Splitting with Near-Circulant Linear Systems: Applications to Total Variation CT and PET

Ernest K. Ryu   S. Ko   J.-H. Won

Outline

Imaging through optimization

Many medical imaging problems are solved via

$$\underset{x \in \mathbb{R}^n}{\text{minimize}} \quad g(Ax - b)$$

- g convex function
- $A \in \mathbb{R}^{m \times n}$ contains something like the Radon transform
- b measurement
- x image to recover

PDHG and ADMM/DRS are popular methods for solving large instances of this problem.

ADMM/DRS

ADMM/DRS

$$x^{k+1} = x^k - \alpha^{-1}(A^TA)^+A^Tu^k$$
$$u^{k+1} = \mathbf{Prox}_{\alpha g^*}(u^k + \alpha(A(2x^{k+1} - x^k) - b))$$

converges for $\alpha > 0$.

ADMM/DRS often requires relatively fewer iterations to converge, but the cost of computing $(AA^T)^+$ can be prohibitively expensive.
Can't perform even a single iteration.

This isn't the usual form of ADMM. We performed a few change of variables.

Why can't we invert?

The $n \times n$ matrix

$$A^T A$$

is difficult to directly (pseudo) invert when, say, $n \geq 512^2$

If $A^T A$ were circulant (spatially invariant) we could use the FFT for the (pseudo) inversion. However, $A^T A$ is not circulant.

PDHG

PDHG

$$
\begin{aligned}
x^{k+1} &= x^k - \beta^{-1} A^T u^k \\
u^{k+1} &= \mathbf{Prox}_{\alpha g^*}(u^k + \alpha(A(2x^{k+1} - x^k) - b))
\end{aligned}
$$

converges for $\beta/\alpha \geq \lambda_{\max}(AA^T)$.

While PDHG has small cost per iteration, it often requires prohibitively many iterations to converge to a good solution.
Requires too many iterations.

Near-circulant matrix

Although $A^T A$ is not circulant, it is *near-circulant*.

In imaging applications, R is a discretization of $\mathcal{R}$, a linear operator on a continuous image.

$$\mathcal{R} \xrightarrow{\text{discretize}} R$$

$\mathcal{R}^*\mathcal{R}$ is a linear spatially invariant operator, so the Fourier transform diagonalizes it:

$$\mathcal{R}^*\mathcal{R}f = F^{-1}\left[\hat{k}(\omega)\hat{f}(\omega)\right]$$

for some $\hat{k}(\omega)$ where $\hat{f} = \mathcal{F}[f]$.

Near-circulant matrix

However, the discretization of $\mathcal{R}^*\mathcal{R}$ (in the Fourier domain) is only approximately equal to $R^T R$.

$$\mathcal{R}^*\mathcal{R} \xrightarrow{\text{discretize}} F^{-1}\operatorname{diag}(h)F \neq R^T R$$

Discretize then transpose $\neq$ Transpose then discretize.

$$\begin{array}{ccc} \mathcal{R} & \xrightarrow{\qquad\qquad\text{discretize}\qquad\qquad} & R \\ \downarrow & & \downarrow \\ \mathcal{R}^*\mathcal{R} \xrightarrow{\text{discretize}} F^{-1}\operatorname{diag}(h)F & \not\longleftrightarrow & R^T R \end{array}$$

(The diagram doesn't commute.)

Near-circulant matrix

Nevertheless,

$$\underbrace{F^{-1}\mathrm{diag}(h)F}_{\text{circulant}} \approx R^T R$$

and $R^T R$ is a near-circulant matrix.

We can use

$$F^{-1}\mathrm{diag}(h)^+ F \approx (R^T R)^+$$

as a computationally efficient approximation to $(R^T R)^+$.

Why not circulant?

$A^T A$ is not circulant for 2 reasons.

Reason 1: Boundary conditions.

When we discretize, we move to a bounded domain. Since we do not use periodic boundary conditions, $A^T A$ is not circulant.

Reason 1 is less important. There are ways to resolve the issue (e.g. zero-padding).

Why not circulant?

Reason 2: Spatial invariance is not preserved in the discretization.

The Radon transform $\mathcal{R}$ maps from Cartesian to polar coordinates, and $\mathcal{R}^*$ maps back to Cartesian coordinates. In continuous space, $\mathcal{R}^*\mathcal{R}$ is spatially invariant.

However, when discretized, the Cartesian $\to$ polar $\to$ Cartesian change of coordinates breaks spatial invariance.

There's no easy way to resolve this issue.

Motivation of main algorithm

Leverage computationally efficient approximate (pseudo) inverse of $A^T A$.

ADMM uses the exact inverse. PDHG uses no inverse.

We want something in between.

Main method NCS

Near-Circulant Splitting (NCS)

$$
\begin{aligned}
x^{k+1} &= x^k - M^+ A^T u^k \\
u^{k+1} &= \mathbf{Prox}_{\alpha g^*}(u^k + \alpha(A(2x^{k+1} - x^k) - b))
\end{aligned}
$$

converges for $M \succeq \alpha A^T A$.

$M = \alpha A^T A$ gives us ADMM.

$M = \beta I$ gives us PDHG.

NCS

$$x^{k+1} = x^k - M^+ A^T u^k$$
$$u^{k+1} = \mathbf{Prox}_{\alpha g^*}(u^k + \alpha(A(2x^{k+1} - x^k) - b))$$

Choose

$$M = \beta I + C$$

where $C \approx \alpha A^T A$ is circulant and $\beta > 0$ is small.

Prior work

- Deng and Yin, JSC, 2016. (Posted 2012.)
  Briefly discusses a similar idea.
- O'Connor and Vandenberghe, SIIMS, 2014.
  Starts with a similar motivation and presents a variety of primal-dual methods.
- Bredies and Sun, J. Math. Imag. Vis. and SINUM, 2015.
  Starts with a similar motivation and presents "Preconditioned DRS".
- O'Connor and Vandenberghe, 2017.
  Presented a reduction, which this work uses for the analysis.

Outline

Assumptions

$$g \text{ is convex and } \mathbf{Prox}_{\alpha g^*} \text{ is well defined.} \tag{A1}$$

$$\text{A primal-dual solution pair exists and strong duality holds.} \tag{A2}$$

$$M \succeq \alpha A^T A \tag{A3}$$

(A1) and (A2) are very standard.

Convergence

Theorem

Assume (A1) *and* (A2). *Then* $x^k \to x^\star$ *and* $u^k \to u^\star$, *where* $x^\star$ *and* $u^\star$ *are primal and dual solutions.*

Rate of convergence

Theorem

Assume (A1) *and* (A2). *Assume* g *is* L*-Lipschitz. Then*

$$
\begin{aligned}
&g(Ax^{k+1} - b) - g(Ax^\star - b) \\
&\qquad \le \frac{1}{\alpha\sqrt{k+1}} \Big(\|u^0 - u^\star - \alpha A(x^0 - x^\star)\| + \|u^\star\| + L \\
&\qquad\qquad + \|x^0 - x^\star\|_{(\alpha M - \alpha^2 A^T A)} \Big)^2
\end{aligned}
$$

where $x^\star$ *and* $u^\star$ *are primal and dual solutions.*

Outline

Total variation CT

In CT imaging with TV regularization, we solve

$$\begin{array}{ll} \text{minimize} & (1/2)\|y\|^2 + \lambda\|z\|_1 \\ \text{subject to} & \begin{bmatrix} R \\ D \end{bmatrix} x - \begin{bmatrix} b \\ 0 \end{bmatrix} = \begin{bmatrix} y \\ z \end{bmatrix} \end{array}$$

Total variation CT

Use the circulant approximation,

$$M = F^{-1}\mathrm{diag}(h)F \approx \alpha R^T R + \alpha D^T D$$

where $\mathrm{diag}(h) \in \mathbb{R}^{N^2 \times N^2}$ corresponds to the $N \times N$ mask defined by

$$\begin{aligned} H_{(j+1)(k+1)} = \gamma + C\alpha \left(\min\{j, N-j\}^2 + \min\{k, N-k\}^2\right)^{-1/2} \\ + 4\beta^2/\alpha \left(\sin^2\left(\frac{j\pi}{N}\right) + 4\sin^2\left(\frac{k\pi}{N}\right)\right) \end{aligned}$$

2nd term corresponds to the Radon transform's ramp filter and 3rd term corresponds to the periodic Laplacian's eigenvalues.

Total variation CT

The method becomes

$$
\begin{aligned}
x^{k+1} &= x^k - F^{-1}\mathrm{diag}(h)^{-1}F(R^T u^k + D^T v^k) \\
u^{k+1} &= \frac{1}{1+\alpha}(u^k + \alpha R(2x^{k+1} - x^k) - \alpha b) \\
v^{k+1} &= \Pi_{[-\lambda,\lambda]}\left(v^k + \alpha D(2x^{k+1} - x^k)\right)
\end{aligned}
$$

The computational bottleneck is multiplication by R and R^T.

Stastical Recon PET

In statistical reconstruction PET with TV regularization, we solve

$$\underset{x\in\mathbb{R}^n}{\text{minimize}} \quad \textstyle\sum_{i=1}^n \ell((Ex)_i; b_i) + \lambda\|Dx\|_1$$

$\ell(\mu; b)$ is the negative log-likelihood for the Poisson distribution with mean μ and observation $b \in \mathbb{Z}$.

Stastical Recon PET

The log-likelihood function is

$$\ell(y; b) = y - b \log y$$

Although ℓ is a differentiable convex function, its domain is not closed and its gradient is not Lipschitz continuous. This makes gradient methods difficult to apply.

However, $\mathbf{Prox}_{\alpha\ell}$ has a closed-form solution.

Stastical Recon PET

The main method becomes

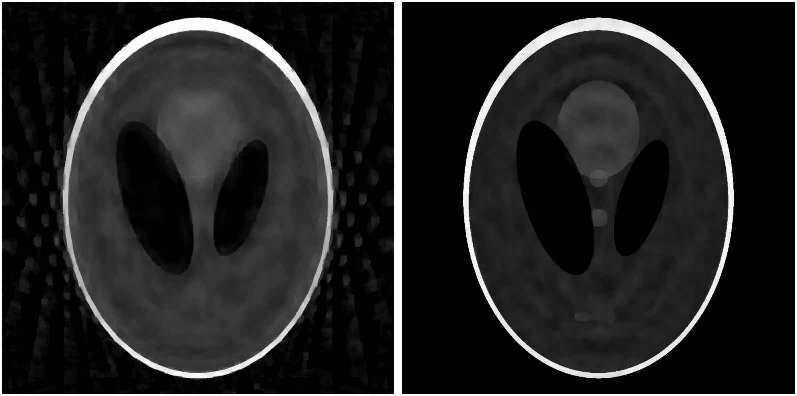
$$
\begin{aligned}
x^{k+1} &= x^k - F^{-1}\mathrm{diag}(h)^{-1}F(E^T u^k + D^T v^k + s^k) \\
u^{k+1} &= S(u^k + \alpha E(2x^{k+1} - x^k); \alpha b_i) \\
v^{k+1} &= \Pi_{[-\lambda,\lambda]}(v^k + \alpha D(2x^{k+1} - x^k))
\end{aligned}
$$

where

$$
S(u; c) = 1 + \frac{u - 1 - \sqrt{(u-1)^2 + 4c}}{2}.
$$

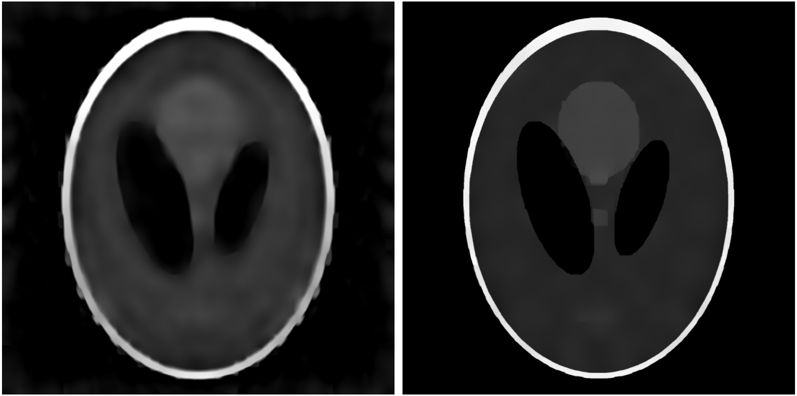
Outline

CT: PDHG



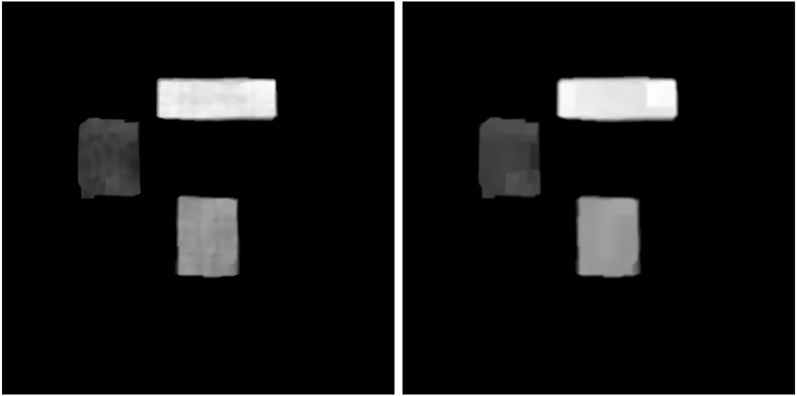
1024×1024, 300 and $1,000$ iterations, 4.67s and 15.4s on TITAN Xp

CT: NCS



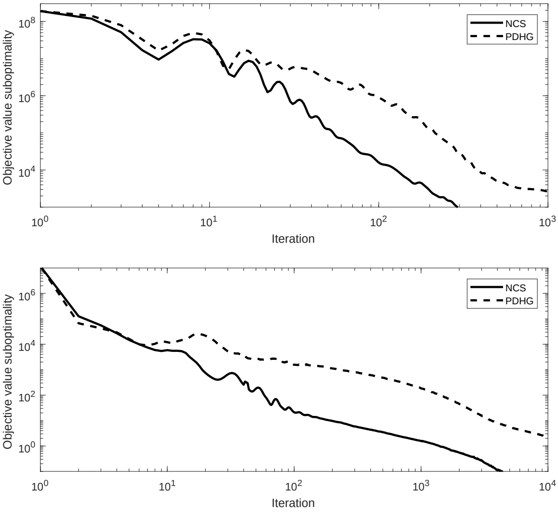
1024×1024, 30 and 100 iterations, 0.51s and 1.59s on TITAN Xp

PET: PDHG and NCS



512×512, 300 iterations for PDHG and 30 iterations for NCS, 2.99s and 0.33s on TITAN Xp

Objective value suboptimality vs. iteration count



GPU acceleration

	Intel Core i7-990X	TITAN Xp	Speedup
CT (128×128)	2.41s	3.86s	0.62x
CT (256×256)	8.51s	4.49s	1.90x
CT (512×512)	38.92s	5.32s	7.32x
CT (1024×1024)	198.53s	14.99s	13.2x
PET (128×128)	27.57s	4.07s	6.77x
PET (256×256)	109.68s	5.09s	21.5x
PET (512×512)	452.8s	9.39s	48.2x

Outline

Conclusion

NCS is a method that efficiently solves

$$\text{minimize} \quad g(Ax - b)$$

by leveraging a circulant approximation to $A^T A$.

Although the optimization problem and the approximate inverse assumption is seemingly very specific, it fits imaging applications very well.

Experiments on synthetic data are promising.