Inverse Problems 1: convolution and deconvolution

Lesson 12: Compressed Sensing and Dictionary Learning

Luca Ratti

October 9, 2019

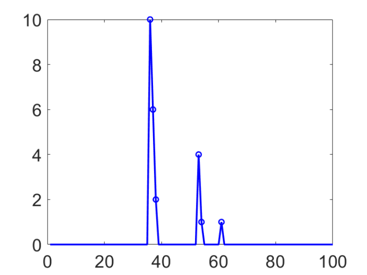
University of Helsinki

Table of contents

What is Compressed Sensing?

The notion of sparsity

A vector $f \in \mathbb{R}^n$ is said to be **sparse** if most of its components are vanishing.

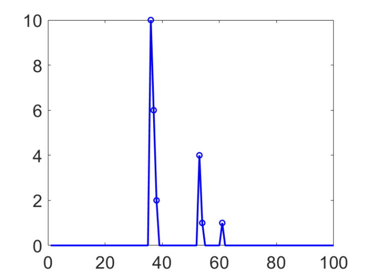


Example: a 1D signal recording the seismical activity in a certain region.

How to describe rigorously the notion of sparsity?

The notion of sparsity

A vector $f \in \mathbb{R}^n$ is said to be **sparse** if most of its components are vanishing.



Example: a 1D signal recording the seismical activity in a certain region.

How to describe rigorously the notion of sparsity?

L^0 "norm" and sparsity

$\|f\|_0 = \#\{j : f_j \neq 0\}$. **Sparse** vector: $f \in \mathbb{R}^n$ s.t. $\|f\|_0 \ll n$.

Warning: $\|\cdot\|_0$ is not actually a norm, and is a delicate tool. Its name comes from the property $\|f\|_0 = \lim_{p\to 0} \|f\|_p^p$, but this does not imply the properties of the norm (e.g. the homogeneity).

Sparsity with respect to a dictionary

Canonical basis in $\mathbb{R}^n$: $\{e_i\}_{i=1}^n$ such that each $e_i \in \mathbb{R}^n$ has all components equal to 0, except the $i-$th one, which is equal to 1; briefly, $(e_i)_j = \delta_{ij}$.

It holds $f \in \mathbb{R}^n \Rightarrow f = \sum_{i=1}^{n} f_i \mathrm{e_i}$.

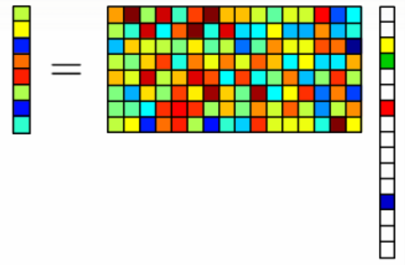
Sparsity with respect to a dictionary

Canonical basis in $\mathbb{R}^n$: $\{e_i\}_{i=1}^n$ such that each $e_i \in \mathbb{R}^n$ has all components equal to 0, except the $i-$th one, which is equal to 1; briefly, $(e_i)_j = \delta_{ij}$.

It holds $f \in \mathbb{R}^n \Rightarrow f = \sum_{i=1}^{n} f_i \mathrm{e_i}$.

Sparsity w.r.t. a dictionary

Consider a matrix $D \in \mathbb{R}^{n \times m}$, and take $f \in \mathbb{R}^n$, $w \in \mathbb{R}^m$ s.t. $f = Dw$. This is equivalent to $f = \sum_{i=1}^{m} w_i \mathrm{d_i}$, being $\mathrm{d_i}$ the columns of D.
A vector $f \in range(D)$ is said to be sparse with respect to D, or to admit a sparse representation in D if $\|w\|_0 \ll n$.



Compressed Sensing: alternative formulations

Original definition of compressed sensing:

$$\text{find } w^* = \arg\min_{w \in \mathbb{R}^m} \left\{ \|w\|_0; \quad w \text{ s.t. } \|Dw - f\|_2^2 \leq \delta \right\}$$

Interpretation: find the sparsest representation of f with respect to D.

We consider the possibility that some inaccuracy may occur on the datum f.

Compressed Sensing: alternative formulations

Original definition of compressed sensing:

$$\text{find } w^* = \underset{w \in \mathbb{R}^m}{\arg\min} \left\{ \|w\|_0; \quad w \text{ s.t. } \|Dw - f\|_2^2 \leq \delta \right\}$$

Interpretation: find the sparsest representation of f with respect to D.
We consider the possibility that some inaccuracy may occur on the datum f.
This is the Morozov formulation of a **Variational Regularization problem** with discrepancy term $\mathcal{D}(w, f)$ and regularization term $\mathcal{R}(w)$:

$$\textit{(Morozov)} \quad w_M^* = \underset{w \in \mathbb{R}^m}{\arg\min} \{\mathcal{R}(w); \quad w \text{ s.t. } \mathcal{D}(w, f) \leq \delta\}$$

$$\textit{(Ivanov)} \quad w_I^* = \underset{w \in \mathbb{R}^m}{\arg\min} \{\mathcal{D}(w, f); \quad w \text{ s.t. } \mathcal{R}(w) \leq \epsilon\}$$

$$\textit{(Tikhonov)} \quad w_T^* = \underset{w \in \mathbb{R}^m}{\arg\min} \{\mathcal{D}(w, f) + \alpha \mathcal{R}(w)\}.$$

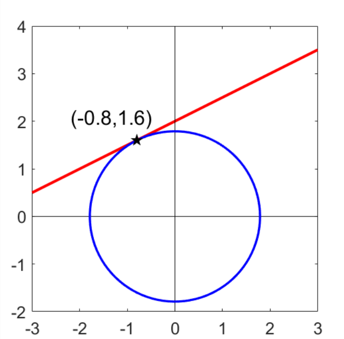
Under certain assumptions, the different formulations are equivalent or lead to similar results. In CS, $\mathcal{D}(w, f) = \|Dw - f\|_2^2$ and $\mathcal{R}(w) = \|w\|_0$.

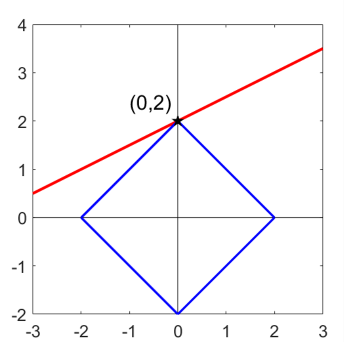
L^1 norm and sparsity

The minimization with respect to $\|\cdot\|_0$ is complicated both theoretically (lack of convexity) and numerically (NP-hard problem).

L^1 norm and sparsity

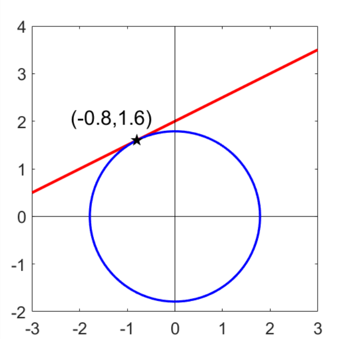
The minimization with respect to $\| \cdot \|_0$ is complicated both theoretically (lack of convexity) and numerically (NP-hard problem).
Under suitable conditions on $\mathcal{D}$, a sparsity promoting regularization is obtained by minimizing $\|w\|_1 = \sum_{i=1}^{m} |w_i|$.

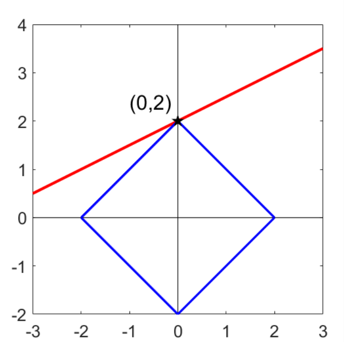




L^1 norm and sparsity

The minimization with respect to $\|\cdot\|_0$ is complicated both theoretically (lack of convexity) and numerically (NP-hard problem).
Under suitable conditions on $\mathcal{D}$, a sparsity promoting regularization is obtained by minimizing $\|w\|_1 = \sum_{i=1}^{m} |w_i|$.





In CS, such conditions are connected to the **Restricted Isometry Property** and to the **incoherence** of the columns of D.

A particular case

Compressed Sensing: Tikhonov formulation with L^1 norm

$$w^* = \underset{w \in \mathbb{R}^n}{\arg\min} \left\{ \|Dw - f\|_2^2 + \alpha \|w\|_1 \right\}$$

A particular case

Compressed Sensing: Tikhonov formulation with L^1 norm

$$w^* = \underset{w \in \mathbb{R}^n}{\arg\min} \left\{ \|Dw - f\|_2^2 + \alpha \|w\|_1 \right\}$$

In case $D \in \mathbb{R}^{n \times m}$ is an orthonormal matrix (so its columns form an orthogonal basis of $\mathbb{R}^n$), then ($D^T D = DD^T = I$):

A particular case

Compressed Sensing: Tikhonov formulation with L^1 norm

$$w^* = \arg\min_{w \in \mathbb{R}^n} \left\{ \|Dw - f\|_2^2 + \alpha \|w\|_1 \right\}$$

In case $D \in \mathbb{R}^{n \times m}$ is an orthonormal matrix (so its columns form an orthogonal basis of $\mathbb{R}^n$), then ($D^T D = DD^T = I$):

$$\|Dw - f\|_2^2 + \alpha \|w\|_1 = \|w - D^T f\|_2^2 + \alpha \|w\|_1 = \sum_{i=1}^{n} \left[(w_i - y_i)^2 + \alpha |w_i| \right],$$

where $y = D^T f$.

A particular case

Compressed Sensing: Tikhonov formulation with L^1 norm

$$w^* = \underset{w \in \mathbb{R}^n}{\arg\min} \left\{ \|Dw - f\|_2^2 + \alpha \|w\|_1 \right\}$$

In case $D \in \mathbb{R}^{n \times m}$ is an orthonormal matrix (so its columns form an orthogonal basis of $\mathbb{R}^n$), then ($D^T D = DD^T = I$):

$$\|Dw - f\|_2^2 + \alpha \|w\|_1 = \|w - D^T f\|_2^2 + \alpha \|w\|_1 = \sum_{i=1}^{n} \left[(w_i - y_i)^2 + \alpha |w_i| \right],$$

where $y = D^T f$. In the minimum problem, the components of w are completely decoupled. Hence, we can solve for each component:

$$w_i = \underset{x \in \mathbb{R}}{\arg\min}\, f(x), \qquad f(x) = (x - y_i)^2 + \alpha |x|.$$

A particular case

Compressed Sensing: Tikhonov formulation with L^1 norm

$$w^* = \arg\min_{w \in \mathbb{R}^n} \left\{ \|Dw - f\|_2^2 + \alpha \|w\|_1 \right\}$$

In case $D \in \mathbb{R}^{n \times m}$ is an orthonormal matrix (so its columns form an orthogonal basis of $\mathbb{R}^n$), then ($D^T D = DD^T = I$):

$$\|Dw - f\|_2^2 + \alpha \|w\|_1 = \|w - D^T f\|_2^2 + \alpha \|w\|_1 = \sum_{i=1}^{n} \left[(w_i - y_i)^2 + \alpha |w_i| \right],$$

where $y = D^T f$. In the minimum problem, the components of w are completely decoupled. Hence, we can solve for each component:

$$w_i = \arg\min_{x \in \mathbb{R}} f(x), \qquad f(x) = (x - y_i)^2 + \alpha |x|.$$

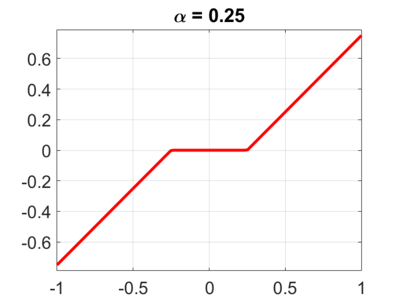
Since f is not differentiable, such problem cannot be solved by (just) computing a derivative. From basic convex analysis, we get

$$w_i = S_\alpha(y_i), \qquad S_\alpha\text{: Soft Thresholding operator}$$

The Soft Thresholding operator

The expression of $S_\alpha(x)$ is as follows:

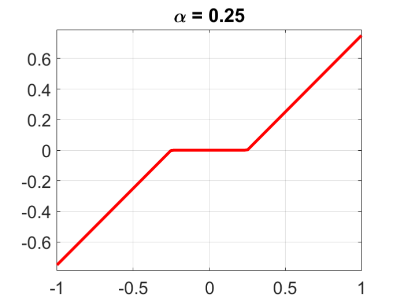
$$S_\alpha(x) = \begin{cases} x + \alpha & \text{if } x \leq -\alpha \\ 0 & \text{if } |x| < \alpha \\ x - \alpha & \text{if } x \geq \alpha \end{cases}$$



The Soft Thresholding operator

The expression of $S_\alpha(x)$ is as follows:

$$S_\alpha(x) = \begin{cases} x+\alpha & \text{if } x \leq -\alpha \\ 0 & \text{if } |x| < \alpha \\ x-\alpha & \text{if } x \geq \alpha \end{cases}$$



In conlcusion, the solution of

$$w^* = \arg\min_{w \in \mathbb{R}^n} \left\{ \|Dw - f\|_2^2 + \alpha \|w\|_1 \right\}$$

in case D is an orthogonal matrix is given by

$$w^* = \mathrm{S}_\alpha(D^T f),$$

where $(\mathrm{S}_\alpha(g))_i = S_\alpha(g_i)$ is the soft thresholding of each component of g.

Sparsity promoting regularization for the deconvolution problem

Sparsity in deconvolution

Consider the deconvolution problem, or any linear inverse problem:

$$\text{given } m_\delta = Af + \varepsilon, \text{ being } \|\varepsilon\| \leq \delta, \text{ determine } f.$$

Sparsity in deconvolution

Consider the deconvolution problem, or any linear inverse problem:

$$\text{given } m_\delta = Af + \varepsilon, \text{ being } \|\varepsilon\| \leq \delta, \text{ determine } f.$$

If we have the additional information that f is sparse with respect to a dictionary $D \in \mathbb{R}^{n \times m}$, we can device a sparsity promoting regularization strategy by looking for a vector $f \in \mathbb{R}^n$ showing

- **data fidelity** $\|Af - m_\delta\|_2$ is small;
- **sparsity** with respect to D: $f = Dw$ for a suitable $w \in \mathbb{R}^m$ and $\|w\|_1$ is small.

Sparsity in deconvolution

Consider the deconvolution problem, or any linear inverse problem:

$$\text{given } m_\delta = Af + \varepsilon, \text{ being } \|\varepsilon\| \leq \delta, \text{ determine } f.$$

If we have the additional information that f is sparse with respect to a dictionary $D \in \mathbb{R}^{n\times m}$, we can device a sparsity promoting regularization strategy by looking for a vector $f \in \mathbb{R}^n$ showing

- **data fidelity** $\|Af - m_\delta\|_2$ is small;
- **sparsity** with respect to D: $f = Dw$ for a suitable $w \in \mathbb{R}^m$ and $\|w\|_1$ is small.

We are mainly focusing on two cases:

1. D is an **orthogonal basis**: $m = n$, $D^T D = DD^T = I$.
2. D is a **tight frame**: $m > n$, $DD^T = I$

Case I: D is an orthonormal base

Examples: $D = I$ (canonical basis), Haar wavelet basis, orhtogonal polinomials, Fourier goniometric polynomials, ...

Case I: D is an orthonormal base

Examples: $D = I$ (canonical basis), Haar wavelet basis, orhtogonal polinomials, Fourier goniometric polynomials, ...

Properties: $range(D) = range(D^T) = \mathbb{R}^n$: every vector f can be written as $f = Dw$, and the weights w can be found as $w = D^T f$.

Case I: D is an orthonormal base

Examples: $D = I$ (canonical basis), Haar wavelet basis, orhtogonal polinomials, Fourier goniometric polynomials, ...

Properties: $range(D) = range(D^T) = \mathbb{R}^n$: every vector f can be written as $f = Dw$, and the weights w can be found as $w = D^T f$.

Two possible formulations:

Sparse deconvolution - synthesis formulation

$$w^* = \arg\min_{w \in \mathbb{R}^m} \{ \|ADw - m_\delta\|_2^2 + \alpha \|w\|_1 \}$$

Sparse deconvolution - analysis formulation

$$f^* = \arg\min_{f \in \mathbb{R}^n} \{ \|Af - m_\delta\|_2^2 + \alpha \|D^T f\|_1 \}$$

Case I: D is an orthonormal base

Examples: $D = I$ (canonical basis), Haar wavelet basis, orhtogonal polinomials, Fourier goniometric polynomials, ...
Properties: $range(D) = range(D^T) = \mathbb{R}^n$: every vector f can be written as $f = Dw$, and the weights w can be found as $w = D^T f$.
Two possible formulations:

Sparse deconvolution - synthesis formulation

$$w^* = \arg\min_{w \in \mathbb{R}^m} \{\|ADw - m_\delta\|_2^2 + \alpha\|w\|_1\}$$

Sparse deconvolution - analysis formulation

$$f^* = \arg\min_{f \in \mathbb{R}^n} \{\|Af - m_\delta\|_2^2 + \alpha\|D^T f\|_1\}$$

Result: if D is orthogonal, the sparsest solution f^* can be obtained by solving the analysis problem or, equivalently, by finding its weights w^* (solution of the synthesis) and computing $f^* = Dw^*$.

Case II: D is a tight frame

Description: $D \in \mathbb{R}^{n \times m}$, $m > n$, $DD^T = I$. This implies that $\forall f \in \mathbb{R}^n$ $\exists w \in \mathbb{R}^m$ s.t. $Dw = f$. Warning: this representation is in general not unique.

Case II: D is a tight frame

Description: $D \in \mathbb{R}^{n\times m}$, $m > n$, $DD^T = I$. This implies that $\forall f \in \mathbb{R}^n$ $\exists w \in \mathbb{R}^m$ s.t. $Dw = f$. Warning: this representation is in general not unique.

Examples: shearlets, curvelets, or more easily any juxtaposition of multiple (orthonormal) bases: $D = \frac{1}{\sqrt{2}}[D_1 \ D_2]$, where D_1 and D_2 are orthogonal matrices.

Case II: D is a tight frame

Description: $D \in \mathbb{R}^{n \times m}$, $m > n$, $DD^T = I$. This implies that $\forall f \in \mathbb{R}^n$ $\exists w \in \mathbb{R}^m$ s.t. $Dw = f$. Warning: this representation is in general not unique.

Examples: shearlets, curvelets, or more easily any juxtaposition of multiple (orthonormal) bases: $D = \frac{1}{\sqrt{2}}[D_1 \ D_2]$, where D_1 and D_2 are orthogonal matrices.

In this case, synthesis and analysis do not coincide:

Sparse deconvolution - synthesis formulation

$$f_S^* = Dw^*, \qquad w^* = \arg\min_{w \in \mathbb{R}^m} \{\|ADw - m_\delta\|_2^2 + \alpha \|w\|_1\}$$

Sparse deconvolution - analysis formulation

$$f_A^* = \arg\min_{f \in \mathbb{R}^n} \{\|Af - m_\delta\|_2^2 + \alpha \|D^T f\|_1\}$$

Case II: D is a tight frame

Description: $D \in \mathbb{R}^{n \times m}$, $m > n$, $DD^T = I$. This implies that $\forall f \in \mathbb{R}^n$ $\exists w \in \mathbb{R}^m$ s.t. $Dw = f$. Warning: this representation is in general not unique.

Examples: shearlets, curvelets, or more easily any juxtaposition of multiple (orthonormal) bases: $D = \frac{1}{\sqrt{2}}[D_1 \; D_2]$, where D_1 and D_2 are orthogonal matrices.

In this case, synthesis and analysis do not coincide:

Sparse deconvolution - synthesis formulation

$$f_S^* = Dw^*, \qquad w^* = \arg\min_{w \in \mathbb{R}^m} \{ \|ADw - m_\delta\|_2^2 + \alpha \|w\|_1 \}$$

Sparse deconvolution - analysis formulation

$$f_A^* = \arg\min_{f \in \mathbb{R}^n} \{ \|Af - m_\delta\|_2^2 + \alpha \|D^T f\|_1 \}$$

It is possible to prove that, in general, $f_A^* \neq f_S^*$. What is the best one? It depends on the problem.

Iterative Soft Thresholding Algorithm

ISTA - introduction

The Iterative Soft Thresholding Algorithm is a strategy to approximate the solution both of the synthesis and of the analysis problems.

ISTA - introduction

The Iterative Soft Thresholding Algorithm is a strategy to approximate the solution both of the synthesis and of the analysis problems.
It was originally proved for linear inverse problems in [Daubechies, Defrise, DeMol, 2004] in a continuous setting. It can also be interpreted in the framework of proximal operator minimization, a well known technique in convex optimization.

ISTA - introduction

The Iterative Soft Thresholding Algorithm is a strategy to approximate the solution both of the synthesis and of the analysis problems.
It was originally proved for linear inverse problems in [Daubechies, Defrise, DeMol, 2004] in a continuous setting. It can also be interpreted in the framework of proximal operator minimization, a well known technique in convex optimization.

Main idea: if w^* is the solution of the synthesis problem,

$$w^* = \underset{w \in \mathbb{R}^m}{\arg\min} \{ \|ADw - m_\delta\|_2^2 + \alpha \|w\|_1 \}.$$

Then, w^* solves the fixed point equation:

$$w^* = \mathsf{S}_\alpha (w^* + D^T A^T m_\delta - D^T A^T A D w^*)$$

ISTA - introduction

The Iterative Soft Thresholding Algorithm is a strategy to approximate the solution both of the synthesis and of the analysis problems.
It was originally proved for linear inverse problems in [Daubechies, Defrise, DeMol, 2004] in a continuous setting. It can also be interpreted in the framework of proximal operator minimization, a well known technique in convex optimization.

Main idea: if w^* is the solution of the synthesis problem,

$$w^* = \arg\min_{w \in \mathbb{R}^m} \{\|ADw - m_\delta\|_2^2 + \alpha\|w\|_1\}.$$

Then, w^* solves the fixed point equation:

$$w^* = \mathsf{S}_\alpha(w^* + D^T A^T m_\delta - D^T A^T ADw^*)$$

ISTA - Synthesis problem

$$\begin{cases} \text{select } w^{(0)} \in \mathbb{R}^m \\ \text{update } w^{(k+1)} = \mathsf{S}_\alpha(w^{(k)} + D^T A^T m_\delta - D^T A^T ADw^{(k)}) \end{cases}$$

ISTA - continuation

Result: under suitable conditions on A and D, for any choice of $w^{(0)}$, the sequence $w^{(k)}$ converges to w^*. ([D,D,D, 2004]).

ISTA - continuation

Result: under suitable conditions on A and D, for any choice of $w^{(0)}$, the sequence $w^{(k)}$ converges to w^*. ([D,D,D, 2004]).
It is also possible to find a formulation of ISTA allowing to approximate the solution of the analysis formulation:

ISTA - Analysis problem

$$\begin{cases} \text{select } f^{(0)} \in \mathbb{R}^m \\ \text{update } f^{(k+1)} = D S_\alpha(D^T(f^{(k)} + A^T m_\delta - A^T A f^{(k)})) \end{cases}$$

ISTA - continuation

Result: under suitable conditions on A and D, for any choice of $w^{(0)}$, the sequence $w^{(k)}$ converges to w^*. ([D,D,D, 2004]).
It is also possible to find a formulation of ISTA allowing to approximate the solution of the analysis formulation:

ISTA - Analysis problem

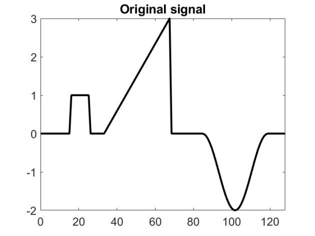
$$\begin{cases} \text{select } f^{(0)} \in \mathbb{R}^m \\ \text{update } f^{(k+1)} = DS_\alpha(D^T(f^{(k)} + A^T m_\delta - A^T A f^{(k)})) \end{cases}$$

It is an easy exercise to check that, if D is orthogonal, the iterations of ISTA for the analysis and synthesis problems coincide: if $f^{(0)} = Dw^{(0)}$, then $f^{(k)} = Dw^{(k)}$ for every k. As a consequence, $f_A^* = Dw^* = f_S^*$, as we already knew.

Implementation

Example 1

1. Consider the same signal f as in Lecture 7. Let $n = 128$. Create the PSF by triple autoconvolution of [111]. Add a Gaussian noise, $\delta = 0.2$.



2. Use ISTA (1000 iterations) to find $f_1 = \arg\min\{\|Af - m\|_2^2 + \alpha\|f\|_1\}$
3. Use ISTA (1000 iterations) to find $f_2 = \arg\min\{\|Af - m\|_2^2 + \alpha\|H^T f\|_1\}$, being H the matrix of the Haar wavelet transform (an orthogonal matrix that can be obtained by the provided function `haarmtx(n,levels)`)
4. Combine the canonical basis and H to form a tight frame D. Compute f_S^* and f_A^* performing 1000 iterations of the two formulations of ISTA.

An example from a recent paper

In [Elad, Figueiredo, Ma, 2010] the authors show the following result obtained by ISTA algorithm applied on a familiar picture.

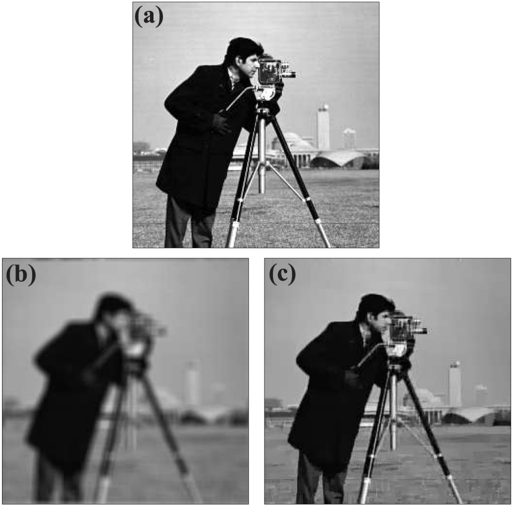
An example from a recent paper

In [Elad, Figueiredo, Ma, 2010] the authors show the following result obtained by ISTA algorithm applied on a familiar picture.

(a) original picture

(b) blurred version (no noise!)

(c) deblurring by ISTA with a tight frame (redundant Haar wavelets)



Dictionary Learning

A more challenging problem

Compressed Sensing (warning: no longer deconvolution!)

Unknown: $w \in \mathbb{R}^m$; Datum: $f = Dw + \varepsilon$;

A priori knowledge: $\|w\|_1 \ll n$, i.e., f is sparse with respect to a dictionary D, which we know.

A more challenging problem

Compressed Sensing (warning: no longer deconvolution!)

Unknown: $w \in \mathbb{R}^m$; Datum: $f = Dw + \varepsilon$;

A priori knowledge: $\|w\|_1 \ll n$, i.e., f is sparse with respect to a dictionary D, which we know.

In **Dictionary Learning** we suppose that the signal is sparse with respect to an unknown dictionary D. Is it possible to learn together D and f (or equivalently w) from m?

A more challenging problem

Compressed Sensing (warning: no longer deconvolution!)

Unknown: $w \in \mathbb{R}^m$; Datum: $f = Dw + \varepsilon$;

A priori knowledge: $\|w\|_1 \ll n$, i.e., f is sparse with respect to a dictionary D, which we know.

In **Dictionary Learning** we suppose that the signal is sparse with respect to an unknown dictionary D. Is it possible to learn together D and f (or equivalently w) from m?

No. That is why we need more data!

A more challenging problem

Compressed Sensing (warning: no longer deconvolution!)

Unknown: $w \in \mathbb{R}^m$; Datum: $f = Dw + \varepsilon$;

A priori knowledge: $\|w\|_1 \ll n$, i.e., f is sparse with respect to a dictionary D, which we know.

In **Dictionary Learning** we suppose that the signal is sparse with respect to an unknown dictionary D. Is it possible to learn together D and f (or equivalently w) from m?

No. That is why we need more data!

Dictionary learning - Morozov formulation

Data: $\{f_j\}_{j=1}^N \subset \mathbb{R}^n$;

Unknowns: $\{w_j\}_{j=1}^N \subset \mathbb{R}^m$, $D \in \mathbb{R}^{n \times m}$

$$\min_{D, w_1, \dots, w_N} \left\{ \sum_{j=1}^N \|w_j\|_1 \quad s.t. \quad \|Dw_j - f_j\|_2^2 \leq \delta \quad j = 1, \dots, N \right\}$$

Alternative formulation

It is possible to provide an alternative formulation of the DL problem.

Alternative formulation

It is possible to provide an alternative formulation of the DL problem.
Collect all the training data $\{f_j\}_{j=1}^N$ in $F \in \mathbb{R}^{n \times N}$, $F = [f_1, \dots, f_N]$.

Alternative formulation

It is possible to provide an alternative formulation of the DL problem. Collect all the training data $\{f_j\}_{j=1}^N$ in $F \in \mathbb{R}^{n \times N}$, $F = [f_1, \dots, f_N]$. The unknowns are $D \in \mathbb{R}^{n \times m}$ and the (hopefully sparse) weight vectors $\{w_i\}$ representing the training data $\{f_j\}$. We can collect $\{w_j\}_{j=1}^N$ in a matrix $W \in \mathbb{R}^{m \times N}$, $W = [w_1, \dots, w_N]$

Alternative formulation

It is possible to provide an alternative formulation of the DL problem. Collect all the training data $\{f_j\}_{j=1}^N$ in $F \in \mathbb{R}^{n\times N}$, $F = [f_1, \dots, f_N]$. The unknowns are $D \in \mathbb{R}^{n\times m}$ and the (hopefully sparse) weight vectors $\{w_i\}$ representing the training data $\{f_j\}$. We can collect $\{w_j\}_{j=1}^N$ in a matrix $W \in \mathbb{R}^{m\times N}$, $W = [w_1, \dots, w_N]$

Dictionary learning - (Tikhonov) matricial formulation

$$\min_{(D,W)\in\mathbb{R}^{n\times m}\times\mathbb{R}^{m\times N}} \{\|DW - F\|_F^2 + \alpha\|W\|_{1,1}\}$$

where we used the following matricial norms:

- the Frobenius norm $\|\cdot\|_F$: $\|A\|_F^2 = \sum_{i=1}^n \sum_{j=1}^m A_{i,j}^2$
- the column-wise 1 norm $\|\cdot\|_{1,1}$: $\|A\|_{1,1} = \sum_{j=1}^m \|A_j\|_1$.

Alternative formulation

It is possible to provide an alternative formulation of the DL problem. Collect all the training data $\{f_j\}_{j=1}^N$ in $F \in \mathbb{R}^{n\times N}$, $F = [f_1, \dots, f_N]$. The unknowns are $D \in \mathbb{R}^{n\times m}$ and the (hopefully sparse) weight vectors $\{w_i\}$ representing the training data $\{f_j\}$. We can collect $\{w_j\}_{j=1}^N$ in a matrix $W \in \mathbb{R}^{m\times N}$, $W = [w_1, \dots, w_N]$

Dictionary learning - (Tikhonov) matricial formulation

$$\min_{(D,W)\in\mathbb{R}^{n\times m}\times\mathbb{R}^{m\times N}} \{\|DW - F\|_F^2 + \alpha\|W\|_{1,1}\}$$

where we used the following matricial norms:

- the Frobenius norm $\|\cdot\|_F$: $\|A\|_F^2 = \sum_{i=1}^{n}\sum_{j=1}^{m} A_{i,j}^2$
- the column-wise 1 norm $\|\cdot\|_{1,1}$: $\|A\|_{1,1} = \sum_{j=1}^{m} \|A_j\|_1$.

Such function of D and W is convex with respect to each of the two variables, but not jointly convex.

The K-SVD algorithm

Input: training data F, initial guess $D^{(0)}$ for the dictionary, $i = 1$.
Iteration: while stopping criterion not satisfied do

1. update the weights W according to $D^{(i-1)}$, i.e.:
$$\text{find } W^{(i)} = \underset{W \in \mathbb{R}^{m \times N}}{\arg\min} \{ \|F - D^{(i-1)} W\|_F^2 + \alpha \|W\|_{1,1} \};$$
this problem can be decoupled in N compressed sensing problems.
2. update D (and again W), preserving the sparsity of W.
This is done each column at a time. The contribution of each column D_j to the term $\|DW - F\|_F^2$ is represented by a rank one matrix $D_j W^j$, where W^j is the $j-$th row of W. Such contribution can be optimized by means of a singular value decomposition, updating D_j and W^j. In order to preserve sparsity, we look for an update restricted to the actual support of W^j.

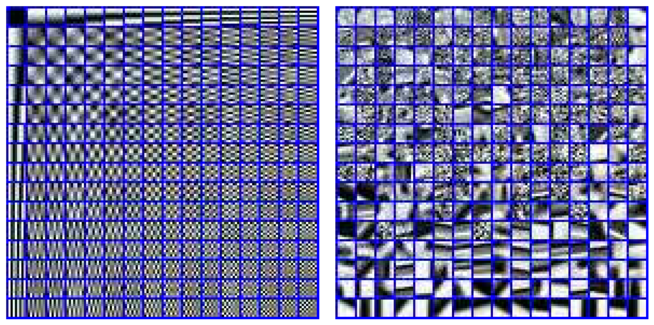
All the details can be found in [Chen, Needell, 2014].

Two examples from recent papers

A very clever idea to apply DL in image denoising is the local patch processing algorithm:

- divide the noisy picture in several smaller patches (e.g.: original picture = 256x256, divide it in 256 patches of size 8x8
- use the 8x8 patches as a training set to learn a dictionary which will allow for the sparse reconstructions of (8x8) pictures
- use K-SVD for the combined learning of the dictionary and of the sparse representation of each patch
- collect all the final patches into a denoised version of the picture

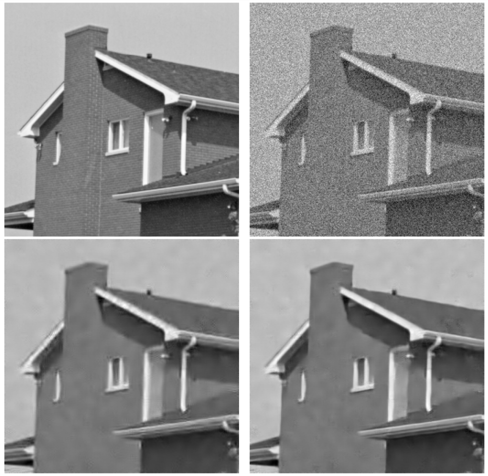
Two examples from recent papers



The initial guess $D^{(0)}$ for the dictionary - The learn dictionary at the last iteration

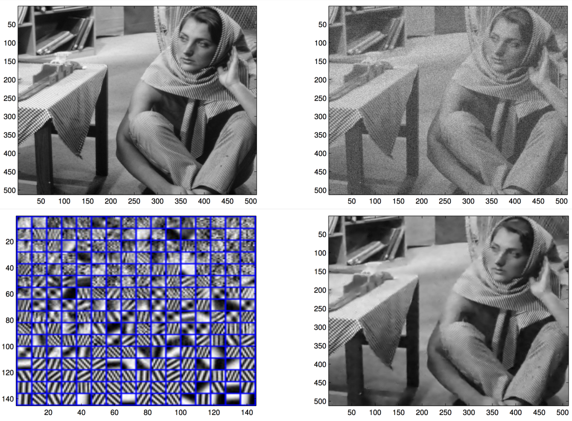
(pictures from [Elad, Figueiredo, Ma, 2010])

Two examples from recent papers



Original picture - Noisy version
Deblurring with initial dictionary - Deblurring with learnt dictionary
(pictures from [Elad, Figueiredo, Ma, 2010])

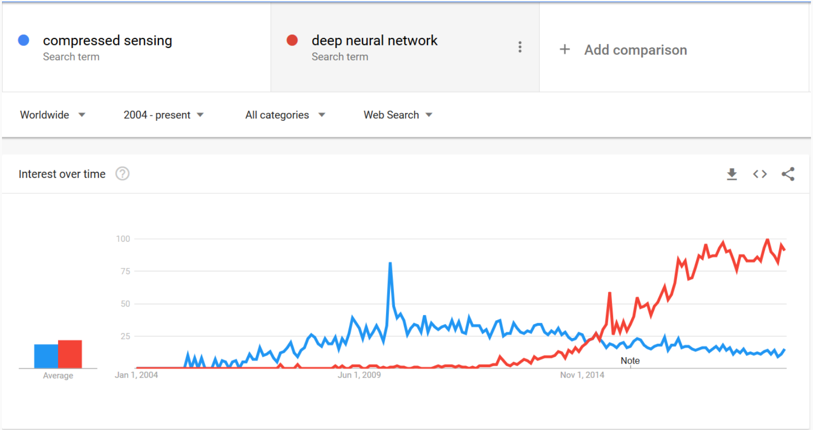
Two examples from recent papers



Original picture - Noisy version
Learnt dictionary - Deblurring with learnt dictionary
(pictures from [Chen, Needell, 2014])

What is the future of Compressed Sensing?

What is the future of Compressed Sensing?



`https://trends.google.com/trends/explore?date=all&q=compressed%20sensing,deep%20neural%20network`