Two-Point Gradient Methods for Nonlinear Ill-Posed Problems

Simon Hubmer and Ronny Ramlau

Johann Radon Institute for Computational and Applied Mathematics (RICAM)
Austrian Academy of Sciences (ÖAW)
Linz, Austria

Joint Fudan - RICAM Seminar, July 8, 2020

Outline

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

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

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:

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

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

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

- 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

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

- Tikhonov = Minimize{ $\Phi(x)$ + 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

General minimization problem

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

Yurii Nesterov: Instead of using gradient descent:

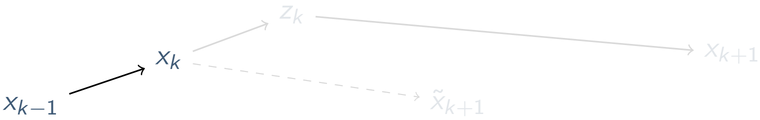
$$x_{k+1} = x_k - \omega \nabla \Phi(x_k)\,,$$

use the following iteration:

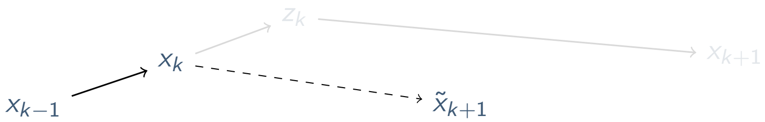
$$z_k = x_k + \tfrac{k-1}{k+\alpha-1}(x_k - x_{k-1})\,,$$
$$x_{k+1} = z_k - \omega \nabla \Phi(z_k)\,.$$

Motivating Picture

Motivating Picture

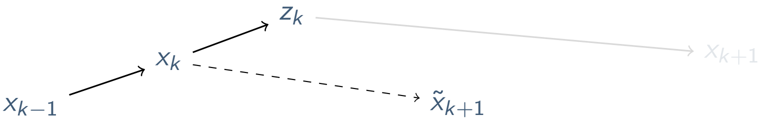


Motivating Picture



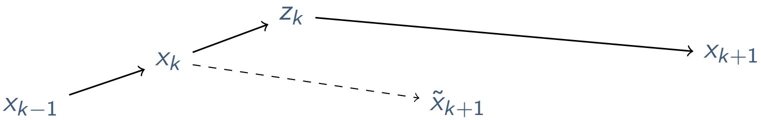
$$\tilde{x}_{k+1} = x_k - \omega \nabla \Phi(x_k)$$

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?

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

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

- **Nesterov Acceleration:**

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

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

For our problem, the method reads

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

There is a generalization to deal with

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

which reads

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

$\Longrightarrow$ Sparsity Constraints, Projections, etc.

Application to Nonlinear Ill-Posed Problems

For our problem, the method reads

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

There is a generalization to deal with

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

which reads

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

$\Longrightarrow$ Sparsity Constraints, Projections, etc.

Neubauer's Linear Results

- **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, Vol. 25, No. 3, 2017

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 $\lambda_k^\delta = \frac{k-1}{k+\alpha-1}$ ← 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
- Yes for $\lambda_k^\delta = \frac{k-1}{k+\alpha-1}$ and a **locally convex residual functional**

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
- Yes for $\lambda_k^\delta = \frac{k-1}{k+\alpha-1}$ and a **locally convex residual functional**

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
- Yes for $\lambda_k^\delta = \frac{k-1}{k+\alpha-1}$ and a **locally convex residual functional**

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
- Yes for $\lambda_k^\delta = \frac{k-1}{k+\alpha-1}$ and a **locally convex residual functional**

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
- Yes for $\lambda_k^\delta = \frac{k-1}{k+\alpha-1}$ and a **locally convex residual functional**

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
- Yes for $\lambda_k^\delta = \frac{k-1}{k+\alpha-1}$ and a **locally convex residual functional**

Two-Point Gradient (TPG) Methods

How about general methods of the form

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

Question: Do they converge under standard assumptions?

- Yes for linear problems and $\lambda_k^\delta = \frac{k-1}{k+\alpha-1}$ ← 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
- Yes for $\lambda_k^\delta = \frac{k-1}{k+\alpha-1}$ and a **locally convex residual functional**

Convergence Conditions I

- 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

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

- Parameters λ_k^δ satisfy

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

Convergence Conditions I

- 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

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

- Parameters λ_k^δ satisfy

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

Convergence Conditions I

- 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

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

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

- 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

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.

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 I

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

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

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

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

Main Result I

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*

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

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

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

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

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

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

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

Convergence Conditions II

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

- Φ^0 is convex in $\mathcal{B}_{6\rho}$

$$\Phi^0(\lambda x_1 + (1-\lambda)x_2) \leq \lambda\Phi^0(x_1) + (1-\lambda)\Phi^0(x_2)$$

- Φ^0 is Lipschitz continuous with constant L in $\mathcal{B}_{6\rho}$

$$\left\|\Phi^0(x_1) - \Phi^0(x_2)\right\| \leq L\left\|x_1 - x_2\right\|$$

- $\alpha > 3$ and the scaling satisfies $0 < \omega \leq 1/L$

Convergence Conditions II

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

- Φ^0 is convex in $\mathcal{B}_{6\rho}$

$$\Phi^0(\lambda x_1 + (1-\lambda)x_2) \leq \lambda\Phi^0(x_1) + (1-\lambda)\Phi^0(x_2)$$

- Φ^0 is Lipschitz continuous with constant L in $\mathcal{B}_{6\rho}$

$$\left\|\Phi^0(x_1) - \Phi^0(x_2)\right\| \leq L\left\|x_1 - x_2\right\|$$

- $\alpha > 3$ and the scaling satisfies $0 < \omega \leq 1/L$

Convergence Conditions II

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

- Φ^0 is convex in $\mathcal{B}_{6\rho}$

$$\Phi^0(\lambda x_1 + (1-\lambda)x_2) \leq \lambda\Phi^0(x_1) + (1-\lambda)\Phi^0(x_2)$$

- Φ^0 is Lipschitz continuous with constant L in $\mathcal{B}_{6\rho}$

$$\left\|\Phi^0(x_1) - \Phi^0(x_2)\right\| \leq L\left\|x_1 - x_2\right\|$$

- $\alpha > 3$ and the scaling satisfies $0 < \omega \leq 1/L$

Main Result II

Adapted Discrepancy Principle:

$$\left\| y^\delta - F(x_{k_*}^\delta) \right\|^2 \leq \frac{2(k+\alpha-1)^2}{k(\alpha-3)} \Delta(\delta) + \tau^2\delta^2 < \left\| y^\delta - F(x_k^\delta) \right\|^2$$

Theorem

Under the above assumptions, one gets weak subsequential convergence of $x_{k_*(\delta,y^\delta)}^\delta$ *to an element* x_* *in the solution set*

Main Result II

Adapted Discrepancy Principle:

$$\left\|y^\delta - F(x_{k_*}^\delta)\right\|^2 \leq \frac{2(k+\alpha-1)^2}{k(\alpha-3)}\Delta(\delta) + \tau^2\delta^2 < \left\|y^\delta - F(x_k^\delta)\right\|^2$$

Theorem

Under the above assumptions, one gets weak subsequential convergence of $x^\delta_{k_*(\delta,y^\delta)}$ *to an element* x_* *in the solution set*

Example Problem: Hammerstein

$$F(x)(s) = \int_0^1 k(s,t)\phi(x(t))\,dt$$

$$x^\dagger(t) = 1 + 10^{-2}(7 - 3t^2 + 2t^3) \qquad x_0(t) = 1$$

Choice of λ_k^δ	k_*	Time
$\lambda_k^\delta = 0$	125	79 s
Backtracking	41	26 s
Explicit	35	22 s
Nesterov	14	9 s

Example Problem: Hammerstein

$$F(x)(s) = \int_0^s x^3(t)\,dt$$

$$x^\dagger(t) = 1 + 10^{-2}(7 - 3t^2 + 2t^3) \qquad x_0(t) = 1$$

Choice of λ_k^δ	k_*	Time
$\lambda_k^\delta = 0$	125	79 s
Backtracking	41	26 s
Explicit	35	22 s
Nesterov	14	9 s

Example Problem: Hammerstein

$$F(x)(s) = \int_0^s x^3(t)\,dt$$

$$x^\dagger(t) = 1 + 10^{-2}(7 - 3t^2 + 2t^3) \qquad x_0(t) = 1$$

Choice of λ_k^δ	k_*	Time
$\lambda_k^\delta = 0$	125	79 s
Backtracking	41	26 s
Explicit	35	22 s
Nesterov	14	9 s

Example Problem: SPECT

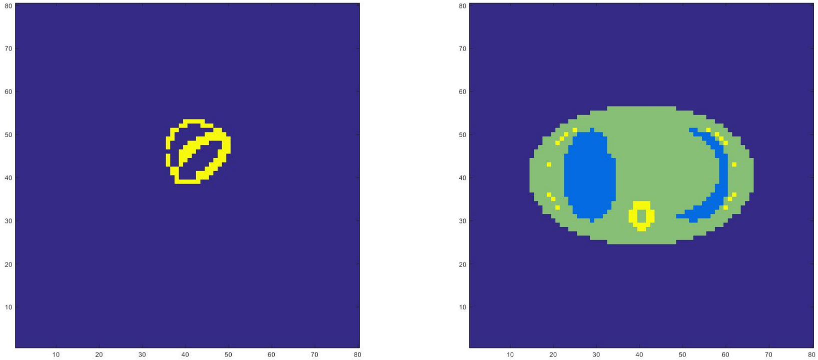


Figure: Activity function f_* (left) and attenuation function μ_* (right).

Example Problem: SPECT

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

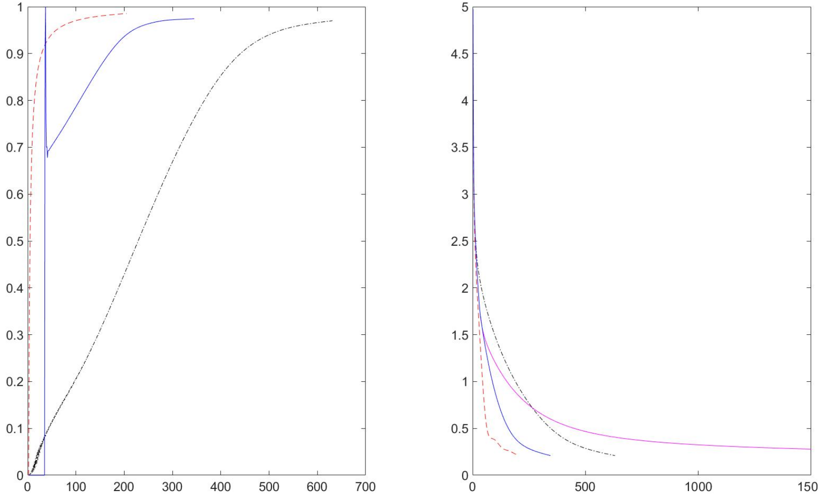
Choice of λ_k^δ	k_*	Time
$\lambda_k^\delta = 0$	3433	489 s
Backtracking	349	77 s
Explicit	631	90 s
Nesterov	205	30 s

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	349	77 s
Explicit	631	90 s
Nesterov	205	30 s

Evolution of λ_k^δ and Residuals



Example Problem: Autoconvolution

$$F(x)(s) = (x * x)(s) = \int_0^1 x(s-t)x(t)\,dt$$

$$x^\dagger = \qquad x_0(t) =$$

Example Problem: Autoconvolution

$$F(x)(s) = (x * x)(s) = \int_0^1 x(s-t)x(t)\,dt$$

$$x^\dagger = 10 + \sqrt{2}\sin(2\pi s) \qquad x_0(t) = 10 + \frac{27}{28}\sqrt{2}\sin(2\pi s)$$

Choice of λ_k^δ	k_*	Time	Rel. Error
$\lambda_k^\delta = 0$	526	57 s	0.0244 %
Nesterov	50	6 s	0.0271 %

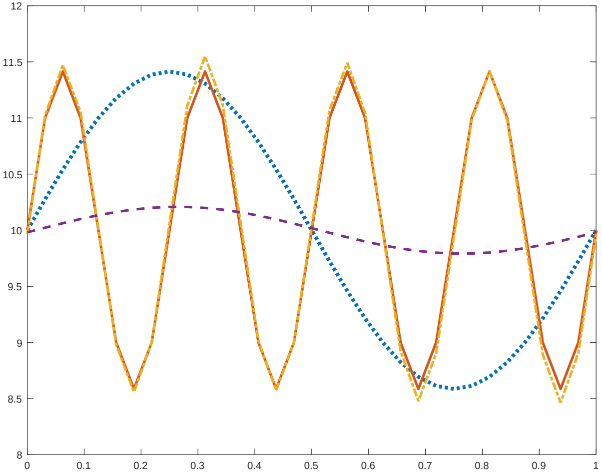
Example Problem: Autoconvolution

$$F(x)(s) = (x * x)(s) = \int_0^1 x(s-t)x(t)\,dt$$

$$x^\dagger = 10 + \sqrt{2}\sin(8\pi s) \qquad x_0(t) = 10 + \sqrt{2}\sin(2\pi s)$$

Choice of λ_k^δ	k_*	Time	Rel. Error
$\lambda_k^\delta = 0$	10000	1067 s	9.57 %
Nesterov	797	87 s	0.65 %

Example Problem: Autoconvolution



Some Recent Developments - 2018

J. Wang, W. Wang, B. Han, Regularization of inverse problems by two-point gradient methods with convex constraints ArXiv preprint, 2018.

Y. Zhang, B. Hofmann, On the second-order asymptotical regularization of linear ill-posed inverse problems Applicable Analysis, 2018.

Some Recent Developments - 2019

- P. Pornsawad, N. Sapsakul, C. Böckmann A Modified Asymptotical Regularization of Nonlinear Ill-Posed Problems Mathematics, 2019.
- G. Gao, B. Han, S. Tong A fast two-point gradient algorithm based on sequential subspace optimization method for nonlinear ill-posed problems arXiv preprint, 2019.
- M. Zhong, W. Wang, Q. Jin, Regularization of inverse problems by two-point gradient methods with convex constraints Numerische Mathematik, 2019.
- H. Long, B. Han., S. Tong, A new Kaczmarz-type method and its acceleration for nonlinear ill-posed problems Inverse Problems, 2019.
- M. Zhong, W. Wang, A regularizing multilevel approach for nonlinear inverse problems Applied Numerical Mathematics, 2019.
- S. Tong, B. Han, H. Long, R. Gu, An accelerated sequential subspace optimization method based on homotopy perturbation iteration for nonlinear ill-posed problems Inverse Problems, 2019.
- J. Wang, W. Wang, B. Han, An iterative regularization method with general convex penalty for nonlinear inverse problems in Banach spaces Journal of Computational and Applied Mathematics, 2019.

Some Recent Developments - 2020

- M. Zhong, W. Wang The two-point gradient methods for nonlinear inverse problems based on Bregman projections Inverse Problems, 2020.
- M. Zhong, J. Liu Recovery of non-smooth radiative coefficient from nonlocal observation by diffusion system Journal of Inverse and Ill-Posed Problems, 2020.
- G. Mittal, A. K. Giri A novel two-point gradient method for Regularization of inverse problems in Banach spaces ArXiv Preprint, 2020.
- H. Long, B. Han, S. Tong A proximal regularized Gauss-Newton-Kaczmarz method and its acceleration for nonlinear ill-posed problems Applied Numerical Mathematics, 2020.
- R. Gong, B. Hofmann, Y. Zhang A new class of accelerated regularization methods, with application to bioluminescence tomography Inverse Problems, 2020.
- Y. Zhang, R. Gong, Second order asymptotical regularization methods for inverse problems in partial differential equations Journal of Computational and Applied Mathematics, 2020.

Summary

Two-Point Gradient (TPG) methods

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

- 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 practice.

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, Inverse Problems, Vol. 33 No. 3, 2017
- S. Hubmer, R. Ramlau Nesterov's Accelerated Gradient Method for Nonlinear Ill-Posed Problems with a Locally Convex Residual Functional Accepted, 2018.
- 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, Vol. 25, No. 3, 2017

Thank you for your attention!

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, Inverse Problems, Vol. 33 No. 3, 2017
- S. Hubmer, R. Ramlau Nesterov's Accelerated Gradient Method for Nonlinear Ill-Posed Problems with a Locally Convex Residual Functional Accepted, 2018.
- 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, Vol. 25, No. 3, 2017

Thank you for your attention!