Two-Point Gradient Methods

for solving

Nonlinear Ill-Posed Problems

Simon Hubmer

Johannes Kepler University, Linz

4.4.2017, DTU Lyngby, HD-Tomo Seminar

Joint work with:

R. Ramlau

Doctoral Program
Computational Mathematics
Numerical Analysis and Symbolic Computation

JOHANNES KEPLER
UNIVERSITY LINZ

The Problem

The Problem

- Hilbert spaces $\mathcal{X}$ and $\mathcal{Y}$, with norms $\|.\|$.

The Problem

- Hilbert spaces $\mathcal{X}$ and $\mathcal{Y}$, with norms $\|.\|$.
- Operator $F : \mathcal{X} \to \mathcal{Y}$, continuously Fréchet differentiable.

The Problem

- Hilbert spaces $\mathcal{X}$ and $\mathcal{Y}$, with norms $\|.\|$.
- Operator $F : \mathcal{X} \to \mathcal{Y}$, continuously Fréchet differentiable.
- Noisy data $y^\delta \in \mathcal{Y}$ and noise level $\delta \in \mathbb{R}^+$.

The Problem

- Hilbert spaces $\mathcal{X}$ and $\mathcal{Y}$, with norms $\|.\|$.
- Operator $F : \mathcal{X} \to \mathcal{Y}$, continuously Fréchet differentiable.
- Noisy data $y^\delta \in \mathcal{Y}$ and noise level $\delta \in \mathbb{R}^+$.

Problem

$$F(x) = y^{(\delta)}$$

The Problem

- Hilbert spaces $\mathcal{X}$ and $\mathcal{Y}$, with norms $\|.\|$.
- Operator $F : \mathcal{X} \to \mathcal{Y}$, continuously Fréchet differentiable.
- Noisy data $y^\delta \in \mathcal{Y}$ and noise level $\delta \in \mathbb{R}^+$.

Problem

$$F(x) = y^{(\delta)}$$

The noisy data y^δ satisfies

$$\left\| y - y^\delta \right\| \leq \delta\,.$$

Examples

Examples

Nonlinear Hammerstein operator:

$$F : H^1[0,1] \to L^2[0,1]\,, \quad F(x)(s) := \int_0^1 k(s,t)\Phi(x(t),t)\,dt\,.$$

Examples

Nonlinear Hammerstein operator:

$$F : H^1[0,1] \to L^2[0,1]\,, \quad F(x)(s) := \int_0^1 k(s,t)\Phi(x(t),t)\,dt\,.$$

Attenuated Radon transform (SPECT):

$$A(f,\mu)(s,\omega) := \int_{\mathbb{R}} f(s\omega^\perp + t\omega)\exp\left(-\int_t^\infty \mu(s\omega^\perp + r\omega)\,dr\right)\,dt\,.$$

Examples

Nonlinear Hammerstein operator:

$$F : H^1[0,1] \to L^2[0,1] \,, \quad F(x)(s) := \int_0^1 k(s,t)\Phi(x(t),t)\, dt \,.$$

Attenuated Radon transform (SPECT):

$$A(f,\mu)(s,\omega) := \int_{\mathbb{R}} f(s\omega^\perp + t\omega) \exp\left(- \int_t^\infty \mu(s\omega^\perp + r\omega)\, dr \right) \, dt \,.$$

Further examples: CT, MRI, MRAI, PI, EIT, AO, . . .

Tikhonov Regularization

Tikhonov Regularization

Required:

- Initial guess x_0 and regularization parameter α.

Tikhonov Regularization

Required:

- Initial guess x_0 and regularization parameter α.

The method:

$$\boxed{\min_x \left\{ \frac{1}{2} \left\| F(x) - y^\delta \right\|^2 + \frac{\alpha}{2} \left\| x - x_0 \right\|^2 \right\} .}$$

Tikhonov Regularization

Required:

- Initial guess x_0 and regularization parameter α.

The method:

$$\min_x \left\{ \frac{1}{2} \left\| F(x) - y^\delta \right\|^2 + \frac{\alpha}{2} \left\| x - x_0 \right\|^2 \right\} .$$

Properties:

Tikhonov Regularization

Required:

- Initial guess x_0 and regularization parameter α.

The method:

$$\boxed{\min_x \left\{ \frac{1}{2} \left\| F(x) - y^\delta \right\|^2 + \frac{\alpha}{2} \left\| x - x_0 \right\|^2 \right\} .}$$

Properties:

+ Weak conditions necessary for analysis.

Tikhonov Regularization

Required:

- Initial guess x_0 and regularization parameter α.

The method:

$$\boxed{\min_x \left\{ \frac{1}{2} \left\| F(x) - y^\delta \right\|^2 + \frac{\alpha}{2} \left\| x - x_0 \right\|^2 \right\} .}$$

Properties:

+ Weak conditions necessary for analysis.
+ Very versatile (different norms, regularization functionals).

Tikhonov Regularization

Required:

- Initial guess x_0 and regularization parameter α.

The method:

$$\boxed{\min_x \left\{ \frac{1}{2} \left\| F(x) - y^\delta \right\|^2 + \frac{\alpha}{2} \left\| x - x_0 \right\|^2 \right\} .}$$

Properties:

+ Weak conditions necessary for analysis.
+ Very versatile (different norms, regularization functionals).
− Computation of the minimum $\leftrightarrow$ HOW??

Landweber Iteration

Landweber Iteration

Required:

- Initial guess x_0 and stopping criterion.

Landweber Iteration

Required:

- Initial guess x_0 and stopping criterion.

The method:

$$\boxed{x_{k+1}^{\delta} = x_k^{\delta} + F'(x_k^{\delta})^*(y^{\delta} - F(x_k^{\delta}))\,.}$$

Landweber Iteration

Required:

- Initial guess x_0 and stopping criterion.

The method:

$$\boxed{x_{k+1}^{\delta} = x_k^{\delta} + F'(x_k^{\delta})^*(y^{\delta} - F(x_k^{\delta}))\,.}$$

Properties:

Landweber Iteration

Required:

- Initial guess x_0 and stopping criterion.

The method:

$$\boxed{x_{k+1}^{\delta} = x_k^{\delta} + F'(x_k^{\delta})^*(y^{\delta} - F(x_k^{\delta}))\,.}$$

Properties:

+ Easy to implement.

Landweber Iteration

Required:

- Initial guess x_0 and stopping criterion.

The method:

$$\boxed{x_{k+1}^\delta = x_k^\delta + F'(x_k^\delta)^*(y^\delta - F(x_k^\delta))\,.}$$

Properties:

+ Easy to implement.
− Strong conditions necessary for analysis.

Landweber Iteration

Required:

- Initial guess x_0 and stopping criterion.

The method:

$$\boxed{x_{k+1}^{\delta} = x_k^{\delta} + F'(x_k^{\delta})^*(y^{\delta} - F(x_k^{\delta}))\,.}$$

Properties:

+ Easy to implement.
– Strong conditions necessary for analysis.
– Slow convergence, i.e., lots of iterations required.

Second Order Methods

Second Order Methods

Levenberg-Marquardt method

$$x_{k+1}^{\delta} = x_k^{\delta} + (F'(x_k^{\delta})^* F'(x_k^{\delta}) + \alpha_k I)^{-1} F'(x_k^{\delta})^* (y^{\delta} - F(x_k^{\delta})) \,.$$

Second Order Methods

Levenberg-Marquardt method

$$x_{k+1}^\delta = x_k^\delta + (F'(x_k^\delta)^* F'(x_k^\delta) + \alpha_k I)^{-1} F'(x_k^\delta)^* (y^\delta - F(x_k^\delta)) \,.$$

Iteratively regularized Gauss-Newton method

$$x_{k+1}^\delta = x_k^\delta + (F'(x_k^\delta)^* F'(x_k^\delta) + \alpha_k I)^{-1} (F'(x_k^\delta)^* (y^\delta - F(x_k^\delta)) + \alpha_k (x_0 - x_k^\delta)) \,.$$

Second Order Methods

Levenberg-Marquardt method

$$x_{k+1}^\delta = x_k^\delta + (F'(x_k^\delta)^* F'(x_k^\delta) + \alpha_k I)^{-1} F'(x_k^\delta)^* (y^\delta - F(x_k^\delta)) \,.$$

Iteratively regularized Gauss-Newton method

$$x_{k+1}^\delta = x_k^\delta + (F'(x_k^\delta)^* F'(x_k^\delta) + \alpha_k I)^{-1} (F'(x_k^\delta)^* (y^\delta - F(x_k^\delta)) + \alpha_k (x_0 - x_k^\delta)) \,.$$

Properties:

Second Order Methods

Levenberg-Marquardt method

$$x_{k+1}^\delta = x_k^\delta + (F'(x_k^\delta)^* F'(x_k^\delta) + \alpha_k I)^{-1} F'(x_k^\delta)^* (y^\delta - F(x_k^\delta)) \,.$$

Iteratively regularized Gauss-Newton method

$$x_{k+1}^\delta = x_k^\delta + (F'(x_k^\delta)^* F'(x_k^\delta) + \alpha_k I)^{-1} (F'(x_k^\delta)^* (y^\delta - F(x_k^\delta)) + \alpha_k (x_0 - x_k^\delta)) \,.$$

Properties:

+ Require much less iterations.

Second Order Methods

Levenberg-Marquardt method

$$x_{k+1}^\delta = x_k^\delta + (F'(x_k^\delta)^* F'(x_k^\delta) + \alpha_k I)^{-1} F'(x_k^\delta)^* (y^\delta - F(x_k^\delta)) \,.$$

Iteratively regularized Gauss-Newton method

$$x_{k+1}^\delta = x_k^\delta + (F'(x_k^\delta)^* F'(x_k^\delta) + \alpha_k I)^{-1} (F'(x_k^\delta)^* (y^\delta - F(x_k^\delta)) + \alpha_k (x_0 - x_k^\delta)) \,.$$

Properties:

+ Require much less iterations.
− Very strong conditions necessary for analysis.

Second Order Methods

Levenberg-Marquardt method

$$x_{k+1}^\delta = x_k^\delta + (F'(x_k^\delta)^* F'(x_k^\delta) + \alpha_k I)^{-1} F'(x_k^\delta)^* (y^\delta - F(x_k^\delta)) \, .$$

Iteratively regularized Gauss-Newton method

$$x_{k+1}^\delta = x_k^\delta + (F'(x_k^\delta)^* F'(x_k^\delta) + \alpha_k I)^{-1} (F'(x_k^\delta)^* (y^\delta - F(x_k^\delta)) + \alpha_k (x_0 - x_k^\delta)) \, .$$

Properties:

+ Require much less iterations.
– Very strong conditions necessary for analysis.
– Require inversion of $(F'(x)^* F'(x) + \alpha I)$ in every iteration step $\rightarrow$ difficult and takes time.

Acceleration Techniques

Acceleration Techniques

- Landweber Iteration with operator approximation:

$$x_{k+1}^\delta = x_k^\delta + \tilde{F}'(x_k^\delta)^*(y^\delta - \tilde{F}(x_k^\delta))\,.$$

Acceleration Techniques

- Landweber Iteration with operator approximation:

$$x_{k+1}^\delta = x_k^\delta + \tilde{F}'(x_k^\delta)^*(y^\delta - \tilde{F}(x_k^\delta)) \,.$$

- Landweber Iteration in Hilbert Scales:

$$x_{k+1}^\delta = x_k^\delta + L^{-2s} F'(x_k^\delta)^*(y^\delta - F(x_k^\delta)) \,.$$

Acceleration Techniques

- Landweber Iteration with operator approximation:

$$x_{k+1}^{\delta} = x_k^{\delta} + \tilde{F}'(x_k^{\delta})^*(y^{\delta} - \tilde{F}(x_k^{\delta}))\,.$$

- Landweber Iteration in Hilbert Scales:

$$x_{k+1}^{\delta} = x_k^{\delta} + L^{-2s} F'(x_k^{\delta})^*(y^{\delta} - F(x_k^{\delta}))\,.$$

- Landweber Iteration with intelligent stepsizes:

$$x_{k+1}^{\delta} = x_k^{\delta} + \alpha_k^{\delta} F'(x_k^{\delta})^*(y^{\delta} - F(x_k^{\delta}))\,.$$

Examples: Steepest Descent, Barzilai-Borwein, Neubauer.

Connection: Residual Functional

$$\Phi(x) = \frac{1}{2}\left\| F(x) - y^{\delta} \right\|^2$$

Connection: Residual Functional

$$\Phi(x) = \frac{1}{2} \left\| F(x) - y^\delta \right\|^2$$

- Tikhonov = Minimize$\{\ \Phi(x) + \text{Regularization}(x)\ \}$.

Connection: Residual Functional

$$\Phi(x) = \frac{1}{2} \left\| F(x) - y^\delta \right\|^2$$

- Tikhonov = Minimize$\{\ \Phi(x) + \text{Regularization}(x)\ \}$.
- Landweber = Gradient Descent for $\Phi(x)$.

Connection: Residual Functional

$$\Phi(x) = \frac{1}{2} \left\| F(x) - y^\delta \right\|^2$$

- Tikhonov = Minimize{ $\Phi(x)$ + Regularization(x) }.
- Landweber = Gradient Descent for $\Phi(x)$.
- Levenberg Marquardt = 2nd order descent for $\Phi(x)$.

Connection: Residual Functional

$$\Phi(x) = \frac{1}{2} \left\| F(x) - y^\delta \right\|^2$$

- Tikhonov = Minimize$\{ \Phi(x) + \text{Regularization}(x) \}$.
- Landweber = Gradient Descent for $\Phi(x)$.
- Levenberg Marquardt = 2nd order descent for $\Phi(x)$.
- Iteratively regularized Gauss-Newton
  = 2nd order descent for $\Phi(x)$ + Tikhonov Type Stabilization.

Nesterov Acceleration

Nesterov Acceleration

General minimization problem

$$\min_x \{\Phi(x)\} \ .$$

Nesterov Acceleration

General minimization problem

$$\min_x \{\Phi(x)\} \; .$$

Yurii Nesterov: Instead of using gradient descent:

$$x_{k+1} = x_k - \omega \nabla \Phi(x_k) \, ,$$

Nesterov Acceleration

General minimization problem

$$\min_x \{\Phi(x)\} .$$

Yurii Nesterov: Instead of using gradient descent:

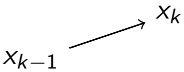
$$x_{k+1} = x_k - \omega \nabla \Phi(x_k) ,$$

use the following iteration:

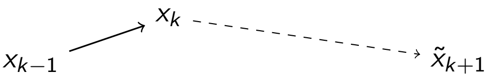
$$z_k = x_k + \tfrac{k-1}{k+\alpha-1}(x_k - x_{k-1}) ,$$
$$x_{k+1} = z_k - \omega \nabla \Phi(z_k) .$$

Motivating Picture

Motivating Picture

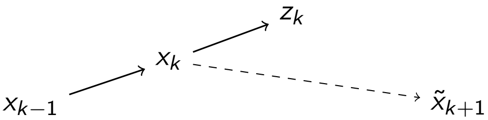


Motivating Picture



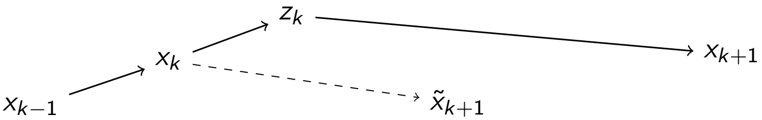
$$\tilde{x}_{k+1} = x_k - \omega \nabla \Phi(x_k)$$

Motivating Picture



$$z_k = x_k + \frac{k-1}{k+\alpha-1}(x_k - x_{k-1})$$

Motivating Picture



$$z_k = x_k + \frac{k-1}{k+\alpha-1}(x_k - x_{k-1})$$

$$x_{k+1} = z_k - \omega \nabla \Phi(z_k)$$

What's so good about that?

What's so good about that?

- **Assume:** Φ is convex.

What's so good about that?

- **Assume:** Φ is convex.
- **Gradient Descent:**

$$\left\| \Phi(x_k) - \Phi(x^\dagger) \right\| = \mathcal{O}(k^{-1})$$

What's so good about that?

- **Assume:** Φ is convex.
- **Gradient Descent:**

$$\left\| \Phi(x_k) - \Phi(x^\dagger) \right\| = \mathcal{O}(k^{-1})$$

- **Nesterov Acceleration:**

$$\left\| \Phi(x_k) - \Phi(x^\dagger) \right\| = \mathcal{O}(k^{-2})$$

What's so good about that?

- **Assume:** Φ is convex.
- **Gradient Descent:**

$$\left\| \Phi(x_k) - \Phi(x^\dagger) \right\| = \mathcal{O}(k^{-1})$$

- **Nesterov Acceleration:**

$$\left\| \Phi(x_k) - \Phi(x^\dagger) \right\| = \mathcal{o}(k^{-2})$$

H. Attouch, J. Peypouquet, The rate of convergence of Nesterov's accelerated forward-backward method is actually $o(k^{-2})$, SIAM Journal on Optimization

Y. Nesterov, A method of solving a convex programming problem with convergence rate $O(1/k^2)$, Soviet Mathematics Doklady, 27, 2, 1983

Application to Nonlinear Ill-Posed Problems

Application to Nonlinear Ill-Posed Problems

For our problem, the method reads as

$$z_k^\delta = x_k^\delta + \tfrac{k-1}{k+\alpha-1}(x_k^\delta - x_{k-1}^\delta)\,,$$
$$x_{k+1}^\delta = z_k^\delta + \alpha_k^\delta\, F'(z_k^\delta)^*(y^\delta - F(z_k^\delta))\,.$$

Application to Nonlinear Ill-Posed Problems

For our problem, the method reads as

$$z_k^\delta = x_k^\delta + \tfrac{k-1}{k+\alpha-1}(x_k^\delta - x_{k-1}^\delta)\,,$$
$$x_{k+1}^\delta = z_k^\delta + \alpha_k^\delta\, F'(z_k^\delta)^*(y^\delta - F(z_k^\delta))\,.$$

There is a generalization to deal with

$$\min\{\Phi(x) + \Psi(x)\}\,,$$

Application to Nonlinear Ill-Posed Problems

For our problem, the method reads as

$$z_k^\delta = x_k^\delta + \tfrac{k-1}{k+\alpha-1}(x_k^\delta - x_{k-1}^\delta)\,,$$
$$x_{k+1}^\delta = z_k^\delta + \alpha_k^\delta\, F'(z_k^\delta)^*(y^\delta - F(z_k^\delta))\,.$$

There is a generalization to deal with

$$\min\{\Phi(x) + \Psi(x)\}\,,$$

which reads as

$$z_k = x_k + \tfrac{k-1}{k+\alpha-1}(x_k - x_{k-1})\,,$$
$$x_{k+1} = \operatorname{prox}_\Psi\left(z_k - \omega\nabla\Phi(z_k)\right)\,.$$

Application to Nonlinear Ill-Posed Problems

For our problem, the method reads as

$$z_k^\delta = x_k^\delta + \tfrac{k-1}{k+\alpha-1}(x_k^\delta - x_{k-1}^\delta)\,,$$
$$x_{k+1}^\delta = z_k^\delta + \alpha_k^\delta F'(z_k^\delta)^*(y^\delta - F(z_k^\delta))\,.$$

There is a generalization to deal with

$$\min\{\Phi(x) + \Psi(x)\}\,,$$

which reads as

$$z_k = x_k + \tfrac{k-1}{k+\alpha-1}(x_k - x_{k-1})\,,$$
$$x_{k+1} = \mathrm{prox}_\Psi\left(z_k - \omega\nabla\Phi(z_k)\right)\,.$$

$\Longrightarrow$ Sparsity Constraints, Projections, etc.

Neubauer strikes again

Neubauer strikes again

- **Assumptions:** Linear operator $F(x) = Tx$, source condition $x^\dagger \in \mathcal{R}((T^*T)^\mu)$, a priori stopping rule.

Neubauer strikes again

- **Assumptions:** Linear operator $F(x) = Tx$, source condition $x^\dagger \in \mathcal{R}((T^*T)^\mu)$, a priori stopping rule.

- If $0 \leq \mu \leq \frac{1}{2}$, then

$$k(\delta) = \mathcal{O}(\delta^{-\frac{1}{2\mu+1}}), \qquad \left\| x^\delta_{k(\delta)} - x^\dagger \right\| = o(\delta^{\frac{2\mu}{2\mu+1}}).$$

Neubauer strikes again

- **Assumptions:** Linear operator $F(x) = Tx$, source condition $x^\dagger \in \mathcal{R}((T^*T)^\mu)$, a priori stopping rule.
- If $0 \leq \mu \leq \frac{1}{2}$, then
$$k(\delta) = \mathcal{O}(\delta^{-\frac{1}{2\mu+1}}) , \qquad \left\| x_{k(\delta)}^\delta - x^\dagger \right\| = o(\delta^{\frac{2\mu}{2\mu+1}}) .$$
- If $\mu > \frac{1}{2}$, then
$$k(\delta) = \mathcal{O}(\delta^{-\frac{2}{2\mu+3}}) , \qquad \left\| x_{k(\delta)}^\delta - x^\dagger \right\| = o(\delta^{\frac{2\mu+1}{2\mu+3}}) .$$

Neubauer strikes again

- **Assumptions:** Linear operator $F(x) = Tx$, source condition $x^\dagger \in \mathcal{R}((T^*T)^\mu)$, a priori stopping rule.

- If $0 \leq \mu \leq \frac{1}{2}$, then

$$k(\delta) = \mathcal{O}(\delta^{-\frac{1}{2\mu+1}})\,, \qquad \left\| x_{k(\delta)}^\delta - x^\dagger \right\| = o(\delta^{\frac{2\mu}{2\mu+1}})\,.$$

- If $\mu > \frac{1}{2}$, then

$$k(\delta) = \mathcal{O}(\delta^{-\frac{2}{2\mu+3}})\,, \qquad \left\| x_{k(\delta)}^\delta - x^\dagger \right\| = o(\delta^{\frac{2\mu+1}{2\mu+3}})\,.$$

- Similar results also when using the discrepancy principle.

Neubauer strikes again

- **Assumptions:** Linear operator $F(x) = Tx$, source condition $x^\dagger \in \mathcal{R}((T^*T)^\mu)$, a priori stopping rule.
- If $0 \leq \mu \leq \frac{1}{2}$, then

$$k(\delta) = \mathcal{O}(\delta^{-\frac{1}{2\mu+1}}) \,, \qquad \left\| x_{k(\delta)}^\delta - x^\dagger \right\| = o(\delta^{\frac{2\mu}{2\mu+1}}) \,.$$

- If $\mu > \frac{1}{2}$, then

$$k(\delta) = \mathcal{O}(\delta^{-\frac{2}{2\mu+3}}) \,, \qquad \left\| x_{k(\delta)}^\delta - x^\dagger \right\| = o(\delta^{\frac{2\mu+1}{2\mu+3}}) \,.$$

- Similar results also when using the discrepancy principle.

A. Neubauer, On Nesterov acceleration for Landweber iteration of linear ill-posed problems, J. Inv. Ill-Posed Problems, to appear.

Two-Point Gradient (TPG) Methods

Two-Point Gradient (TPG) Methods

How about general methods of the form

$$\begin{aligned} z_k^\delta &= x_k^\delta + \lambda_k^\delta (x_k^\delta - x_{k-1}^\delta)\,, \\ x_{k+1}^\delta &= z_k^\delta + \alpha_k^\delta F'(z_k^\delta)^* (y^\delta - F(z_k^\delta))\,. \end{aligned}$$

Two-Point Gradient (TPG) Methods

How about general methods of the form

$$z_k^\delta = x_k^\delta + \lambda_k^\delta (x_k^\delta - x_{k-1}^\delta)\,,$$
$$x_{k+1}^\delta = z_k^\delta + \alpha_k^\delta F'(z_k^\delta)^* (y^\delta - F(z_k^\delta))\,.$$

Question: Do they converge under standard assumptions?

Two-Point Gradient (TPG) Methods

How about general methods of the form

$$z_k^\delta = x_k^\delta + \lambda_k^\delta (x_k^\delta - x_{k-1}^\delta)\,,$$
$$x_{k+1}^\delta = z_k^\delta + \alpha_k^\delta F'(z_k^\delta)^*(y^\delta - F(z_k^\delta))\,.$$

Question: Do they converge under standard assumptions?

- Yes for linear problems and $\lambda_k^\delta = \frac{k-1}{k+\alpha-1}$ $\leftarrow$ Neubauer

Two-Point Gradient (TPG) Methods

How about general methods of the form

$$z_k^\delta = x_k^\delta + \lambda_k^\delta (x_k^\delta - x_{k-1}^\delta)\,,$$
$$x_{k+1}^\delta = z_k^\delta + \alpha_k^\delta F'(z_k^\delta)^*(y^\delta - F(z_k^\delta))\,.$$

Question: Do they converge under standard assumptions?

- Yes for linear problems and $\lambda_k^\delta = \frac{k-1}{k+\alpha-1}$ $\leftarrow$ Neubauer
- Yes for $\lambda_k^\delta \to 0$ fast enough.

Two-Point Gradient (TPG) Methods

How about general methods of the form

$$z_k^\delta = x_k^\delta + \lambda_k^\delta (x_k^\delta - x_{k-1}^\delta) \,,$$
$$x_{k+1}^\delta = z_k^\delta + \alpha_k^\delta F'(z_k^\delta)^* (y^\delta - F(z_k^\delta)) \,.$$

Question: Do they converge under standard assumptions?

- Yes for linear problems and $\lambda_k^\delta = \frac{k-1}{k+\alpha-1}$ $\leftarrow$ Neubauer
- Yes for $\lambda_k^\delta \to 0$ fast enough.
- Yes for some explicit choices of λ_k^δ.

Two-Point Gradient (TPG) Methods

How about general methods of the form

$$z_k^\delta = x_k^\delta + \lambda_k^\delta (x_k^\delta - x_{k-1}^\delta)\,,$$
$$x_{k+1}^\delta = z_k^\delta + \alpha_k^\delta F'(z_k^\delta)^* (y^\delta - F(z_k^\delta))\,.$$

Question: Do they converge under standard assumptions?

- Yes for linear problems and $\lambda_k^\delta = \frac{k-1}{k+\alpha-1}$ $\leftarrow$ Neubauer
- Yes for $\lambda_k^\delta \to 0$ fast enough.
- Yes for some explicit choices of λ_k^δ.
- Yes for λ_k^δ defined via a backtracking search.

Two-Point Gradient (TPG) Methods

How about general methods of the form

$$z_k^\delta = x_k^\delta + \lambda_k^\delta (x_k^\delta - x_{k-1}^\delta)\,,$$
$$x_{k+1}^\delta = z_k^\delta + \alpha_k^\delta F'(z_k^\delta)^*(y^\delta - F(z_k^\delta))\,.$$

Question: Do they converge under standard assumptions?

- Yes for linear problems and $\lambda_k^\delta = \frac{k-1}{k+\alpha-1}$ $\leftarrow$ Neubauer
- Yes for $\lambda_k^\delta \to 0$ fast enough.
- Yes for some explicit choices of λ_k^δ.
- Yes for λ_k^δ defined via a backtracking search.

Open: Convergence for nonlinear problems and $\lambda_k^\delta = \frac{k-1}{k+\alpha-1}$

Convergence Conditions

Convergence Conditions

- Nonlinearity Condition

$$\left\| F(x) - F(\tilde{x}) - F'(x)(x - \tilde{x}) \right\| \leq \eta \left\| F(x) - F(\tilde{x}) \right\| ,$$

$$x, \tilde{x} \in \mathcal{B}_{4\rho}(x_0) \subset \mathcal{D}(F) , \qquad \eta < \frac{1}{2} .$$

Convergence Conditions

- Nonlinearity Condition

$$\left\| F(x) - F(\tilde{x}) - F'(x)(x - \tilde{x}) \right\| \leq \eta \left\| F(x) - F(\tilde{x}) \right\| , \quad x, \tilde{x} \in \mathcal{B}_{4\rho}(x_0) \subset \mathcal{D}(F) , \qquad \eta < \frac{1}{2} .$$

- Parameters $0 \leq \lambda_k^\delta \leq 1$ and stepsizes $\alpha_k^\delta \geq \alpha_{\min} > 0$

Convergence Conditions

- Nonlinearity Condition

$$\left\| F(x) - F(\tilde{x}) - F'(x)(x - \tilde{x}) \right\| \leq \eta \left\| F(x) - F(\tilde{x}) \right\| \,,$$

$$x, \tilde{x} \in \mathcal{B}_{4\rho}(x_0) \subset \mathcal{D}(F)\,, \qquad \eta < \frac{1}{2}\,.$$

- Parameters $0 \leq \lambda_k^\delta \leq 1$ and stepsizes $\alpha_k^\delta \geq \alpha_{\min} > 0$ satisfy

$$\lambda_k^\delta(\lambda_k^\delta + 1) \left\| x_k^\delta - x_{k+1}^\delta \right\|^2 - \left(1 + \frac{\Psi}{\mu}\right) \alpha_k^\delta \left\| F(z_k^\delta) - y^\delta \right\|^2$$

$$+ (\alpha_k^\delta)^2 \left\| F'(z_k^\delta)^* (F(z_k^\delta) - y^\delta) \right\|^2 \leq 0\,.$$

Convergence Conditions

- Nonlinearity Condition

$$\left\| F(x) - F(\tilde{x}) - F'(x)(x - \tilde{x}) \right\| \leq \eta \left\| F(x) - F(\tilde{x}) \right\| \,, \qquad x, \tilde{x} \in \mathcal{B}_{4\rho}(x_0) \subset \mathcal{D}(F) \,, \qquad \eta < \frac{1}{2} \,.$$

- Parameters $0 \leq \lambda_k^\delta \leq 1$ and stepsizes $\alpha_k^\delta \geq \alpha_{\min} > 0$ satisfy

$$\lambda_k^\delta(\lambda_k^\delta + 1) \left\| x_k^\delta - x_{k+1}^\delta \right\|^2 - \left(1 + \frac{\Psi}{\mu} \right) \alpha_k^\delta \left\| F(z_k^\delta) - y^\delta \right\|^2 + (\alpha_k^\delta)^2 \left\| F'(z_k^\delta)^*(F(z_k^\delta) - y^\delta) \right\|^2 \leq 0 \,.$$

- Parameters λ_k^δ satisfy

$$\sum_{k=0}^{\infty} \lambda_k^0 \left\| x_k^0 - x_{k-1}^0 \right\| < \infty \,.$$

Some Possible Choices

Some Possible Choices

For the stepsizes α_k^δ, one can use

Some Possible Choices

For the stepsizes α_k^δ, one can use

- a constant stepsize $\alpha_k^\delta = \omega$,

Some Possible Choices

For the stepsizes α_k^δ, one can use

- a constant stepsize $\alpha_k^\delta = \omega$,
- the steepest descent stepsize or the minimal error stepsize.

Some Possible Choices

For the stepsizes α_k^δ, one can use

- a constant stepsize $\alpha_k^\delta = \omega$,
- the steepest descent stepsize or the minimal error stepsize.

The parameters λ_k^δ can be chosen

Some Possible Choices

For the stepsizes α_k^δ, one can use

- a constant stepsize $\alpha_k^\delta = \omega$,
- the steepest descent stepsize or the minimal error stepsize.

The parameters λ_k^δ can be chosen

- as any sequence decaying sufficiently fast,

Some Possible Choices

For the stepsizes α_k^δ, one can use

- a constant stepsize $\alpha_k^\delta = \omega$,
- the steepest descent stepsize or the minimal error stepsize.

The parameters λ_k^δ can be chosen

- as any sequence decaying sufficiently fast,
- explicitly via

$$\lambda_k^\delta = \min\left\{ -\frac{1}{2} + \sqrt{\frac{1}{4} + \frac{\Psi(\tau\delta)^2}{\mu\bar{\omega}^2 \left\| x_k^\delta - x_{k-1}^\delta \right\|^2}}, 1 \right\},$$

Some Possible Choices

For the stepsizes α_k^δ, one can use

- a constant stepsize $\alpha_k^\delta = \omega$,
- the steepest descent stepsize or the minimal error stepsize.

The parameters λ_k^δ can be chosen

- as any sequence decaying sufficiently fast,
- explicitly via

$$\lambda_k^\delta = \min\left\{ -\frac{1}{2} + \sqrt{\frac{1}{4} + \frac{\Psi(\tau\delta)^2}{\mu\bar{\omega}^2 \left\| x_k^\delta - x_{k-1}^\delta \right\|^2}}, 1 \right\},$$

- via a backtracking algorithm.

Main Result

Main Result

Discrepancy Principle:

$$\left\| y^\delta - F(z^\delta_{k_*}) \right\| \leq \tau\delta < \left\| y^\delta - F(z^\delta_k) \right\| , \qquad 0 \leq k < k_* = k_*(\delta, y^\delta) .$$

Main Result

Discrepancy Principle:

$$\left\| y^\delta - F(z_{k_*}^\delta) \right\| \leq \tau\delta < \left\| y^\delta - F(z_k^\delta) \right\| , \qquad 0 \leq k < k_* = k_*(\delta, y^\delta) .$$

Theorem

Under the above assumptions, there holds

$$\lim_{\delta \to 0} z_{k_*(\delta, y^\delta)}^\delta = x_* .$$

Main Result

Discrepancy Principle:

$$\left\| y^\delta - F(z^\delta_{k_*}) \right\| \leq \tau\delta < \left\| y^\delta - F(z^\delta_k) \right\| \,, \qquad 0 \leq k < k_* = k_*(\delta, y^\delta)\,.$$

Theorem

Under the above assumptions, there holds

$$\lim_{\delta\to 0} z^\delta_{k_*(\delta,y^\delta)} = x_*\,.$$

If additionally $\mathcal{N}(F'(x^\dagger)) \subset \mathcal{N}(F'(x))$*, then we have*

$$\lim_{\delta\to 0} z^\delta_{k_*(\delta,y^\delta)} = x^\dagger\,.$$

Proof Steps

Proof Steps

1. While the discrepancy principle is not satisfied,

$$\left\| x_{k+1}^{\delta} - x_* \right\| \leq \left\| x_k^{\delta} - x_* \right\| .$$

Proof Steps

1. While the discrepancy principle is not satisfied,

$$\left\| x_{k+1}^{\delta} - x_* \right\| \leq \left\| x_k^{\delta} - x_* \right\| .$$

2. The discrepancy principle yields a well defined stopping rule.

Proof Steps

1. While the discrepancy principle is not satisfied,

$$\left\| x_{k+1}^{\delta} - x_* \right\| \leq \left\| x_k^{\delta} - x_* \right\| .$$

2. The discrepancy principle yields a well defined stopping rule.
3. For exact data, i.e., for $\delta = 0$ or $y = y^{\delta}$, one has

$$\lim_{k \to \infty} x_k^0 \to x_*$$

Proof Steps

1. While the discrepancy principle is not satisfied,

$$\left\|x_{k+1}^\delta - x_*\right\| \leq \left\|x_k^\delta - x_*\right\| .$$

2. The discrepancy principle yields a well defined stopping rule.
3. For exact data, i.e., for $\delta = 0$ or $y = y^\delta$, one has

$$\lim_{k\to\infty} x_k^0 \to x_* \qquad \text{and} \qquad \lim_{k\to\infty} z_k^0 \to x_* .$$

Proof Steps

1. While the discrepancy principle is not satisfied,

$$\left\| x_{k+1}^{\delta} - x_* \right\| \leq \left\| x_k^{\delta} - x_* \right\| .$$

2. The discrepancy principle yields a well defined stopping rule.
3. For exact data, i.e., for $\delta = 0$ or $y = y^{\delta}$, one has

$$\lim_{k \to \infty} x_k^0 \to x_* \qquad \text{and} \qquad \lim_{k \to \infty} z_k^0 \to x_* .$$

4. Combine everything and use continuity to get

$$\lim_{\delta \to 0} z_{k_*(\delta, y^{\delta})}^{\delta} = x_* .$$

Proof Steps

1. While the discrepancy principle is not satisfied,

$$\left\| x_{k+1}^{\delta} - x_* \right\| \leq \left\| x_k^{\delta} - x_* \right\| .$$

2. The discrepancy principle yields a well defined stopping rule.
3. For exact data, i.e., for $\delta = 0$ or $y = y^{\delta}$, one has

$$\lim_{k \to \infty} x_k^0 \to x_* \qquad \text{and} \qquad \lim_{k \to \infty} z_k^0 \to x_* .$$

4. Combine everything and use continuity to get

$$\lim_{\delta \to 0} z_{k_*(\delta, y^{\delta})}^{\delta} = x_* .$$

5. If $\mathcal{N}(F'(x^{\dagger})) \subset \mathcal{N}(F'(x))$, use a special property of $x^{\dagger}$.

Backtracking Algorithm

Backtracking Algorithm

For many stepsizes α_k^δ, the *coupling condition* above reduces to

$$\lambda_k^\delta(\lambda_k^\delta + 1)\left\|x_k^\delta - x_{k-1}^\delta\right\|^2 \leq \frac{\Psi}{\mu}\alpha_k^\delta\left\|y^\delta - F(z_k^\delta)\right\|^2 .$$

Backtracking Algorithm

For many stepsizes α_k^δ, the *coupling condition* above reduces to

$$\lambda_k^\delta(\lambda_k^\delta + 1) \left\| x_k^\delta - x_{k-1}^\delta \right\|^2 \leq \frac{\Psi}{\mu} \alpha_k^\delta \left\| y^\delta - F(z_k^\delta) \right\|^2 .$$

Idea: Given a *summable* sequence $(q_n)_n$, choose λ_k^δ via

$$\lambda_k^\delta = \min \left\{ \frac{q_{n_k}}{\left\| x_k^\delta - x_{k-1}^\delta \right\|}, 1 \right\} ,$$

Backtracking Algorithm

For many stepsizes α_k^δ, the *coupling condition* above reduces to

$$\lambda_k^\delta(\lambda_k^\delta + 1) \left\| x_k^\delta - x_{k-1}^\delta \right\|^2 \leq \frac{\Psi}{\mu} \alpha_k^\delta \left\| y^\delta - F(z_k^\delta) \right\|^2 .$$

Idea: Given a *summable* sequence $(q_n)_n$, choose λ_k^δ via

$$\lambda_k^\delta = \min \left\{ \frac{q_{n_k}}{\left\| x_k^\delta - x_{k-1}^\delta \right\|}, 1 \right\} ,$$

where the subsequence $(q_{n_k})_k$ is chosen such that the above inequality is satisfied.

Backtracking Algorithm

For many stepsizes α_k^δ, the *coupling condition* above reduces to

$$\lambda_k^\delta(\lambda_k^\delta + 1) \left\| x_k^\delta - x_{k-1}^\delta \right\|^2 \leq \frac{\Psi}{\mu} \alpha_k^\delta \left\| y^\delta - F(z_k^\delta) \right\|^2 .$$

Idea: Given a *summable* sequence $(q_n)_n$, choose λ_k^δ via

$$\lambda_k^\delta = \min \left\{ \frac{q_{n_k}}{\left\| x_k^\delta - x_{k-1}^\delta \right\|}, 1 \right\} ,$$

where the subsequence $(q_{n_k})_k$ is chosen such that the above inequality is satisfied.With this choice, one also has

$$\sum_{k=0}^{\infty} \lambda_k^0 \left\| x_k^0 - x_{k-1}^0 \right\| < \infty .$$

Example Problem: SPECT

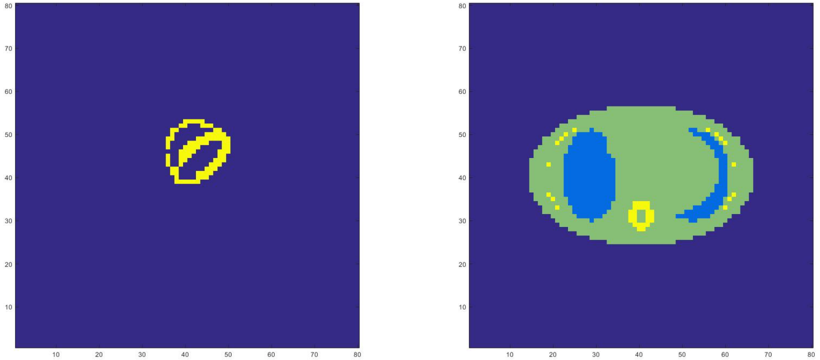
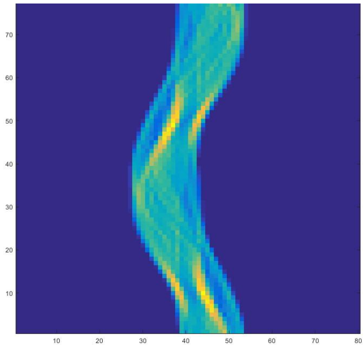
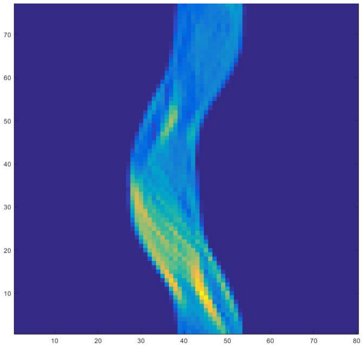


Figure: Activity function f_* (left) and attenuation function μ_* (right).

SPECT Data


70
60
50
40
30
20
10
10
20
30
40
50
60
70
80


70
60
50
40
30
20
10
10
20
30
40
50
60
70
80

Example Problem: SPECT

$$A(f,\mu)(s,\omega) := \int_{\mathbb{R}} f(s\omega^{\perp} + t\omega) \exp\left(- \int_{t}^{\infty} \mu(s\omega^{\perp} + r\omega)\, dr \right)\, dt \,.$$

Example Problem: SPECT

$$A(f,\mu)(s,\omega) := \int_{\mathbb{R}} f(s\omega^\perp + t\omega) \exp\left(- \int_t^\infty \mu(s\omega^\perp + r\omega)\, dr \right) dt \,.$$

Choice of λ_k^δ	k_*	Time
$\lambda_k^\delta = 0$	3433	489 s
Backtracking	631	90 s
Explicit	345	77 s
Nesterov	205	30 s

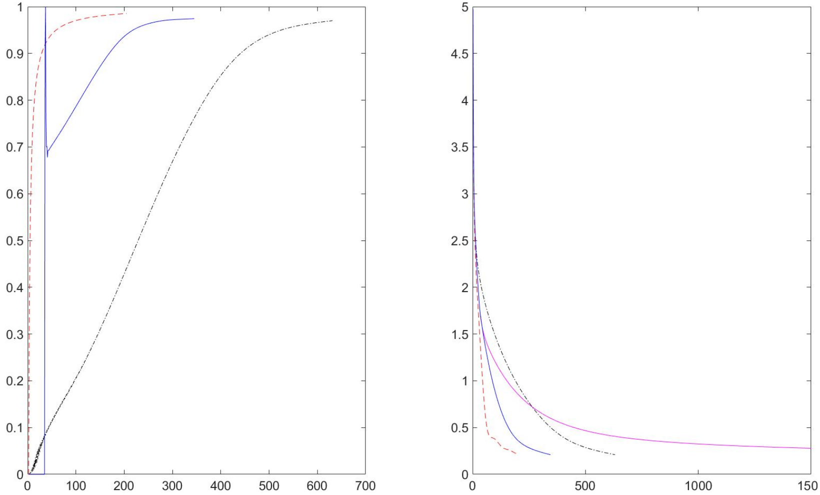
Example Problem: SPECT

$$A(f,\mu)(s,\omega) := \int_{\mathbb{R}} f(s\omega^{\perp} + t\omega) \exp\left(- \int_{t}^{\infty} \mu(s\omega^{\perp} + r\omega)\, dr \right) dt\,.$$

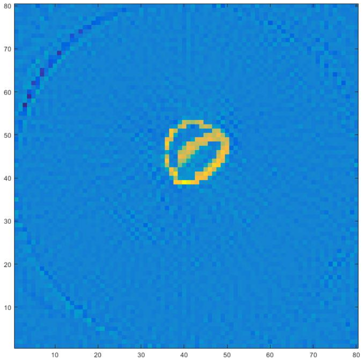
Choice of λ_k^δ	k_*	Time
$\lambda_k^\delta = 0$	3433	489 s
Backtracking	631	90 s
Explicit	345	77 s
Nesterov	205	30 s

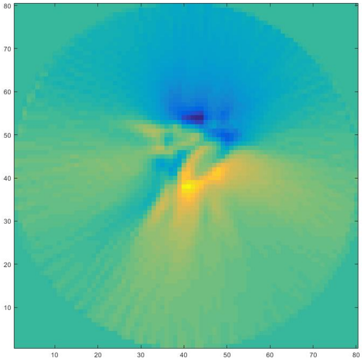
S. Hubmer, R. Ramlau Convergence Analysis of a Two-Point Gradient Method for Nonlinear Ill-Posed Problems, DK Preprint Series (2017)

Evolution of λ_k^δ and Residuals



SPECT Reconstruction





Conclusion

Two-Point Gradient (TPG) methods

$$z_k^\delta = x_k^\delta + \lambda_k^\delta (x_k^\delta - x_{k-1}^\delta) \,,$$
$$x_{k+1}^\delta = z_k^\delta + \alpha_k^\delta F'(z_k^\delta)^* (y^\delta - F(z_k^\delta)) \,,$$

Conclusion

Two-Point Gradient (TPG) methods

$$z_k^\delta = x_k^\delta + \lambda_k^\delta (x_k^\delta - x_{k-1}^\delta)\,,$$
$$x_{k+1}^\delta = z_k^\delta + \alpha_k^\delta F'(z_k^\delta)^*(y^\delta - F(z_k^\delta))\,,$$

- converge under standard assumptions,

Conclusion

Two-Point Gradient (TPG) methods

$$z_k^\delta = x_k^\delta + \lambda_k^\delta (x_k^\delta - x_{k-1}^\delta)\,,$$
$$x_{k+1}^\delta = z_k^\delta + \alpha_k^\delta F'(z_k^\delta)^*(y^\delta - F(z_k^\delta))\,,$$

- converge under standard assumptions,
- are very easy to implement,

Conclusion

Two-Point Gradient (TPG) methods

$$z_k^\delta = x_k^\delta + \lambda_k^\delta (x_k^\delta - x_{k-1}^\delta)\,,$$
$$x_{k+1}^\delta = z_k^\delta + \alpha_k^\delta F'(z_k^\delta)^*(y^\delta - F(z_k^\delta))\,,$$

- converge under standard assumptions,
- are very easy to implement,
- require no more computation time than Landweber iteration,

Conclusion

Two-Point Gradient (TPG) methods

$$z_k^\delta = x_k^\delta + \lambda_k^\delta (x_k^\delta - x_{k-1}^\delta)\,,$$
$$x_{k+1}^\delta = z_k^\delta + \alpha_k^\delta F'(z_k^\delta)^* (y^\delta - F(z_k^\delta))\,,$$

- converge under standard assumptions,
- are very easy to implement,
- require no more computation time than Landweber iteration,
- and lead to a considerable speed-up in practise.

Future Work

Future Work

- Convergence rates of the form

$$\left\| x_{k_*}^{\delta} - x^{\dagger} \right\| = \mathcal{O}\left(\delta^{\frac{2\mu}{2\mu+1}}\right), \qquad k_*(\delta, y^{\delta}) = \psi(\delta).$$

Future Work

- Convergence rates of the form

$$\left\| x_{k_*}^\delta - x^\dagger \right\| = \mathcal{O}\left(\delta^{\frac{2\mu}{2\mu+1}} \right) , \qquad k_*(\delta, y^\delta) = \psi(\delta) \,.$$

- Nonlinear problems and original Nesterov parameter

$$\lambda_k^\delta = \frac{k-1}{k+\alpha-1} \,.$$

Future Work

- Convergence rates of the form

$$\left\|x_{k_*}^\delta - x^\dagger\right\| = \mathcal{O}\left(\delta^{\frac{2\mu}{2\mu+1}}\right), \qquad k_*(\delta, y^\delta) = \psi(\delta)\,.$$

- Nonlinear problems and original Nesterov parameter

$$\lambda_k^\delta = \frac{k-1}{k+\alpha-1}\,.$$

- Analysis only under local convexity assumption.

Future Work

- Convergence rates of the form

$$\left\| x_{k_*}^\delta - x^\dagger \right\| = \mathcal{O}\left(\delta^{\frac{2\mu}{2\mu+1}} \right) , \qquad k_*(\delta, y^\delta) = \psi(\delta) \,.$$

- Nonlinear problems and original Nesterov parameter

$$\lambda_k^\delta = \frac{k-1}{k+\alpha-1} \,.$$

- Analysis only under local convexity assumption.
- Weakening of convexity assumption to, e.g., quasi-convexity.

References

- H. Attouch, J. Peypouquet, The rate of convergence of Nesterov's accelerated forward-backward method is actually $o(k^{-2})$, SIAM Journal on Optimization
- S. Hubmer, R. Ramlau Convergence Analysis of a Two-Point Gradient Method for Nonlinear Ill-Posed Problems, DK Preprint Series (2017)
- Y. Nesterov, A method of solving a convex programming problem with convergence rate $O(1/k^2)$, Soviet Mathematics Doklady, 27, 2, 1983
- A. Neubauer, On Nesterov acceleration for Landweber iteration of linear ill-posed problems, J. Inv. Ill-Posed Problems, to appear.

References

- H. Attouch, J. Peypouquet, The rate of convergence of Nesterov's accelerated forward-backward method is actually $o(k^{-2})$, SIAM Journal on Optimization
- S. Hubmer, R. Ramlau Convergence Analysis of a Two-Point Gradient Method for Nonlinear Ill-Posed Problems, DK Preprint Series (2017)
- Y. Nesterov, A method of solving a convex programming problem with convergence rate $O(1/k^2)$, Soviet Mathematics Doklady, 27, 2, 1983
- A. Neubauer, On Nesterov acceleration for Landweber iteration of linear ill-posed problems, J. Inv. Ill-Posed Problems, to appear.

Thank you for your attention!