Proximal Gradient Algorithm in the Presence of Adjoint Mismatch

Emilie Chouzenoux[1] (joint work with M. Savanier[1,2], J.-C. Pesquet[1], C. Riddell[2] and Y. Trousset[2])

(1) Université Paris-Saclay, Inria, CentraleSupélec, Centre de Vision Numérique, Gif-sur-Yvette, France.
(2) GE Healthcare, Buc, France.

February 2nd, 2021

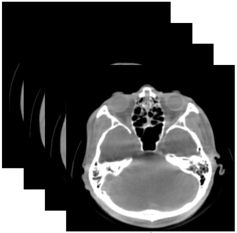
LE SÉMINAIRE
PALAISIEN

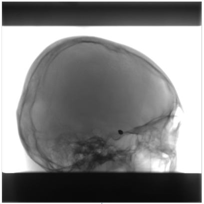
Introduction

Inverse problem formulation

Large panel of applications:
medical imaging, fluorescence microscopy, astronomy, image restoration, etc.

Object of interest $\bar{\mathbf{x}} \in \mathbb{R}^N$ → System → Observation $\mathbf{y} = \mathbf{H}\bar{\mathbf{x}} + \mathbf{b} \in \mathbb{R}^M$





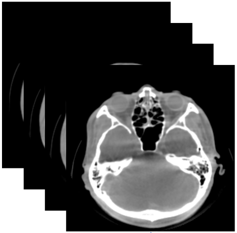
How?

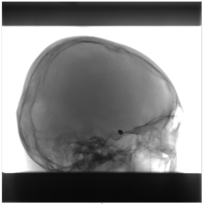
$$\underset{\mathbf{x} \in \mathbb{R}^N}{\text{minimize}} \quad \underbrace{f_1(\mathbf{x})}_{\text{Datafit}} + \underbrace{f_2(\mathbf{x})}_{\text{Regularization}}$$

Inverse problem formulation

Large panel of applications:
medical imaging, fluorescence microscopy, astronomy, image restoration, etc.

Object of interest $\bar{\mathbf{x}} \in \mathbb{R}^N$ → System → Observation $\mathbf{y} = \mathbf{H}\bar{\mathbf{x}} + \mathbf{b} \in \mathbb{R}^M$





How?

$$\underset{\mathbf{x}\in\mathbb{R}^N}{\text{minimize}} \quad \underbrace{f_1(\mathbf{x})}_{\text{Datafit}} + \underbrace{f_2(\mathbf{x})}_{\text{Regularization}}$$

Iterative approaches for penalized least-squares

- Tikhonov problem:

$$\underset{\mathbf{x}\in\mathbb{R}^N}{\text{minimize}} \quad \underbrace{\frac{1}{2}\|\mathbf{y}-\mathbf{H}\mathbf{x}\|^2}_{f_1(\mathbf{x})} + \underbrace{\frac{\kappa}{2}\|\mathbf{x}\|^2}_{f_2(\mathbf{x})}, \tag{P0}$$

  where $\kappa \in [0,+\infty[$

- Gradient algorithm:

$$\boxed{\mathbf{x}_{n+1} = (1-\gamma\kappa)\mathbf{x}_n - \gamma\mathbf{H}^\top(\mathbf{H}\mathbf{x}_n - \mathbf{y}) \qquad \text{(A0)}}$$

  $\Rightarrow$ Linear iteration

- Only differentiable regularization can be added
- Variants: Algebraic iterative methods (SIRT, SART, etc)

Iterative approaches for penalized least-squares

- Penalized least-squares:

$$\underset{\mathbf{x}\in\mathbb{R}^N}{\text{minimize}} \quad \underbrace{\frac{1}{2}\|\mathbf{y}-\mathbf{H}\mathbf{x}\|^2}_{f_1(\mathbf{x})} + \underbrace{g(\mathbf{x}) + \frac{\kappa}{2}\|\mathbf{x}\|^2}_{f_2(\mathbf{x})}, \tag{P1}$$

  where g convex possibly non-smooth regularization function and $\kappa \in [0,+\infty[$.
  $\Rightarrow$ Embed more sophisticated priors (e.g., range constraint, ℓ_1, total variation)

- Simple yet efficient approach = Proximal Gradient Algorithm (PGA)

$$\boxed{\mathbf{x}_{n+1} = \mathbf{x}_n + \theta_n\left(\mathrm{prox}_{\gamma g}((1-\gamma\kappa)\mathbf{x}_n - \gamma\mathbf{H}^\top(\mathbf{H}\mathbf{x}_n-\mathbf{y})) - \mathbf{x}_n\right) \qquad \text{(A1)}}$$

  $\Rightarrow$ reduces to (A0) when $g = 0$ and $\theta_n \equiv 1$

Proposition (Convergence of PGA)

- $\gamma \in]0, 2/(|||\mathbf{H}|||^2+\kappa)[$,
- $\theta_n \in [\epsilon, 1]$ with $\epsilon \in]0,1[$

Then the sequence $(\mathbf{x}_n)_{n\in\mathbb{N}}$ generated by Algorithm (A1) converges to a solution to Problem (P1) when such a solution exists.

Iterative approaches for penalized least-squares

- Penalized least-squares:

$$\underset{\mathbf{x}\in\mathbb{R}^N}{\text{minimize}} \quad \underbrace{\frac{1}{2}\|\mathbf{y}-\mathbf{H}\mathbf{x}\|^2}_{f_1(\mathbf{x})} + \underbrace{g(\mathbf{x}) + \frac{\kappa}{2}\|\mathbf{x}\|^2}_{f_2(\mathbf{x})}, \tag{P1}$$

where g is a suitable convex possibly non-smooth regularization function and $\kappa \in [0, +\infty[$.

- Simple yet efficient approach = Proximal Gradient Algorithm (PGA)

$$\boxed{\mathbf{x}_{n+1} = \mathbf{x}_n + \theta_n\big(\mathrm{prox}_{\gamma g}((1-\gamma\kappa)\mathbf{x}_n - \gamma\mathbf{K}(\mathbf{H}\mathbf{x}_n - \mathbf{y})) - \mathbf{x}_n\big), \qquad \text{(A2)}}$$

Question

What can be said about the sequence $(\mathbf{x}_n)_{n\in\mathbb{N}}$ generated by Algorithm (A2)?
$\Rightarrow$ when $g \equiv 0$ see [EH18, ZG00, DHHBR19, LRS18].

- This talk : Characterization of PGA convergence and limit points in the general case of a convex function g in the presence of **an adjoint mismatch**.

Motivation: Image reconstruction in X-ray tomography

- Hardware optimization

  GPU-friendly implementation (ASTRA Toolbox), limited memory budget
- Different discretizations to cope with changes in sampling rates

  Rotation in tomographic reconstruction (parallel and divergent geometry)
  $\Rightarrow$ different strategies suitable for iterative and analytical reconstruction
- Improved conditioning of **KH**

  SPECT (attenuation not modeled in the backprojector), embedding of the ramp filter

Proposed stability analysis

Mathematical background

Definition (Proximity operator)

Let $\Gamma_0(\mathbb{R}^N)$ be the set of functions which take values in $\mathbb{R} \cup \{+\infty\}$ and are proper convex, lower semicontinuous on $\mathbb{R}^N$. The proximity operator of $g \in \Gamma_0(\mathbb{R}^N)$ at $\mathbf{x} \in \mathbb{R}^N$ is defined as

$$\mathrm{prox}_g(\mathbf{x}) = \underset{\mathbf{z}\in\mathbb{R}^N}{\mathrm{argmin}} \left(g(\mathbf{z}) + \frac{1}{2}\|\mathbf{x}-\mathbf{z}\|^2 \right).$$

- $\mathbf{p} = \mathrm{prox}_g(\mathbf{x}) \quad \Leftrightarrow \quad \mathbf{x}-\mathbf{p} \in \partial g(\mathbf{p})$.

Notation

Let $\mathbf{L} = \mathbf{KH} + \kappa\mathbf{Id}$ with $\kappa > 0$.
Let $\lambda_{\min}$ (resp. $\lambda_{\max}$) be the minimum (resp. maximum) eigenvalue of $(\mathbf{L}+\mathbf{L}^\top)/2$.
Let $\lambda_{\min}^{+}$ be the minimum positive eigenvalue of $(\mathbf{L}+\mathbf{L}^\top)/2$ and let $\beta = |||\mathbf{L}-\mathbf{L}^\top|||/2$.
Let T_γ be the operator defined as

$$\begin{aligned} T_\gamma \colon \mathbb{R}^N &\to \mathbb{R}^N \\ \mathbf{x} &\mapsto \mathrm{prox}_{\gamma g}((1-\gamma\kappa)\mathbf{x} - \gamma\mathbf{K}(\mathbf{Hx}-\mathbf{y})) \end{aligned}$$

Fixed point algorithm

Equivalent formulation

Proximal Gradient Algorithm (PGA):

$$
\begin{aligned}
\mathbf{x}_{n+1} &= \mathbf{x}_n + \theta_n\big(\mathrm{prox}_{\gamma g}((1-\gamma\kappa)\mathbf{x}_n - \gamma\mathbf{K}(\mathbf{H}\mathbf{x}_n - \mathbf{y})) - \mathbf{x}_n\big)\\
&= \mathbf{x}_n + \theta_n\big(T_\gamma(\mathbf{x}_n) - \mathbf{x}_n\big)
\end{aligned}
$$

Fixed point operator

Analysis of the properties of:

$$
\begin{aligned}
T_\gamma : \mathbb{R}^N &\to \mathbb{R}^N\\
\mathbf{x} &\mapsto \mathrm{prox}_{\gamma g}((1-\gamma\kappa)\mathbf{x} - \gamma\mathbf{K}(\mathbf{H}\mathbf{x} - \mathbf{y}))
\end{aligned}
$$

or, equivalently,

$$
\begin{aligned}
T_\gamma : \mathbb{R}^N &\to \mathbb{R}^N\\
\mathbf{x} &\mapsto \mathrm{prox}_{\gamma g}(\mathbf{x} - \gamma\mathbf{L}\mathbf{x} + \gamma\mathbf{K}\mathbf{y})
\end{aligned}
$$

Methodology

Non-linear iteration

$\Rightarrow$ tools from convex analysis, monotone operators and fixed point theory.
$\Rightarrow$ Proposed analysis pipeline:

- Cocoercivity of $\mathbf{L}$?
- Existence and uniqueness of the fixed points of T_γ ?

Definition (Cocoercivity)

Operator $A\colon \mathbb{R}^N \to \mathbb{R}^N$ is η-cocoercive with $\eta \in [0, +\infty[$ if, for every $(\mathbf{x}, \mathbf{y}) \in (\mathbb{R}^N)^2$,

$$\eta \|A\mathbf{x} - A\mathbf{y}\|^2 \leqslant \langle \mathbf{x} - \mathbf{y}, A\mathbf{x} - A\mathbf{y} \rangle.$$

Example: the gradient of a η-Lipschitz differentiable function is a η-cocoercive operator.

Recover the cocoercivity of $\mathbf{L}$

Proposition

- $\mathbf{L}$ is η-**cocoercive** with $\eta \in]0,+\infty[$ if and only if $\lambda_{\min} \geqslant 0$, $\mathrm{Ker}(\mathbf{L}+\mathbf{L}^\top) = \mathrm{Ker}\,\mathbf{L}$, and

$$\eta \leqslant \overline{\eta} = \frac{2}{|||(\mathbf{Id} + (\mathbf{L}-\mathbf{L}^\top)(\mathbf{L}+\mathbf{L}^\top)^{-1})(\mathbf{L}+\mathbf{L}^\top)^{1/2}|||^2}.$$

- Assume that $\lambda_{\min} \geqslant 0$.
  If $\mathrm{Ker}(\mathbf{L}+\mathbf{L}^\top) = \mathrm{Ker}\,\mathbf{L}$, then $\mathbf{L}$ is $\underline{\eta}$-**cocoercive** with

$$\underline{\eta} = 1\Big/\left(\sqrt{\lambda_{\max}} + \frac{\beta}{\sqrt{\lambda_{\min}^{+}}}\right)^2.$$

  If $\beta = 0$, then $\mathbf{L}$ is $(1/\lambda_{\max})$-cocoercive.

Example

A simple condition for ensuring that $\lambda_{\min}$ is positive is to choose $\kappa > -\widetilde{\lambda}_{\min}$ where $\widetilde{\lambda}_{\min}$ is the minimum eigenvalue of $(\mathbf{KH}+\mathbf{H}^\top\mathbf{K}^\top)/2$. In this case, $\lambda_{\min}^{+} = \lambda_{\min} > 0$, while $\mathrm{Ker}(\mathbf{L}+\mathbf{L}^\top)$ and $\mathrm{Ker}\,\mathbf{L}$ reduce to the null space.

Fixed points of T_γ

Let $\gamma \in]0, +\infty[$ and let $\widetilde{\mathbf{x}} \in \mathbb{R}^N$. We have $\widetilde{\mathbf{x}} \in \operatorname{Fix} T_\gamma$ if and only if $\widetilde{\mathbf{x}}$ belongs to

$$\mathcal{F} = \left\{ \mathbf{x} \in \mathbb{R}^N \;\middle|\; 0 \in \mathbf{K}(\mathbf{H}\mathbf{x} - \mathbf{y}) + \partial g(\mathbf{x}) + \kappa \mathbf{x} \right\}.$$

Proposition (Existence)

- If $\lambda_{\min} \geqslant 0$, then $\mathcal{F}$ is a closed and convex set.
- Assume that $\mathbf{L}$ is cocoercive. $\mathcal{F}$ is **nonempty** if one of the following condition holds:
  - $\operatorname{dom} \partial g = \mathbb{R}^N$ and
    $$\mathbf{x} \mapsto \frac{1}{2} \langle \mathbf{x} \mid \mathbf{L}\mathbf{x} \rangle + g(\mathbf{x})$$
    is coercive;
  - $\operatorname{dom} g$ is bounded.

Fixed points of T_γ

Let $\gamma \in]0, +\infty[$ and let $\widetilde{\mathbf{x}} \in \mathbb{R}^N$. We have $\widetilde{\mathbf{x}} \in \mathrm{Fix}\, T_\gamma$ if and only if $\widetilde{\mathbf{x}}$ belongs to

$$\mathcal{F} = \left\{ \mathbf{x} \in \mathbb{R}^N \;\middle|\; 0 \in \mathbf{K}(\mathbf{H}\mathbf{x} - \mathbf{y}) + \partial g(\mathbf{x}) + \kappa \mathbf{x} \right\}.$$

Proposition (Uniqueness)

$\mathcal{F}$ is **a singleton** if $\lambda_{\text{min}} \geqslant 0$ and one of the following condition holds:

- $\lambda_{\text{min}} \neq 0$;
- g is strongly convex.

Convergence results

Theorem

Assume that the following hold.

(i) $\mathbf{L}$ *is cocoercive.*

(ii) Let $\nu \in [0,+\infty[$ *be the strong convexity modulus of* g*. Either* $\nu > 0$ *or* $\lambda_{\min} \neq 0$*.*

(iii) $\widehat{\mathbf{x}} = \underset{\mathbf{x}\in\mathbb{R}^N}{\operatorname{argmin}} \; \frac{1}{2}\|\mathbf{y}-\mathbf{H}\mathbf{x}\|^2 + g(\mathbf{x}) + \frac{\kappa}{2}\|\mathbf{x}\|^2$

Then there **exists a unique solution** $\widetilde{x} \in \mathcal{F}$ *and:*

$$\|\widetilde{\mathbf{x}} - \widehat{\mathbf{x}}\| \leqslant \chi \, |||\mathbf{H}^\top - \mathbf{K}||| \, \|\mathbf{H}\widehat{\mathbf{x}} - \mathbf{y}\|$$

where $\chi = (\nu + 2\lambda_{\min})^{-1}$.

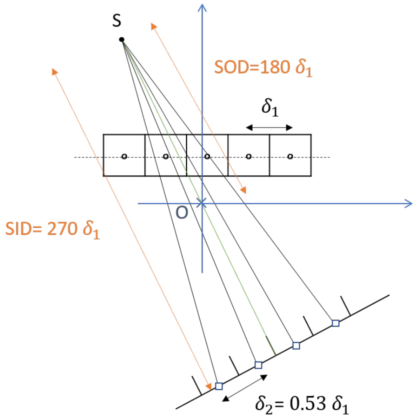
$\Rightarrow$ Trade-off between fidelity and bias.

Proposition (Convergence)

Assume that $\mathbf{L}$ is η-cocoercive. Let $\gamma \in]0, 2\eta[$ and $\delta = 2 - \gamma/(2\eta)$. Let $(\theta_n)_{n\in\mathbb{N}}$ be a sequence in $[0,\delta]$ such that $\sum_{n\in\mathbb{N}} \theta_n(\delta - \theta_n) = +\infty$. Suppose that $\mathcal{F} \neq \varnothing$.
Then, the sequence $(\mathbf{x}_n)_{n\in\mathbb{N}}$ generated by Algorithm (A2) **converges** to a point $\widetilde{\mathbf{x}} \in \mathcal{F}$. In addition, if $\lambda_{\min} \neq 0$ and, for every $n \in \mathbb{N}$, $\theta_n \in [\underline{\theta}, 1]$ with $\underline{\theta} \in]0,+\infty[$, then $(\mathbf{x}_n)_{n\in\mathbb{N}}$ converges linearly.

Numerical experiment

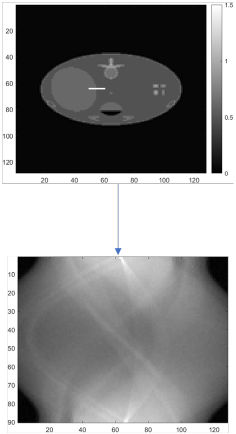
Tomographic reconstruction



Detector has 128 bins => **Truncation**

$$\begin{cases} M = 90 \times 128 \\ N = 128 \times 128 \end{cases}$$

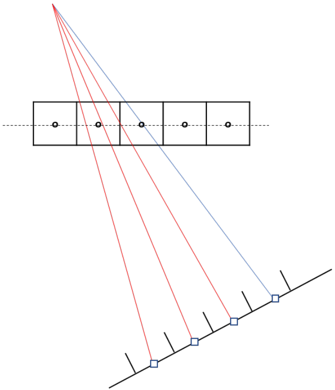
$\bar{\mathbf{x}} \in \mathbb{R}^N$



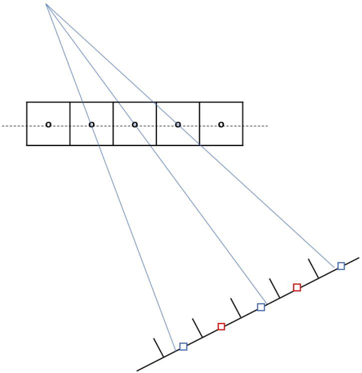
$$\mathbf{y} = \mathbf{H}\bar{\mathbf{x}} + \mathbf{b} \in \mathbb{R}^M$$

$$\mathbf{b} \sim \mathcal{N}(\mathbf{0}, \mathbf{0.2})$$

Origin of the mismatch



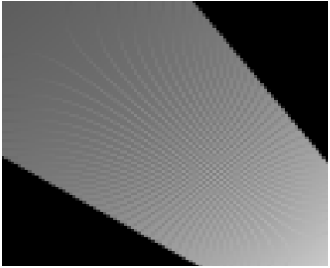
H



K

Origin of the mismatch

- Zoom on the backprojection of a uniform view:



$\mathbf{H}^\top$ (sparsity $s = |||\mathbf{H}|||_0/MN = 1.0778\%$)

$\mathbf{K}$ (sparsity $s = |||\mathbf{K}|||_0/MN = 0.89\%$)

$$\mu = \langle \mathbf{Hu}, \mathbf{v} \rangle / \langle \mathbf{u}, \mathbf{Kv} \rangle \approx 1.0076$$

$$|||\mathbf{K}|||_0 = \sum_{i=1}^{N} \sum_{j=1}^{M} \mathbb{I}(K_{ij} \neq 0)$$

Modeling and parameterization

$$\underset{\mathbf{x}\in\mathbb{R}^N}{\text{minimize}} \quad \frac{1}{2}\|\mathbf{y}-\mathbf{H}\mathbf{x}\|^2 + \underbrace{\lambda\|\mathbf{W}\mathbf{x}\|_1}_{g(\mathbf{x})} + \frac{\kappa}{2}\|\mathbf{x}\|^2,$$

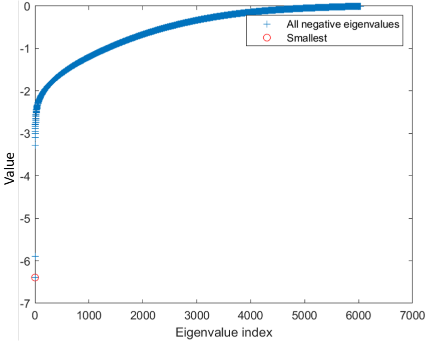
where $\mathbf{W}$ being the orthogonal Symlet 2 wavelet transform, $\lambda = 0.45$, $\kappa = \{\kappa_1, \kappa_2\}$ with $\kappa_1 = 0.01$, $\kappa_2 = 6.5$.

Two scenarios

$\mathbf{L}$ is cocoercive for:

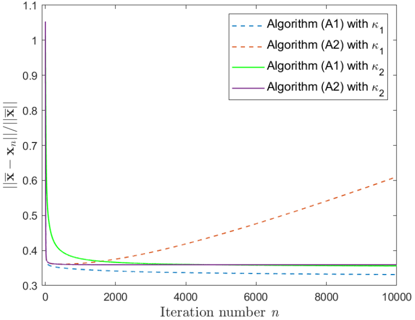
- ☒ κ_1
- ☑ κ_2

Settings:
$\gamma = 1.9/(|||\mathbf{H}|||^2 + \kappa)$,
$\theta_n \equiv 1$.

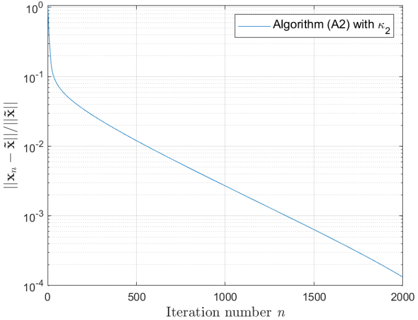


Distribution of the negative eigenvalues of $(\mathbf{K}\mathbf{H} + \mathbf{H}^\top\mathbf{K}^\top)/2$

Results

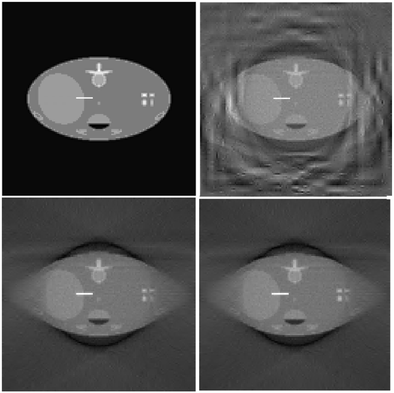


Decay of the error along iterations for Algorithms (A1) and (A2) and two choices of κ parameter.



Decay of the distance to the optimum along iterations for Algorithms (A2) with $\kappa = \kappa_2$.

Results



Original phantom (top left), and reconstruction results using Algorithm (A2) with κ_1 (top right), Algorithm (A1) with κ_2 (bottom left) and Algorithm (A2) with κ_2 (bottom right)

κ	$\mathbf{H}^\top\mathbf{H}$	$\mathbf{KH}$
κ_1	24.41	22.32
κ_2	26.02	25.06

SNR (dB) over central ROI

Conclusion and perspectives

- Characterization of the fixed points of PGA in the presence of an adjoint mismatch.
- Conditions of convergence with new bounds on the gradient step-size and on the regularization parameters.
- Characterization of the distance from the generated fixed point of the algorithm to a "true" minimizer of the original objective function.
- Validation of these results on an image reconstruction task.

$\Rightarrow$ reconcile theory with practical implementations of PGA iteration in the context of X-ray tomographic imaging.

Future work:

- Extend our analysis to other types of data fidelity: convex ℓ_1 or more robust non-convex ℓ_p potentials ($p < 1$).

Thank You!

References

Emilie Chouzenoux, Jean-Christophe Pesquet, Cyril Riddell, Marion Savanier, and Yves Trousset, *Convergence of proximal gradient algorithm in the presence of adjoint mismatch*, Inverse Problems (to appear) (2021).

Yiqiu Dong, Per Christian Hansen, Michiel E. Hochstenbach, and Nicolai André Brogaard Riis, *Fixing nonconvergence of algebraic iterative reconstruction with an unmatched backprojector*, SIAM Journal on Scientific Computing **41** (2019), no. 3, A1822–A1839.

Tommy Elfving and Per Christian Hansen, *Unmatched projector/backprojector pairs: Perturbation and convergence analysis*, SIAM Journal on Scientific Computing **40** (2018), no. 1, A573–A591.

Dirk A. Lorenz, Sean Rose, and Frank Schöpfer, *The randomized Kaczmarz method with mismatched adjoint*, BIT Numerical Mathematics **58** (2018), 1079–1098.

Gengsheng L. Zeng and Grant T. Gullberg, *Unmatched projector/backprojector pairs in an iterative reconstruction algorithm*, IEEE Transactions on Medical Imaging (2000).