Image Denoising and Deblurring

Applied Math 515 Final Project

Samuel Rudy, Kelsey Maass, Riley Molloy, and Kevin Mueller

March 12 2016

Contents

Image Denoising and Deblurring

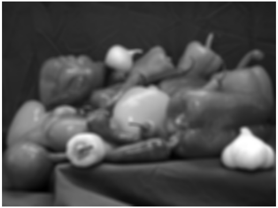
Mathematical Formulation

$$\underset{\text{Blur operator}}{A}\overset{\text{True image}}{x} + \underset{\text{Noise}}{w} = \overset{\text{Observed image}}{b}$$

- Blur: Ax is a discrete convolution of the true image with a Gaussian kernel (reflexive boundary conditions).
- Noise: w is noise drawn from a Gaussian or Student's t distribution.

Naive Solution: $x = A^{-1}b$

True image

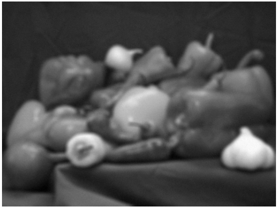


Blurred image

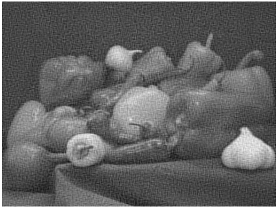
Recovered image

Naive Solution: $x = A^{-1}(b - w)$

True image



Blurred and noisy image



Recovered image

Since the blur operator is ill-conditioned, a better approach is to minimize a regularized objective function.

General Objective Function

$$L_b(x) \;=\; \underbrace{f(Ax-b)}_{\text{Fidelity term}} + \underbrace{\lambda R(x)}_{\text{Regularization}}$$

Fidelity Terms

$$f = \begin{cases} \|\cdot\|_F^2 \\ h_\gamma(\cdot)\text{we should define this} \\ \gamma^{-1}\log(\cosh(\gamma\cdot)) \end{cases}$$

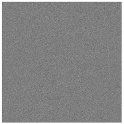
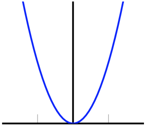
Regularization Terms

$$R = \begin{cases} \|Wx\|_1 \\ TV(x) \end{cases}$$

Fidelity Term Penalty Functions

The fidelity term $f(Ax - b)$ measures how well our results comply with the linear blurring model. Depending upon the type of noise present in the observed image, the choice of penalty function may influence the efficacy of our deblurring/denoising procedure.

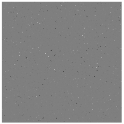
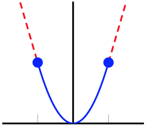
Gaussian Noise

Due to the lack of outliers, the quadratic penalty is sufficient

$$f(z) = \frac{1}{2}\|z\|^2$$

Student's t Noise

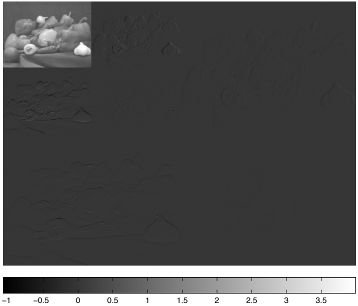



Huber penalty preferred since it is more robust to heavy-tailed noise

$$f(z) = \min_y \frac{1}{2}\|z - y\|^2 + \gamma\|y\|_1$$

L1 Wavelet Regularization

$$R(x) = \|Wx\|_1$$



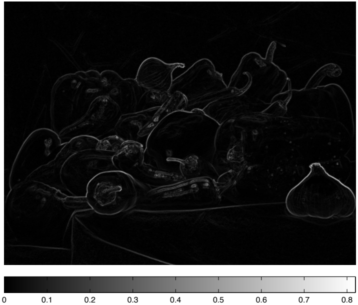
Haar Transform (2 Levels)

Images are often sparse in wavelet domains, so the L1 wavelet regularizer can be used to encourage this property in our recovered image.

For our examples, we let W be the orthogonal 2D Haar or Daubechies wavelet transform using 5 levels.

Total Variation Regularization

$$R(x) = TV(x)$$



$$|x_{ij} - x_{i+1,j}| + |x_{ij} - x_{i,j+1}|$$

words words

L1 Wavelet Regularization

Objective Function

$$L_b(x) = f(Ax - b) + \lambda \|Wx\|_1$$

Proximal Gradient Step

$$\begin{aligned} x^{k+1} &= \mathrm{prox}_{\alpha^{-1}\lambda\|W\cdot\|_1}\left(x^k - \alpha^{-1}A^T\nabla f(Ax^k - b)\right) \\ &= W^*\mathrm{prox}_{\alpha^{-1}\lambda\|\cdot\|_1}\left(W\left(x^k - \alpha^{-1}A^T\nabla f(Ax^k - b)\right)\right) \end{aligned}$$

FISTA L1 Wavelet Regularization

FISTA(b, f, λ)

$y^1 = x^0 = b;\ t^1 = 1$

$\alpha \geq Lip(\nabla f)$

for $k = 1 : N$ **do**

$$u^k = y^k - \alpha^{-1} A^T \nabla f(Ay^k - b)$$

$$x^k = W^* \mathrm{sgn}(Wu^k) \max(0, |Wu^k| - \alpha^{-1}\lambda)$$

$$t^{k+1} = \frac{1+\sqrt{1+4t_2^k}}{2}$$

$$y^{k+1} = x^k + \left(\frac{t^k - 1}{t^{k+1}}\right)(x^k - x^{k-1})$$

end for

return x^N

Note on Blur Operators

For our assumed reflexive boundary conditions, the matrix A is a Kronecker product of Toeplitz-plus-Hankel matrices which can be diagonalized by the discrete cosine transform

$$A = C^T \Lambda C,$$

where the eigenvalues are determined by the blur kernel. This means that to blur an image, we don't actually have to construct A:

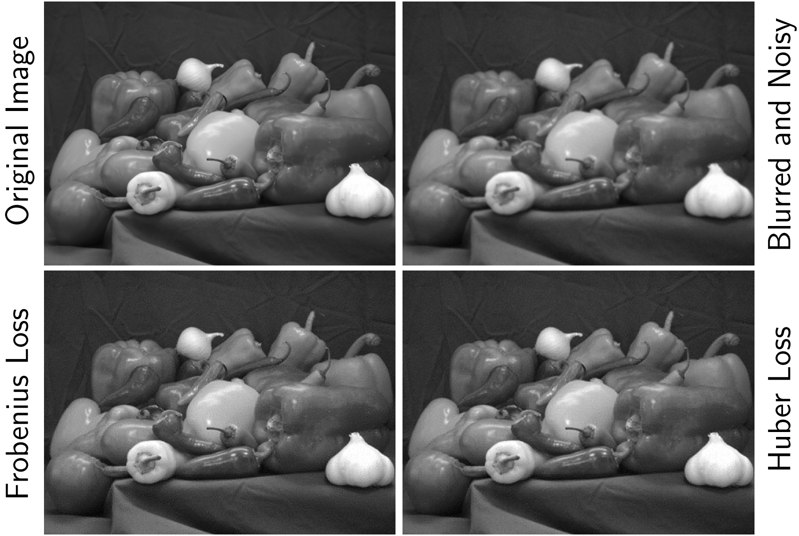
$$Ax = \texttt{idct2}\big(\Lambda \texttt{dct2}(x)\big)$$

Note on Blur Operators

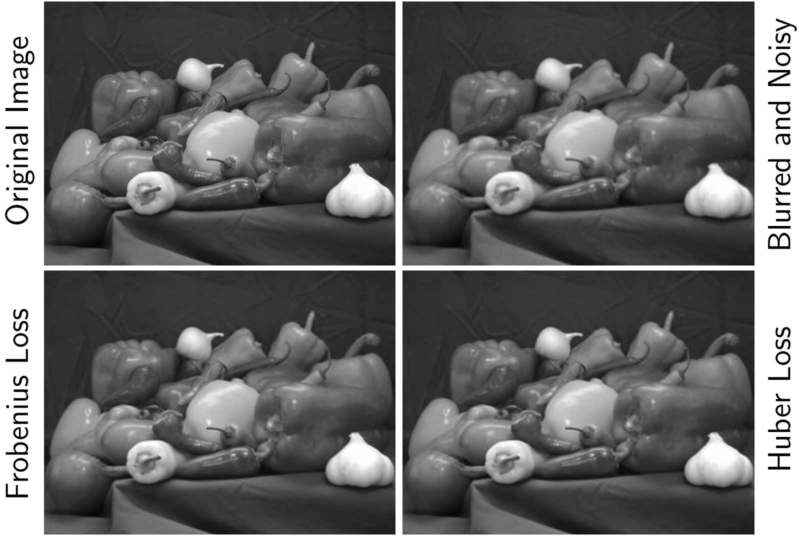
When $f = \|Ax - b\|^2$ we can exploit the fact that C is unitary:

$$
\begin{aligned}
f(Ax - b) &= \|Ax - b\|^2 \\
&= \|C^T \Lambda C x - b\|^2 \\
&= \|\Lambda C x - Cb\|^2 \\
&= \|\Lambda \hat{x} - \hat{b}\|^2
\end{aligned}
$$

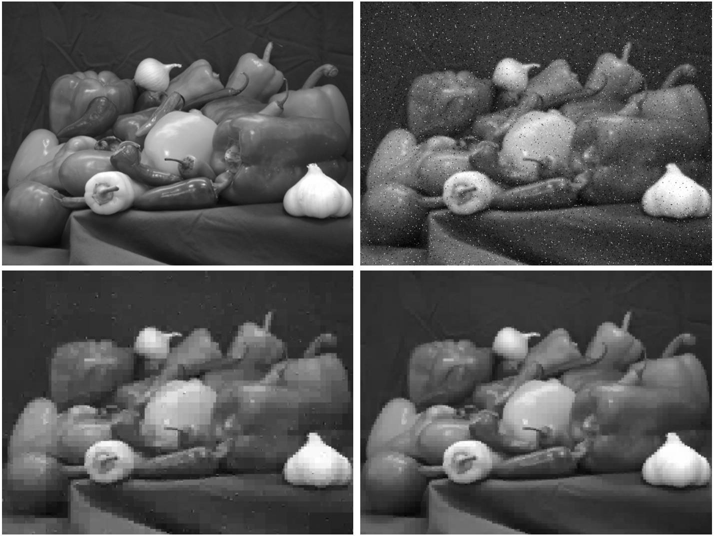
Results: Gaussian Noise (Haar)



Results: Gaussian Noise (Daubechies)



Results: Student's t Noise (Haar)



Original Image

Blurred and Noisy

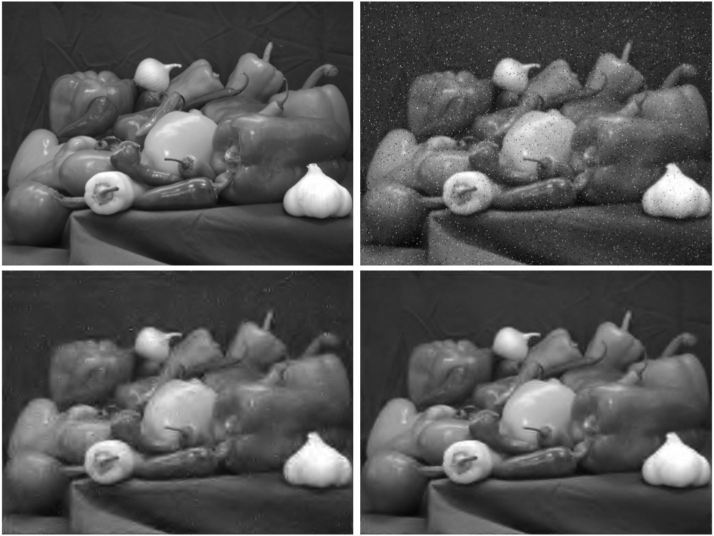
Frobenius Loss

Huber Loss

Results: Student's t Noise (Daubechies)

Original Image

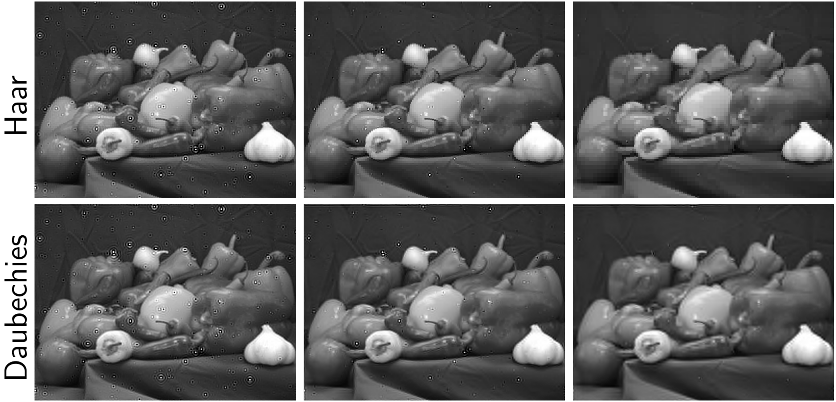
Blurred and Noisy



Frobenius Loss

Huber Loss

Choosing λ



$\lambda = 10^{-4}$ $\lambda = 10^{-3}$ $\lambda = 10^{-2}$

Total Variation Regularization

Objective Function

$$L_b(x) = f(Ax - b) + \lambda \mathrm{TV}(x) + \delta(x|[0,1])$$

Proximal Gradient Step

$$\begin{aligned} x^{k+1} &= \mathrm{prox}_{\alpha^{-1}(\lambda\|\cdot\|_{TV} + \delta_{[0,1]})}\underbrace{(x^k - \alpha^{-1}A^T\nabla f(Ax^k - b))}_{u^k} \\ &= \underset{z}{arg\ min}\left(\|u^k - z\|_F^2 + \alpha^{-1}\lambda\|z\|_{TV} + \delta(z|[0,1])\right) \\ &= P_{[0,1]}\left(\underset{z}{arg\ min}\left(\|u^k - z\|_F^2 + \alpha^{-1}\lambda\|z\|_{TV}\right)\right) \end{aligned}$$

Dual Form of Total Variation

A Few Definitions

- $\mathcal{P} = \{(p,q) \in \mathbb{R}^{(m-1)\times n} \times \mathbb{R}^{m\times(n-1)} : |p_{i,j}| \leq 1, |p_{i,j}| \leq 1\}$,
- $\mathcal{L} : \mathbb{R}^{(m-1)\times n} \times \mathbb{R}^{m\times(n-1)} \to \mathbb{R}^{m\times n}$ such that

$$\mathcal{L}(p,q)_{i,j} = p_{i,j} + q_{i,j} - p_{i-1,j} - q_{i,j-1}$$

  for $i = 1, \ldots, m$, $j = 1, \ldots, n$, and
  $p_{0,j} = p_{m,j} = q_{i,0} = q_{i,n} = 0$.
- P_C is the usual projection operator onto the set C

Total Variation

$$\mathrm{TV}(x) = max_{p,q\in\mathcal{P}} T(x,p,q) \implies T(x,p,q) = \mathrm{Tr}(\mathcal{L}(p,q)^T x).$$

Dual Form of TV Denoising with $\|\cdot\|_F^2$

The problem:

$$\min_{x \in C} \|x - b\|_F^2 + 2\lambda \mathrm{TV}(x),\, C = [0,1]$$

Dual problem:

$$\min_{(p,q)\in\mathcal{P}} \underbrace{-\|H_C(b - \lambda\mathcal{L}(p,q))\|_F^2 + \|b - \lambda\mathcal{L}(p,q)\|_F^2}_{h(p,q)}$$

$$H_C(x) = \underbrace{x - P_C(x)}_{\text{prox}}$$

Optimality conditions:

$$x = P_C(b - \lambda\mathcal{L}(p,q)).$$

Optimization of Dual Form

Problem Statement

$$\min_{(p,q)\in\mathcal{P}} \left\{ \|b - \lambda\mathcal{L}(p,q)\|_F^2 - \|(I - P_{[0,1]})(b - \lambda\mathcal{L}(p,q))\|_F^2 \right\}$$

$$\min_{(p,q)} \left\{ \underbrace{\|b - \lambda\mathcal{L}(p,q)\|_F^2 - \|(I - P_{[0,1]})(b - \lambda\mathcal{L}(p,q))\|_F^2}_{=h(p,q)} + \delta((p,q)|\mathcal{P}) \right\}$$

$$\nabla h(p,q) = -2\lambda\mathcal{L}^T P_{[0,1]}(b - \lambda\mathcal{L}(p,q))$$

Lipschitz with constant $\leq 16\lambda^2$

$\Rightarrow$ Use projected gradient

Optimization in Dual Form

Projection onto $\mathcal{P}$

Recal $\mathcal{P} = (p, q) \in [-1, 1]^{m-1 \times n} \times [-1, 1]^{m \times n-1}$

$$P_{\mathcal{P}}(p, q) = (r, s) \text{ with } \begin{cases} r_{ij} = sgn(p_{ij}) \min\{1, |p_{ij}|\} \\ s_{ij} = sgn(q_{ij}) \min\{1, |q_{ij}|\} \end{cases}$$

Projected Gradient Step

$$(p^{k+1}, q^{k+1}) = P_{\mathcal{P}} \left((p^k, q^k) + \frac{1}{8\lambda} \mathcal{L}^T P_{[0,1]}(b - \lambda \mathcal{L}(p, q)) \right)$$

Monotone FISTA TV Regularization

MFISTA(b, f, λ)

$y^1 = x^0 = b;\ t^1 = 1$

$\alpha \geq Lip(\nabla f)$

for $k = 1 : N$ **do**

$$u^k = y^k - \frac{A^T \nabla f(Ay^k - b)}{\alpha}$$

$$z^k = FGP(u^k, \tfrac{\lambda}{2\alpha})$$

$$x^k = \operatorname*{argmin}_{x \in \{x^{k-1}, z^k\}} L_b(x)$$

$$t^{k+1} = \frac{1 + \sqrt{1 + 4{t^k}^2}}{2}$$

$$y^{k+1} = x^k + \frac{t^k}{t^{k+1}}(z^k - x^k) + \frac{t^k - 1}{t^{k+1}}(z^k - x^k)$$

end for

return x^N

FGP(b, λ)

$(r_{ij}^1, s_{ij}^1) = (p_{ij}^0, q_{ij}^0) = 0;\ t^1 = 1$

for $k = 1 : N$ **do**

$$(p^k, q^k) = P_{\mathcal{P}}\left((r^k, s^k) - \frac{\mathcal{L}^T P_{[0,1]}(b - \lambda \mathcal{L}(r^k, s^k))}{8\lambda}\right)$$

$$t^{k+1} = \frac{1 + \sqrt{1 + 4{t^k}^2}}{2}$$

$$(r^k, s^k) = (p^k, q^k) + \frac{t^k - 1}{t^{k+1}}(p^k - p^{k-1}, q^k - q^{k-1})$$

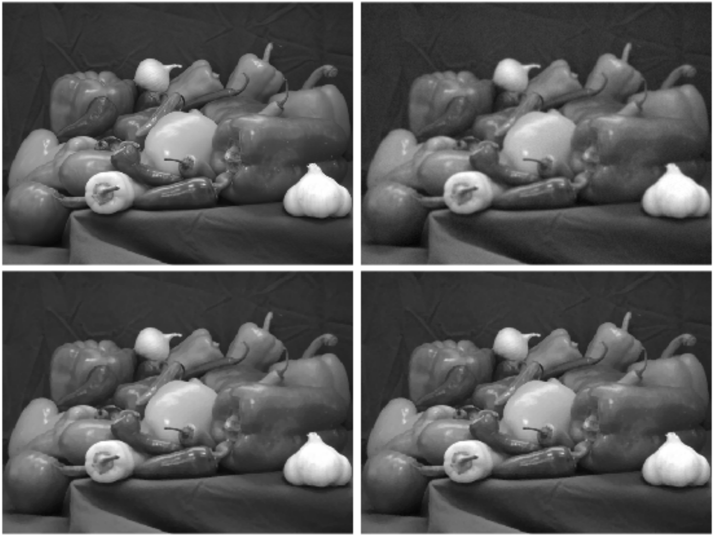
end for

return $P_{[0,1]}(b - \lambda \mathcal{L}(p^N, q^N))$

Results: Gaussian Noise

Original Image

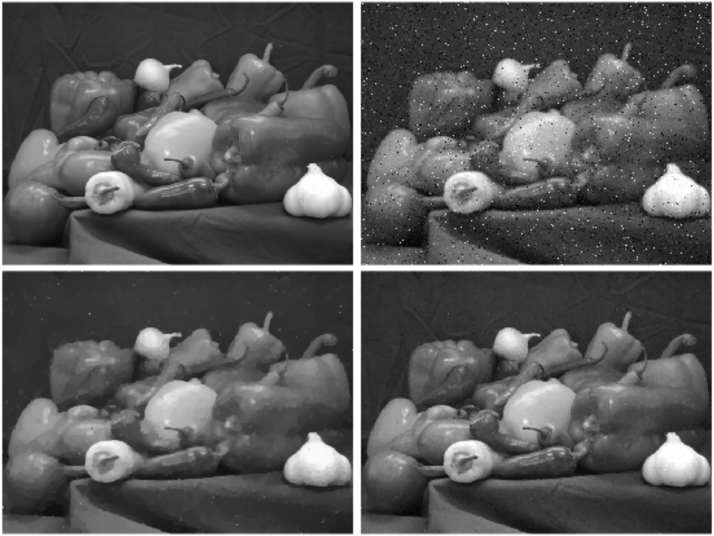
Blurred and Noisy



Frobenius Loss

Huber Loss

Results: Student's t Noise



Conclusions

Wavelet vs. Total Variation

- Total variation seems to do better for high amounts of noise.
- **How do they compare on timing???? Any advantages of Wavelet???**

Frobenius vs Huber

- On Gaussian noise, the two are comparable.
- Huber outperforms Frobenius on noise with heavier tail.

Challenges

- Ideal parameter values change image to image.
- How can we quantitatively evaluate performance?
- How can we optimize parameters without performance metric?

Questions?

J. G. Nagy, P. C. Hansen, and D. P. O'Leary. *Deblurring Images: Matrices, Spectra, and Filtering.* SIAM, Philadelphia, 2006.

A. Beck and M. Teboulle. A fast iterative shrinkage-thresholding algorithm for linear inverse problem. *SIAM Journal of Imaging Sciences*, 2(1):183-202, 2009.

A. Beck and M. Teboulle. Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems. *IEEE Transactions on Image Proccessing*, 18(11):2419-2434, 2009.