Tomographic Image Reconstruction Using Training Images

Per Christian Hansen

Joint work with Sara Soltani and Martin S. Andersen

Department of Applied Mathematics and Computer Science
Technical University of Denmark

July 2016

HDtomo
HIGH - DEFINITION TOMOGRAPHY

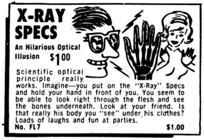
DTU Compute
Department of Applied Mathematics and Computer Science

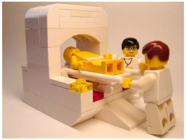
DTU

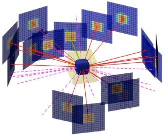
X-Ray Computed Tomography

Tomography is the science of seeing inside objects.

X-rays are sent through an object from many different angles. The response of the object to the signal is measured. We use the data + a mathematical model to compute an image of the object's interior.







Data b_i associated with the ith X-ray through the domain:

$$b_i = \int_{\mathrm{ray}_i} \xi(x,y)\, d\ell, \qquad \xi(x,y) = \text{attenuation coef.}$$

Assume $\xi(x,y)$ is a constant x_j in pixel j. This leads to:

$$b_i = \textstyle\sum_j a_{ij}\, x_j, \qquad a_{ij} = \begin{cases} \text{length of ray } i \text{ in pixel } j \\ 0 \text{ if ray } i \text{ does not intersect pixel } j. \end{cases}$$

This leads to a large sparse system $A\, x \approx b$.

Outline

The use of training images for image reconstruction is not a new idea.

Our main interest is in formulations that lead to efficient and robust *computational algorithms* based on numerical optimization methods.

S. Soltani, M. S. Andersen, and P. C. Hansen, *Tomographic image reconstruction using training images*, submitted to J. Comput. Appl. Math.

The Linear Inverse Problem of Image Reconstruction

Discretization of the tomographic imaging problem:

Unknown image   Data   Exact data   Exact image

$$A\,x \approx b, \qquad b = \bar{b} + \epsilon, \qquad \bar{b} = A\,\bar{x}$$

Forward projection model   White noise

The vector $x \in \mathbb{R}^n$ represents the image.

The matrix A is large, sparse, and ill-conditioned $\rightarrow$ need regularization!

A basic reconstruction problem:

$$\min_x \|A\,x - b\|_2^2 + \mathcal{R}(x),$$

where $\mathcal{R}(x)$ is an appropriate regularization term.

Regularization is necessary when the problem is underdetermined.

More About the Regularization Term

$$\min_x \|A\,x - b\|_2^2 + \mathcal{R}(x)$$

We can use the regularization term $\mathcal{R}(x)$ to incorporate smoothness requirements on the image x:

- $\|x\|_2^2$ gives a smooth image.
- Total Variation (TV) gives a piecewise smooth image.

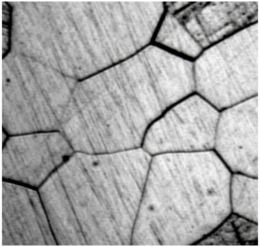
But some priors – e.g., about texture or other visual features of x – are hard to represent in this form.

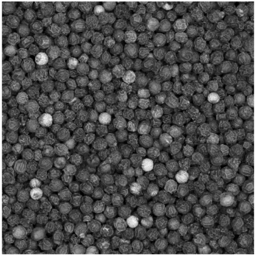
An alternative is to use a carefully constructed set of images, store them as columns of a matrix W, and compute a sparse representation or approximation using this set:

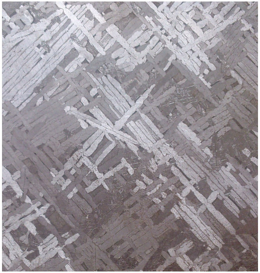
$$x = W\,\alpha, \qquad \alpha \text{ sparse.}$$

Priors in the Form of Training Images

Prior information is sometimes available in the form of *training images* which represent geometrical or visual features.



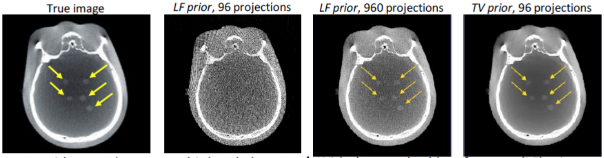




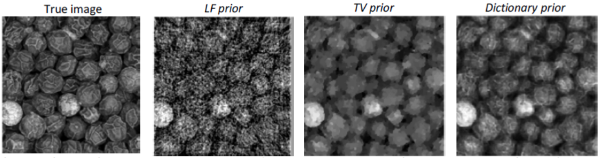
Key idea: Incorporate prior information from the training images in the form of a *dictionary of image patches* that sparsely encodes the training data [Olshausen & Field 1996].

We use *image patches* because working with full images is computationally overwhealming/infeasible.

A Word of Warning About Priors



Experiments with an anthropomorphic head phamtom[4]. With the standard *low-freqency (LF) prior* we need 960 projections to identify 5 small regions of soft tissue (marked by the arrows). With a *total variation (TV) prior* we only need 96 projections, thus reducing the X-ray dose by a factor of 10.



Artifical example. Neither *LF* nor *TV* priors provide a reconstruction with correct texture, but a *dictionary prior* developed by us (similar to ideas described in this proposal) gives a reconstruction with reasonable texture[22].

While a *dictionary prior* is potentially well suited for texture images, an incorrect type of dictionary image enforces a wrong texture in the reconstruction (P.C. Hansen & S. Soltani, DTU Compute).

Constructing/Learning the Dictionary

Some dictionary learning *applications* in imaging:

- Sparse representation of signals and images [Elad 2010].
- Image denoising [Elad & Ahron 2006], inpainting [Mairal et al. 2008], and deblurring [Liu et al. 2013].
- Low-dose CT reconstruction [Xu et al. 2012].

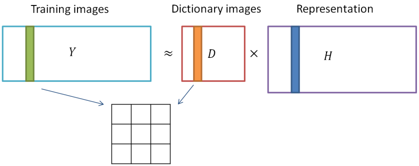
Dictionary learning *methods:* Sparse coding [Olshausen & Field 1996], Nonnegative matrix factorization [Lee & Seung 1999], K-SVD [Elad & Aharon 2006] and many more . . .

Two different approaches to creating the dictionary:

- Learning from the given data during the reconstruction (NP hard).
- Learning from given (high-res) images before the reconstruction.

Dictionary Learning via Nonnegative Matrix Factorization

- Use small image patches of size $P \times Q$.
- Organize each of the t training image patches as columns of a matrix $Y \in \mathbb{R}^{p \times t}$, with $p = PQ$.
- **Nonnegative matrix factorization** $\boxed{Y \approx DH}$ $D =$ dictionary.



- Such a factorization is not unique $\rightarrow$ regularize the problem.
- Our generic dictionary learning problem:

$$\min_{D,H} \ \tfrac{1}{2}\|Y - DH\|_{\mathrm{F}}^2 + \Phi_{\mathrm{dic}}(D) + \Phi_{\mathrm{rep}}(H)$$

Φ_{dic} and Φ_{rep} incorporate priors on <u>dic</u>tionary and <u>rep</u>resentation.

The Regularized Dictionary Learning Problem

Our specific discionary learning problem takes the form:

$$\min_{D,H} \quad \frac{1}{2}\|Y - DH\|_{\mathrm{F}}^2 + \lambda\|H\|_{\mathrm{sum}} \quad \text{s.t.} \quad D \in \mathcal{D},\ H \in \mathbb{R}_+^{s\times t}$$

where λ = regularization parameter and sparsity is enforced by

$$\|H\|_{\mathrm{sum}} \equiv \sum_{i=1}^{s}\sum_{j=1}^{t} |H_{ij}| = \|H(:)\|_1.$$

Dictionary $D = [d_1, d_2, \ldots, d_s]$ and two choices of the set $\mathcal{D}$:

$$\mathcal{D}_\infty = \{D \in \mathbb{R}_+^{p\times s} \,|\, \|d_j\|_\infty \leq 1\},$$

$$\mathcal{D}_2 = \{D \in \mathbb{R}_+^{p\times s} \,|\, \|d_j\|_2 \leq \sqrt{p}\}.$$

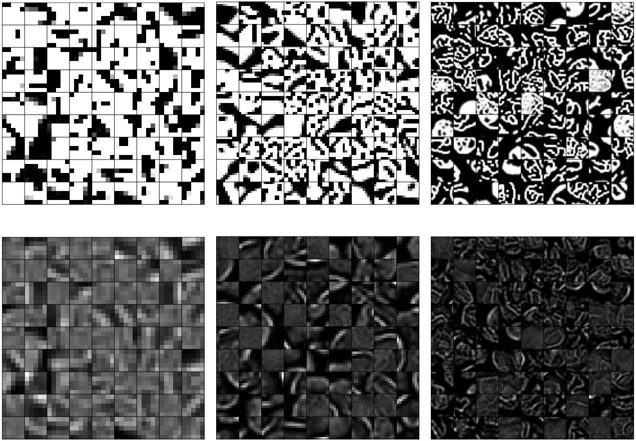
We use Alternating Direction Method of Multiplies (ADMM) [Boyd et al. 2011] to compute a (local) minimum for this problem.

Illustration of Dictionary Elements (Image Patches)

Examples of different dictionaries – the patches have different size $P \times Q$, and $s =$ no. of dictionary elements (dictionary size).

Top: $D \in \mathcal{D}_\infty$

Bottom: $D \in \mathcal{D}_2$



$5 \times 5,\ s = 100$ $\quad$ $10 \times 10,\ s = 300$ $\quad$ $20 \times 20,\ s = 800$

The Generic Reconstruction Problem

All previous algorithm use overlapping patches, which increases the amount of computation work considerably.

- We partition the image into $q = {}^{M}/_{P} \cdot {}^{N}/_{Q}$ *non-overlapping* patches.
- Global dictionary $W = \Pi \, (I \otimes D)$ from learned patch dictionary D.
- The permutation Π puts image patches in the right locations.
- Write the solution as

$$x = W \, \alpha = \Pi \begin{pmatrix} D & & \\ & \ddots & \\ & & D \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_q \end{pmatrix} .$$

Find a regularized solution that solves

$$\min_\alpha \tfrac{1}{2m} \| A \, W \alpha - b \|_2^2 + \Phi_{\rm sp}(\alpha) + \Phi_{\rm ip}(\alpha),$$

where $\Phi_{\rm sp}$ = sparsity prior and $\Phi_{\rm ip}$ = image prior.

Our Regularized Reconstruction Problem

- Sparsity prior on coefficient vector α via

$$\Phi_{\mathrm{sp}} = \mu \, \|\alpha\|_1 / q, \qquad \mu = \text{reg. parameter.}$$

- Non-overlapping patches give rise to block artifacts.
- Hence we introduce an image prior

$$\Phi_{\mathrm{ip}}(\alpha) = \delta^2 \, \|L\, W \alpha\|_2^2 / \nu, \qquad \mu = \text{normalization constant}$$

  that penalizes block artifacts; the matrix L contains finite-difference approximations across the block edges.

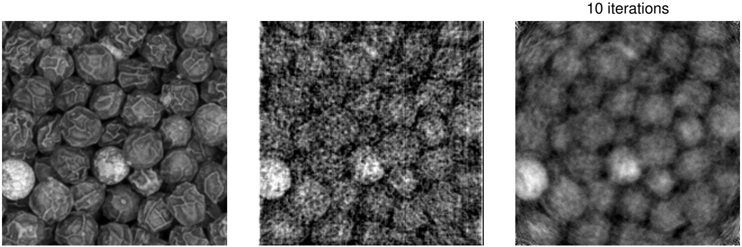
The reconstruction problem is a convex optimization problem:

$$\min_\alpha \ \frac{1}{2m} \|A\, W\alpha - b\|_2^2 + \frac{\mu}{q} \|\alpha\|_1 + \delta^2 \, \|L\, W\alpha\|_2^2/\nu \quad \text{s.t.} \quad \alpha \geq 0.$$

Solve with software package TFOCS (Templates for First-Order Conic Solvers) [Becker et al. 2011].

Underdetermined CT Reconstruction Problem – 2D

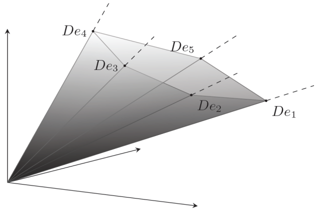
- Parallel beam, 25 projections in $[0°, 180°)$, each of 283 rays.
- The image is 200×200.
- The system matrix $A \in \mathbb{R}^{7{,}075 \times 40{,}000}$ $\leftarrow$ highly underdetermined.
- Additive white Gaussian noise e with $\|e\|_2 / \|A\,\bar{x}\|_2 = 0.01$.
- 50 000 training images from a much larger image of peppers.
- Patch sizes: 5×5, 10×10, and 20×20.
- Corresponding dictionary matrices $D^{(5)}$, $D^{(10)}$, $D^{(20)} \in \mathcal{D}_2$ of size 25×100, 100×300, and 400×800.



The 200×200 exact image $\bar{x}$; FBP and ART reconstructions (bad).

Beware of the Approximation Error

- We require $\alpha \geq 0 \Rightarrow x \geq 0$.
- The cone defined by our patch dictionary: $\mathcal{C} = \{D\alpha \mid \alpha \in \mathbb{R}^s_+\}$.
- Typically $\mathcal{C} \subset \mathbb{R}^p_+$, i.e., a proper subset!



Note that the exact image patches may not lie in the cone $\mathcal{C}$, thus leading to an *approximation error*.

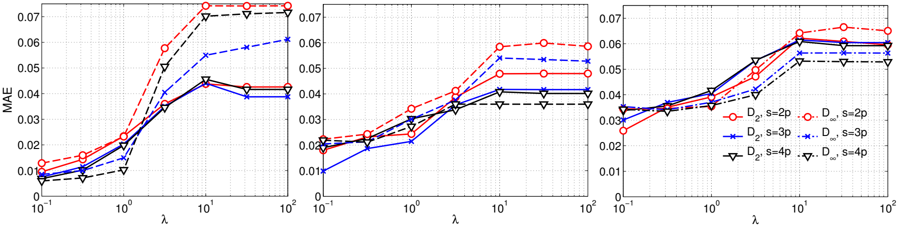
- Reconstruction error = regularization error + approximation error.
- Desired: approximation error $\ll$ regularization error.

Dictionary Learning: Approximation Errors

Mean approximation error:

$$\text{MAE} = \frac{1}{q}\sum_{j=1}^{q} \frac{1}{\sqrt{p}} \left\| P_{\mathcal{C}}(\bar{x}_j) - \bar{x}_j \right\|_2,$$

where $P_{\mathcal{C}}(\bar{x}_j) = D\alpha_j^\star$ = best approximation of the jth block in the cone.



Patch size $P \times Q = 5 \times 5,\ \ 10 \times 10,\ \ 20 \times 20.$

Number of dictionary elements $s = \{2, 3, 4\}\, p$ with $p = P\, Q$.

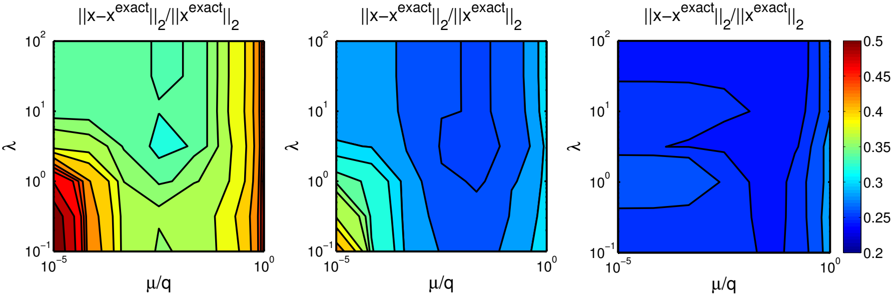
Dictionaries $D \in \mathcal{D}_\infty$ and $D \in \mathcal{D}_2$.

MAE is somewhat independent of the block size, and for larger block sizes less dependent on λ (the dictionary learning reg. parameter).

Reconstruction Error (RE) versus Reg. Parameters, I

$$\min_{D,H} \; {}^1\!/_2 \|Y - DH\|_{\mathrm{F}}^2 + \lambda \|H\|_{\mathrm{sum}}$$

$$\min_{\alpha} {}^1\!/_{(2m)} \|A\,\Pi^T (I \otimes D)\alpha - b\|_2^2 + {}^\mu\!/_q \|\alpha\|_1 \qquad \delta = 0$$

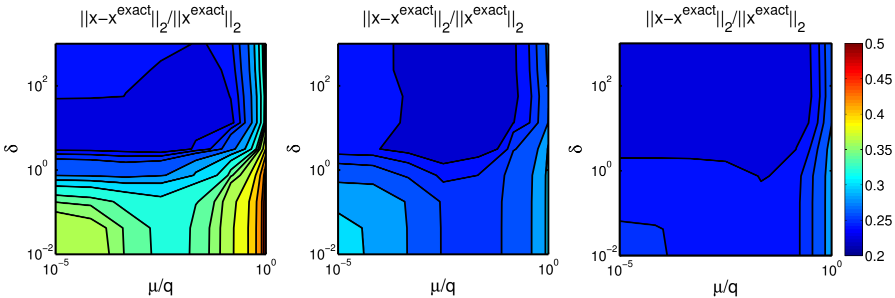


Blocks are 5×5, 10×10, 20×20. For 20×20 the RE is not so dependent on the regularization parameters λ and μ. Smallest errors for $\lambda \approx 3$.

Reconstruction Error (RE) versus Reg. Parameters, II

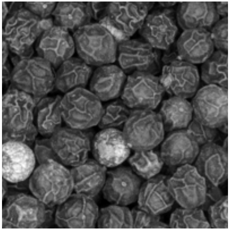
$$\min_{D,H} \ 1/2\|Y - DH\|_F^2 + \lambda\|H\|_{\text{sum}}, \qquad \lambda = 0.3$$

$$\min_{\alpha} 1/(2m)\|A\,\Pi^T(I \otimes D)\alpha - b\|_2^2 + \mu/q\|\alpha\|_1 + \delta^2\,\|L\,W\alpha\|_2^2/\nu$$



Blocks are 5×5, 10×10, 20×20. Smaller RE for larger patch sizes where the RE is almost independent of δ and quite independent of μ.

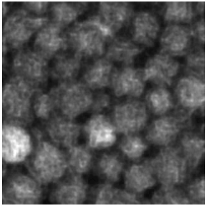
Reconstructions



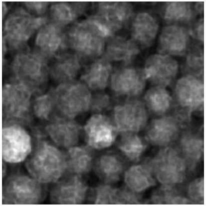
Exact image

$D \in \mathcal{D}_2$, $\lambda = 3$,
and $\mu/q = 0.022$.

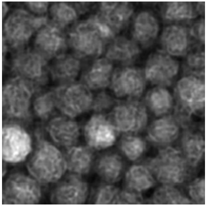
We obtain more structure and perhaps more correct edges than TV.



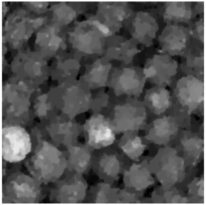
5×5



10×10



20×20



TV

Sensitivity to More Noise and Limited-Angle Data

$\lambda = 3$

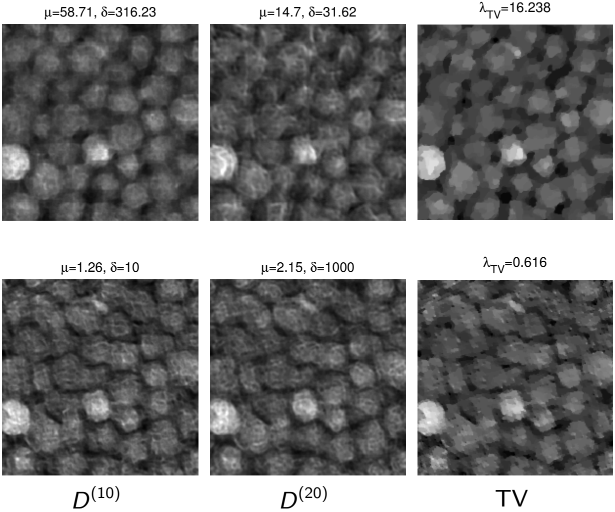
Top:

$N_p = 25 \in [0^\circ, 180^\circ]$, noise 5%.

Bottom:

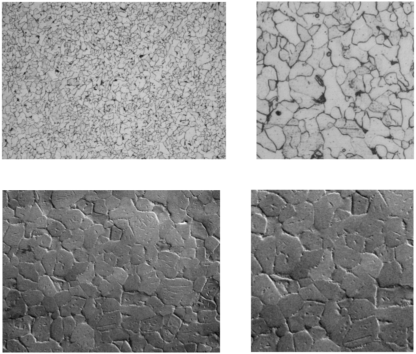
$N_p = 25 \in [0^\circ, 120^\circ]$, noise 1%.

In both cases we do much better than TV; FBP and ART (not shown) are terrible.



A Large Test Problems: The Data

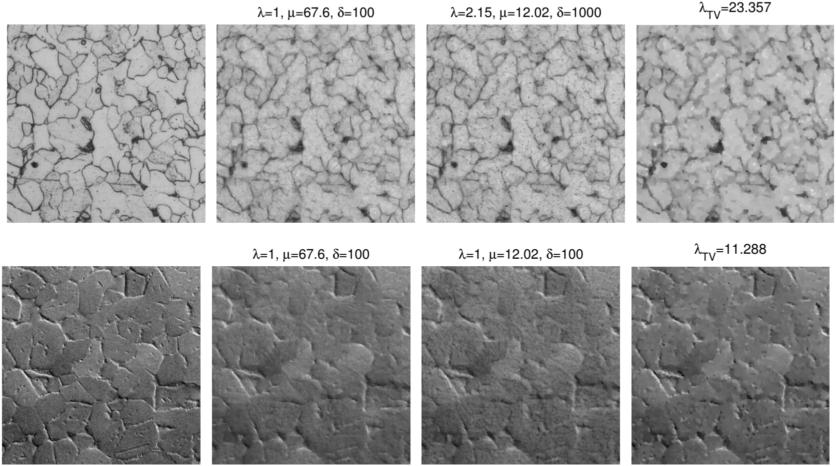
- Parallel-beam tomography with 50 projections in $[0^\circ, 180^\circ)$, each of 735 rays.
- Reconstructed images is 520×520.
- Problem size:
  $m =$ 36,750 data values,
  $n =$ 270,400 pixels.
- Patch size is 20×20.
- Dictionary $D^{(20)} \in \mathcal{D}_2, \mathcal{D}_\infty$ of size 400×800.
- Relative Gaussian noise level 0.01.



Left: Training images of steel micro-structure (top) and zirconium grains (bottom).

Right: The exact images of size 520×520.

A Large Test Problem: Reconstructions



Exact | $D^{20} \in \mathcal{D}_2$ | $D^{20} \in \mathcal{D}_\infty$ | TV

Conclusions

- Efficient reconstruction algorithm that uses training images as priors.
- Working with *nonoverlapping* blocks reduces the computational work compared to previous algorithms that use overlapping blocks around every pixel in the image.
- Training images provide a strong prior.
- Sparsity prior + nonnegativity constraints $\Rightarrow$ additional regularization.
- Reconstructions are not very sensitive to regularization parameters.
- Promising results for artificial test problems.
- Tensor formulation [Soltani, Kilmer, H 2016]: