Sparse optimization and machine learning in image reconstruction

Aleksandra Pižurica

Group for Artificial Intelligence and Sparse Modelling (GAIM)

Department Telecommunications and Information Processing

Ghent University

2nd Annual Meeting
B-Q MINDED
Siemens Healthineers, Belgium, January 29-30, 2020

Outline

1. Model-based iterative reconstruction algorithms
   - Sparse optimization
   - Solution strategies: greedy methods vs. convex optimization
   - Optimization methods in sparse image reconstruction
2. Structured sparsity
   - Wavelet-tree sparsity
   - Markov Random Field (MRF) priors
3. Machine learning in image reconstruction
   - Main ideas and current trends

A fairly general formulation

Reconstruct a signal (image) $\mathbf{x} \in X$ from data $\mathbf{y} \in Y$ where

$$\mathbf{y} = \mathcal{T}(\mathbf{x}) + \mathbf{n}$$

X and Y are Hilbert spaces, $\mathcal{T} : X \mapsto Y$ is the forward operator and $\mathbf{n}$ is noise.
A common model-driven approach is to minimize the negative log-likelihood $\mathcal{L}$:

$$\min_{x \in X} \mathcal{L}(\mathcal{T}(\mathbf{x}), \mathbf{y})$$

Typically, ill-posed and leads to over-fitting. Variational regularization, also called model-based iterative reconstruction seeks to minimize a regularized objective function

$$\min_{x \in X} \mathcal{L}(\mathcal{T}(\mathbf{x}), \mathbf{y}) + \tau \phi(\mathbf{x})$$

$\phi : X \mapsto \mathbb{R} \cup \{-\infty, \infty\}$ is a regularization functional. $\tau \geq 0$ governs the influence of the a priori knowledge against the need to fit the data.

Linear inverse problems

Many image reconstruction problems can be formulated as a linear inverse problem. A noisy indirect observation **y**, of the original image **x** is then

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{n}$$

Matrix **A** is the forward operator. $\mathbf{x} \in \mathbb{R}^n$; $\mathbf{y}, \mathbf{n} \in \mathbb{R}^m$ (or $\mathbf{x} \in \mathbb{C}^n$; $\mathbf{y}, \mathbf{n} \in \mathbb{C}^m$). Here, image pixels are stacked into vectors (raster scanning). In general, $m \neq n$.

Some examples

- CT: **A** is the system matrix modeling the X-ray transformation
- MRI: **A** is (partially sampled) Fourier operator
- OCT: **A** is the first Born approximation for the scattering
- Compressed sensing: **A** is a measurement matrix (dense or sparse)

Linear inverse problems

For the linear inverse problem $\mathbf{y} = \mathbf{Ax} + \mathbf{n}$, model-based reconstruction seeks to solve:

$$\min_{\mathbf{x}} \frac{1}{2}\|\mathbf{Ax} - \mathbf{y}\|_2^2 + \tau\phi(\mathbf{x}) \qquad \text{(Tikhonov formulation)}$$

Alternatively,

$$\min_{\mathbf{x}} \phi(\mathbf{x}) \quad \text{subject to} \quad \|\mathbf{Ax} - \mathbf{y}\|_2^2 \leq \epsilon \qquad \text{(Morozov formulation)}$$

$$\min_{\mathbf{x}} \|\mathbf{Ax} - \mathbf{y}\|_2^2 \quad \text{subject to} \quad \phi(\mathbf{x}) \leq \delta \qquad \text{(Ivanov formulation)}$$

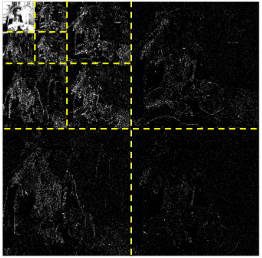
Under mild conditions, these are all **equivalent** [Figueiredo and Wright, 2013], and which one is more convenient is problem-dependent.

Outline

1. Model-based iterative reconstruction algorithms
   - Sparse optimization
   - Solution strategies: greedy methods vs. convex optimization
   - Optimization methods in sparse image reconstruction
2. Structured sparsity
   - Wavelet-tree sparsity
   - Markov Random Field (MRF) priors
3. Machine learning in image reconstruction
   - Main ideas and current trends

Sparse optimization

- A common assumption: $\mathbf{x}$ is sparse in a well-chosen transform domain.
- We refer to a **wavelet** representation meaning any wavelet-like multiscale representation, including curvelets and shearlets..



$$\mathbf{x} = \mathbf{\Psi}\boldsymbol{\theta}, \ \boldsymbol{\theta} \in \mathbb{R}^d, \ \mathbf{\Psi} \in \mathbb{R}^{n \times d}$$

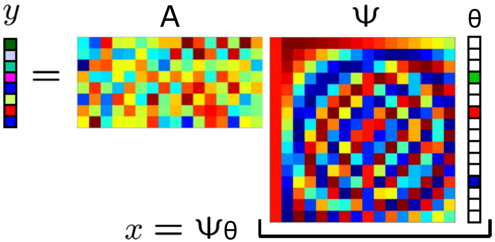
The columns of $\mathbf{\Psi}$ are the elements of a **wavelet frame** (an orthogonal basis or an overcomplete dictionary)

- The main results hold for **learned dictionaries**, trained on a set of representative examples to yield optimally sparse representation for a particular class of images.

Compressed sensing

Consider $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{n}$, with $\mathbf{y}, \mathbf{n} \in \mathbb{R}^m$, $\mathbf{x} = \boldsymbol{\Psi}\boldsymbol{\theta}$, $\mathbf{x} \in \mathbb{R}^n$, $\boldsymbol{\theta} \in \mathbb{R}^d$, $m < n$

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \frac{1}{2}\|\mathbf{A}\boldsymbol{\Psi}\boldsymbol{\theta} - \mathbf{y}\|_2^2 + \tau\phi(\boldsymbol{\theta}), \qquad \hat{\mathbf{x}} = \boldsymbol{\Psi}\hat{\boldsymbol{\theta}}$$



Commonly: $\min_{\boldsymbol{\theta}} \|\boldsymbol{\theta}\|_0$ s.t. $\|\mathbf{A}\boldsymbol{\Psi}\boldsymbol{\theta} - \mathbf{y}\|_2^2 \leq \epsilon$ or $\min_{\boldsymbol{\theta}} \frac{1}{2}\|\mathbf{A}\boldsymbol{\Psi}\boldsymbol{\theta} - \mathbf{y}\|_2^2 + \tau\|\boldsymbol{\theta}\|_1$

[Candès et al., 2006], [Donoho, 2006], [Lustig et al., 2007]

Compressed sensing: recovery guarantees

Consider $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{n}$, with $\mathbf{y}, \mathbf{n} \in \mathbb{R}^m$, $\mathbf{x} = \mathbf{\Psi}\theta$, $\mathbf{x} \in \mathbb{R}^n$, $m < n$

Matrix $\mathbf{\Phi} = \mathbf{A}\mathbf{\Psi}$ has *K-restricted isometry property* (**K-RIP**) with constant $\epsilon_K < 1$ if $\forall$ K-sparse (having only K non-zero entries) θ:

$$(1 - \epsilon_K)\|\theta\|_2^2 \leq \|\mathbf{\Phi}\theta\|_2^2 \leq (1 + \epsilon_K)\|\theta\|_2^2$$

Suppose matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$ is formed by subsampling a given sampling operator $\bar{\mathbf{A}} \in \mathbb{R}^{n \times n}$. The **mutual coherence** between $\bar{\mathbf{A}}$ and $\mathbf{\Psi}$:

$$\mu(\bar{\mathbf{A}}, \mathbf{\Psi}) = \max_{i,j} |a_i^\top \psi_j|$$

If $m > C\mu^2(\bar{\mathbf{A}}, \mathbf{\Psi})Kn\log(n)$, for some constant $C > 0$, then

$$\min_{\theta} \frac{1}{2}\|\mathbf{A}\mathbf{\Psi}\theta - \mathbf{y}\|_2^2 + \tau\|\theta\|_1$$

recovers $\mathbf{x}$ with high probability, given the K-RIP holds for $\mathbf{\Phi} = \mathbf{A}\mathbf{\Psi}$.

Analysis vs. synthesis formulation

Consider $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{n}$, with $\mathbf{y} \in \mathbb{R}^m$, $\mathbf{x} = \mathbf{\Psi}\boldsymbol{\theta}$, $\mathbf{x} \in \mathbb{R}^n$, $\boldsymbol{\theta} \in \mathbb{R}^d$

Synthesis approach:

$$\min_{\boldsymbol{\theta}} \frac{1}{2}\|\mathbf{A}\mathbf{\Psi}\boldsymbol{\theta} - \mathbf{y}\|_2^2 + \tau\phi(\boldsymbol{\theta})$$

Analysis approach:

$$\min_{\mathbf{x}} \frac{1}{2}\|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2^2 + \tau\phi(\mathbf{x})$$

Analysis vs. synthesis formulation

Consider $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{n}$, with $\mathbf{y} \in \mathbb{R}^m$, $\mathbf{x} = \boldsymbol{\Psi}\boldsymbol{\theta}$, $\mathbf{x} \in \mathbb{R}^n$, $\boldsymbol{\theta} \in \mathbb{R}^d$

Synthesis approach:

$$\min_{\boldsymbol{\theta}} \frac{1}{2}\|\mathbf{A}\boldsymbol{\Psi}\boldsymbol{\theta} - \mathbf{y}\|_2^2 + \tau\phi(\boldsymbol{\theta})$$

or in a constrained form:

$$\min_{\boldsymbol{\theta}} \phi(\boldsymbol{\theta}) \quad \text{subject to} \quad \|\mathbf{A}\boldsymbol{\Psi}\boldsymbol{\theta} - y\|_2^2 \leq \epsilon$$

Analysis approach:

$$\min_{\mathbf{x}} \frac{1}{2}\|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2^2 + \tau\phi(\mathbf{x})$$

or in a constrained form:

$$\min_{\mathbf{x}} \phi(\mathbf{x}) \quad \text{subject to} \quad \|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2^2 \leq \epsilon$$

Analysis vs. synthesis formulation

Consider $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{n}$, with $\mathbf{y} \in \mathbb{R}^m$, $\mathbf{x} = \boldsymbol{\Psi}\boldsymbol{\theta}$, $\mathbf{x} \in \mathbb{R}^n$, $\boldsymbol{\theta} \in \mathbb{R}^d$

Synthesis approach:

$$\min_{\boldsymbol{\theta}} \frac{1}{2}\|\mathbf{A}\boldsymbol{\Psi}\boldsymbol{\theta} - \mathbf{y}\|_2^2 + \tau\phi(\boldsymbol{\theta})$$

or in a constrained form:

$$\min_{\boldsymbol{\theta}} \phi(\boldsymbol{\theta}) \quad \text{subject to} \quad \|\mathbf{A}\boldsymbol{\Psi}\boldsymbol{\theta} - y\|_2^2 \leq \epsilon$$

Analysis approach that also applies to wavelet regularization

$$\min_{\mathbf{x}} \frac{1}{2}\|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2^2 + \tau\phi(\mathbf{P}\mathbf{x})$$

or in a constrained form:

$$\min_{\mathbf{x}} \phi(\mathbf{P}\mathbf{x}) \quad \text{subject to} \quad \|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2^2 \leq \epsilon$$

Analysis vs. synthesis formulation

Consider $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{n}$, with $\mathbf{y} \in \mathbb{R}^m$, $\mathbf{x} = \mathbf{\Psi}\boldsymbol{\theta}$, $\mathbf{x} \in \mathbb{R}^n$, $\boldsymbol{\theta} \in \mathbb{R}^d$

Synthesis approach:

$$\min_{\boldsymbol{\theta}} \frac{1}{2}\|\mathbf{A}\mathbf{\Psi}\boldsymbol{\theta} - \mathbf{y}\|_2^2 + \tau\phi(\boldsymbol{\theta})$$

or in a constrained form:

$$\min_{\boldsymbol{\theta}} \phi(\boldsymbol{\theta}) \quad \text{subject to} \quad \|\mathbf{A}\mathbf{\Psi}\boldsymbol{\theta} - y\|_2^2 \leq \epsilon$$

Analysis approach that also applies to wavelet regularization

$$\min_{\mathbf{x}} \frac{1}{2}\|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2^2 + \tau\phi(\mathbf{P}\mathbf{x})$$

or in a constrained form:

$$\min_{\mathbf{x}} \phi(\mathbf{P}\mathbf{x}) \quad \text{subject to} \quad \|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2^2 \leq \epsilon$$

$\mathbf{P}$: a wavelet transform operator or $\mathbf{P} = \mathbf{I}$ (standard analysis)

Outline

1. Model-based iterative reconstruction algorithms
   - Sparse optimization
   - Solution strategies: greedy methods vs. convex optimization
   - Optimization methods in sparse image reconstruction
2. Structured sparsity
   - Wavelet-tree sparsity
   - Markov Random Field (MRF) priors
3. Machine learning in image reconstruction
   - Main ideas and current trends

Solution strategies: greedy methods vs. convex optimization

Solution strategy is problem-dependent. For non-convex problems like

$$\min_{x} \|\mathbf{x}\|_0 \quad \text{subject to} \quad \|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2^2 \leq \epsilon$$

Greedy algorithms, e.g.,

- **Matching Pursuit** (**MP**) [Mallat and Zhang, 1993]
- **OMP**[Tropp, 2004], **CoSaMP** [Needell and Tropp, 2009]
- **IHT** [Blumensath and Davies, 2009]

or **convex relaxation** can be applied leading to:

$$\min_{x} \|\mathbf{x}\|_1 \quad \text{subject to} \quad \|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2^2 \leq \epsilon$$

$$\min_{x} \frac{1}{2}\|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2^2 + \tau\phi(\mathbf{x})$$

known as **LASSO** [Tibshirani, 1996] or **BPDN** [Chen et al., 2001] problem.

Greedy methods: OMP

OMP algorithm for solving $\min_{x} \|\mathbf{x}\|_0$ subject to $\mathbf{Ax} = \mathbf{y}$

Require: $k = 1, \mathbf{r}^{(1)} = \mathbf{y}, \Lambda^{(0)} = \emptyset$

1: **repeat**
2: $\quad \lambda^{(k)} = \arg\max_j |\mathbf{A}_j \cdot \mathbf{r}^{(k)}|$
3: $\quad \Lambda^{(k)} = \Lambda^{(k-1)} \cup \{\lambda^{(k)}\}$
4: $\quad \mathbf{x}^{(k)} = \arg\min_x \|\mathbf{A}_{\Lambda_k}\mathbf{x} - \mathbf{y}\|_2$
5: $\quad \hat{\mathbf{y}}^{(k)} = \mathbf{A}_{\Lambda_k}\mathbf{x}^{(k)}$
6: $\quad \mathbf{r}^{(k+1)} = \mathbf{r}^{(k)} - \hat{\mathbf{y}}^{(k)}$
7: $\quad k = k + 1$
8: **until** stopping criterion satisfied

$\mathbf{A}_j$ is the j-th column of $\mathbf{A}$, and $\mathbf{A}_\Lambda$ a sub-matrix of $\mathbf{A}$ with columns indicated in Λ.

Greedy methods: OMP

OMP algorithm for solving $\min_{\mathbf{x}} \|\mathbf{x}\|_0$ subject to $\mathbf{Ax} = \mathbf{y}$

Require: $k = 1, \mathbf{r}^{(1)} = \mathbf{y}, \Lambda^{(0)} = \emptyset$

1: **repeat**

2: $\lambda^{(k)} = \arg\max_j |\mathbf{A}_j \cdot \mathbf{r}^{(k)}|$ — identify the 'important' column of $\mathbf{A}$

3: $\Lambda^{(k)} = \Lambda^{(k-1)} \cup \{\lambda^{(k)}\}$ — augment the index set

4: $\mathbf{x}^{(k)} = \arg\min_{\mathbf{x}} \|\mathbf{A}_{\Lambda_k}\mathbf{x} - \mathbf{y}\|_2$ — solve the least square problem

5: $\hat{\mathbf{y}}^{(k)} = \mathbf{A}_{\Lambda_k}\mathbf{x}^{(k)}$ — express the portion of $\mathbf{y}$ being explained by $\mathbf{A}_{\Lambda_k}\mathbf{x}^{(k)}$

6: $\mathbf{r}^{(k+1)} = \mathbf{r}^{(k)} - \hat{\mathbf{y}}^{(k)}$ — update the residual by removing the explained portion of $\mathbf{y}$

7: $k = k + 1$

8: **until** stopping criterion satisfied

$\mathbf{A}_j$ is the j-th column of $\mathbf{A}$, and $\mathbf{A}_\Lambda$ a sub-matrix of $\mathbf{A}$ with columns indicated in Λ.
'Important column' = that with max absolute value of correlation with the residual

Outline

1. Model-based iterative reconstruction algorithms
   - Sparse optimization
   - Solution strategies: greedy methods vs. convex optimization
   - Optimization methods in sparse image reconstruction
2. Structured sparsity
   - Wavelet-tree sparsity
   - Markov Random Field (MRF) priors
3. Machine learning in image reconstruction
   - Main ideas and current trends

Proximal operator

Many state-of-the-art image reconstruction algorithms solve problems of the kind

$$\min_{x} \frac{1}{2}\|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2^2 + \tau\phi(\mathbf{P}\mathbf{x})$$

making use of the proximity operator i.e., the **Moreau proximal mapping** [Combettes and Wajs, 2005]

$$\mathrm{prox}_{\tau\phi}(\mathbf{y}) = \mathrm{argmin}_{x} \frac{1}{2}\|\mathbf{x} - \mathbf{y}\|_2^2 + \tau\phi(\mathbf{x})$$

For certain choices of $\phi(\mathbf{x})$, this operator has a closed-form, e.g.,

- $\phi(\mathbf{x}) = \|\mathbf{x}\|_1 \rightarrow \mathrm{prox}_{\tau\ell_1}(\mathbf{y}) = \mathrm{soft}(\mathbf{y}, \tau)$ **component-wise** soft thresholding
- $\phi(\mathbf{x}) = \|\mathbf{x}\|_0 \rightarrow \mathrm{prox}_{\tau\ell_0}(\mathbf{y}) = \mathrm{hard}(\mathbf{y}, \sqrt{2\tau})$ **component-wise** hard thresholding

Another common regularization function is total variation (TV):

- $\phi(\mathbf{x}) = \|\mathbf{x}\|_{TV} \rightarrow \mathrm{prox}_{\tau TV}(\mathbf{y})$ **Chambolle's algorithm** [Chambolle, 2004]

Iterative shrinkage/thresholding (IST)

The standard algorithm for solving

$$\min_{\mathbf{x}} \frac{1}{2}\|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2^2 + \tau\phi(\mathbf{x})$$

is **iterative shrinkage/thresholding (IST)** algorithm [Figueiredo and Nowak, 2003], [Daubechies et al., 2004]:

$$\mathbf{x}^{k+1} = \mathrm{prox}_{\tau\phi}\Big(\mathbf{x}^k - \frac{1}{\gamma} \underbrace{\mathbf{A}^H(\mathbf{A}\mathbf{x}^k - \mathbf{y})}_{\text{gradient of the data fidelity term}}\Big)$$

Its key ingredient is the **proximity operator** $\mathrm{prox}_{\tau\phi}(\mathbf{y}) = \operatorname{argmin}_{\mathbf{x}} \frac{1}{2}\|\mathbf{x} - \mathbf{y}\|_2^2 + \tau\phi(\mathbf{x})$

A general approach in [Daubechies et al., 2004]: $\phi(\mathbf{x})$ is a weighted ℓ_p norm of the coefficients of $\mathbf{x}$ with respect to a wavelet basis.

Iterative shrinkage/thresholding (IST) and extensions

The standard algorithm for solving

$$\min_{x} \frac{1}{2}\|\mathbf{Ax} - \mathbf{y}\|_2^2 + \tau\phi(\mathbf{x})$$

is **iterative shrinkage/thresholding (IST)** algorithm [Figueiredo and Nowak, 2003], [Daubechies et al., 2004]:

$$\mathbf{x}^{k+1} = \mathrm{prox}_{\tau\phi}\Big(\mathbf{x}^k - \frac{1}{\gamma} \underbrace{\mathbf{A}^H(\mathbf{Ax}^k - \mathbf{y})}_{\text{gradient of the data fidelity term}}\Big)$$

Different accelerated versions:

- **TwIST** [Bioucas-Dias and Figueiredo, 2007]
- **FISTA** [Beck and Teboulle, 2009]
- **SpaRSA** [Wright et al., 2009]

Variable splitting

A very old idea (back to at least [Courant, 1943]): Represent $\min_{\mathbf{x}} f_1(\mathbf{x}) + f_2(\mathbf{Gx})$ as

$$\min_{\mathbf{x}} f_1(\mathbf{x}) + f_2(\mathbf{z}) \quad \text{subject to} \quad \mathbf{Gx} = \mathbf{z}$$

The rationale: it may be easier to solve the constrained problem.

Variable splitting (**VS**) together with the augmented Lagrangian method (**ALM**) and non linear block Gauss-Seidel (**NLBGS**) leads to a form of Alternating Direction Method of Multipliers (**ADMM**). It is this interpretation:

$$(\mathbf{VS} + \mathbf{ALM} + \mathbf{NLBGS}) \rightarrow \mathbf{ADMM}$$

that we give in the next few slides, following [Afonso et al., 2010]

Variable splitting and Augmented Lagrangian Method

A very old idea (back to at least [Courant, 1943]): Represent $\min_{\mathbf{x}} f_1(\mathbf{x}) + f_2(\mathbf{G}\mathbf{x})$ as

$$\min_{\mathbf{x}} f_1(\mathbf{x}) + f_2(\mathbf{z}) \quad \text{subject to} \quad \mathbf{G}\mathbf{x} = \mathbf{z}$$

$$\mathrm{L}_\mu(\mathbf{x}, \mathbf{z}, \boldsymbol{\lambda}) = \underbrace{f_1(\mathbf{x}) + f_2(\mathbf{z}) + \boldsymbol{\lambda}^T(\mathbf{G}\mathbf{x} - \mathbf{z})}_{\text{Lagrangian}} + \underbrace{\frac{\mu}{2}\|\mathbf{G}\mathbf{x} - \mathbf{z}\|_2^2}_{\text{“augmentation”}}$$

Basic augmented Lagrangian method (**ALM**), a.k.a., method of multipliers (**MM**),:

$$\begin{aligned}
(\mathbf{x}^{(k)}, \mathbf{z}^{(k)}) &= \arg\min_{(\mathbf{x},\mathbf{z})} \mathrm{L}_\mu(\mathbf{x}, \mathbf{z}, \boldsymbol{\lambda}^{(k-1)}) \\
\boldsymbol{\lambda}^{(k)} &= \boldsymbol{\lambda}^{(k-1)} + \mu(\mathbf{G}\mathbf{x}^{(k)} - \mathbf{z}^{(k)})
\end{aligned}$$

Goes back to at least [Hestenes, 1969], [Powell, 1969]

Variable splitting and Augmented Lagrangian Method

A very old idea (back to at least [Courant, 1943]): Represent $\min_x f_1(\mathbf{x}) + f_2(\mathbf{Gx})$ as

$$\min_x f_1(\mathbf{x}) + f_2(\mathbf{z}) \quad \text{subject to} \quad \mathbf{Gx} = \mathbf{z}$$

$$\mathrm{L}_\mu(\mathbf{x}, \mathbf{z}, \boldsymbol{\lambda}) = \underbrace{f_1(\mathbf{x}) + f_2(\mathbf{z}) + \boldsymbol{\lambda}^T(\mathbf{Gx} - \mathbf{z})}_{\text{Lagrangian}} + \underbrace{\frac{\mu}{2}\|\mathbf{Gx} - \mathbf{z}\|_2^2}_{\text{"augmentation"}}$$

After simple "complete-the-squares" **ALM**/**MM** yields [Afonso et al., 2010]:

$$\begin{aligned}(\mathbf{x}^{(k)}, \mathbf{z}^{(k)}) &= \underset{(x,z)}{\operatorname{argmin}}\, f_1(\mathbf{x}) + f_2(\mathbf{z}) + \frac{\mu}{2}\|\mathbf{Gx} - \mathbf{z} - \mathbf{d}^{(k-1)}\|_2^2 \\ \mathbf{d}^{(k)} &= \mathbf{d}^{(k-1)} - (\mathbf{Gx}^{(k)} - \mathbf{z}^{(k)})\end{aligned}$$

ADMM as Variable splitting and ALM

Use variable splitting (**VS**) to represent $\min_x f_1(\mathbf{x}) + f_2(\mathbf{Gx})$ as

$$\min_x f_1(\mathbf{x}) + f_2(\mathbf{z}) \quad \text{subject to} \quad \mathbf{Gx} = \mathbf{z}$$

ALM/**MM** yields :

$$\begin{aligned} (\mathbf{x}^{(k)}, \mathbf{z}^{(k)}) &= \operatorname*{argmin}_{(x,z)} f_1(\mathbf{x}) + f_2(\mathbf{z}) + \frac{\mu}{2}\|\mathbf{Gx} - \mathbf{z} - \mathbf{d}^{(k-1)}\|_2^2 \qquad (P) \\ \mathbf{d}^{(k)} &= \mathbf{d}^{(k-1)} - (\mathbf{Gx}^{(k)} - \mathbf{z}^{(k)}) \end{aligned}$$

Solve (P) with one step of **NLBGS** → "scaled" **ADMM** version [Boyd et al., 2011]:

$$\begin{aligned} \mathbf{x}^{(k)} &= \operatorname*{argmin}_{x} f_1(\mathbf{x}) + \frac{\mu}{2}\|\mathbf{Gx} - \mathbf{z}^{(k-1)} - \mathbf{d}^{(k-1)}\|_2^2 \\ \mathbf{z}^{(k)} &= \operatorname*{argmin}_{z} f_2(\mathbf{z}) + \frac{\mu}{2}\|\mathbf{Gx}^{(k-1)} - \mathbf{z} - \mathbf{d}^{(k-1)}\|_2^2 \\ \mathbf{d}^{(k)} &= \mathbf{d}^{(k-1)} - (\mathbf{Gx}^{(k)} - \mathbf{z}^{(k)}) \end{aligned}$$

ADMM algorithm

ADMM algorithm for solving: $\min_{\mathbf{x}} f_1(\mathbf{x}) + f_2(\mathbf{G}\mathbf{x})$

Require: $k = 0, \mu > 0, \mathbf{z}^{\{0\}}, \mathbf{d}^{\{0\}}$

1: **repeat**

2: $\mathbf{x}^{(k)} = \underset{\mathbf{x}}{\mathrm{argmin}}\, f_1(\mathbf{x}) + \frac{\mu}{2}\|\mathbf{G}\mathbf{x} - \mathbf{z}^{(k-1)} - \mathbf{d}^{(k-1)}\|_2^2$

3: $\mathbf{z}^{(k)} = \underset{\mathbf{z}}{\mathrm{argmin}}\, f_2(\mathbf{z}) + \frac{\mu}{2}\|\mathbf{G}\mathbf{x}^{(k-1)} - \mathbf{z} - \mathbf{d}^{(k-1)}\|_2^2$

4: $\mathbf{d}^{(k)} = \mathbf{d}^{(k-1)} - (\mathbf{G}\mathbf{x}^{(k)} - \mathbf{z}^{(k)})$

5: $k = k + 1$

6: **until** stopping criterion is satisfied

Equivalent to **split-Bregman** method [Goldstein and Osher, 2009].
Connections with Douglas-Raschford splitting [Setzer, 2009].

Instantiate ADMM to our linear inverse problem: $\min_x \|\mathbf{Ax} - \mathbf{y}\|_2^2 + \tau\phi(\mathbf{Px})$

Require: $k = 0, \mu > 0, \mathbf{z}^{\{0\}}, \mathbf{d}^{\{0\}}$

1: **repeat**

2: $\mathbf{x}^{(k)} = \underset{x}{\operatorname{argmin}} \|\mathbf{Ax} - \mathbf{y}\|_2^2 + \frac{\mu}{2}\|\mathbf{Px} - \mathbf{z}^{(k-1)} - \mathbf{d}^{(k-1)}\|_2^2$

3: $\mathbf{z}^{(k)} = \underset{z}{\operatorname{argmin}}\, \tau\phi(\mathbf{z}) + \frac{\mu}{2}\|\mathbf{Px}^{(k-1)} - \mathbf{z} - \mathbf{d}^{(k-1)}\|_2^2 = \operatorname{prox}_{\tau\phi/\mu}(\mathbf{Px}^{(k-1)} - \mathbf{d}^{(k-1)})$

4: $\mathbf{d}^{(k)} = \mathbf{d}^{(k-1)} - (\mathbf{Px}^{(k)} - \mathbf{z}^{(k)})$

5: $k = k + 1$

6: **until** stopping criterion is satisfied

A variant of ADMM algorithm for more than two functions

Consider $\min_{\mathbf{x}\in\mathbb{R}^n} \sum_{j=1}^{J} g_j(\mathbf{H}_j\mathbf{x})$ and map it into the previous: $\min_{\mathbf{x}} f_1(\mathbf{x}) + f_2(\underbrace{\mathbf{G}\mathbf{x}}_{\mathbf{z}})$

$$f_1(\mathbf{x}) = 0,\; f_2(\mathbf{z}) = \sum_{j=1}^{J} g_j(\mathbf{z}_j),\; \mathbf{G} = \begin{bmatrix} \mathbf{H}_1 \\ \vdots \\ \mathbf{H}_J \end{bmatrix} \in \mathbb{R}^{p\times n},\; \mathbf{z}^{(k)} = \begin{bmatrix} \mathbf{z}_1^{(k)} \\ \vdots \\ \mathbf{z}_J^{(k)} \end{bmatrix},\; \mathbf{d}^{(k)} = \begin{bmatrix} \mathbf{d}_1^{(k)} \\ \vdots \\ \mathbf{d}_J^{(k)} \end{bmatrix},$$

$$\mathbf{x}^{(k)} = \left(\sum_{j=1}^{J}((\mathbf{H}_j)^\top \mathbf{H}_j\right)^{-1}\left(\sum_{j=1}^{J}(\mathbf{H}_j)^\top(\mathbf{z}_j^{(k-1)} + \mathbf{d}_j^{k-1})\right)$$

$$\mathbf{z}_1^{(k)} = \mathrm{prox}_{g_1\mu}(\mathbf{H}_1\mathbf{x}^{(k-1)} - \mathbf{d}_1^{(k-1)})$$

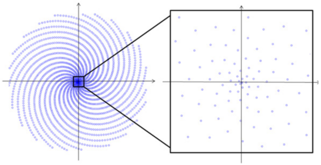
$\vdots$

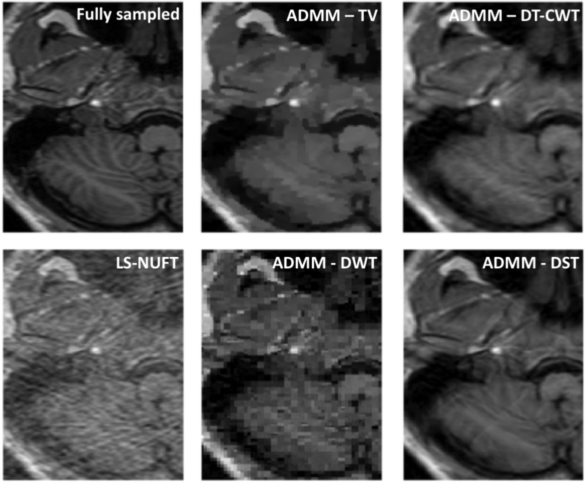
$$\mathbf{z}_J^{(k)} = \mathrm{prox}_{g_J\mu}(\mathbf{H}_J\mathbf{x}^{(k-1)} - \mathbf{d}_J^{(k-1)})$$

C-SALSA [Afonso et al., 2011]

$$\mathbf{d}^{(k)} = \mathbf{d}^{(k-1)} - (\mathbf{G}\mathbf{x}^{(k)} - \mathbf{z}^{(k)})$$

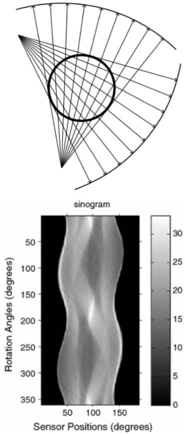
Example: MRI reconstruction with shearlet regularization

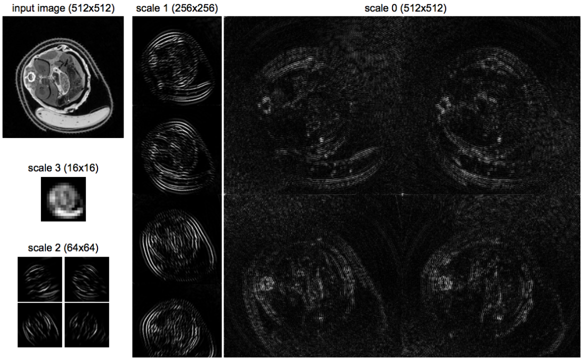




A transversal slice of a FLAIR sequence, resampled along a non-Cartesian trajectory based on an Archimedean spiral (sampling rate 15%). [Aelterman et al., 2011].

Example: CT reconstruction with shearlet regularization

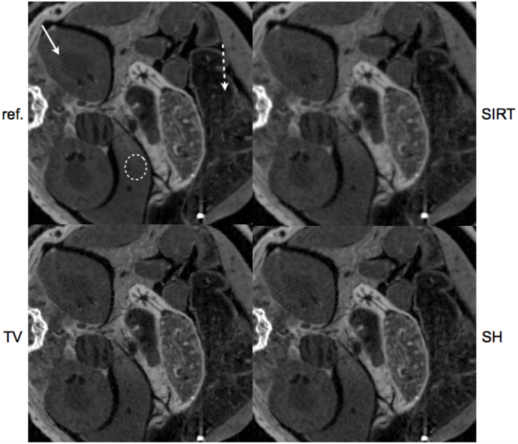




$$\hat{\mathbf{x}} = \arg\min_{x} \frac{1}{2} \|\mathbf{C}^{-1}(\mathbf{A}\mathbf{x} - \mathbf{y})\|_2^2 + \tau \|\mathbf{P}\mathbf{x}\|_1$$

Matrix **C** is a "prewhitener" for the acquisition system [Vandeghinste et al., 2013].

Example: CT reconstruction with shearlet regularization



Top left: reference; **Top right:** SIRT; **Bottom left:** ADMM with TV regularization; **Bottom right:** ADMM with shearlet regularization [Vandeghinste et al., 2013]

Modelling structured sparsity

Two main approaches to modelling structured sparsity in image reconstruction

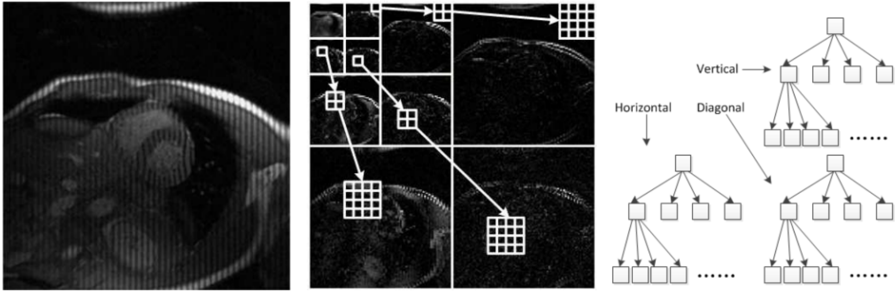
- in the **acquisition** stage
- in the **reconstruction** stage

In the following we only focus on the second approach.

For the the improved design of the sampling patterns/sampling trajectories making use of the structured sparsity, see [Roman et al., 2015], [Adcock et al., 2017], [Gozcu, 2018]

Outline

Wavelet tree sparsity



[Jacob et al., 2009], [He and Carin, 2009], [Rao et al., 2011].

Application to MRI [Chen and Huang, 2014].

Standard CS approach with ℓ_1 regularization:

$$\hat{\theta} = \min_{\theta} \frac{1}{2}\|\mathbf{A\Psi}\theta - \mathbf{y}\|_2^2 + \tau\|\theta\|_1$$

Wavelet-tree sparsity approach:

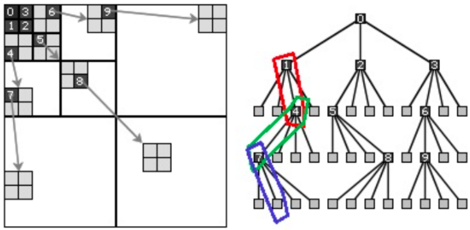


Illustration from [Rao et al., 2011]

A group Lasso estimator:

$$\hat{\theta}_{GL} = \min_{\theta} \frac{1}{2}\|\mathbf{A\Psi}\theta - \mathbf{y}\|_2^2 + \tau \sum_{g\in\mathcal{G}} \|\theta_g\|_2$$

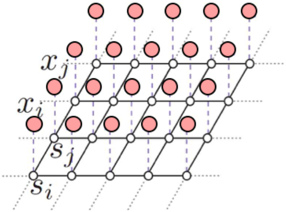
$\mathcal{G}$ – the collection of all parent-child groups;
g is one such group.

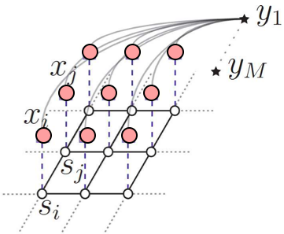
Whole groups are set to zero if their ℓ_2 norm is relatively small.

Outline

Sparse reconstruction with Markov Random Field priors

Consider (a little simpler): $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{n}$, with $\mathbf{y}, \mathbf{n} \in \mathbb{R}^m$, $\mathbf{x} \in \mathbb{R}^n$, $s_i \in \{0, 1\}$

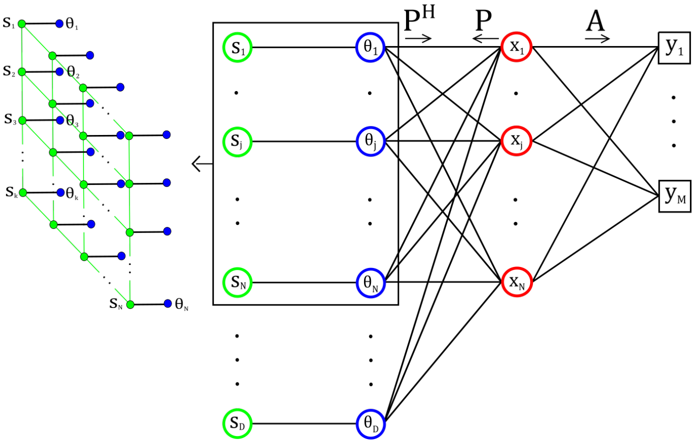




$$P(\mathbf{s}) = \frac{1}{Z}\exp\left[-\left(\sum_i \alpha s_i + \sum_{\langle i,j\rangle} \beta s_i s_j\right)\right] \qquad P(\mathbf{s}, \mathbf{x}, \mathbf{y}) = P(\mathbf{y}|\mathbf{x})P(\mathbf{x}|\mathbf{s})P(\mathbf{s})$$

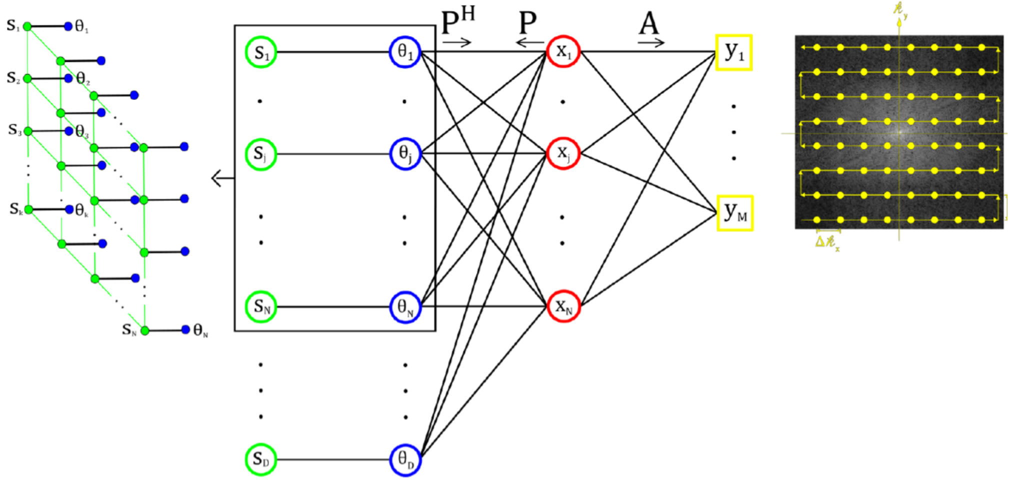
$$[\hat{\mathbf{x}}, \hat{\mathbf{s}}] = \arg\max_{\mathbf{x},\mathbf{s}} \left\{ \sum_{\langle i,j\rangle} \beta s_i s_j + \sum_i [\alpha s_i + \log(p(\mathbf{x}_i|s_i))] - \frac{1}{2\sigma^2}\|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 \right\}$$

Sparse reconstruction with Markov Random Field priors

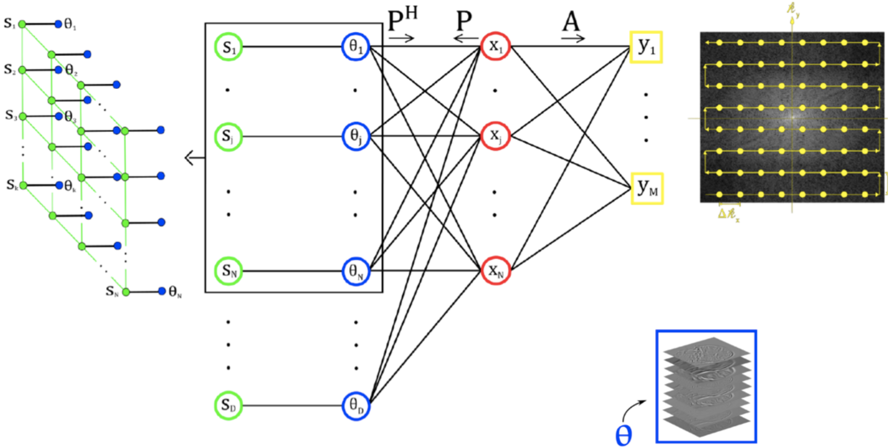


Use Markov Random Field (MRF) as a statistical model for the spatial clustering of important wavelet coefficients [Cevher et al., 2010], [Pižurica et al., 2011]

Sparse MRI reconstruction with MRF priors



Sparse MRI reconstruction with MRF priors



$$\underset{\mathbf{x}\in\mathbb{C}^N}{\text{minimize}}\ \phi(\overbrace{\mathbf{Px}}^{\theta}) \quad \text{subject to} \quad \|\mathbf{Ax}-\mathbf{y}\|_2^2 \le \epsilon$$

Sparse MRI reconstruction with MRF priors

Consider: $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{n}$, with $\mathbf{y}, \mathbf{n} \in \mathbb{C}^m$, $\mathbf{x} \in \mathbb{C}^n$, $\mathbf{x} = \mathbf{\Psi}\boldsymbol{\theta}$, $\boldsymbol{\theta} \in \mathbb{C}^d$, $s_i \in \{0, 1\}$

Let $\Omega_s = \{i \in \mathcal{N} : s_i = 1\}$. Define a **model** for $\boldsymbol{\theta}$ that conforms to the support **s**:

$$\mathcal{M}_s = \{\boldsymbol{\theta} \in \mathbb{C}^D : \mathrm{supp}(\boldsymbol{\theta}) = \Omega_s\}$$

Our objective is:

$$\min_{x \in \mathbb{C}^N} \|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2^2 \quad \text{subject to} \quad \mathbf{P}\mathbf{x} \in \mathcal{M}_{\hat{s}}$$

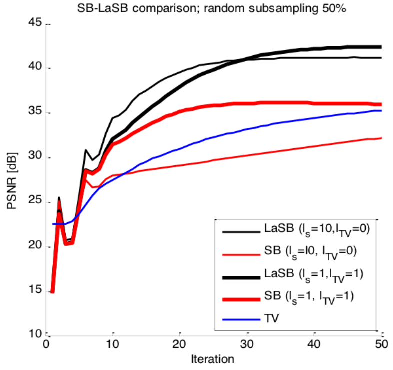
or equivalently:

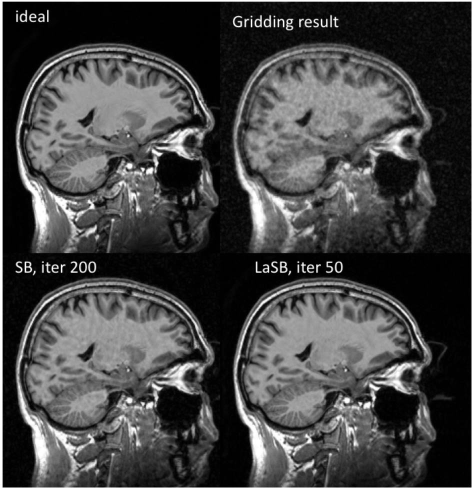
$$\min_{x \in \mathbb{C}^N} \|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2^2 \quad + \iota_{\Omega_s}(\mathrm{supp}(\mathbf{P}\mathbf{x}))$$

where

$$\iota_{\mathcal{Q}}(\mathbf{q}) = \begin{cases} 0, & \mathbf{q} \in \mathcal{Q} \\ +\infty, & \text{otherwise} \end{cases}$$

LaSB [Pižurica et al., 2011], **GreeLa** [Panić et al., 2016], **LaSAL** [Panić et al., 2017]

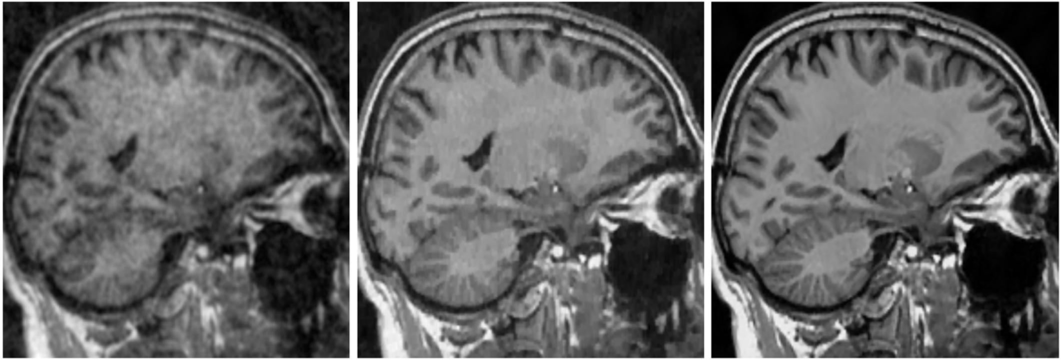




SB (split-Bregman) and **LaSB** implemented with the same shearlet transform.

Example: CS-MRI with MRF priors

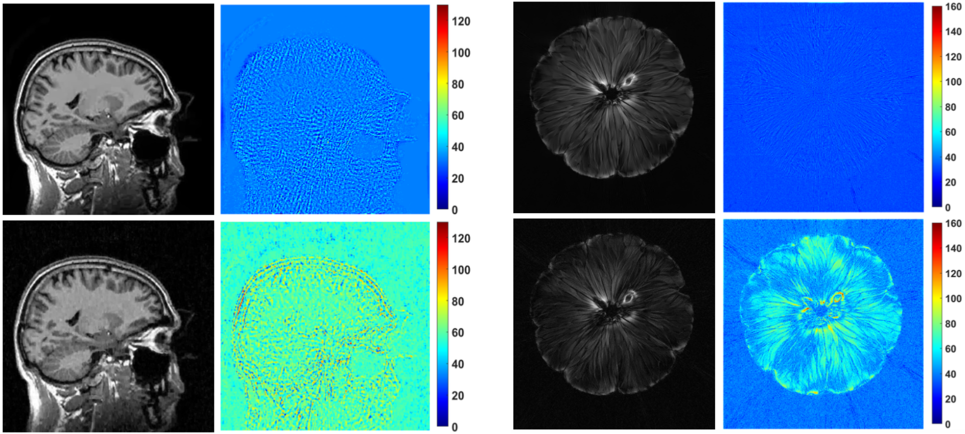
20% measurements, with variable density random sampling



Left: zero fill (PSNR = 19.87 dB)
Middle: **WaTMRI** [Chen and Huang, 2014] (wavelet-tree; PSNR = 28.78 dB)
Right: **LaSAL** [Panić et al., 2017] (MRF-based; PSNR = 33.43 dB)

Example: CS-MRI with MRF priors

Reconstructions from 20% measurements, with radial sampling



Left: reconstructions; **Right**: error images; **Top**: **LaSAL**, **Bottom**: **WaTMRI**

Outline

1. Model-based iterative reconstruction algorithms
   - Sparse optimization
   - Solution strategies: greedy methods vs. convex optimization
   - Optimization methods in sparse image reconstruction
2. Structured sparsity
   - Wavelet-tree sparsity
   - Markov Random Field (MRF) priors
3. Machine learning in image reconstruction
   - Main ideas and current trends

Machine learning in image reconstruction

Covered in many recent workshops, special sessions and special issues of journals:

IEEE TRANSACTIONS ON MEDICAL IMAGING, VOL. 37, NO. 6, JUNE 2018 1289

Image Reconstruction Is a New Frontier of Machine Learning

Ge Wang, *Fellow, IEEE*, Jong Chu Ye, *Senior Member, IEEE*, Klaus Mueller, *Senior Member, IEEE*, and Jeffrey A. Fessler, *Fellow, IEEE*

Three main direction have been proposed

- **learned postprocessing** or **learned denoisers**;
- **learn a regularizer** and use it in a classical variational regularization scheme;
- **learning the full reconstruction operator**

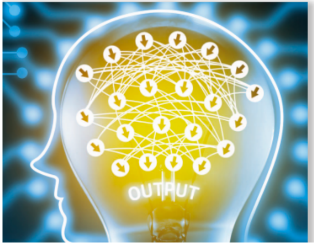
Machine learning in image reconstruction: Special issues

DEEP LEARNING FOR VISUAL UNDERSTANDING

Michael T. McCann, Kyong Hwan Jin,
and Michael Unser

Convolutional Neural Networks for Inverse Problems in Imaging

A review



In this article, we review recent uses of convolutional neural networks (CNNs) to solve inverse problems in imaging. It has recently become feasible to train deep CNNs on large databases of images, and they have shown outstanding performance on object classification and segmentation tasks. Motivated by these successes, researchers have begun to apply CNNs to the resolution of inverse problems such as denoising, deconvolution, super-resolution, and medical image reconstruction, and they have started to report improvements over state-of-the-art methods, including sparsity-based techniques such as compressed sensing. Here, we review the recent experimental work in these areas, with a focus on the critical design decisions:

- From where do the training data come?
- What is the architecture of the CNN?
- How is the learning problem formulated and solved?

We also mention a few key theoretical papers that offer perspectives on why CNNs are appropriate for inverse problems, and we point to some next steps in the field.

IEEE Signal Processing Magazine,
November 2017

Deep Learning for Visual Understanding, Part 1

Machine learning in image reconstruction: Special issues

DEEP LEARNING FOR VISUAL UNDERSTANDING: PART 2

Alice Lucas, Michael Iliadis, Rafael Molina, and Aggelos K. Katsaggelos

Using Deep Neural Networks for Inverse Problems in Imaging

Beyond analytical methods

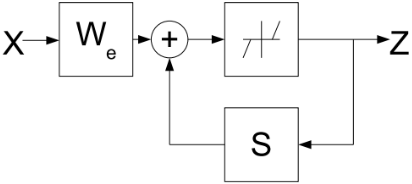
Traditionally, analytical methods have been used to solve imaging problems such as image restoration, inpainting, and superresolution (SR). In recent years, the fields of machine and deep learning have gained a lot of momentum in solving such imaging problems, often surpassing the performance provided by analytical approaches. Unlike analytical methods for which the problem is explicitly defined and domain-knowledge carefully engineered into the solution, deep neural networks (DNNs) do not benefit from such prior knowledge and instead make use of large data sets to learn the unknown solution to the inverse problem. In this article, we review deep-learning techniques for solving such inverse problems in imaging. More

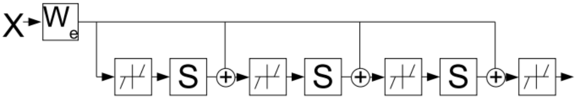
IEEE Signal Processing Magazine, January 2018

Deep Learning for Visual Understanding, Part 2

Learning fast approximations of sparse coding

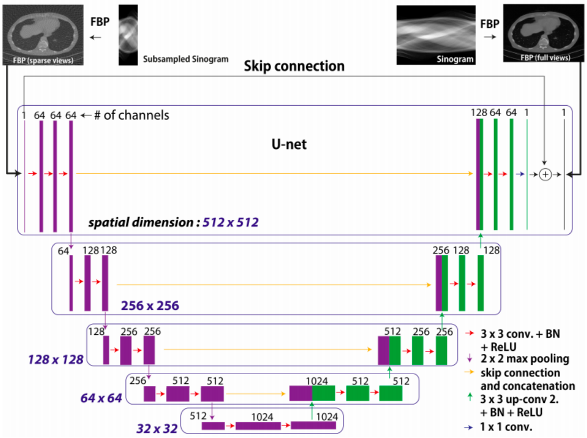
Core idea: time-unfolded version of an iterative reconstruction algorithm, like **IST**, truncated to a fixed number of iterations.
Representatives: **LISTA** [Gregor and LeCun, 2010],[Moreau and Bruna, 2017]





Deep CNN models in image reconstruction

A central question is whether one can combine elements of model and data driven approaches for solving ill-posed inverse problems.



[McCann et al., 2017]

Summary

- Sparse optimization is a fundamental concept in inverse problems like image reconstruction.
- We covered some basic components of sparse image recovery algorithms, including ADMM-based methods.
- The concept of structured sparsity was underlined with particular attention to using Markov Random Field priors in sparse image recovery.
- A new frontier: machine learning in image reconstruction. Great potential, a huge variability of approaches.

Adcock, B., Hansen, A., Poon, C., and Roman, B. (2017).
Breaking the coherence barrier: A new theory for compressed sensing.
Forum of Mathematics, Sigma, 5(e4).

Aelterman, J., Luong, H. Q., Goossens, B., Pižurica, A., and Philips, W. (2011).
Augmented Lagrangian based reconstruction of non-uniformly sub-Nyquist sampled MRI data.
Signal Processing, 91(12):2731–2742.

Afonso, M. V., Bioucas-Dias, J. M., and Figueiredo, M. A. (2010).
Fast image recovery using variable splitting and constrained optimization.
IEEE Transactions on Image Processing, 19(9):2345–2356.

Afonso, M. V., Bioucas-Dias, J. M., and Figueiredo, M. A. T. (2011).
An augmented Lagrangian approach to the constrained optimization formulation of imaging inverse problems.
IEEE Transactions on Image Processing, 20(3):681–695.

Beck, A. and Teboulle, M. (2009).

A fast iterative shrinkage-thresholding algorithm for linear inverse problems.
SIAM Journal on Imaging Sciences, 2(1):183–202.

Bioucas-Dias, J. M. and Figueiredo, M. A. (2007).
A new twist: two-step iterative shrinkage/thresholding algorithms for image restoration.
IEEE Transactions on Image Processing, 16(12):2992–3004.

Blumensath, T. and Davies, M. (2009).
Sampling theorems for signals from the union of finite-dimensional linear subspaces.
IEEE Transactions on Information Theory, 55(4):1872–1882.

Boyd, S., Parikh, N., Chu, E., Peleato, B., and Eckstein, J. (2011).
Distributed optimization and statistical learning via the alternating direction method of multipliers.
Foundations and Trends in Machine Learning, 3(1):1–122.

Candès, E. J., Romberg, J., and Tao, T. (2006).

Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information.
IEEE Transactions on Information Theory, 52(2):489–509.

Cevher, V., Indyk, P., Carin, L., and Baraniuk, R. G. (2010).
Sparse signal recovery and acquisition with graphical models.
IEEE Signal Processing Magazine, 27(6):92–103.

Chambolle, A. (2004).
An algorithm for total variation minimization and applications.
Journal of Mathematical Imaging and Vision, 20(1-2):89–97.

Chen, C. and Huang, J. (2014).
Exploiting the wavelet structure in compressed sensing MRI.
Magnetic Resonance Imaging, 32(10):1377–1389.

Chen, S., Donoho, D., and Saunders, M. (2001).
Atomic decomposition by basis pursuit.
SIAM Review, 43(1):129–159.

Combettes, P. L. and Wajs, V. R. (2005).
Signal recovery by proximal forward-backward splitting.
Multiscale Modeling & Simulation, 4(4):1168–1200.

Courant, R. (1943).
Variational methods for the solution of problems of equilibrium and vibrations.
Bulletin of the American Mathematical Society, 49(1):1–23.

Daubechies, I., Defrise, M., and De Mol, C. (2004).
An iterative thresholding algorithm for linear inverse problems with a sparsity constraint.
Communications on Pure and Applied Mathematics, 57(11):1413–1457.

Donoho, D. L. (2006).
Compressed sensing.
IEEE Transactions on Information Theory, 52(4):1289–1306.

Figueiredo, M. and Nowak, R. (2003).
An EM algorithm for wavelet-based image restoration.

IEEE Transactions on Image Processing, 12(8):906–916.

Figueiredo, M. and Wright, S. (2013).
Applications of sparse optimization.

Goldstein, T. and Osher, S. (2009).
The split Bregman method for ℓ_1-regularized problems.
SIAM Journal on Imaging Sciences, 2(2):323–343.

Gozcu, B. e. a. (2018).
Learning-based compressive MRI.
IEEE Transactions on Medical Imaging, 37(6).

Gregor, K. and LeCun, Y. (2010).
Learning fast approximations of sparse coding.
In *Proceedings of the 27th International Conference on Machine Learning*.

He, L. and Carin, L. (2009).
Exploiting structure in wavelet-based bayesian compressive sensing.
IEEE Transactions on Signal Processing, 57(9):3488–3497.

Hestenes, M. R. (1969).
Multiplier and gradient methods.
Journal of Optimization Theory and Applications, 4(5):303–320.

Jacob, L., G., O., and Vert, J. (2009).
Group Lasso with overlap and graph Lasso.
In *Proceedings of the 26th International Conference on Machine Learning*.

Lustig, M., Donoho, D., and Pauly, J. M. (2007).
Sparse MRI: The application of compressed sensing for rapid MR imaging.
Magnetic Resonance in Medicine, 58(6):1182–1195.

Mallat, S. and Zhang, Z. (1993).
Matching pursuits with time-frequency dictionaries.
IEEE Transactions on Image Processing, 41(12):3397–3415.

McCann, M., Jin, K., and Unser, M. (2017).
Convolutional neural networks for inverse problems in imaging.
IEEE Signal Processing Magazine, 34(6):85–95.

Moreau, T. and Bruna, J. (2017).
Understanding trainable sparse coding via matrix factorization.
In *Proceedings ICLR*.

Needell, D. and Tropp, J. A. (2009).
Cosamp: Iterative signal recovery from incomplete and inaccurate samples.
Applied and Computational Harmonic Analysis, 26(3):301–321.

Panić, M., Aelterman, J., Crnojević, V., and Pižurica, A. (2017).
Sparse recovery in magnetic resonance imaging with a Markov random field prior.
IEEE Transactions on Medical Imaging, 36(10):2104–2115.

Panić, M., Vukobratović, D., Crnojević, V., and Pižurica, A. (2016).
Sparse MRI with a Markov random field prior for the subband coefficients.
In *Proceedings of the third "international Traveling Workshop on Interactions between Sparse models and Technology" (iTWIST'16)*, pages 56–58.

Pižurica, A., Aelterman, J., Bai, F., Vanloocke, S., Luong, Q., Goossens, B., and Philips, W. (2011).

On structured sparsity and selected applications in tomographic imaging.
In *SPIE Conference on Wavelets and Sparsity XIV*, volume 8138, pages 81381D–1–12.

Powell, M. (1969).
A method for nonlinear constraints in minimization problems.
In Fletcher, R., editor, *Optimization*, pages 283–298. Academic Press, New York.

Rao, N., Nowak, R., Wright, S., and Kingsbury, N. (2011).
Convex approaches to model wavelet sparsity patterns.
In *IEEE international Conference on Image Processing*.

Roman, B., Adcock, B., and Hansen, A. (2015).
On asymptotic structure in compressed sensing,.
In *Proceedings of the third "international Traveling Workshop on Interactions between Sparse models and Technology" (iTWIST'16)*.

Setzer, S. (2009).
Split Bregman algorithm, Douglas-Rachford splitting and frame shrinkage.

In *International Conference on Scale Space and Variational Methods in Computer Vision*, pages 464–476. Springer.

Tibshirani, R. (1996).
Regression shrinkage and selection via the Lasso.
Journal of the Royal Statistical Society, Series B, 58(1):267–288.

Tropp, J. (2004).
Greed is good: Algorithmic results for sparse approximation.
IEEE Transactions on Information Theory, 50(10):2231–2242.

Vandeghinste, B., Goossens, B., Van Holen, R., Vanhove, C., Pižurica, A., Vandenberghe, S., and Staelens, S. (2013).
Iterative CT reconstruction using shearlet-based regularization.
IEEE Transactions on Nuclear Science, 60(5):3305–3317.

Wright, S. J., Nowak, R. D., and Figueiredo, M. A. (2009).
Sparse reconstruction by separable approximation.
IEEE Transactions on Signal Processing, 57(7):2479–2493.