EE 367 / CS 448I Computational Imaging and Display
Notes: Compressive Imaging and Regularized Image Reconstruction (lecture 11)

Gordon Wetzstein
gordon.wetzstein@stanford.edu

This document serves as a supplement to the material discussed in lecture 11. The document is not meant to be a comprehensive review of compressive sensing or compressive imaging. It is supposed to be an intuitive introduction to the basic mathematical concepts of compressive image reconstruction for the single pixel camera. More information can be found in the paper by Wakin et al. [2006], which is representative for an entire class of research papers published throughout the last decade.

1 Image Formation

Given a vectorized image $\mathbf{x} \in \mathbb{R}^N$ and a measurement matrix $\mathbf{A} \in \mathbb{R}^{M \times N}$, the vectorized measurements $\mathbf{b} \in \mathbb{R}^M$ are formed as

$$\mathbf{b} = \mathbf{A}\mathbf{x} + \eta. \tag{1}$$

Here, η is an additive, signal-independent noise term. For the single pixel camera and other compressive imaging applications, the number of measurements is usually smaller than the number of unknowns $M < N$, which makes the linear system under-determined. There are infinitely many solutions that result in the correct measurements, so which one should we pick? We have to impose a prior on the unknown image $\mathbf{x}$ that will help determine which of the infinitely many feasible solutions is the one we seek.

The reconstruction problem is

$$\underset{\{\mathbf{x}\}}{\text{minimize}} \quad \frac{1}{2} \|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2^2 + \lambda \Gamma(\mathbf{x}), \tag{2}$$

where $\Gamma(\mathbf{x})$ is the prior and λ is its relative weight compared to the data fidelity term. A popular choice for the prior is the ℓ_2-norm on the image $\|\mathbf{x}\|_2^2$, such that we seek a feasible solution a small ℓ_2 norm. Similar to the normal equations for the least squares solution ($\mathbf{x}_{ls} = \left(\mathbf{A}^T\mathbf{A}\right)^{-1} \mathbf{A}^T\mathbf{b}$), a least norm solution is often also computed by the closed form solution to the problem minimize $\|\mathbf{x}\|_2$ s.t. $\mathbf{b} = \mathbf{A}\mathbf{x}$ which is $\mathbf{x}_{ln} = \mathbf{A}^T \left(\mathbf{A}\mathbf{A}^T\right)^{-1} \mathbf{b}$. Although the least norm solution has many important applications, it may not be the solution we are looking for in many cases. Figure 1 shows simulations of the single pixel camera for binary noise patterns with Gaussian sensor noise. These solutions are computed with Matlab's implementation of the conjugate gradient method (via the function *pcg* with matrix-free operations).

2 Regularized Image Reconstruction for Compressive Imaging and Beyond

Equation 2 can be efficiently solved with ADMM. We already encountered this type of a reconstruction problem in lecture 6 and used ADMM to solve it. The difference between the deconvolution problem in lecture 6 and Equation 2 is that there is no algebraic inverse of $\mathbf{A}$ in this context. In the following, we derive a general ADMM-based image reconstruction method that applies to compressive imaging (i.e., $\mathbf{A}$ is "fat") but also to all other cases ($\mathbf{A}$ is square or "skinny"). We show to implement total variation (TV) and non-local means (NLM) as specific image priors.

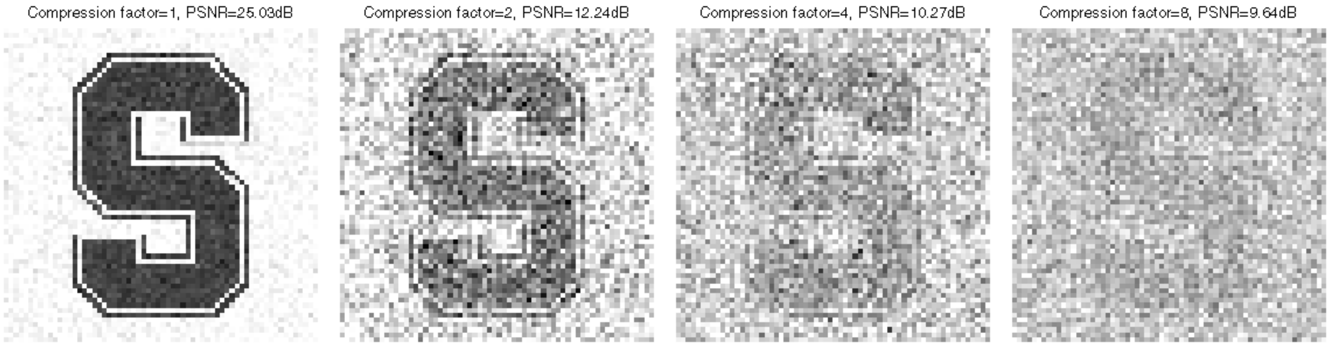

Figure 1: *Least norm solution for single pixel camera. We simulate measurements with $M = N$, $M = N/2$, $M = N/4$, and $M = N/8$ measurements. Note that even for the case where $M = N$, a perfect reconstruction may not be possible because the measurement matrix contains random but only binary values (i.e. it may not be full-rank) and the measurements are further corrupted by noise. In all cases, the reconstruction is noisy and for larger compression ratios, the reconstructed image quality is poor.*

Without loss of generality, we introduce a slack variable $\mathbf{z} \in \mathbb{R}^L$ and split Equation 2 as

$$\begin{aligned} &\underset{\{x\}}{\text{minimize}} \;\; \underbrace{\frac{1}{2}\left\|\mathbf{A}\mathbf{x}-\mathbf{b}\right\|_2^2}_{f(x)} + \underbrace{\lambda\Gamma\left(\mathbf{z}\right)}_{g(z)} \\ &\text{subject to } \mathbf{K}\mathbf{x}-\mathbf{z}=0 \end{aligned} \tag{3}$$

Here, $\mathbf{K} \in \mathbb{R}^{N \times L}$ is a general linear operator. This is useful, because it can model a finite difference approximation of the image gradients or other specific operators while keeping the inverse model general for now.

Following the general ADMM strategy, the Augmented Lagrangian of Equation 3 is formed as

$$L_\rho\left(\mathbf{x},\mathbf{z},\mathbf{y}\right) = f\left(\mathbf{x}\right) + g\left(\mathbf{z}\right) + y^T\left(\mathbf{K}\mathbf{x}-\mathbf{z}\right) + \frac{\rho}{2}\left\|\mathbf{K}\mathbf{x}-\mathbf{z}\right\|_2^2 \tag{4}$$

As discussed in more detail in Chapter 3.1 of [Boyd et al. 2001], using the scaled form of the Augmented Lagrangian, the following iterative updates rules can be derived:

$$\begin{aligned} &\mathbf{x} \leftarrow \mathbf{prox}_{\|\cdot\|_2,\rho}\left(\mathbf{v}\right) = \underset{\{\mathbf{x}\}}{\arg\min}\; L_\rho\left(\mathbf{x},\mathbf{z},\mathbf{y}\right) = \underset{\{\mathbf{x}\}}{\arg\min}\; \frac{1}{2}\left\|\mathbf{A}\mathbf{x}-\mathbf{b}\right\|_2^2 + \frac{\rho}{2}\left\|\mathbf{K}\mathbf{x}-\mathbf{v}\right\|_2^2, \quad \mathbf{v}=\mathbf{z}-\mathbf{u} \\ &\mathbf{z} \leftarrow \mathbf{prox}_{\Gamma,\rho}\left(\mathbf{v}\right) = \underset{\{\mathbf{z}\}}{\arg\min}\; L_\rho\left(\mathbf{x},\mathbf{z},\mathbf{y}\right) = \underset{\{\mathbf{z}\}}{\arg\min}\; \lambda\Gamma\left(\mathbf{z}\right) + \frac{\rho}{2}\left\|\mathbf{v}-\mathbf{z}\right\|_2^2, \quad \mathbf{v}=\mathbf{K}\mathbf{x}+\mathbf{u} \\ &\mathbf{u} \leftarrow \mathbf{u}+\mathbf{K}\mathbf{x}-\mathbf{z} \end{aligned} \tag{5}$$

where $\mathbf{u} = (1/\rho)\mathbf{y}$. The $\mathbf{x}$ and $\mathbf{z}$-updates are performed via the proximal operators $\mathbf{prox}_{\cdot,\rho}$ The interested reader is referred to [Boyd et al. 2001] for more details.

2.1 Efficient Implementation of x-Update

For the $\mathbf{x}$-update, we need to derive the proximal operator $\mathbf{prox}_{\|\cdot\|_2,\rho}$, which is the following a quadratic program:

$$\mathbf{prox}_{\|\cdot\|_2,\rho}\left(\mathbf{v}\right) = \underset{\{\mathbf{x}\}}{\arg\min}\; \frac{1}{2}\left\|\mathbf{A}\mathbf{x}-\mathbf{b}\right\|_2^2 + \frac{\rho}{2}\left\|\mathbf{K}\mathbf{x}-\mathbf{v}\right\|_2^2, \quad \mathbf{v}=\mathbf{z}-\mathbf{u} \tag{6}$$

To make it easy to follow the derivation step-by-step, we write the objective function out as

$$\frac{1}{2}\left(\mathbf{A}\mathbf{x}-\mathbf{b}\right)^T\left(\mathbf{A}\mathbf{x}-\mathbf{b}\right)+\frac{\rho}{2}\left(\mathbf{K}\mathbf{x}-\mathbf{v}\right)^T\left(\mathbf{K}\mathbf{x}-\mathbf{v}\right) \tag{7}$$
$$=\frac{1}{2}\left(\mathbf{x}^T\mathbf{A}^T\mathbf{A}\mathbf{x}-2\mathbf{x}^T\mathbf{A}^T\mathbf{b}+\mathbf{b}^T\mathbf{b}\right)+\frac{\rho}{2}\left(\mathbf{x}^T\mathbf{K}^T\mathbf{K}\mathbf{x}-2\mathbf{x}^T\mathbf{K}^T\mathbf{v}+\mathbf{v}^T\mathbf{v}\right)$$

The gradient of this expression (i.e. the objective function in Eq. 6) is

$$\mathbf{A}^T\mathbf{A}\mathbf{x}-\mathbf{A}^T\mathbf{b}+\rho\mathbf{K}^T\mathbf{K}\mathbf{x}-\rho\mathbf{K}^T\mathbf{v}, \tag{8}$$

which, equated to zero, results in the normal equations that allow us to derive an expression for estimating $\widetilde{\mathbf{x}}$ as

$$\mathbf{prox}_{\|\cdot\|_2,\rho}\left(\mathbf{v}\right)=\left(\underbrace{\mathbf{A}^T\mathbf{A}+\rho\mathbf{K}^T\mathbf{K}}_{\widetilde{\mathbf{A}}}\right)^{-1}\left(\underbrace{\mathbf{A}^T\mathbf{b}+\rho\mathbf{K}^T\mathbf{v}}_{\widetilde{b}}\right). \tag{9}$$

Unfortunately, it is not easily possible to derive a closed-form solution for the inverse of $\widetilde{\mathbf{A}}$ as it is for the convolution problem in lecture 6. Hence, we will use an iterative solver to compute this proximal operator by solving $\widetilde{\mathbf{A}}\mathbf{x}=\widetilde{\mathbf{b}}$. Since $\widetilde{\mathbf{A}}$ is symmetric and positive semi-definite, the conjugate gradient (CG) method is an adequate choice. CG is implemented in the Matlab Optimization Toolbox as the function *pcg*. The *pcg* function also supports matrix-free operations – one simply supplies a function handle that computes $\widetilde{\mathbf{A}}\mathbf{x}$ without forming the matrix. Due to symmetry of the matrix ($\widetilde{\mathbf{A}}^T=\widetilde{\mathbf{A}}$) we do not have to specify the adjoint operation.

2.2 Implementation of z-Update

Updating $\mathbf{z}$ requires a proximal operator for the specific prior (or sum of priors). In general, this prior is

$$\mathbf{prox}_{\Gamma,\rho}\left(\mathbf{v}\right)=\underset{\{\mathbf{z}\}}{\arg\min}\; L_\rho\left(\mathbf{x},\mathbf{z},\mathbf{y}\right)=\underset{\{\mathbf{z}\}}{\arg\min}\;\lambda\Gamma\left(\mathbf{z}\right)+\frac{\rho}{2}\left\|\mathbf{v}-\mathbf{z}\right\|_2^2,\quad \mathbf{v}=\mathbf{K}\mathbf{x}+\mathbf{u} \tag{10}$$

2.2.1 Total Variation Prior

Total variation uses a sparsity prior on the image gradients, i.e. $\Gamma\left(\mathbf{x}\right)=\left\|\mathbf{D}\mathbf{x}\right\|_1$. Here, $\mathbf{K}=\mathbf{D}=\left[\mathbf{D}_x^T\;\mathbf{D}_y^T\right]^T\in\mathbb{R}^{2N\times N}$, $\mathbf{z},\mathbf{u}\in\mathbb{R}^{2N}$. In the $\mathbf{z}$-update, the ℓ_1-norm is convex but not differentiable. Nevertheless, a closed form solution for the proximal operator exists, such that

$$\mathbf{prox}_{\Gamma,\rho}\left(\mathbf{v}\right)=\underset{\{\mathbf{z}\}}{\arg\min}\;\lambda\left\|\mathbf{z}\right\|_1+\frac{\rho}{2}\left\|\mathbf{v}-\mathbf{z}\right\|_2^2=\mathcal{S}_{\lambda/\rho}\left(\mathbf{v}\right), \tag{11}$$

with $\mathbf{v}=\mathbf{D}\mathbf{x}+\mathbf{u}$ and $\mathcal{S}_\kappa\left(\cdot\right)$ being the element-wise soft thresholding operator

$$\mathcal{S}_\kappa\left(v\right)=\begin{cases} v-\kappa & v>\kappa \\ 0 & |v|\leq\kappa \\ v+\kappa & v<-\kappa \end{cases}=\left(v-\kappa\right)_+-\left(-v-\kappa\right) \tag{12}$$

that can be implemented very efficiently. Here $\kappa=\lambda/\rho$.

2.2.2 Non-local Means Prior

As derived in the notes for lecture 10, we can also use self similarity priors (i.e. Gaussian denoising algorithms) as general image priors. For the specific case of non-local means (NLM), the proximal operator is

$$\mathbf{prox}_{\Gamma,\rho}\left(\mathbf{v}\right)=\text{NLM}\left(\mathbf{v},\frac{\lambda}{\rho}\right). \tag{13}$$

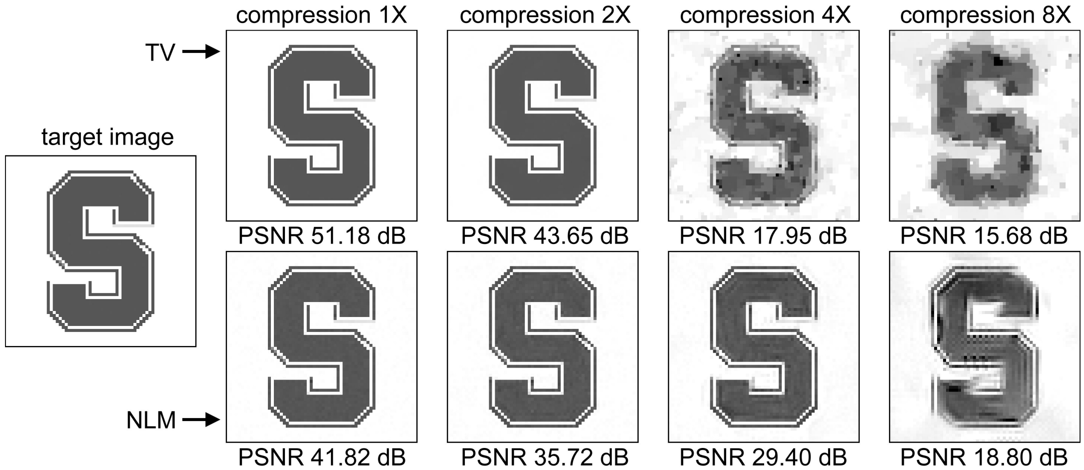

Figure 2: *Regularized ADMM solution for single pixel camera using TV (top row) and NLM (bottom row) priors. We simulate measurements with $M = N$, $M = N/2$, $M = N/4$, and $M = N/8$ measurements (compressions of $1, 2, 4, 8\times$, respectively), as seen in each column. This image has sparse gradients, so an almost perfect reconstruction is obtained with TV for small compression ratios. The reconstructed image quality for larger compression ratios rapidly drops for the TV prior. The NLM prior works significantly better for larger compression ratios.*

However, any other Gaussian denoising strategy, such as video NLM, BM3D and others, can be used instead of non-local means.

Here, $\mathbf{K} = \mathbf{I} \in \mathbb{R}^{N \times N}$, $\mathbf{z}, \mathbf{u} \in \mathbb{R}^N$.

2.3 Pseudo Code and Additional Information

For additional information on implementation, please see the example code provided on the ADMM website `http://stanford.edu/~boyd/papers/admm/`.

Further, Algorithm 1 outlines pseudo code for the TV-regularized ADMM deconvolution.

Algorithm 1 ADMM for TV-regularized deconvolution

initialize ρ and λ
$\mathbf{x} = zeros\,(W, H)\,;$
$\mathbf{z} = zeros\,(W, H, 2)\,;$
$\mathbf{u} = zeros\,(W, H, 2)\,;$
for $k = 1$ **to** $maxIters$ **do**
  $\mathbf{x} = \mathbf{prox}_{\|\cdot\|_2,\rho}(\mathbf{v}) = \underset{\{\mathbf{x}\}}{\arg\min}\ \frac{1}{2}\|\mathbf{Ax} - \mathbf{b}\|_2^2 + \frac{\rho}{2}\|\mathbf{Kx} - \mathbf{v}\|_2^2, \quad \mathbf{v} = \mathbf{z} - \mathbf{u}$
  $\mathbf{z} = \mathbf{prox}_{\Gamma,\rho}(\mathbf{v}) = \underset{\{\mathbf{z}\}}{\arg\min}\ \lambda\Gamma(\mathbf{z}) + \frac{\rho}{2}\|\mathbf{v} - \mathbf{z}\|_2^2, \quad \mathbf{v} = \mathbf{Kx} + \mathbf{u}$
  $\mathbf{u} = \mathbf{u} + \mathbf{Kx} - \mathbf{z}$
end for

References

BOYD, S., PARIKH, N., CHU, E., PELEATO, B., AND ECKSTEIN, J. 2001. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends in Machine Learning 3*, 1, 1–122.

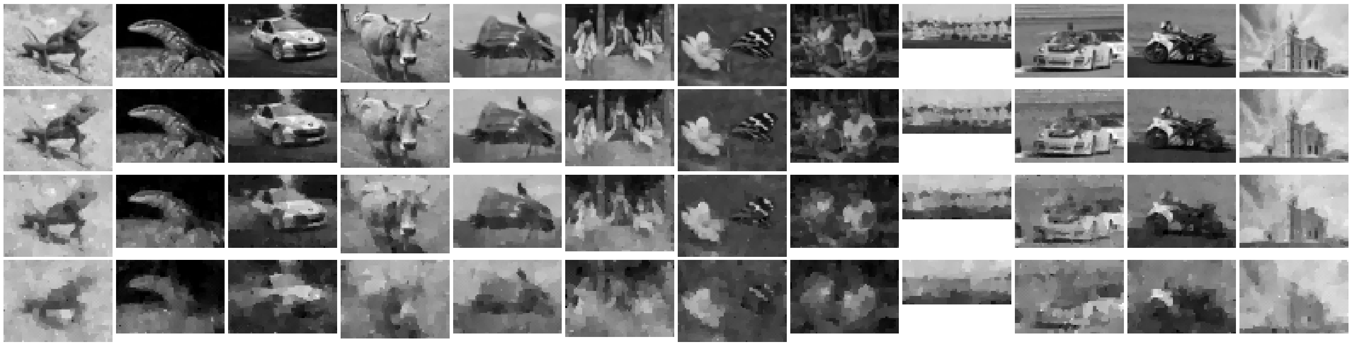

Figure 3: *Overview of compressive imaging reconstruction performance for 12 small images (columns). Each row shows a TV-reconstruction for different compression ratios (number of measurements), i.e.* $N/M = 1, 2, 4, 8$, *respectively. We see that with a decreasing number of measurements, the results become more "patchy", which is due to the TV prior. However, compression factors* N/M *of 2-4 can be achieved for a reasonable image quality for this simple image prior.*

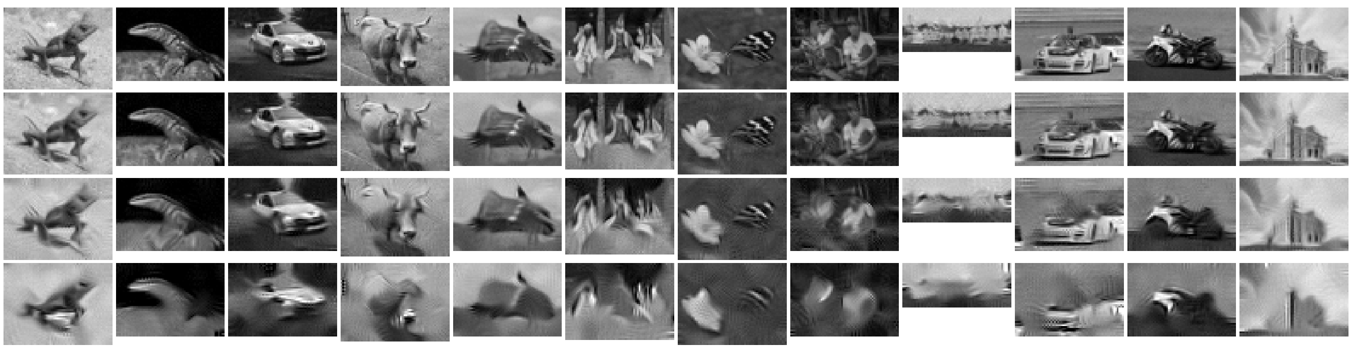

Figure 4: *Overview of compressive imaging reconstruction performance for 12 small images (columns). Each row shows a NLM-reconstruction for different compression ratios (number of measurements), i.e.* $N/M = 1, 2, 4, 8$, *respectively. We see that with a decreasing number of measurements, the results are significantly better than those of the TV prior.*

WAKIN, M., LASKA, J., DUARTE, M., BARON, D., SARVOTHAM, S., TAKHAR, D., KELLY, K., AND BARANIUK, R. 2006. An architecture for compressive imaging. In *Proc. International Conference on Image Processing (ICIP)*.