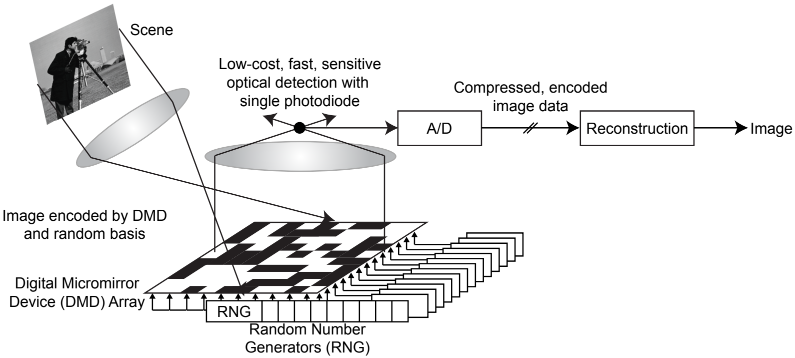
The Dark Side of Compressed Sensing: Minimax bounds under Poisson noise

Rebecca Willett
Xin Jiang and Garvesh Raskutti

Compressive optical systems[1]



If we fix our total data acquisition time to T, then we have an explicit tradeoff between the number of projections, n, and the number of photons collected per projection, $O(T/n)$. As n increases, photon-limitations dominate errors.

[1] Duarte, Davenport, Laska, Sun, Takhar, Sarvotham, Baron, Wakin & Kelly, Baraniuk (2006)

Sensing model

We observe

$$y \sim \text{Poisson}(TAf^*)$$

$$y_i \sim \text{Poisson}\left(T\sum_{j=1}^{p} A_{i,j} f_j^*\right), \qquad i \in \{1, \dots, n\},$$

where

- $y \in \mathbb{Z}_+^n$
- $T \in \mathbb{R}_+$ is the total data acquisition time
- $A \in [0,1]^{n \times p}$ is a known sensing matrix
- $f^* \in \mathcal{F}$, where

$$\mathcal{F} = \left\{ f \in \mathbb{R}_+^p \ : \ \|f\|_1 = 1, \ \|D^T f\|_0 \le s+1 \right\}$$

for an orthonormal basis D

Our goal is to reconstruct f^* from y. How does performance depend on n, p, T, D, and A? What fundamentally limits our sensing capabilities?

1. Previous work established **upper** bounds; were these bounds **tight**?
2. Is it better to have a **lot of high-noise measurements** (big n), or a **few low-noise measurements?**
3. What are the key **ramifications** of Poisson compressed sensing? How is it different from typical settings?

Spoilers

CS conventional wisdom (for Gaussian noise settings) tells us rates are

- Independent of sparsifying basis
- Not much worse than if we collected non-compressive measurements

In Poisson noise settings, because of the interaction between physical constraints and sparsity assumptions

- Rates are highly dependent on sparsifying basis
- Depending on the sparsity assumptions, we can do far better using non-compressive measurements

This is not your ordinary CS problem

Sensing matrix A has several physical constraints

Think of $A_{i,j}$ as likelihood of photon from location j in f^* hitting detector at location i:

$$
\begin{aligned}
A_{i,j} &\in [0,1] \\
\mathbb{1}^\top A &\preceq \mathbb{1} && \text{(columns sum to at most one)} \\
\|Af\|_1 &\leq \|f\|_1 && \forall f
\end{aligned}
$$

Typical CS sensing matrices do not satisfy these constraints!

Sensing matrix

Start with a sensing matrix $\widetilde{A} \in \frac{1}{\sqrt{n}}\{-1,1\}^{n\times p}$ such that the product $\widetilde{A}D$ satisfies the RIP:

$$(1-\delta_s)\|\theta\|_2^2 \leq \|\widetilde{A}D\theta\|_2^2 \leq (1+\delta_s)\|\theta\|_2^2 \quad \forall \quad 2s - \text{sparse } \theta$$

Let

$$A \triangleq (\widetilde{A} + \frac{3}{\sqrt{n}}\mathbb{1}_{n\times p})/4\sqrt{n}.$$

Sensing matrix

Start with a sensing matrix $\widetilde{A} \in \frac{1}{\sqrt{n}}\{-1,1\}^{n \times p}$ such that the product $\widetilde{A}D$ satisfies the RIP:

$$(1-\delta_s)\|\theta\|_2^2 \leq \|\widetilde{A}D\theta\|_2^2 \leq (1+\delta_s)\|\theta\|_2^2 \quad \forall \quad 2s - \text{sparse } \theta$$

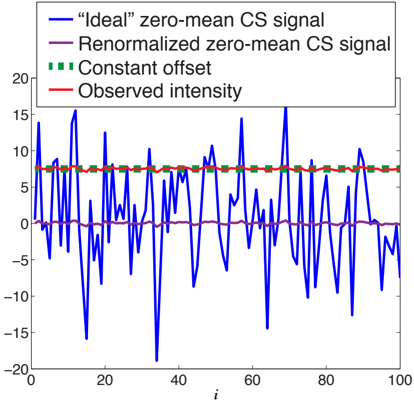
Let

$$A \triangleq (\widetilde{A} + \frac{3}{\sqrt{n}}\mathbb{1}_{n\times p})/4\sqrt{n}.$$

We observe

$$y \sim \text{Poisson}(TAf^*)$$

$$\sim \text{Poisson}\left(\frac{T\widetilde{A}f^*}{4\sqrt{n}} + \underbrace{\frac{3T}{4n}\mathbb{1}_{n\times 1}}_{\text{determines variance}}\right)$$



Rates for high-intensity settings (large T) [2]

Theorem:

$$\inf_{\widehat{f}} \sup_{f^* \in \mathcal{F}} \mathbb{E}[\|\widehat{f} - f^*\|_2^2] \asymp \frac{s \log p}{T}$$

where

$$\mathcal{F} = \left\{ f \in \mathbb{R}_+^p \ : \ \|f\|_1 = 1, \ \|D^T f\|_0 \leq s + 1 \right\}$$

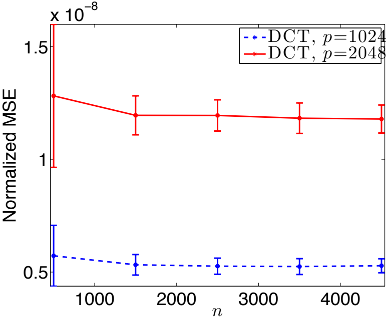
- The data acquisition time T, which reflects the signal-to-noise ratio, controls the error decay
- Once the number of measurements, n, is sufficiently large to ensure a RIP-like sensing matrix, it does not impact errors

[2] Jiang, Raskutti & Willett (2014)

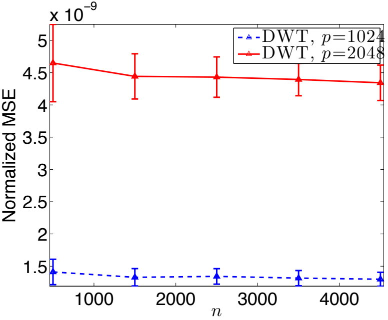
MSE vs. measurements

If we fix our total data acquisition time to T, then we have an explicit tradeoff between the number of projections, n, and the number of photons collected per projection, $O(T/n)$. As n increases, photon-limitations dominate errors.

Is it better to have a lot of high-noise measurements (big n), or a few low-noise measurements?

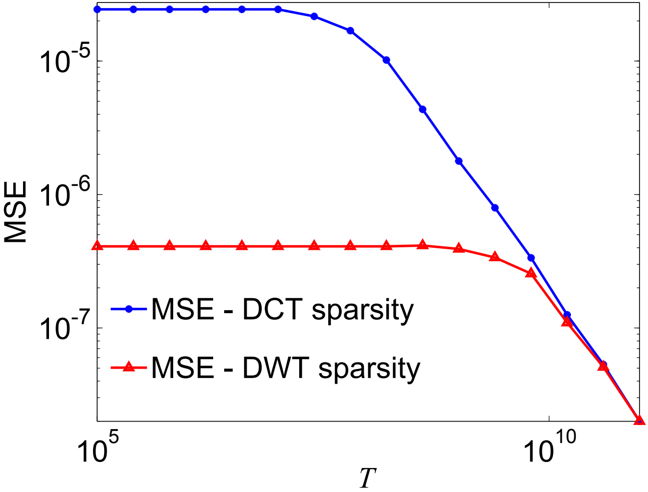


DCT sparsity



DWT sparsity

MSE vs. T: An elbow in the rates



So far we have only considered high-intensity (large T) settings. What happens in low intensities?

Low-intensity settings (small T) [3]

Let $\overline{f^*} \equiv \mathbb{1}_{p\times 1}/\sqrt{p}$ be the average of f^*. Then

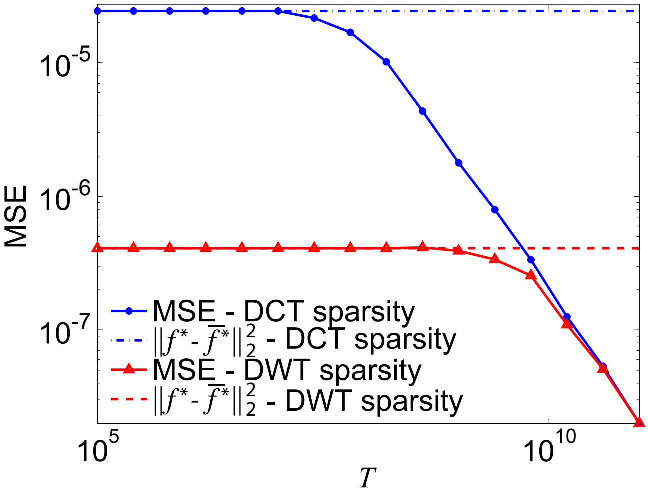
$$\mathbb{E}[\|\widehat{f} - f^*\|_2^2] \asymp \|f^* - \overline{f^*}\|_2^2$$

Rates depend on how much f^* deviates from its mean ("residual energy"), subject to the constraint that $\|f^*\|_1 = 1$ for $f^* \in \mathcal{F}$.

For different sparsifying bases D, this residual energy falls in different ranges, giving different rates.

[3] Jiang, Raskutti & Willett (2014)

MSE vs. T: An elbow in the rates



Upper bounds for low T

- We need an upper bound on the "residual energy":

$$\|f^* - \overline{f}^*\|_2^2 \leq s \max_{j,k} |D_{j,k}|^2 \|f^*\|_1^2 = s\|D\|_{\max}^2$$

- The quantity $s\|D\|_{\max}$ characterizes how closely we can approximate a δ-function with s non-zero basis coefficients.

Lower bounds for low T

Let the first column of basis D correspond to a constant signal; denote the remaining $p-1$ columns $\overline{D}$.

k-sparse localization constant[4]:

$$\lambda_k = \lambda_k(\overline{D}) \triangleq \max_{\substack{\beta \in \{-1,0,1\}^{p-1} \\ \|\beta\|_0 = k}} \|\overline{D}\beta\|_\infty$$

This quantity is the highest amplitude of any zero-mean signal which is k-sparse in $\overline{D}$ and has uniform amplitude non-zero coefficients.

This is where the geometry of our constraints (*i.e.*, nonnegativity) and the geometry of sparsity interact. Different bases D can have very different $\lambda_k(\overline{D})$.

[4] inspired by similar constant proposed by P. Reynaud-Bouret, 2003

Rates for low-intensity settings (small T) [5]

Theorem:

Lower bound: Dominated by spread-out, low-amplitude signals in $\mathcal{F}$ (with low residual energy), which is reflected by the basis-dependent $\lambda_k(\overline{D})$ for $k = 1, \ldots, s$:

$$\min_{\widehat{f}} \max_{f^* \in \mathcal{F}} \mathbb{E}[\|\widehat{f} - f^*\|_2^2] \geq C \max_{1 \leq k \leq s} \frac{k}{p^2 \lambda_k^2(\overline{D})}$$

Upper bound: Dominated by the highest amplitude signals in $\mathcal{F}$ (with high residual energy), which is reflected in the basis-dependent $\|D\|_{\max}^2$:

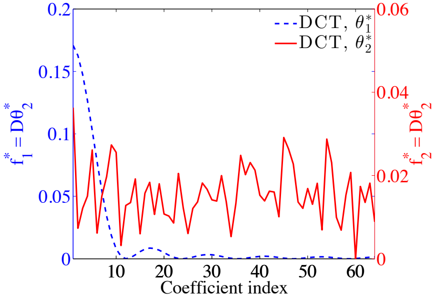
$$\min_{\widehat{f}} \max_{f^* \in \mathcal{F}} \mathbb{E}[\|\widehat{f} - f^*\|_2^2] \leq C' \min(s\|D\|_{\max}^2, 1)$$

[5] Jiang, Raskutti & Willett (2014)

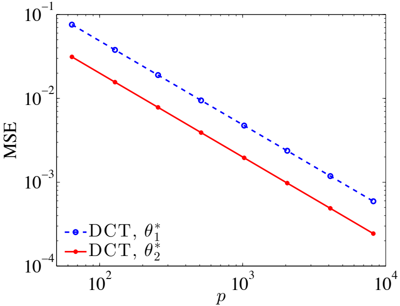
Low-intensity rates in two common bases

	Lower bound	Upper bound
DCT	$\frac{1}{p}$	$\frac{s}{p}$
DWT	$\frac{s}{p^2}$	1
	Dominated by flattest s-sparse signals in $\mathcal{F}$	Dominated by peakiest s-sparse signals in $\mathcal{F}$

Fourier sparsity



Two signals in our function class...

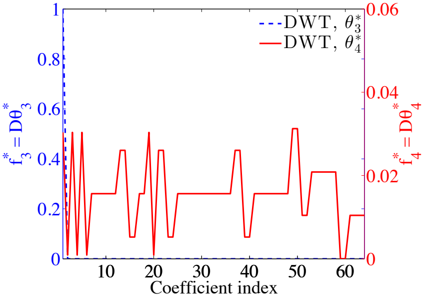


...and their rates

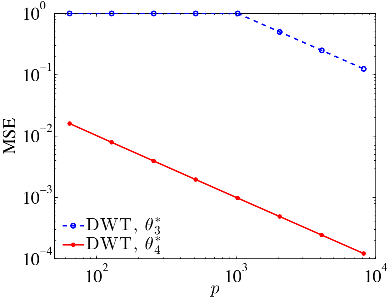
Red signal (residual energy 0.0191) controls lower bounds, blue signal controls upper bounds (residual energy 0.1151)

As predicted by theory, both signals have same scaling with p but different scaling with s

Wavelet sparsity



Two signals in our function class...



...and their rates

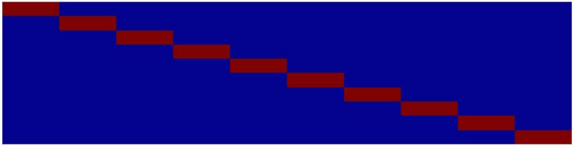
Red signal (residual energy 0.0188) controls lower bounds, blue signal controls upper bounds (residual energy 1.0000)

As predicted by theory, signal with high amplitude yields much slower rates than more diffuse signal

CS can be suboptimal at low intensities

Consider the special case where our signal is sparse in a DWT basis, and s' of the s nonzero coefficients are at coarse scales.

We will compare the CS paradigm from earlier with a simple downsampling system A^{DS}:



Measuring $A^{\mathrm{DS}} f^*$ is equivalent to measuring coarse-scale Haar wavelet coefficients of f^*.

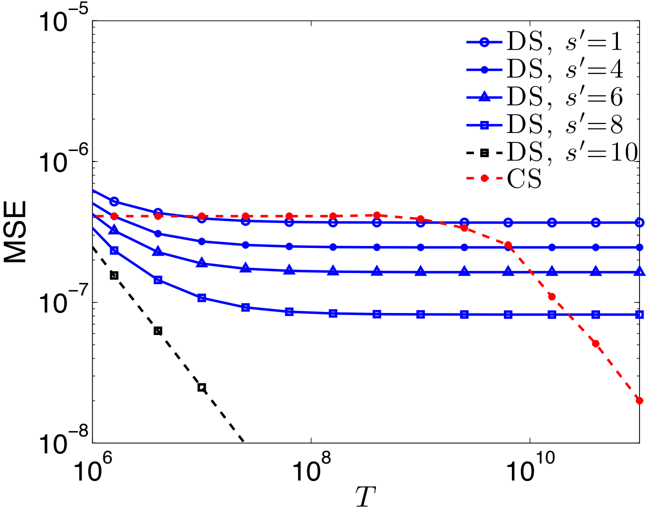
CS can be suboptimal at low intensities

Consider the special case where our signal is sparse in a Haar basis, and s' of the s nonzero coefficients are at coarse scales.

Compressive sampling	Direct measurement of coarse-scale info
$\min\left(\frac{s \log p}{T}, \frac{s}{p^2}\right),$	$\frac{s'n}{Tp^2} + \frac{s-s'}{\lambda_s^2 p^2}$
independent of s'	dependent on s'

These curves cross; for small T and moderate s', downsampling can perform significantly better than compressive sampling.

Compressive vs. direct measurements



Ramifications

CS conventional wisdom (for Gaussian noise settings) tells us rates are

- Independent of sparsifying basis
- Not much worse than if we collected non-compressive measurements

In Poisson noise settings, because of the interaction between physical constraints and sparsity assumptions

- Rates are highly dependent on sparsifying basis
- Depending on the sparsity assumptions, we can do far better using non-compressive measurements

BECOME SEDUCED BY THE DARK SIDE OF CS

Addressing photon limitations in compressed sensing poses exciting mathematical challenges

http://willett.ece.wisc.edu/
arXiv:1403.6532