A Class of Fast Algorithms for TV Based Image Reconstruction

Yin Zhang

Department of CAAM

Rice University

Joint work with: Junfeng Yang, Yilun Wang and Wotao Yin

Outline

- Introduction
  - Image formation equation
  - Maximum likelihood estimation
  - Maximum *a posteriori* estimation
  - Regularization
- A fast alternating algorithm
  - Motivation and algorithm
  - Relation with half-quadratic technique
  - Optimality and convergence results
- Numerical results and extensions

Introduction

- Image formation equation.

$$f = K\bar{u} + \omega$$

  - $\bar{u}$: original image
  - K: convolution operator
  - ω: random noise
  - f: observation

Our purpose is to recover $\bar{u}$ from f (deconvolve and denoise) as well as possible.

- Deconvolution is severely ill-conditioned. Let $\mathbf{F}$ be the 2D Fourier transform matrix. The equation is equivalent to

$$\hat{f} = \hat{K}\hat{\hat{u}} + \hat{\omega},$$

where $\hat{f} = \mathbf{F}f$ and $\hat{K} = \mathbf{F}K\mathbf{F}^{-1}$(diagonal). A tempting solution would be

$$u^{\mathrm{direct}} = \mathbf{F}^{-1}(\hat{K}^{-1}\hat{f}) = \mathbf{F}^{-1}(\hat{\hat{u}} + \hat{K}^{-1}\hat{\omega}).$$

Does this work?

Experiment 1. Blur: ('gaussian',11,5); Noise: $\mathcal{N}(0, 10^{-8})$.

Noise is amplified!

Cut off high frequencies (Weiner Filter):

$$\hat{u}_i = \begin{cases} 0, & \text{if } |\hat{f}_i/\hat{K}_{ii}| > M; \\ \hat{f}_i/\hat{K}_{ii}, & \text{otherwise.} \end{cases}$$

Result of experiment 1 after cutting off some high frequencies:

Cut off high frequencies

Statistics Interpretation:

- Maximum likelihood estimation. Given $\omega \sim \mathcal{N}(0, \sigma^2)$, the MLE of $\bar{u}$ is

$$
\begin{aligned}
u^{\mathrm{MLE}} &= \arg\max_u \Pr\{f|u\} \\
&= \arg\min_u \left(-\log(\Pr\{f|u\})\right) \\
&= \arg\min_u \|Ku - f\|^2.
\end{aligned}
$$

Thus, MLE, LS and direct inverse are all equivalent. They do not work. When noise is correlated, i.e., $\omega \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma})$, MLE becomes weighted LS.

Another Statistics Viewpoint:

- Maximum *a posteriori* estimation. Given $\omega \sim \mathcal{N}(0, \sigma^2)$, the MAP of $\bar{u}$ is

$$
\begin{aligned}
u^{\mathrm{MAP}} &= \arg\max_u \Pr\{u|f\} \\
&= \arg\max_u \frac{\Pr\{u\}\Pr\{f|u\}}{\Pr\{f\}} \\
&= \arg\min_u \{-\log(\Pr\{u\}) - \log(\Pr\{f|u\})\} \\
&= \arg\min_u \Phi_{\mathrm{prior}}(u) + \|Ku - f\|^2.
\end{aligned}
$$

Thus, $\Phi_{\mathrm{prior}}(u)$ enforces some prior constraints on $\bar{u}$, which is called regularization. Qusetion: what kind of *prior* do we need?

- Regularization.

$$\min_u \Phi_{\text{reg}}(u) + \mu \|Ku - f\|_2^2$$

  - Tikhonov-like regularization (notice 2-norm squared)

$$\Phi_{\text{reg}}(u) = \Phi_{\text{Tik}}(u) \triangleq \sum_{j \in J} \|D^{(j)} u\|_2^2,$$

    for some $J \subset \{0, 1, 2, \ldots\}$, where

    * $D^{(0)}$: identity matrix
    * $D^{(j)}, j = 1, 2$: the 1st order finite difference matrices
    * $D^{(j)}, j = 3, 4, 5$: the 2nd order . . . (used by MATLAB “deconvreg”).

    The solution satisfies

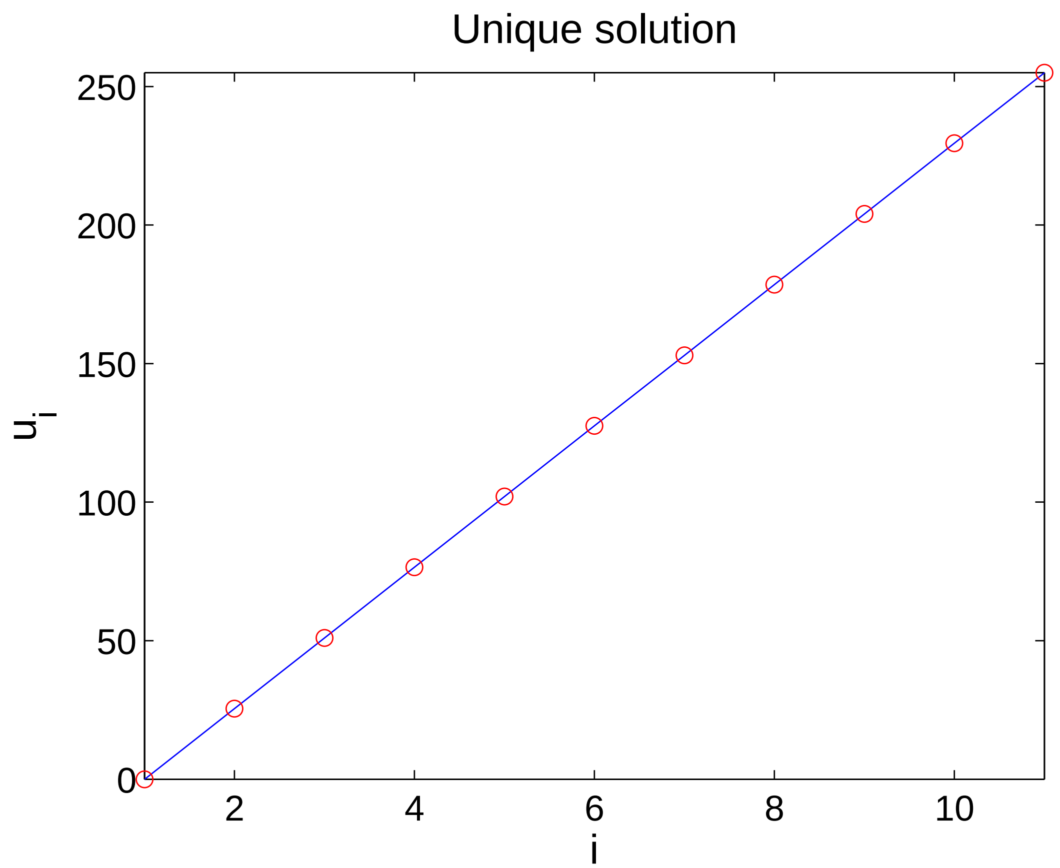
$$\left(\sum_{j \in J} (D^{(j)})^\top D^{(j)} + \mu K^\top K \right) u = \mu K^\top f.$$

Experiment 2. Result of Tikhonov regularization. Blur: ('gaussian',21,11); Noise: $\mathcal{N}(0, 10^{-6})$.

Advantages: Not so sensitive to noise, easy to compute.

Disadvantage of Square: Incapable of recovering image discontinuities.

$$\min_{u\in R^{11}} \phi(u) = \sum_i |u_{i+1} - u_i|^2, \text{ s.t. } u_1 = 0, u_{11} = 255.$$



– Total variation regularization (Rudin, Osher and Fatemi, 1992).

$$\Phi_{\text{reg}}(u) = \text{TV}(u) \triangleq \sum_i \|D_i u\|.$$

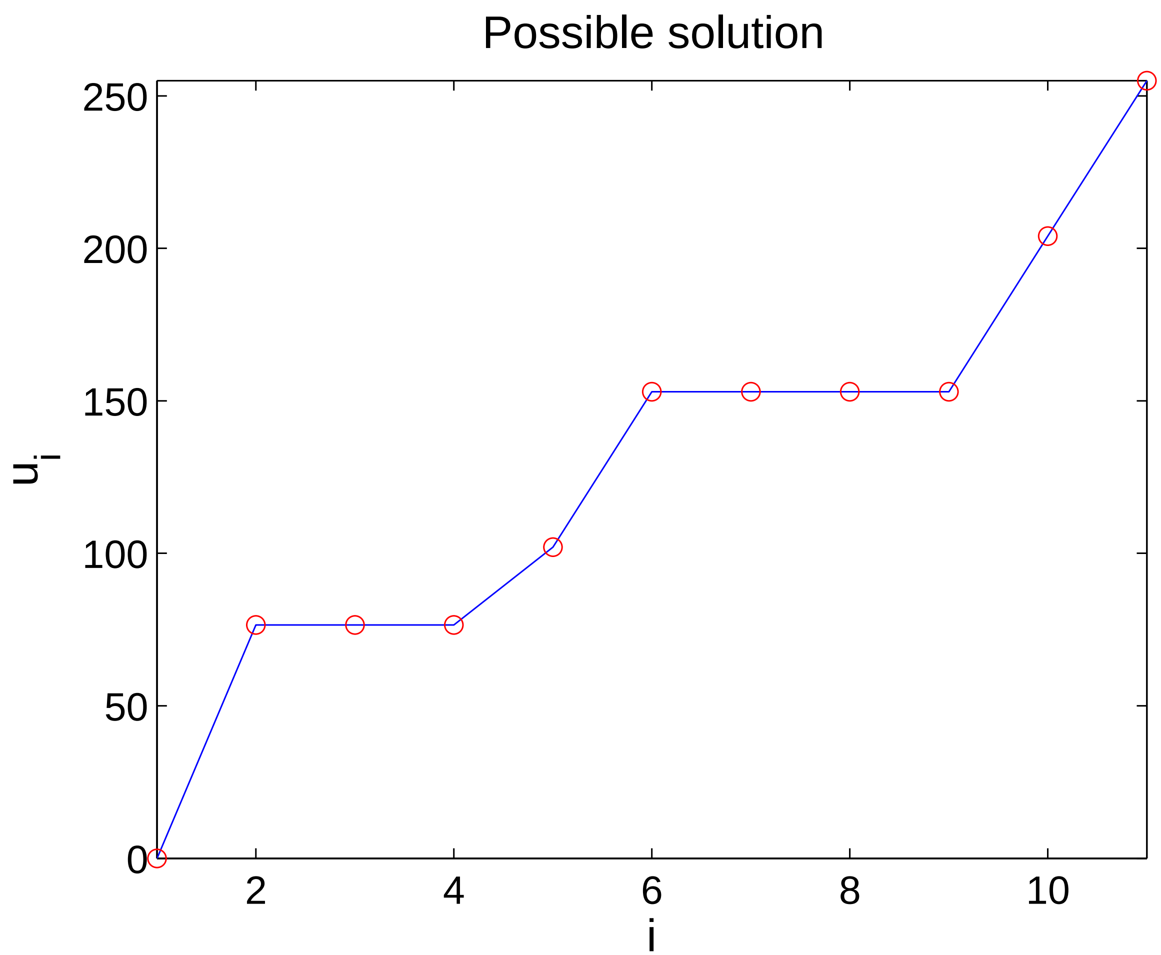
* $\|D_i u\|$: the variation of u at pixel i, where

$$D_i u = \begin{pmatrix} (D^{(1)}u)_i \\ (D^{(2)}u)_i \end{pmatrix} \in \mathbf{R}^2.$$

* $\sum_i$ is taken over all pixels.
* The sum represents a 1-norm.
* $\| \cdot \|$: the 2-norm (isotropic) or the 1-norm (anisotropic).

Advantage of 1-norm: Permits sharp edges in images.

$$\min_{u \in R^{11}} \mathrm{TV}(u) = \sum_i |u_{i+1} - u_i|, \text{ s.t. } u_1 = 0, u_{11} = 255.$$



Experiment 3. Compare Tikhonov with TV regularization. The same inputs as in experiment 2.

Disadvantages: More expensive in computation, stair-casing effect.

$$TV/L^2: \quad \min_u \sum_i \|D_i u\|_2 + \frac{\mu}{2}\|Ku - f\|^2.$$

It's a convex program, large-scale, still ill-conditioned and requires "real-time" processing.

- Some existing methods.
  - Lagged diffusivity method (Vogel & Oman, 1995). Given u^k, u^{k+1} is determined by solving

$$\sum_i D_i^\top \frac{D_i u}{\|D_i u^k\|_\alpha} + \mu K^\top (Ku - f) = 0,$$

which is a linearization to the optimality condition of

$$\min_u \sum_i \|D_i u\|_\alpha + \frac{\mu}{2}\|Ku - f\|^2.$$

Here $\|\cdot\|_\alpha \triangleq \sqrt{\|\cdot\|^2 + \alpha}$ for some small $\alpha > 0$.
Most earlier methods were based on solving (Euler-Langrange) PDE.

– Iterative Shrinkage/Thresholding based methods (Daubechies, Defrise & De Mol, 2004). Given u_k, the original IST method iterates as

$$u_{k+1} = \Psi_\mu \left(u_k - \lambda_k K^\top (K u_k - f) \right),$$

where $\lambda_k > 0$ and

$$\Psi_\mu(\xi) \triangleq \arg\min_u \mathrm{TV}(u) + \frac{\mu}{2} \|u - \xi\|^2.$$

There exist several variants of IST methods, e.g., TwIST (Bioucas-Dias & Figueiredo, 2007).

– Second-order cone programming approach (Goldfarb & Yin, 2005).

– Iterative Denoising (Michael Ng *et al* 2007). Much faster, but ......

A Fast Alternating Algorithm

- Motivation. The problem is

$$\min_u \sum_i \|D_i u\| + \frac{\mu}{2}\|Ku - f\|^2.$$

By introducing $\mathbf{w}_i \in \mathbb{R}^2$, TV/L^2 is approximated by, for $\beta \gg 0$,

$$\min_{\mathbf{w}_i, u} \sum_i \left(\|\mathbf{w}_i\| + \frac{\beta}{2}\|\mathbf{w}_i - D_i u\|^2 \right) + \frac{\mu}{2}\|Ku - f\|^2.$$

The approximation problem allows very fast alternating minimization.

A simple and important lemma:

Lemma 1 *Given a positive integer d. For any $\beta > 0$ and $\mathbf{t} \in \mathbf{R}^d$, it holds*

$$\max\left\{\|\mathbf{t}\| - \frac{1}{\beta}, 0\right\} \frac{\mathbf{t}}{\|\mathbf{t}\|} = \arg\min_{\mathbf{s}\in\mathbf{R}^d} \left\{\|\mathbf{s}\| + \frac{\beta}{2}\|\mathbf{s}-\mathbf{t}\|^2\right\},$$

where we follow the convention $0 \cdot (0/0) = 0$.

An important Observation: Finite differences, $D^{(1)}$ and $D^{(2)}$ can be treated as discrete convolution under suitable boundary conditions.

Consequently, $D^{(1)}$ and $D^{(2)}$ and K are circulant matrices under the periodic boundary conditions for u, and all can be diagonalized by FFT.

- Our Simple Algorithm:
  - w-subproblem. Fixing u, minimizing w.r.t. $\mathbf{w}$ reduces to

$$\min_{\mathbf{w}_i} \|\mathbf{w}_i\| + \frac{\beta}{2}\|\mathbf{w}_i - D_i u\|^2, \quad \forall i.$$

Separate and closed form solutions at all pixels i:

$$\mathbf{w}_i = \max\left\{\|D_i u\| - \frac{1}{\beta}, 0\right\} \frac{D_i u}{\|D_i u\|}, \quad \forall i.$$

Linear time complexity: $O(n^2)$.

– u-subproblem. Fixing $\{\mathbf{w}_i\}$, minimizing w.r.t. u reduces to

$$\min_u \frac{\beta}{2}\sum_i \|\mathbf{w}_i - D_i u\|^2 + \frac{\mu}{2}\|Ku - f\|^2.$$

Its normal equations are

$$\left(\sum_i D_i^\top D_i + \frac{\mu}{\beta}K^\top K\right)u = \sum_i D_i^\top \mathbf{w}_i + \frac{\mu}{\beta}K^\top f$$

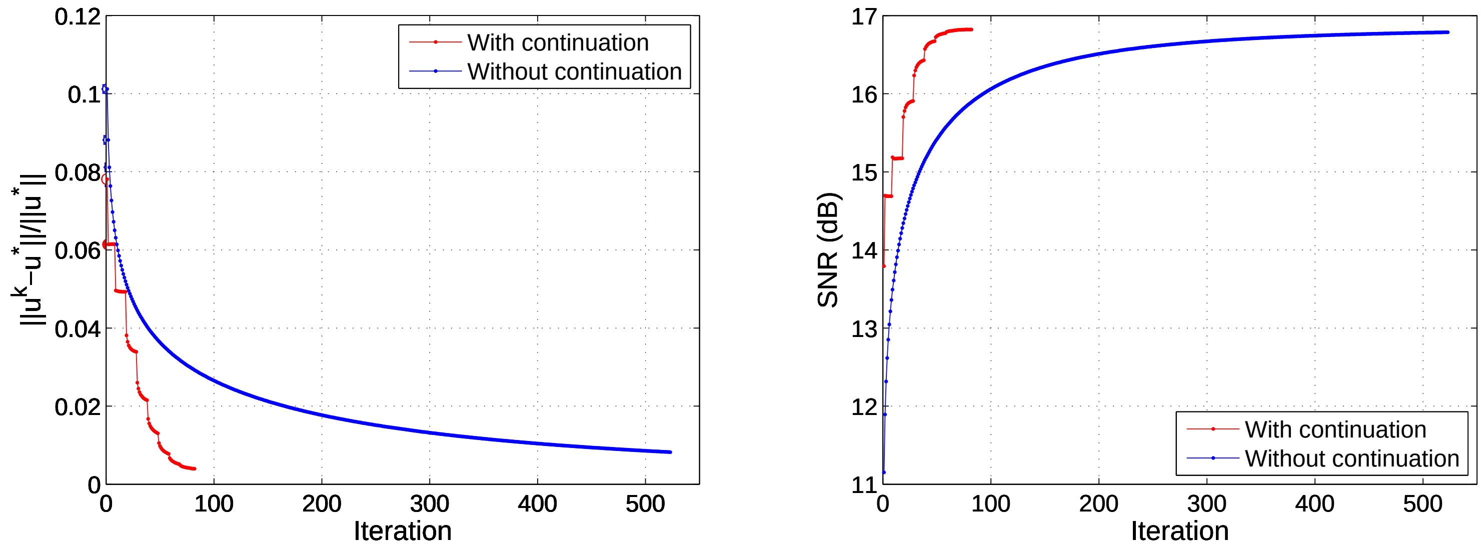
or equivalently

$$\left(\sum_{j=1}^{2}(D^{(j)})^\top D^{(j)} + \frac{\mu}{\beta}K^\top K\right)u = \sum_{j=1}^{2}(D^{(j)})^\top w_j + \frac{\mu}{\beta}K^\top f,$$

where $w_j = \{\mathbf{w}_i(j) : i = 1, \ldots, n^2\}$ for $j = 1, 2$.

This system can be solved by 2 FFTs at a cost of $O(n^2 \log n)$.

– Continuation/path-following. Initialize β small, and then increase it gradually. The previous solution is used to warm-start the next problem.



Test on continuation: $\beta = 2^0, 2^1, \ldots, 2^{10}$.

Continuation not only accelerates the speed, but also, unexpectedly, enhances solution robustness.

Given $\beta > 0$, we solve the approximation problem by alternately minimizing w.r.t. $\mathbf{w}$ and u.

– **FTVd** (Fast TV deconvolution). Input K, f, $\mu > 0$, $\beta_{max} \gg 0$ and $\gamma > 1$; Initialize $\beta = \beta_0 > 0$ and $u = u_0$.

While $\beta <= \beta_{max}$, **Do**

1) Solve the approximation to certain accuracy for u_β.

2) Update $u \leftarrow u_\beta$, $\beta \leftarrow \gamma * \beta$.

End Do

- Relation with half-quadratic technique. Given $\beta > 0$, FTVd solves

$$\min_{u,\mathbf{w}} \sum_i \left\{ \|\mathbf{w}_i\| + \frac{\beta}{2}\|D_i u - \mathbf{w}_i\|^2 \right\} + \frac{\mu}{2}\|Ku - f\|^2.$$

The above is equivalent to

$$\min_u \sum_i \phi\left(D_i u\right) + \frac{\mu}{2}\|Ku - f\|^2,$$

where $\phi(\mathbf{t})$, $\mathbf{t} \in \mathbf{R}^2$, is defined as

$$\phi(\mathbf{t}) = \begin{cases} \frac{\beta}{2}\|\mathbf{t}\|^2, & \text{if } \|\mathbf{t}\| \leq 1/\beta; \\ \|\mathbf{t}\| - \frac{1}{2\beta}, & \text{otherwise.} \end{cases}$$

This is an extension to the half-quadratic transform (German and Yang 1995).

- Optimality. A pair $(\mathbf{w}, u)$ solves the approximation problem iff

$$\begin{cases} \mathbf{w}_i/\|\mathbf{w}_i\| + \beta(\mathbf{w}_i - D_i u) = 0 & i \in I_1 \triangleq \{i : \mathbf{w}_i \neq \mathbf{0}\}, \\ \beta\|D_i u\| \leq 1 & i \in I_2 \triangleq \{i : \mathbf{w}_i = \mathbf{0}\}, \end{cases}$$

$$\beta D^\top (Du - w) + \mu K^\top (Ku - f) = 0.$$

Eliminating $\mathbf{w}$, the final equations become

$$\sum_{i \in I_1} D_i^\top \frac{D_i u}{\|D_i u\|} + \sum_{i \in I_2} D_i^\top h_i + \mu K^\top (Ku - f) = 0,$$

where $h_i = \beta D_i u$ satisfies $\|h_i\| \leq 1$, which is an approximation to the optimality condition of TV/L^2.

- Convergence results. Let $D = \left(D^{(1)}; D^{(2)}\right)$,

$$M = D^\top D + (\mu/\beta) \cdot K^\top K \text{ and } T = DM^{-1}D^\top.$$

Assuming $\mathcal{N}(D) \cap \mathcal{N}(K) = \{0\}$, for fixed β we have

1. The sequence $\{(w^k, u^k)\}$ generated by FTVd converges to a solution (w^*, u^*) of the approximation problem.
2. Finite convergence. $\mathbf{w}_L^k \equiv \mathbf{w}_L^*$ in finite number of iterations.
3. q-linear convergence. For k sufficiently large, there hold
   (a) $\|D(u^{k+1} - u^*)\| \leq \sqrt{\|(T^2)_{EE}\|} \cdot \|D(u^k - u^*)\|$;
   (b) $\|w^{k+1} - w^*\| \leq \sqrt{\|(T^2)_{EE}\|} \cdot \|w^k - w^*\|$;
   (c) $\|u^{k+1} - u^*\|_M \leq \sqrt{\|T_{EE}\|} \cdot \|u^k - u^*\|_M$.

Here $L = \{i, \|D_i u^*\| < 1/\beta\}$ and $E = \{1, 2, \ldots, n^2\} \setminus L$.

Numerical results and extensions

- Restoration of grayscale images. Kernel: ('gaussian',21,10); Noise: Gaussian white with mean zero and `std`$= 10^{-3}$.

Blurry&Noisy. SNR: 7.7dB

Lag.D. SNR: 14.4dB, CPU: 1918.0s

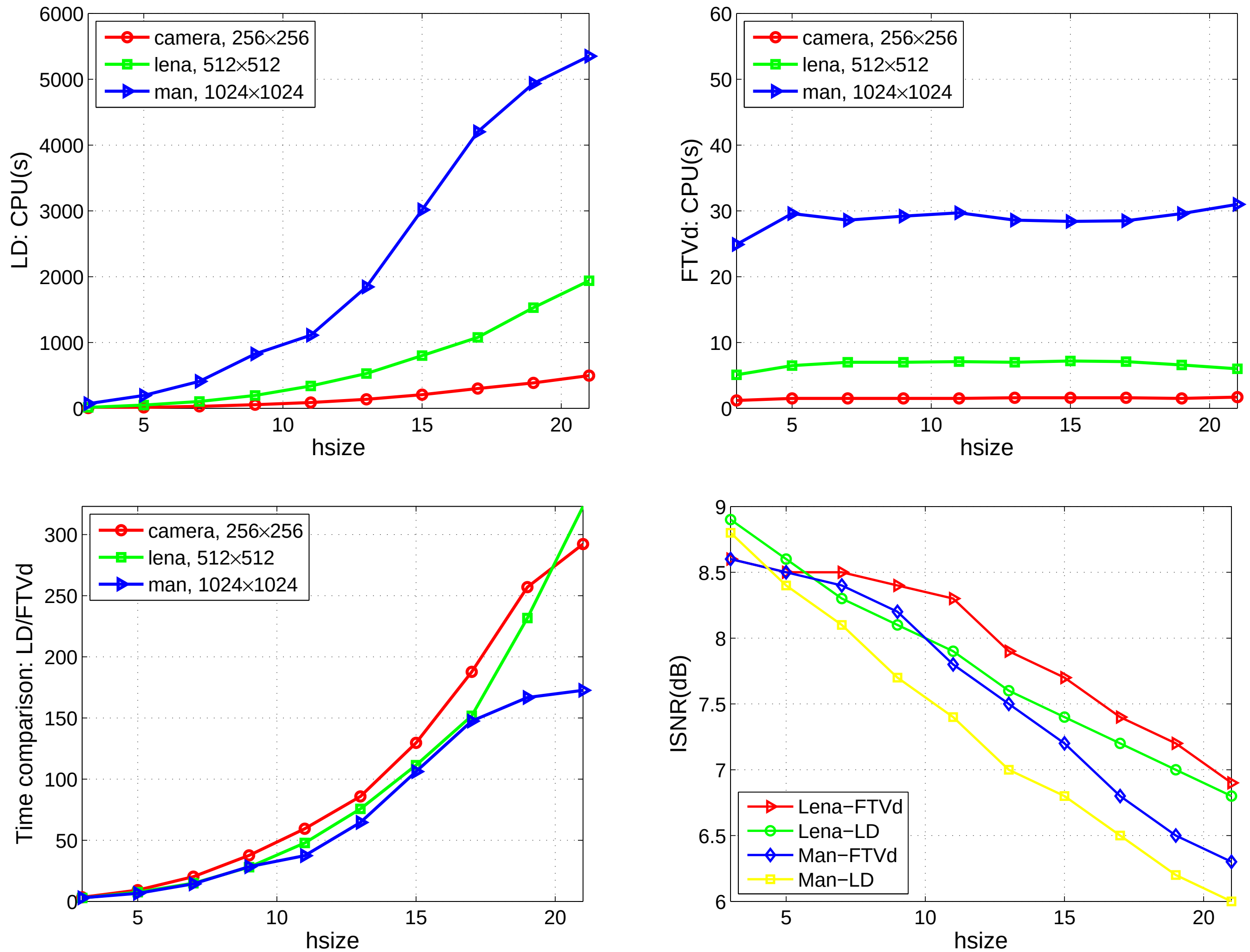
FTVd. SNR: 14.8dB, Iter: 9, CPU: 7.0s

Blurry&Noisy. SNR: 9.1dB

Lag.D. SNR: 15.0dB, CPU: 7306.7s

FTVd. SNR: 15.5dB, Iter: 10, CPU: 31.2s

- Speed comparison with Lagged Diffusivity method. Noise: Gaussian, mean zero and $\texttt{std} = 10^{-3}$; Blur: ('gaussian',hsize,10).



- Multichannel image deconvolution. Let u be a RGB image. The image formulation equation $f = Ku + \omega$ becomes

$$\begin{pmatrix} f^r \\ f^g \\ f^b \end{pmatrix} = \begin{pmatrix} K_{rr} & K_{rg} & K_{rb} \\ K_{gr} & K_{gg} & K_{gb} \\ K_{br} & K_{bg} & K_{bb} \end{pmatrix} \begin{pmatrix} u^r \\ u^g \\ u^b \end{pmatrix} + \begin{pmatrix} \omega^r \\ \omega^g \\ \omega^b \end{pmatrix}.$$

TV is extended to

$$\mathrm{MTV}(u) \triangleq \sum_i \|(I_3 \otimes D_i)u\|,$$

where

$$(I_3 \otimes D_i)u = \left[D^{(1)}u^r, D^{(2)}u^r, D^{(1)}u^g, D^{(2)}u^g, D^{(1)}u^b, D^{(2)}u^b\right]_i \in \mathbf{R}^6.$$

Generally, let $u \in \mathbf{R}^{mn^2}$ be a m-channel image and $K = [K_{jk}]_{jk=1}^{m}$ be a cross-channel blurring matrix. The TV/L^2 model is extended as

$$\min_u \sum_i \alpha_i \|G_i u\| + \frac{\mu}{2}\|Ku - f\|^2,$$

where $G_i = I_m \otimes \mathcal{D}_i$, and $\mathcal{D}_i$ is a 1st and/or higher order local finite difference operator. It is approximated by

$$\min_{u,\mathbf{w}} \sum_i \left(\alpha_i \|\mathbf{w}_i\| + \frac{\beta}{2}\|\mathbf{w}_i - G_i u\|^2 \right) + \frac{\mu}{2}\|Ku - f\|^2.$$

– Fixing u, the minimizer function for $\mathbf{w}$ is given explicitly by:

$$\mathbf{w}_i = \max\left\{ \|G_i u\| - \frac{\alpha_i}{\beta}, 0 \right\} \frac{G_i u}{\|G_i u\|}, \quad \forall i.$$

– The u-subproblem is equivalent to

$$\left(\sum_j (G^{(j)})^\top G^{(j)} + \frac{\mu}{\beta} K^\top K\right) u = \sum_j (G^{(j)})^\top w_j + \frac{\mu}{\beta} K^\top f.$$

By pre- and post- multiplying $I_m \otimes \mathbf{F}$ and its inverse, respectively, the coefficient matrix becomes

$$\begin{pmatrix} \Lambda_{11} & \Lambda_{12} & \dots & \Lambda_{1m} \\ \Lambda_{21} & \Lambda_{22} & \dots & \Lambda_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ \Lambda_{m1} & \Lambda_{m2} & \dots & \Lambda_{mm} \end{pmatrix},$$

with each Λ_{ij} a diagonal matrix. Thus u-subproblem is easily solved by FFTs and low complexity Gaussian elimination.

Restoration from cross-channel blur and Gaussian noise:

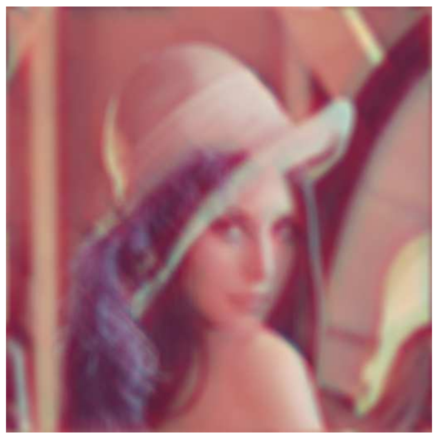
Original

Blurry&Noisy. SNR: 6.70dB

FTVd: SNR: 18.49dB, t = 4.29s

Original

Blurry&Noisy. SNR: 8.01dB



FTVd: SNR: 19.54dB, t = 16.86s

- Deconvolution in the presence of impulsive noise. Cameraman degraded by convolution and 10% salt-and-pepper noise. Right: solution of TV/L^2.

For impulsive noise, the ℓ_1-norm fidelity is more suitable. We recover $\bar{u}$ as the solution of the TV/L^1 model:

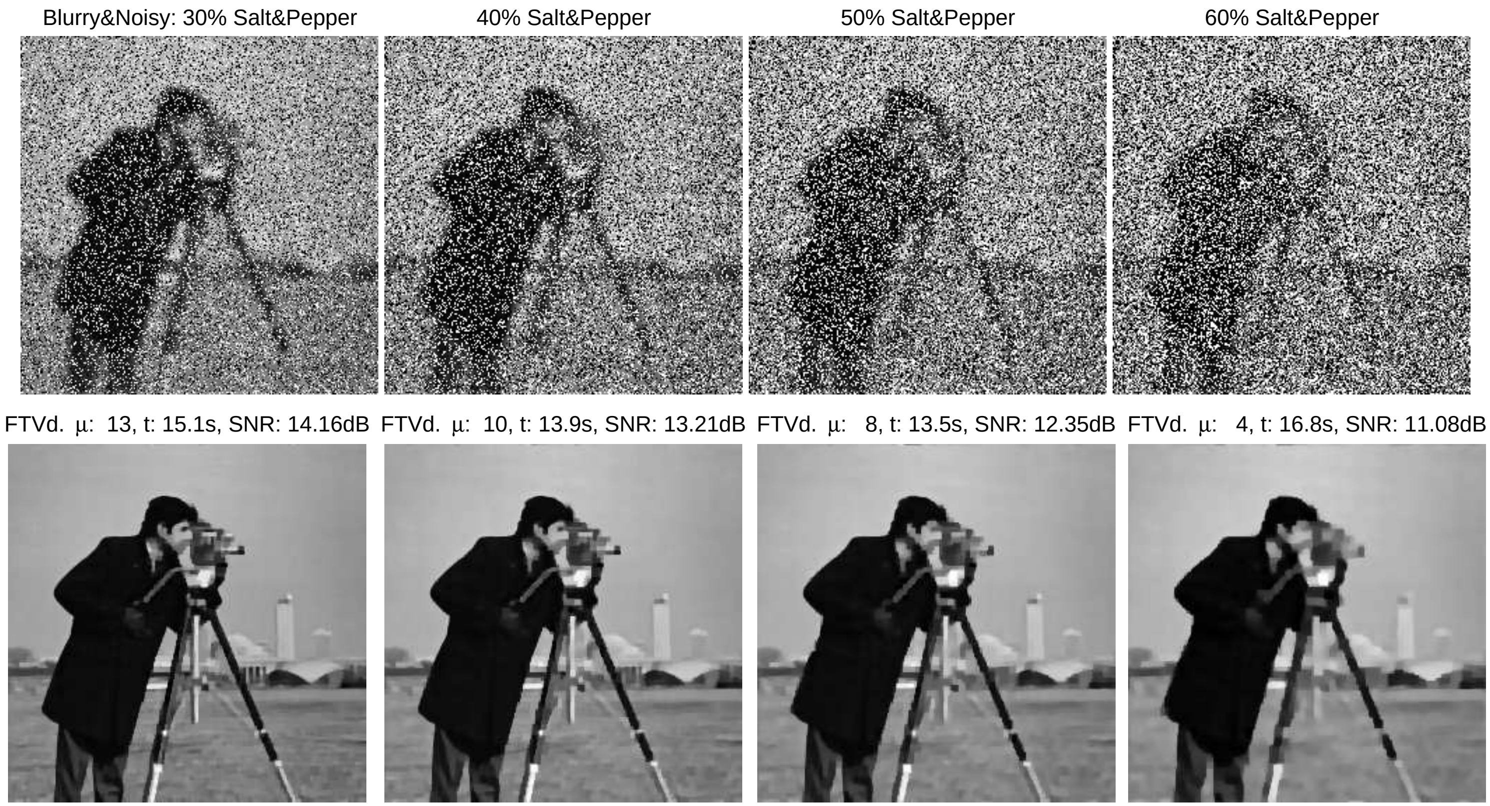
$$\min_u \sum_i \|D_i u\| + \mu \|Ku - f\|_1.$$

The approximation problem is given by

$$\min_{\mathbf{w},z,u} \sum_i \left(\|\mathbf{w}_i\| + \frac{\beta}{2} \|\mathbf{w}_i - D_i u\|^2 \right) + \mu \left(\|z\|_1 + \frac{\gamma}{2} \|z - (Ku - f)\|^2 \right).$$

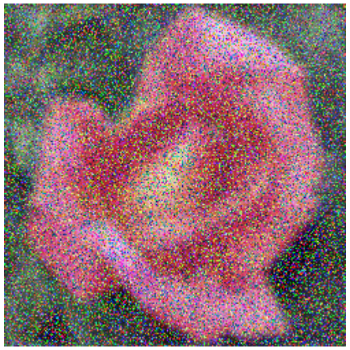
Minimization w.r.t. $\mathbf{w}$, z and u each is easy!

Restoration from Gaussian blur and salt-and-pepper noise:

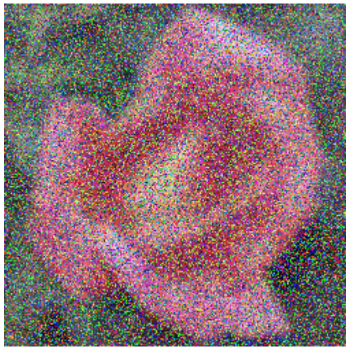


Restoration from cross-channel blur and random-valued noise:

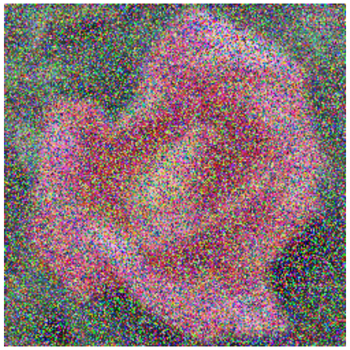
40% RV



50% RV



60% RV



μ: 8, t: 117s, SNR: 16.04dB

μ: 4, t: 138s, SNR: 14.06dB

μ: 2, t: 136s, SNR: 10.60dB

- MRI reconstruction. In MR imaging system, MR scanner collects data:

$$f_p = \mathcal{F}_p \bar{u} + \omega \in \mathcal{C}^M, \quad M \ll N.$$

Without noise, under certain desirable conditions, it holds

$$\bar{u} = \arg\min_u \left\{ \mathrm{TV}(u) : \mathcal{F}_p u = f_p \right\}.$$

In the presence of noise, we recover $\bar{u}$ via

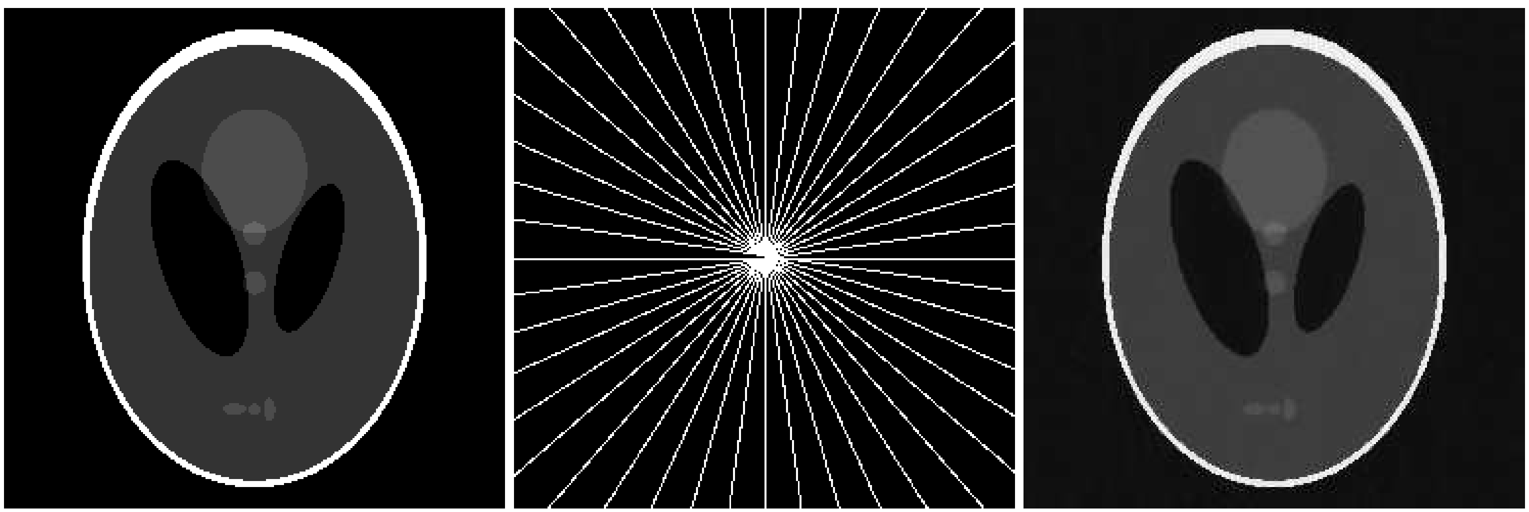
$$\min_u \mathrm{TV}(u) + \frac{\mu}{2}\|\mathcal{F}_p u - f_p\|^2.$$

When $\bar{u}$ has sparse/compressible representation under certain wavelet basis, we recover it via

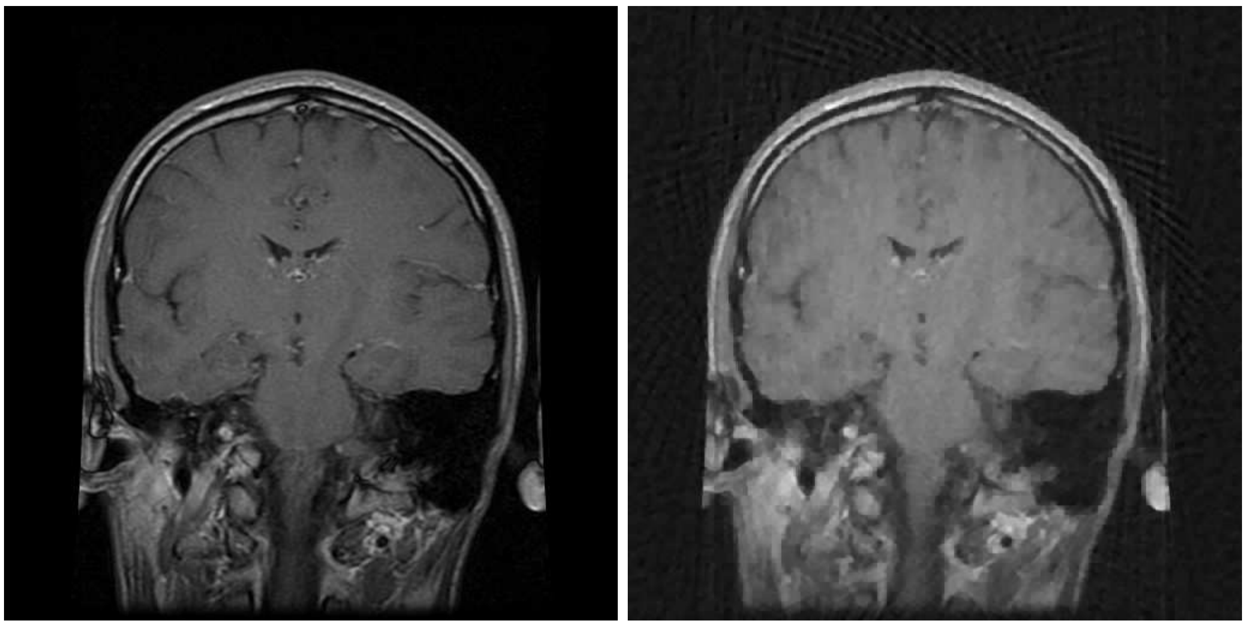
$$\min_u \mathrm{TV}(u) + \tau\|\Psi^\top u\|_1 + \frac{\mu}{2}\|\mathcal{F}_p u - f_p\|^2.$$

FTVd can be extended to solve the above TVL^1-L^2 problem.

Sparse (under TV) image reconstruction. Left to right: Original, Fourier domain samples (9.36%), reconstructed image (RelErr: 4.48%). Gaussian noise with mean zero and std$= .01$.



Compressible (under wavelet) image reconstruction. Sample ratio: 9.64%; Noise: Gaussian, mean zero, $\texttt{std} = .01$; Left: original brain image; Right: reconstructed (RelErr: 11.58%).



- Summary.
  - FTVd converges without the assumption of strictly convexity.
  - Finite convergence of auxiliary variables is established.
  - Linear convergence rate is established and the convergence factor depends on a submatrix.
  - FTVd is fast for TV based problem because it fully exploits problem structure and utilizes FFT.

References

[1] L. Rudin and S. Osher and E. Fatemi, *Nonlinear total variation based noise removal algorithms*, Phys. D, vol.60, 259-268, 1992.

[2] C. R. Vogel and M. E. Oman, *Fast total variation based image reconstruction*, Proc. ASME design engineering conferences, **3**, 1009-1015, 1995.

[3] D. Goldfarb and W. Yin, *Second-order cone programming methods for total variation-based image restoration*, SIAM J. Sci. Comput., vol.27, **2**, 622-645, 2005.

[4] I. Daubechies, M. Defriese, and C. De Mol, *An iterative thresholding algorithm for linear inverse problems with a sparsity constraint*, Commun. Pure Appl. Math., vol. LVII, pp. 14131457, 2004.

[5] J. Bioucas-Dias, and M. Figueiredo, *A new TwIST: Two-step iterative thresholding algorithm for image restoration*, IEEE Trans. Imag. Process., vol. 16, no. 12, pp. 2992–3004, 2007.

[6] Y. Wang, J. Yang, W. Yin and Y. Zhang, *A new alternating minimization algorithm for total variation image reconstruction*, SIAM J. Imag. Sci., vol. 1, no. 3, 248–272, 2008.

[7] J. Yang, W. Yin, Y. Zhang, and Y. Wang, *A fast algorithm for edge-preserving variational multichannel image restoration*, TR08-09, CAAM, Rice University, Submitted to SIIMS.

[8] J. Yang, Y. Zhang, and W. Yin, *An efficient TVL1 algorithm for deblurring of multichannel images corrupted by impulsive noise*, TR08-12, CAAM, Rice University, Submitted to SISC.

[9] J. Yang, Y. Zhang, and W. Yin, *A fast TVL1-L2 minimization algorithm for signal reconstruction from partial Fourier data*, TR08-27, CAAM, Rice University, Submitted to IEEE JSTSP.

Codes available at:

`http://www.caam.rice.edu/~optimization/L1/ftvd`

Acknowledgments

- Junfeng Yang has been supported by the Chinese Scholarship Council during his visit to Rice University.
- Wotao Yin has been supported in part by ONR Grant N00014-08-1-1101 and NSF CAREER Award DMS-0748839.
- Yin Zhang has been supported in part by NSF Grant DMS-0811188 and ONR Grant N00014-08-1-1101.

Thank you!