

Adaptive regularization methods for dynamic MRI image reconstruction



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

Imaging Summit: Medical Imaging in the era of AI

2017-07-10

Commercial availability of iterative methods for human scanners per FDA 510(k) dates:

- ▶ **PET/SPECT**
Unregularized OS-EM \approx 1997
- ▶ **X-ray CT**
Regularized MBIR [2011-11-09 for GE Veo]
(Installed at UM in Jan. 2012)
- ▶ **PET**
Regularized EM variant (Q.Clear) 2014-03-21
- ▶ **MRI**
Compressed sensing! (Sparsity-based regularization)
[2017-01-27 for Siemens Cardiac Cine]
[2017-04-20 for GE HyperSense]
- ▶ **Ultrasound?**

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

Deep reconstruction using imaging physics

Summary

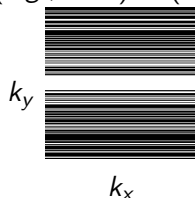
Backup

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

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

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

- ▶ compressed sensing (e.g., MRI) (\mathbf{A} “random” rows of DFT)



- ▶ deblurring (restoration)
- ▶ in-painting
- ▶ denoising (not ill posed)

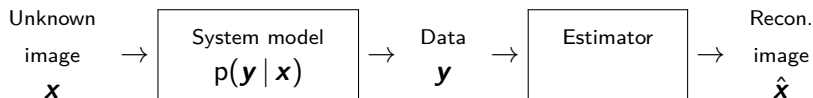
(\mathbf{A} Toeplitz)

(\mathbf{A} subset of rows of \mathbf{I})

($\mathbf{A} = \mathbf{I}$)

- ▶ Reduce scan time (?)
 - Patient comfort
 - Scan cost / throughput
 - Motion artifacts (Philips at ISMRM 2017)
- ▶ Improve spatial resolution (collect higher k-space lines)
- ▶ Improve scan diversity for quantitative MRI
- ▶ Improve temporal resolution trade-off in dynamic MRI

(But under-sampling leads to ill-posed inverse problems...)



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

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

For gaussian measurement errors and a linear forward model:

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

where $\|\mathbf{y}\|_{\mathbf{W}}^2 = \mathbf{y}'\mathbf{W}\mathbf{y}$

and $\mathbf{W}^{-1} = \text{Cov}\{\mathbf{y} | \mathbf{x}\}$ is known
(\mathbf{A} from physics, \mathbf{W} from statistics)

- ▶ If all images \mathbf{x} are “plausible” (have non-zero probability) then

$$p(\mathbf{x}) \propto e^{-R(\mathbf{x})} \implies -\log p(\mathbf{x}) \equiv R(\mathbf{x})$$

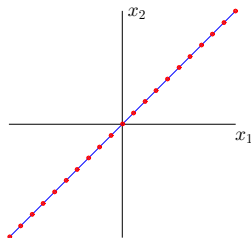
(from fantasy / imagination / wishful thinking / data)

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

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

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

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



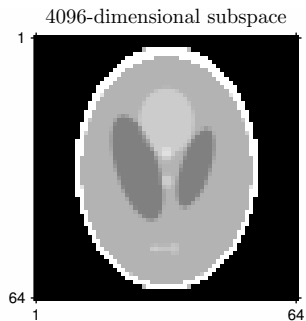
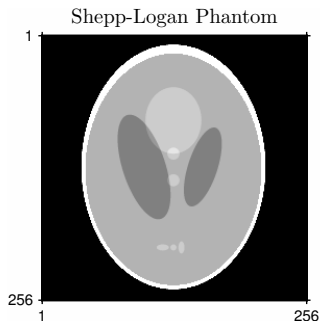
Assume $\mathbf{x} = \mathbf{D}\mathbf{z}$ where $\mathbf{D} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ and $\mathbf{z} \in \mathbb{R}^1$

(\mathbf{z} has only one nonzero element so very sparse!?)

Estimate coefficient(s): $\hat{\mathbf{z}} = \arg \min_{\mathbf{z}} \|\mathbf{y} - \mathbf{A}\mathbf{D}\mathbf{z}\|_2^2$, then $\hat{\mathbf{x}} = \mathbf{D}\hat{\mathbf{z}}$,
where usually $\text{cond}(\mathbf{D}'\mathbf{A}'\mathbf{A}\mathbf{D}) \ll \text{cond}(\mathbf{A}'\mathbf{A})$.

Why not use subspace models?

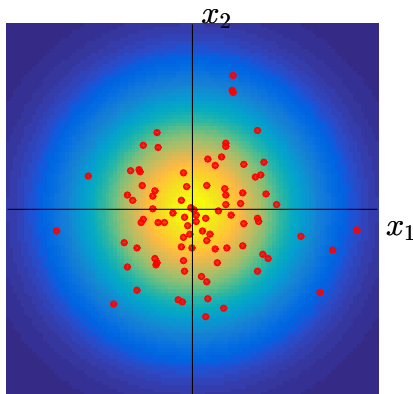
Candès and Romberg (2005) [1] used 22 (noiseless) CT projection views (*i.e.*, 22 pseudo-radial lines in MRI), each with 256 samples.
 $\implies 22 \cdot 256 = 5632$ measured values,
vs $256^2 = 65536$ unknown pixels



Subspace representation (using pixel basis) is undesirably coarse.

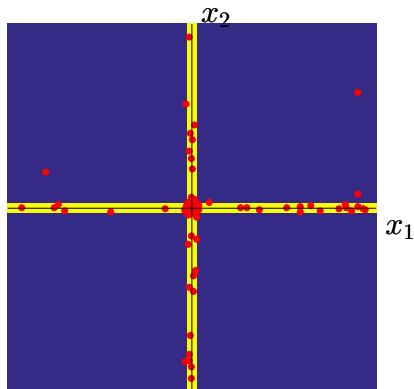
- ▶ Tikhonov regularization (IID gaussian prior)
- ▶ Roughness penalty (Basic MRF prior)
- ▶ Sparsity in ambient space
- ▶ Edge-preserving regularization
- ▶ Total-variation (TV) regularization
- ▶ Black-box denoiser like NLM

$$R(\mathbf{x}) = \beta \|\mathbf{x}\|_2^2$$



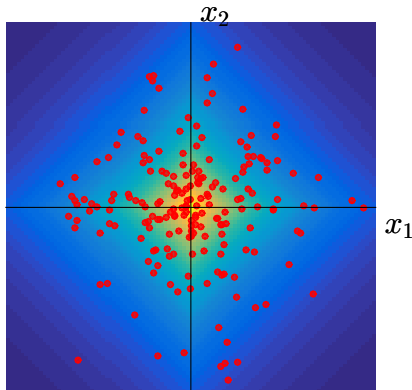
- ▶ Colors show equivalent (normalized) prior $p(\mathbf{x}) / p(\mathbf{0}) = e^{-R(\mathbf{x})}$
- ▶ Equivalent to IID gaussian prior on \mathbf{x}
- ▶ Makes any ill-conditioned / ill-posed problem well conditioned
- ▶ Ignores correlations between pixels

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

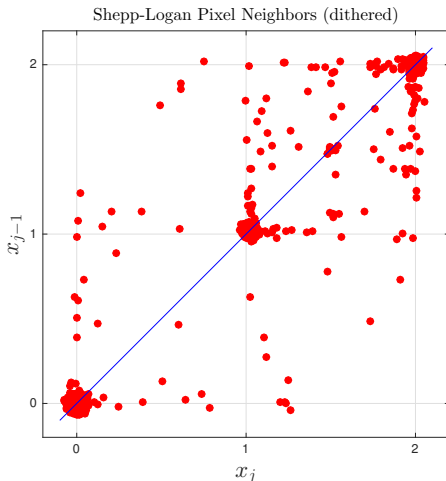
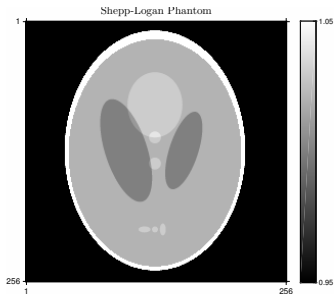


- ▶ Approximate Bayesian interpretation
- ▶ Non-convex
- ▶ IID \implies also ignores correlations

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



- ▶ Equivalent to IID Laplacian prior on \mathbf{x}
- ▶ Also ignores correlations

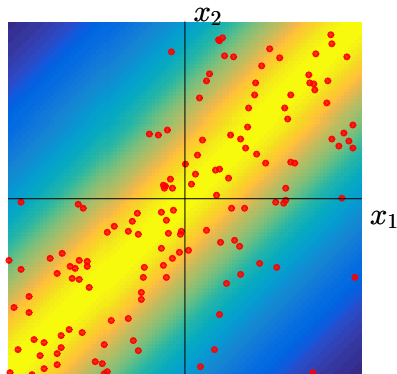
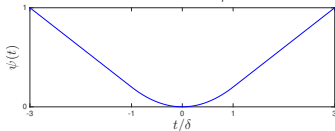


Caution: Shepp-Logan phantom [2] was designed for testing non-Bayesian methods, not for designing signal models. Q: What causes the spread??

Neighboring pixels tend to have similar values except near edges:

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

Potential function ψ :



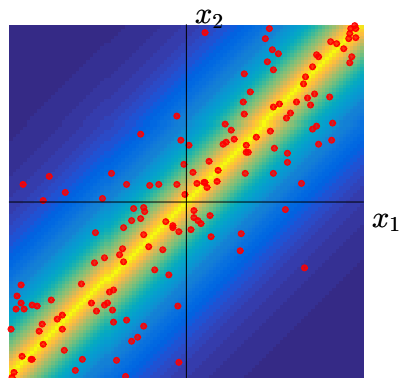
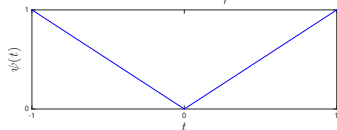
- Equivalent to improper prior (agnostic to DC value)
- Accounts for spatial correlations, but only very locally
- Used clinically now for low-dose X-ray CT image reconstruction

Total-variation (TV) regularization

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

$$\begin{aligned} R(\mathbf{x}) &= \beta \text{TV}(\mathbf{x}) = \beta \|\mathbf{C}\mathbf{x}\|_1 \\ &= \beta \sum_j |x_j - x_{j-1}| \end{aligned}$$

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...

Noisy image \rightarrow Denoiser \rightarrow Denoised image

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

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

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

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

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

Deep reconstruction using imaging physics

Summary

Backup

Wikipedia (2017-07-09)

https://en.wikipedia.org/wiki/Artificial_intelligence

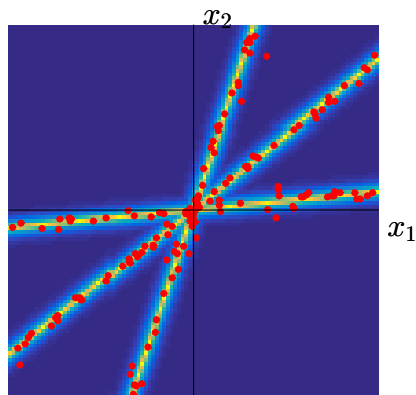
AI = "... any device that **perceives its environment** and takes actions that maximize its chance of success ..."

AI = trained ?

AI = adaptive !

Next:

- dictionary learning review
- new DL algorithm
- joint DL and image reconstruction ("blind compressed sensing")



- ▶ 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 $\mathbf{D}_1, \dots, \mathbf{D}_K$
(dictionaries with full column rank \implies tall).

Assume $\mathbf{x} \approx \mathbf{D}_k \mathbf{z}_k$ for some k and some (non-sparse) coefficients \mathbf{z}_k .

Natural regularizer for this model is:

$$\begin{aligned} R(\mathbf{x}) &= \underbrace{\min_k}_{\text{"classification"}} \underbrace{\min_{\mathbf{z}_k} \beta \frac{1}{2} \|\mathbf{x} - \mathbf{D}_k \mathbf{z}_k\|_2^2}_{\text{regression}} \\ &= \min_k \beta \frac{1}{2} \|\mathbf{x} - \mathbf{D}_k \mathbf{D}_k^+ \mathbf{x}\|_2^2. \end{aligned}$$

- ▶ $R(\mathbf{x}) = 0$ if \mathbf{x} lies in the span of any of the dictionaries $\{\mathbf{D}_k\}$.
- ▶ 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**.

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

- ▶ \mathbf{D} is a dictionary (often over-complete \implies wide)
- ▶ \mathbf{z} is a *sparse* coefficient vector (subset of columns of \mathbf{D}).

Corresponding regularizers:

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

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

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

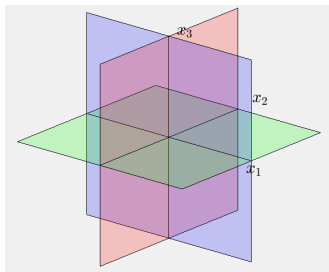
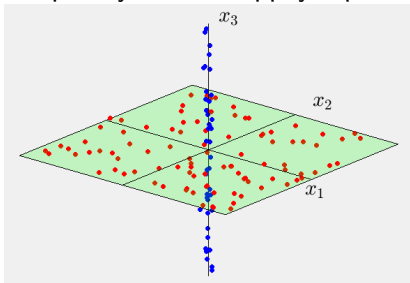
Union-of-subspaces vs sparse-coding-with-dictionary

Consider union-of-subspaces model with $D_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}$, $D_2 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$.

So D_1 spans x-y plane and D_2 spans z-axis.

A dictionary model with $D = [D_1 \ D_2] = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

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 D from data**, say $\mathbf{X} \in \mathbb{C}^{d \times N}$
- Write sparse representation as Sum of Outer Products (SOUP):

$$\mathbf{X} \approx \mathbf{DZ} = \mathbf{DC}' = \sum_{j=1}^J \mathbf{d}_j \mathbf{c}_j'$$

where $\mathbf{Z}' = \mathbf{C} = [\mathbf{c}_1 \dots \mathbf{c}_J] \in \mathbb{R}^{N \times J}$ (coefficients for each atom)

- Replace individual atom sparsity constraint $\|\mathbf{z}_n\|_0 \leq s$ of K-SVD with aggregate sparsity regularizer: $\|\mathbf{Z}\|_0 = \|\mathbf{C}\|_0$.
 - ▶ Natural for Dictionary Learning (DIL) from training data.
 - ▶ Unnatural for image compression using sparse coding.

SOUP-DIL ℓ_0 formulation:

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

SOUP-DIL formulation [9–11]:

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

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

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

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

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

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

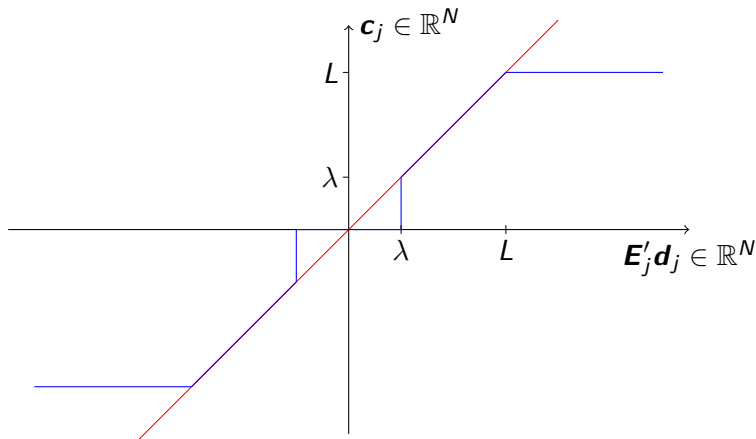
Truncated (via L) hard thresholding of $\mathbf{E}'_j \mathbf{d}_j$ with threshold λ .

- Dictionary atom step: update \mathbf{d}_j

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

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

Truncated hard thresholding for SOUP-DIL



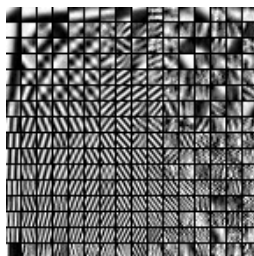
(Acts element-wise.) (In practice take $L = \infty$.)

(Algorithm also provides a simple sparse coding method.)

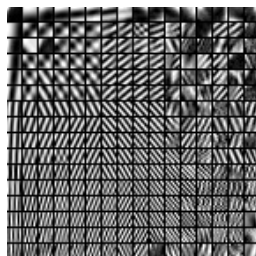
Example: dictionary learning for Barbara



Barbara



K-SVD D



SOUP-DIL 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

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

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

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

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

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

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

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

Deep reconstruction using imaging physics

Summary

Backup

Dictionary-blind MR image reconstruction:

$$\hat{\mathbf{x}} = \arg \min_{\mathbf{x}} \frac{1}{2} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 + \beta \mathbf{R}(\mathbf{x})$$
$$\mathbf{R}(\mathbf{x}) = \min_{\mathbf{D} \in \mathcal{D}} \min_{\mathbf{Z}' \in \mathcal{C}} \sum_{m=1}^M \left(\|\mathbf{P}_m \mathbf{x} - \mathbf{D}\mathbf{z}_m\|_2^2 + \lambda^2 \|\mathbf{z}_m\|_0 \right)$$

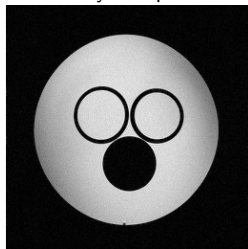
where \mathbf{P}_m extracts m th of M image patches.

In words: of the many images...

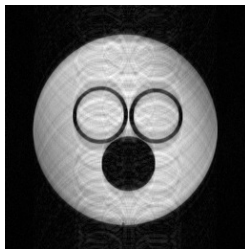
Alternating (nested) minimization:

- ▶ Fixing \mathbf{x} and \mathbf{D} , update each \mathbf{z}_j via hard-thresholding
- ▶ Fixing \mathbf{x} and \mathbf{Z} , update \mathbf{D} using SOUP-DIL
- ▶ Fixing \mathbf{Z} and \mathbf{D} , updating \mathbf{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].

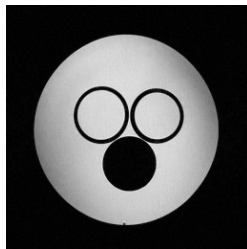
Fully Sampled



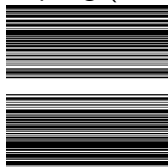
Zero-Filled



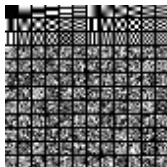
SOUP-DILLO-MRI



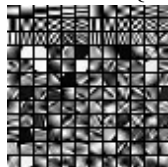
Sampling ($2.5\times$)



Initial D



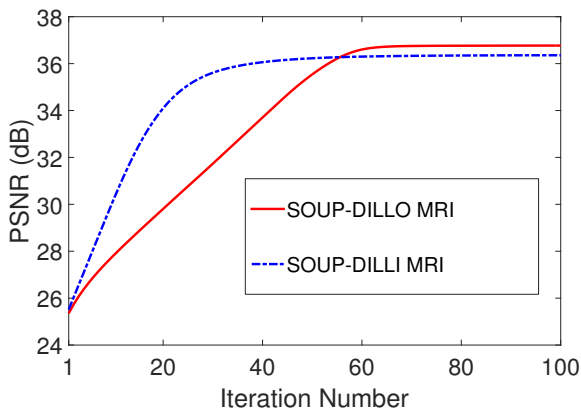
Learned real $\{D\}$



6×6 patches

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

[9]



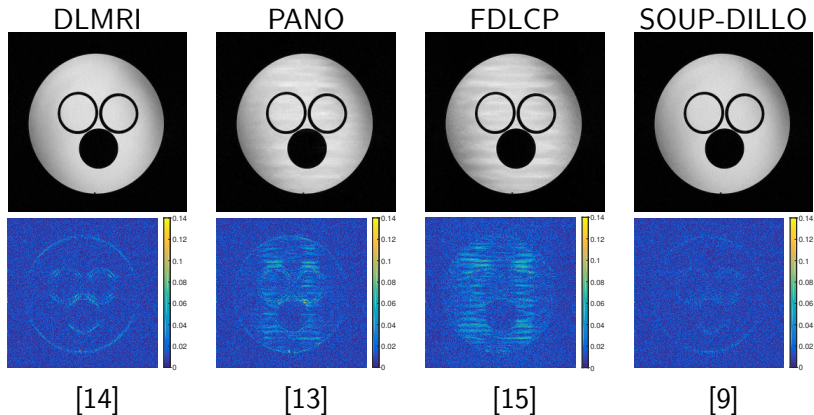
(SNR compared to fully sampled image.)

Using $\|\mathbf{z}_m\|_0$ leads to higher SNR than $\|\mathbf{z}_m\|_1$.

Adaptive case is non-convex anyway...



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]



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

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

Deep reconstruction using imaging physics

Summary

Backup

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

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

Recent extension [10]

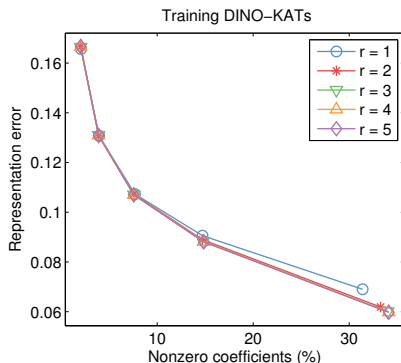
DIctioNary with lOw-rank AToms (DINO-KAT) model:

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

where $\text{reshape}(\mathbf{d}_j)$ reshapes dictionary atom \mathbf{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 $\|\mathbf{X} - \mathbf{DC}'\|_F / \|\mathbf{X}\|_F$ versus sparsity λ for several atom ranks r for $8 \times 8 \times 5$ space-time patches from (fully sampled) cardiac perfusion images. ▶

Dictionary with low-rank AToms (DINO-KAT) model:

$$D^* = \arg \min_{D \in \mathbb{R}^{d \times J}} \min_{C \in \mathbb{R}^{N \times J}} \|X - DC'\|_F^2 + \lambda^2 \|C\|_0 \quad \text{s.t.} \quad \begin{aligned} &\|d_j\|_2 = 1 \quad \forall j \\ &\|c_j\|_\infty \leq L \quad \forall j \\ &\text{rank}\{\text{reshape}(d_j)\} \leq r, \end{aligned}$$

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

- Sparse coding step for C uses same truncated hard thresholding
- Dictionary atom update step for d_j :

$$\arg \min_{d_j} \|E_j - d_j c_j'\|_F^2 \quad \text{s.t.} \quad \|d_j\|_2 = 1, \text{rank}\{\text{reshape}(d_j)\} \leq r$$

Simple solution: $\text{reshape}(d_j) = \frac{U_r \Sigma_r V_r'}{\|\Sigma_r\|_F}$

$U_r \Sigma_r V_r'$ is the rank- r truncated SVD of $\text{reshape}(E_j c_j)$.

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

Deep reconstruction using imaging physics

Summary

Backup

“dynamic” = changing over time = motion [16–19]

▶ Nuisance motions:

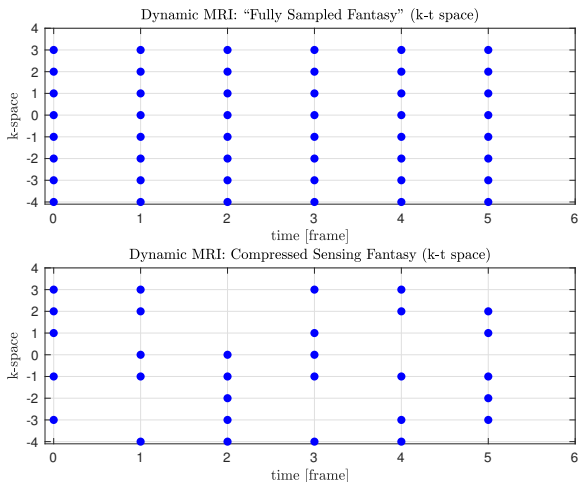
- Breathing
- Cardiac
- Peristalsis
- Tremors
- Kids ...

⇒ Faster scans (shorter time) can help reduce motion blur

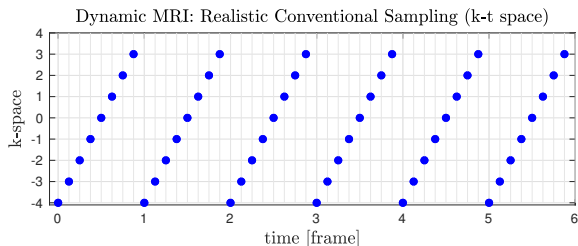
▶ Motions of interest (true “dynamic” scans):

- Vocalization (for speech studies)
- Cardiac (for function)
- Joint articulation (musculoskeletal scans)
- Contrast agent (blood flow / perfusion)
- Diffusion

⇒ Trade-offs between temporal resolution and spatial resolution

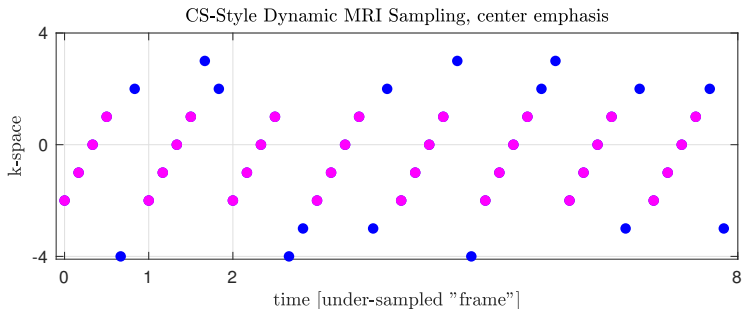


- ▶ Scan "twice as fast" !?
- ▶ Matrix completion problem!? \implies ... robust PCA (L+S) ...
[20, 21]



- ▶ 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



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

Deep reconstruction using imaging physics

Summary

Backup

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

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

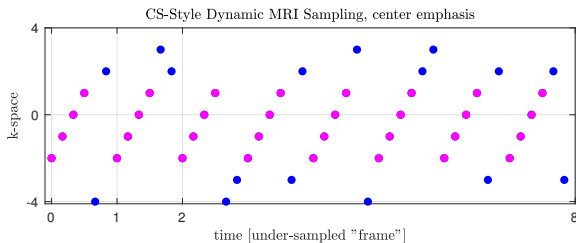
$$\mathbf{R}(\mathbf{x}) = \min_{\mathbf{D} \in \mathbb{C}^{d \times J}} \min_{\mathbf{Z} \in \mathbb{C}^{J \times M}} \sum_{m=1}^M \|\mathbf{P}_m \mathbf{x} - \mathbf{D} \mathbf{z}_m\|_2^2 + \lambda^2 \|\mathbf{z}_m\|_0$$

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

Block coordinate descent (BCD) algorithm (monotone descent)

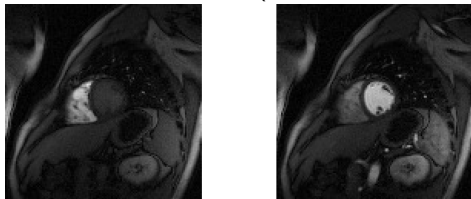
- Update coefficients \mathbf{Z} : sparse coding via hard thresholding
- Dictionary atom update of \mathbf{d}_j : uses residual, SVD
- Image update uses FFT (single coil Cartesian) or CG

- ▶ Latent signal vector $\mathbf{x} \in \mathbb{C}^{n_y n_x n_t}$ 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 $\mathbf{y} \in \mathbb{C}^{n_{\text{sample}} n_c}$ acquired using n_c coils.
- ▶ Sensing matrix \mathbf{A} includes:
 - coil sensitivity maps,
 - 2D or 3D spatial Fourier transform,
 - k -space sampling pattern.
- ▶ \mathbf{y} is undersampled, so regularization is required to estimate dynamic image sequence \mathbf{x} .



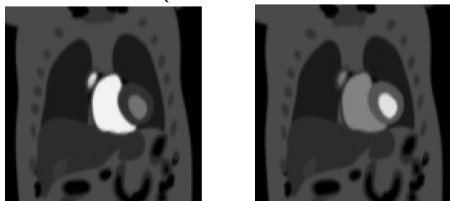
- ▶ **Low-rank and sparse (k-t SLR)** [22]:
 - Model: \mathbf{x} reshaped into an $N \times n_t$ space-time matrix, is both low-rank and (transform) sparse
- ▶ **Low-rank plus sparse ($\mathbf{L} + \mathbf{S}$)** [20, 21]
 - model: $\mathbf{x} = \mathbf{x}_L + \mathbf{x}_S$,
 - \mathbf{x}_L reshaped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

Cardiac perfusion data (ref. frames 7, 13)



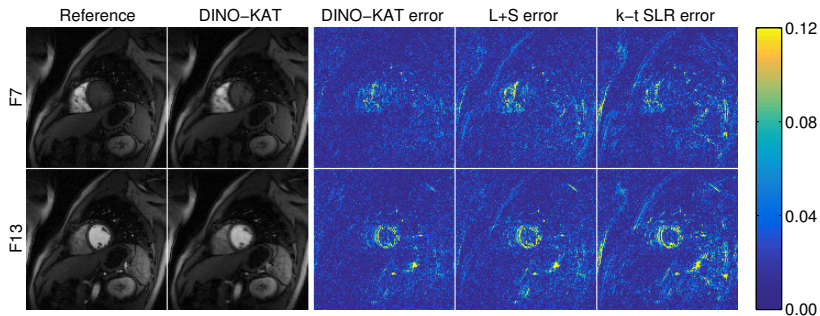
$128^2 \times 40$ fr.
 $3.2^2 \times 8$ mm³
12 coil
 $\Delta T = 307$ ms
Otazo et al. [21] (L+S)

PINCAT data (reference frames 16, 25)

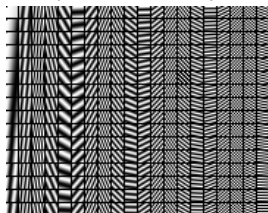


$128^2 \times 50$ fr.
 1.5 mm²
1 coil?
 $9 \times$ acc.
Lingala et al. [22] (k-t SLR)

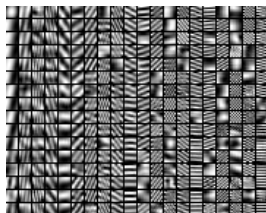
8× acceleration



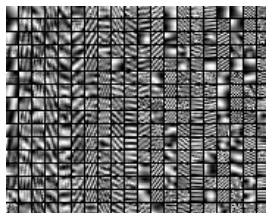
Initial atoms
(DCT matrix)



Real-part
of learned atoms

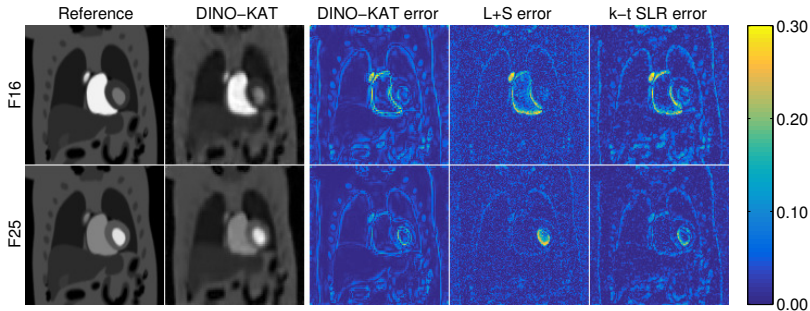


Imaginary-part
of learned atoms



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

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

8× acceleration

9× 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

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

Deep reconstruction using imaging physics

Summary

Backup

Recall image reconstruction with dictionary-based patch regularizer:

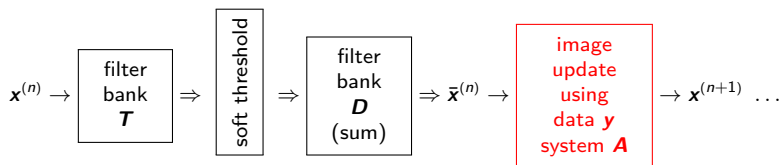
$$\hat{\mathbf{x}} = \arg \min_{\mathbf{x}} \min_{\mathbf{Z}} \frac{1}{2} \|\mathbf{y} - \mathbf{Ax}\|_2^2 + \beta \sum_{m=1}^M \left(\|\mathbf{P}_m \mathbf{x} - \mathbf{Dz}_m\|_2^2 + \lambda^2 \|\mathbf{z}_m\|_1 \right).$$

Alternating minimization algorithm:

- Apply dictionary atoms $\mathbf{T} = \mathbf{D}'$ to current image patches (convolution)
- Soft-threshold the resulting coefficients \mathbf{Z}
- Synthesize a “denoised” image: $\bar{\mathbf{x}}^{(n)} = \mathbf{DZ}$ (convolution then sum)
- Update image using system model:

$$\mathbf{x}^{(n+1)} = \arg \min_{\mathbf{x}} \|\mathbf{y} - \mathbf{Ax}\|_2^2 + \beta \|\mathbf{x} - \bar{\mathbf{x}}^{(n)}\|_2^2$$

One “iteration” (or layer):



System physics used as part of each update.

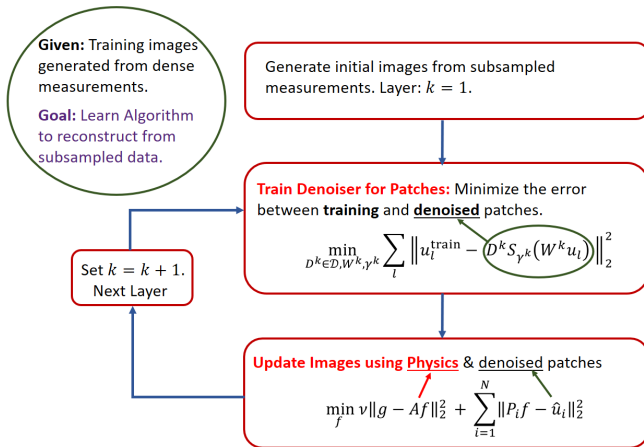
Learn from training data:

- transform T
- threshold (individual for each coefficient, *cf.* ReLU)
- dictionary D

Train each layer one by one using relatively simple updates.

No SGD or other tricks like “drop out” needed.

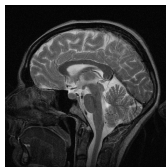
Works directly with complex MRI data.



- ▶ Multiple iterations or layers with trained model in each layer.
Training and application guided by physics of system.

- Trained with 4 of 32 complex 512^2 brain images from M. Lustig, UC Berkeley.
- Emulation of single coil MRI
- 10x under-sampling
- 8x8 patches, 256 atoms for D and T (4x over-complete)
- 256 soft-thresholds (one per filter)

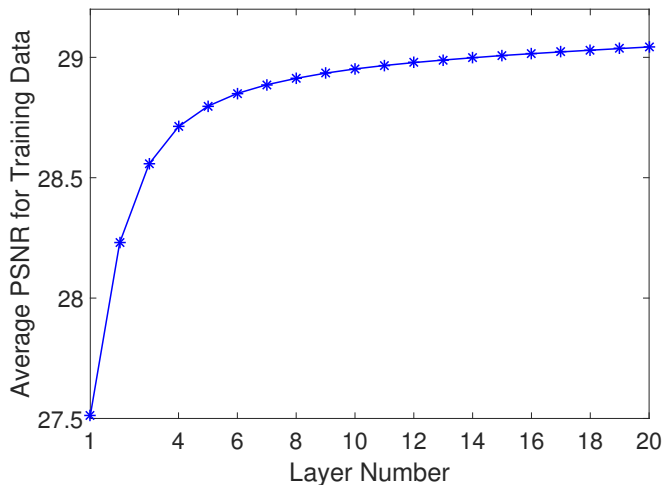
Image:



K-space Sampling:

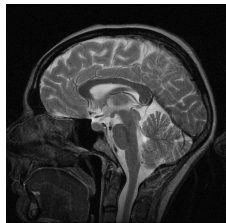


Effects of multiple layers during training



First layer similar to recent work: [24, 25].

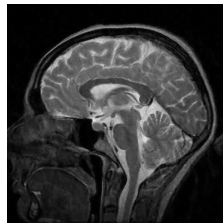
Physics-Driven Deep MRI Result: 10x Undersampling



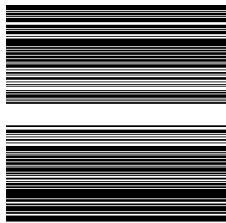
Reference



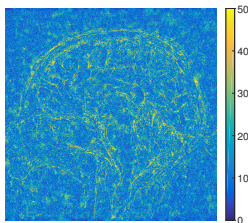
Wavelets + TV



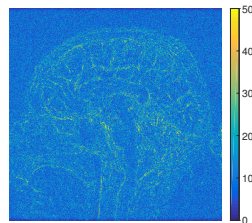
Trained Method



Sampling Mask

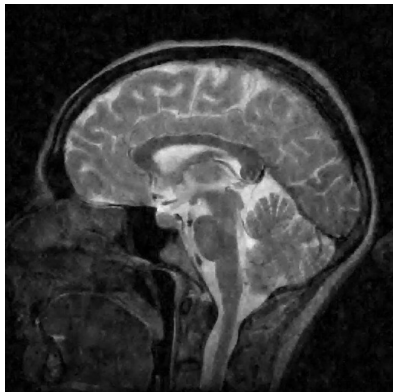


Error



Error

Non-adaptive vs adaptive: Zoom



Wavelets + TV



Trained

- ▶ 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 [26–29]
 - Adaptive methods versus trained (deep) methods?
 - Prospective use!
 - T-MI special issue on Machine-Learning for Image Reconstruction

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

Deep reconstruction using imaging physics

Summary

Backup

Consider the patch-based regularization term:

$$f(\mathbf{x}) = \sum_{m=1}^M \|\mathbf{P}_m \mathbf{x} - \mathbf{r}_m\|_2^2$$

where $\mathbf{r}_m = \mathbf{D}\mathbf{z}_m$ is a synthesized patch.

Using patch stride of one pixel and circulant boundary conditions:

$\sum_{m=1}^M \mathbf{P}'_m \mathbf{P}_m = N\mathbf{I}$, where N is number of pixels in patch. Thus

$$\begin{aligned} f(\mathbf{x}) &\equiv \sum_{m=1}^M \mathbf{x}' \mathbf{P}'_m \mathbf{P}_m \mathbf{x} - 2\mathbf{x}' \mathbf{P}'_m \mathbf{r}_m \equiv N \left(\mathbf{x}' \mathbf{x} - 2\mathbf{x}' \frac{1}{N} \sum_{m=1}^M \mathbf{P}'_m \mathbf{r}_m \right) \\ &\equiv N \|\mathbf{x} - \bar{\mathbf{x}}\|_2^2, \quad \bar{\mathbf{x}} = \frac{1}{N} \sum_{m=1}^M \mathbf{P}'_m \mathbf{r}_m, \end{aligned}$$

where $\bar{\mathbf{x}}$ denotes a patch combination image.

- [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, J. Ma, J. Wang, Y. Liu, H. Lu, and Z. Liang. “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. Szelam, 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, B. Moore, R. R. Nadakuditi, and J. A. Fessler. “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, Y. Hou, F. Lam, D. Guo, J. Zhong, and Z. Chen. “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, J-F. Cai, D. Guo, Y. Liu, Z. Chen, and X. Qu. “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, A. Koenig, S. Roux, P. Hugonnard, R. Guillemaud, and P. Grangeat. “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, X. Li, Y. Ma, and J. Wright. “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, Y. Hu, E. DiBella, and M. Jacob. “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, B. E. Moore, R. R. Nadakuditi, and J. A. Fessler. “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. Wang, Z. Su, L. Ying, X. Peng, S. Zhu, F. Liang, D. Feng, and D. Liang. “Accelerating magnetic resonance imaging via deep learning.” In: *Proc. IEEE Intl. Symp. Biomed. Imag.* 2016, 514–7.
- [25] S. Wang, Z. Su, L. Ying, X. Peng, and D. Liang. “Exploiting deep convolutional neural network for fast magnetic resonance imaging.” In: *Proc. Intl. Soc. Mag. Res. Med.* 2016, p. 1778.
- [26] S. Ravishankar, B. Wen, and Y. Bresler. “Online sparsifying transform learning - Part I: algorithms.” In: *ieee-jstsp* 9.4 (June 2015), 625–36.
- [27] S. Ravishankar and Y. Bresler. “Online sparsifying transform learning - Part II: convergence analysis.” In: *ieee-jstsp* 9.4 (June 2015), 637–46.

- [28] S. Ravishankar, B. E. Moore, R. R. Nadakuditi, and J. A. Fessler. “Efficient online dictionary adaptation and image reconstruction for dynamic MRI.” In: *Proc., IEEE Asilomar Conf. on Signals, Systems, and Comp.* To appear. 2017.
- [29] 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.