Theorem I and II remain true as long as $\overline{\mu} < \infty$.

Outline

Example (Numerical example)

- $\mathcal{Y} = L^2(0,1)$, $\mathcal{X} = L^{1.1}(0,1)$ and

$$[A\,x](t) := \int_0^t x(\tau)\,d\tau, \qquad \tau \in [0,1]$$

- two different types of exact solutions:

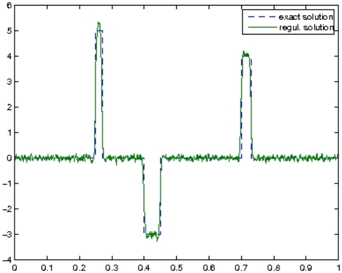
$$x_1^\dagger(t) := 3(t-0.5)^2 + 0.2,$$

and

$$x_2^\dagger(t) := \begin{cases} 5, & t \in [0.25, 0.27], \\ -3, & t \in [0.4, 0.45], \\ 4, & t \in [0.7, 0.73], \\ 0, & \text{else.} \end{cases}$$

- $x_0 \equiv 0$, $s = p = 2$, $\tau := 1.2$
- discretization: $k = 1000$ DoF

Example



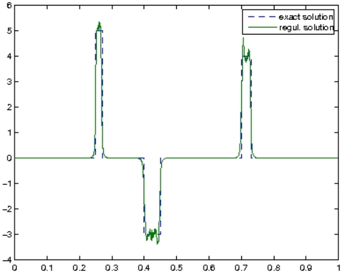


Figure: Regularized solutions for $x_2^\dagger$ with $\mathcal{X} = L^2$ (left plot) and $\mathcal{X} = L^{1.1}$ (right plot), $\delta_{rel} = 0.01$

Example

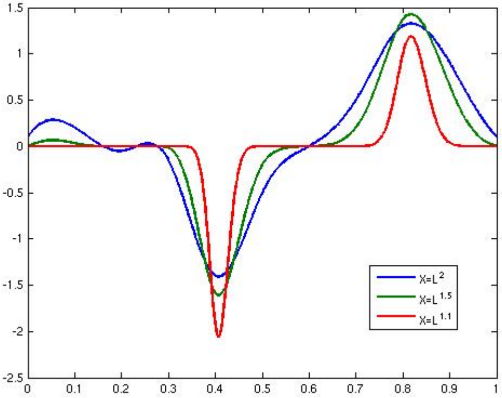


Figure: Application of duality mappings on x^* (blue) for $\mathcal{X}^* = L^3 \;\Leftrightarrow\; \mathcal{X} = L^{1.5}$ (green) and $\mathcal{X}^* = L^{10} \;\Leftrightarrow\; \mathcal{X} = L^{1.1}$ (red)

Example

	$\mu_n = const.$		Algorithm II		Algorithm I	
δ_{rel}	$N(\delta, y^\delta)$	time	$N(\delta, y^\delta)$	time	$N(\delta, y^\delta)$	time
0.05	863	0.85	63	0.16	28	0.18
0.01	7530	6.56	335	0.40	93	0.33
10^{-3}	79120	69.01	2065	2.29	451	2.05
10^{-4}	$> 10^6$	–	24548	26.27	2068	8.69
10^{-5}	–	–	118823	126.81	12479	49.97

Calculation times for sample function $x_1^\dagger$

- The reconstruction error is similar to Tikhonov regularization:

$$x_\alpha^\delta := \operatorname{argmin} \left\{ \frac{1}{2} \|Ax - y^\delta\|_{L^2}^2 + \frac{\alpha}{2} \|x\|_{L^{1.1}}^2 \ : \ x \in \mathcal{X} \right\}$$

- Matrix-vector multiplications in this example with $\mathcal{O}(k)$ operations

Example

	$\mu_n = const.$		Algorithm II		Algorithm I	
δ_{rel}	$N(\delta, y^\delta)$	time	$N(\delta, y^\delta)$	time	$N(\delta, y^\delta)$	time
0.05	4023	3.52	253	0.35	104	0.29
0.01	36720	31.98	1520	1.35	358	1.28
10^{-3}	457270	391.40	11022	12.01	963	4.17
10^{-4}	$> 10^6$	–	94315	101.77	6729	27.10
10^{-5}	–	–	606582	653.37	50890	205.01

Calculation times for sample function $x_2^\dagger$

Outline

Summary

- We presented an accelerated Landweber-type method for the regularization of linear ill-posed problems
- choice of the step size by solving a (simple) one-dimensional minimization problem
- number of necessary iterations as well as calculation time can be reduced significantly
- generalization to nonlinear equations possible using the η-condition as restriction of the nonlinearity
- open problem: convergence rates

Nonlinear Equations

- suppose $F : \mathcal{D}(F) \subseteq \mathcal{X} \longrightarrow \mathcal{Y}$ satisfies

$$\|F(\tilde{x}) - F(x) - F'(x)(\tilde{x} - x)\| \leq L\,\|F(\tilde{x}) - F(x)\|$$

  for some $L < 1$. Then we derive

$$\begin{aligned} &\Delta_s(x^\dagger, J^*_{s^*}(x^*_n - \mu\psi^*_n)) - \Delta_s(x^\dagger, x^\delta_n) \\ &\quad\leq \frac{1}{s^*}\|x^*_n - \mu\psi^*_n\|^{s^*} - \frac{1}{s^*}\|x^*_n\|^{s^*} + \mu\langle\psi^*_n, x^\delta_n\rangle \\ &\qquad -\mu\left((1-L)\|F(x^\delta_n) - y^\delta\|^p - (1+L)\delta\|F(x^\delta_n) - y^\delta\|^{p-1}\right) \end{aligned}$$

- hence: choose $\tau > \frac{1+L}{1-L}$
- further assumptions: $\|F'(x)\| \leq K$ uniformly and $F'(x)$ depends continuously on x