Accelerating image recovery using variable splitting methods

Jeffrey A. Fessler
EECS Dept., BME Dept., Dept. of Radiology
University of Michigan

web.eecs.umich.edu/~fessler

AA OSA
Apr. 16, 2013
Disclosure

- Research support from GE Healthcare
- Research support to GE Global Research
- Supported in part by NIH grants R01 HL-098686 and P01 CA-87634
- Equipment support from Intel
Statistical image reconstruction: a CT revolution

- A picture is worth 1000 words
- (and perhaps several 1000 seconds of computation?)

Thin-slice FBP
- Seconds

ASIR
- A bit longer

Statistical
- Much longer

(Same sinogram, so all at same dose)
Outline

- **Image denoising** (review)

- **Image restoration**
  Antonios Matakos, Sathish Ramani, JF, IEEE T-IP 2013 (To appear)
  Accelerated edge-preserving image restoration without boundary artifacts

- **Low-dose X-ray CT image reconstruction**
  A splitting-based iterative algorithm for accelerated statistical X-ray CT reconstruction

- **Model-based MR image reconstruction**
  Parallel MR image reconstruction using augmented Lagrangian methods
Image denoising
Denoising using sparsity

Measurement model:

\[ \tilde{y} \quad \text{observed} \quad = \quad \tilde{x} \quad \text{unknown} \quad + \quad \tilde{\varepsilon} \quad \text{noise} \]

Object model: assume \( Qx \) is sparse (compressible) for some orthogonal sparsifying transform \( Q \), such as an orthogonal wavelet transform (OWT).

Sparsity regularized estimator:

\[ \hat{x} = \arg \min_x \frac{1}{2} \| y - x \|_2^2 + \beta \| Qx \|_p \]

Regularization parameter \( \beta \) determines trade-off.

Equivalently (because \( Q^{-1} = Q' \) is an orthonormal matrix):

\[ \hat{x} = Q'\hat{\theta}, \quad \hat{\theta} = \arg \min_\theta \frac{1}{2} \| Qy - \theta \|_2^2 + \beta \| \theta \|_p = \text{shrink}(Qy : \beta, p) \]

Non-iterative solution!

But sparsity in orthogonal transforms often yields artifacts.
Hard thresholding example

\[ p = 0, \text{ orthonormal Haar wavelets} \]
Sparsity using shift-invariant models

Analysis form:
Assume \(Rx\) is sparse for some sparsifying transform \(R\).
Often \(R\) is a “tall” matrix, e.g., finite differences along horizontal and vertical directions, i.e., anisotropic total variation (TV).
Often \(R\) is shift invariant: \(\|Rx\|_p = \|R \text{ circshift}(x)\|_p\) and \(R'R\) is circulant.

\[
\hat{x} = \arg \min_x \frac{1}{2} \|y - x\|_2^2 + \beta \|Rx\|_p.
\]

Synthesis form
Assume \(x = S\theta\) where coefficient vector \(\theta\) is sparse.
Often \(S\) is a “fat” matrix (over-complete dictionary) and \(S'S\) is circulant.

\[
\hat{x} = S\hat{\theta}, \quad \hat{\theta} = \arg \min_\theta \frac{1}{2} \|y - S\theta\|_2^2 + \beta \|\theta\|_p
\]

Analysis form preferable to synthesis form?
(Elad et al., Inv. Prob., June 2007)
Constrained optimization

Unconstrained estimator (analysis form for illustration):

$$\hat{x} = \arg \min_x \frac{1}{2} \|y - x\|_2^2 + \beta \|Rx\|_p.$$  

(Nonnegativity constraint or box constraints easily added.)

Equivalent **constrained** optimization problem:

$$\min_{x,v} \frac{1}{2} \|y - x\|_2^2 + \beta \|v\|_p \text{ sub. to } v = Rx.$$  

(Y. Wang *et al.*, SIAM J. Im. Sci., 2008)  
(M Afonso, J Bioucas-Dias, M Figueiredo, IEEE T-IP, Sep. 2010)

(The auxiliary variable $v$ is discarded after optimization; keep only $\hat{x}$.)

**Penalty** approach:

$$\hat{x} = \arg \min_x \min_v \frac{1}{2} \|y - x\|_2^2 + \beta \|v\|_p + \frac{\mu}{2} \|v - Rx\|_2^2.$$  

Large $\mu$ better enforces the constraint $v = Rx$, but can worsen conditioning.

Preferable (?) approach: augmented Lagrangian.
Augmented Lagrangian method: V1

General linearly constrained optimization problem:
\[
\min_u \Psi(u) \text{ sub. to } Cu = b.
\]

Form \textit{augmented Lagrangian}:
\[
L(u, \gamma) \triangleq \Psi(u) + \gamma^T (Cu - b) + \frac{\rho}{2} \|Cu - b\|_2^2
\]
where \(\gamma\) is the \textit{dual variable} or \textit{Lagrange multiplier vector}.

AL method alternates between minimizing over \(u\) and gradient ascent on \(\gamma\):
\[
\begin{align*}
    u^{(n+1)} &= \arg \min_u L(u, \gamma^{(n)}) \\
    \gamma^{(n+1)} &= \gamma^{(n)} + \rho (Cu^{(n+1)} - b).
\end{align*}
\]

Desirable convergence properties.
AL penalty parameter \(\rho\) affects convergence \textit{rate}, not solution!

Unfortunately, minimizing over \(u\) is impractical here:
\[
v = Rx \quad \text{equivalent to} \quad Cu = b, \quad C = [R \quad -I], \quad u = \begin{bmatrix} x \\ v \end{bmatrix} , \quad b = 0.
\]
Augmented Lagrangian method: V2

General linearly constrained optimization problem:

$$\min_u \Psi(u) \text{ sub. to } Cu = b.$$  

Form (modified) **augmented Lagrangian** by completing the square:

$$L(u, \eta) \triangleq \Psi(u) + \frac{\rho}{2} \|Cu - \eta\|_2^2 + C\eta,$$

where $\eta \triangleq b - \frac{1}{\rho} \gamma$ is a modified **dual variable** or **Lagrange multiplier vector**.

AL method alternates between minimizing over $u$ and gradient ascent on $\eta$:

$$u^{(n+1)} = \arg \min_u L(u, \gamma^{(n)})$$

$$\eta^{(n+1)} = \eta^{(n)} - (Cu^{(n+1)} - b).$$

Desirable convergence properties.

AL penalty parameter $\rho$ affects convergence **rate**, not solution!

Unfortunately, minimizing over $u$ is impractical here:

$$v = Rx \quad \text{equivalent to} \quad Cu = b, \quad C = [R \quad -I], \quad u = \begin{bmatrix} x \\ v \end{bmatrix}, \quad b = 0.$$
Alternating direction method of multipliers (ADMM)

When \( u \) has multiple component vectors, e.g., \( u = \begin{bmatrix} x \\ v \end{bmatrix} \), rewrite (modified) augmented Lagrangian in terms of all component vectors:

\[
L(x, v; \eta) = \Psi(x, v) + \frac{\rho}{2} \| Rx - v - \eta \|^2 \\
= \frac{1}{2} \| y - x \|^2 + \beta \| v \|_p + \frac{\rho}{2} \| Rx - v - \eta \|^2
\]

cf. penalty!

because here \( Cu = Rx - v \).

Alternate between minimizing over each component vector:

\[
x^{(n+1)} = \arg \min_x L(x, v^{(n)}, \eta^{(n)}) \\
v^{(n+1)} = \arg \min_v L(x^{(n+1)}, v, \eta^{(n)}) \\
\eta^{(n+1)} = \eta^{(n)} + (Rx^{(n+1)} - v^{(n+1)})
\]

Reasonably desirable convergence properties. (Inexact inner minimizations!) 

Sufficient conditions on matrix \( C \).


ADMM for image denoising

Augmented Lagrangian:

\[ L(x, v; \eta) = \frac{1}{2} \|y - x\|_2^2 + \beta \|v\|_p + \frac{\rho}{2} \|Rx - v - \eta\|_2^2 \]

Update of primal variable (unknown image):

\[ x^{(n+1)} = \arg \min_x L(x, v^{(n)}, \eta^{(n)}) = [I + \rho R'R]^{-1} (y + \rho R' (v^{(n)} + \eta^{(n)})) \]

Update of auxiliary variable: (No “corner rounding” needed for \(\ell_1\).)

\[ v^{(n+1)} = \arg \min_v L(x^{(n+1)}, v, \eta^{(n)}) = \text{shrink}(Rx^{(n+1)} - \eta^{(n)}; \beta / \rho, p) \]

Update of multiplier: \( \eta^{(n+1)} = \eta^{(n)} + (Rx^{(n+1)} - v^{(n+1)}) \)

Equivalent to “split Bregman” approach.

(Goldstein & Osher, SIAM J. Im. Sci. 2009)

Each update is simple and exact (non-iterative) if \([I + \rho R'R]^{-1}\) is easy.
ADMM image denoising example

$\mathbf{R} :$ horizontal and vertical finite differences (anisotropic TV), $p = 1$ (i.e., $\ell_1$), $\beta = 1/2$, $\rho = 1$ (condition number of $(\mathbf{I} + \rho \mathbf{R}'\mathbf{R})$ is 9)
ADMM image denoising iterates

admm denoising movie
Image restoration
Unrealistic model:

\[ y = Ax + x + \mathcal{E} \]

Measured blurry image \( y \) and unknown image \( x \) have the same size. \( A \) is a circulant matrix corresponding to a shift-invariant blur model.

Somewhat more realistic measurement model:

\[ y = TAx + \mathcal{E} \]

Measured blurry image \( y \) is smaller than unknown image \( x \). \( T \) is a (fat) “truncation” matrix, akin to \([0 \ I \ 0]\).

Image restoration with sparsity regularization

Regularized estimator:

\[ \hat{x} = \arg\min_x \frac{1}{2} \| y - TAx \|_2^2 + \beta \| Rx \|_p. \]

Basic equivalent constrained optimization problem:

\[ \min_{x,v} \frac{1}{2} \| y - TAx \|_2^2 + \beta \| v \|_p \text{ sub. to } v = Rx. \]

Corresponding (modified) augmented Lagrangian (cf. “split Bregman”):

\[ L(x, v; \eta) = \frac{1}{2} \| y - TAx \|_2^2 + \beta \| v \|_p + \frac{\rho}{2} \| Rx - v - \eta \|_2^2 \]

ADMM update of primal variable (unknown image):

\[ x^{(n+1)} = \arg\min_x L(x, v^{(n)}, \eta^{(n)}) = \left[ A'T'TA + \rho R'R \right]^{-1} (A'T'y + \rho R' (v^{(n)} + \eta^{(n)})) \]

Simple if \( A'A \) and \( R'R \) are circulant and \( T = I \) (unrealistic). Otherwise need iterative inner (quadratic) minimization: PCG.
Improved ADMM for image restoration

New equivalent constrained optimization problem:

\[
\min_{x,u,v} \frac{1}{2} \|y - Tu\|_2^2 + \beta \|v\|_p \quad \text{sub. to } v = Rx, \quad u = Ax.
\]


Corresponding (modified) augmented Lagrangian:

\[
L(x, u, v; \eta_1, \eta_2) = \frac{1}{2} \|y - Tu\|_2^2 + \beta \|v\|_p + \frac{\rho_1}{2} \|Rx - v - \eta_1\|_2^2 + \frac{\rho_2}{2} \|Ax - u - \eta_2\|_2^2
\]

ADMM update of primal variable (unknown image):

\[
\arg\min_x L(x, u, v, \eta_1, \eta_2) = \left[\rho_2A'\!\!A + \rho_1R'R\right]^{-1} \left(\rho_1R'(v + \eta_1) + \rho_2A'(u + \eta_2)\right)
\]

Simple if \(A'A\) and \(R'R\) are circulant. No inner iterations needed!

ADMM update of new auxiliary variable \(u\):

\[
\arg\min_u L(x, u, v, \eta_1, \eta_2) = \left[T'T + \rho_2I\right]^{-1} \left(T'y + \rho_2(Ax - \eta_2)\right)
\]

\(v\) update is shrinkage again. Very easy to code!
Image restoration results: quality

Measurement $y$

Using circulant model with boundary preprocessing

ADMM with Reeves model

$15 \times 15$ pixel uniform blur, $50\text{dB BSNR} = 10 \log \left( \frac{\text{Var} \{ TAx \}}{\sigma^2} \right)$, isotropic TV regularization, $\beta = 2^{-17}$

Qualitatively confirms Reeves model is preferable.
Image restoration results: iterations
Image restoration results: speed

Proposed ADMM is fast due to non-iterative inner updates.
X-ray CT image reconstruction
Low-dose X-ray CT image reconstruction

Regularized estimator:

\[ \hat{x} = \arg \min_{x \geq 0} \frac{1}{2} \| y - Ax \|_W^2 + \beta \| Rx \|_p. \]

Complications:

- \( A'A \) is not circulant (but “approximately Toeplitz” in 2D)
- \( A'W'A \) is highly shift variant due to huge dynamic range of weighting \( W \)
- Non-quadratic (edge-preserving) regularization \( \| \cdot \|_p \)
- Nonnegativity constraint
- Large problem size
Basic ADMM for X-ray CT

Basic equivalent constrained optimization problem (cf. split Bregman):

\[
\min_{x \geq 0, v} \frac{1}{2} \| y - Ax \|_W^2 + \beta \| v \|_p \quad \text{sub. to } v = Rx.
\]

Corresponding (modified) augmented Lagrangian (cf. “split Bregman”):

\[
L(x, v; \eta) = \frac{1}{2} \| y - Ax \|_W^2 + \beta \| v \|_p + \frac{\rho}{2} \| Rx - v - \eta \|_2^2
\]

ADMM update of primal variable (unknown image):

\[
x^{(n+1)} = \arg \min_x L(x, v^{(n)}, \eta^{(n)}) = \left[ A'WA + \rho R'R \right]^{-1} (A'Wy + \rho R' (v^{(n)} + \eta^{(n)}))
\]

- Ignoring nonnegativity constraint
- \( \left[ A'WA + \rho R'R \right]^{-1} \) requires iteration (e.g., PCG) but hard to precondition
- Auxiliary variable \( v = Rx \) is enormous in 3D CT
Improved ADMM for X-ray CT

\[
\min_{x \geq 0, u, v} \frac{1}{2} \| y - u \|_W^2 + \beta \| v \|_p \quad \text{sub. to } v = Rx, \quad u = Ax.
\]

Corresponding (modified) augmented Lagrangian:

\[
L(x, u, v; \eta_1, \eta_2) = \frac{1}{2} \| y - u \|_W^2 + \beta \| v \|_p + \frac{\rho_1}{2} \| Rx - v - \eta_1 \|_2^2 + \frac{\rho_2}{2} \| Ax - u - \eta_2 \|_2^2
\]

ADMM update of primal variable (ignoring nonnegativity):

\[
\arg \min_{x} L(x, u, v, \eta_1, \eta_2) = \left[ \rho_2 A'A + \rho_1 R'R \right]^{-1} \left( \rho_1 R' (v + \eta_1) + \rho_2 A'(u + \eta_2) \right)
\]

For 2D CT, \( \left[ \rho_2 A'A + \rho_1 R'R \right]^{-1} \) is approximately Toeplitz so a circulant preconditioner is very effective.

ADMM update of auxiliary variable \( u \):

\[
\arg \min_{u} L(x, u, v, \eta_1, \eta_2) = \left[ W + \rho_2 I \right]^{-1} \left( Wy + \rho_2 (Ax - \eta_2) \right)
\]

\( v \) update is shrinkage again. Reasonably simple to code.

(Sathish Ramani & JF, IEEE T-MI, Mar. 2012)
PWLS with $\ell_1$ regularization of shift-invariant Haar wavelet transform. No nonnegativity constraint, but probably unimportant if well-regularized.
Circulant preconditioner for $\left[\rho_2 A' A + \rho_1 R' R\right]^{-1}$ is crucial to acceleration.

Similar results for real head CT scan in paper.
Lower-memory ADMM for X-ray CT

\[
\min_{x,u,z \geq 0} \frac{1}{2} \|y - u\|^2_w + \beta \|Rz\|_p \quad {\text{sub. to}} \quad z = x, \quad u = Ax.
\]

(M McGaffin, S Ramani, JF, SPIE 2012)

Corresponding (modified) augmented Lagrangian:

\[
L(x, u, z; \eta_1, \eta_2) = \frac{1}{2} \|y - u\|^2_w + \beta \|Rz\|_p + \frac{\rho_1}{2} \|x - z - \eta_1\|^2_2 + \frac{\rho_2}{2} \|Ax - u - \eta_2\|^2_2
\]

ADMM update of primal variable (nonnegativity not required, use PCG):

\[
\arg \min_x L(x, u, z, \eta_1, \eta_2) = \left[\rho_2A'A + \rho_1I\right]^{-1} \left(\rho_1 (z + \eta_1) + \rho_2A'(u + \eta_2)\right).
\]

ADMM update of auxiliary variable \(z\):

\[
\arg \min_{z \geq 0} L(x, u, z, \eta_1, \eta_2) = \arg \min_{z \geq 0} \frac{\rho_1}{2} \|x - z - \eta_1\|^2_2 + \beta \|Rz\|_p.
\]

Use nonnegatively constrained, edge-preserving image denoising.

ADMM updates of auxiliary variables \(u\) and \(v\) same as before.

Variations...
3D X-ray CT image reconstruction results

Awaiting better preconditioner for $\left[ \rho_2 A^T A + \rho_1 I \right]^{-1}$
Image reconstruction for parallel MRI
Model-based image reconstruction in parallel MR

Undersampled Cartesian k-space, multiple receive coils, ...

Regularized estimator:

\[
\hat{x} = \arg\min_x \frac{1}{2} \|y - FSx\|_2^2 + \beta \|Rx\|_p.
\]

\(F\) is under-sampled DFT matrix (fat)

Features:

- coil sensitivity matrix \(S\) is block diagonal \(\text{(Pruessmann et al., MRM, Nov. 1999)}\)
- \(FF'\) is circulant

Complications:

- Data-fit Hessian \(S'F'FS\) is highly shift variant due to coil sensitivity maps
- Non-quadratic (edge-preserving) regularization \(\|\cdot\|_p\)
- Complex quantities
- Large problem size (if 3D)
Basic ADMM for parallel MRI

Basic equivalent constrained optimization problem (cf. split Bregman):

$$\min_{x,v} \frac{1}{2} \| y - FSx \|_2^2 + \beta \| v \|_p \quad \text{sub. to } v = Rx.$$ 

Corresponding (modified) augmented Lagrangian (cf. “split Bregman”):

$$L(x, v; \eta) = \frac{1}{2} \| y - FSx \|_2^2 + \beta \| v \|_p + \frac{\rho}{2} \| Rx - v - \eta \|_2^2$$

(Skipping technical details about complex vectors.)

ADMM update of primal variable (unknown image):

$$x^{(n+1)} = \arg \min_x L(x, v^{(n)}, \eta^{(n)}) = \left[ S'F'FS + \rho R'R \right]^{-1} \left( S'F'y + \rho R' \left( v^{(n)} + \eta^{(n)} \right) \right)$$

- $\left[ S'F'FS + \rho R'R \right]^{-1}$ requires iteration (e.g., PCG) but hard to precondition
- (Trivial for single coil case with $S = I$.)
- The “problem” matrix is on opposite side:
  - MRI: $FS$
  - Restoration: $TA$
Improved ADMM for parallel MRI

\[
\min_{x,u,v,z} \frac{1}{2} \|y - F u\|_2^2 + \beta \|v\|_p \quad \text{sub. to} \quad v = Rz, \quad u = Sx, \quad z = x.
\]

Corresponding (modified) augmented Lagrangian:

\[
\frac{1}{2} \|y - F u\|_2^2 + \beta \|v\|_p + \frac{\rho_1}{2} \|Rz - v - \eta_1\|_2^2 + \frac{\rho_2}{2} \|Sx - u - \eta_2\|_2^2 + \frac{\rho_3}{2} \|x - z - \eta_3\|_2^2
\]

ADMM update of primal variable

\[
\arg\min_x L(x, u, v, z; \eta_1, \eta_2, \eta_3) = \left[\rho_2 S'S + \rho_3 I\right]^{-1} (\rho_2 S'(u + \eta_2) + \rho_3 (z + \eta_3))
\]

ADMM update of auxiliary variables:

\[
\arg\min_u L(x, u, v, z; \eta_1, \eta_2, \eta_3) = \left[F'F + \rho_2 I\right]^{-1} (F'y + \rho_2 (Sx - \eta_2))
\]

\[
\arg\min_z L(x, u, v, z; \eta_1, \eta_2, \eta_3) = \left[\rho_1 R'R + \rho_3 I\right]^{-1} (\rho_1 R'(v + \eta_1) + \rho_3 (x - \eta_3))
\]

\(v\) update is shrinkage again.

Simple, but does not satisfy sufficient conditions.

(Sathish Ramani & JF, IEEE T-MI, Mar. 2011)
2.5D parallel MR image reconstruction results: data

Fully sampled body coil image of human brain
Poisson-disk-based k-space sampling, 16% sampling (acceleration 6.25)
Square-root of sum-of-squares inverse FFT of zero-filled k-space data
2.5D parallel MR image reconstruction results: IQ

- Fully sampled body coil image of human brain
- Regularized reconstruction $x^{(\infty)}$ (1000s of iterations of MFISTA)
  (A Beck & M Teboulle, SIAM J. Im. Sci, 2009)
- Combined TV and $\ell_1$ norm of two-level undecimated Haar wavelets
- Difference image magnitude
2.5D parallel MR image reconstruction results: speed

AL approach converges to $x^{(\infty)}$ much faster than MFISTA and CG
Current and future directions with ADMM

- Motion-compensated image reconstruction: \( y = A T(\alpha)x + \epsilon \)
  (J H Cho, S Ramani, JF, 2nd CT meeting, 2012)
- Dynamic image reconstruction
- Improved preconditioners for ADMM for 3D CT
  (M McGaffin and JF, Submitted to Fully 3D 2013)
- Combining ADMM with ordered subsets (OS) methods
  (H Nien and JF, Submitted to Fully 3D 2013)
- Generalize parallel MRI algorithm to include spatial support constraint
  (M Le, S Ramani, JF, To appear at ISMRM 2013)
- Non-Cartesian MRI (combine optimization transfer and variable splitting)
  (S Ramani and JF, ISBI 2013, to appear.)
- SPECT-CT reconstruction with non-local means regularizer
  (S Y Chun, Y K Dewaraja, JF, Submitted to Fully 3D 2013)
- Estimation of coil sensitivity maps (quadratic problem!)
- L1-SPIRiT for non-Cartesian parallel MRI
- Multi-frame super-resolution
- Selection of AL penalty parameter \( \rho \) to optimize convergence rate

- Other non-ADMM methods...
Acknowledgements

CT group
- Jang Hwan Cho
- Donghwan Kim
- Jungkuk Kim
- Madison McGaffin
- Hung Nien
- Stephen Schmitt

MR group
- Michael Allison
- Mai Le
- Antonis Matakos
- Matthew Muckley

Post-doctoral fellows
- Se Young Chun
- Sathish Ramani
- Daniel Weller

UM collaborators
- Doug Noll
- Jon-Fredrik Nielsen
- Mitch Goodsitt
- Ella Kazerooni
- Tom Chenevert
- Charles Meyer

GE collaborators
- Bruno De Man
- Jean-Baptiste Thibault
Bibliography


