Dynamic MRI image reconstruction using adaptive regularization methods

Jeffrey A. Fessler
William L. Root Professor of EECS

EECS Dept., BME Dept., Dept. of Radiology
University of Michigan

http://web.eecs.umich.edu/~fessler

With Sai Ravishankar, Brian Moore, & Raj Nadakuditi

Neurospin Seminar

2017-06-01

Outline

Ill-posed problems and regularization

- Classical "hand crafted" regularizers
- Data-driven (adaptive / learned) regularizers
- Data-driven regularized MRI via dictionary learning
- Extension: learning low-rank atoms
  - DIctioNary with lOw-ranK ATOms (DINO-KAT)

Dynamic MR imaging

- DINO-KAT for dynamic MR

Summary

Backup

- DINO-KAT convergence guarantees
- Online method

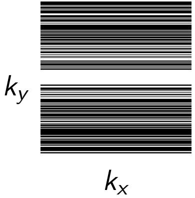
Ill-posed inverse problems

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{\varepsilon}$$

$\boldsymbol{y}$: measurements　　$\boldsymbol{\varepsilon}$: noise

$\boldsymbol{x}$: unknown image　　$\boldsymbol{A}$: system matrix (typically wide)

- compressed sensing (*e.g.*, MRI)　　($\boldsymbol{A}$ "random" rows of DFT)



- deblurring (restoration)　　($\boldsymbol{A}$ Toeplitz)
- in-painting　　($\boldsymbol{A}$ subset of rows of $\boldsymbol{I}$)
- denoising (not ill posed)　　($\boldsymbol{A} = \boldsymbol{I}$)

Inverse problems via MAP estimation

Unknown image $\boldsymbol{x}$ $\rightarrow$ [System model $\mathrm{p}(\boldsymbol{y} \,|\, \boldsymbol{x})$] $\rightarrow$ Data $\boldsymbol{y}$ $\rightarrow$ [Estimator] $\rightarrow$ Recon. image $\hat{\boldsymbol{x}}$

If we have a prior $\mathrm{p}(\boldsymbol{x})$, then the MAP estimate is:

$$\hat{\boldsymbol{x}} = \arg\max_{\boldsymbol{x}} \mathrm{p}(\boldsymbol{x} \,|\, \boldsymbol{y}) = \arg\max_{\boldsymbol{x}} \log \mathrm{p}(\boldsymbol{y} \,|\, \boldsymbol{x}) + \log \mathrm{p}(\boldsymbol{x}) \,.$$

For gaussian measurement errors and a linear forward model:

$$-\log \mathrm{p}(\boldsymbol{y} \,|\, \boldsymbol{x}) \equiv \frac{1}{2} \|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\|_{\boldsymbol{W}}^2$$

where $\|\boldsymbol{y}\|_{\boldsymbol{W}}^2 = \boldsymbol{y}' \boldsymbol{W} \boldsymbol{y}$ and $\boldsymbol{W}^{-1} = \mathrm{Cov}\{\boldsymbol{y} \,|\, \boldsymbol{x}\}$ is known

($\boldsymbol{A}$ from physics, $\boldsymbol{W}$ from statistics)

Priors for MAP estimation

- If all images $\boldsymbol{x}$ are "plausible" (have non-zero probability) then

$$\mathrm{p}(\boldsymbol{x}) \propto e^{-\mathsf{R}(\boldsymbol{x})} \Longrightarrow -\log \mathrm{p}(\boldsymbol{x}) \equiv \mathsf{R}(\boldsymbol{x})$$

(from fantasy / imagination / wishful thinking / data)

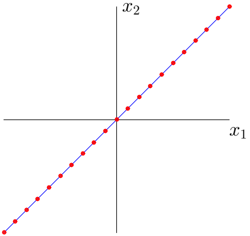
- MAP $\equiv$ regularized weighted least-squares (WLS) estimation:

$$\begin{aligned}\hat{\boldsymbol{x}} &= \arg\max_{\boldsymbol{x}} \ \log \mathrm{p}(\boldsymbol{y} \mid \boldsymbol{x}) + \log \mathrm{p}(\boldsymbol{x}) \\ &= \arg\min_{\boldsymbol{x}} \ \frac{1}{2} \|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\|_{\boldsymbol{W}}^2 + \mathsf{R}(\boldsymbol{x})\end{aligned}$$

- A regularizer $\mathsf{R}(\boldsymbol{x})$, aka log prior, is essential for high-quality solutions to ill-conditioned / ill-posed inverse problems.
- Why ill-posed? Often high ambitions...

Subspace model: Alternative to regularization

Assuming $\boldsymbol{x}$ lies in a sufficiently low-dimensional subspace could make an inverse problem well conditioned.

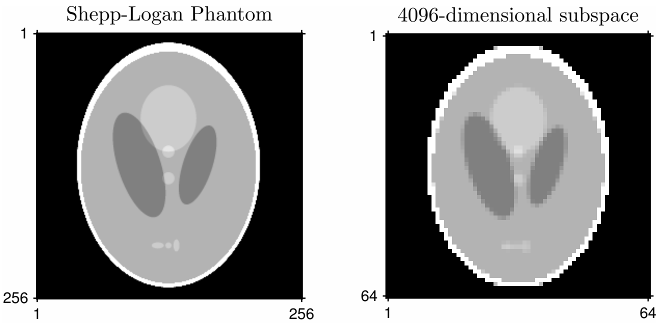


Assume $\boldsymbol{x} = \boldsymbol{D}\boldsymbol{z}$ where $\boldsymbol{D} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ and $\boldsymbol{z} \in \mathbb{R}^1$

($\boldsymbol{z}$ has only one nonzero element so very sparse!?)
Estimate coefficient(s): $\hat{\boldsymbol{z}} = \arg\min_{\boldsymbol{z}} \|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{D}\boldsymbol{z}\|_2^2$, then $\hat{\boldsymbol{x}} = \boldsymbol{D}\hat{\boldsymbol{z}}$,
where usually $\text{cond}(\boldsymbol{D}'\boldsymbol{A}'\boldsymbol{A}\boldsymbol{D}) \ll \text{cond}(\boldsymbol{A}'\boldsymbol{A})$.

Why not use subspace models?

Candès and Romberg (2005) [1] used 22 (noiseless) CT projection views (*i.e.*, 22 pseudo-radial lines in MRI), each with 256 samples.
$\Longrightarrow 22 \cdot 256 = 5632$ measured values,
vs $256^2 = 65536$ unknown pixels



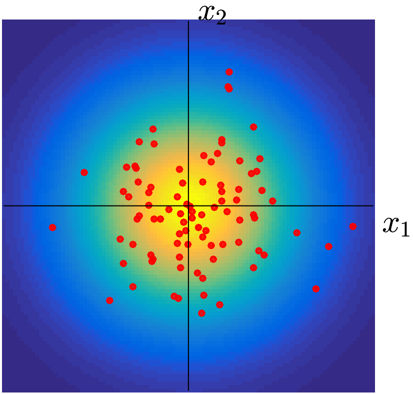
Subspace representation (using pixel basis) is undesirably coarse.

Classical regularizers (“hand crafted”)

- Tikhonov regularization (IID gaussian prior)
- Roughness penalty (Basic MRF prior)
- Sparsity in ambient space
- Edge-preserving regularization
- Total-variation (TV) regularization
- Black-box denoiser like NLM

Tikhonov regularization

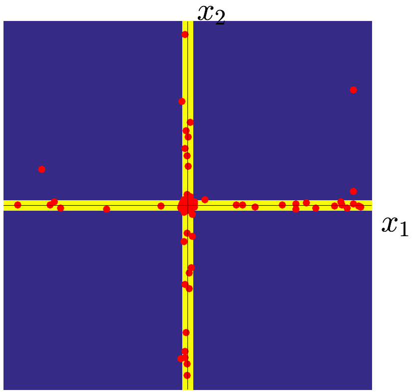
$\mathsf{R}(\boldsymbol{x}) = \beta \, \|\boldsymbol{x}\|_2^2$



- Colors show equivalent (normalized) prior $\mathsf{p}(\boldsymbol{x}) \,/\, \mathsf{p}(\mathbf{0}) = \mathrm{e}^{-\mathsf{R}(\boldsymbol{x})}$
- Equivalent to IID gaussian prior on $\boldsymbol{x}$
- Makes any ill-conditioned / ill-posed problem well conditioned
- Ignores correlations between pixels

Sparsity regularization in ambient space

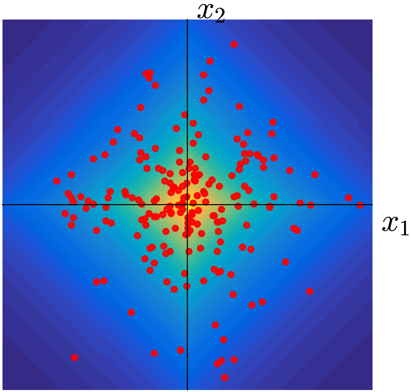
$$R(\boldsymbol{x}) = \beta \|\boldsymbol{x}\|_0 = \beta \sum_j \mathbb{I}_{\{x_j \neq 0\}}$$



- Approximate Bayesian interpretation
- Non-convex
- IID $\Longrightarrow$ also ignores correlations

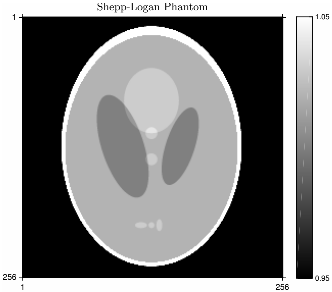
Sparsity regularization: convex relaxation

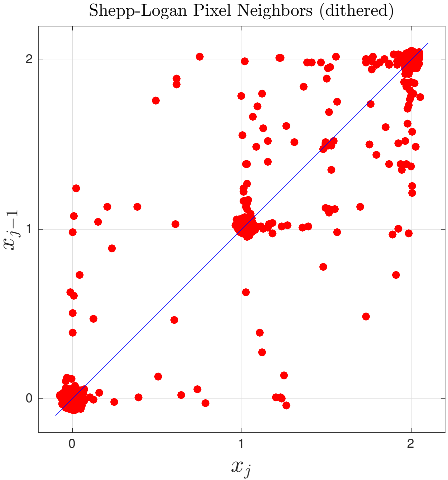
$$\mathsf{R}(\boldsymbol{x}) = \beta \|\boldsymbol{x}\|_1 = \beta \sum_j |x_j|$$



- Equivalent to IID Laplacian prior on $\boldsymbol{x}$
- Also ignores correlations

Correlation





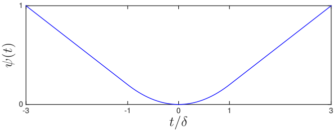
Caution: Shepp-Logan phantom [2] was designed for testing non-Bayesian methods, not for designing signal models. Q: What causes the spread??

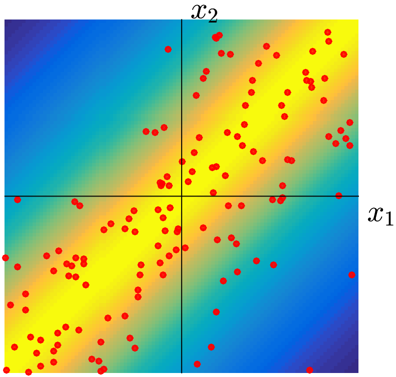
Edge-preserving regularization

Neighboring pixels tend to have similar values except near edges:

$$\mathsf{R}(\boldsymbol{x}) = \beta \sum_j \psi(x_j - x_{j-1})$$

Potential function ψ:



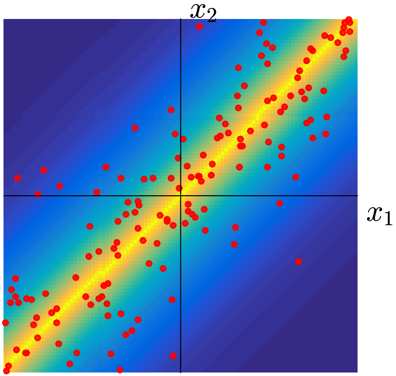


- Equivalent to improper prior (agnostic to DC value)
- Accounts for spatial correlations, but only very locally
- Used clinically now for low-dose X-ray CT image reconstruction

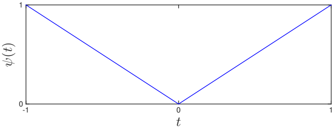
Total-variation (TV) regularization

Neighboring pixels tend to have similar values except near edges ("gradient sparsity"):

$$\mathsf{R}(\boldsymbol{x}) = \beta \operatorname{TV}(\boldsymbol{x}) = \beta \left\| \boldsymbol{C x} \right\|_1$$
$$= \beta \sum_j \left| x_j - x_{j-1} \right|$$



Potential function ψ:



- Equivalent to improper prior (agnostic to DC value)
- Accounts for correlations, but only very locally
- Well-suited to piece-wise constant Shepp-Logan phantom!
- Used in many academic publications...

Black-box denoiser as a regularizer

$$\text{Noisy image} \rightarrow \boxed{\text{Denoiser}} \rightarrow \text{Denoised image}$$

- Example: Non-local means (NLM)
- Corresponding regularizer [3–5]:

$$\mathsf{R}(\boldsymbol{x}) = \beta \frac{1}{2} \|\boldsymbol{x} - \text{NLM}(\boldsymbol{x})\|_2^2$$

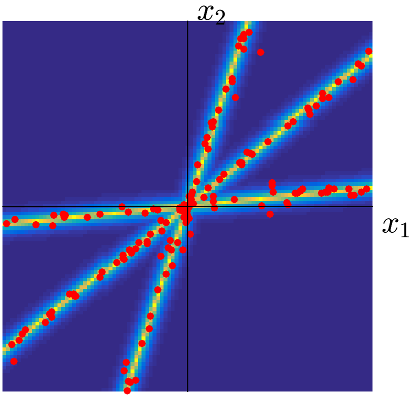
- Encourages self-consistency with denoised version of image
- No evident Bayesian interpretation
- Variable splitting can facilitate minimization [6].

Many more regularizers / priors

- Transforms: wavelets, curvelets, ...
- Markov random field models
- Graphical models
- ...

Outline

Union of subspaces model



- Dimensionality reduction?
- *cf.* classification / clustering motivation [7]
- (Extension to union of "flats" (linear varieties) is possible [8].)

Union of subspaces regularization

Given (?) collection of K subspace bases $\boldsymbol{D}_1, \ldots, \boldsymbol{D}_K$ (dictionaries with full column rank $\Longrightarrow$ tall).
Assume $\boldsymbol{x} \approx \boldsymbol{D}_k \boldsymbol{z}_k$ for some k and some (non-sparse) coefficients $\boldsymbol{z}_k$.
Natural regularizer for this model is:

$$\mathsf{R}(\boldsymbol{x}) = \underbrace{\min_k}_{\text{“classification”}} \underbrace{\min_{\boldsymbol{z}_k} \beta \frac{1}{2} \|\boldsymbol{x} - \boldsymbol{D}_k \boldsymbol{z}_k\|_2^2}_{\text{regression}}$$

$$= \min_k \beta \frac{1}{2} \left\| \boldsymbol{x} - \boldsymbol{D}_k \boldsymbol{D}_k^+ \boldsymbol{x} \right\|_2^2 .$$

- $\mathsf{R}(\boldsymbol{x}) = 0$ if $\boldsymbol{x}$ lies in the span of any of the dictionaries $\{\boldsymbol{D}_k\}$.
- Otherwise, distance to nearest subspace (discourage, not constrain).
- Non-convex (highly?) (*cf.* preceding picture) due to min
- Apply to image patches to be practical.
- Equivalent Bayesian interpretation? (Not a mixture model here.)
- Given? Learned from training data.

Regularizers using sparse coding with dictionaries

Assume $\boldsymbol{x} \approx \boldsymbol{D}\boldsymbol{z}$ where

- $\boldsymbol{D}$ is a dictionary (often over-complete $\Longrightarrow$ wide)
- $\boldsymbol{z}$ is a *sparse* coefficient vector (subset of columns of $\boldsymbol{D}$).

Corresponding regularizers:

$$\mathsf{R}(\boldsymbol{x}) = \min_{\boldsymbol{z} \,:\, \|\boldsymbol{z}\|_p \leq s} \beta \frac{1}{2} \|\boldsymbol{x} - \boldsymbol{D}\boldsymbol{z}\|_2^2, \qquad \text{or:}$$

$$\mathsf{R}(\boldsymbol{x}) = \min_{\boldsymbol{z}} \left(\beta_1 \frac{1}{2} \|\boldsymbol{x} - \boldsymbol{D}\boldsymbol{z}\|_2^2 + \beta_2 \|\boldsymbol{z}\|_p \right).$$

- Convex in $\boldsymbol{z}$ (for given $\boldsymbol{x}$) if $p \geq 1$ and $\boldsymbol{D}$ given.
- $\mathsf{R}(\boldsymbol{x})$ typically non-convex in $\boldsymbol{x}$, due to $\|\cdot\|_p$.
- Could be equivalent to a union-of-subspaces regularizer if $\boldsymbol{D} = [\boldsymbol{D}_1 \;\ldots\; \boldsymbol{D}_K]$ and if we constrain coefficient vector $\boldsymbol{z}$ in a non-standard way.

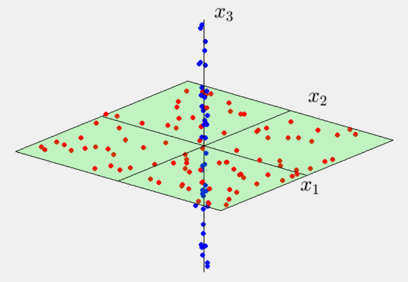
Union-of-subspaces vs sparse-coding-with-dictionary

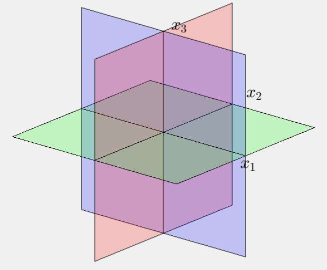
Consider union-of-subspaces model with $\boldsymbol{D}_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}$, $\boldsymbol{D}_2 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$.

So $\boldsymbol{D}_1$ spans x-y plane and $\boldsymbol{D}_2$ spans z-axis.

A dictionary model with $\boldsymbol{D} = [\boldsymbol{D}_1 \ \boldsymbol{D}_2] = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

and sparsity $s = 2$, happily represents all three cardinal planes.





Thus dictionary model seems "less constrained" than union-of-subspaces model.

(Still, focus on sparse dictionary representation hereafter.)

New dictionary learning method (SOUP-DIL)

Joint work with Sai Ravishankar and Raj Nadakuditi [9–11]

- In practice, must learn $\boldsymbol{D}$ from data, say $\boldsymbol{X}$
- Write sparse representation as Sum of OUter Products (SOUP):

$$\boldsymbol{X} \approx \boldsymbol{D}\boldsymbol{Z} = \boldsymbol{D}\boldsymbol{C}' = \sum_{j=1}^{J} \boldsymbol{d}_j \boldsymbol{c}_j'$$

  where $\boldsymbol{Z}' = \boldsymbol{C} = [\boldsymbol{c}_1 \; \ldots \; \boldsymbol{c}_J] \in \mathbb{R}^{N \times J}$ (coefficients for each atom)
- Replace individual atom sparsity constraint $\|\boldsymbol{z}_n\|_0 \leq s$ of K-SVD with aggregate sparsity regularizer: $\|\boldsymbol{Z}\|_0 = \|\boldsymbol{C}\|_0$.
  - Natural for DIctionary Learning (DIL) from training data.
  - Unnatural for image compression using sparse coding.

SOUP-DIL ℓ_0 formulation:

$$\boldsymbol{D}^* = \operatorname*{arg\,min}_{\boldsymbol{D} \in \mathbb{R}^{d \times J}} \min_{\boldsymbol{C} \in \mathbb{R}^{N \times J}} \|\boldsymbol{X} - \boldsymbol{D}\boldsymbol{C}'\|_F^2 + \lambda^2 \|\boldsymbol{C}\|_0 \quad \text{s.t.} \quad \begin{array}{l} \|\boldsymbol{d}_j\|_2 = 1 \; \forall j \\ \|\boldsymbol{c}_j\|_\infty \leq L \; \forall j \end{array}$$

SOUP-DIL algorithm

SOUP-DIL formulation [9–11]:

$$\boldsymbol{D}^* = \underset{\boldsymbol{D}\in\mathbb{R}^{d\times J}}{\arg\min} \min_{\boldsymbol{C}\in\mathbb{R}^{N\times J}} \|\boldsymbol{X} - \boldsymbol{D}\boldsymbol{C}'\|_F^2 + \lambda^2\|\boldsymbol{C}\|_0 \quad \text{s.t.} \quad \begin{array}{l} \|\boldsymbol{d}_j\|_2 = 1 \ \forall j \\ \|\boldsymbol{c}_j\|_\infty \le L \ \forall j \end{array}$$

- Block coordinate descent (BCD) algorithm
  - Sparse coding step for $\boldsymbol{C}$
  - Dictionary update step for $\boldsymbol{D}$
- Very simple update rules (low compute cost)
- Monotone descent of cost function $\Psi(\boldsymbol{D}, \boldsymbol{C})$
- Convergence theorem: for any given initialization $(\boldsymbol{D}^0, \boldsymbol{C}^0)$, all accumulation points of sequence $(\boldsymbol{D}, \boldsymbol{C})$
  - are critical points of cost Ψ and
  - are equivalent (reach same cost function value Ψ^*).
  - Furthermore: $\left\{\left\|\boldsymbol{D}^{(k)} - \boldsymbol{D}^{(k-1)}\right\|\right\} \to 0$. Same for $\left\{\boldsymbol{C}^{(k)}\right\}$.

SOUP-DIL updates

$$\boldsymbol{D}^* = \underset{\boldsymbol{D}\in\mathbb{R}^{d\times J}}{\arg\min} \min_{\boldsymbol{C}\in\mathbb{R}^{N\times J}} \|\boldsymbol{X} - \boldsymbol{D}\boldsymbol{C}'\|_F^2 + \lambda^2 \|\boldsymbol{C}\|_0 \quad \text{s.t.} \quad \begin{array}{l} \|\boldsymbol{d}_j\|_2 = 1 \ \forall j \\ \|\boldsymbol{c}_j\|_\infty \leq L \ \forall j \end{array}$$

Alternate: update one column $\boldsymbol{c}_j$ of $\boldsymbol{C}$ then one column $\boldsymbol{d}_j$ of $\boldsymbol{D}$.

- Sparse coding step: update $\boldsymbol{c}_j$ with residual $\boldsymbol{E}_j \triangleq \boldsymbol{X} - \sum_{k\neq j} \boldsymbol{d}_k \boldsymbol{c}_k'$:

$$\min_{\boldsymbol{c}_j} \|\boldsymbol{E}_j - \boldsymbol{d}_j \boldsymbol{c}_j'\|_F^2 + \lambda^2 \|\boldsymbol{c}_j\|_0 \quad \text{s.t.} \quad \|\boldsymbol{c}_j\|_\infty \leq L.$$

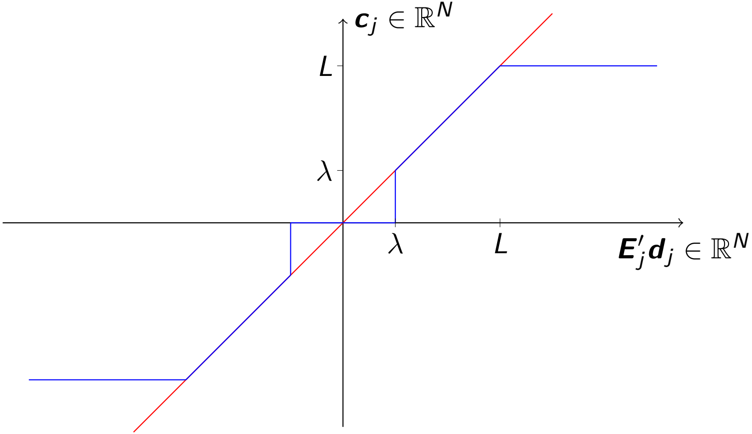
  Truncated (via L) hard thresholding of $\boldsymbol{E}_j'\boldsymbol{d}_j$ with threshold λ.

- Dictionary atom step: update $\boldsymbol{d}_j$

$$\min_{\boldsymbol{d}_j} \|\boldsymbol{E}_j - \boldsymbol{d}_j \boldsymbol{c}_j'\|_F^2 \quad \text{s.t.} \quad \|\boldsymbol{d}_j\|_2 = 1.$$

  Constrained least-squares solution: $\boldsymbol{d}_j = (\boldsymbol{E}_j \boldsymbol{c}_j) / \|\boldsymbol{E}_j \boldsymbol{c}_j\|_2$.

Truncated hard thresholding for SOUP-DIL

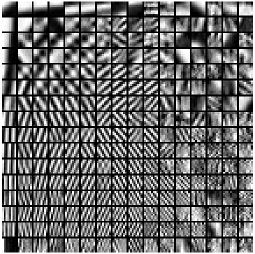


(Acts element-wise.) (In practice take $L = \infty$.)

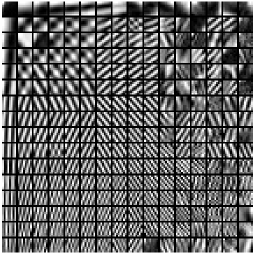
(Algorithm also provides a simple sparse coding method.)

Example: dictionary learning for Barbara

Barbara



K-SVD $\boldsymbol{D}$



SOUP-DIL $\boldsymbol{D}$

Denoising PSNR (dB) from [9]

σ	Noisy	O-DCT	K-SVD	SOUP-DIL
20	22.13	29.95	**30.83**	30.79
25	20.17	28.68	29.63	**29.64**
30	18.59	27.62	28.54	**28.63**
100	8.11	21.87	21.87	**21.97**

SOUP-DIL faster than K-SVD

Regularization using SOUP-DIL

- Large image $\boldsymbol{x} \Longrightarrow$ extract M patches $\boldsymbol{X} = [\boldsymbol{P}_1\boldsymbol{x} \; \ldots \; \boldsymbol{P}_M\boldsymbol{x}]$.
- Assume patch $\boldsymbol{x}_m = \boldsymbol{P}_m\boldsymbol{x} \approx \boldsymbol{D}\boldsymbol{z}_m$ has (aggregate) sparse representation in dictionary $\boldsymbol{D} \in \mathbb{R}^{d \times J}$ where d is patch size.
- Two variations:
  - Use dictionary $\boldsymbol{D}$ from training data:

$$\mathsf{R}(\boldsymbol{x}) = \mathsf{R}(\boldsymbol{X}) = \min_{\boldsymbol{C} \in \mathcal{C}} \|\boldsymbol{X} - \boldsymbol{D}\boldsymbol{C}'\|_F^2 + \lambda^2 \|\boldsymbol{C}\|_0$$

  - Learn $\boldsymbol{D}$ while reconstructing (blind / adaptive)

$$\mathsf{R}(\boldsymbol{x}) = \min_{\boldsymbol{D} \in \mathcal{D}} \min_{\boldsymbol{C} \in \mathcal{C}} \|\boldsymbol{X} - \boldsymbol{D}\boldsymbol{C}'\|_F^2 + \lambda^2 \|\boldsymbol{C}\|_0$$

$$\mathcal{D} = \left\{\boldsymbol{D} \in \mathbb{R}^{d \times J} : \|\boldsymbol{d}_j\|_2 = 1 \; \forall j\right\}, \; \mathcal{C} = \left\{\boldsymbol{C} \in \mathbb{R}^{M \times J} : \|\boldsymbol{c}_j\|_\infty \leq L \; \forall j\right\}$$

- $\mathsf{R}(\boldsymbol{x}) \approx 0$ if patches can be represented closely with "sufficiently few" non-zero coefficients (depends on λ).
- Ignore constraint $\|\boldsymbol{c}_j\|_\infty \leq L$ in practice.
- Bayesian interpretation?

Outline

Ill-posed problems and regularization
- Classical "hand crafted" regularizers
- Data-driven (adaptive / learned) regularizers
- Data-driven regularized MRI via dictionary learning
- Extension: learning low-rank atoms
  - DIctioNary with lOw-ranK ATars (DINO-KAT)

Dynamic MR imaging
- DINO-KAT for dynamic MR

Summary

Backup
- DINO-KAT convergence guarantees
- Online method

MR reconstruction using adaptive dictionary regularizer

Dictionary-blind MR image reconstruction:

$$\hat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}} \frac{1}{2} \|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\|_2^2 + \beta \, \mathsf{R}(\boldsymbol{x})$$

$$\mathsf{R}(\boldsymbol{x}) = \min_{\boldsymbol{D} \in \mathcal{D}} \min_{\boldsymbol{Z}' \in \mathcal{C}} \sum_{m=1}^{M} \left(\|\boldsymbol{P}_m \boldsymbol{x} - \boldsymbol{D}\boldsymbol{z}_m\|_2^2 + \lambda^2 \|\boldsymbol{z}_m\|_0 \right)$$

where $\boldsymbol{P}_m$ extracts mth of M image patches.

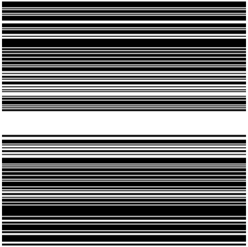
In words: of the many images...

Alternating (nested) minimization:

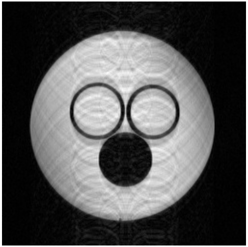
- Fixing $\boldsymbol{x}$ and $\boldsymbol{D}$, update each $\boldsymbol{z}_j$ via hard-thresholding
- Fixing $\boldsymbol{x}$ and $\boldsymbol{Z}$, update $\boldsymbol{D}$ using SOUP-DIL
- Fixing $\boldsymbol{Z}$ and $\boldsymbol{D}$, updating $\boldsymbol{x}$ is a quadratic problem.
  - Efficient FFT solution for single-coil Cartesian MRI.
  - Use CG for non-Cartesian and/or parallel MRI.
- Non-convex, but monotone decreasing and some convergence theory [9].

2D CS MRI results I

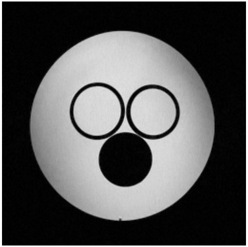
Sampling ($2.5\times$)



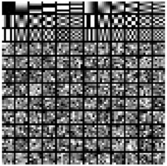
Zero-Filled



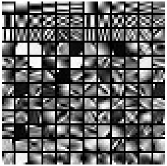
SOUP-DILLO-MRI



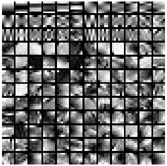
Initial $\boldsymbol{D}$



Learned real$\{\boldsymbol{D}\}$



Learned imag$\{\boldsymbol{D}\}$

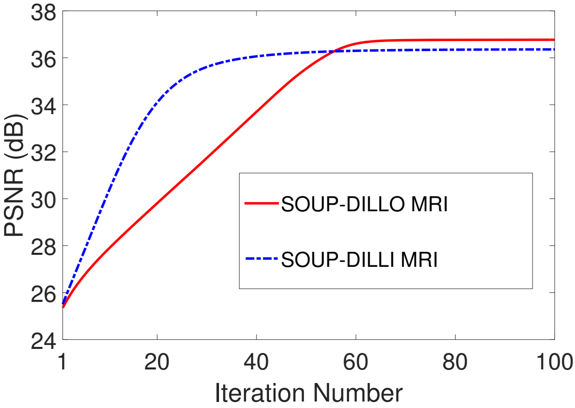


6×6 patches

$\boldsymbol{D} \in \mathbb{C}^{6^2 \times 144}$

[9]

2D CS MRI results II

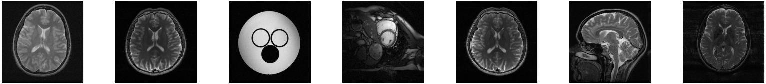


(SNR compared to fully sampled image.)

Using $\|\boldsymbol{z}_m\|_0$ leads to higher SNR then $\|\boldsymbol{z}_m\|_1$.

Adaptive case is non-convex anyway...

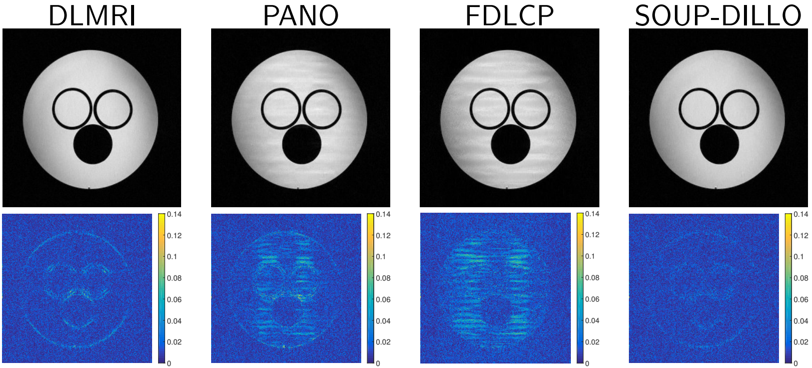
2D CS MRI results III



(a) (b) (c) (d) (e) (f) (g)

Im.	Samp.	Acc.	0-fill	Sparse MRI	PANO	DLMRI	SOUP-DILLI	SOUP-DILLO
a	Cart.	7x	27.9	28.6	**31.1**	**31.1**	30.8	**31.1**
b	Cart.	2.5x	27.7	31.6	41.3	40.2	38.5	**42.3**
c	Cart.	2.5x	24.9	29.9	34.8	36.7	36.6	**37.3**
c	Cart.	4x	25.9	28.8	**32.3**	32.1	32.2	**32.3**
d	Cart.	2.5x	29.5	32.1	36.9	38.1	36.7	**38.4**
e	Cart.	2.5x	28.1	31.7	40.0	38.0	37.9	**41.5**
f	2D rand.	5x	26.3	27.4	30.4	30.5	30.3	**30.6**
g	Cart.	2.5x	32.8	39.1	41.6	41.7	42.2	**43.2**
Ref.				[12]	[13]	[14]	[9]	[9]

2D CS MRI results IV



[14] [13] [15] [9]

Summary: 2D static MR reconstruction from under-sampled data with adaptive dictionary learning and convergent algorithm, faster than K-SVD approach of DLMRI.

Outline

- Ill-posed problems and regularization
  - Classical “hand crafted” regularizers
  - Data-driven (adaptive / learned) regularizers
  - Data-driven regularized MRI via dictionary learning
  - Extension: learning low-rank atoms
    - DIctioNary with lOw-ranK AToms (DINO-KAT)
- Dynamic MR imaging
  - DINO-KAT for dynamic MR
- Summary
- Backup
  - DINO-KAT convergence guarantees
  - Online method

Recall SOUP-DIL ℓ_0 formulation for dictionary learning from data $\boldsymbol{X}$:

$$\boldsymbol{D}^* = \underset{\boldsymbol{D}\in\mathbb{R}^{d\times J}}{\arg\min} \min_{\boldsymbol{C}\in\mathbb{R}^{N\times J}} \|\boldsymbol{X} - \boldsymbol{D}\boldsymbol{C}'\|_F^2 + \lambda^2 \|\boldsymbol{C}\|_0 \quad \text{s.t.} \quad \begin{array}{l} \|\boldsymbol{d}_j\|_2 = 1 \; \forall j \\ \|\boldsymbol{c}_j\|_\infty \le L \; \forall j. \end{array}$$

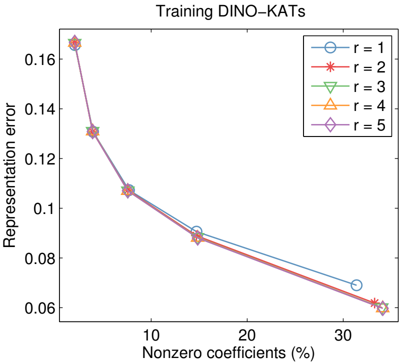
Recent extension [10]
DIctioNary with lOw-ranK AToms (DINO-KAT) model:

$$\boldsymbol{D}^* = \underset{\boldsymbol{D}\in\mathbb{R}^{d\times J}}{\arg\min} \min_{\boldsymbol{C}\in\mathbb{R}^{N\times J}} \|\boldsymbol{X} - \boldsymbol{D}\boldsymbol{C}'\|_F^2 + \lambda^2 \|\boldsymbol{C}\|_0 \quad \text{s.t.} \quad \begin{array}{l} \|\boldsymbol{d}_j\|_2 = 1 \; \forall j \\ \|\boldsymbol{c}_j\|_\infty \le L \; \forall j \\ \operatorname{rank}\{\texttt{reshape}(\boldsymbol{d}_j)\} \le r, \end{array}$$

where $\texttt{reshape}(\boldsymbol{d}_j)$ reshapes dictionary atom $\boldsymbol{d}_j$ into a 2D array.

DINO-KAT: why low-rank atoms?

- Low-rank atoms are less prone to over-fitting.
- Model structure (*e.g.*, temporal correlation) of dynamic data.
- Learned dictionary atoms on patch data often have only a few dominant singular values.



Representation error $\|\boldsymbol{X} - \boldsymbol{DC}'\|_F / \|\boldsymbol{X}\|_F$ versus sparsity λ for several atom ranks r for $8 \times 8 \times 5$ space-time patches from (fully sampled) cardiac perfusion images.

DINO-KAT algorithm

DIctioNary with lOw-ranK ATtoms (DINO-KAT) model:

$$\boldsymbol{D}^* = \arg\min_{\boldsymbol{D}\in\mathbb{R}^{d\times J}} \min_{\boldsymbol{C}\in\mathbb{R}^{N\times J}} \|\boldsymbol{X} - \boldsymbol{D}\boldsymbol{C}'\|_F^2 + \lambda^2 \|\boldsymbol{C}\|_0 \quad \text{s.t.} \quad \begin{array}{l} \|\boldsymbol{d}_j\|_2 = 1 \ \forall j \\ \|\boldsymbol{c}_j\|_\infty \le L \ \forall j \\ \text{rank}\{\texttt{reshape}(\boldsymbol{d}_j)\} \le r, \end{array}$$

Block coordinate descent (BCD) algorithm (monotone descent) with simple update rules (low compute cost)

- Sparse coding step for $\boldsymbol{C}$ uses same truncated hard thresholding
- Dictionary atom update step for $\boldsymbol{d}_j$:

$$\arg\min_{\boldsymbol{d}_j} \|\boldsymbol{E}_j - \boldsymbol{d}_j\boldsymbol{c}_j'\|_F^2 \quad \text{s.t.} \quad \|\boldsymbol{d}_j\|_2 = 1, \ \text{rank}\{\texttt{reshape}(\boldsymbol{d}_j)\} \le r$$

Simple solution: $\texttt{reshape}(\boldsymbol{d}_j) = \frac{\boldsymbol{U}_r\boldsymbol{\Sigma}_r\boldsymbol{V}_r'}{\|\boldsymbol{\Sigma}_r\|_F}$
$\boldsymbol{U}_r\boldsymbol{\Sigma}_r\boldsymbol{V}_r'$ is the rank-r truncated SVD of $\texttt{reshape}(\boldsymbol{E}_j\boldsymbol{c}_j)$.

Outline

Dynamic MRI overview

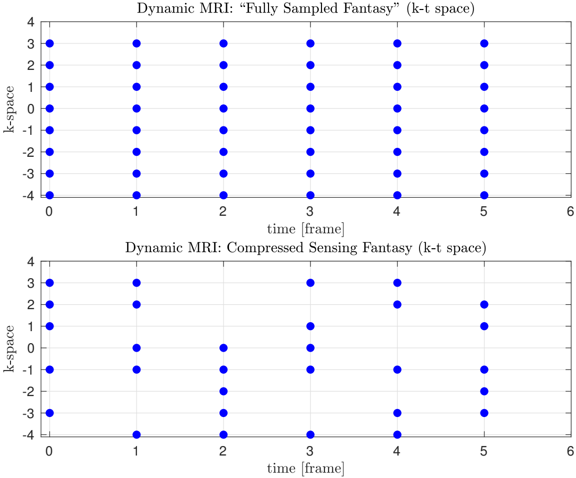
"dynamic" = changing over time = motion [16–19]

- Nuisance motions:
  - Breathing
  - Cardiac
  - Peristalsis
  - Tremors
  - Kids ...

  $\Longrightarrow$ Faster scans (shorter time) can help reduce motion blur
- Motions of interest (true "dynamic" scans):
  - Vocalization (for speech studies)
  - Cardiac (for function)
  - Joint articulation (musculoskeletal scans)
  - Contrast agent (blood flow / perfusion)
  - Diffusion

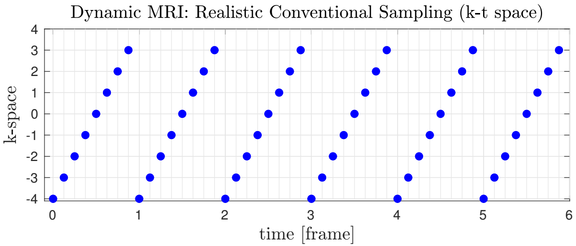
$\Longrightarrow$ Trade-offs between temporal resolution and spatial resolution

Dynamic MRI sampling: Fantasy edition



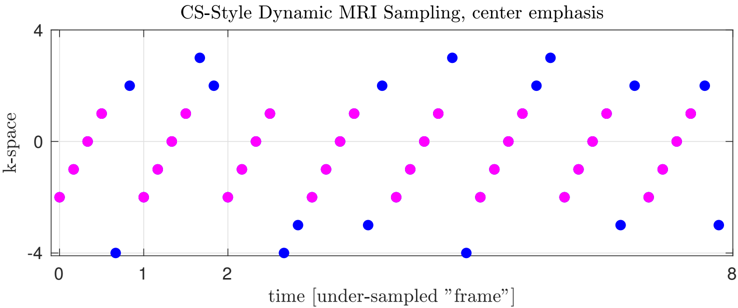
- Scan "twice as fast" !?
- Matrix completion problem!? $\Longrightarrow$... robust PCA (L+S) ... [20, 21]

Dynamic MRI sampling: Reality I



- All 3D dynamic MRI data is *inherently under-sampled*
- No real "fully sampled" data exists, now or ever
- Unlikely to satisfy any "matrix completion" sufficient conditions (N measurements but N^2 unknowns per frame)
- Retrospective "under sampling" of "fully sampled" dynamic data is dubious
- Opportunity: powerful signal models needed for reconstruction from such data
- Challenge: validation of signal models given such highly incomplete data (low-rank / locally low rank / tensors / wavelets / non-local patches / ...)

Dynamic MRI sampling: Reality II



Outline

DINO-KAT as an adaptive (data-driven) regularizer:

$$\hat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}} \frac{1}{2} \|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\|_2^2 + \beta \, \mathsf{R}(\boldsymbol{x})$$

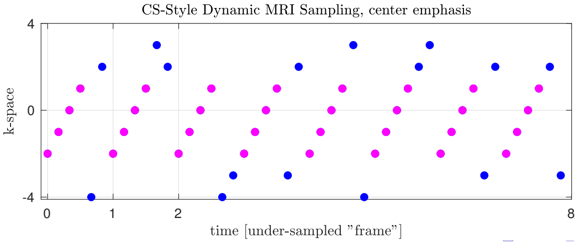
$$\mathsf{R}(\boldsymbol{x}) = \min_{\boldsymbol{D} \in \mathbb{C}^{d \times J}} \min_{\boldsymbol{Z} \in \mathbb{C}^{J \times M}} \sum_{m=1}^{M} \|\boldsymbol{P}_m \boldsymbol{x} - \boldsymbol{D}\boldsymbol{z}_m\|_2^2 + \lambda^2 \|\boldsymbol{z}_m\|_0$$

$$\text{s.t. } \|\boldsymbol{d}_j\|_2 = 1 \; \forall j, \; \|\boldsymbol{z}_m\|_\infty \leq L \; \forall m, \; \text{rank}\{\texttt{reshape}(\boldsymbol{d}_j)\} \leq r$$

Block coordinate descent (BCD) algorithm (monotone descent)

- Update coefficients $\boldsymbol{Z}$: sparse coding via hard thresholding
- Dictionary atom update of $\boldsymbol{d}_j$: uses residual, SVD
- Image update uses FFT (single coil Cartesian) or CG

- Latent signal vector $\boldsymbol{x} \in \mathbb{C}^{n_y n_x n_t}$ modeled as n_t frames, each of dimension $N = n_x \times n_y$ or $N = n_x \times n_y \times n_z$.
- k-space data $\boldsymbol{y} \in \mathbb{C}^{n_{\text{sample}} n_c}$ acquired using n_c coils.
- Sensing matrix $\boldsymbol{A}$ includes:
  - coil sensitivity maps,
  - 2D or 3D spatial Fourier transform,
  - k-space sampling pattern.
- $\boldsymbol{y}$ is undersampled, so regularization is required to estimate dynamic image sequence $\boldsymbol{x}$.

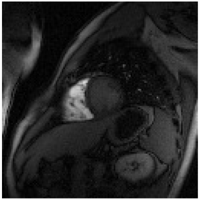
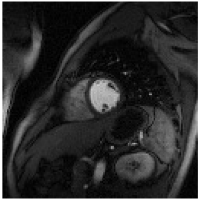


Dynamic MRI models

- **Low-rank and sparse (k-t SLR)** [22]:
  - Model: $\boldsymbol{x}$ reshaped into an $N \times n_t$ space-time matrix, is both low-rank and (transform) sparse
- **Low-rank plus sparse (L + S)** [20, 21]
  - model: $\boldsymbol{x} = \boldsymbol{x}_\mathrm{L} + \boldsymbol{x}_\mathrm{S}$,
  - $\boldsymbol{x}_\mathrm{L}$ reshaped into a $N \times n_t$ space-time matrix is low-rank,
  - $\boldsymbol{x}_\mathrm{S}$ is (transform) sparse.
- **DINO-KAT for dynamic MRI** [10, 23]:
  - Extract $p \times p \times q$ patches of $\boldsymbol{x}_\mathrm{S}$.
  - Model patches as sparse w.r.t. an adaptive (learned) dictionary $\boldsymbol{D}$.
  - Model dictionary atoms $\{\boldsymbol{d}_j\}$ as low-rank when reshaped into $p^2 \times q$ space-time matrices.
  - Blind compressed sensing model [14].

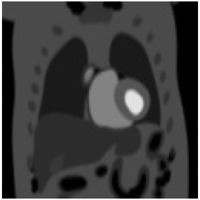
Dynamic MRI data

Cardiac perfusion data (ref. frames 7, 13)

$128^2 \times 40$ fr.
$3.2^2 \times 8\ \mathrm{mm}^3$
12 coil
$\Delta T = 307$ ms
Otazo et al. [21] (L+S)

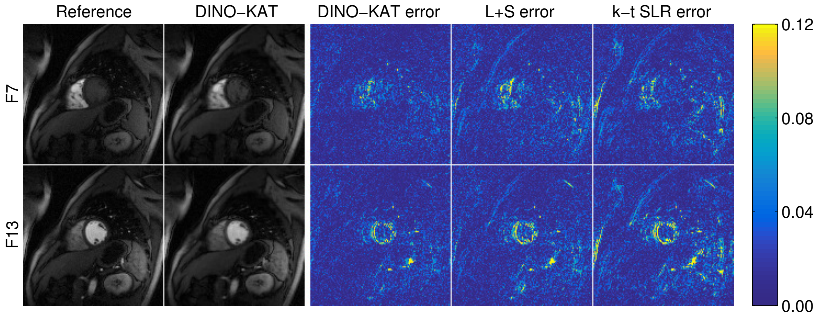
PINCAT data (reference frames 16, 25)

$128^2 \times 50$ fr.
1.5 mm^2
1 coil?
$9 \times$ acc.
Lingala et al. [22] (k-t SLR)

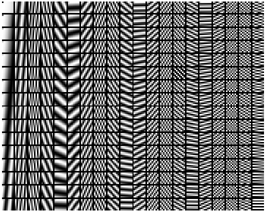
Dynamic MRI reconstruction results

8× acceleration

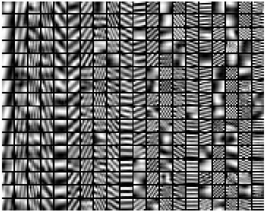


Learned dynamic patch dictionary: cardiac perfusion

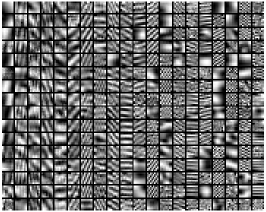
Initial atoms (DCT matrix)



Real-part of learned atoms



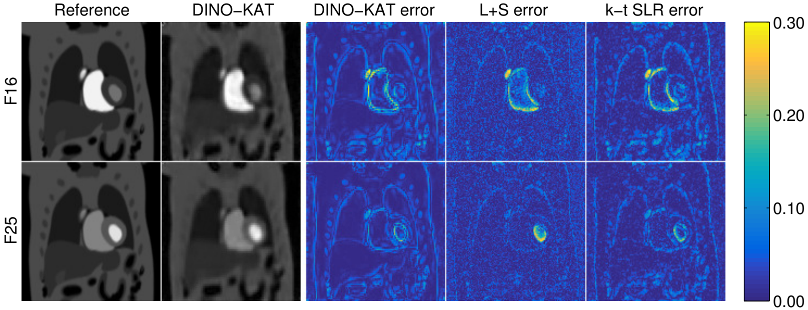
Imaginary-part of learned atoms



- First temporal slices of $8 \times 8 \times 5$ atoms
- Learned atoms adapt to structure of data

Reconstructions: PINCAT phantom

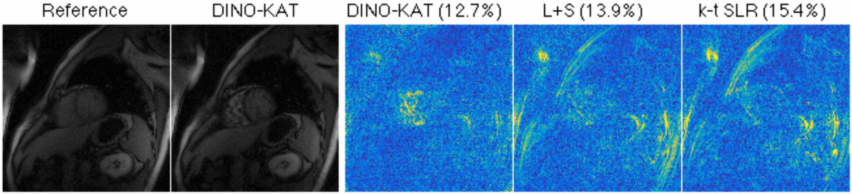
9× acceleration



- Two representative frames of each reconstruction
- DINO-KAT method shows less error
  than the L+S and k-t SLR (L&S) methods

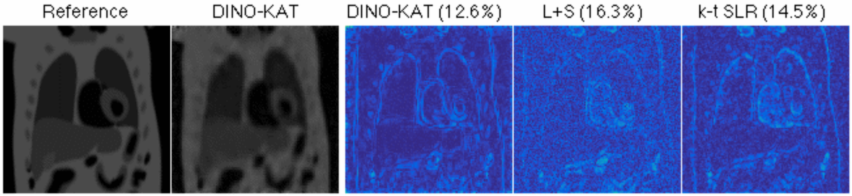
Reconstructions: cardiac perfusion

8× acceleration



Reconstructions: PINCAT phantom

$9\times$ acceleration



Quantitative results for cardiac perfusion data

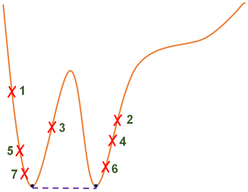
Acceleration	4x	8x	12x	16x	20x	24x
NRMSE (L+S) %	10.93	14.00	15.80	18.87	21.33	23.36
NRMSE (Fixed D) %	11.29	13.76	15.33	18.31	20.77	22.82
NRMSE ($r = 5$) %	10.85	13.08	14.37	17.01	19.19	21.35
NRMSE ($r = 1$) %	**10.57**	**12.90**	**14.20**	**16.77**	**18.74**	**20.91**
Gain over $L + S$ (dB)	0.29	0.71	0.92	1.03	1.13	0.96
Gain over $r = 5$ (dB)	0.23	0.12	0.10	0.13	0.21	0.18

Summary

- Data-driven / adaptive regularization
  - Beneficial for under-sampled MRI reconstruction
  - Dictionary atom structure (*e.g.*, low rank) further helpful
  - SOUP provides reasonably computationally efficient methods (vs KSVD)
  - Convergence theory (unlike KSVD)
- Future work:
  - Synthesis (*e.g.*, dictionary) vs analysis (*e.g.*, transform learning) formulations
  - Online methods for reduced memory, better adaptation [24–27]
  - Other machine-learning methods (deep...) ?
  - T-MI special issue on Machine-Learning for Image Reconstruction

Outline

Convergence guarantees



Theorem (Convergence guarantees)

[10, 23, 27]
Let $\{D^t, C^t\}$ be the iterate sequence generated by the BCD algorithm for DINO-KAT. Then

- *The cost function $\Psi(D^t, C^t)$ is monotone decreasing and converges to a finite value, say $\Psi^\star$*
- *The iterate sequence $\{D^t, C^t\}$ is bounded and its accumulation points have common cost value $\Psi^\star$*

Online Transform Learning Formulation

- For frame $t = 1, 2, 3, \ldots,$ solve

$$\text{(P3)} \quad \left\{\hat{W}_t, \hat{x}_t\right\} = \underset{W, x_t}{\arg\min} \frac{1}{t} \sum_{j=1}^{t} \Big\{ \overbrace{\|Wy_j - x_j\|_2^2}^{\text{Sparsification Error}} + \overbrace{\lambda_j v(W)}^{\text{Regularizer}} \Big\}$$

$$\text{s.t.} \quad \|x_t\|_0 \leq s, \; x_j = \hat{x}_j, \; 1 \leq j \leq t-1.$$

- Minimize the average cost computed over the observed signals.
- $\hat{W}_t^{-1} \hat{x}_t$ is an (e.g., denoised) estimate of y_t.
- For non-stationary data, use forgetting factor $\rho \in [0, 1]$, to diminish the influence of old data.

$$\frac{1}{\sum_{j=1}^{t} \rho^{t-j}} \sum_{j=1}^{t} \rho^{t-j} \left\{\|Wy_j - x_j\|_2^2 + \lambda_j v(W)\right\}$$

Fast Online Transform Learning Algorithm I

- **Sparse Coding:** solve for x_t in (P3) with fixed $W = \hat{W}_{t-1}$.

$$\min_{x_t} \|Wy_t - x_t\|_2^2 \text{ s.t. } \|x_t\|_0 \leq s$$

  - **Cheap Solution:** $\hat{x}_t = H_s(Wy_t)$.

Fast Online Transform Learning Algorithm II

- **Transform Update:** solve for W with $x_t = \hat{x}_t$.

$$\min_{W} \frac{1}{t} \sum_{j=1}^{t} \left\{ \|Wy_j - x_j\|_2^2 + \lambda_j \left(\|W\|_F^2 - \log|\det W| \right) \right\}$$

$$\hat{W}_t = 0.5 R_t \left(\Sigma_t + \left(\Sigma_t^2 + 2\beta_t I \right)^{\frac{1}{2}} \right) Q_t^T L_t^{-1}$$

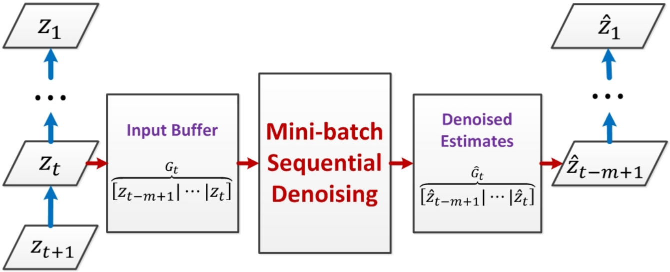
  - $t^{-1} \sum_{j=1}^{t} \left(y_j y_j^T + \lambda_0 \|y_j\|_2^2 I \right) = L_t L_t^T$. **Perform rank-1 update.**
  - $\beta_t = \lambda_0 t^{-1} \sum_{j=1}^{t} \|y_j\|_2^2$. $Q_t \Sigma_t R_t^T$ is full SVD of $L_t^{-1} \Theta_t = t^{-1} \sum_{j=1}^{t} L_t^{-1} y_j x_j^T$.
    - $L_t^{-1} \Theta_t \approx (1 - t^{-1}) L_{t-1}^{-1} \Theta_{t-1} + t^{-1} L_t^{-1} y_t x_t^T$
      $\Longrightarrow$ **efficient rank-1 SVD update.**

Online Transform Learning (OTL) Convergence Result

- **Assumption:** y_t are i.i.d. random samples from the sphere $S^{n-1} = \{y \in \mathbb{R}^n : \|y\|_2 = 1\}$, for an absolutely continuous probability measure p.
- We consider minimizing the expected [28] learning cost:

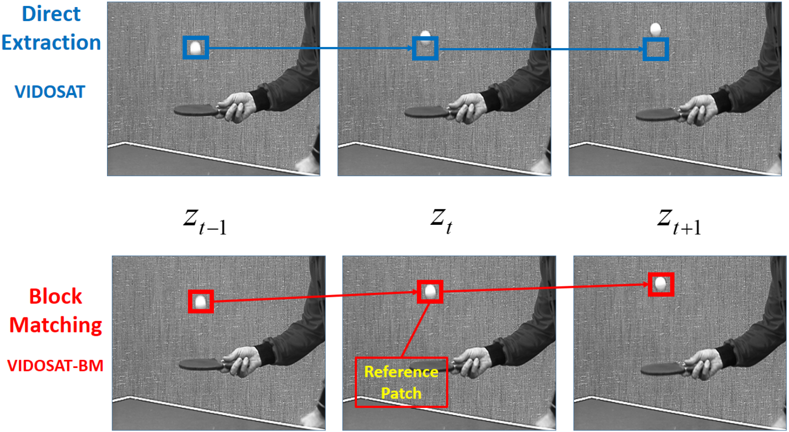
$$g(W) = \mathbb{E}_y \left[\|Wy - H_s(Wy)\|_2^2 + \lambda_0 \|y\|_2^2 \, v(W) \right].$$

- **Main Result [25]: $\hat{W}_t$ in OTL converges to the set of stationary points of $g(W)$ almost surely.**
  $\hat{W}_{t+1} - \hat{W}_t \sim O(1/t)$.



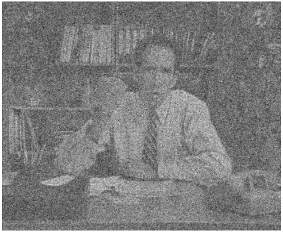
- z_t: noisy frame, $\hat{z}_t$: denoised version.
- G_t: 3D array with m frames formed using a sliding window scheme.
- Overlapping 3D patches in the G_t's are denoised sequentially with OTL.
- Denoised patches averaged at 3D locations to yield frame estimates.

3D Patches in Proposed VIDOSAT Method



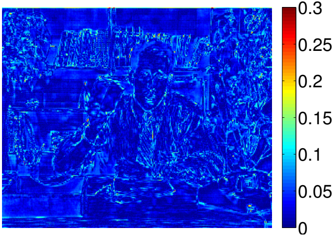
- Transform is learned online from sequentially extracted 3D patches.

Video Denoising Example: Salesman

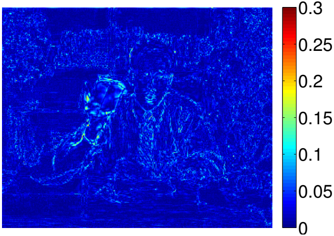


Noisy frame

VIDOSAT (PSNR = 30.97 d

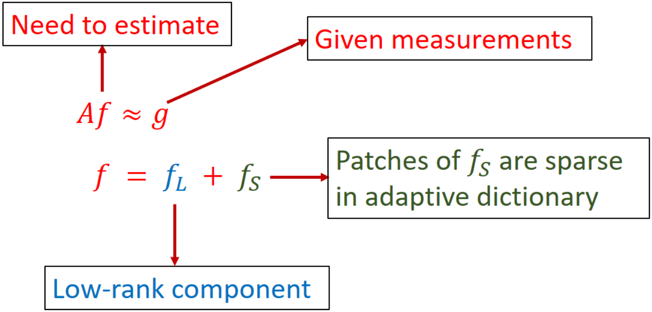


VBM4D [29] Error (PSNR = 27.20 dB)



VIDOSAT Error

LASSI Model for Dynamic Image Reconstruction



- Low-rank + **Adaptive** Sparse SIgnal (LASSI) model for dynamic data.
- **Goal:** learn dictionary and reconstruct (f_L, f_S) from limited g.
- Efficient algorithm for LASSI estimation proposed recently [23]

Dynamic MRI 8x Cartesian Undersampling (12 coils)

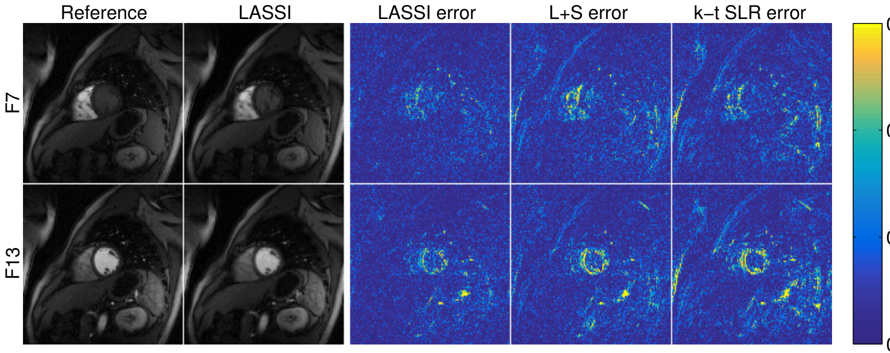
```
open
~/tex/paper/done/17/ravishankar-17-lra/gifs/tmi_otazo_8x_lassi_vs_lps.gif
```

```
open
```

```
~/tex/conf/misc/ivmsp,16/ravishankar-lassi/talk/tmi_otazo_8x_lassi_vs_lps.avi
```

Click for LASSI and L+S Results

Cardiac Perfusion Results: 8x Acceleration



- **Error maps for LASSI show smaller distortions than for the k-t SLR [22] (L & S) method and the L+S [21] method.**

Bibliography I

[1] E. Candès and J. K. Romberg. “Signal recovery from random projections.” In: *Proc. SPIE 5674 Computational Imaging III*. 2005, 76–86.

[2] L. A. Shepp and B. F. Logan. “The Fourier reconstruction of a head section.” In: *IEEE Trans. Nuc. Sci.* 21.3 (June 1974), 21–43.

[3] M. Mignotte. “A non-local regularization strategy for image deconvolution.” In: *Pattern Recognition Letters* 29.16 (Dec. 2008), 2206–12.

[4] Z. Yang and M. Jacob. “Nonlocal regularization of inverse problems: A unified variational framework.” In: *IEEE Trans. Im. Proc.* 22.8 (Aug. 2013), 3192–203.

[5] H. Zhang et al. “Statistical image reconstruction for low-dose CT using nonlocal means-based regularization.” In: *Computerized Medical Imaging and Graphics* 38.6 (Sept. 2014), 423–35.

[6] S. Y. Chun, Y. K. Dewaraja, and J. A. Fessler. “Alternating direction method of multiplier for tomography with non-local regularizers.” In: *IEEE Trans. Med. Imag.* 33.10 (Oct. 2014), 1960–8.

[7] R. Vidal. “Subspace clustering.” In: *IEEE Sig. Proc. Mag.* 28.2 (Mar. 2011), 52–68.

[8] T. Zhang, A. Szlam, and G. Lerman. “Median K-Flats for hybrid linear modeling with many outliers.” In: *Proc. Intl. Conf. Comp. Vision*. 2009, 234–41.

[9] S. Ravishankar, R. R. Nadakuditi, and J. A. Fessler. “Efficient sum of outer products dictionary learning (SOUP-DIL) and its application to inverse problems.” In: *IEEE Trans. Computational Imaging* (2017). To appear.

[10] S. Ravishankar et al. “Efficient learning of dictionaries with low-rank atoms.” In: *IEEE GlobalSIP*. 2016, 222–6.

[11] S. Ravishankar, R. R. Nadakuditi, and J. A. Fessler. *Efficient sum of outer products dictionary learning (SOUP-DIL) - The ℓ_0 method*. arxiv 1511.08842. 2015.

[12] M. Lustig and J. M. Pauly. “SPIRiT: Iterative self-consistent parallel imaging reconstruction from arbitrary k-space.” In: *Mag. Res. Med.* 64.2 (Aug. 2010), 457–71.

[13] X. Qu et al. “Magnetic resonance image reconstruction from undersampled measurements using a patch-based nonlocal operator.” In: *Med. Im. Anal.* 18.6 (Aug. 2014), 843–56.

[14] S. Ravishankar and Y. Bresler. “MR image reconstruction from highly undersampled k-space data by dictionary learning.” In: *IEEE Trans. Med. Imag.* 30.5 (May 2011), 1028–41.

[15] Z. Zhan et al. “Fast multiclass dictionaries learning with geometrical directions in MRI reconstruction.” In: *IEEE Trans. Biomed. Engin.* 63.9 (Sept. 2016), 1850–61.

[16] R. S. Lawson. “Application of mathematical methods in dynamic nuclear medicine studies.” In: *Phys. Med. Biol.* 44.4 (Apr. 1999), R57–98.

[17] A. R. Padhani. “Dynamic contrast-enhanced MRI in clinical oncology: Current status and future directions.” In: *J. Mag. Res. Im.* 16.4 (Oct. 2002), 407–22.

[18] S. Bonnet et al. “Dynamic X-ray computed tomography.” In: *Proc. IEEE* 91.10 (Oct. 2003), 1574–87.

[19] S. G. Lingala and M. Jacob. *Accelerated Dynamic MRI using adaptive signal models*. MRI: Physics, Image Reconstruction, and Analysis, CRC Press (Book Chapter). 2015.

[20] E. J. Candès et al. “Robust principal component analysis?” In: *J. Assoc. Comput. Mach.* 58.3 (May 2011), 1–37.

[21] R. Otazo, E. Candès, and D. K. Sodickson. “Low-rank plus sparse matrix decomposition for accelerated dynamic MRI with separation of background and dynamic components.” In: *Mag. Res. Med.* 73.3 (Mar. 2015), 1125–36.

[22] S. G. Lingala et al. “Accelerated dynamic MRI exploiting sparsity and low-rank structure: k-t SLR.” In: *IEEE Trans. Med. Imag.* 30.5 (May 2011), 1042–54.

[23] S. Ravishankar et al. "Low-rank and adaptive sparse signal (LASSI) models for highly accelerated dynamic imaging." In: *IEEE Trans. Med. Imag.* 36.5 (May 2017), 1116–28.

[24] S. Ravishankar, B. Wen, and Y. Bresler. "Online sparsifying transform learning - Part I: algorithms." In: *ieee-jstsp* 9.4 (June 2015), 625–36.

[25] S. Ravishankar and Y. Bresler. "Online sparsifying transform learning - Part II: convergence analysis." In: *ieee-jstsp* 9.4 (June 2015), 637–46.

[26] S. Ravishankar et al. "Efficient online dictionary adaptation and image reconstruction for dynamic MRI." In: *Proc., IEEE Asilomar Conf. on Signals, Systems, and Comp.* Submitted. 2017.

[27] B. E. Moore and S. Ravishankar. "Online data-driven dynamic image restoration using DINO-KAT models." In: *Proc. IEEE Intl. Conf. on Image Processing*. To appear. 2017.

[28] L. Bottou. "Online algorithms and stochastic approximations." In: *Online Learning and Neural Networks*. Ed. by D. Saad. UK: Cambridge, 1998.

[29] M. Maggioni et al. "Video denoising, deblocking, and enhancement through separable 4-D nonlocal spatiotemporal transforms." In: *IEEE Trans. Im. Proc.* 21.9 (Sept. 2012), 3952–66.