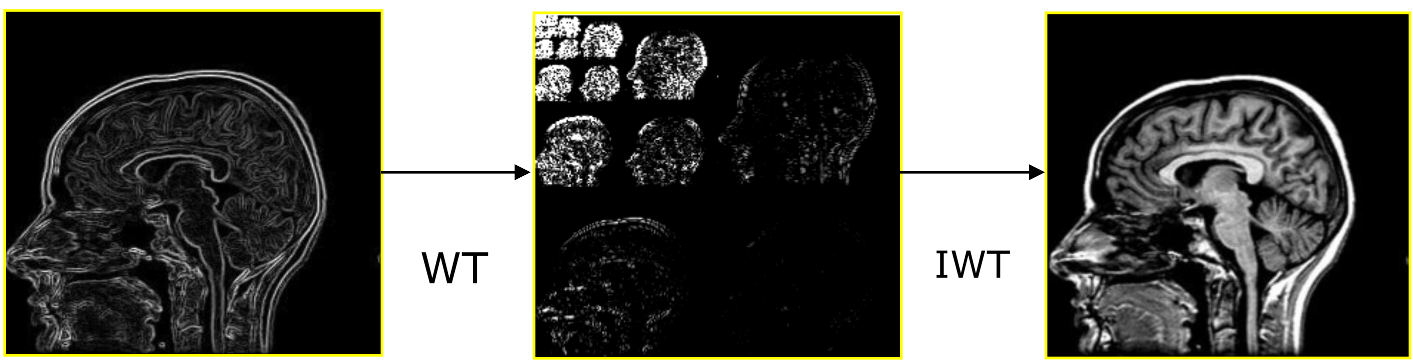
Multi-Task Compressive Sensing with Dirichlet Process Priors

Yuting Qi[1], Dehong Liu[1], David Dunson[2], and Lawrence Carin[1]

[1]Department of Electrical and Computer Engineering
[2]Department of Statistical Science
Duke University, USA

Overviews of CS - 1/6

- Nyquist Sampling
  - $f_{sampling} >= 2\, f_{max}$
  - In many applications, $f_{sampling}$ is **very** high.
  - Most digital signals are highly compressible, only encode few large coefficients and throw away most of them.

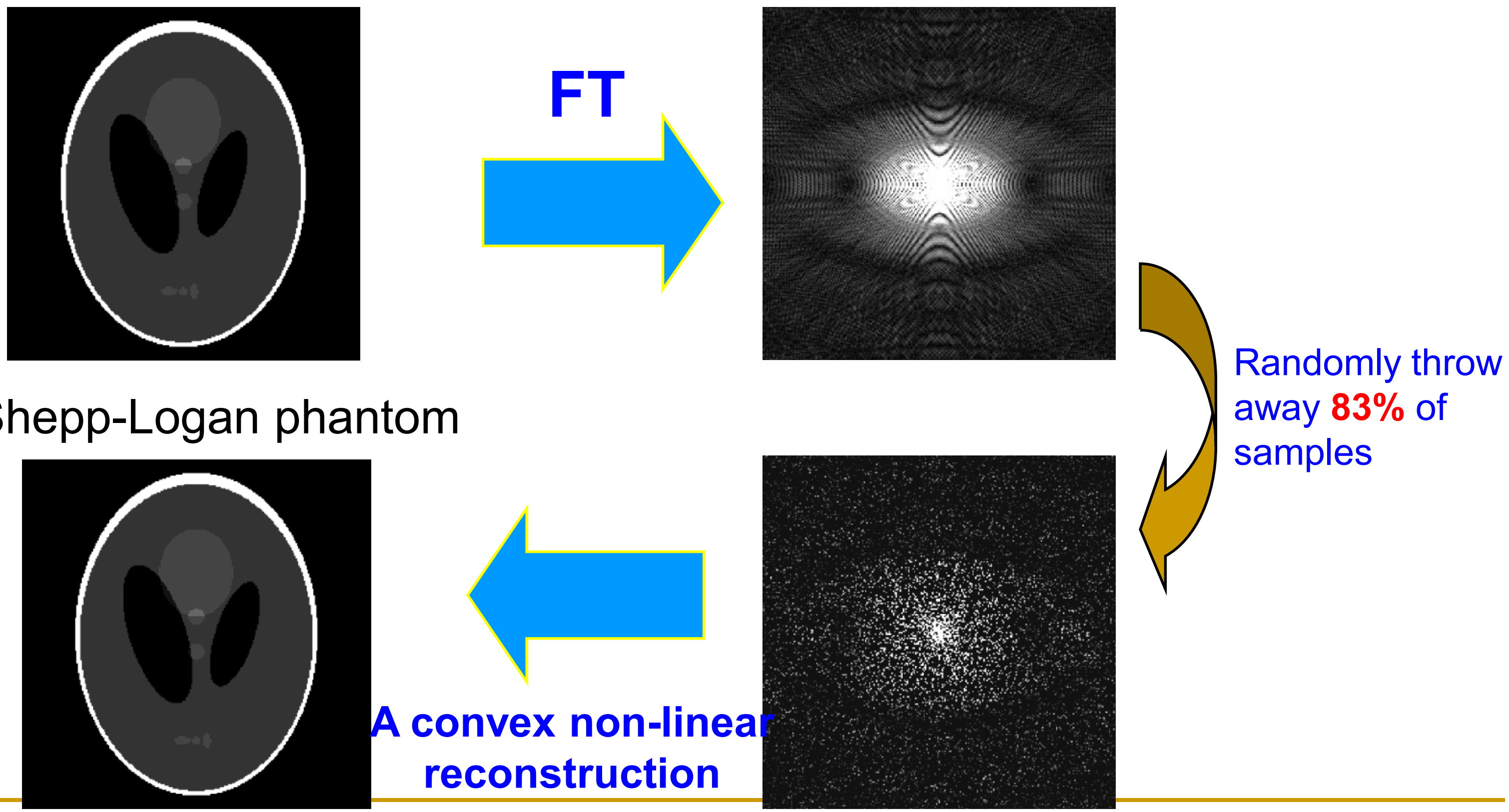


512*512 pixels

Wavelet coefficients

Overviews of CS - 2/6

- Why waste so many measurements if eventually most are discarded?
- A surprising experiment:



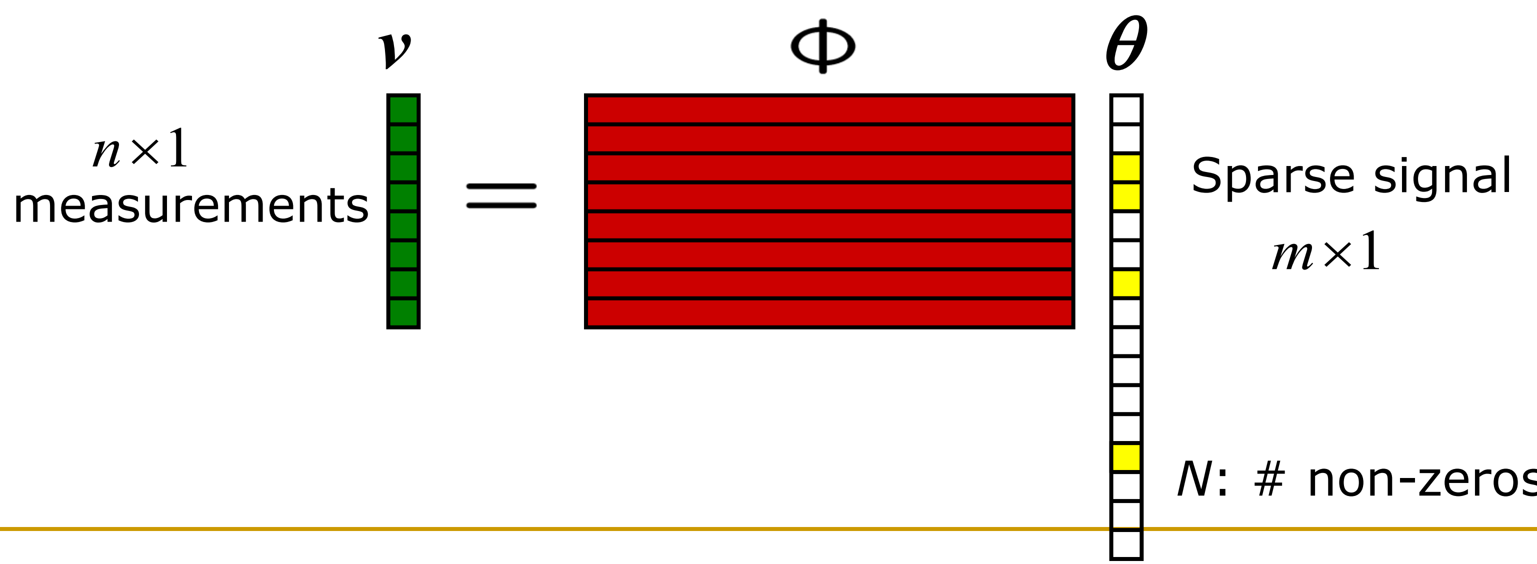
Shepp-Logan phantom

Overviews of CS - 3/6

- Basic Idea (*Donoho, Candes, Tao, et al*):
  - Assume an $m\times 1$ compressible signal $u=\boldsymbol{\Psi\theta}$, with $\boldsymbol{\Psi}$ an orthonormal basis and θ sparse coefficients.
  - In CS, we measure $\boldsymbol{v}$, a compact form of signal $\boldsymbol{u}$,

    $$\boldsymbol{v} = \mathbf{T}\boldsymbol{u} = \mathbf{T}\boldsymbol{\Psi\theta} = \boldsymbol{\Phi\theta}$$

    $\boldsymbol{T}$ is $n\times m$ with elements constituted randomly.



Overviews of CS - 4/6

- The theory of Candes et al. (*2006*):
  - With overwhelming probability, $\boldsymbol{\theta}$ (hence $\boldsymbol{u}$) is recovered with

    $$\boxed{\min ||\boldsymbol{\theta}||_{l_1}, \quad \text{s.t.}, \quad \boldsymbol{v} = \boldsymbol{\Phi\theta},}$$

    if the number of CS measurements

    $$\boxed{n > C \cdot N \cdot \log m}$$

    (C is a constant and N is number of non-zeros in $\boldsymbol{\theta}$)

  - If N is small, *i.e.*, $\boldsymbol{u}$ is highly compressible, $m << n$.
  - The problem may be solved by linear programming or greedy algorithms.

Overviews of CS - 5/6

- Bayesian CS (*Ji and Carin, 2007*)
  - Recall

$$\boxed{\min ||\boldsymbol{\theta}||_{l_1}, \quad \text{s.t.}, \quad \boldsymbol{v} = \boldsymbol{\Phi}\boldsymbol{\theta},}$$

  - Connection to linear regression

$$\boldsymbol{v} = \Phi\boldsymbol{\theta} + \mathbf{n}_e$$

  - BCS
    - Put sparse prior over $\boldsymbol{\theta}$,

$$\boldsymbol{\theta} \sim N(0, \boldsymbol{\alpha}^{-1}), \qquad \boldsymbol{\alpha} \sim Ga(c, d), \quad \mathbf{n}_e \sim N(0, \boldsymbol{\alpha}_0^{-1})$$

    - Given observation v, $p(\boldsymbol{\theta}|v)$?

Overviews of CS - 6/6

- Multi-Task CS
  - *M* CS tasks.
  - Reduce measurement number by exploiting relationships among tasks.
  - Existing methods assume all tasks fully shared.
  - In practice, not all signals are satisfied with this assumption.
  - Can we expect an algorithm *simultaneously* discovers sharing structure of all tasks and perform the CS inversion of the underlying signals within each group?

DP Multi-task CS - 1/4

- DP MT CS
  - M sets of CS measurements

$$\boldsymbol{v}_i = \boldsymbol{\Phi}_i \boldsymbol{\theta}_i + \boldsymbol{\epsilon}_i$$

v_i: CS measurements of i-th task
θ_i: underlying sparse signal of i-th task
Φ_i: random projection matrix of i-th task
ϵ_i: measurement error of i-th task

  - May be from different scenarios.
    - some are heart MRI, some are skeleton MRI.
  - What we want?
    - Share information among all sets of CS tasks when sharing is appropriate.
    - Reduce measurement number.

DP Multi-task CS - 2/4

- DP MT CS Formula

$$\boldsymbol{v}_i = \boldsymbol{\Phi}_i \boldsymbol{\theta}_i + \boldsymbol{\epsilon}_i$$

  - Put sparseness prior over sparse signal θ_i:

$$\theta_{i,j} \mid \alpha_{i,j} \sim N(0, \alpha_{i,j}^{-1})$$

  - Encourage sharing of α_i, variance of sparse prior, via DP prior:

$$\alpha_i \mid G \sim G$$
$$G \mid \gamma . G_0 \sim DP(\gamma, G_0)$$

  - If necessary, then sharing; otherwise, no.
  - Estimate signals and learn sharing structure automatically & simultaneously.

DP Multi-Task CS - 3/4

- Choice of G_0
  - $G_0 = \prod_{j=1}^{m} Ga(c,d)$
  - Sparseness promoting prior:
    - $p(\boldsymbol{\theta}_i|c,d) = \prod_{j=1}^{m} \int \mathcal{N}(\theta_{i,j}|0, {\alpha^*_{k,j}}^{-1}) Ga(\alpha^*_{k,j}|c,d) d\alpha^*_{k,j}$
    - Automatic relevance determination (ARD) prior which enforces the sparsity over parameters.
    - If c=d, this becomes a student-t distribution t(0,1).

DP Multi-Task CS - 4/4

- Mathematical representation:

$$\boldsymbol{v}_i|\boldsymbol{\theta}_i, \alpha_0 \sim \mathcal{N}(\boldsymbol{\Phi}_i\boldsymbol{\theta}_i, \alpha_0^{-1}I), \quad i = 1, \cdots, M,$$

$$\theta_{i,j}|z_i, \{\boldsymbol{\alpha}_k^*\}_{k=1,K} \sim \mathcal{N}(0, {\alpha^*_{z_i,j}}^{-1}), \quad j = 1, \cdots, m, \quad i = 1, \cdots, M,$$

Sparseness Prior

$$z_i|\{w_k\}_{k=1,K} \overset{iid}{\sim} Multinomial(\{w_k\}_{k=1,K}), \quad i = 1, \cdots, M,$$

Association

$$w_k = \pi_k \prod_{l=1}^{k-1}(1-\pi_l), \quad k = 1, \cdots, K,$$

$$\pi_k \overset{iid}{\sim} Beta(1, \lambda), \quad k = 1, \cdots, K,$$

$$\lambda|e, f \sim Ga(e, f),$$

$$\boldsymbol{\alpha}_k^*|c, d \overset{iid}{\sim} \prod_{j=1}^{m} Ga(c, d), \quad i = 1, \cdots, M,$$

$$\alpha_0 \sim Ga(a, b),$$

Stick-breaking components

Hyper priors over parameters

Inference

- Variational Bayesian Inference
  - Bayes rule:

$$p(\Phi|\mathbf{X},\Psi) = \frac{p(\mathbf{X}|\Phi)p(\Phi|\Psi)}{\int p(\mathbf{X}|\Phi)p(\Phi|\Psi)d\Phi}$$

Marginal likelihood = ?

  - Introduce $q(\Phi)$ to approximate $p(\Phi|\mathbf{X},\Psi)$;
  - Log marginal likelihood

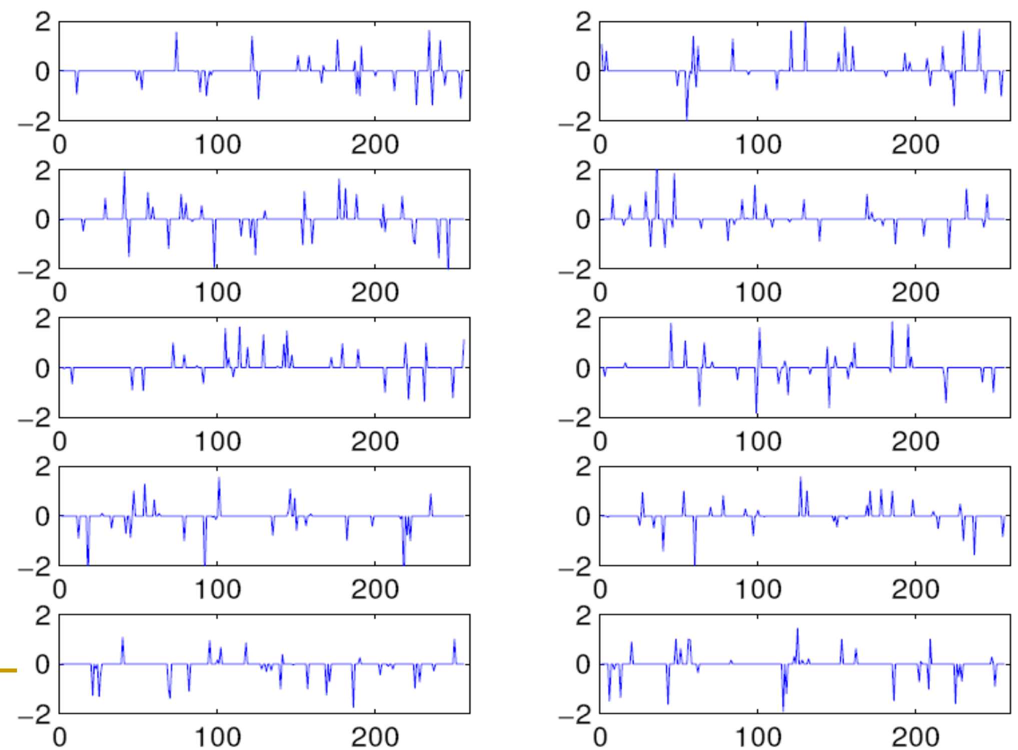
$$\log p(\mathbf{X}|\Psi) = \int q(\Phi)\log\frac{q(\Phi)}{p(\Phi|\mathbf{X},\Psi)}d\Phi + \int q(\Phi)\log\frac{p(\mathbf{X}|\Phi)p(\Phi|\Psi)}{q(\Phi)}d\Phi$$

$$\min\ \mathcal{D}_{KL}(q(\Phi)||p(\Phi|\mathbf{X},\Psi)) \Longleftrightarrow \max \mathcal{L}(q(\Phi))$$

  - $q(\Phi)$ is obtained by maximizing $\mathcal{L}(q(\Phi))$, which is computationally tractable.

Experiments - 1/6

- Synthetic data
  - Data are generated from 10 underlying clusters.
  - Each cluster is generated from one signal template.
  - Each template has a length of 256, with 30 spikes drawn from N(0,1); locations are random too.
  - Correlation of any two templates is zero.
  - From each template, we generate 5 signals: random move 3 spikes.
  - Total 50 sparse signals.



Experiments - 2/6

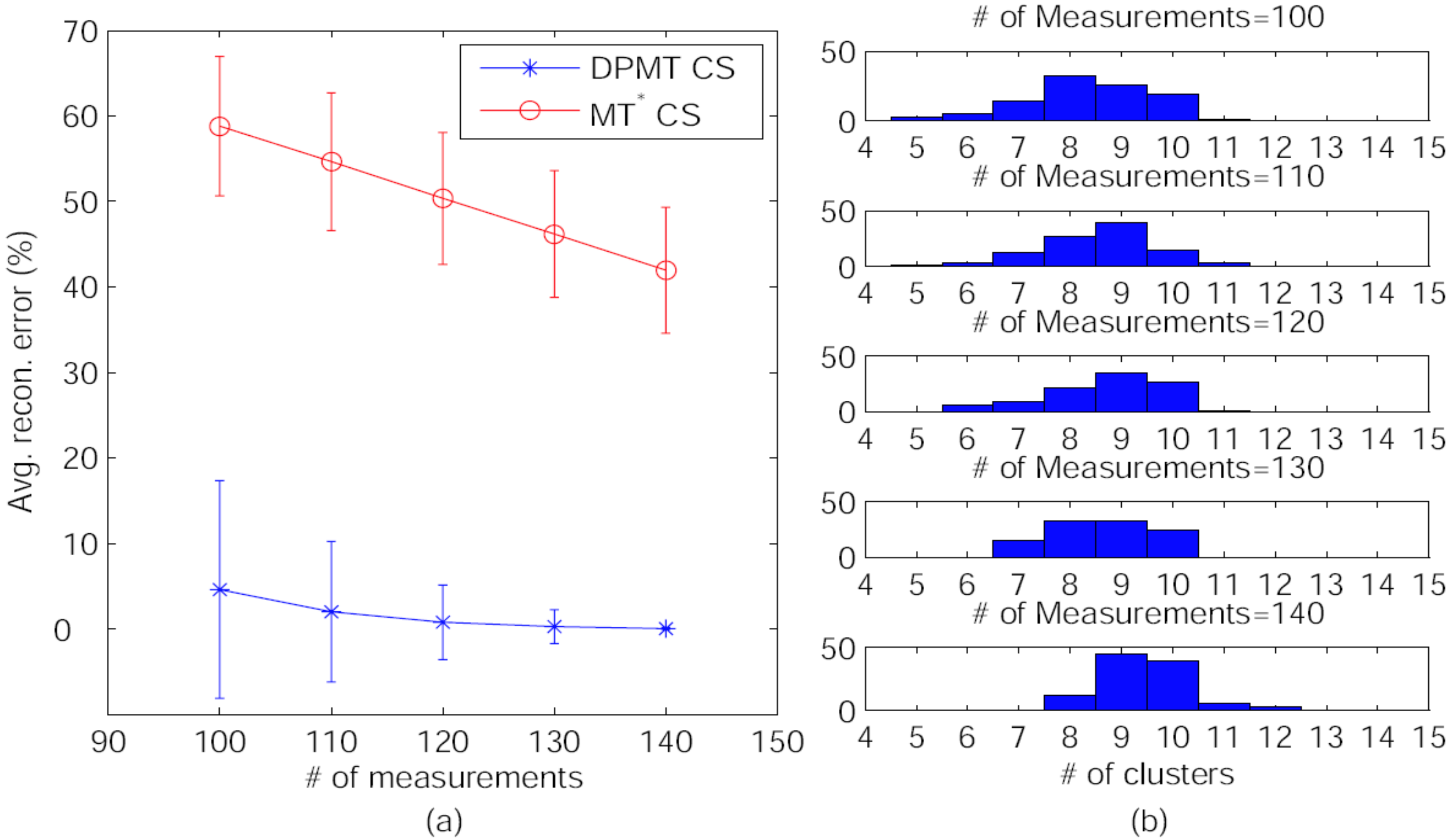


Figure.2 (a) Reconstruction error: DPMT CS and fully sharing MT CS (100 runs). (b) Histogram of number of clusters inferred from DPMT CS (100 runs).

Experiments - 3/6

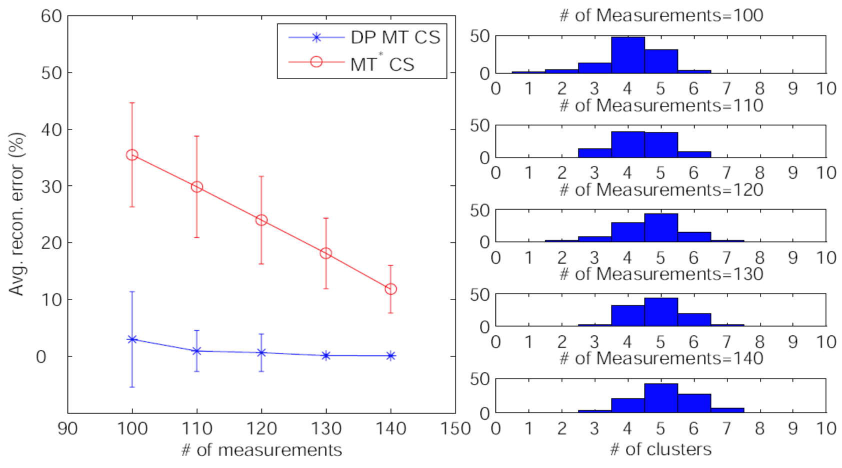


Figure.3 Five underlying clusters

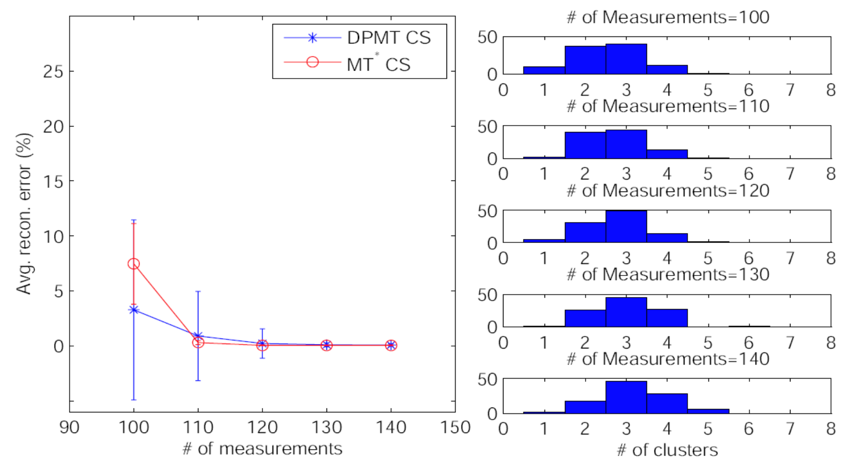


Figure.4 Three underlying clusters

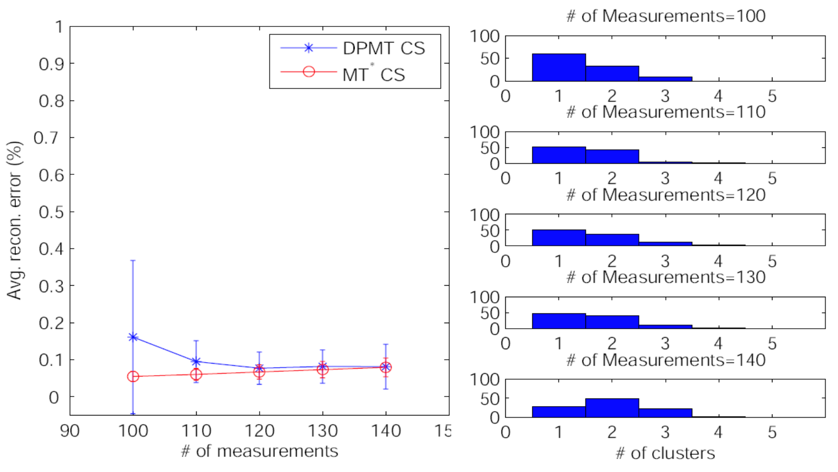


Figure.5 Two underlying clusters

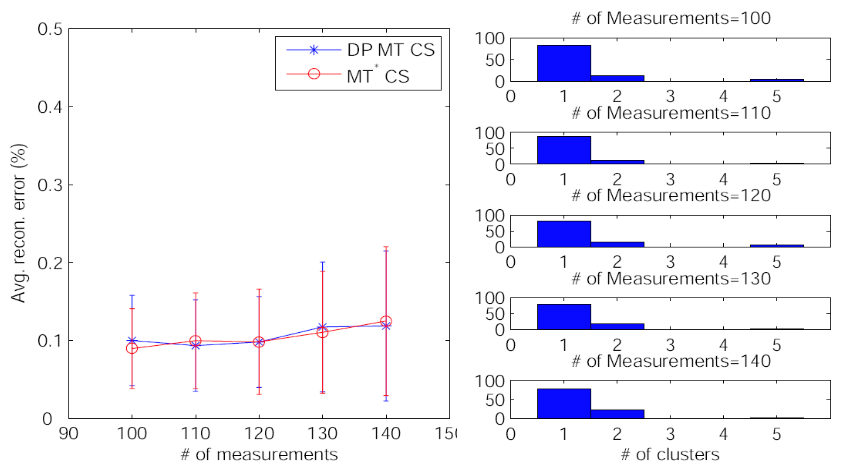
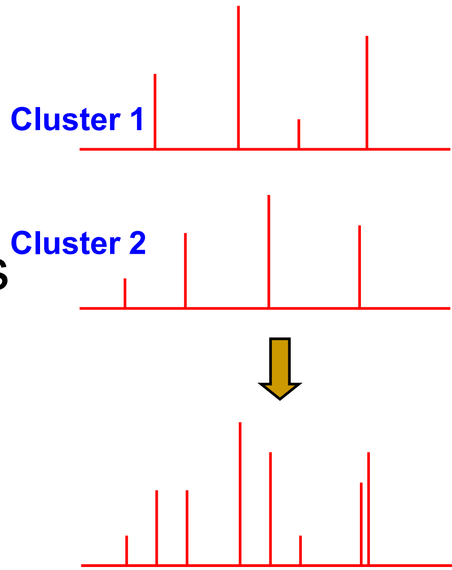


Figure.6 One underlying cluster

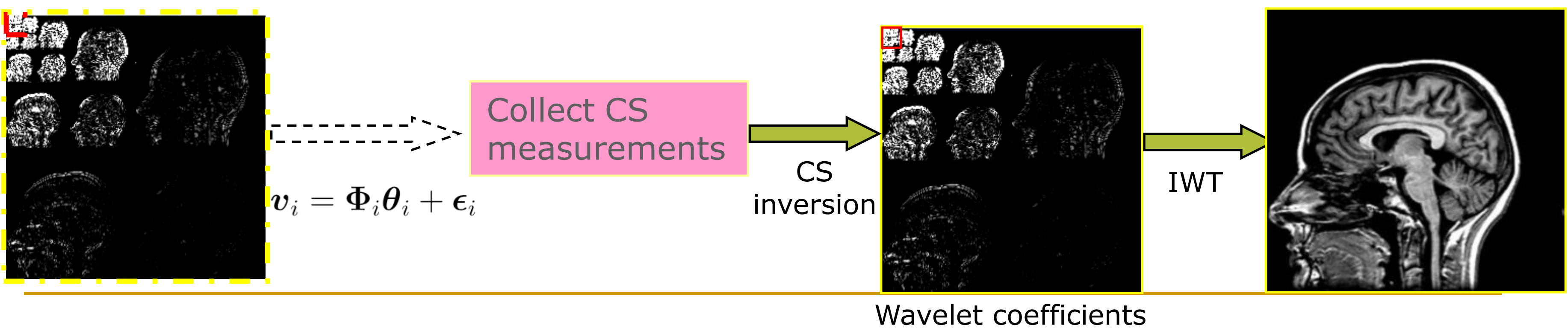
Experiments - 4/6

- Interesting observations:
  - As # of underlying clusters decreases, the difference of DP-MT and global-sharing MT CS decreases.
  - Sparseness sharing means sharing non-zero components **AND** zero components
  - Each cluster has distinct non-zero components, BUT they share large amount of zero components.
  - One global sparseness prior is enough to describe two clusters by treating them as one cluster.
  - However, for ten-cluster case, ten templates do not cumulatively share the same set of zero-amplitude coefficients; So global sparseness prior is inappropriate.



Experiments - 5/6

- Real image
  - 12 images of 256 by 256, Sparse in wavelet domain.
  - Image reconstruction:
    - Collect CS measurements (random projection of θ); estimate θ via CS inversion; reconstruct image by inverse wavelet transformation.
  - Hybrid scheme:
    - Assume finest wavelet coefficients zero, only estimate other 4096 coefficients (θ);
    - Assume all coarsest coefficients are measured;
    - CS measurements are performed on coefficients other than finest and coarsest ones.



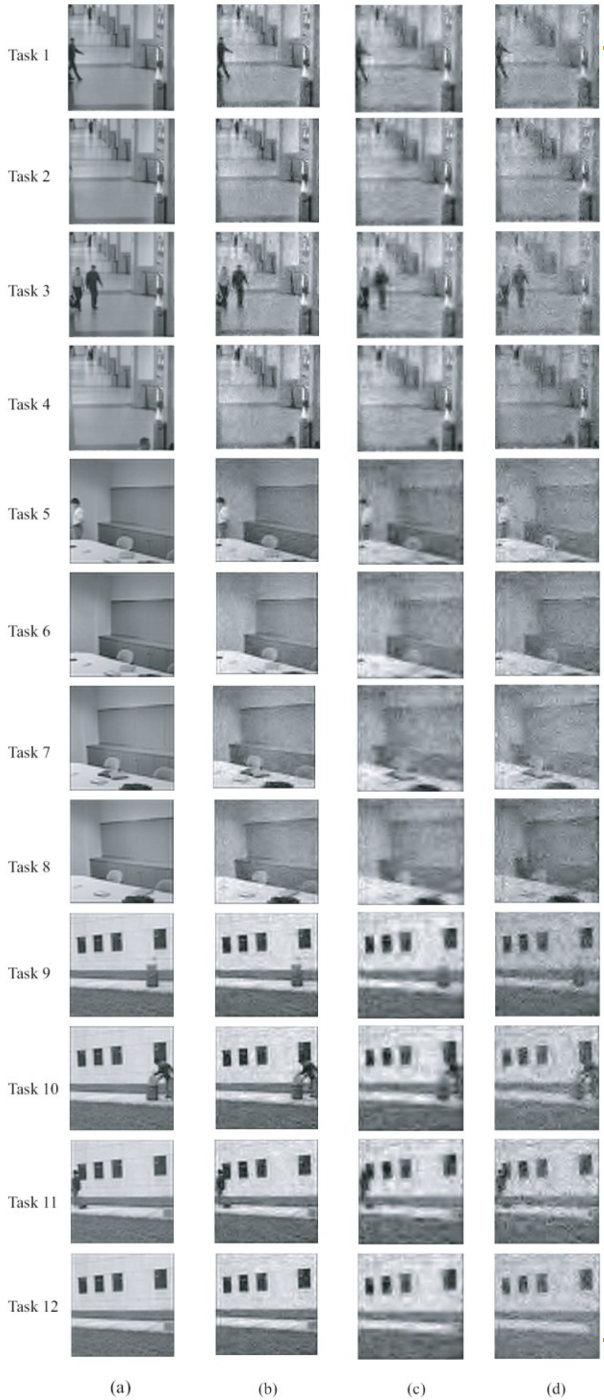


Figure 8. CS recon., (a) Linear, (b) DP-MT, (c) MT*, (d) ST

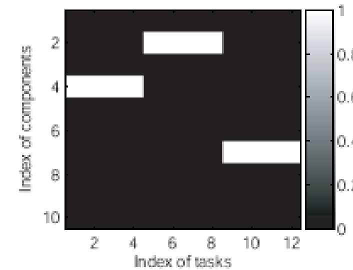


Figure 9. Sharing mechanism for 12 tasks in Figure 8 yielded by DP-MT CS.

Table 1 Reconstruction Error

	Task 1	Task 2	Task 3	Task 4	Task 5	Task 6
DP-MT	8.79	7.89	9.69	8.04	14.33	13.22
MT*	10.19	9.14	11.49	9.18	16.94	15.59
ST	10.28	10.37	12.81	10.28	18.37	16.18
Linear	6.66	6.20	7.08	6.14	12.41	11.70

	Task 7	Task 8	Task 9	Task 10	Task 11	Task 12
	15.18	14.54	15.51	16.71	16.11	15.19
	17.46	16.50	18.62	19.82	19.34	18.03
	18.65	17.67	20.77	22.24	21.19	19.59
	12.43	11.99	13.83	14.41	14.10	13.53

Conclusions

- A DP-based multi-task compressive sensing framework is developed for jointly performing multiple CS inversion tasks.
- This new method can simultaneously discover sharing structure of all tasks and perform the CS inversion of the underlying signals within each group.
- A variational Bayesian inference is developed for computational efficiency.
- Both synthetic and real image data show MT CS works at least as well as ST CS and outperforms full-sharing MT CS when the assumption of full-sharing is not true.

Thanks for your attention!

Questions?