Compressed Sensing and High-Resolution Image Inversion

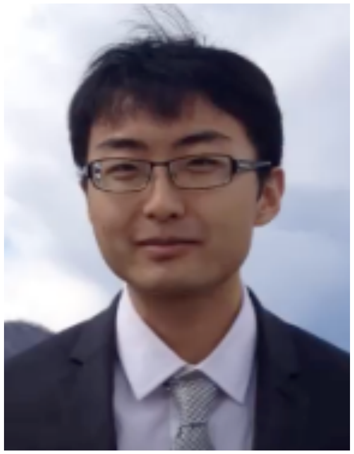
Ali Pezeshki
ECE and Mathematics
Colorado State University

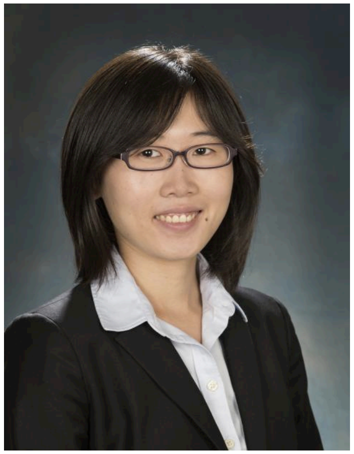
Matheon Workshop 2013
Compressed Sensing and Its Applications
TU Berlin, Berlin, Germany

December 9, 2013

Acknowledgements


Pooria Pakrooh


Wenbing Dang


Yuejie Chi


Louis Scharf

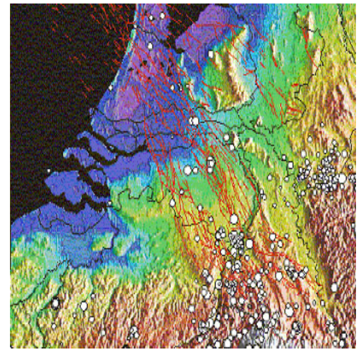
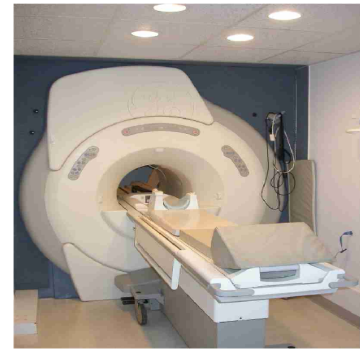

Robert Calderbank


Edwin Chong

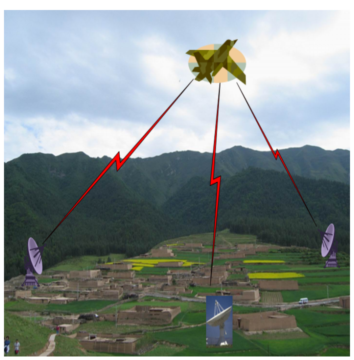
Supported by NSF under grants CCF-1018472 and CCF-1017431

Image Inversion



$$y(t) = \sum_{i=1}^{k} \psi(\nu_i)\theta_i \quad \longrightarrow \quad \{y(t_i)\}_{i=1}^{m} \quad \longrightarrow \quad \{\hat{\nu}_i, \hat{\theta}_i\}_{i=1}^{k}$$




Classical Inversion vs Compressed Sesning

- **Classical: Matched filtering**
  - Sequence of rank-one subspaces, or one-dimensional test images, is matched to the measured image by filtering or correlating or phasing.

Classical Inversion vs Compressed Sesning

- **Classical: Matched filtering**
  - Sequence of rank-one subspaces, or one-dimensional test images, is matched to the measured image by filtering or correlating or phasing.
  - Test images is generated by scanning a prototype image (e.g., a waveform or a steering vector) through frequency, wavenumber, doppler, and/or delay.

Classical Inversion vs Compressed Sesning

- **Classical: Matched filtering**
  - Sequence of rank-one subspaces, or one-dimensional test images, is matched to the measured image by filtering or correlating or phasing.
  - Test images is generated by scanning a prototype image (e.g., a waveform or a steering vector) through frequency, wavenumber, doppler, and/or delay.
  - Extends to subspace matching for those cases in which the model for the image is comprised of several dominant modes.

Classical Inversion vs Compressed Sesning

- **Classical: Matched filtering**
  - Sequence of rank-one subspaces, or one-dimensional test images, is matched to the measured image by filtering or correlating or phasing.
  - Test images is generated by scanning a prototype image (e.g., a waveform or a steering vector) through frequency, wavenumber, doppler, and/or delay.
  - Extends to subspace matching for those cases in which the model for the image is comprised of several dominant modes.
  - Extends to whitened matched filter, or minimum variance unbiased (MVUB) filter, or generalized sidelobe canceller.

Classical Inversion vs Compressed Sesning

- **Classical: Estimation in Separable Model**
  - Low-order separable modal representation for the field.
  - Estimates of linear parameters (complex amplitudes of modes) and nonlinear mode parameters (frequency, wavenumber, delay, and/or doppler) are extracted, usually based on maximum likelihood, or some variation on linear prediction, using l2 minimization.

Classical Inversion vs Compressed Sesning

- **Classical: Estimation in Separable Model**
  - Low-order separable modal representation for the field.
  - Estimates of linear parameters (complex amplitudes of modes) and nonlinear mode parameters (frequency, wavenumber, delay, and/or doppler) are extracted, usually based on maximum likelihood, or some variation on linear prediction, using l2 minimization.
  - SNR, Fisher Information, Cramer-Rao Bound (CRB), Kullback-Leibler divergence, Bayesian CRB, Threshold Effects.

Classical Inversion vs Compressed Sesning

- **Classical: Estimation in Separable Model**
  - Low-order separable modal representation for the field.
  - Estimates of linear parameters (complex amplitudes of modes) and nonlinear mode parameters (frequency, wavenumber, delay, and/or doppler) are extracted, usually based on maximum likelihood, or some variation on linear prediction, using l2 minimization.
  - SNR, Fisher Information, Cramer-Rao Bound (CRB), Kullback-Leibler divergence, Bayesian CRB, Threshold Effects.
- **Sampling:** Any subsampling of the measured image has consequences for resolution (or bias) and for variability (or variance).

Classical Inversion vs Compressed Sesning

- **Compressed Sensing:**
  - Subsampling has manageable consequences for image inversion provided we have known sparsity structure.
  - Typically employs randomly drawn linear combinations.

Agenda

1. **Fisher Information:** What is the impact of compressive sampling on Fisher information and Cramer-Rao bound (CRB) for estimating nonlinear parameters?

Agenda

1. **Fisher Information:** What is the impact of compressive sampling on Fisher information and Cramer-Rao bound (CRB) for estimating nonlinear parameters?

2. **Breakdown Thresholds:** What is the impact of compressive sampling on SNR thresholds at which mean-squared error in estimating parameters deviate sharply from the CRB?

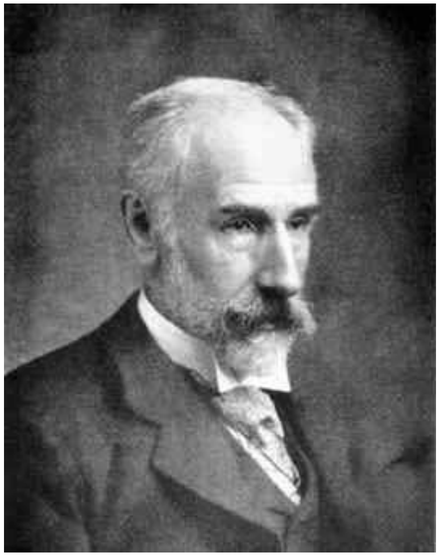
Agenda

1. **Fisher Information:** What is the impact of compressive sampling on Fisher information and Cramer-Rao bound (CRB) for estimating nonlinear parameters?

2. **Breakdown Thresholds:** What is the impact of compressive sampling on SNR thresholds at which mean-squared error in estimating parameters deviate sharply from the CRB?

3. **Model mismatch:** What is the sensitivity of compressed sensing to model mismatch? Can these sensitivities be mitigated?

Fisher Information and Cramer-Rao Bound

Fisher Information and Cramer-Rao Bound

- **Fisher Information:**

$$\mathbf{J}(\boldsymbol{\theta}) = E[(\frac{\partial \log f(\mathbf{y};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}})(\frac{\partial \log f(\mathbf{y};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}})^H].$$

- **Cramer-Rao Bound:** The inverse $\mathbf{J}^{-1}(\boldsymbol{\theta})$ lower bounds the error covariance matrix for any unbiased estimator of $\boldsymbol{\theta}$.

CS, Fisher Information, and CRB

- Complex Normal model:

$$\mathbf{y} = \mathbf{s}(\boldsymbol{\theta}) + \mathbf{n} \in \mathbb{C}^n; \quad \mathbf{y} = \mathcal{CN}_n[\mathbf{s}(\boldsymbol{\theta}), \mathbf{R}]$$

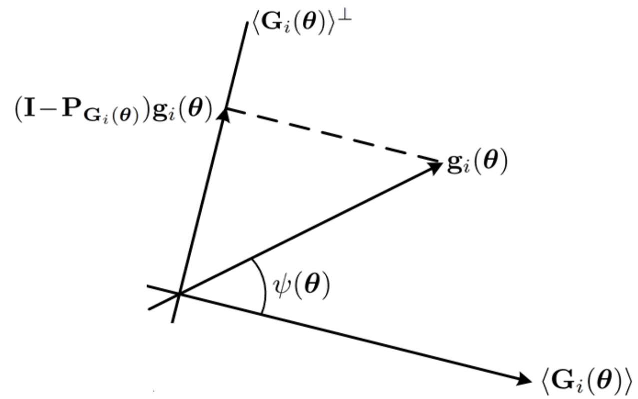
- Fisher information matrix:

$$\begin{aligned}\mathbf{J}(\boldsymbol{\theta}) &= \mathbf{G}^H(\boldsymbol{\theta})\mathbf{R}^{-1}\mathbf{G}(\boldsymbol{\theta}) \\ &= \frac{1}{\sigma^2}\mathbf{G}^H(\boldsymbol{\theta})\mathbf{G}(\boldsymbol{\theta}), \quad \text{when} \quad \mathbf{R} = \sigma^2\mathbf{I}\end{aligned}$$

$$\mathbf{G}(\boldsymbol{\theta}) = [\mathbf{g}_1(\boldsymbol{\theta}), \ldots, \mathbf{g}_k(\boldsymbol{\theta})]; \quad \mathbf{g}_i(\boldsymbol{\theta}) = \frac{\partial \mathbf{s}(\boldsymbol{\theta})}{\partial \theta_i}$$

- Cramer-Rao lower bound:

$$(\mathbf{J}^{-1}(\boldsymbol{\theta}))_{ii} = \sigma^2(\mathbf{g}_i^T(\boldsymbol{\theta})(\mathbf{I} - \mathbf{P}_{\mathbf{G}_i(\boldsymbol{\theta})})\mathbf{g}_i(\boldsymbol{\theta}))^{-1}$$



When one sensitivity looks likes a linear combination of others, performance is poor.

CS, Fisher Information, and CRB

- Compressive measurement:

$$\mathbf{z} = \boldsymbol{\Phi}\mathbf{y} = \boldsymbol{\Phi}[\mathbf{s}(\boldsymbol{\theta}) + \mathbf{n}] \in \mathbb{C}^m;$$

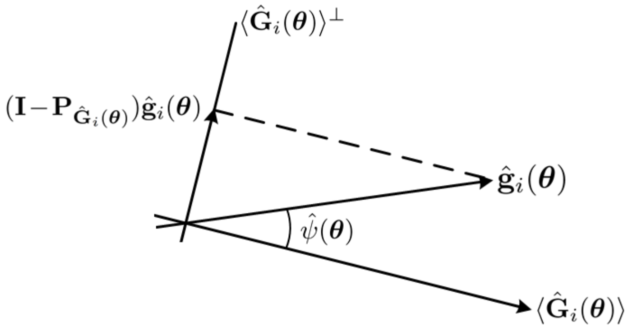
- Fisher information matrix

$$\hat{\mathbf{J}}(\boldsymbol{\theta}) = \frac{1}{\sigma^2}\mathbf{G}^H(\boldsymbol{\theta})\mathbf{P}_{\boldsymbol{\Phi}^H}\mathbf{G}(\boldsymbol{\nu}) = \hat{\mathbf{G}}^H(\boldsymbol{\theta})\hat{\mathbf{G}}(\boldsymbol{\theta})$$

$$\hat{\mathbf{G}}(\boldsymbol{\theta}) = [\hat{\mathbf{g}}_1(\boldsymbol{\theta}), \ldots, \hat{\mathbf{g}}_k(\boldsymbol{\theta})]; \quad \hat{\mathbf{g}}_i(\boldsymbol{\theta}) = \mathbf{P}_{\boldsymbol{\Phi}^H}\frac{\partial \mathbf{s}(\boldsymbol{\theta})}{\partial \theta_i}$$

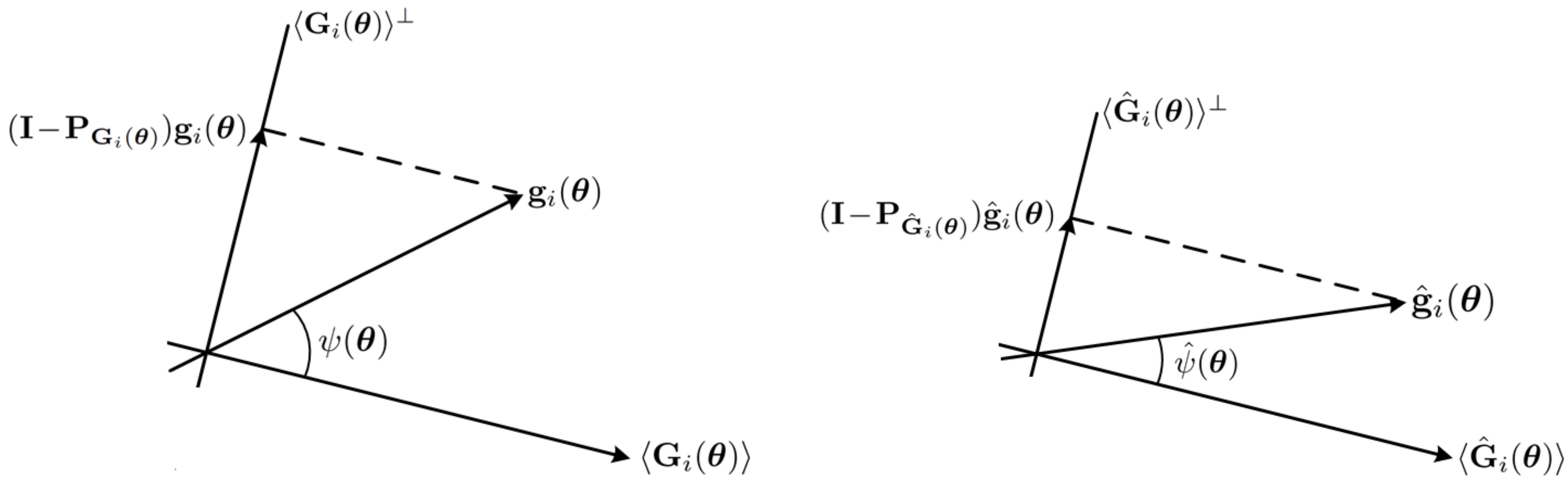
- Cramer-Rao lower bound:

$$(\hat{\mathbf{J}}^{-1}(\boldsymbol{\theta}))_{ii} = \sigma^2(\hat{\mathbf{g}}_i^T(\boldsymbol{\theta})(\mathbf{I} - \mathbf{P}_{\hat{\mathbf{G}}_\mathbf{i}(\boldsymbol{\theta})})\hat{\mathbf{g}}_i(\boldsymbol{\theta}))^{-1}$$



Compressive measurement reduces the distance between subspaces: loss of information.

CS, Fisher Information, and CRB



- **Question:** What is the impact of compressive sampling on the Fisher information matrix, Cramer-Rao bound (CRB), and Kullback-Leibler divergence for estimating parameters?

Johnson-Lindenstrauss Lemma

- **JL Lemma:** For any $\epsilon \in (0,1)$, a random linear transformation $\mathbf{\Phi} : \mathbb{R}^n \to \mathbb{R}^m$ is said to satisfy an $\epsilon-$JL type Lemma over a set of vectors $\mathcal{Q} \subset \mathbb{R}^n$ with probability at least $1-\delta$ if

$$Pr\left(\forall \mathbf{q} \in \mathcal{Q} : (1-\epsilon)\|\mathbf{q}\|_2^2 \leq \|\mathbf{\Phi q}\|_2^2 \leq (1+\epsilon)\|\mathbf{q}\|_2^2\right) \geq 1-\delta.$$

- For random matrices with i.i.d. $\mathcal{N}(0,1/m)$ entries Φ_{ij}, we have $\delta \leq 2|\mathcal{Q}|e^{-mc_0(\epsilon)}$ where $c_0(\epsilon) = \epsilon^2/4 - \epsilon^3/6$ [Baraniuk, Davenport, Devore, and Wakin '08; Dasgupta and Gupta '02].

Subspace Johnson-Lindenstrauss Lemma

- **Subspace JL Lemma:** [Sarlos '06] Let $\mathbf{\Phi} : \mathbb{R}^n \to \mathbb{R}^m$, $m < n$, and $\epsilon \in (0, 1)$. Then $\mathbf{\Phi}$ satisfies the $\epsilon-$JL type Lemma over any arbitrary p-dimensional subspace $\langle \mathbf{V} \rangle$ of $\mathbb{R}^n$ with probability at least $1 - \delta$, provided that it satisfies the $\epsilon'-$JL type Lemma over any set $\mathcal{Q} \subset \mathbb{R}^n$ of $\lceil (2\sqrt{p}/\epsilon')^p \rceil$ vectors with probability at least $1 - \delta$, where ϵ' satisfies

$$(\frac{3\epsilon'}{1-\epsilon'})^2 + 2(\frac{3\epsilon'}{1-\epsilon'}) = \epsilon.$$

CS, Fisher Information, and CRB

Theorem: [Pakrooh, P., Scharf, Chi '13]

(a) For any compression matrix, we have

$$(\mathbf{J}^{-1}(\boldsymbol{\theta}))_{ii} \leq (\hat{\mathbf{J}}^{-1}(\boldsymbol{\theta}))_{ii} \leq 1/\lambda_{min}(\mathbf{G}^T(\boldsymbol{\theta})\mathbf{P}_{\boldsymbol{\Phi}^T}\mathbf{G}(\boldsymbol{\theta}))$$

(b) For a random compression matrix, we have

$$(\hat{\mathbf{J}}^{-1}(\boldsymbol{\theta}))_{ii} \leq \frac{\lambda_{max}(\mathbf{J}^{-1}(\boldsymbol{\theta}))}{C(1-\epsilon)}$$

with probability at least $1-\delta-\delta'$, where

- $1-\delta$ is the lower bound on the probability that $\boldsymbol{\Phi}$ satisfies the $\epsilon-$JL type lemma for any p-dimensional subspace, and
- $1-\delta'$ is the probability that $\lambda_{min}((\boldsymbol{\Phi}\boldsymbol{\Phi}^T)^{-1})$ is larger than C.

CS, Fisher Information, and CRB

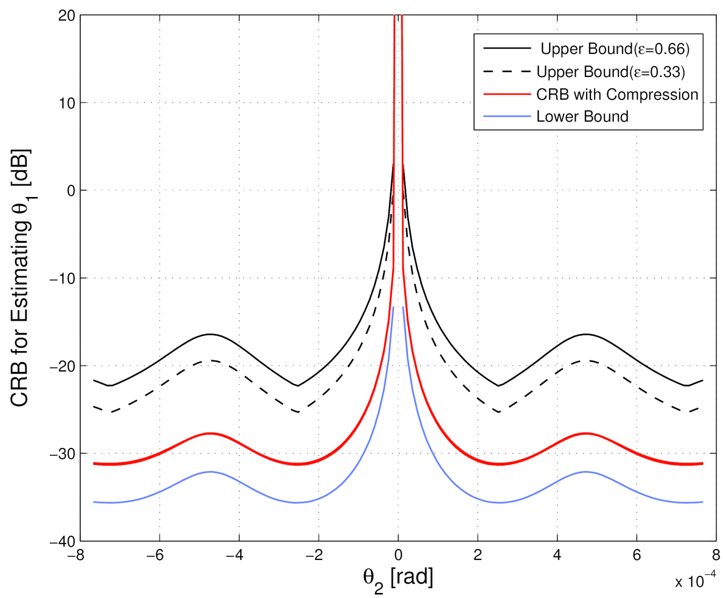
- For $\mathrm{tr}\,(\hat{\mathbf{J}}^{-1}(\boldsymbol{\theta}))$ we have

$$\mathrm{tr}(\mathbf{J}^{-1}(\boldsymbol{\theta})) \leq \mathrm{tr}(\hat{\mathbf{J}}^{-1}(\boldsymbol{\theta})) \leq \frac{p\lambda_{max}(\mathbf{J}^{-1}(\boldsymbol{\theta}))}{C(1-\epsilon)}$$

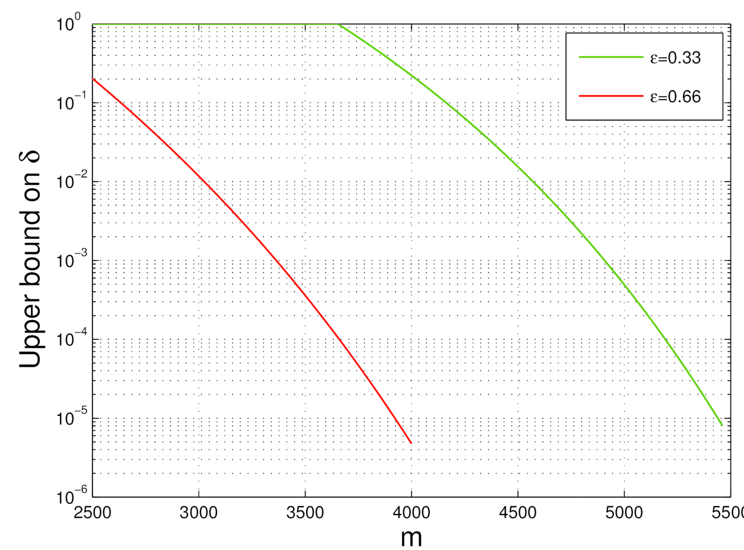
  where again the upper bound holds with probability at least $1-\delta-\delta'$.

- We can also bound $\det(\hat{\mathbf{J}}^{-1}(\boldsymbol{\theta}))$.

CRB after Compression



Bounds on the CRB for $-2\pi/n \leq \theta_2 \leq 2\pi/n$, $m = 3000$, $n = 8192$



Upper bounds on δ versus the number of measurements m for $n = 8192$ and $\epsilon = 0.66$ (red) and $\epsilon = 0.33$ (green)

CS and Kullback-Leibler Divergence

KL divergence between $\mathcal{N}(\mathbf{x}(\boldsymbol{\theta}), \mathbf{R})$ and $\mathcal{N}(\mathbf{x}(\boldsymbol{\theta}'), \mathbf{R})$:

$$D(\boldsymbol{\theta}, \boldsymbol{\theta}') = \frac{1}{2}[(\mathbf{x}(\boldsymbol{\theta}) - \mathbf{x}(\boldsymbol{\theta}'))^T \mathbf{R}^{-1} (\mathbf{x}(\boldsymbol{\theta}) - \mathbf{x}(\boldsymbol{\theta}'))].$$

- After compression with $\boldsymbol{\Phi}$:

$$\hat{D}(\boldsymbol{\theta}, \boldsymbol{\theta}') = \frac{1}{2}[(\mathbf{x}(\boldsymbol{\theta}) - \mathbf{x}(\boldsymbol{\theta}'))^T \boldsymbol{\Phi}^T (\boldsymbol{\Phi} \mathbf{R} \boldsymbol{\Phi}^T)^{-1} \boldsymbol{\Phi} (\mathbf{x}(\boldsymbol{\theta}) - \mathbf{x}(\boldsymbol{\theta}'))].$$

- With white noise $\mathbf{R} = \sigma^2 \mathbf{I}$:

$$\hat{D}(\boldsymbol{\theta}, \boldsymbol{\theta}') = \frac{1}{2\sigma^2}[(\mathbf{x}(\boldsymbol{\theta}) - \mathbf{x}(\boldsymbol{\theta}'))^T \mathbf{P}_{\boldsymbol{\Phi}^T} (\mathbf{x}(\boldsymbol{\theta}) - \mathbf{x}(\boldsymbol{\theta}'))].$$

Theorem: [Pakrooh, P., Scharf, and Chi (ICASSP'13)]

$$C(1-\epsilon) D(\boldsymbol{\theta}, \boldsymbol{\theta}') \leq \hat{D}(\boldsymbol{\theta}, \boldsymbol{\theta}') \leq D(\boldsymbol{\theta}, \boldsymbol{\theta}')$$

with probability at least $1 - \delta - \delta'$, where δ, δ'.

References on CS, Fisher Infromation, and CRB

- L. L. Scharf, E. K. P. Chong, A. Pezeshki, and J. R. Luo, "Compressive sensing and sparse inversion in signal processing: Cautionary notes," *in Proc. 7th Workshop on Defence Applications of Signal Processing (DASP), Coolum, Queensland, Australia, Jul. 10-14, 2011.*

- L. L. Scharf, E. K. P. Chong, A. Pezeshki, and J. R. Luo, "Sensitivity considerations in compressed sensing," *in Conf. Rec. 45th Annual Asilomar Conf. Signals, Systs., Computs., Pacific Grove, CA,, Nov. 2011, pp. 744–748.*

- P. Pakrooh, L. L. Scharf, A. Pezeshki and Y. Chi, "Analysis of Fisher information and the Cramer-Rao bound for nonlinear parameter estimation after compressed sensing", *in Proc. 2013 IEEE Int. Conf. on Acoust., Speech and Signal Process. (ICASSP), Vancouver May 26-31, 2013.*

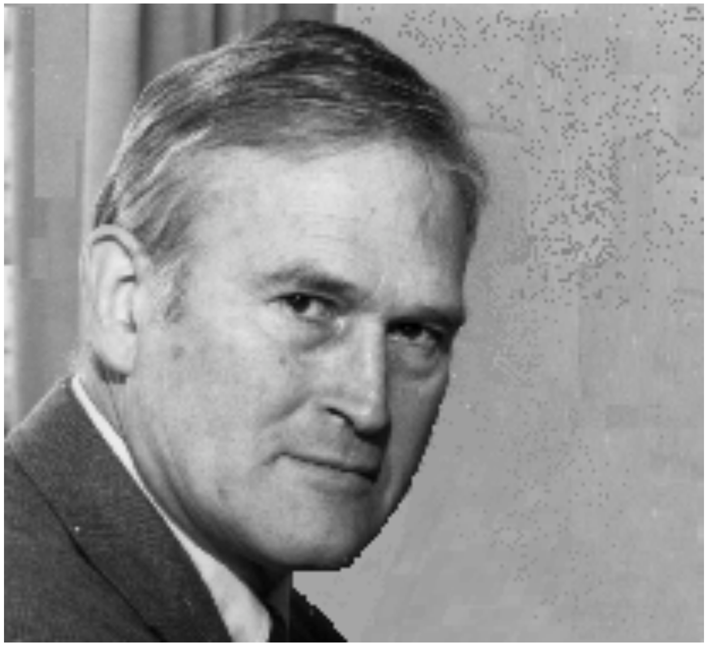
Work by Others

- Nielsen, Christensen, and Jensen (ICASSP'12): Bounds on mean value of Fisher Information after random compression.
- Ramasamy, Venkateswaran, and Madhow (Asilomar'12): Bounds on Fisher information after compression in a different noisy model.
- Babadi, Kalouptsidis, and Tarokh (TSP 2009): Existence of an estimator ("Joint Typicality Estimator") that asymptotically achieves the CRB in linear parameter estimation with random Gaussian compression matrices.

Breakdown Threshold and Subspace Swaps

Breakdown Threshold and Subspace Swaps

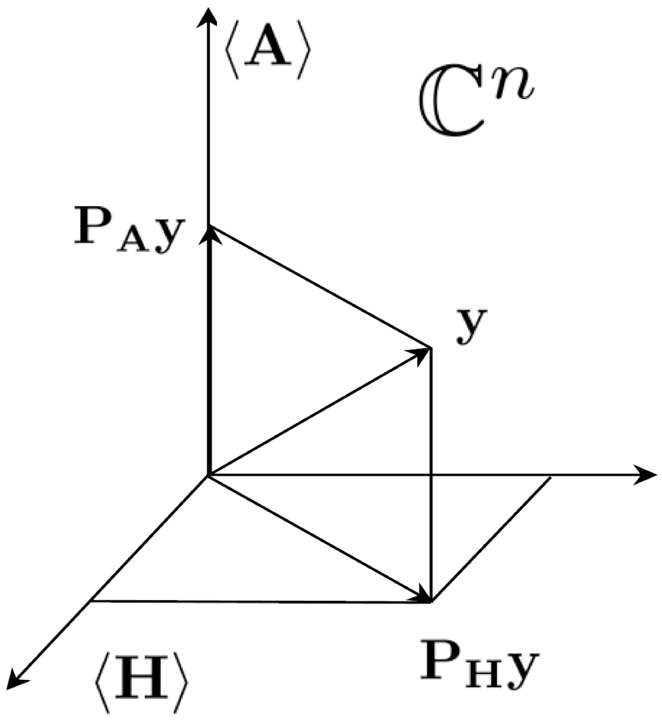
- Threshold effect: Sharp deviation of Mean Squared Error (MSE) performance from Cramer-Rao Bound (CRB).
- Breakdown threshold: SNR at which a threshold effect occurs with non-negligible probability.



Donald W. Tufts (1933-2012)

Breakdown Threshold and Subspace Swaps

- Subspace Swap: Event in which measured data is more accurately resolved by one or more modes of an *orthogonal subspace* to the signal subspace.
- Cares only about what the data itself is saying.
- Bound probability of a subspace swap to predict breakdown SNRs.



Signal Model: Mean Case

- Before compression:

$$\mathbf{y} : \mathcal{CN}_n[\mathbf{Hu}, \sigma^2 \mathbf{I}]$$

- After compression with left-orthogonal $\mathbf{\Phi} \in \mathbb{C}^{m \times n}$, $m < n$:

$$\mathbf{y} : \mathcal{CN}_m[\mathbf{Gu}, \sigma^2 \mathbf{I}], \ \ \mathbf{G} = \mathbf{\Phi H}$$

Signal Model: Covariance Case

- Before compression:

$$\mathbf{y} : \mathcal{CN}_n[\mathbf{0}, \mathbf{H}\mathbf{R}_{uu}\mathbf{H}^H + \sigma^2\mathbf{I}]$$

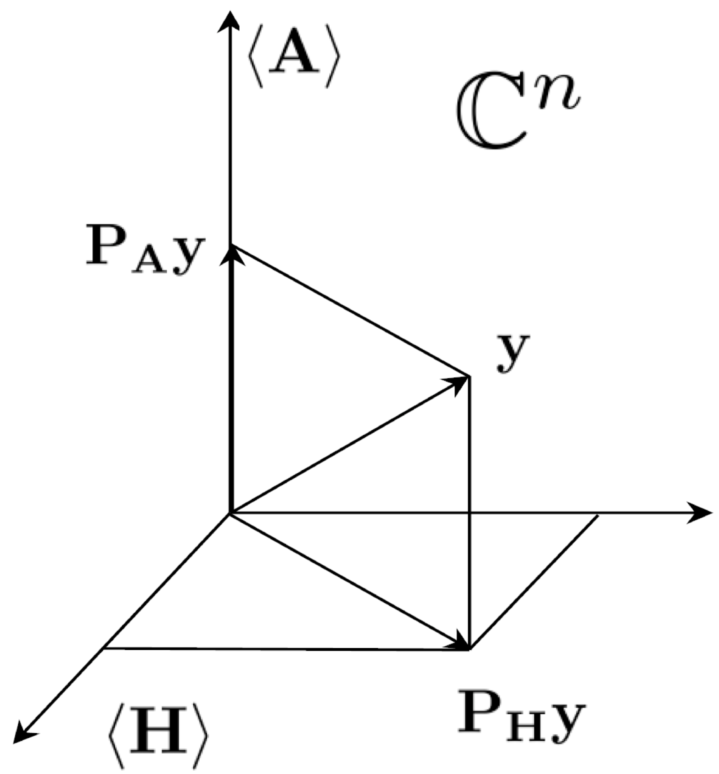
- After compression with left-orthogonal $\mathbf{\Phi} \in \mathbb{C}^{m \times n}$, $m < n$:

$$\mathbf{y} : \mathrm{CN}_m[\mathbf{0}, \mathbf{G}\mathbf{R}_{uu}\mathbf{G}^H + \sigma^2\mathbf{I}], \ \ \mathbf{G} = \mathbf{\Phi}\mathbf{H}$$

  Assume data consists of L iid realizations of $\mathbf{y}$ arranged as $\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, \cdots, \mathbf{y}_L]$.

Subspace Swap Events

- Subspace Swap Event E: One or more modes of the orthogonal subspace $\langle \mathbf{A} \rangle$ resolves more energy than one or more modes of the noise-free signal subspace $\langle \mathbf{H} \rangle$.



Subspace Swap Events

- Subevent F: Average energy resolved in the orthogonal subspace $\langle \mathbf{A} \rangle$ is greater than the average energy resolved in the noise-free signal subspace $\langle \mathbf{H} \rangle$.

$$\min_i |\mathbf{h}_i^H \mathbf{y}|^2 \leq \frac{1}{p} \mathbf{y}^H \mathbf{P}_H \mathbf{y} < \frac{1}{n-p} \mathbf{y}^H \mathbf{P}_A \mathbf{y} \leq \max_i |\mathbf{a}_i^H \mathbf{y}|^2$$

- Subevent G: Energy resolved in the apriori minimum mode $\mathbf{h}_{min}$ of the noise-free signal subspace $\langle \mathbf{H} \rangle$ is smaller than the average energy resolved in the orthogonal subspace $\langle \mathbf{A} \rangle$.

$$|\mathbf{h}_{min}^H \mathbf{y}|^2 < \frac{1}{n-p} \mathbf{y}^H \mathbf{P}_A \mathbf{y} \leq \max_i |\mathbf{a}_i^H \mathbf{y}|^2.$$

Probability of Subspace Swap: Mean Case

Theorem: [Pakrooh, P., Scharf (GlobalSIP'13)]

- (a) Before compression:

$$P_{ss} \geq 1 - P[\frac{\mathbf{y}^H \mathbf{P}_H \mathbf{y}/p}{\mathbf{y}^H \mathbf{P}_A \mathbf{y}/(n-p)} > 1]$$
$$= 1 - P[F_{2p,2(n-p)}(\|\mathbf{Hu}\|_2^2/\sigma^2) > 1]$$

$\|\mathbf{Hu}\|_2^2/\sigma^2$ is the SNR before compression.

- (b) After compression:

$$P_{ss} \geq 1 - P[F_{2p,2(m-p)}(\|\mathbf{Gu}\|_2^2/\sigma^2) > 1]$$

$\|\mathbf{Gu}\|_2^2/\sigma^2$ is the SNR after compression, $\mathbf{G} = \mathbf{\Phi H}$.

Probablity of Subspace Swap: Covariance Case

Theorem: [Pakrooh, P., Scharf (GlobalSIP'13)]

- (a) Before compression:

$$P_{ss} \geq 1 - P[\frac{tr(\mathbf{Y}^H \mathbf{P}_H \mathbf{Y}/pL)}{tr(\mathbf{Y}^H \mathbf{P}_A \mathbf{Y}/(n-p)L)} > 1]$$
$$= 1 - P[F_{2pL,2(n-p)L} > \frac{1}{1+\lambda_p/\sigma^2}].$$

$\lambda_p = ev_{min}(\mathbf{H}\mathbf{R}_{uu}\mathbf{H}^H)$

λ_p/σ^2: Effective SNR before compression

- (b) After compression:

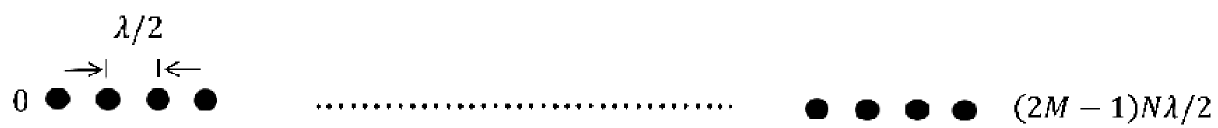
$$P_{ss} \geq 1 - P[F_{2pL,2(m-p)L} > \frac{1}{1+\lambda'_p/\sigma^2}].$$

$\lambda'_p = ev_{min}(\mathbf{G}\mathbf{R}_{uu}\mathbf{G}^H)$

λ'_p/σ^2: Effective SNR after compression

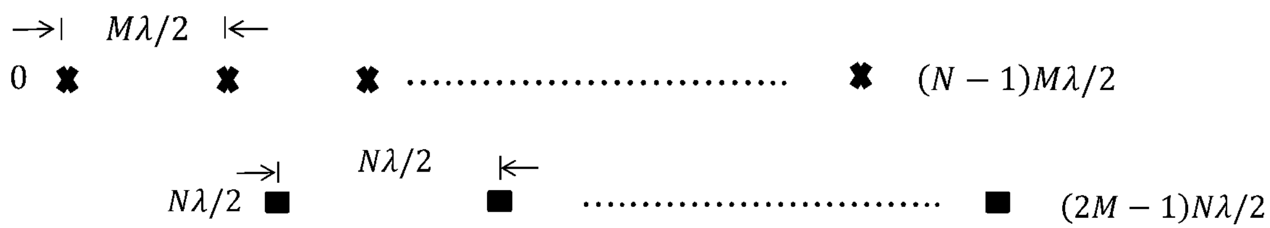
Sensor Array Processing: Dense, Gaussian, and Co-prime

- Dense array



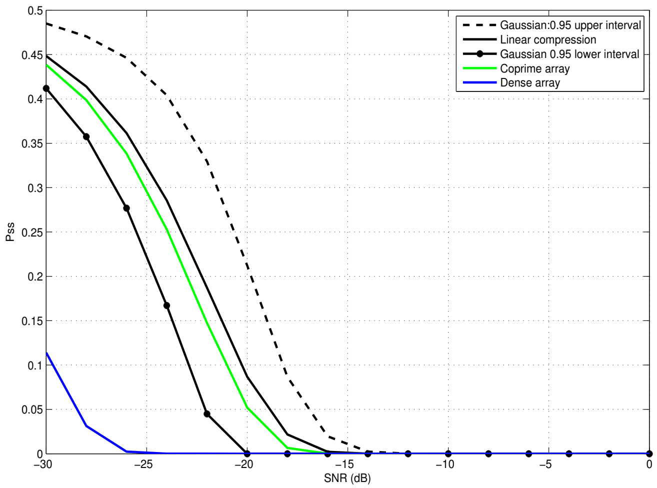
- Guassian compression

- Co-prime compression [Pal and Vaidyanathan (2011)]

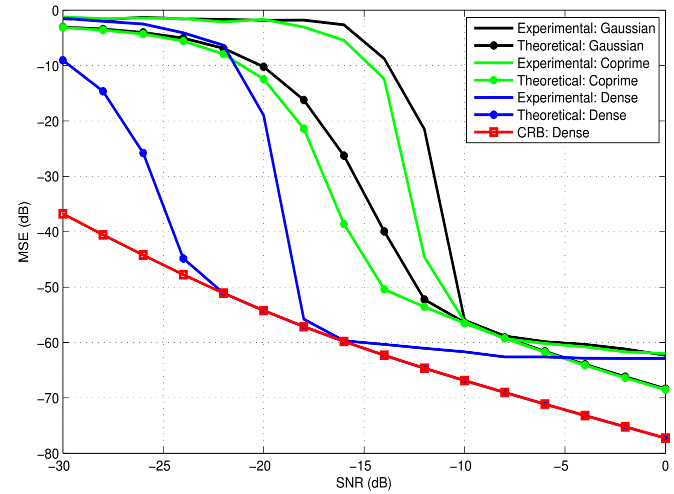


At $N = 11$ and $M = 9$, $(2M-1)N\lambda/2 = 187\lambda/2$.

Sensor Array Processing–Mean Case

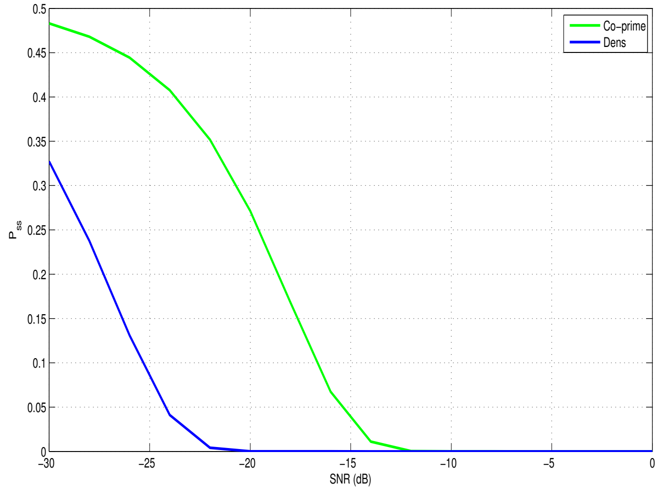


Analytical lower bounds for the probability of subspace swap.

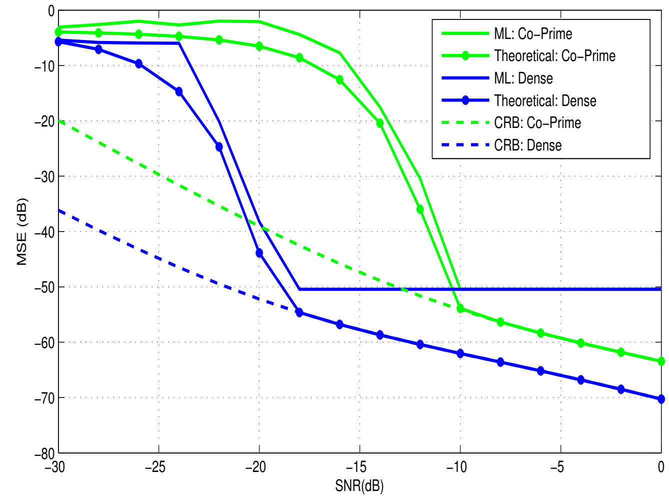


MSE and MSE bounds; Interfering source at $\theta_2 = \pi/188$; Average over 200 trials.

Sensor Array Processing–Covariance Case



Analytical lower bounds for the probability of subspace swap.



MSE and MSE bounds; Interfering source at $\theta_2 = \pi/188$; 200 snapshots; Averaged over 500 trials.

References on Breakdown Thresholds

- P. Pakrooh, A. Pezeshki, and L. L. Scharf, "Threshold effects in parameter estimation from compressed data," *Proc. 1st IEEE Global Conference on Signal and Information Processing*, Austin, TX, Dec. 2013.
- D. Tufts, A. Kot, and R. Vaccaro, The threshold effect in signal processing algorithms which use an estimated subspace, *SVD and Signal Processing II: Algorithms, Analysis and Applications*, New York: Elsevier, 1991, pp. 301320.
- J. K. Thomas, L. L. Scharf, and D. W. Tufts, The probability of a subspace swap in the SVD, *IEEE Transactions on Signal Processing*, vol. 43, no. 3, pp. 730736, Mar. 1995.
- B. A. Johnson, Y. I. Abramovich, and X. Mestre, MUSIC, G-MUSIC, and maximum-likelihood performance breakdown, *IEEE Transactions on Signal Processing*, vol. 56, no. 8, pp. 3944-3958, Aug. 2008.

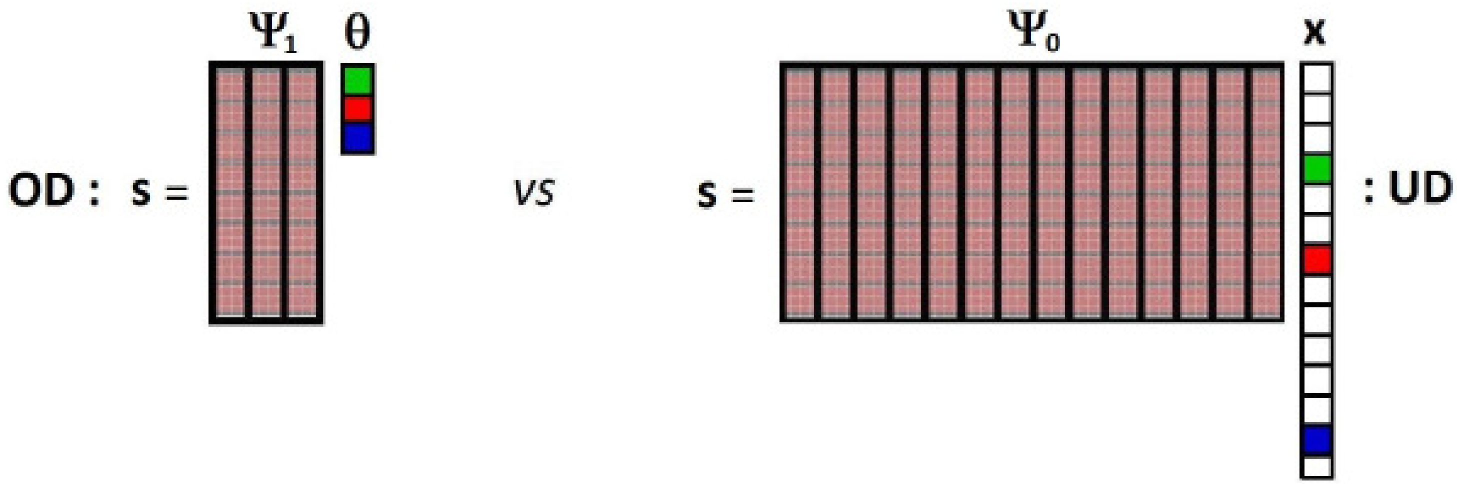
Intermediate Recap

- Compression, whether by linear maps (eg, Gaussian or Bernoulli) or by subsampling (eg, co-prime) has performance consequences.
- The CR bound increases and the onset of threshold SNR increases. These increases may be quantified to determine where compressive sampling is viable.

Model Mismatch

From Over-determined to Under-determined

$$\mathbf{y} = \sum_{i=1}^{K} \psi(\boldsymbol{\nu}_i)\theta_i \qquad\qquad \mathbf{y} \approx [\psi(\omega_1), \cdots, \psi(\omega_n)] \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$$

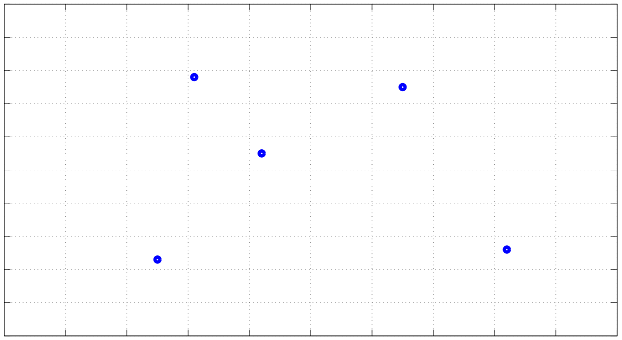


Model Mismatch

Mathematical (CS) model:

$$\mathbf{s} = \mathbf{\Psi}_0 \mathbf{x}$$

The basis (or frame) $\mathbf{\Psi}_0$ is assumed, typically a gridded imaging matrix (e.g., n point DFT matrix or identity matrix), and $\mathbf{x}$ is presumed to be k-sparse.



Physical (true) model:

$$\mathbf{s} = \mathbf{\Psi}_1 \boldsymbol{\theta}$$

The basis $\mathbf{\Psi}_1$ is unknown, and is determined by a point spread function, a Green's function, or an impulse response, and $\boldsymbol{\theta}$ is a k-sparse and unknown.

Key transformation:

$$\mathbf{x} = \mathbf{\Psi} \boldsymbol{\theta} = \mathbf{\Psi}_0^{-1} \mathbf{\Psi}_1 \boldsymbol{\theta}$$

$\mathbf{x}$ is sparse in the unknown $\mathbf{\Psi}$ basis, not in the identity basis.

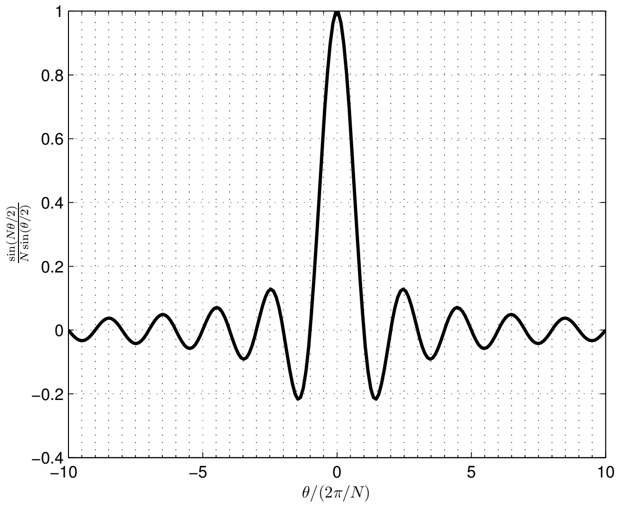
Model Mismatch: From Sparse to Incompressible

DFT Grid Mismatch:

$$\mathbf{\Psi} = \mathbf{\Psi}_0^{-1}\mathbf{\Psi}_1 = \begin{bmatrix} L(\Delta\theta_0 - 0) & L(\Delta\theta_1 - \frac{2\pi(n-1)}{n}) & \cdots & L(\Delta\theta_{n-1} - \frac{2\pi}{n}) \\ L(\Delta\theta_0 - \frac{2\pi}{n}) & L(\Delta\theta_1 - 0) & \cdots & L(\Delta\theta_{n-1} - \frac{2\pi\cdot 2}{n}) \\ \vdots & \vdots & \ddots & \vdots \\ L(\Delta\theta_0 - \frac{2\pi(n-1)}{n}) & L(\Delta\theta_1 - \frac{2\pi(n-2)}{n}) & \cdots & L(\Delta\theta_{n-1} - 0) \end{bmatrix}$$

where $L(\theta)$ is the Dirichlet kernel:

$$L(\theta) = \frac{1}{n}\sum_{\ell=0}^{n-1} e^{j\ell\theta} = \frac{1}{n}e^{j\frac{\theta(n-1)}{2}}\frac{\sin(\theta n/2)}{\sin(\theta/2)}.$$



Slow decay of the Dirichlet kernel means that the presumably sparse vector $\mathbf{x} = \mathbf{\Psi}\boldsymbol{\theta}$ is in fact incompressible.

Sensitivity to Model Mismatch

- Question: What is the consequence of assuming that $\mathbf{x}$ is k-sparse in $\mathbf{I}$, when in fact it is only k-sparse in an *unknown* basis $\mathbf{\Psi}$, which is determined by the mismatch between $\mathbf{\Psi}_0$ and $\mathbf{\Psi}_1$?
- CS Inverter: Basis pursuit solution satisfies

$$\begin{aligned} \text{Noise-free:} \quad & \|\mathbf{x}^* - \mathbf{x}\|_1 \leq C_0 \|\mathbf{x} - \mathbf{x}_k\|_1 \\ \text{Noisy:} \quad & \|\mathbf{x}^* - \mathbf{x}\|_2 \leq C_0 k^{-1/2} \|\mathbf{x} - \mathbf{x}_k\|_1 + C_1 \epsilon \end{aligned}$$

  where $\mathbf{x}_k$ is the best k-term approximation to $\mathbf{x}$.
- **Key:** Analyze the sensitivity of $\|\mathbf{x} - \mathbf{x}_k\|_1$ to basis mismatch.

Sensitivity to Model Mismatch

Theorem: [Chi, Scharf, P., Calderbank (TSP 2011)] Let $\mathbf{\Psi} = \mathbf{\Psi}_0^{-1}\mathbf{\Psi}_1 = \mathbf{I} + \mathbf{E}$, where $\mathbf{x} = \mathbf{\Psi}\boldsymbol{\theta}$. Let $1 \le p, q \le \infty$ and $1/p + 1/q = 1$.

- If the rows $\mathbf{e}_\ell^T \in \mathbb{C}^{1\times n}$ of $\mathbf{E}$ are bounded as $\|\mathbf{e}_\ell\|_p \le \beta$, then

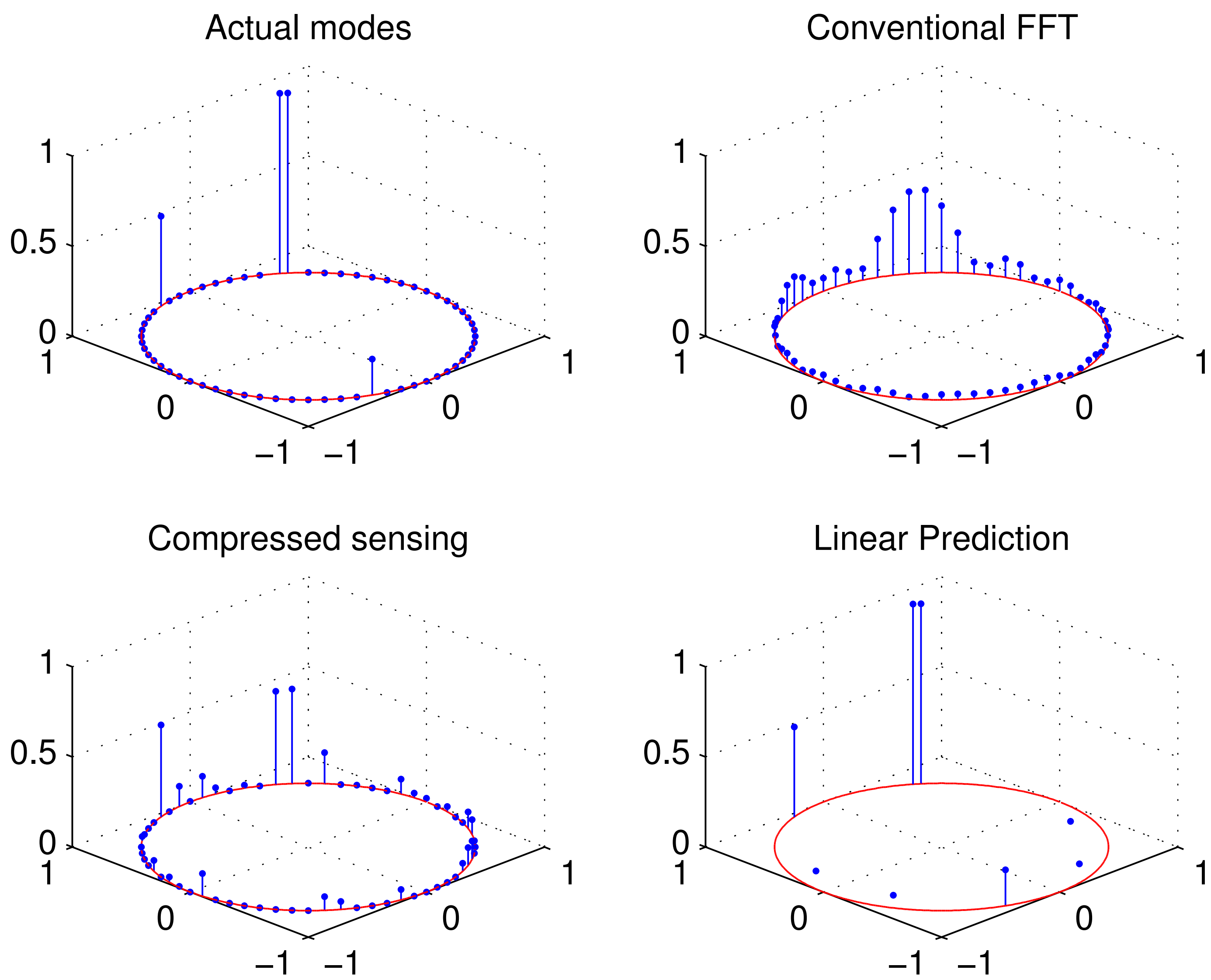
$$\|\mathbf{x} - \mathbf{x}_k\|_1 \le \|\boldsymbol{\theta} - \boldsymbol{\theta}_k\|_1 + (n-k)\beta\|\boldsymbol{\theta}\|_q.$$

  The bound is achieved when the entries of $\mathbf{E}$ satisfy

$$e_{mn} = \pm\beta \cdot e^{j(\arg(\theta_m) - \arg(\theta_n))} \cdot (|\theta_n|/\|\boldsymbol{\theta}\|_q)^{q/p}.$$

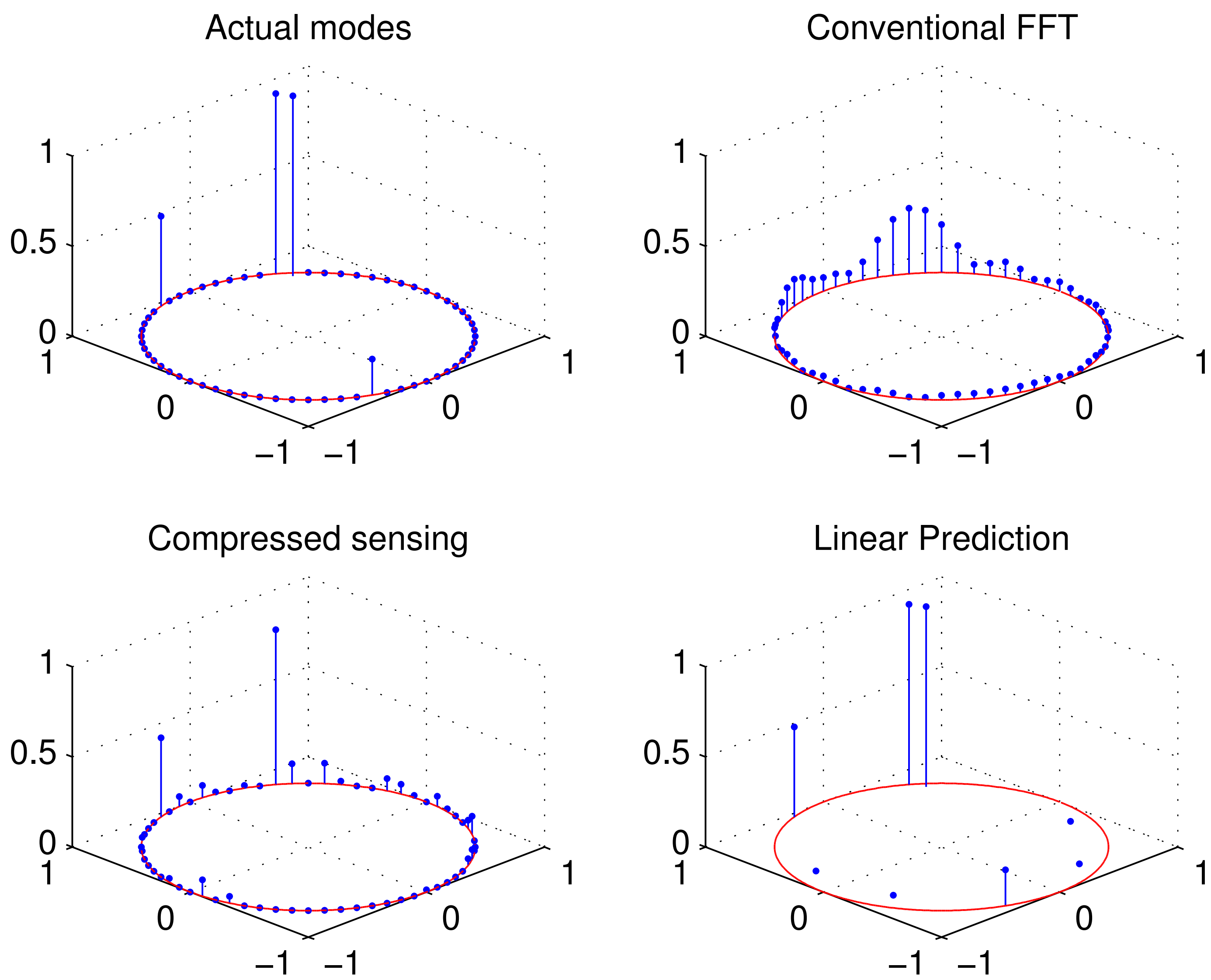
Message: In the presence of basis mismatch, *exact or near-exact sparse recovery cannot be guaranteed*. Recovery *may* suffer large errors.

Mismatch in Modal Analysis



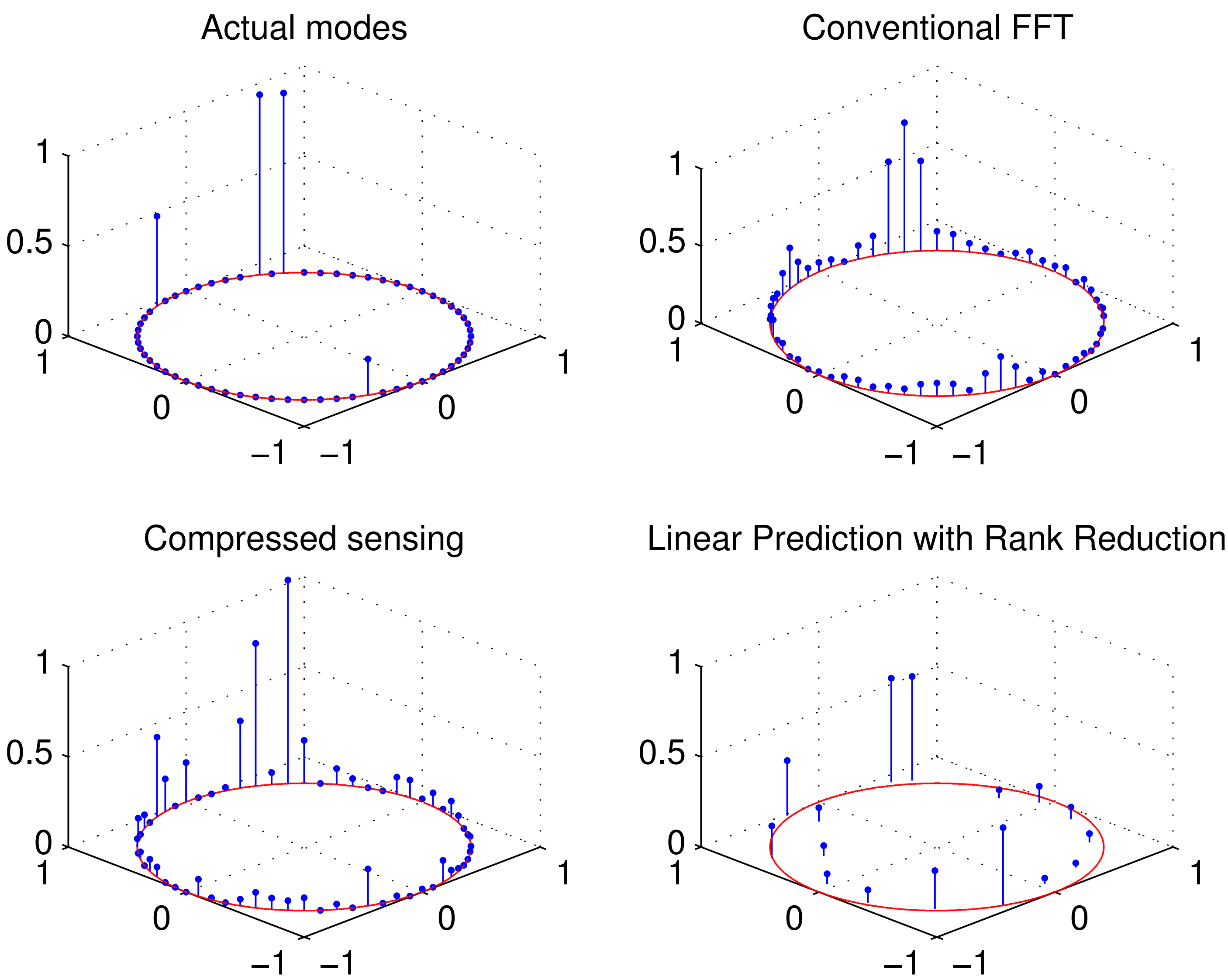
Frequency mismatch

Mismatch in Modal Analysis



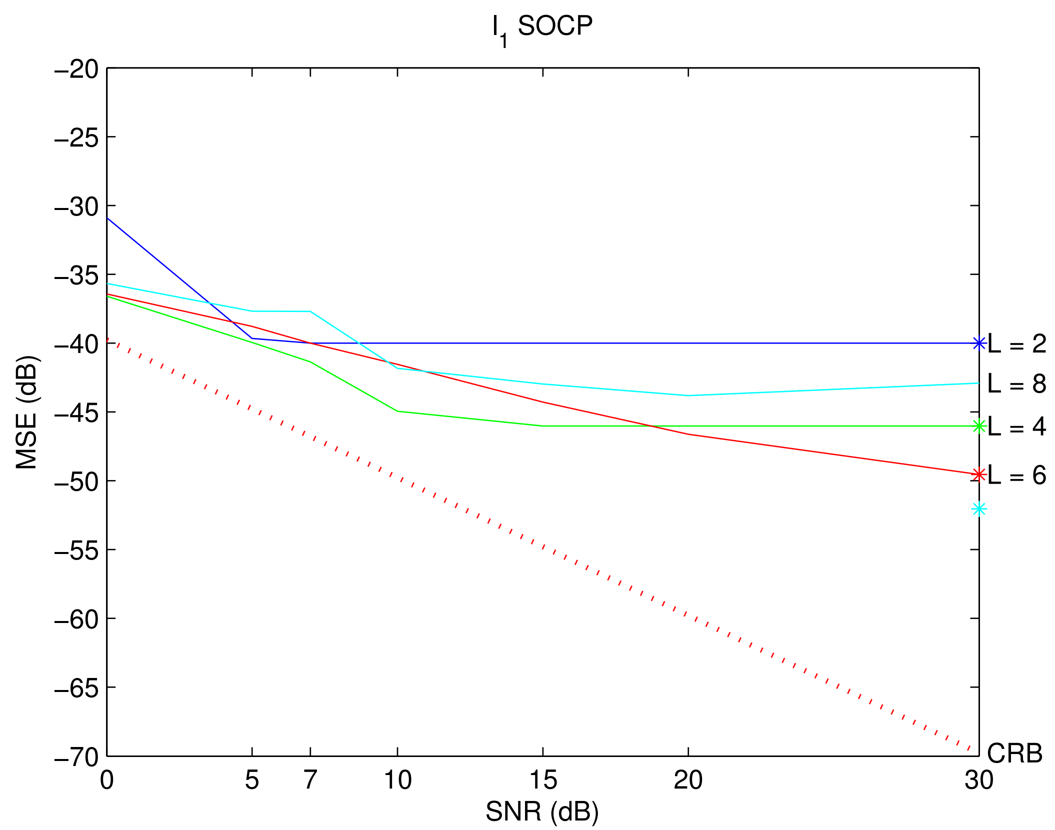
Damping mismatch

Mismatch in Modal Analysis



Frequency mismatch–noisy measurements

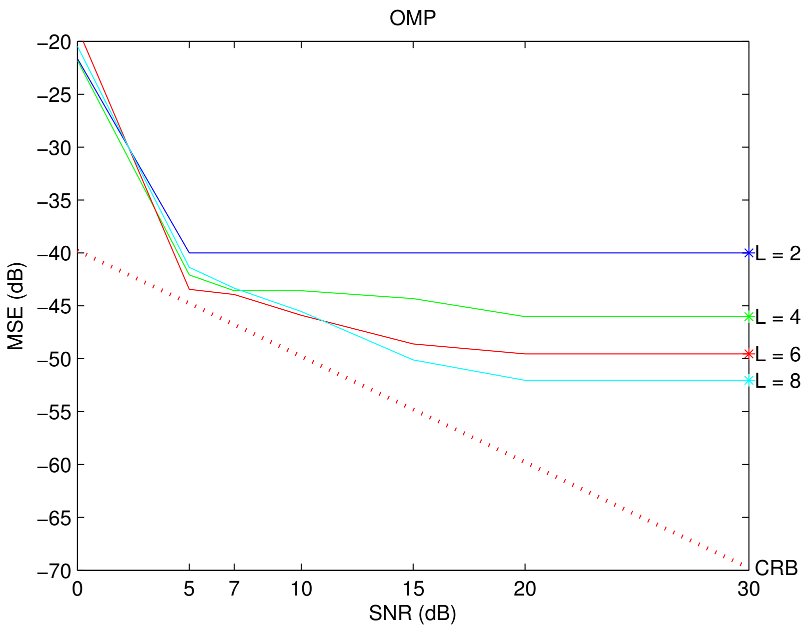
Noise Limited, Quantization Limited, or Null Space Limited



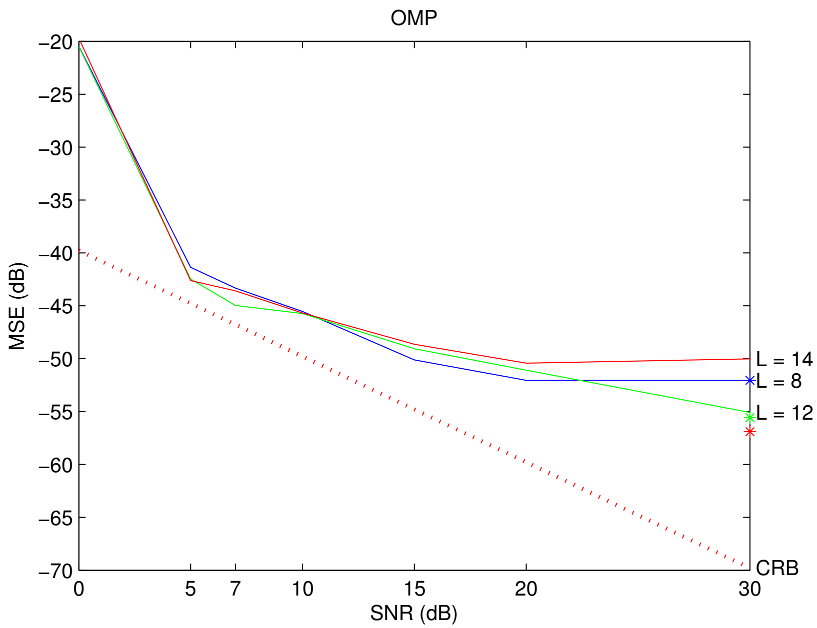
ℓ_1 inversions for $L = 2, 4, 6, 8$

- $f_1 = 0.5$ Hz, $f_2 = 0.52$ Hz, $m = 25$ samples, complex Gaussian noise of variance σ^2. DFT frame $\mathbf{\Psi} \in \mathbb{C}^{n \times mL}$ with half-cell width $(1/2nL)$.

Noise Limited, Quantization Limited, or Null Space Limited

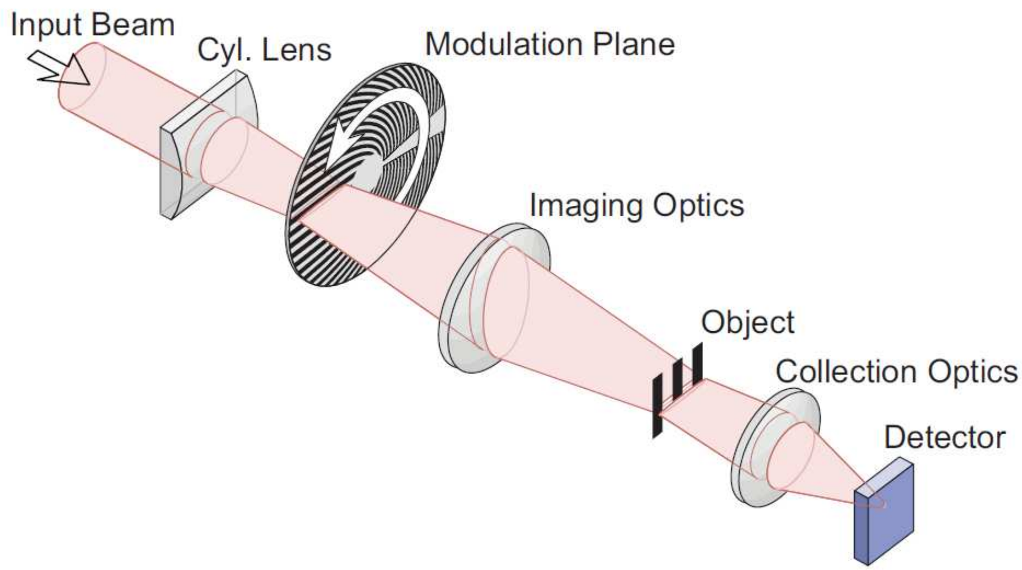


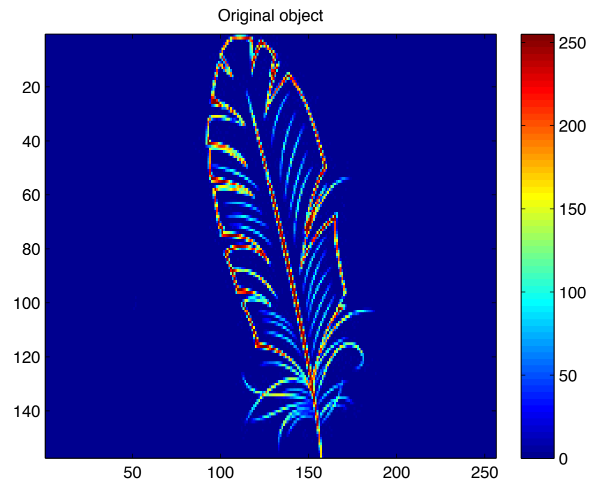
(a) OMP for $L = 2, 4, 6, 8$

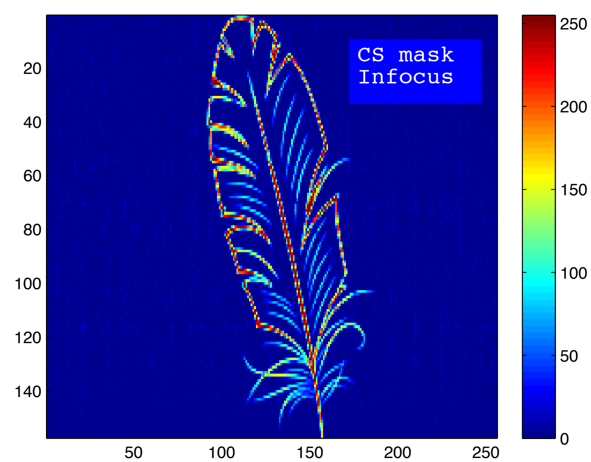


(b) OMP for $L = 8, 12, 14$

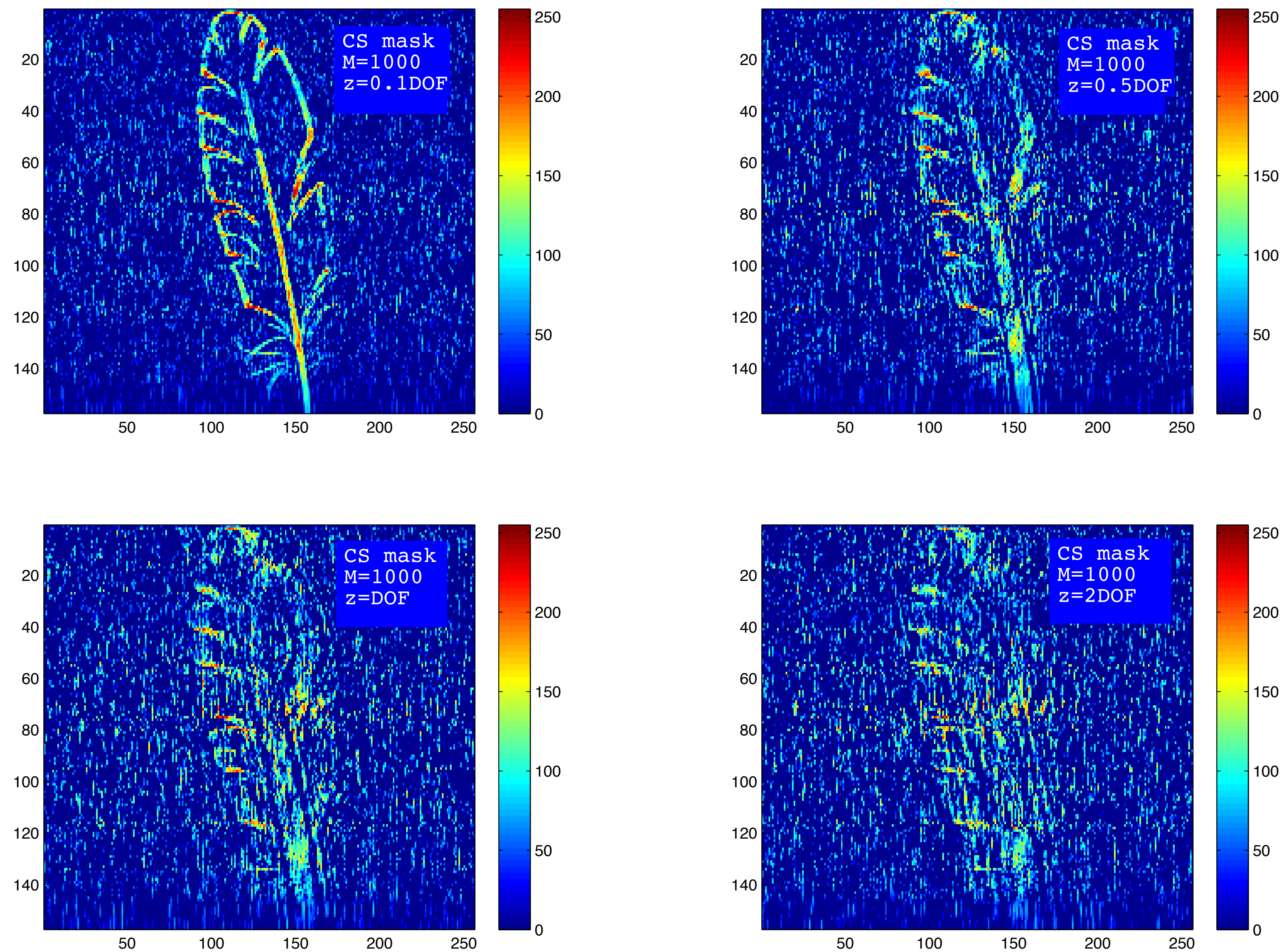
Misfocus in Optical Imaging







Misfocus in Optical Imaging



References on Model Mismatch in CS

- Y. Chi, A. Pezeshki, L. L. Scharf, and R. Calderbank, “Sensitivity to basis mismatch in compressed sensing,” *in Proc. 2010 IEEE Int. Conf. on Acoust., Speech and Signal Process.(ICASSP), Dallas, TX, Mar. 2010, pp. 3930 –3933.*

- Y. Chi, L.L. Scharf, A. Pezeshki, and A.R. Calderbank, “Sensitivity to basis mismatch in compressed sensing,” *IEEE Transactions on Signal Processing, vol. 59, no. 5, pp. 2182–2195, May 2011.*

- L. L. Scharf, E. K. P. Chong, A. Pezeshki, and J. R. Luo, “Sensitivity considerations in compressed sensing,” *in Conf. Rec. 45th Annual Asilomar Conf. Signals, Systs., Computs., Pacific Grove, CA,, Nov. 2011, pp. 744–748.*

Work by Others on Model Mismatch in CS

- M. A. Herman and T. Strohmer, "General deviants: An analysis of perturbations in compressed sensing," *IEEE J. Selected Topics in Signal Processing*, vol. 4, no. 2, pp. 342349, Apr. 2010.
- D. H Chae, P. Sadeghi, and R. A. Kennedy, "Effects of basis-mismatch in compressive sampling of continuous sinusoidal signals," *Proc. Int. Conf. on Future Computer and Commun.*, Wuhan, China, May 2010.
- H. Zhu, G. Leus, and G. B. Giannakis, "Sparsity-cognizant total least-squares for perturbed compressive sampling," *IEEE Transactions on Signal Processing*, vol. 59, May 2011 (to appear).
- Marco F. Duarte and Richard G. Baraniuk, "Spectral compressive sensing," *IEEE Trans. Signal Processing*, submitted Sep. 2010.

Compressed Sensing Off The Grid

Atomic Norm Decomposition

- Model:

$$\mathbf{y} = \sum_{i=1}^{k} \boldsymbol{\psi}_{\mathbf{k}} \theta_k; \quad \{\psi_k\} : \text{ Atoms}$$

- Atomic norm [Chandrasekaran, Recht, Parrilo, and Willsky (Allerton 2010)]:

$$\|\mathbf{y}\|_{\mathcal{A}} = \inf_{(\boldsymbol{\theta},\boldsymbol{\psi})} \sum_{i=1}^{k} |\theta_k|$$

- Atomic norm decomposition:

$$\min \|\boldsymbol{\eta}\|_{\mathcal{A}} \quad \text{s.t.} \quad \mathcal{P}_{\Omega}(\mathbf{y}) = \mathcal{P}_{\Omega}(\boldsymbol{\eta})$$

- Atomic set:

$$\mathcal{A} = \left\{ \begin{pmatrix} e^{j\phi} \\ e^{j(2\pi\nu+\phi)} \\ \vdots \\ e^{j(n-1)2\pi\nu+\phi)} \end{pmatrix} : \ \nu \in [0,1), \ \phi \in [0, 2\pi) \right\}$$

Atomic Norm Decomposition

Line spectra resolution:

- **Theorem:** [Candes and Fernandez-Granda 2012] A line spectrum with minimum frequency separation $\Delta_f > 4/k$ can be recovered from the first $2k$ Fourier coefficients via atomic norm minimization.
- **Theorem:** [Tang, Bhaskar, Shah, and Recht 2012] A line spectrum with minimum frequency separation $\Delta_f > 4/n$ can be recovered from most subsets of the first n Fourier coefficients of size at least $m = \mathcal{O}(k \log(k) log(n))$.
- **Theorem:** [Chi and Chen 2013] A 2D line spectrum with minimum frequency separation $\Delta_f > 4/(\sqrt{n_1 n_2})$ can be recovered from most subsets of the first n Fourier coefficients of size at least $m = \mathcal{O}(k \log(k) log(n))$.

Refernces on Off-Grid CS

- V. Chandrasekaran, B. Recht, P. Parrilo, and A. Willsky, "The convex algebraic geometry of linear inverse problems," in *Proc. 48th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, 2010, pp. 699-703.

- G. Tang, B. N. Bhaskar, P. Shah, and B. Recht, "Compressed sensing off the grid," July 2012. Available: http://arxiv.org/abs/1207.6053.

- E. Candes and C. Fernandez-Granda, Towards a mathematical theory of super-resolution, March 2012. Available: http://arxiv.org/abs/1203.5871.

- Y. Chi and Y. Chen, "Compressive Recovery of 2-D Off-Grid Frequencies ," *Conf. Rec. Asilomar Conference on Signals, Systems, and Computers*, Pacific Grove, CA, Nov. 2013.

Concluding Remarks

- Compression, whether by linear maps (eg, Gaussian or Bernoulli) or by subsampling (eg, co-prime) has performance consequences. The CR bound increases and the onset of breakdown threshold increases.