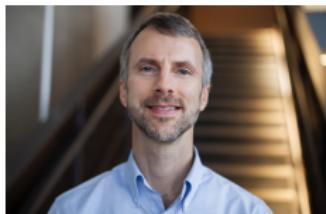


Generalizing Nesterov's Scheme and Magical High-order Methods

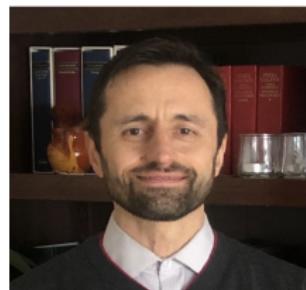
Tao Hong

working with Prof. Jeffrey A. Fessler and Prof. Luis Hernandez Garcia

fMRI Lab & EECS
University of Michigan, Ann Arbor
Email: tahong@umich.edu



Jeff



Luis



Outline

Generalizing Nesterov's Scheme

Motivation

General Nest. Acc.

Numerical Tests

Magical High-order Methods → CS MRI Reco.

Problem Formulation

Our Suggestion

Numerical Results



Generalizing Nesterov's Scheme

Motivation

General Nest. Acc.

Numerical Tests

Magical High-order Methods → CS MRI Reco.

Problem Formulation

Our Suggestion

Numerical Results



Optimization Problems

$\min_{\mathbf{x} \in \mathbb{R}^N} f(\mathbf{x})$: f smooth & convex & L Lip. Const.



Optimization Problems

$\min_{\mathbf{x} \in \mathbb{R}^N} f(\mathbf{x})$: f smooth & convex & L Lip. Const.

Solver: plain \rightarrow gradient descent $O(\frac{1}{k})$ $\mathbf{x}_{k+1} = \mathbf{x}_k - \frac{1}{L} \nabla_{\mathbf{x}} f(\mathbf{x}_k)$



Optimization Problems

$\min_{\mathbf{x} \in \mathbb{R}^N} f(\mathbf{x})$: f smooth & convex & L Lip. Const.

Solver: plain \rightarrow gradient descent $O(\frac{1}{k})$ $\mathbf{x}_{k+1} = \mathbf{x}_k - \frac{1}{L} \nabla_{\mathbf{x}} f(\mathbf{x}_k)$

Nest. Acc. $O(\frac{1}{k^2})$ $\mathbf{x}_{k+1} = \mathbf{v}_k - \frac{1}{L} \nabla_{\mathbf{x}} f(\mathbf{v}_k)$, $\mathbf{v}_{k+1} = t_k^1 \mathbf{x}_k + t_k^2 \mathbf{x}_{k+1}$



Optimization Problems

$\min_{\mathbf{x} \in \mathbb{R}^N} f(\mathbf{x})$: f smooth & convex & L Lip. Const.

Solver: plain \rightarrow gradient descent $O(\frac{1}{k})$ $\mathbf{x}_{k+1} = \mathbf{x}_k - \frac{1}{L} \nabla_{\mathbf{x}} f(\mathbf{x}_k)$

Nest. Acc. $O(\frac{1}{k^2})$ $\mathbf{x}_{k+1} = \mathbf{v}_k - \frac{1}{L} \nabla_{\mathbf{x}} f(\mathbf{v}_k)$, $\mathbf{v}_{k+1} = t_k^1 \mathbf{x}_k + t_k^2 \mathbf{x}_{k+1}$
 \rightarrow Cubic reg. (Nest. 06 MP)



Optimization Problems

$\min_{\mathbf{x} \in \mathbb{R}^N} f(\mathbf{x})$: f smooth & convex & L Lip. Const.

Solver: plain \rightarrow gradient descent $O(\frac{1}{k})$ $\mathbf{x}_{k+1} = \mathbf{x}_k - \frac{1}{L} \nabla_{\mathbf{x}} f(\mathbf{x}_k)$

Nest. Acc. $O(\frac{1}{k^2})$ $\mathbf{x}_{k+1} = \mathbf{v}_k - \frac{1}{L} \nabla_{\mathbf{x}} f(\mathbf{v}_k)$, $\mathbf{v}_{k+1} = t_k^1 \mathbf{x}_k + t_k^2 \mathbf{x}_{k+1}$
 \rightarrow Cubic reg. (Nest. 06 MP) \rightarrow Acc. version (Nest. 08 MP)



Optimization Problems

$\min_{\mathbf{x} \in \mathbb{R}^N} f(\mathbf{x})$: f smooth & convex & L Lip. Const.

Solver: plain \rightarrow gradient descent $O(\frac{1}{k})$ $\mathbf{x}_{k+1} = \mathbf{x}_k - \frac{1}{L} \nabla_{\mathbf{x}} f(\mathbf{x}_k)$

Nest. Acc. $O(\frac{1}{k^2})$ $\mathbf{x}_{k+1} = \mathbf{v}_k - \frac{1}{L} \nabla_{\mathbf{x}} f(\mathbf{v}_k)$, $\mathbf{v}_{k+1} = t_k^1 \mathbf{x}_k + t_k^2 \mathbf{x}_{k+1}$
 \rightarrow Cubic reg. (Nest. 06 MP) \rightarrow Acc. version (Nest. 08 MP)

$\min_{\mathbf{x} \in \mathbb{R}^N} f(\mathbf{x}) + g(\mathbf{x})$: g nonsmooth & convex & L Lip. Const. of $f(\mathbf{x})$



Optimization Problems

$$\min_{\mathbf{x} \in \mathbb{R}^N} f(\mathbf{x}): f \text{ smooth \& convex \& } L \text{ Lip. Const.}$$

Solver: plain \rightarrow gradient descent $O(\frac{1}{k}) \quad \mathbf{x}_{k+1} = \mathbf{x}_k - \frac{1}{L} \nabla_{\mathbf{x}} f(\mathbf{x}_k)$

Nest. Acc. $O(\frac{1}{k^2}) \quad \mathbf{x}_{k+1} = \mathbf{v}_k - \frac{1}{L} \nabla_{\mathbf{x}} f(\mathbf{v}_k), \quad \mathbf{v}_{k+1} = t_k^1 \mathbf{x}_k + t_k^2 \mathbf{x}_{k+1}$
 \rightarrow Cubic reg. (Nest. 06 MP) \rightarrow Acc. version (Nest. 08 MP)

$$\min_{\mathbf{x} \in \mathbb{R}^N} f(\mathbf{x}) + g(\mathbf{x}): g \text{ nonsmooth \& convex \& } L \text{ Lip. Const. of } f(\mathbf{x})$$

Define $\text{Prox}_g(\mathbf{x}) = \operatorname{argmin}_{\mathbf{v}} \frac{1}{2} \|\mathbf{v} - \mathbf{x}\|_2^2 + g(\mathbf{v})$

Solver: plain \rightarrow proximal gradient method

$$\mathbf{x}_{k+1} = \text{Prox}_g\left(\mathbf{x}_k - \frac{1}{L} \nabla_{\mathbf{x}} f(\mathbf{x}_k)\right)$$



Optimization Problems

$$\min_{\mathbf{x} \in \mathbb{R}^N} f(\mathbf{x}): f \text{ smooth \& convex \& } L \text{ Lip. Const.}$$

Solver: plain \rightarrow gradient descent $O(\frac{1}{k}) \quad \mathbf{x}_{k+1} = \mathbf{x}_k - \frac{1}{L} \nabla_{\mathbf{x}} f(\mathbf{x}_k)$

Nest. Acc. $O(\frac{1}{k^2}) \quad \mathbf{x}_{k+1} = \mathbf{v}_k - \frac{1}{L} \nabla_{\mathbf{x}} f(\mathbf{v}_k), \quad \mathbf{v}_{k+1} = t_k^1 \mathbf{x}_k + t_k^2 \mathbf{x}_{k+1}$
 \rightarrow Cubic reg. (Nest. 06 MP) \rightarrow Acc. version (Nest. 08 MP)

$$\min_{\mathbf{x} \in \mathbb{R}^N} f(\mathbf{x}) + g(\mathbf{x}): g \text{ nonsmooth \& convex \& } L \text{ Lip. Const. of } f(\mathbf{x})$$

Define $\text{Prox}_g(\mathbf{x}) = \operatorname{argmin}_{\mathbf{v}} \frac{1}{2} \|\mathbf{v} - \mathbf{x}\|_2^2 + g(\mathbf{v})$

Solver: plain \rightarrow proximal gradient method

$$\mathbf{x}_{k+1} = \text{Prox}_g\left(\mathbf{x}_k - \frac{1}{L} \nabla_{\mathbf{x}} f(\mathbf{x}_k)\right)$$

Nest. Acc. (APM or FISTA)

$$\mathbf{x}_{k+1} = \text{Prox}_g\left(\mathbf{v}_k - \frac{1}{L} \nabla_{\mathbf{x}} f(\mathbf{v}_k)\right), \quad \mathbf{v}_{k+1} = t_k^1 \mathbf{x}_k + t_k^2 \mathbf{x}_{k+1}$$



Optimization Problems

$$\min_{\mathbf{x} \in \mathbb{R}^N} f(\mathbf{x}): f \text{ smooth \& convex \& } L \text{ Lip. Const.}$$

Solver: plain \rightarrow gradient descent $O(\frac{1}{k})$ $\mathbf{x}_{k+1} = \mathbf{x}_k - \frac{1}{L} \nabla_{\mathbf{x}} f(\mathbf{x}_k)$

Nest. Acc. $O(\frac{1}{k^2})$ $\mathbf{x}_{k+1} = \mathbf{v}_k - \frac{1}{L} \nabla_{\mathbf{x}} f(\mathbf{v}_k), \mathbf{v}_{k+1} = t_k^1 \mathbf{x}_k + t_k^2 \mathbf{x}_{k+1}$
 \rightarrow Cubic reg. (Nest. 06 MP) \rightarrow Acc. version (Nest. 08 MP)

$$\min_{\mathbf{x} \in \mathbb{R}^N} f(\mathbf{x}) + g(\mathbf{x}): g \text{ nonsmooth \& convex \& } L \text{ Lip. Const. of } f(\mathbf{x})$$

Define $\text{Prox}_g(\mathbf{x}) = \operatorname{argmin}_{\mathbf{v}} \frac{1}{2} \|\mathbf{v} - \mathbf{x}\|_2^2 + g(\mathbf{v})$

Solver: plain \rightarrow proximal gradient method

$$\mathbf{x}_{k+1} = \text{Prox}_g\left(\mathbf{x}_k - \frac{1}{L} \nabla_{\mathbf{x}} f(\mathbf{x}_k)\right)$$

Nest. Acc. (APM or FISTA)

$$\mathbf{x}_{k+1} = \text{Prox}_g\left(\mathbf{v}_k - \frac{1}{L} \nabla_{\mathbf{x}} f(\mathbf{v}_k)\right), \mathbf{v}_{k+1} = t_k^1 \mathbf{x}_k + t_k^2 \mathbf{x}_{k+1}$$

$$\min_{\mathbf{x}_1, \mathbf{x}_2, \dots} f(\mathbf{x}_1, \mathbf{x}_2, \dots): \text{block coordinate descent and acc. version}$$

Optimization Problems

$$\min_{\mathbf{x} \in \mathbb{R}^N} f(\mathbf{x}): f \text{ smooth \& convex \& } L \text{ Lip. Const.}$$

Solver: plain \rightarrow gradient descent $O(\frac{1}{k}) \quad \mathbf{x}_{k+1} = \mathbf{x}_k - \frac{1}{L} \nabla_{\mathbf{x}} f(\mathbf{x}_k)$

Nest. Acc. $O(\frac{1}{k^2}) \quad \mathbf{x}_{k+1} = \mathbf{v}_k - \frac{1}{L} \nabla_{\mathbf{x}} f(\mathbf{v}_k), \quad \mathbf{v}_{k+1} = t_k^1 \mathbf{x}_k + t_k^2 \mathbf{x}_{k+1}$
 \rightarrow Cubic reg. (Nest. 06 MP) \rightarrow Acc. version (Nest. 08 MP)

$$\min_{\mathbf{x} \in \mathbb{R}^N} f(\mathbf{x}) + g(\mathbf{x}): g \text{ nonsmooth \& convex \& } L \text{ Lip. Const. of } f(\mathbf{x})$$

Define $\text{Prox}_g(\mathbf{x}) = \operatorname{argmin}_{\mathbf{v}} \frac{1}{2} \|\mathbf{v} - \mathbf{x}\|_2^2 + g(\mathbf{v})$

Solver: plain \rightarrow proximal gradient method

$$\mathbf{x}_{k+1} = \text{Prox}_g\left(\mathbf{x}_k - \frac{1}{L} \nabla_{\mathbf{x}} f(\mathbf{x}_k)\right)$$

Nest. Acc. (APM or FISTA)

$$\mathbf{x}_{k+1} = \text{Prox}_g\left(\mathbf{v}_k - \frac{1}{L} \nabla_{\mathbf{x}} f(\mathbf{v}_k)\right), \quad \mathbf{v}_{k+1} = t_k^1 \mathbf{x}_k + t_k^2 \mathbf{x}_{k+1}$$

$$\min_{\mathbf{x}_1, \mathbf{x}_2, \dots} f(\mathbf{x}_1, \mathbf{x}_2, \dots): \text{block coordinate descent and acc. version}$$

Many others...



Generalizing Nesterov's Acceleration

Can we accelerate an abstract solver by Nesterov's scheme?



Generalizing Nesterov's Acceleration

Can we accelerate an abstract solver by Nesterov's scheme?

Looks possible and we have some answers



Generalizing Nesterov's Acceleration

Can we accelerate an abstract solver by Nesterov's scheme?

Looks possible and we have some answers

DOI: 10.1002/nla.2417

RESEARCH ARTICLE

WILEY

On adapting Nesterov's scheme to accelerate iterative methods for linear problems

Tao Hong[✉] | Irad Yavneh

Department of Computer Science,
Technion - Israel Institute of Technology,
Haifa, Israel

Abstract

Nesterov's well-known scheme for accelerating gradient descent in convex opti-

T. Hong and I. Yavneh, NLAA, 2022.



Generalizing Nesterov's Scheme

Motivation

General Nest. Acc.

Numerical Tests

Magical High-order Methods → CS MRI Reco.

Problem Formulation

Our Suggestion

Numerical Results



Problem Formulation

We consider ($\mathbf{A} \succeq 0$)

$$\min_{\mathbf{x} \in \mathbb{R}^N} \frac{1}{2} \mathbf{x}^T \mathbf{A} \mathbf{x} - \mathbf{f}^T \mathbf{x} \Leftrightarrow \mathbf{A} \mathbf{x} = \mathbf{f}$$



Problem Formulation

We consider ($\mathbf{A} \succeq 0$)

$$\min_{\mathbf{x} \in \mathbb{R}^N} \frac{1}{2} \mathbf{x}^T \mathbf{A} \mathbf{x} - \mathbf{f}^T \mathbf{x} \Leftrightarrow \mathbf{A} \mathbf{x} = \mathbf{f}$$

Nest. formulation:

$$\begin{aligned}\mathbf{x}_{k+1} &= \mathbf{Bv}_k + \text{Constant} \\ \mathbf{v}_{k+1} &= \mathbf{x}_{k+1} + c_k(\mathbf{x}_{k+1} - \mathbf{x}_k)\end{aligned}$$



Problem Formulation

We consider ($\mathbf{A} \succeq 0$)

$$\min_{\mathbf{x} \in \mathbb{R}^N} \frac{1}{2} \mathbf{x}^T \mathbf{A} \mathbf{x} - \mathbf{f}^T \mathbf{x} \Leftrightarrow \mathbf{A} \mathbf{x} = \mathbf{f}$$

Nest. formulation:

$$\begin{aligned}\mathbf{x}_{k+1} &= \mathbf{Bv}_k + \text{Constant} \\ \mathbf{v}_{k+1} &= \mathbf{x}_{k+1} + c_k(\mathbf{x}_{k+1} - \mathbf{x}_k)\end{aligned}$$

\mathbf{B} iteration matrix, e.g., $\mathbf{B} = \mathbf{I} - a_k \mathbf{A}$ (gradient descent)



Problem Formulation

We consider ($\mathbf{A} \succeq 0$)

$$\min_{\mathbf{x} \in \mathbb{R}^N} \frac{1}{2} \mathbf{x}^T \mathbf{A} \mathbf{x} - \mathbf{f}^T \mathbf{x} \Leftrightarrow \mathbf{A} \mathbf{x} = \mathbf{f}$$

Nest. formulation:

$$\begin{aligned}\mathbf{x}_{k+1} &= \mathbf{Bv}_k + \text{Constant} \\ \mathbf{v}_{k+1} &= \mathbf{x}_{k+1} + c_k(\mathbf{x}_{k+1} - \mathbf{x}_k)\end{aligned}$$

\mathbf{B} iteration matrix, e.g., $\mathbf{B} = \mathbf{I} - a_k \mathbf{A}$ (gradient descent)

Next question: how to choose c_k & can we get acceleration?

Problem Formulation

We consider ($\mathbf{A} \succeq 0$)

$$\min_{\mathbf{x} \in \mathbb{R}^N} \frac{1}{2} \mathbf{x}^T \mathbf{A} \mathbf{x} - \mathbf{f}^T \mathbf{x} \Leftrightarrow \mathbf{A} \mathbf{x} = \mathbf{f}$$

Nest. formulation:

$$\begin{aligned}\mathbf{x}_{k+1} &= \mathbf{Bv}_k + \text{Constant} \\ \mathbf{v}_{k+1} &= \mathbf{x}_{k+1} + c_k(\mathbf{x}_{k+1} - \mathbf{x}_k)\end{aligned}$$

\mathbf{B} iteration matrix, e.g., $\mathbf{B} = \mathbf{I} - a_k \mathbf{A}$ (gradient descent)

Next question: how to choose c_k & can we get acceleration?

The answer is positive at least for some \mathbf{B}



Optimal Acceleration → The Choice of c_k

Assumption: \mathbf{B} only has real eigenvalues.



Optimal Acceleration → The Choice of c_k

Assumption: \mathbf{B} only has real eigenvalues.

$$-1 < b_1 \leq \dots \leq b_N < 1 \text{ (eigenvalue } \mathbf{B} \text{) \& } c_{cr}(b) = \frac{1-\sqrt{1-b}}{1+\sqrt{1-b}}$$

Theorem

Let $-1 < b_1 \leq b_N < 1 \rightarrow c^* = c_{cr}(g(b_1, b_N))$

$$g(b_1, b_N) = \begin{cases} b_N, & b_N \geq -3b_1, \\ -\frac{8b_N b_1 (b_1 + b_N)}{(b_1 - b_N)^2}, & -\frac{1}{3}b_1 < b_N < -3b_1, \\ b_1, & b_N \leq -\frac{1}{3}b_1, \end{cases}$$

yielding conv. factor

$$r^* = \begin{cases} 1 - \sqrt{1 - b_N}, & b_N \geq -3b_1, \\ r(c^*, b_1) = r(c^*, b_N), & -\frac{1}{3}b_1 < b_N < -3b_1, \\ \sqrt{1 - b_1} - 1, & b_N \leq -\frac{1}{3}b_1. \end{cases}$$

$$r(c, b) = \frac{1}{2} \left| (1+c)b + \operatorname{sgn}(b) \sqrt{(1+c)^2 b^2 - 4cb} \right|$$

Remarks

- $c_k \rightarrow c$



Remarks

- $c_k \rightarrow c$
- Nest. can converge even for some divergent \mathbf{B} (whose spectral radii are larger than 1). Relax assumption from
 $-1 < b_1 \leq b_N < 1$ to $-3 < b_1 \leq b_N < 1$.



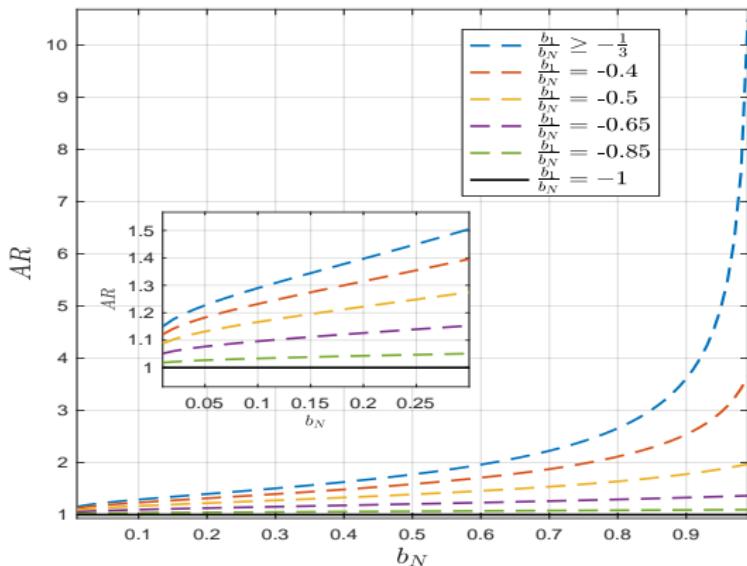
Remarks

- $c_k \rightarrow c$
- Nest. can converge even for some divergent \mathbf{B} (whose spectral radii are larger than 1). Relax assumption from $-1 < b_1 \leq b_N < 1$ to $-3 < b_1 \leq b_N < 1$.
- Optimal step-size $\frac{4}{3L}$



Remarks

- $c_k \rightarrow c$
- Nest. can converge even for some divergent \mathbf{B} (whose spectral radii are larger than 1). Relax assumption from $-1 < b_1 \leq b_N < 1$ to $-3 < b_1 \leq b_N < 1$.
- Optimal step-size $\frac{4}{3L}$
- $AR = \frac{\log r^*}{\log b_N}$ b_N conv. factor (plain)



B Complex Eigenvalues

Denote by $b^c = \bar{b}^c e^{j\theta}$

j : imaginary unit; \bar{b}^c : modulus; $\theta \in (-\pi, \pi]$: argument



B Complex Eigenvalues

Denote by $b^c = \bar{b}^c e^{j\theta}$

j : imaginary unit; \bar{b}^c : modulus; $\theta \in (-\pi, \pi]$: argument

Theorem

In addition to $-1 < b_1 \leq \dots \leq b_N < 1$ of **B**,

B also has complex eigenvalues.

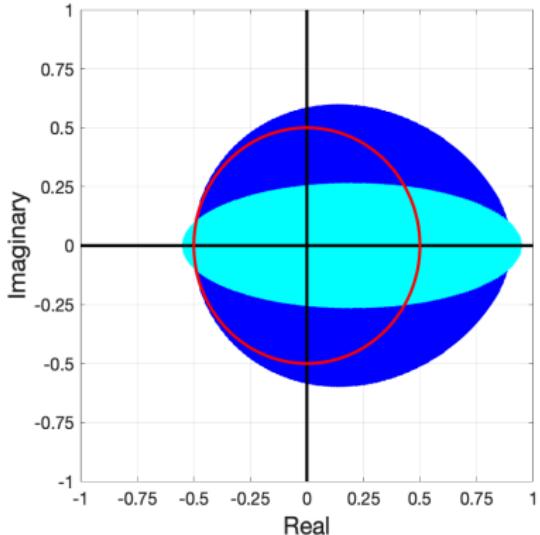
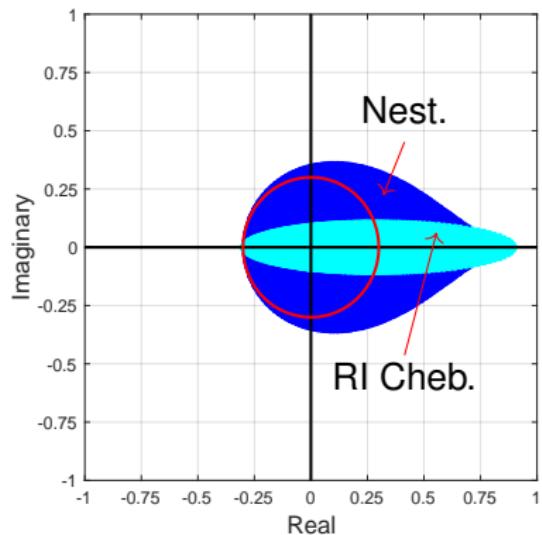
Then, c^* and r^* remain valid if the modulus of all complex eigenvalues satisfies

$$\bar{b}^c \leq \begin{cases} \frac{1}{3}b_N & b_N \geq -3b_1 \\ \min(|b_1|, |b_N|) & -\frac{1}{3}b_1 < b_N < -3b_1 \\ -\frac{1}{3}b_1 & b_N \leq -\frac{1}{3}b_1 \end{cases}$$



Compare with RI Chebyshev Acc.

Left: $b_1 = -0.3$ and $b_N = 0.9$; Right: $b_1 = -0.5$ and $b_N = 0.9$.



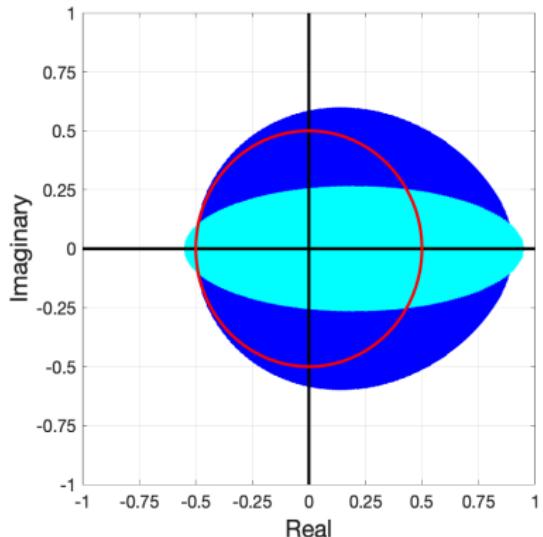
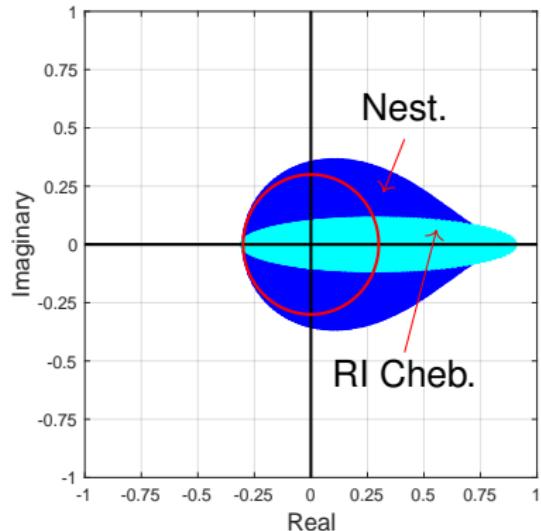
Cheb. Acc.

$$\mathbf{x}_1 = \gamma(\mathbf{Bx}_0 + \text{Constant}) + (1 - \gamma)\mathbf{x}_0,$$

$$\mathbf{x}_{k+1} = \beta_{k+1} \{ \gamma(\mathbf{Bx}_k + \text{Constant}) + (1 - \gamma)\mathbf{x}_k \} + (1 - \beta_{k+1})\mathbf{x}_{k-1}$$

Compare with RI Chebyshev Acc.

Left: $b_1 = -0.3$ and $b_N = 0.9$; Right: $b_1 = -0.5$ and $b_N = 0.9$.



Nest. and RI Cheb. are different

Generalizing Nesterov's Scheme

Motivation

General Nest. Acc.

Numerical Tests

Magical High-order Methods → CS MRI Reco.

Problem Formulation

Our Suggestion

Numerical Results



Numerical Tests

Consider a diffusion equation:

$$-\nabla(\sigma(x,y)\nabla u(x,y)) = f(x,y)$$



Numerical Tests

Consider a diffusion equation:

$$-\nabla(\sigma(x,y)\nabla u(x,y)) = f(x,y)$$

Test on three examples:

- 1) $\sigma(x,y) = 1$;
- 2) $\sigma(x,y)$ sampled from log-normal distribution;
- 3) $\sigma(x,y)$ sampled from uniform distribution;



Numerical Tests

Consider a diffusion equation:

$$-\nabla(\sigma(x,y)\nabla u(x,y)) = f(x,y)$$

Test on three examples:

- 1) $\sigma(x,y) = 1$;
- 2) $\sigma(x,y)$ sampled from log-normal distribution;
- 3) $\sigma(x,y)$ sampled from uniform distribution;

Discretized on a 1024×1024 uniform grid $\rightarrow \mathbf{A}\mathbf{u} = \mathbf{f}$

Residual vector: $\mathbf{r}_k = \mathbf{f} - \mathbf{A}\mathbf{u}_k$; Residual norm: $\|\mathbf{r}_k\|_2$;

Conv. factor: $\lim_{k \rightarrow \infty} \frac{\|\mathbf{r}_{k+1}\|_2}{\|\mathbf{r}_k\|_2}$



Numerical Tests

Consider a diffusion equation:

$$-\nabla(\sigma(x,y)\nabla u(x,y)) = f(x,y)$$

Test on three examples:

- 1) $\sigma(x,y) = 1$;
- 2) $\sigma(x,y)$ sampled from log-normal distribution;
- 3) $\sigma(x,y)$ sampled from uniform distribution;

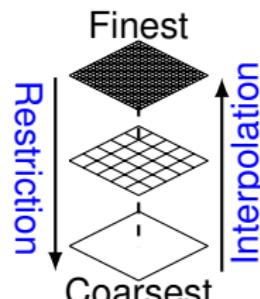
Discretized on a 1024×1024 uniform grid $\rightarrow \mathbf{A}\mathbf{u} = \mathbf{f}$

Residual vector: $\mathbf{r}_k = \mathbf{f} - \mathbf{A}\mathbf{u}_k$; Residual norm: $\|\mathbf{r}_k\|_2$;

Conv. factor: $\lim_{k \rightarrow \infty} \frac{\|\mathbf{r}_{k+1}\|_2}{\|\mathbf{r}_k\|_2}$

B: multigrid methods

Local relaxation
Restriction
Interpolation



Test 1: $\sigma(x, y) = 1$

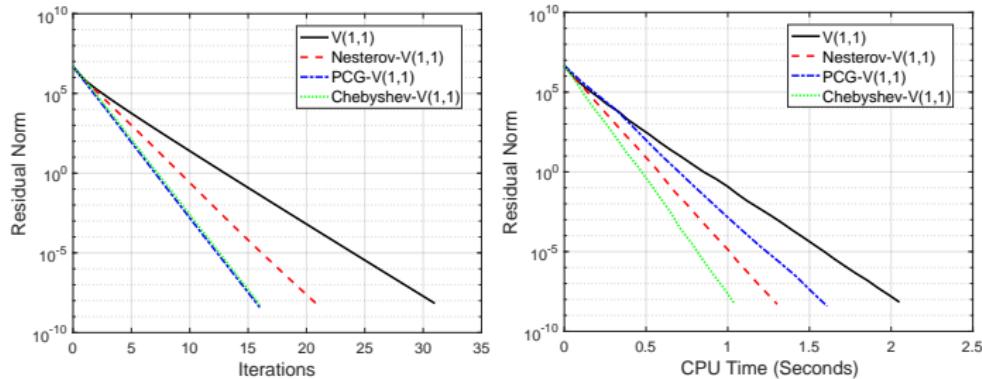
The Poisson problem

Damped Jacobi relaxation $\rightarrow \mathbf{B}$ only has real eigenvalues

Test 1: $\sigma(x, y) = 1$

The Poisson problem

Damped Jacobi relaxation $\rightarrow \mathbf{B}$ only has real eigenvalues



PCG: preconditioned conjugate gradient

$V(1, 1)$: multigrid method

Test 1: $\sigma(x, y) = 1$

The Poisson problem

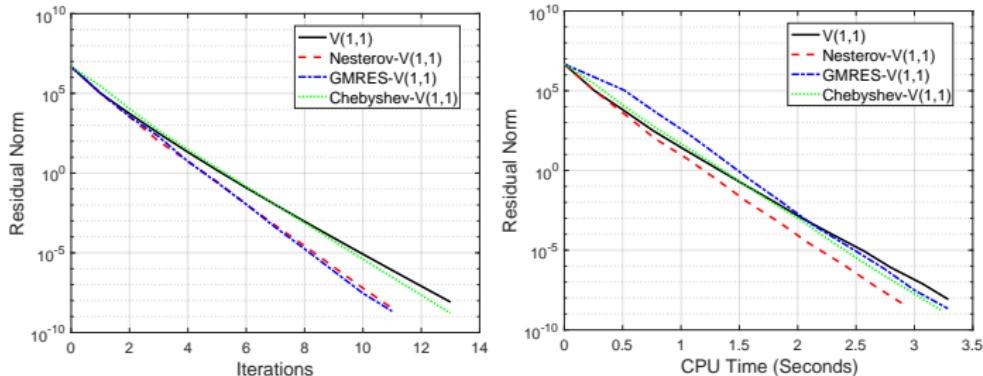
Red-Black relaxation $\rightarrow \mathbf{B}$ contains complex & real eigenvalues
PCG: invalid



Test 1: $\sigma(x, y) = 1$

The Poisson problem

Red-Black relaxation $\rightarrow \mathbf{B}$ contains complex & real eigenvalues
PCG: invalid



Test 2: $\sigma(x, y)$ log-normal distribution

B: Black Box multigrid method (complex)

Cheb. Acc.: invalid

$V(1, 1)$: multigrid method

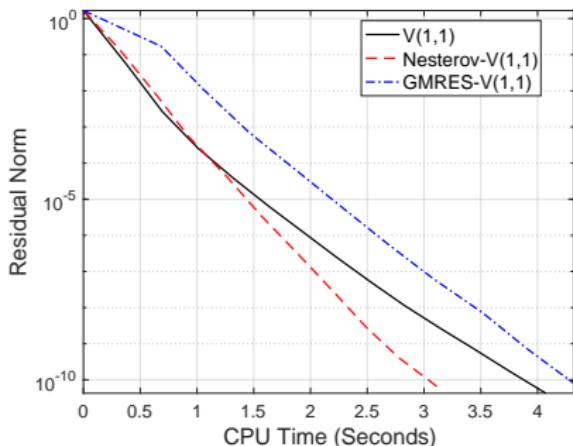
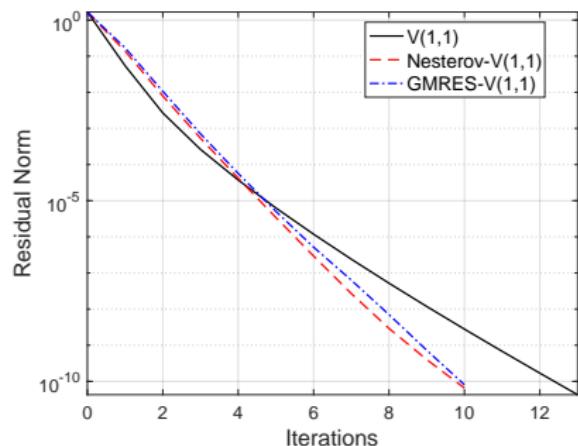


Test 2: $\sigma(x, y)$ log-normal distribution

B: Black Box multigrid method (complex)

Cheb. Acc.: invalid

$V(1, 1)$: multigrid method



Test 3: $\sigma(x, y)$ normal distribution

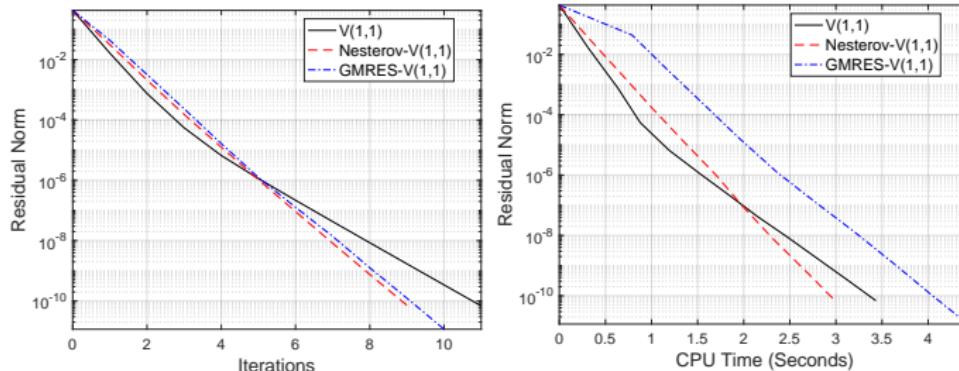
B: Black Box multigrid method (complex)

Cheb. Acc.: invalid

Test 3: $\sigma(x, y)$ normal distribution

B: Black Box multigrid method (complex)

Cheb. Acc.: invalid



Generalizing Nesterov's Scheme

Motivation

General Nest. Acc.

Numerical Tests

Magical High-order Methods → CS MRI Reco.

Problem Formulation

Our Suggestion

Numerical Results



MRI: Magnetic Resonance Imaging

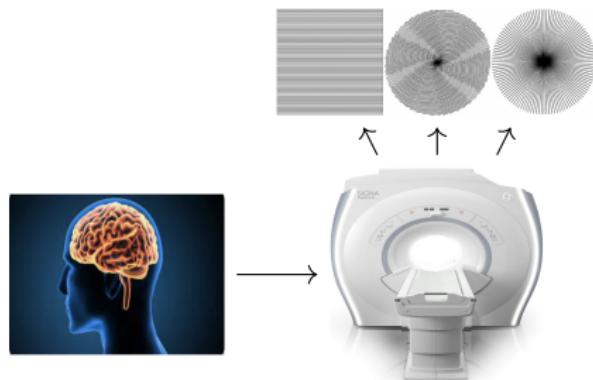


MRI: Magnetic Resonance Imaging

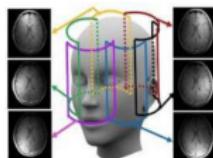


\mathbf{x}^{true}

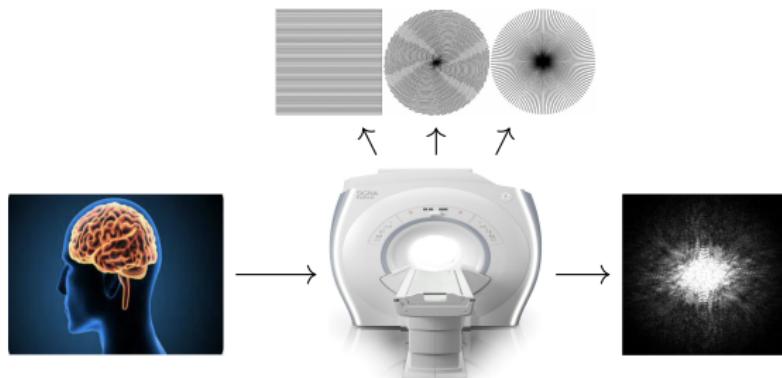
MRI: Magnetic Resonance Imaging



x^{true} forward process **A**

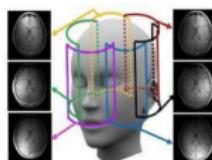


MRI: Magnetic Resonance Imaging

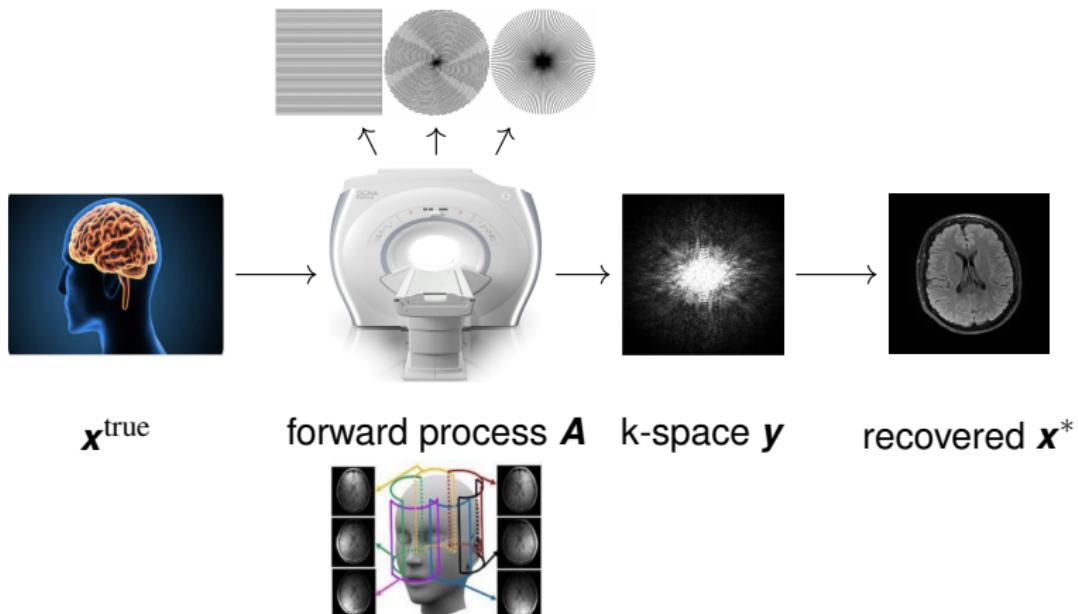


x^{true}

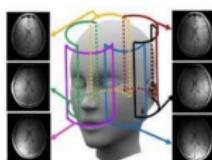
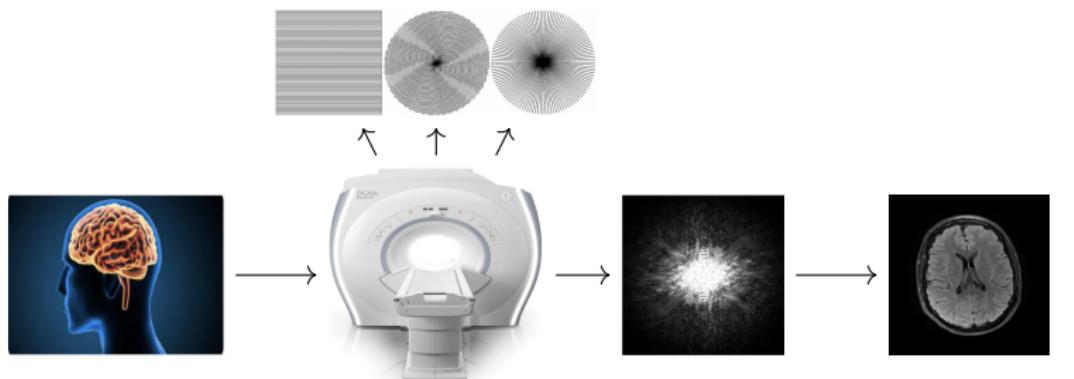
forward process \mathbf{A} k-space y



MRI: Magnetic Resonance Imaging



MRI: Magnetic Resonance Imaging



$$y = \mathbf{A}x^{\text{true}} + \text{noise}$$

Problem Formulation $\mathbf{y} = \mathbf{A}\mathbf{x}^{\text{true}} + \text{noise}$

Fully sample $\mathbf{x}^* = \text{IFFT}(\mathbf{y})$



Problem Formulation $\mathbf{y} = \mathbf{A}\mathbf{x}^{\text{true}} + \text{noise}$

Fully sample $\mathbf{x}^* = \text{IFFT}(\mathbf{y})$

Reco. in CS MRI \rightarrow Composite Optimization Problem

$$\min_{\mathbf{x} \in \mathbb{C}^N} \underbrace{\frac{1}{2} \|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2^2}_{f(\mathbf{x})} + R(\mathbf{x})$$



Problem Formulation $\mathbf{y} = \mathbf{A}\mathbf{x}^{\text{true}} + \text{noise}$

Fully sample $\mathbf{x}^* = \text{IFFT}(\mathbf{y})$

Reco. in CS MRI \rightarrow Composite Optimization Problem

$$\min_{\mathbf{x} \in \mathbb{C}^N} \underbrace{\frac{1}{2} \|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2^2}_{f(\mathbf{x})} + R(\mathbf{x})$$

$$\mathbf{A} = [\mathbf{A}_i, \dots], \mathbf{A}_i = \mathbf{PFS}_i$$



Problem Formulation $\mathbf{y} = \mathbf{Ax}^{\text{true}} + \text{noise}$

Fully sample $\mathbf{x}^* = \text{IFFT}(\mathbf{y})$

Reco. in CS MRI \rightarrow Composite Optimization Problem

$$\min_{\mathbf{x} \in \mathbb{C}^N} \underbrace{\frac{1}{2} \|\mathbf{Ax} - \mathbf{y}\|_2^2}_{f(\mathbf{x})} + R(\mathbf{x})$$

$$\mathbf{A} = [\mathbf{A}_i, \dots], \mathbf{A}_i = \mathbf{PFS}_i$$

\mathbf{P} : downsample; \mathbf{F} : (nonuniform) FFT; \mathbf{S}_i : sensitivity mapping



Problem Formulation $\mathbf{y} = \mathbf{A}\mathbf{x}^{\text{true}} + \text{noise}$

Fully sample $\mathbf{x}^* = \text{IFFT}(\mathbf{y})$

Reco. in CS MRI \rightarrow Composite Optimization Problem

$$\min_{\mathbf{x} \in \mathbb{C}^N} \underbrace{\frac{1}{2} \|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2^2}_{f(\mathbf{x})} + R(\mathbf{x})$$

$$\mathbf{A} = [\mathbf{A}_i, \dots], \mathbf{A}_i = \mathbf{PFS}_i$$

\mathbf{P} : downsample; \mathbf{F} : (nonuniform) FFT; \mathbf{S}_i : sensitivity mapping

$R(\mathbf{x})$: regularizer

We consider wavelet, TV, or both.

Wavelet and TV

Wavelet Reco. $\lambda > 0$

$$\mathbf{x}^* = \operatorname{argmin}_{\mathbf{x} \in \mathbb{C}^M} \frac{1}{2} \|\mathbf{A}\mathbf{T}^{-1}\mathbf{x} - \mathbf{y}\|_2^2 + \lambda \|\mathbf{x}\|_1$$

\mathbf{T} : wavelet transform, image $\mathbf{T}^{-1}\mathbf{x}^*$



Wavelet and TV

Wavelet Reco. $\lambda > 0$

$$\mathbf{x}^* = \operatorname{argmin}_{\mathbf{x} \in \mathbb{C}^M} \frac{1}{2} \|\mathbf{A}\mathbf{T}^{-1}\mathbf{x} - \mathbf{y}\|_2^2 + \lambda \|\mathbf{x}\|_1$$

\mathbf{T} : wavelet transform, image $\mathbf{T}^{-1}\mathbf{x}^*$

TV Reco.

$$\mathbf{x}^* = \operatorname{argmin}_{\mathbf{x} \in \mathbb{C}^N} \frac{1}{2} \|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2^2 + \lambda \operatorname{TV}(\mathbf{x})$$

image \mathbf{x}^*



Wavelet and TV

Wavelet Reco. $\lambda > 0$

$$\mathbf{x}^* = \operatorname{argmin}_{\mathbf{x} \in \mathbb{C}^M} \frac{1}{2} \|\mathbf{A}\mathbf{T}^{-1}\mathbf{x} - \mathbf{y}\|_2^2 + \lambda \|\mathbf{x}\|_1$$

\mathbf{T} : wavelet transform, image $\mathbf{T}^{-1}\mathbf{x}^*$

TV Reco.

$$\mathbf{x}^* = \operatorname{argmin}_{\mathbf{x} \in \mathbb{C}^N} \frac{1}{2} \|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2^2 + \lambda \operatorname{TV}(\mathbf{x})$$

image \mathbf{x}^*

Wavelet and TV Reco.

$$\mathbf{x}^* = \operatorname{argmin}_{\mathbf{x} \in \mathbb{C}^N} \frac{1}{2} \|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2^2 + \lambda\alpha \|\mathbf{T}\mathbf{x}\|_1 + \lambda(1-\alpha) \operatorname{TV}(\mathbf{x}), \quad \alpha \in (0, 1)$$

image \mathbf{x}^*

$$\text{Solvers } \min_{\mathbf{x} \in \mathbb{C}^N} \underbrace{\frac{1}{2} \|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2^2}_{f(\mathbf{x})} + R(\mathbf{x})$$

$$\text{Solvers } \min_{\mathbf{x} \in \mathbb{C}^N} \underbrace{\frac{1}{2} \|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2^2}_{f(\mathbf{x})} + R(\mathbf{x})$$

$$\text{Wavelet Reco. } \min_{\mathbf{x} \in \mathbb{C}^M} \frac{1}{2} \|\mathbf{A}\mathbf{T}^{-1}\mathbf{x} - \mathbf{y}\|_2^2 + \lambda \|\mathbf{x}\|_1 \rightarrow \text{FISTA (APM)}$$



$$\text{Solvers } \min_{\mathbf{x} \in \mathbb{C}^N} \underbrace{\frac{1}{2} \|\mathbf{Ax} - \mathbf{y}\|_2^2}_{f(\mathbf{x})} + R(\mathbf{x})$$

Wavelet Reco. $\min_{\mathbf{x} \in \mathbb{C}^M} \frac{1}{2} \|\mathbf{AT}^{-1}\mathbf{x} - \mathbf{y}\|_2^2 + \lambda \|\mathbf{x}\|_1 \rightarrow \text{FISTA (APM)}$

TV Reco.

$$\min_{\mathbf{x} \in \mathbb{C}^N} \frac{1}{2} \|\mathbf{Ax} - \mathbf{y}\|_2^2 + \lambda \underbrace{\|\mathbf{Dx}\|_1}_{\text{TV}(\mathbf{x})} \rightarrow \text{APM iter. proximal mapping}$$



Solvers $\min_{\mathbf{x} \in \mathbb{C}^N} \underbrace{\frac{1}{2} \|\mathbf{Ax} - \mathbf{y}\|_2^2}_{f(\mathbf{x})} + R(\mathbf{x})$

Wavelet Reco. $\min_{\mathbf{x} \in \mathbb{C}^M} \frac{1}{2} \|\mathbf{AT}^{-1}\mathbf{x} - \mathbf{y}\|_2^2 + \lambda \|\mathbf{x}\|_1 \rightarrow \text{FISTA (APM)}$

TV Reco.

$$\min_{\mathbf{x} \in \mathbb{C}^N} \frac{1}{2} \|\mathbf{Ax} - \mathbf{y}\|_2^2 + \lambda \underbrace{\|\mathbf{Dx}\|_1}_{\text{TV}(\mathbf{x})} \rightarrow \text{APM iter. proximal mapping}$$

Wavelet and TV Reco.

$$\min_{\mathbf{x} \in \mathbb{C}^N} \frac{1}{2} \|\mathbf{Ax} - \mathbf{y}\|_2^2 + \lambda \alpha \|\mathbf{Tx}\|_1 + \lambda(1 - \alpha) \text{TV}(\mathbf{x}) \rightarrow \text{APM&ADMM}$$

Solvers $\min_{\mathbf{x} \in \mathbb{C}^N} \underbrace{\frac{1}{2} \|\mathbf{Ax} - \mathbf{y}\|_2^2}_{f(\mathbf{x})} + R(\mathbf{x})$

Wavelet Reco. $\min_{\mathbf{x} \in \mathbb{C}^M} \frac{1}{2} \|\mathbf{AT}^{-1}\mathbf{x} - \mathbf{y}\|_2^2 + \lambda \|\mathbf{x}\|_1 \rightarrow \text{FISTA (APM)}$

TV Reco.

$$\min_{\mathbf{x} \in \mathbb{C}^N} \frac{1}{2} \|\mathbf{Ax} - \mathbf{y}\|_2^2 + \lambda \underbrace{\|\mathbf{Dx}\|_1}_{\text{TV}(\mathbf{x})} \rightarrow \text{APM iter. proximal mapping}$$

Wavelet and TV Reco.

$$\min_{\mathbf{x} \in \mathbb{C}^N} \frac{1}{2} \|\mathbf{Ax} - \mathbf{y}\|_2^2 + \lambda \alpha \|\mathbf{Tx}\|_1 + \lambda(1 - \alpha) \text{TV}(\mathbf{x}) \rightarrow \text{APM&ADMM}$$

Proximal gradient descent:

$$\mathbf{x}_{k+1} = \underbrace{\arg\min_{\mathbf{u}} f(\mathbf{x}_k) + \langle \nabla f(\mathbf{x}_k), \mathbf{u} - \mathbf{x}_k \rangle + \frac{1}{2a_k} \|\mathbf{u} - \mathbf{x}_k\|_2^2}_{\text{Prox}_{a_k R}(\mathbf{x})} + R(\mathbf{u})$$

$$\text{Prox}_{a_k R}(\mathbf{x}) = \arg\min_{\mathbf{u}} a_k R(\mathbf{u}) + \frac{1}{2} \|\mathbf{u} - \mathbf{x}\|_2^2$$



Solvers $\min_{\mathbf{x} \in \mathbb{C}^N} \underbrace{\frac{1}{2} \|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2^2}_{f(\mathbf{x})} + R(\mathbf{x})$

Wavelet Reco. $\min_{\mathbf{x} \in \mathbb{C}^M} \frac{1}{2} \|\mathbf{AT}^{-1}\mathbf{x} - \mathbf{y}\|_2^2 + \lambda \|\mathbf{x}\|_1 \rightarrow \text{FISTA (APM)}$

TV Reco.

$$\min_{\mathbf{x} \in \mathbb{C}^N} \frac{1}{2} \|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2^2 + \lambda \underbrace{\|\mathbf{D}\mathbf{x}\|_1}_{\text{TV}(\mathbf{x})} \rightarrow \text{APM iter. proximal mapping}$$

Wavelet and TV Reco.

$$\min_{\mathbf{x} \in \mathbb{C}^N} \frac{1}{2} \|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2^2 + \lambda \alpha \|\mathbf{T}\mathbf{x}\|_1 + \lambda(1 - \alpha) \text{TV}(\mathbf{x}) \rightarrow \text{APM&ADMM}$$

Proximal gradient descent:

$$\mathbf{x}_{k+1} = \underbrace{\arg\min_{\mathbf{u}} f(\mathbf{x}_k) + \langle \nabla f(\mathbf{x}_k), \mathbf{u} - \mathbf{x}_k \rangle + \frac{1}{2a_k} \|\mathbf{u} - \mathbf{x}_k\|_2^2}_{\text{Prox}_{a_k R}(\mathbf{x})} + R(\mathbf{u})$$

$$\text{Prox}_{a_k R}(\mathbf{x}) = \arg\min_{\mathbf{u}} a_k R(\mathbf{u}) + \frac{1}{2} \|\mathbf{u} - \mathbf{x}\|_2^2$$

Nest. Acc.

$$\mathbf{x}_{k+1} = \text{Prox}_{a_k R}(\mathbf{v}_k - a_k \nabla_{\mathbf{x}} f(\mathbf{v}_k))$$

$$\mathbf{v}_{k+1} = t_k^1 \mathbf{x}_k + t_k^2 \mathbf{x}_{k+1}$$



Generalizing Nesterov's Scheme

Motivation

General Nest. Acc.

Numerical Tests

Magical High-order Methods → CS MRI Reco.

Problem Formulation

Our Suggestion

Numerical Results



Our Suggestion – Complex Quasi-Newton Proximal Methods (CQNPMs)

$$\min_{\mathbf{x} \in \mathbb{C}^N} \underbrace{\frac{1}{2} \|\mathbf{Ax} - \mathbf{y}\|_2^2}_{f(\mathbf{x})} + \lambda R(\mathbf{x})$$



Our Suggestion – Complex Quasi-Newton Proximal Methods (CQNPMs)

$$\min_{\mathbf{x} \in \mathbb{C}^N} \underbrace{\frac{1}{2} \|\mathbf{Ax} - \mathbf{y}\|_2^2}_{f(\mathbf{x})} + \lambda R(\mathbf{x})$$

$$\mathbf{x}_{k+1} = \underbrace{\arg\min_{\mathbf{u}} f(\mathbf{x}_k) + \langle \nabla f(\mathbf{x}_k), \mathbf{u} - \mathbf{x}_k \rangle + \frac{1}{2a_k} (\mathbf{u} - \mathbf{x}_k)^H \mathbf{B}_k (\mathbf{u} - \mathbf{x}_k) + R(\mathbf{u})}_{\text{Prox}_{a_k R}^{\mathbf{W}}(\mathbf{x}) = \arg\min_{\mathbf{u}} a_k R(\mathbf{u}) + \frac{1}{2} \|\mathbf{u} - \mathbf{x}\|_{\mathbf{W}}^2, \mathbf{W} = \mathbf{B}_k}$$



Our Suggestion – Complex Quasi-Newton Proximal Methods (CQNPMs)

$$\min_{\mathbf{x} \in \mathbb{C}^N} \underbrace{\frac{1}{2} \|\mathbf{Ax} - \mathbf{y}\|_2^2}_{f(\mathbf{x})} + \lambda R(\mathbf{x})$$

$$\mathbf{x}_{k+1} = \underbrace{\arg\min_{\mathbf{u}} f(\mathbf{x}_k) + \langle \nabla f(\mathbf{x}_k), \mathbf{u} - \mathbf{x}_k \rangle + \frac{1}{2a_k} (\mathbf{u} - \mathbf{x}_k)^H \mathbf{B}_k (\mathbf{u} - \mathbf{x}_k) + R(\mathbf{u})}_{\text{Prox}_{a_k R}^{\mathbf{W}}(\mathbf{x}) = \arg\min_{\mathbf{u}} a_k R(\mathbf{u}) + \frac{1}{2} \|\mathbf{u} - \mathbf{x}\|_{\mathbf{W}}^2, \mathbf{W} = \mathbf{B}_k}$$

$$\mathbf{x}_{k+1} = \text{Prox}_{a_k R}^{\mathbf{B}_k}(\mathbf{x}_k - a_k \mathbf{B}_k^{-1} \nabla_{\mathbf{x}} f(\mathbf{x}_k))$$

$\mathbf{B}_k \approx \mathbf{A}^H \mathbf{A}$, we use the SR1 method

Our Suggestion – Complex Quasi-Newton Proximal Methods (CQNPMs)

$$\min_{\mathbf{x} \in \mathbb{C}^N} \underbrace{\frac{1}{2} \|\mathbf{Ax} - \mathbf{y}\|_2^2}_{f(\mathbf{x})} + \lambda R(\mathbf{x})$$

$$\mathbf{x}_{k+1} = \underbrace{\arg\min_{\mathbf{u}} f(\mathbf{x}_k) + \langle \nabla f(\mathbf{x}_k), \mathbf{u} - \mathbf{x}_k \rangle + \frac{1}{2a_k} (\mathbf{u} - \mathbf{x}_k)^H \mathbf{B}_k (\mathbf{u} - \mathbf{x}_k) + R(\mathbf{u})}_{\text{Prox}_{a_k R}^{\mathbf{W}}(\mathbf{x}) = \arg\min_{\mathbf{u}} a_k R(\mathbf{u}) + \frac{1}{2} \|\mathbf{u} - \mathbf{x}\|_{\mathbf{W}}^2, \mathbf{W} = \mathbf{B}_k}$$

$$\mathbf{x}_{k+1} = \text{Prox}_{a_k R}^{\mathbf{B}_k}(\mathbf{x}_k - a_k \mathbf{B}_k^{-1} \nabla_{\mathbf{x}} f(\mathbf{x}_k))$$

$\mathbf{B}_k \approx \mathbf{A}^H \mathbf{A}$, we use the SR1 method

CQNPMs converge faster than APMs \rightarrow iterations

Our Suggestion – Complex Quasi-Newton Proximal Methods (CQNPMs)

$$\min_{\mathbf{x} \in \mathbb{C}^N} \underbrace{\frac{1}{2} \|\mathbf{Ax} - \mathbf{y}\|_2^2}_{f(\mathbf{x})} + \lambda R(\mathbf{x})$$

$$\mathbf{x}_{k+1} = \underbrace{\arg\min_{\mathbf{u}} f(\mathbf{x}_k) + \langle \nabla f(\mathbf{x}_k), \mathbf{u} - \mathbf{x}_k \rangle + \frac{1}{2a_k} (\mathbf{u} - \mathbf{x}_k)^H \mathbf{B}_k (\mathbf{u} - \mathbf{x}_k) + R(\mathbf{u})}_{\text{Prox}_{a_k R}^{\mathbf{W}}(\mathbf{x}) = \arg\min_{\mathbf{u}} a_k R(\mathbf{u}) + \frac{1}{2} \|\mathbf{u} - \mathbf{x}\|_{\mathbf{W}}^2, \mathbf{W} = \mathbf{B}_k}$$

$$\mathbf{x}_{k+1} = \text{Prox}_{a_k R}^{\mathbf{B}_k}(\mathbf{x}_k - a_k \mathbf{B}_k^{-1} \nabla_{\mathbf{x}} f(\mathbf{x}_k))$$

$\mathbf{B}_k \approx \mathbf{A}^H \mathbf{A}$, we use the SR1 method

CQNPMs converge faster than APMs → iterations
Wall (CPU or GPU) time?

Our Suggestion – Complex Quasi-Newton Proximal Methods (CQNPMs)

$$\min_{\mathbf{x} \in \mathbb{C}^N} \underbrace{\frac{1}{2} \|\mathbf{Ax} - \mathbf{y}\|_2^2}_{f(\mathbf{x})} + \lambda R(\mathbf{x})$$

$$\mathbf{x}_{k+1} = \underbrace{\arg\min_u f(\mathbf{x}_k) + \langle \nabla f(\mathbf{x}_k), \mathbf{u} - \mathbf{x}_k \rangle + \frac{1}{2a_k} (\mathbf{u} - \mathbf{x}_k)^H \mathbf{B}_k (\mathbf{u} - \mathbf{x}_k) + R(\mathbf{u})}_{\text{Prox}_{a_k R}^W(\mathbf{x}) = \arg\min_u a_k R(\mathbf{u}) + \frac{1}{2} \|\mathbf{u} - \mathbf{x}\|_W^2, W = \mathbf{B}_k}$$
$$\mathbf{x}_{k+1} = \text{Prox}_{a_k R}^{\mathbf{B}_k}(\mathbf{x}_k - a_k \mathbf{B}_k^{-1} \nabla_{\mathbf{x}} f(\mathbf{x}_k))$$

$\mathbf{B}_k \approx \mathbf{A}^H \mathbf{A}$, we use the SR1 method

CQNPMs converge faster than APMs \rightarrow iterations

Wall (CPU or GPU) time?

Slow because of $\text{Prox}_{a_k R}^{\mathbf{B}_k}$

Our Suggestion – Complex Quasi-Newton Proximal Methods (CQNPMs)

$$\min_{\mathbf{x} \in \mathbb{C}^N} \underbrace{\frac{1}{2} \|\mathbf{Ax} - \mathbf{y}\|_2^2}_{f(\mathbf{x})} + \lambda R(\mathbf{x})$$

$$\mathbf{x}_{k+1} = \underbrace{\arg\min_{\mathbf{u}} f(\mathbf{x}_k) + \langle \nabla f(\mathbf{x}_k), \mathbf{u} - \mathbf{x}_k \rangle + \frac{1}{2a_k} (\mathbf{u} - \mathbf{x}_k)^H \mathbf{B}_k (\mathbf{u} - \mathbf{x}_k) + R(\mathbf{u})}_{\text{Prox}_{a_k R}^W(\mathbf{x}) = \arg\min_{\mathbf{u}} a_k R(\mathbf{u}) + \frac{1}{2} \|\mathbf{u} - \mathbf{x}\|_W^2, W = \mathbf{B}_k}$$
$$\mathbf{x}_{k+1} = \text{Prox}_{a_k R}^{\mathbf{B}_k}(\mathbf{x}_k - a_k \mathbf{B}_k^{-1} \nabla_{\mathbf{x}} f(\mathbf{x}_k))$$

$\mathbf{B}_k \approx \mathbf{A}^H \mathbf{A}$, we use the SR1 method

CQNPMs converge faster than APMs → iterations

Wall (CPU or GPU) time?

Slow because of $\text{Prox}_{a_k R}^{\mathbf{B}_k}$



A Complex Quasi-Newton Proximal Method for Image Reconstruction in Compressed Sensing MRI

Challenging Issues – Compute $\text{Prox}_R^W(\mathbf{x})$

$$\text{Prox}_R^W(\mathbf{x}) = \operatorname{argmin}_{\mathbf{u}} R(\mathbf{u}) + \frac{1}{2} \|\mathbf{u} - \mathbf{x}\|_W^2$$

$$R(\mathbf{u}) = \|\cdot\|_1, \text{TV}, \alpha\|\cdot\|_1 + (1 - \alpha)\text{TV}$$

Challenging Issues – Compute $\text{Prox}_R^W(\mathbf{x})$

$$\text{Prox}_R^W(\mathbf{x}) = \operatorname{argmin}_{\mathbf{u}} R(\mathbf{u}) + \frac{1}{2} \|\mathbf{u} - \mathbf{x}\|_W^2$$

$$R(\mathbf{u}) = \|\cdot\|_1, \text{TV}, \alpha\|\cdot\|_1 + (1-\alpha)\text{TV}$$

Define $\mathcal{L} : \mathbb{C}^{(I-1) \times J} \times \mathbb{C}^{I \times (J-1)} \rightarrow \mathbb{C}^{I \times J}$ that

$$\mathcal{L}(\mathbf{P}, \mathbf{Q})_{i,j} = \mathbf{P}_{i,j} + \mathbf{Q}_{i,j} - \mathbf{P}_{i-1,j} - \mathbf{Q}_{i,j-1}, \forall i, j,$$

The adjoint operator of $\mathcal{L} : \mathbb{C}^{I \times J} \rightarrow \mathbb{C}^{(I-1) \times J} \times \mathbb{C}^{I \times (J-1)}$ is

$$\mathcal{L}^T(\mathbf{X}) = (\mathbf{P}, \mathbf{Q}),$$

that $\mathbf{P}_{i,j} = \mathbf{X}_{i,j} - \mathbf{X}_{i+1,j}$, $\mathbf{Q}_{i,j} = \mathbf{X}_{i,j} - \mathbf{X}_{i,j+1}$, $\forall i, j$.

Compute $\text{Prox}_R^W(\mathbf{x})$

For complex x, y , we have

$$\begin{aligned}\sqrt{|x|^2 + |y|^2} &= \max_{p_1, p_2 \in \mathbb{C}} \{ \Re(p_1^* x + p_2^* y) : |p_1|^2 + |p_2|^2 \leq 1 \} \\ |x| &= \max_{p \in \mathbb{C}} \{ \Re(p^* x) : |p| \leq 1 \}\end{aligned}$$

Compute $\text{Prox}_R^W(\mathbf{x})$

For complex x, y , we have

$$\begin{aligned}\sqrt{|x|^2 + |y|^2} &= \max_{p_1, p_2 \in \mathbb{C}} \{ \Re(p_1^* x + p_2^* y) : |p_1|^2 + |p_2|^2 \leq 1 \} \\ |x| &= \max_{p \in \mathbb{C}} \{ \Re(p^* x) : |p| \leq 1 \}\end{aligned}$$

Then

$$\text{TV}(\mathbf{x}) = \max_{(\mathbf{P}, \mathbf{Q}) \in \mathcal{P}} \Re \left\{ \text{vec}(\mathcal{L}(\mathbf{P}, \mathbf{Q}))^{\mathcal{H}} \mathbf{x} \right\},$$

$$\|\mathbf{T}\mathbf{x}\|_1 = \max_{\mathbf{z} \in \mathcal{Z}} \Re \left\{ \mathbf{z}^{\mathcal{H}} \mathbf{T}\mathbf{x} \right\}$$

\mathcal{P}, \mathcal{Z} : convex sets

Compute $\text{Prox}_R^W(\mathbf{x})$

Consider wavelet and TV: $\text{Prox}_R^W(\mathbf{x}) = \arg \min_{\mathbf{u}} R(\mathbf{u}) + \frac{1}{2} \|\mathbf{u} - \mathbf{x}\|_W^2$

$$\min_{\mathbf{x} \in \mathbb{C}^N} \max_{\substack{\mathbf{z} \in \mathcal{Z} \\ (\mathbf{P}, \mathbf{Q}) \in \mathcal{P}}} \|\mathbf{x} - \mathbf{v}_k\|_{\mathbf{B}_k}^2 + 2\lambda g(\mathbf{x}, \mathbf{z}, \mathbf{P}, \mathbf{Q})$$

$\mathbf{v}_k = \mathbf{x}_k - a_k \mathbf{B}_k^{-1} \nabla_{\mathbf{x}} f(\mathbf{x}_k)$ and

$$g(\mathbf{x}, \mathbf{z}, \mathbf{P}, \mathbf{Q}) = \Re \left\{ \alpha \langle \mathbf{T}\mathbf{x}, \mathbf{z} \rangle + (1 - \alpha) \text{vec}(\mathcal{L}(\mathbf{P}, \mathbf{Q}))^\mathcal{H} \mathbf{x} \right\}$$

Compute $\text{Prox}_R^W(\mathbf{x})$

Consider wavelet and TV: $\text{Prox}_R^W(\mathbf{x}) = \arg \min_{\mathbf{u}} R(\mathbf{u}) + \frac{1}{2} \|\mathbf{u} - \mathbf{x}\|_W^2$

$$\min_{\mathbf{x} \in \mathbb{C}^N} \max_{\substack{\mathbf{z} \in \mathcal{Z} \\ (\mathbf{P}, \mathbf{Q}) \in \mathcal{P}}} \|\mathbf{x} - \mathbf{v}_k\|_{\mathbf{B}_k}^2 + 2\lambda g(\mathbf{x}, \mathbf{z}, \mathbf{P}, \mathbf{Q})$$

$\mathbf{v}_k = \mathbf{x}_k - a_k \mathbf{B}_k^{-1} \nabla_{\mathbf{x}} f(\mathbf{x}_k)$ and

$$g(\mathbf{x}, \mathbf{z}, \mathbf{P}, \mathbf{Q}) = \Re \left\{ \alpha \langle \mathbf{T}\mathbf{x}, \mathbf{z} \rangle + (1 - \alpha) \text{vec}(\mathcal{L}(\mathbf{P}, \mathbf{Q}))^\mathcal{H} \mathbf{x} \right\}$$

$$\max_{\substack{\mathbf{z} \in \mathcal{Z}, \\ (\mathbf{P}, \mathbf{Q}) \in \mathcal{P}}} \min_{\mathbf{x} \in \mathbb{C}^N} \|\mathbf{x} - \mathbf{w}_k(\mathbf{z}, \mathbf{P}, \mathbf{Q})\|_{\mathbf{B}_k}^2 - \|\mathbf{w}_k(\mathbf{z}, \mathbf{P}, \mathbf{Q})\|_{\mathbf{B}_k}^2$$

$$\mathbf{w}_k(\mathbf{z}, \mathbf{P}, \mathbf{Q}) = \mathbf{v}_k - \lambda \mathbf{B}_k^{-1} \left(\alpha \mathbf{T}^\mathcal{H} \mathbf{z} + (1 - \alpha) \text{vec}(\mathcal{L}(\mathbf{P}, \mathbf{Q})) \right)$$

Compute Prox $_R^W(\mathbf{x})$

$$\max_{\substack{\mathbf{z} \in \mathcal{Z}, \\ (\mathbf{P}, \mathbf{Q}) \in \mathcal{P}}} \min_{\mathbf{x} \in \mathbb{C}^N} \|\mathbf{x} - \mathbf{w}_k(\mathbf{z}, \mathbf{P}, \mathbf{Q})\|_{\mathbf{B}_k}^2 - \|\mathbf{w}_k(\mathbf{z}, \mathbf{P}, \mathbf{Q})\|_{\mathbf{B}_k}^2$$

$$\mathbf{w}_k(\mathbf{z}, \mathbf{P}, \mathbf{Q}) = \mathbf{v}_k - \lambda \mathbf{B}_k^{-1} \left(\alpha \mathbf{T}^{\mathcal{H}} \mathbf{z} + (1 - \alpha) \text{vec}(\mathcal{L}(\mathbf{P}, \mathbf{Q})) \right)$$

$$\mathbf{x}^* = \mathbf{w}_k(\mathbf{z}, \mathbf{P}, \mathbf{Q})$$

$$(\mathbf{z}^*, \mathbf{P}^*, \mathbf{Q}^*) = \operatorname*{argmin}_{\substack{\mathbf{z} \in \mathcal{Z}, \\ (\mathbf{P}, \mathbf{Q}) \in \mathcal{P}}} \|\mathbf{w}_k(\mathbf{z}, \mathbf{P}, \mathbf{Q})\|_{\mathbf{B}_k}^2.$$

Compute Prox $_R^W(\mathbf{x})$

$$\max_{\mathbf{z} \in \mathcal{Z}, \quad \mathbf{x} \in \mathbb{C}^N} \quad \min_{(\mathbf{P}, \mathbf{Q}) \in \mathcal{P}} \quad \|\mathbf{x} - \mathbf{w}_k(\mathbf{z}, \mathbf{P}, \mathbf{Q})\|_{\mathbf{B}_k}^2 - \|\mathbf{w}_k(\mathbf{z}, \mathbf{P}, \mathbf{Q})\|_{\mathbf{B}_k}^2$$

$$\mathbf{w}_k(\mathbf{z}, \mathbf{P}, \mathbf{Q}) = \mathbf{v}_k - \lambda \mathbf{B}_k^{-1} \left(\alpha \mathbf{T}^{\mathcal{H}} \mathbf{z} + (1 - \alpha) \text{vec}(\mathcal{L}(\mathbf{P}, \mathbf{Q})) \right)$$

$$\mathbf{x}^* = \mathbf{w}_k(\mathbf{z}, \mathbf{P}, \mathbf{Q})$$

$$(\mathbf{z}^*, \mathbf{P}^*, \mathbf{Q}^*) = \operatorname{argmin}_{\substack{\mathbf{z} \in \mathcal{Z}, \\ (\mathbf{P}, \mathbf{Q}) \in \mathcal{P}}} \|\mathbf{w}_k(\mathbf{z}, \mathbf{P}, \mathbf{Q})\|_{\mathbf{B}_k}^2.$$

Gradient

$$-2\lambda \begin{bmatrix} \alpha \mathbf{T} \\ (1 - \alpha) \mathcal{L}^T \end{bmatrix} \mathbf{w}_k(\mathbf{z}, \mathbf{P}, \mathbf{Q}) \quad \text{APM, iter.}$$

Compute Prox R^W (\mathbf{x})

$$\max_{\mathbf{z} \in \mathcal{Z}, \atop (\mathbf{P}, \mathbf{Q}) \in \mathcal{P}} \min_{\mathbf{x} \in \mathbb{C}^N} \|\mathbf{x} - \mathbf{w}_k(\mathbf{z}, \mathbf{P}, \mathbf{Q})\|_{\mathbf{B}_k}^2 - \|\mathbf{w}_k(\mathbf{z}, \mathbf{P}, \mathbf{Q})\|_{\mathbf{B}_k}^2$$

$$\mathbf{w}_k(\mathbf{z}, \mathbf{P}, \mathbf{Q}) = \mathbf{v}_k - \lambda \mathbf{B}_k^{-1} \left(\alpha \mathbf{T}^{\mathcal{H}} \mathbf{z} + (1 - \alpha) \text{vec}(\mathcal{L}(\mathbf{P}, \mathbf{Q})) \right)$$

$$\mathbf{x}^* = \mathbf{w}_k(\mathbf{z}, \mathbf{P}, \mathbf{Q})$$

$$(\mathbf{z}^*, \mathbf{P}^*, \mathbf{Q}^*) = \underset{\mathbf{z} \in \mathcal{Z}, \atop (\mathbf{P}, \mathbf{Q}) \in \mathcal{P}}{\text{argmin}} \|\mathbf{w}_k(\mathbf{z}, \mathbf{P}, \mathbf{Q})\|_{\mathbf{B}_k}^2.$$

Gradient

$$-2\lambda \begin{bmatrix} \alpha \mathbf{T} \\ (1 - \alpha) \mathcal{L}^T \end{bmatrix} \mathbf{w}_k(\mathbf{z}, \mathbf{P}, \mathbf{Q}) \quad \text{APM, iter.}$$

But

$$\mathbf{w}_k(\mathbf{z}, \mathbf{P}, \mathbf{Q}) = \mathbf{v}_k - \lambda \mathbf{B}_k^{-1} \left(\alpha \mathbf{T}^{\mathcal{H}} \mathbf{z} + (1 - \alpha) \text{vec}(\mathcal{L}(\mathbf{P}, \mathbf{Q})) \right)$$

Structure of $\mathbf{B}_k = \mathbf{D} + \sigma \mathbf{u} \mathbf{u}^{\mathcal{H}}$

Define $\sigma = 1 / \langle \mathbf{m}_k - \mathbf{H}_0 \mathbf{s}_k, \mathbf{s}_k \rangle$ & $\mathbf{D} = \mathbf{H}_0$

Algorithm 1 SR1

Initialization: $\gamma > 1$, $\delta = 10^{-8}$, $\Xi > 0$ a fixed real scalar, \mathbf{x}_k , \mathbf{x}_{k-1} , $\nabla f(\mathbf{x}_k)$, and $\nabla f(\mathbf{x}_{k-1})$.

- 1: \vdots
 - 2: Set $\mathbf{s}_k \leftarrow \mathbf{x}_k - \mathbf{x}_{k-1}$ and $\mathbf{m}_k \leftarrow \nabla f(\mathbf{x}_k) - \nabla f(\mathbf{x}_{k-1})$.
 - 3: Compute $\tau_k \leftarrow \gamma \frac{\|\mathbf{m}_k\|_2^2}{\langle \mathbf{s}_k, \mathbf{m}_k \rangle}$. % $\langle \mathbf{a}, \mathbf{b} \rangle = \mathbf{b}^{\mathcal{H}} \mathbf{a}$
 - 4: \vdots
 - 5: $\mathbf{H}_0 \leftarrow \tau_k \mathbf{I}$.
 - 6: $\mathbf{u}_k \leftarrow \mathbf{m}_k - \mathbf{H}_0 \mathbf{s}_k$.
 - 7: $\mathbf{B}_k \leftarrow \mathbf{H}_0 + \frac{\mathbf{u}_k \mathbf{u}_k^{\mathcal{H}}}{\langle \mathbf{m}_k - \mathbf{H}_0 \mathbf{s}_k, \mathbf{s}_k \rangle}$.
-

Only Wavelet

$$(\mathbf{z}^*, \mathbf{P}^*, \mathbf{Q}^*) = \operatorname{argmin}_{\substack{\mathbf{z} \in \mathcal{Z}, \\ (\mathbf{P}, \mathbf{Q}) \in \mathcal{P}}} \|\mathbf{w}_k(\mathbf{z}, \mathbf{P}, \mathbf{Q})\|_{\mathbf{B}_k}^2.$$

Gradient

$$-2\lambda \begin{bmatrix} 1 \mathbf{T} \\ (1 - 1)\mathcal{L}^T \end{bmatrix} \mathbf{w}_k(\mathbf{z}, \mathbf{P}, \mathbf{Q})$$



Only Wavelet

$$(\mathbf{z}^*, \mathbf{P}^*, \mathbf{Q}^*) = \operatorname{argmin}_{\substack{\mathbf{z} \in \mathcal{Z}, \\ (\mathbf{P}, \mathbf{Q}) \in \mathcal{P}}} \|\mathbf{w}_k(\mathbf{z}, \mathbf{P}, \mathbf{Q})\|_{\mathbf{B}_k}^2.$$

Gradient

$$-2\lambda \begin{bmatrix} 1 \mathbf{T} \\ (1 - 1)\mathcal{L}^T \end{bmatrix} \mathbf{w}_k(\mathbf{z}, \mathbf{P}, \mathbf{Q})$$

Solve $\min_{\mathbf{x} \in \mathbb{C}^M} \frac{1}{2} \|\mathbf{A}\mathbf{T}^{-1}\mathbf{x} - \mathbf{y}\|_2^2 + \lambda \|\mathbf{x}\|_1$



Only Wavelet

$$(\mathbf{z}^*, \mathbf{P}^*, \mathbf{Q}^*) = \operatorname{argmin}_{\substack{\mathbf{z} \in \mathcal{Z}, \\ (\mathbf{P}, \mathbf{Q}) \in \mathcal{P}}} \|\mathbf{w}_k(\mathbf{z}, \mathbf{P}, \mathbf{Q})\|_{\mathbf{B}_k}^2.$$

Gradient

$$-2\lambda \begin{bmatrix} 1 \mathbf{T} \\ (1 - 1)\mathcal{L}^T \end{bmatrix} \mathbf{w}_k(\mathbf{z}, \mathbf{P}, \mathbf{Q})$$

Solve $\min_{\mathbf{x} \in \mathbb{C}^M} \frac{1}{2} \|\mathbf{A}\mathbf{T}^{-1}\mathbf{x} - \mathbf{y}\|_2^2 + \lambda \|\mathbf{x}\|_1$

Theorem (Becker19 SIAMOPT: real \rightarrow complex)

Let $\mathbf{W} = \mathbf{D} \pm \mathbf{u}\mathbf{u}^H$. Then,

$$\operatorname{Prox}_{\lambda R}^{\mathbf{W}}(\mathbf{x}) = \operatorname{Prox}_{\lambda R}^{\mathbf{D}}(\mathbf{x} \mp \mathbf{D}^{-1}\mathbf{u}\beta^*),$$

where $\beta^* \in \mathbb{C}$ is the unique zero of the following nonlinear equation

$$\mathbb{J}(\beta) : \mathbf{u}^H (\mathbf{x} - \operatorname{Prox}_{\lambda R}^{\mathbf{D}}(\mathbf{x} \mp \mathbf{D}^{-1}\mathbf{u}\beta)) + \beta.$$

Only Wavelet

$$(\mathbf{z}^*, \mathbf{P}^*, \mathbf{Q}^*) = \operatorname{argmin}_{\substack{\mathbf{z} \in \mathcal{Z}, \\ (\mathbf{P}, \mathbf{Q}) \in \mathcal{P}}} \|\mathbf{w}_k(\mathbf{z}, \mathbf{P}, \mathbf{Q})\|_{\mathbf{B}_k}^2.$$

Gradient

$$-2\lambda \begin{bmatrix} 1 \mathbf{T} \\ (1 - 1)\mathcal{L}^T \end{bmatrix} \mathbf{w}_k(\mathbf{z}, \mathbf{P}, \mathbf{Q})$$

Solve $\min_{\mathbf{x} \in \mathbb{C}^M} \frac{1}{2} \|\mathbf{A}\mathbf{T}^{-1}\mathbf{x} - \mathbf{y}\|_2^2 + \lambda \|\mathbf{x}\|_1$

Theorem (Becker19 SIAMOPT: real \rightarrow complex)

Let $\mathbf{W} = \mathbf{D} \pm \mathbf{u}\mathbf{u}^H$. Then,

$$\operatorname{Prox}_{\lambda R}^{\mathbf{W}}(\mathbf{x}) = \operatorname{Prox}_{\lambda R}^{\mathbf{D}}(\mathbf{x} \mp \mathbf{D}^{-1}\mathbf{u}\beta^*),$$

where $\beta^* \in \mathbb{C}$ is the unique zero of the following nonlinear equation

$$\mathbb{J}(\beta) : \mathbf{u}^H (\mathbf{x} - \operatorname{Prox}_{\lambda R}^{\mathbf{D}}(\mathbf{x} \mp \mathbf{D}^{-1}\mathbf{u}\beta)) + \beta.$$

$\mathbf{B}_k = \mathbf{D} + \sigma \mathbf{u}\mathbf{u}^H$ and $\sigma = 1 / \langle \mathbf{m}_k - \mathbf{H}_0 \mathbf{s}_k, \mathbf{s}_k \rangle$ is real



What is More: TV + Wavelet?

Wavelet and TV:

Gradient

$$-2\lambda \begin{bmatrix} \alpha \mathbf{T} \\ (1 - \alpha) \mathcal{L}^T \end{bmatrix} \mathbf{w}_k(\mathbf{z}, \mathbf{P}, \mathbf{Q})$$



What is More: TV + Wavelet?

Wavelet and TV:

Gradient

$$-2\lambda \begin{bmatrix} \alpha \mathbf{T} \\ (1-\alpha) \mathcal{L}^T \end{bmatrix} \mathbf{w}_k(\mathbf{z}, \mathbf{P}, \mathbf{Q})$$

Partially smooth:

$$\min_{\mathbf{x} \in \mathbb{C}^N} \underbrace{\frac{1}{2} \|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2^2 + \lambda \alpha \cdot S^\eta(\|\mathbf{T}\mathbf{x}\|_1)}_{f(\mathbf{x})} + \underbrace{\lambda(1-\alpha) \text{TV}(\mathbf{x})}_{R(\mathbf{x})}$$

$$S^\eta(\|\mathbf{x}\|_1) = \sum_{n=1}^N \sqrt{\mathbf{x}_n^2 + \eta}$$



Generalizing Nesterov's Scheme

Motivation

General Nest. Acc.

Numerical Tests

Magical High-order Methods → CS MRI Reco.

Problem Formulation

Our Suggestion

Numerical Results



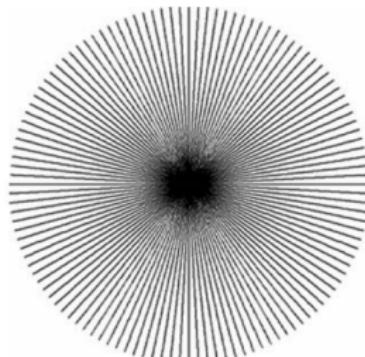
Experimental Settings

- Took k-space data from NYU fastMRI dataset
- Applied the ESPIRiT algorithm to recover the complex images
- Cropped the images to size 256×256 with maximum magnitude scaled to one
- Formulated the simulated k-space data with a given trajectory
- Added Gaussian noise with mean zero and variance 10^{-2} to all coils to form the final measurements

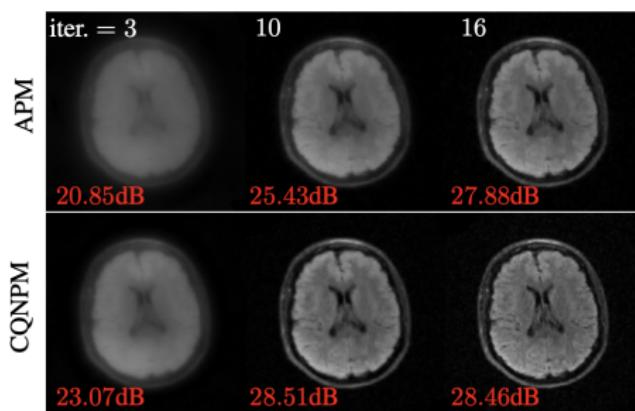
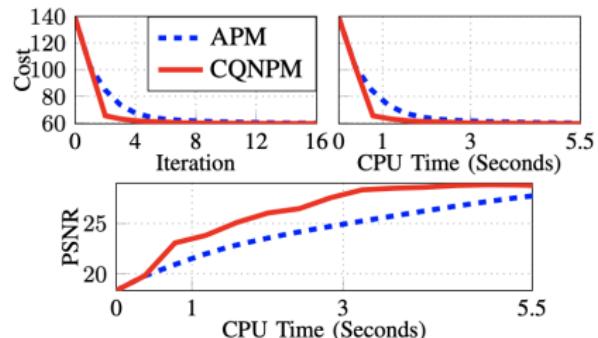
Experimental Settings

- Took k-space data from NYU fastMRI dataset
- Applied the ESPIRiT algorithm to recover the complex images
- Cropped the images to size 256×256 with maximum magnitude scaled to one
- Formulated the simulated k-space data with a given trajectory
- Added Gaussian noise with mean zero and variance 10^{-2} to all coils to form the final measurements

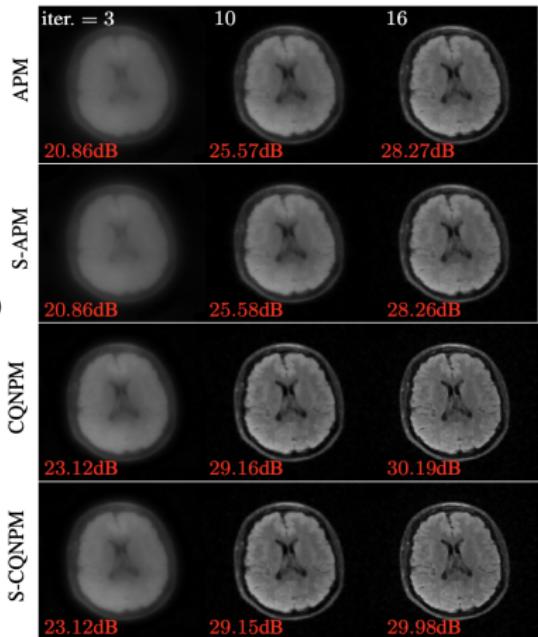
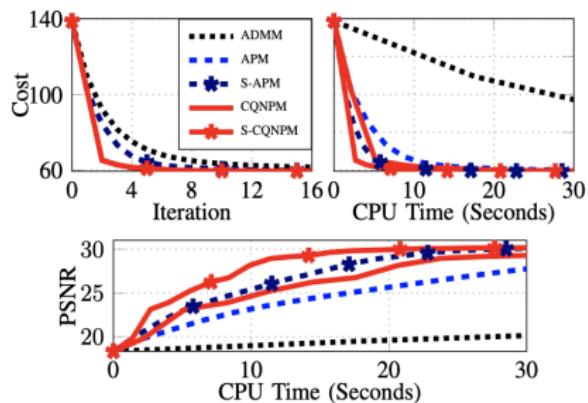
96 radial projections, 512 readout points, and 12 coils



Wavelet



Ours: CQNPBM & S-CQNPBM





😊 Thanks & Questions? 😊

