From Denoising to Compressed Sensing

Christopher A. Metzler
Arian Maleki
Richard G. Baraniuk

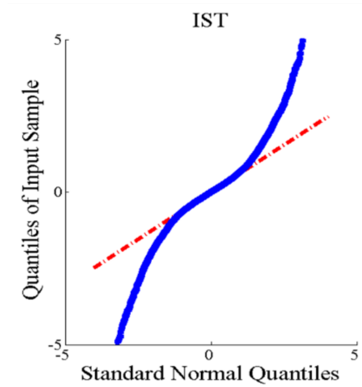
Review by: Alireza Naderi

From Compressed Sensing…

$$x_{n\times 1} \in \Sigma_k, \qquad y_{m\times 1} = \mathbf{\Phi}_{m\times n} x_{n\times 1} + w_{m\times 1}, \qquad y_{m\times 1} \overset{\widetilde{\mathbb{P}}_1}{\Rightarrow} x_{n\times 1}$$

- ISTA

$$\begin{cases} x^{t+1} = \eta_\tau(x^t + \mathbf{\Phi}^H z^t) \\ z^t = y - \mathbf{\Phi} x^t \end{cases}$$

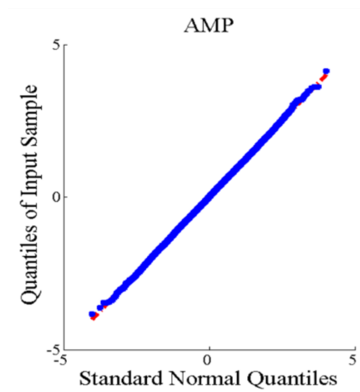


- AMP

$$\begin{cases} x^{t+1} = \eta_\tau(x^t + \mathbf{\Phi}^H z^t) \\ z^t = y - \mathbf{\Phi} x^t + \frac{1}{\delta} z^{t-1} \langle \eta'_\tau(x^{t-1} + \mathbf{\Phi}^H z^{t-1}) \rangle \end{cases}$$

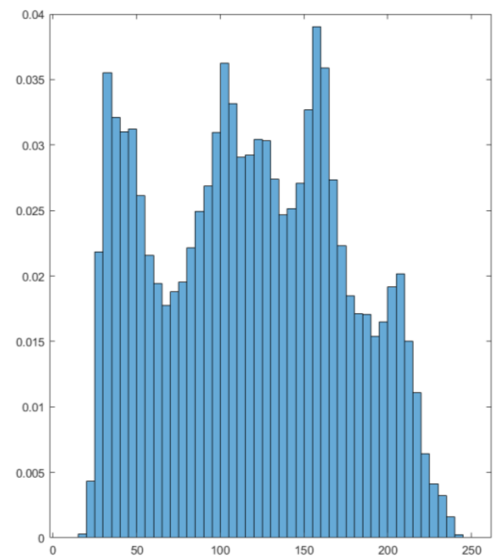
Onsager correction term

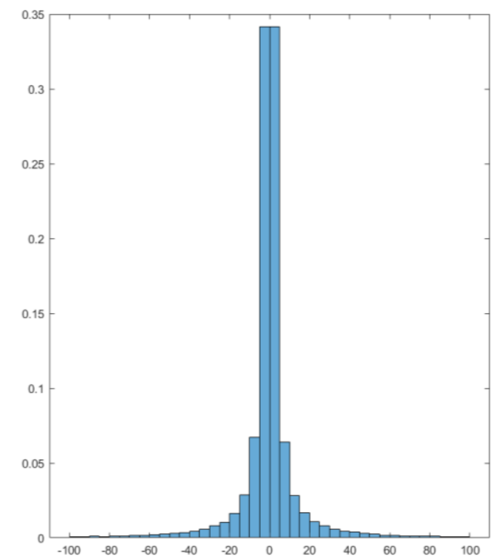
$(\delta = \frac{m}{n})$



… to Denoising

- But, what if x is not sparse?





$$x_{n\times 1} \in \mathcal{C}, \qquad y_{m\times 1} = \mathbf{\Phi}_{m\times n} x_{n\times 1} + w_{m\times 1}, \qquad y_{m\times 1} \overset{?}{\Rightarrow} x_{n\times 1}$$

D-AMP

$$x_{n\times 1} \in \mathcal{C}, \qquad y_{m\times 1} = \mathbf{\Phi}_{m\times n} x_{n\times 1} + w_{m\times 1}, \qquad y_{m\times 1} \overset{?}{\Rightarrow} x_{n\times 1}$$

- If there exists a family of denoisers $\{D_\sigma : \sigma > 0\}$ for $\mathcal{C} \subset \mathbb{R}^n$;

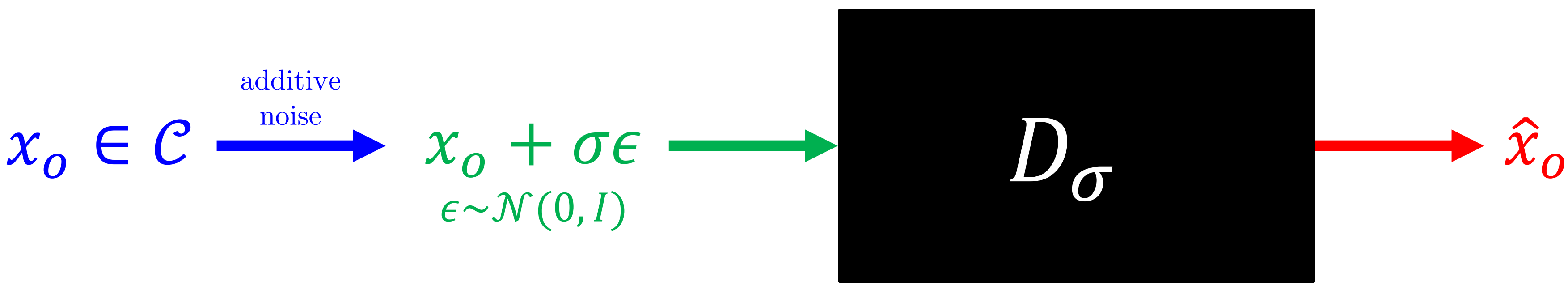
  - D-AMP

$$\begin{cases} x^{t+1} = D_{\hat{\sigma}^t}(x^t + \mathbf{\Phi}^H z^t) \\ z^t = y - \mathbf{\Phi} x^t + z^{t-1} \operatorname{div}[D_{\hat{\sigma}^{t-1}}(x^{t-1} + \mathbf{\Phi}^H z^{t-1})]/m \\ (\hat{\sigma}^t)^2 = \dfrac{\|z^t\|_2^2}{m} \end{cases}$$

$$\operatorname{div} D(x) = \sum_{i=1}^{n} \frac{\partial D(x_i)}{\partial x_i}$$

will do the work!

Denoiser



- $\{D_\sigma\}$ is:
  - κ-proper $\Leftrightarrow \forall \sigma > 0$: $\sup_{x_o \in \mathcal{C}} \frac{\mathbb{E}\|D_\sigma(x_o+\sigma\epsilon)-x_o\|_2^2}{n} \le \kappa\sigma^2$
  - (κ, B)-near proper $\Leftrightarrow \forall \sigma > 0$: $\sup_{x_o \in \mathcal{C}} \frac{\mathbb{E}\|D_\sigma(x_o+\sigma\epsilon)-x_o\|_2^2}{n} \le \kappa\sigma^2 + B$
  - monotone $\Leftrightarrow \forall x_o \in \mathcal{C}$: $\frac{\mathbb{E}\|D_\sigma(x_o+\sigma\epsilon)-x_o\|_2^2}{n}$ is non-decreasing in σ^2
  - Lipschitz continuous

$\{D_\sigma\}$ can always be modified to satisfy this

Some Examples

$\mathcal{C} = \mathbb{R}^n$

$D_\sigma =$ ML estimator

$$\frac{\mathbb{E}\|D_\sigma(x_o + \sigma\epsilon) - x_o\|_2^2}{n}$$

$$= \frac{\mathbb{E}\|(x_o + \sigma\epsilon) - x_o\|_2^2}{n}$$

$$= \frac{n\sigma^2}{n} = \sigma^2 \Rightarrow \kappa = 1$$

$\boxed{\mathcal{C} \text{ has no structure} \Rightarrow \text{the worst case}}$

$\mathcal{C} =$ k-dimensional subspace of $\mathbb{R}^n$

$D_\sigma =$ projection onto $\mathcal{C}$

$$\frac{\mathbb{E}\|D_\sigma(x_o + \sigma\epsilon) - x_o\|_2^2}{n}$$

$$= \frac{\mathbb{E}\|P_\mathcal{C}(x_o) + \sigma P_\mathcal{C}(\epsilon) - x_o\|_2^2}{n}$$

$$= \frac{\sigma^2}{n}\mathbb{E}\|P_\mathcal{C}(\epsilon)\|_2^2 = \frac{k}{n}\sigma^2$$

$$\Rightarrow \kappa = \frac{k}{n}$$

follows χ^2 distribution with k DOF

The Main Finding

- D-AMP

$$\begin{cases} x^{t+1} = D_{\hat{\sigma}^t}(x^t + \mathbf{\Phi}^H z^t) \\ z^t = y - \mathbf{\Phi} x^t + z^{t-1} \mathrm{div}[D_{\hat{\sigma}^{t-1}}(x^{t-1} + \mathbf{\Phi}^H z^{t-1})]/m \\ (\hat{\sigma}^t)^2 = \frac{\|z^t\|_2^2}{m} \end{cases}$$

$$= x_o + \text{effective noise} = x_o + v^t$$

- **Assumption: $\boldsymbol{v^t}$ resembles i.i.d. Gaussian noise!**
  - No proof. Based on extensive simulations.
  - Builds up a framework for theoretical analysis of algorithm.

State Evolution (SE)

$$\begin{cases} \theta^0 = \dfrac{\|x^0\|_2^2}{n} \\ \theta^{t+1}(x_o, \delta, \sigma_w^2) = \dfrac{1}{n}\mathbb{E}\left\|D_{\sigma^t}(x_o + \sigma^t \epsilon) - x_o\right\|_2^2 \\ (\sigma^t)^2(x_o, \delta, \sigma_w^2) = \dfrac{\theta^t}{\delta} + \sigma_w^2 \end{cases}$$

$$\left.\begin{array}{r} \text{D-AMP starts from } x_0 = 0 \\ m, n \gg 1 \text{ while } \frac{m}{n} = \delta \\ \Phi_{ij} \sim \mathcal{N}(0, \frac{1}{m^2}) \text{ and independent} \\ w \text{ is i.i.d. Gaussian} \\ \{D_\sigma\} \text{ is Lipschitz continuous} \end{array}\right\} \Longrightarrow \theta^t \approx \frac{1}{n}\|x^t - x_o\|_2^2$$

SE predicts the
MSE of D-AMP

D-AMP Without Noise

- Noiseless setting: $\sigma_w = 0 \implies \begin{cases} \theta^{t+1}(x_o, \delta, 0) = \frac{1}{n} \mathbb{E} \left\| D_{\sigma^t}(x_o + \sigma^t \epsilon) - x_o \right\|_2^2 \\ (\sigma^t)^2 = \theta^t / \delta \end{cases}$

- Two scenarios: (i) $\theta^t(x_o, \delta, 0) \to 0$ D-AMP success

  (ii) $\theta^t(x_o, \delta, 0) \nrightarrow 0$ D-AMP failure

- What are the success and failure regions (in terms of δ)?

- Lemma 1: For monotone denoisers,

$$\theta^t(x_o, \delta_0, 0) \to 0 \quad \Rightarrow \quad \forall \delta > \delta_0: \quad \theta^t(x_o, \delta, 0) \to 0$$

- $\delta^*(x_o) = \inf_{\delta \in (0,1)} \{\delta: \ \theta^t(x_o, \delta, 0) \to 0\}$

kind of Phase Transition

D-AMP Without Noise

- Minimum of δ for recovery of every $x_o \in \mathcal{C}$?
- Proposition 1: $\{D_\sigma\}$ is κ-proper $\Rightarrow \sup_{x_o \in \mathcal{C}} \delta^*(x_o) \leq \kappa$.
- What to say about nearly proper denoisers?
  - Perfect recovery may not be possible!
- Lemma 2: If $\{D_\sigma\}$ is (κ, B)-near proper,

$$\delta > \kappa \quad \Rightarrow \quad \lim_{t \to \infty} \left(\sigma^t(x_o, \delta, 0)\right)^2 \leq \frac{B}{\delta - \kappa}$$

D-AMP With Noise

- Noisy setting: $\sigma_w \neq 0 \quad \Rightarrow \quad$ recovery is not exact

$$\Rightarrow \text{D-AMP fixed point} = \underbrace{\theta^{\infty}(x_o, \delta, \sigma_w^2)}_{\substack{\text{recovery error} \\ \text{in noisy setting}}} \neq 0$$

- Noise sensitivity: $\mathrm{NS}(\delta, \sigma_w^2) = \sup_{x_o \in \mathcal{C}} \theta^{\infty}(x_0, \delta, \sigma_w^2)$

- Proposition 2: If $\{D_\sigma\}$ is (κ, B)-near proper,

$$\delta > \kappa \quad \Rightarrow \quad \mathrm{NS}(\delta, \sigma_w^2) \leq \frac{\kappa \sigma_w^2 + B}{1 - \frac{\kappa}{\delta}}$$

Parameter Tuning

- Practical denoisers have a few free parameters…
  - Example: τ in soft thresholding

- State Evolution also depends on the parameters used in each iteration:

$$\theta^{t+1}(x_o, \delta, \sigma_w^2; \tau^0, \tau^1, \ldots, \tau^t) = \frac{1}{n}\mathbb{E}\left\|D_{\sigma^t,\tau^t}(x_o + \sigma^t \epsilon) - x_o\right\|_2^2$$

$$(\sigma^t)^2 = \frac{\theta^t}{\delta} + \sigma_w^2$$

- Optimal sequence of parameters at iteration $t+1$:

$$(\tau_*^0, \tau_*^1, \ldots, \tau_*^t) = \underbrace{\arg\min_{\tau^0,\ldots,\tau^t}}_{\text{joint optimization}} \theta^{t+1}(x_o, \delta, \sigma_w^2; \tau^0, \tau^1, \ldots, \tau^t)$$

- Lemma 3: Greedy parameter tuning is optimal for D-AMP.
  - No joint optimization. Can be done one by one.

Optimality of D-AMP

- Is there any algorithm that outperforms D-AMP?

- Uniform optimality:
  - $\mathcal{E}_\kappa = \left\{ \mathcal{C} : \exists \{ D_\sigma^{\mathcal{C}} \} \quad \sup_{\sigma^2} \sup_{x_o \in \mathcal{C}} \frac{\mathbb{E}\| D_\sigma(x_o + \sigma\epsilon) - x_o \|_2^2}{n\sigma^2} \leq \kappa \right\}$
  - D-AMP recovers all $x_o \in \mathcal{C} \in \mathcal{E}_\kappa$ from $\delta > \kappa$ measurements.
  - Can it be done with fewer measurements?

- Proposition 3: $\sup_{\mathcal{C} \in \mathcal{E}_\kappa} \frac{m^*(\mathcal{C})}{n} \geq \kappa$.
  - D-AMP is uniformly optimal!

Optimality of D-AMP

- Is there any algorithm that outperforms D-AMP?

- Single class optimality:
  - $\{D_\sigma^*\}$ is minimax optimal for D-AMP
    $\Leftrightarrow D_\sigma^* = \arg\min_{D_\sigma} \sup_{x_o \in \mathcal{C}} \theta_D^\infty(x_o, \delta, \sigma_w^2)$
  - $\{D_\sigma^M\}$ is minimax optimal for $\mathcal{C}$
    $\Leftrightarrow D_\sigma^M = \arg\min_{D_\sigma} \sup_{x_o \in \mathcal{C}} \mathbb{E}\|D_\sigma(x_o + \sigma\epsilon) - x_o\|_2^2$

- Proposition 4: $\{D_\sigma^*\} = \{D_\sigma^M\} \Rightarrow$ D*-AMP requires at least

$$n\kappa_{MM} = \sup_{\sigma^2} \frac{\min_{D_\sigma} \sup_{x_o \in \mathcal{C}} \mathbb{E}\|D_\sigma(x_o+\sigma\epsilon)-x_o\|_2^2}{\sigma^2}$$ measurements.

  - Not optimal in this sense.
    Some classes enjoy algorithms with fewer measurements.

Onsager Correction Term

- How to calculate $z^{t-1}\text{div}[D_{\hat{\sigma}^{t-1}}(x^{t-1} + \mathbf{\Phi}^H z^{t-1})]/m$ for an arbitrary denoiser?
  - Good denoisers don't have an explicit input-output relation!

- Monte Carlo method: If $b \sim \mathcal{N}(0, I)$, then

$$\text{div}\, D_{\sigma,\tau}(x) = \lim_{\epsilon \to 0} \mathbb{E}_b \left\{ b^T \left(\frac{D_{\sigma,\tau}(x+\epsilon b) - D_{\sigma,\tau}(x)}{\epsilon} \right) \right\}$$

$$\approx \mathbb{E}_b \left\{ \frac{1}{\epsilon} b^T \left(D_{\sigma,\tau}(x+\epsilon b) - D_{\sigma,\tau}(x) \right) \right\},\ \text{for very small } \epsilon.$$

in practice, replaced by mean of values for $b^1, \dots, b^M$. $M = 1$ would suffice for images.

Image Denoising Algorithms

- Gaussian kernel
- Bilateral filter
- Non-local means (NLM)
- Wavelet thresholding
- BLS-GSM
- BM3D
- BM3D-SAPCA

Simulation Results

10% Sampling	Lena	Barbara	Boat	Fingerprint	House	Peppers
AMP	18.47	17.67	18.96	15.87	19.98	17.50
Turbo-AMP	18.35	17.46	18.62	16.30	21.77	17.01
ALSB	25.30	24.01	22.44	16.25	31.09	24.01
NLR-CS	**26.74**	**24.95**	23.97	18.11	**34.46**	**25.21**
BM3D-IT	5.68	5.97	5.43	4.70	4.93	5.72
NLM-AMP	21.81	20.17	21.43	17.69	24.81	20.42
BLS-GSM-AMP	24.92	23.35	23.98	17.53	30.52	24.09
BM3D-AMP	26.01	24.24	**24.07**	**18.24**	34.12	24.41
BM3D-SAPCA-AMP	15.04	24.28	22.62	18.17	32.74	23.99

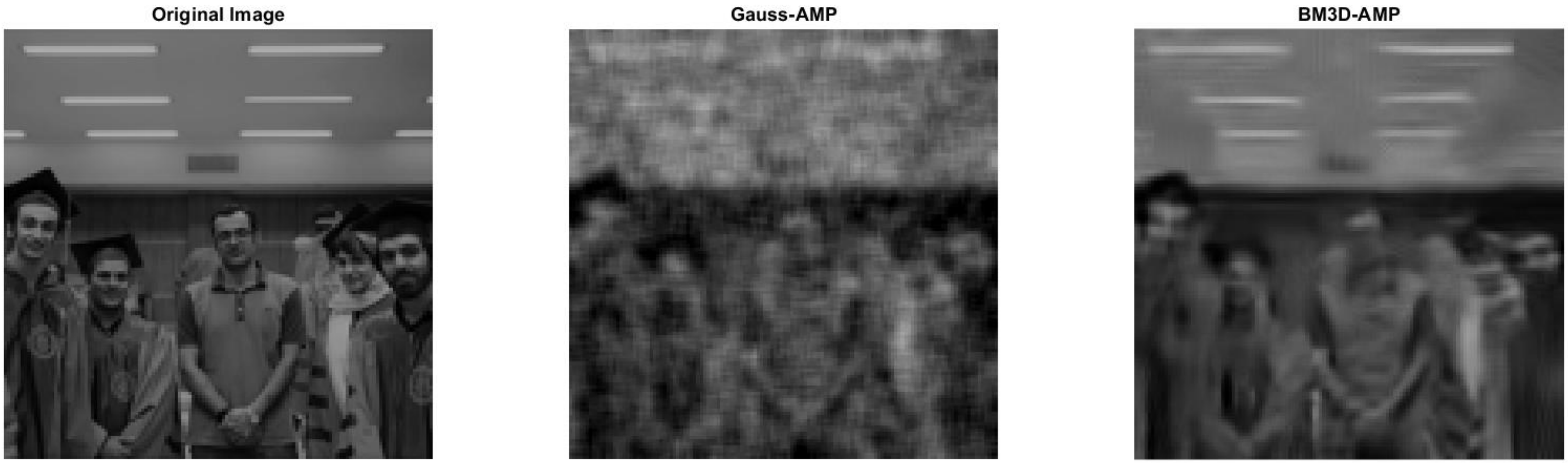
20% Sampling	Lena	Barbara	Boat	Fingerprint	House	Peppers
AMP	21.26	20.08	21.62	16.86	22.97	20.27
Turbo-AMP	23.48	21.45	23.36	16.31	28.20	21.78
ALSB	28.66	27.98	26.09	17.42	36.28	28.12
NLR-CS	31.88	30.31	26.96	21.10	38.70	30.42
BM3D-IT	25.64	24.38	23.79	6.63	32.92	23.87
NLM-AMP	27.73	24.27	23.97	19.72	31.75	23.70
BLS-GSM-AMP	29.77	28.00	27.06	18.45	35.76	29.14
BM3D-AMP	31.12	29.83	**27.58**	21.14	38.30	30.00
BM3D-SAPCA-AMP	**32.15**	**30.41**	27.35	**22.02**	**38.94**	**31.09**

30% Sampling	Lena	Barbara	Boat	Fingerprint	House	Peppers
AMP	23.90	22.70	23.67	17.57	26.15	23.12
Turbo-AMP	25.88	24.35	24.80	16.33	32.18	24.58
ALSB	31.91	30.69	28.69	22.76	38.51	31.85
NLR-CS	35.86	33.78	30.27	23.01	41.15	34.80
BM3D-IT	28.16	27.21	24.43	18.44	35.48	25.49
NLM-AMP	29.94	29.39	27.67	20.81	36.49	29.92
BLS-GSM-AMP	33.29	31.06	29.95	19.20	39.17	32.73
BM3D-AMP	34.87	33.14	30.60	22.95	40.92	33.83
BM3D-SAPCA-AMP	**36.21**	**34.18**	**31.22**	**23.71**	**41.55**	**34.91**

40% Sampling	Lena	Barbara	Boat	Fingerprint	House	Peppers
AMP	26.35	24.77	25.41	18.65	29.20	25.36
Turbo-AMP	27.91	26.11	26.98	16.65	35.37	26.83
ALSB	34.17	34.19	30.92	24.14	41.13	35.15
NLR-CS	39.07	36.99	32.75	24.78	43.45	37.63
BM3D-IT	29.50	28.22	25.13	19.47	36.94	28.86
NLM-AMP	32.58	32.15	28.94	21.53	38.62	31.47
BLS-GSM-AMP	36.50	34.33	32.41	20.32	40.84	35.86
BM3D-AMP	38.05	35.94	32.77	24.59	42.97	36.77
BM3D-SAPCA-AMP	**39.33**	**37.05**	**33.56**	**25.01**	**43.86**	**38.06**

50% Sampling	Lena	Barbara	Boat	Fingerprint	House	Peppers
AMP	28.12	27.19	27.44	19.84	31.86	27.99
Turbo-AMP	30.64	27.69	28.80	19.24	37.54	29.17
ALSB	36.95	37.10	32.96	25.80	42.76	38.11
NLR-CS	42.05	**39.86**	35.31	26.26	45.65	40.51
BM3D-IT	30.95	29.18	27.14	20.24	38.19	29.56
NLM-AMP	35.09	34.72	31.45	25.34	39.71	34.10
BLS-GSM-AMP	38.92	36.42	34.72	21.61	42.34	38.72
BM3D-AMP	40.89	38.21	35.07	25.99	44.91	39.38
BM3D-SAPCA-AMP	**42.12**	39.49	**36.05**	**26.76**	**45.70**	**40.61**

Simulation Results



δ = 0.1 PSNR = 23.5822 PSNR = 26.9664

Conclusion

- Generalization of CS problem to arbitrary classes of signals
- SE as an accurate tool for performance prediction
- D-AMP is robust to noise, easy to tune, and optimal
- Potential future works
  - Proof of Gaussianity of v^t (in high-dimensional problems)
  - Characterizing the classes of signals for which D-AMP is optimal
  - Extension to other measurement matrices such as Fourier matrix