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

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

Motivation and algorithm

Theory

Applications

Experiments

Conclusion
Imaging through optimization

Many medical imaging problems are solved via

$$\min_{x \in \mathbb{R}^n} g(Ax - b)$$

- $g$ 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.

Motivation and algorithm
ADMM/DRS

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

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.

Motivation and algorithm
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

\[ x^{k+1} = x^k - \beta^{-1} A^T u^k \]

\[ u^{k+1} = \text{Prox}_{\alpha g^*}(u^k + \alpha (A(2x^{k+1} - x^k) - b)) \]

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 $R^T R$.

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

Discretize then transpose $\neq$ Transpose then discretize.

(The diagram doesn’t commute.)
Near-circulant matrix

Nevertheless,

\[ F^{-1} \text{diag}(h) F \approx R^T R \]

and \( R^T R \) is a near-circulant matrix.

We can use

\[ F^{-1} \text{diag}(h)^+ F \approx (R^T R)^+ \]

as a computationally efficient approximation to \((R^T R)^+\).
Why not circulant?

$A^TA$ 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, $A^TA$ 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 $R$ maps from Cartesian to polar coordinates, and $R^*$ maps back to Cartesian coordinates. In continuous space, $R^*R$ is spatially invariant.

However, when discretized, the Cartesian $\rightarrow$ polar $\rightarrow$ 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)

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

converges for \( M \succeq \alpha A^T A \).

\( M = \alpha A^T A \) gives us ADMM.

\( M = \beta I \) gives us PDHG.
\[ x^{k+1} = x^k - M^+ A^T u^k \]
\[ u^{k+1} = \text{Prox}_{\alpha g^*}(u^k + \alpha(A(2x^{k+1} - x^k) - b)) \]

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.

  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

Motivation and algorithm

Theory

Applications

Experiments

Conclusion
Assumptions

$g$ is convex and $\text{Prox}_{\alpha g^*}$ is well defined. (A1)
A primal-dual solution pair exists and strong duality holds. (A2)

$$M \succeq \alpha A^T A$$  (A3)

(A1) and (A2) are very standard.
Convergence

Theorem

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

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

$$g(Ax^{k+1} - b) - g(Ax^* - b) \leq \frac{1}{\alpha \sqrt{k + 1}} \left( \|u^0 - u^* - \alpha A(x^0 - x^*)\| + \|u^*\| + L \right)$$

$$+ \left( \frac{1}{\alpha M - \alpha^2 A^T A} \right)^2$$

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

Motivation and algorithm

Theory

Applications

Experiments

Conclusion
Total variation CT

In CT imaging with TV regularization, we solve

\[
\begin{align*}
\text{minimize} & \quad \frac{1}{2} \|y\|^2 + \lambda \|z\|_1 \\
\text{subject to} & \quad \begin{bmatrix} R \\ D \end{bmatrix} x - \begin{bmatrix} b \\ 0 \end{bmatrix} = \begin{bmatrix} y \\ z \end{bmatrix}
\end{align*}
\]
Use the circulant approximation,

\[ M = F^{-1} \text{diag}(h) F \approx \alpha R^T R + \alpha D^T D \]

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

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

2nd term corresponds to the Radon transform’s ramp filter and 3rd term corresponds to the periodic Laplacian’s eigenvalues.
The method becomes

\[ x^{k+1} = x^k - F^{-1} \text{diag}(h)^{-1} F (R^T u^k + D^T v^k) \]
\[ u^{k+1} = \frac{1}{1 + \alpha} (u^k + \alpha R (2x^{k+1} - x^k) - \alpha b) \]
\[ v^{k+1} = \Pi_{[-\lambda, \lambda]} (v^k + \alpha D (2x^{k+1} - x^k)) \]

The computational bottleneck is multiplication by \( R \) and \( R^T \).
In statistical reconstruction PET with TV regularization, we solve

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

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

Applications 24
The log-likelihood function is

\[ \ell(y; b) = y - b \log y \]

Although \( \ell \) 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, \( \text{Prox}_{\alpha \ell} \) has a closed-form solution.
The main method becomes

\[ x^{k+1} = x^k - F^{-1} \text{diag}(h)^{-1} F(ETu^k + DTv^k + s^k) \]

\[ u^{k+1} = S(u^k + \alpha E(2x^{k+1} - x^k); \alpha b_i) \]

\[ v^{k+1} = \Pi_{[-\lambda, \lambda]}(v^k + \alpha D(2x^{k+1} - x^k)) \]

where

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

Motivation and algorithm

Theory

Applications

Experiments

Conclusion
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

<table>
<thead>
<tr>
<th>Iteration</th>
<th>Objective value suboptimality</th>
<th>NCS</th>
<th>PDHG</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>10^8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>10^6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>10^4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>10^2</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Experiments

The graphs above illustrate the objective value suboptimality vs. iteration count for NCS and PDHG algorithms. The plots show the decrease in suboptimality as the iteration count increases, with NCS generally outperforming PDHG.
## GPU acceleration

<table>
<thead>
<tr>
<th></th>
<th>Intel Core i7-990X</th>
<th>TITAN Xp</th>
<th>Speedup</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>CT (128 × 128)</strong></td>
<td>2.41s</td>
<td>3.86s</td>
<td>0.62x</td>
</tr>
<tr>
<td><strong>CT (256 × 256)</strong></td>
<td>8.51s</td>
<td>4.49s</td>
<td>1.90x</td>
</tr>
<tr>
<td><strong>CT (512 × 512)</strong></td>
<td>38.92s</td>
<td>5.32s</td>
<td>7.32x</td>
</tr>
<tr>
<td><strong>CT (1024 × 1024)</strong></td>
<td>198.53s</td>
<td>14.99s</td>
<td>13.2x</td>
</tr>
<tr>
<td><strong>PET (128 × 128)</strong></td>
<td>27.57s</td>
<td>4.07s</td>
<td>6.77x</td>
</tr>
<tr>
<td><strong>PET (256 × 256)</strong></td>
<td>109.68s</td>
<td>5.09s</td>
<td>21.5x</td>
</tr>
<tr>
<td><strong>PET (512 × 512)</strong></td>
<td>452.8s</td>
<td>9.39s</td>
<td>48.2x</td>
</tr>
</tbody>
</table>
Outline

Motivation and algorithm

Theory

Applications

Experiments

Conclusion
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.