Analysis versus Synthesis in Signal Priors

Ron Rubinstein

Computer Science Department
Technion – Israel Institute of Technology
Haifa, 32000 Israel

October 2006

Agenda

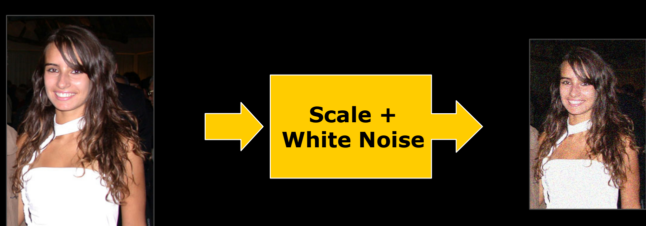
- **Inverse Problems – Two Bayesian Approaches**
  Introducing MAP-Analysis and MAP-Synthesis
- Geometrical Study: Why there is no Equivalence
  Geometry reveals underlying gaps
- From Theoretical Gap to Practical Results
  Finding where the differences hurt the most
- Algebra at Last: Characterizing the Gap
  Bound provides new insight
- What Next: Current and Future Work

Inverse Problem Formulation

- We consider the following general inverse problem:

$$\underline{y} = T\{\underline{x}\} + \underline{n}$$

- T is the degradation operator (not necessarily linear)
- Additive white Gaussian noise: $\underline{n} \sim \exp\left\{-\alpha \|\underline{n}\|_2^2\right\}$



Bayesian Estimation

- The statistical model: $P(\underline{y} \mid \underline{x}) = Const \cdot \exp\left\{-\alpha \left\|\underline{y} - T\{\underline{x}\}\right\|_2^2\right\}$
- Maximum A Posterior (MAP) estimator

$$\begin{aligned}\hat{x}_{MAP} &= \arg\max_{\underline{x}} P(\underline{x} \mid \underline{y}) = \arg\max_{\underline{x}} P(\underline{y} \mid \underline{x}) P(\underline{x}) \\ &= \arg\min_{\underline{x}} \left\{ \left\|\underline{y} - T\{\underline{x}\}\right\|_2^2 - \lambda \cdot \log P(\underline{x}) \right\}\end{aligned}$$

$$\hat{x}_{MAP} = \arg\min_{\underline{x}} \left\{ \left\|\underline{y} - T\{\underline{x}\}\right\|_2^2 + \lambda R\{\underline{x}\} \right\}$$

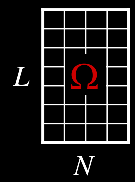
Analysis Priors ("MAP-Analysis")

- Analysis priors suggest a regularization of the form:

$$\underline{x}_{MAP-A} = \underset{\underline{x}}{\arg\min} \left\{ \left\| \underline{y} - T\{\underline{x}\} \right\|_2^2 + \lambda \cdot \left\| \Omega \underline{x} \right\|_p^p \right\} \qquad \left(\Omega \in \mathbb{R}^{LxN} \right)$$

- The *analyzing operator* Ω can be of any size, but is usually overcomplete $(L \geq N)$.
- Typically $1 \leq p \leq 2$
- This regularization is explained by the prior

$$P(\underline{x}) = Const \cdot \exp\left\{ -\alpha \left\| \Omega \underline{x} \right\|_p^p \right\}$$

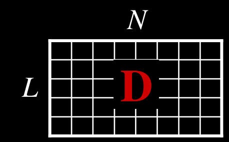


Synthesis Priors ("MAP-Synthesis")

- Synthesis priors stem from the concept of sparse representation in overcomplete dictionaries (Chen, Donoho & Saunders):

$$\underline{x}_{MAP-S} = \mathbf{D} \cdot \underset{\underline{\gamma}}{\arg\min} \left\{ \left\| \underline{y} - T\{\mathbf{D}\underline{\gamma}\} \right\|_2^2 + \lambda \cdot \left\| \underline{\gamma} \right\|_p^p \right\}$$

- $\mathbf{D}$ is generally overcomplete $(L \geq N)$:
- Typically $0 \leq p \leq 1$
- Can also be explained in terms of MAP estimation.



Analysis versus Synthesis

- The two approaches are algebraically very similar:

$$\underline{x}_{MAP-A} = \underset{\underline{x}}{\arg\min} \left\{ \left\| \underline{y} - T\{\underline{x}\} \right\|_2^2 + \lambda \cdot \left\| \Omega \underline{x} \right\|_p^p \right\}$$

$$\underline{x}_{MAP-S} = \mathbf{D} \cdot \underset{\underline{\gamma}}{\arg\min} \left\{ \left\| \underline{y} - T\{\mathbf{D}\underline{\gamma}\} \right\|_2^2 + \lambda \cdot \left\| \underline{\gamma} \right\|_p^p \right\}$$

- Both methods are motivated by the same principal of *representational sparsity.*

Analysis versus Synthesis

- MAP-Synthesis:
  - Supported by empirical evidence (Olshausen & Field)
  - Constructive form
  - Seems to benefit from high redundancy
  - Supported by a wealth of theoretical results: Donoho & Huo, Elad & Bruckstein, Gribonval & Nielsen, Fuchs, Donoho Elad & Temlyakov, Tropp…
- MAP-Analysis:
  - Significantly simpler to solve
  - Potentially more stable (all atoms contribute)

Some Algebra: Could the two be Related?

- Using the pseudo-inverse, the two formulations can *almost* be brought to the same form:

$$\Omega\underline{x}=\underline{\gamma} \quad \Leftrightarrow \quad \underline{x}=\Omega^{+}\underline{\gamma}$$

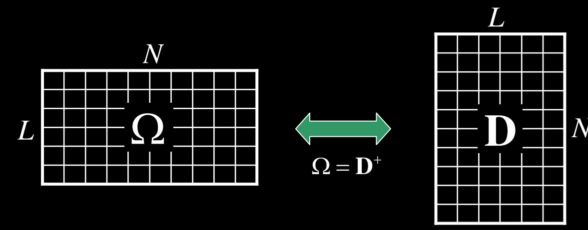
$$\underline{x}_{MAP-A} = \arg\min_{\underline{x}} \left\{ \left\|\underline{y}-T\{\underline{x}\}\right\|_2^2 + \lambda\cdot\left\|\Omega\underline{x}\right\|_p^p \right\}$$

$$\underline{x}_{MAP-A} = \Omega^{+}\cdot\arg\min_{\underline{\gamma}} \left\{ \left\|\underline{y}-T\{\Omega^{+}\underline{\gamma}\}\right\|_2^2 + \lambda\cdot\left\|\underline{\gamma}\right\|_p^p \right\} \quad s.t.\ \ \Omega\Omega^{+}\underline{\gamma}=\underline{\gamma}$$

- This is precisely the MAP-Synthesis formulation, but with the added constraint since $\underline{\gamma}$ must be in the column-span of Ω in the MAP-Analysis case.
- Though sometimes close, the two solutions are generally different.

Specific Cases of Equivalence

- In the square case, as well as the under-complete denoising case, the two formulations become equivalent.



- The pseudo-inverse also obtains equivalence in the overcomplete p=2 case. For other values of p, however, simulations show that the pseudo-inverse relation fails.

Analysis versus Synthesis

- Contradicting approaches in literature:

"…MAP-Synthesis is very 'trendy'. It is a promising approach and provides superior results over MAP-Analysis"

"…The two methods are much closer. In fact, one can be used to approximate the other."

- Are the two prior types related?
- Which approach is better?

Agenda

- Inverse Problems – Two Bayesian Approaches
  Introducing MAP-Analysis and MAP-Synthesis
- **Geometrical Study: Why there is no Equivalence**
  Geometry reveals underlying gaps
- From Theoretical Gap to Practical Results
  Finding where the differences hurt the most
- Algebra at Last: Characterizing the Gap
  Bound provides new insight
- What Next: Current and Future Work

The General Problem Is Difficult

- Searching for the most general relationship, we find ourselves with a large number of unknowns:
  - The relation between Ω and $\mathbf{D}$ is unknown.
  - The regularizing parameter λ may not be the same for the two problems.
  - Even the value of p may vary between the two approaches.

Simplification

- Concentrate on p=1. Motivation for this choice:
  - The "meeting point" between the two approaches.
  - One of the most common choices for both methods, provides a combination of convexity and robustness.
  - For MAP-Synthesis, it is known to be a good approximation of p=0 (true sparsity) in many cases.
- Replace regularization with a constraint:

$$\underline{x}_{MAP-A}(a) = \arg\min_{\underline{x}} \|\Omega\underline{x}\|_1 \quad s.t. \quad \|\underline{y} - T\{\underline{x}\}\|_2^2 \le a$$

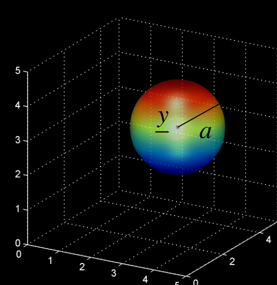
$$\underline{x}_{MAP-S}(a) = \mathbf{D}\cdot\arg\min_{\underline{\gamma}} \|\underline{\gamma}\|_1 \quad s.t. \quad \|\underline{y} - T\{\mathbf{D}\underline{\gamma}\}\|_2^2 \le a$$

A Geometrical Analysis

- Both problems seek a solution over the same domain: a region of "radius" a about the input.
- In this region, each method aims to minimize a different target function:

$$f_{MAP-A}(\underline{x}) = \|\Omega\underline{x}\|_1$$

$$f_{MAP-S}(\underline{x}) = \min_{\{\underline{\gamma}:\, \underline{x}=\mathbf{D}\underline{\gamma}\}} \|\underline{\gamma}\|_1$$

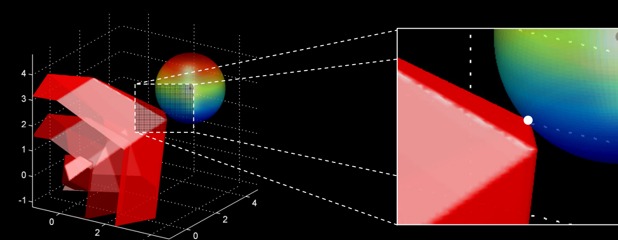


A Geometrical Analysis

- The iso-surfaces of the MAP-Analysis target function form a set of coinciding, centro-symmetric polytopes:

$$\{f_{MAP-A}(\underline{x}) \le c\} = \{\underline{x} : \|\Omega\underline{x}\|_1 \le c\}$$

- Imagine a very small iso-surface, being inflated it until first touching the ball; this will be the MAP-Analysis solution!



The MAP Defining Polytopes

- A similar description applies to MAP-Synthesis, where

$$\left\{ f_{MAP-S}\left(\underline{x}\right) \le c \right\} = \mathbf{D} \cdot \left\{ \underline{\gamma} : \left\|\underline{\gamma}\right\|_1 \le c \right\}$$

- For both methods, the coinciding polytopes are *similar*, and can be determined from the iso-surface with $c = 1$:

$$\left\{ f_{MAP}\left(\underline{x}\right) \le c \right\} = c \cdot \left\{ f_{MAP}\left(\underline{x}\right) \le 1 \right\}$$

The MAP Defining Polytopes

- Conclusion: we can characterize each of the MAP priors using a single polytope!
- We define the *MAP defining polytopes* as

MAP-Analysis Defining Polytope

$$\Psi\left(\Omega\right) = \left\{ \underline{x} : \left\|\Omega\underline{x}\right\|_1 \le 1 \right\}$$

MAP-Synthesis Defining Polytope

$$\Phi\left(\mathbf{D}\right) = \mathbf{D} \cdot \left\{ \underline{\gamma} : \left\|\underline{\gamma}\right\|_1 \le 1 \right\}$$

- We now have a basis for comparing the two approaches.

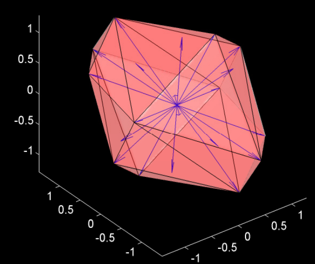
MAP-Synthesis Defining Polytope

- Obtained as the convex hull of the columns of $\mathbf{D}$ and their antipodes, $\left\{ \pm d_i \right\}$.

$$\mathbf{D} = \begin{pmatrix} 1.09 & 1.43 & -0.25 & -1.30 & -0.42 & 0.32 & -0.56 & -0.24 & 0.90 & 0.47 \\ -0.73 & 0.32 & 1.20 & -0.66 & 1.51 & -0.51 & 0.28 & -0.19 & -0.22 & 0.55 \\ -0.56 & 0.32 & 0.98 & 0.48 & 0.37 & 1.29 & 0.76 & -0.95 & 0.39 & -0.58 \end{pmatrix}$$

Redundant!

Conclusion: any row in $\mathbf{D}$ which is the convex combination of the remaining columns (and their antipodes) can be discarded.



MAP-Analysis Defining Polytope

- Highly complex polytope, whose faces are obtained as null-spaces of rows in Ω.
- Some properties of this polytope:
  - Exponential worst-case vertex count: $N_v = \Theta\binom{L}{N-1}$

    Also the expected number of vertices when the directions of the rows in Ω are uniformly distributed.
  - Highly regular structure

    Faces are arranged in very specific structures. Highly organized neighborliness patterns.

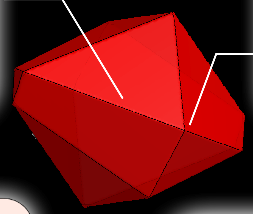
MAP-Analysis Defining Polytope

Edge loops

The edges are arranged in planar loops about the origin.

Neighborliness

Every vertex has exactly $2(N-1)=4$ neighbors.



Vertex count

$$\frac{1}{2}N_v=\binom{L}{N-1}=\binom{4}{2}=6$$

$$\Omega=\begin{pmatrix} -0.204 & -0.905 & -0.005 \\ 0.111 & -0.324 & 0.608 \\ 0.860 & -0.242 & -0.432 \\ -0.455 & -0.131 & -0.667 \end{pmatrix}$$

Comparison: MAP Defining Polytopes

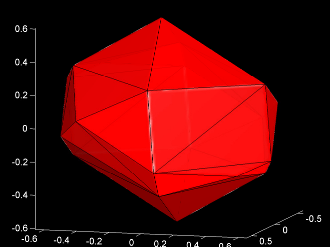
	MAP-Analysis	MAP-Synthesis
Expected Vertex #	High: $O\binom{L}{N-1}$	Low: $O(L)$
Neighborliness (u,v are non-antipodes)	Low: $P\{e(u,v)\}\to 0$ $as\ N\to\infty$	High: $P\{e(u,v)\}\to 1$ $as\ N\to\infty$
Regularity	High	None

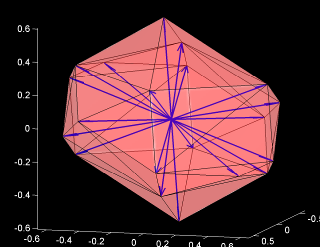
- The neighborliness property for MAP-Synthesis defining polytopes has been recently proven by Donoho, and is obtained for dictionaries in which $L=O(N)$, and under certain randomness assumptions.

Translating Analysis to Synthesis

$$\Omega=\begin{pmatrix} 0.88 & -0.09 & 0.06 \\ 0.16 & 0.20 & 0.48 \\ -0.30 & -0.10 & -0.42 \\ -0.32 & -0.28 & 0.77 \\ -0.08 & 0.93 & 0.09 \end{pmatrix} \Rightarrow \mathbf{D}=\begin{pmatrix} -.08 & -.08 & .04 & -.05 & -.24 & -.46 & -.60 & -.31 & -.52 & .56 \\ -.62 & -.66 & .61 & -.06 & -.49 & .45 & -.08 & -.58 & -.08 & .03 \\ .29 & .21 & .24 & .61 & .29 & -.03 & .24 & .08 & .39 & .25 \end{pmatrix}$$

Vertices of the MAP-Analysis defining polytope





Analysis as a Subset of Synthesis

- Any MAP-Analysis problem can be reformulated as an identical MAP-Synthesis one.
- However, the translation leads to an *exponentially large* dictionary; a feasible equivalence does not exist!
- The other direction does not hold: many MAP-Synthesis problems have no equivalent MAP-Analysis form.

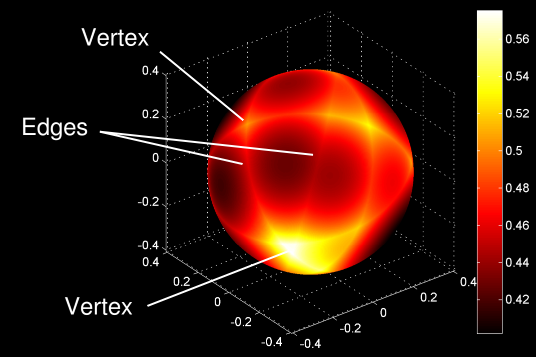
Theorem: Any L_1 MAP-Analysis problem has an equivalent L_1 MAP-Synthesis one. The reverse is not true.

Favorable MAP Signals

- For MAP-Synthesis, we think of the dictionary atoms as the “ideal” signals. Other favorable signals are sparse combinations of these signals.
- What are the favorable MAP-Analysis signals?
- Observation: for MAP-Synthesis, the dictionary atoms are the vertices of its defining polytope, and their sparse combinations are its low-dimensional faces.
- The favorable signals of a MAP prior can be found on its low-dimensional faces!

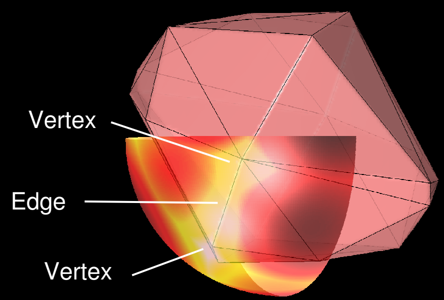
Favorable MAP Signals

- Sample MAP distribution on the unit sphere:



Favorable MAP Signals

- The MAP favorable signals are located on the *low-dimensional faces* of the MAP defining polytope.
- This is, however, only a *necessary* condition!



Intermediate Summary

- We have studied the two formulations from a geometrical perspective. This viewpoint has led to the following conclusions:
  - The geometrical structure underlying the two formulations is substantially different (of asymptotic nature).
  - MAP-Analysis can only represent a small part of the problems representable by MAP-Synthesis.
- But how significant are these differences in practice?

Agenda

- Inverse Problems – Two Bayesian Approaches
  Introducing MAP-Analysis and MAP-Synthesis
- Geometrical Study: Why there is no Equivalence
  Geometry reveals underlying gaps
- **From Theoretical Gap to Practical Results**
  Finding where the differences hurt the most
- Algebra at Last: Characterizing the Gap
  Bound provides new insight
- What Next: Current and Future Work

Synthetic Experiments: Setup

- Dictionary: 128x256 Identity-Hadamard, $\mathbf{D} = \frac{1}{\sqrt{2}}[\mathbf{I} \quad \mathbf{H}]$
  Analysis operator: the pseudo-inverse, $\Omega = \mathbf{D}^{T}$
- Motivation for this choice –
  - Simple two-ortho structure for both operators. Since $\mathbf{D}$ is a tight-frame, pseudo-inversion is obtained through direct matrix transpose.
  - The dictionary is a near-optimal Grassmanian frame, and so is a preferred choice for MAP-Synthesis.
- Reminder: the Hadamard transform is given by

$$\mathbf{H}_2 = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}, \quad \mathbf{H}_{k+1} = \mathbf{H}_2 \otimes \mathbf{H}_k$$

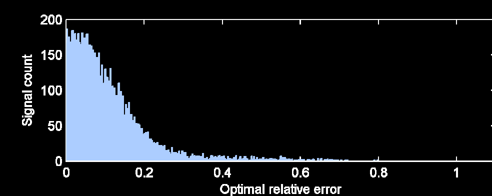
Synthetic Experiments: Setup

- Dataset:
  - 10,000 MAP-Analysis principal signals
  - 256 MAP-Synthesis principal signals
  - Additional sets of sparse MAP-Synthesis signals (to compensate for the small number of principal signals): 1,000 2-atom, 1,000 3-atom, and so on up to 12-atom.
- Procedure:
  - Generate noisy versions of all signals.
  - Apply both MAP methods to the noisy signals, setting a to its optimal value for each signal individually (this value was determined by brute-force search).
  - Collect the optimal errors obtained by each method for these signals.

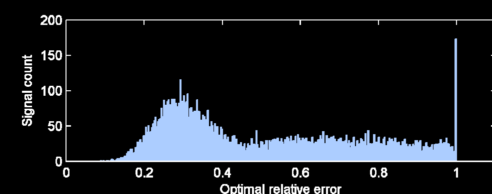
Synthetic Experiments: Results

- Distribution of optimal errors obtained for MAP-Analysis principal signals:

MAP-Analysis Denoising:



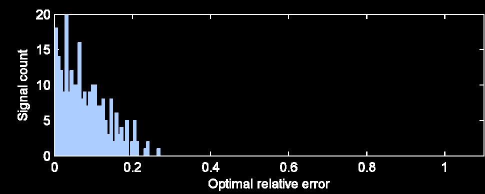
MAP-Synthesis Denoising:



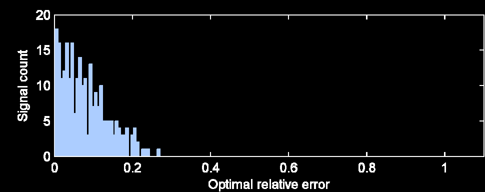
Synthetic Experiments: Results

- Distribution of optimal errors obtained for MAP-Synthesis principal signals:

MAP-Analysis Denoising:



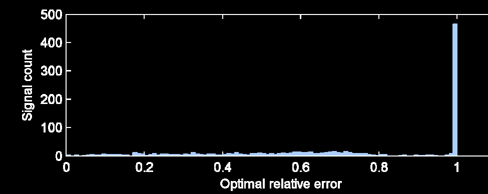
MAP-Synthesis Denoising:



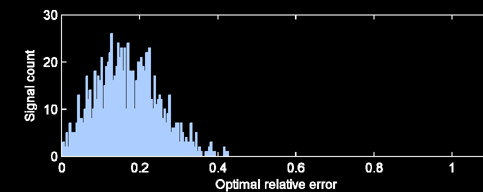
Synthetic Experiments: Results

- Distribution of optimal errors obtained for 2-atom MAP-Synthesis signals:

MAP-Analysis Denoising:



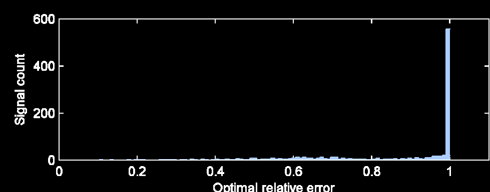
MAP-Synthesis Denoising:



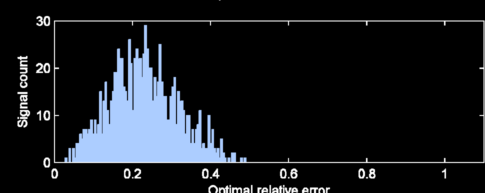
Synthetic Experiments: Results

- Distribution of optimal errors obtained for 3-atom MAP-Synthesis signals:

MAP-Analysis Denoising:

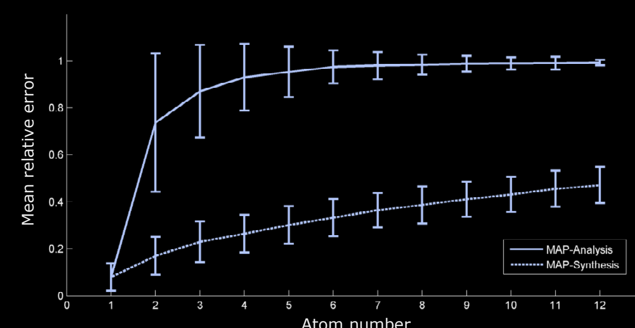


MAP-Synthesis Denoising:



Synthetic Experiments: Results

- Summary of results for MAP-Synthesis favorable signals (mean denoising error vs. number of atoms):



Synthetic Experiments: Discussion

- The geometrical model correctly predicted the favorable signals of each method.
- However, each method favors *different* sets of signals.
- There is a large difference in the *number* of favorable signals between the two prior forms; this is due to the asymptotical gaps between them.
- The pseudo-inverse does not bridge the gap between the two methods!

Real-World Experiments: Setup

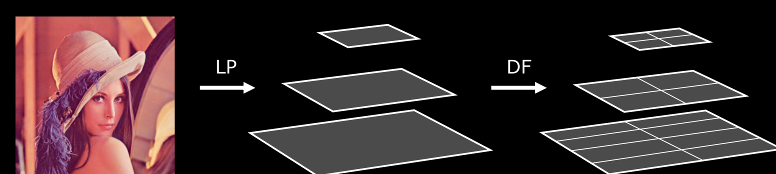
- Dictionary: overcomplete DCT, contourlet.
  Analysis operator: the pseudo-inverse (transpose)
- Motivation –
  - Commonly used in image processing
  - Tight frames
  - Variety of redundancy factors
- Dataset: standard test images (*Lenna, Barbara, Mandrill…*), rescaled to 128x128 using bilinear interpolation.
- Procedure: add white noise (PSNR=25dB), denoise using both methods, compare.

Overcomplete DCT Transform

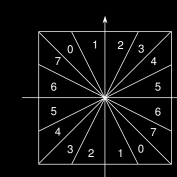
- Forward transform: block DCT with overlapping (amount of overlap may be adjusted).
- Backward transform: inverse DCT + averaging.

Contourlet Transform (Do & Vetterli)

- Forward transform: Laplacian pyramid + directional filtering (level-dependent).

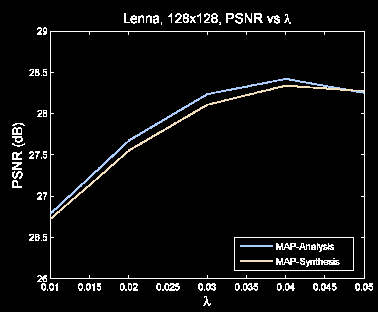


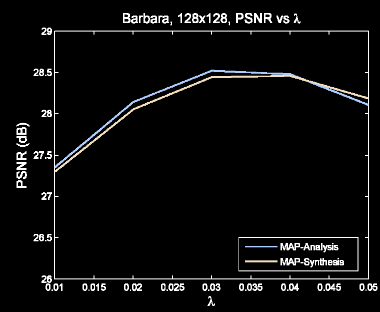
- Directional filtering partitions the image to differently oriented filtered regions:
- DF is critically-sampled (invertible).
- Backward transform: pseudo-inverse.



Real-World Experiments: Results

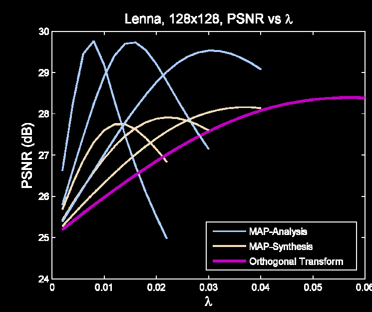
- Contourlet results (overcompleteness of 4:3):

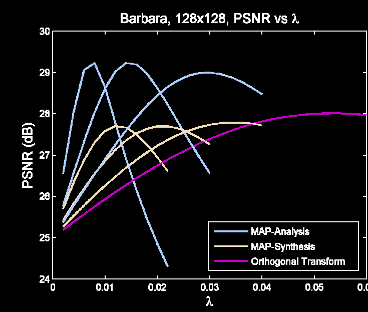




Real-World Experiments: Results

- DCT results (overcompleteness of x4, x16, x64):





Real-World Experiments: Discussion

- MAP-Analysis is beating MAP-Synthesis in every test!
- Furthermore, MAP-Analysis gains from the redundancy, while MAP-Synthesis does not.
- Conclusion: there is a real gap between the two methods in the overcomplete case.
- The gap increases with the overcompleteness.
- Despite recent trend toward MAP-Synthesis, MAP-Analysis should also be considered for inverse problem regularization.

Agenda

- Inverse Problems – Two Bayesian Approaches
  Introducing MAP-Analysis and MAP-Synthesis
- Geometrical Study: Why there is no Equivalence
  Geometry reveals underlying gaps
- From Theoretical Gap to Practical Results
  Finding where the differences hurt the most
- **Algebra at Last: Characterizing the Gap**
  Bound provides new insight
- What Next: Current and Future Work

Some Algebra

- Consider the two methods in the following denoising setup:

$$\underline{x}_{MAP-A} = \arg\min_{\underline{x}} \left\{ \tfrac{1}{2}\left\|\underline{y}-\underline{x}\right\|_2^2 + \lambda \cdot \left\|\Omega\underline{x}\right\|_1 \right\}$$

$$\underline{x}_{MAP-S} = \mathbf{D} \cdot \arg\min_{\underline{\gamma}} \left\{ \tfrac{1}{2}\left\|\underline{y}-\mathbf{D}\underline{\gamma}\right\|_2^2 + \lambda \cdot \left\|\underline{\gamma}\right\|_1 \right\}$$

- Taking a gradient we obtain equations for the optimum,

$$\underline{x}_A - \underline{y} + \lambda \cdot \Omega^T sign\left(\Omega\underline{x}_A\right) = 0$$

$$\mathbf{D}^T\left(\mathbf{D}\underline{\gamma}_S - \underline{y}\right) + \lambda \cdot sign\left(\underline{\gamma}_S\right) = 0$$

Some Algebra

- Now, assume $\mathbf{D}$ is a *left-inverse* of Ω. Multiplying the second equation by Ω^T, we obtain

$$\underline{x}_A - \underline{y} + \lambda \cdot \Omega^T sign\left(\Omega\underline{x}_A\right) = 0$$

$$\mathbf{D}^T\left(\mathbf{D}\underline{\gamma}_S - \underline{y}\right) + \lambda \cdot sign\left(\underline{\gamma}_S\right) = 0$$

$$\underline{x}_S - \underline{y} + \lambda \cdot \Omega^T sign\left(\underline{\gamma}_S\right) = 0$$

-

$$\underline{x}_A - \underline{x}_S = \lambda \cdot \Omega^T \left(sign\left(\underline{\gamma}_S\right) - sign\left(\underline{\gamma}_A\right)\right)$$

Some Algebra

- We have an upper bound on the distance between the two methods (for a fixed λ):

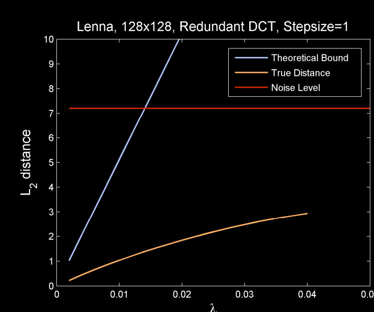
$$\left\|\underline{x}_A - \underline{x}_S\right\|_p \le 2\lambda \cdot \left\|\Omega^T\right\|_p \left\|\underline{1}\right\|_p \qquad (p \ge 1)$$

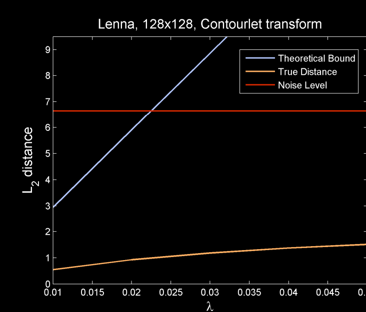
- Specifically,

$$\left\|\underline{x}_A - \underline{x}_S\right\|_2 \le 2\sqrt{L}\lambda\,\rho(\Omega) \qquad \left\|\underline{x}_A - \underline{x}_S\right\|_\infty \le 2\lambda\left\|\Omega\right\|_1$$

Numerical Simulations

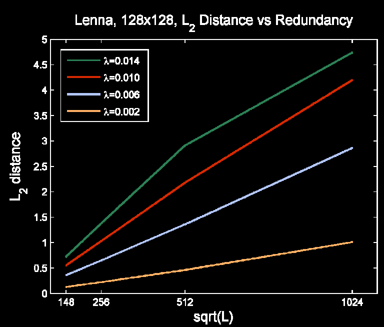
- Simulations show that the bound is very pessimistic; nonetheless, it remains informative (i.e. below the noise level) for small λ values:





Numerical Simulations

- Observation: the ℓ^2 bound predicts a linear dependence in λ and $\sqrt{L}$:



Transform	$\sqrt{L}$
Contourlet (x4/3)	148
DCT-4 (x4)	256
DCT-2 (x16)	512
DCT-1 (x64)	1024

Wrap-Up

- MAP-Analysis and MAP-Synthesis both emerge from the same Bayesian (MAP) methodology.
- The two are equivalent in simple cases, but not in the general (overcomplete) case.
- The difference between the two increases with the redundancy. For the denoising case, this distance is approximately proportional to $\sqrt{L}$.
- None of the two has a clear advantage; rather, each performs best on different types of signals. Though recent trend favors MAP-Synthesis, MAP-Analysis still remains a very worthy candidate.

Agenda

- Inverse Problems – Two Bayesian Approaches
  Introducing MAP-Analysis and MAP-Synthesis
- Geometrical Study: Why there is no Equivalence
  Geometry reveals underlying gaps
- From Theoretical Gap to Practical Results
  Finding where the differences hurt the most
- Algebra at Last: Characterizing the Gap
  Bound provides new insight
- **What Next: Current and Future Work**

Learning MAP-Analysis Operators

- Efficient algorithms exist for learning MAP-Synthesis dictionaries (Olshausen & Field, Lewicki & Sejnowski, Aharon & Elad)
- The success of MAP-Analysis motivates the development of parallel training algorithms for the analysis operator.
- Related work done by Black & Roth; assume a distribution of the form

$$P(X) = const \cdot \exp\left\{ -\sum_k \sum_i \alpha_i \, \varphi(w_i^T x_k) \right\}, \qquad \varphi(z) = \ln\left(1 + \frac{1}{2} z^2\right)$$

Learning MAP-Analysis Operators

- Suggestion: minimize the Haber-Tenorio penalty function.
- We assume a Θ-parameterized recovery method

$$\hat{x} = \mathbf{R}(y;\Theta)$$

- Given the set of training data (x_i, y_i), the Haber-Tenorio supervised learning approach finds the parameter set minimizing the recovery MSE of the data:

$$\hat{\Theta} = \mathrm{Arg}\min_{\Theta} \sum_i \left\| x_i - \mathbf{R}(y_i;\Theta) \right\|_2^2$$

Learning MAP-Analysis Operators

- Example: the K-SVD algorithm (MAP-Synthesis) can be interpreted as special case of the Haber-Tenorio approach.
- We assume a denoising method of the form

$$\mathbf{R}(y;\mathbf{D}) = \mathbf{D} \cdot \mathrm{Arg}\min_{\|\gamma\|_0 \le L} \left\| y - \mathbf{D}\gamma \right\|_2^2$$

- The training set $\{x_i\}$ is assumed to contain near-perfect signals (yet allowed a small amount of noise). Substituting these as both the clean *and* noisy signals, we obtain

$$\hat{\mathbf{D}}_{K-SVD} = \mathrm{Arg}\min_{\mathbf{D}} \mathrm{Min}_{\mathbf{\Gamma}} \left\| \mathbf{X} - \mathbf{D\Gamma} \right\|_F^2 \quad s.t. \quad \left\| \mathbf{\Gamma}_i \right\|_0 \le L$$

Learning MAP-Analysis Operators

- Can the same method be reproduced for MAP-Analysis?
- Unfortunately, no! Beginning with the denoising process

$$\mathbf{R}(y;\mathbf{D}) = \mathrm{Arg}\min_{x} \left\| x - y \right\|_2^2 \quad s.t. \quad \left\| \mathbf{\Omega} x \right\|_p^p \le L$$

- We set $\{x_i\}$ as both the clean and noisy signals, obtaining

$$\hat{\mathbf{\Omega}} = \mathrm{Arg}\min_{\mathbf{\Omega}} \mathrm{Min}_{\mathbf{Z}} \left\| \mathbf{X} - \mathbf{Z} \right\|_2^2 \quad s.t. \quad \left\| \mathbf{\Omega Z}_i \right\|_p^p \le L$$

- This is clearly useless…

Learning MAP-Analysis Operators

- The H-T approach fails when attempting to reproduce the K-SVD approximation using MAP-Analysis.
- Conclusion: we must consider pairs (x_i, y_i) after all.
- Returning to the original MAP-Analysis formulation, our target is to minimize

$$\hat{\mathbf{\Omega}} = \mathrm{Arg}\min_{\mathbf{\Omega}} \sum_i \left\| x_i - \hat{x}_i \right\|_2^2$$

$$\hat{x}_i = \mathrm{Arg}\min_{x} \left\| \mathbf{\Omega} x \right\|_1 \quad s.t. \quad \left\| x - y_i \right\|_2 \le a$$

- How can this target function be minimized?

Learning MAP-Analysis Operators

- Suggested solution:
- Assume we have some initial guess for $\mathbf{\Omega}$
- Using this guess, we compute

$$\hat{x}_i = \text{Arg}\min_x \|\mathbf{\Omega}x\|_1 \quad s.t. \quad \|x - y_i\|_2 \le a \quad \rightarrow \quad \hat{x}_i \ne x_i$$

- Since x_i is also in the feasible region (let's assume a is large enough), the reason for this must be that

$$\|\mathbf{\Omega}\hat{x}_i\|_1 < \|\mathbf{\Omega}x_i\|_1$$

Learning MAP-Analysis Operators

- Idea: correct $\mathbf{\Omega}$ by minimizing

$$f(\mathbf{\Omega}) = \|\mathbf{\Omega}x_i\|_1 - \|\mathbf{\Omega}\hat{x}_i\|_1 \quad \rightarrow \quad f(\mathbf{\Omega}) = \sum_i \|\mathbf{\Omega}x_i\|_1 - \|\mathbf{\Omega}\hat{x}_i\|_1$$

- Gradient descent now suggests the update step:

$$\mathbf{\Omega}_{new} = \mathbf{\Omega}_{old} - \eta \cdot \sum_i \left\{ \text{sign}(\mathbf{\Omega}x_i) x_i^T - \text{sign}(\mathbf{\Omega}\hat{x}_i)\hat{x}_i^T \right\}$$

Learning MAP-Analysis Operators

- More generally, we can consider any function of the form

$$f(\mathbf{\Omega}) = \varphi(\|\mathbf{\Omega}x_i\|) - \varphi(\|\mathbf{\Omega}\hat{x}_i\|)$$

- $\varphi : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ is monotonically increasing
- The update rule becomes

$$\mathbf{\Omega}_{new} = \mathbf{\Omega}_{old} - \eta \cdot \sum_i \left\{ \varphi'(\|\mathbf{\Omega}x_i\|)\text{sign}(\mathbf{\Omega}x_i)x_i^T - \varphi'(\|\mathbf{\Omega}\hat{x}_i\|)\text{sign}(\mathbf{\Omega}\hat{x}_i)\hat{x}_i^T \right\}$$

Algorithm Summary

Init: $\mathbf{\Omega} := \mathbf{\Omega}_0$

Iterate until converge:

(1) For all i, compute

$$\hat{x}_i = \text{Arg}\min_x \|\mathbf{\Omega}x\|_1 \quad s.t. \quad \|x - y_i\|_2 \le a$$

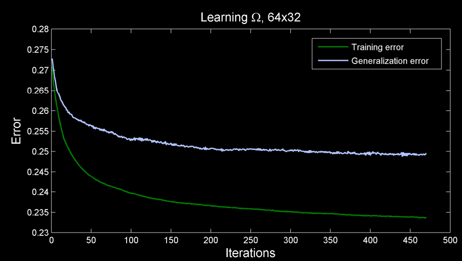
(2) Determine descent direction

$$d = \sum_i \left\{ \varphi'(\|\mathbf{\Omega}x_i\|)\text{sign}(\mathbf{\Omega}x_i)x_i^T - \varphi'(\|\mathbf{\Omega}\hat{x}_i\|)\text{sign}(\mathbf{\Omega}\hat{x}_i)\hat{x}_i^T \right\}$$

(3) Update: $\mathbf{\Omega}_{new} = \mathbf{\Omega}_{old} - \eta \cdot d$

Initial Results are Encouraging

- We used $\varphi(x) = \sqrt{x}$
- Dataset: random 64x32 Ω operator, from which 1500 MAP-Analysis vertices were computed.
- 1300 for training, 200 for validation
- Adding low-intensity noise leads to the input pairs (x_i, y_i)



Future Directions

- Improving the MAP-Analysis prior by learning.
- Beyond the Bayesian methodology: learning problem-based regularizers.
- MAP-Analysis versus MAP-Synthesis: how do they compare for specific applications?
- Learning structured priors and fast transforms.
- Redundancy: how much is good? The benefits of each approach from overcompleteness.
- Generalizing the regularization and degradation models.

Thank You!

Questions?

MAP-Analysis Defining Polytope

- Let $\underline{x} \in \partial\Psi(\Omega)$, and let $k(\underline{x})$ denote the rank of the rows in Ω to which $\underline{x}$ is orthogonal to, then it resides strictly within a face of dimension $N - k(\underline{x}) - 1$ of the MAP-Analysis defining polytope.

$\underline{x}$ is a vertex	$\Leftrightarrow$	$\underline{x}$ is orthogonal to N-1 independent rows in Ω
$\underline{x}$ is on an edge	$\Leftrightarrow$	$\underline{x}$ is orthogonal to N-2 independent rows in Ω
	...	
$\underline{x}$ is on an facet	$\Leftrightarrow$	$\underline{x}$ is orthogonal to 0 rows in Ω

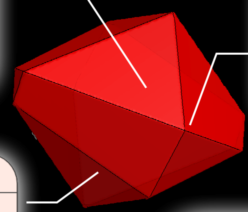
MAP-Analysis Defining Polytope

Facet (2D-face)

$$\underline{x} = \begin{pmatrix} -0.13 \\ -0.23 \\ 0.44 \end{pmatrix}, \quad \Omega\underline{x} = \begin{pmatrix} 0.23 \\ 0.33 \\ -0.24 \\ -0.21 \end{pmatrix}$$

Vertex (0D-face)

$$\underline{x} = \begin{pmatrix} -0.10 \\ -0.70 \\ 0.20 \end{pmatrix}, \quad \Omega\underline{x} = \begin{pmatrix} 0.65 \\ 0.34 \\ \mathbf{0} \\ \mathbf{0} \end{pmatrix}$$



Edge (1D-face)

$$\underline{x} = \begin{pmatrix} -0.49 \\ -0.30 \\ -0.07 \end{pmatrix}, \quad \Omega\underline{x} = \begin{pmatrix} 0.37 \\ \mathbf{0} \\ -0.32 \\ 0.31 \end{pmatrix}$$

$$\Omega = \begin{pmatrix} -0.204 & -0.905 & -0.005 \\ 0.111 & -0.324 & 0.608 \\ 0.860 & -0.242 & -0.432 \\ -0.455 & -0.131 & -0.667 \end{pmatrix}$$

Regularity of MAP-A Defining Polytope

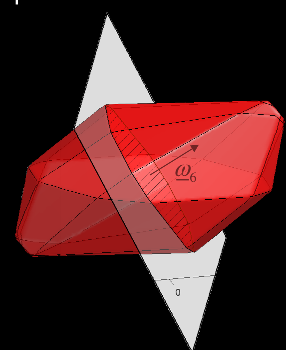
- The MAP-Analysis defining polytope displays a structural regularity which has a recursive description:

1. Its edges are arranged in *planar edge loops* about the origin.
2. For $k \geq 3$, every $N-k$ independent rows from define a k-D null-space, whose intersection with the polytope is a k-D polytope exhibiting itself *the same MAP-Analysis polytopal regularity*.

Regularity of MAP-A Defining Polytope

- Example: In the 3D case, each row corresponds to a planar edge loop of the polytope:

$$\Omega = \begin{pmatrix} 0.39 & 0.26 & 0.88 \\ -0.99 & 0.12 & 0.01 \\ 0.96 & 0.02 & 0.27 \\ -0.56 & -0.17 & 0.81 \\ 0.24 & -0.48 & -0.84 \\ 0.10 & 0.89 & -0.45 \end{pmatrix} \quad \underline{\omega}_6$$



Principal Signals

- Definition: the *principal signals* of a MAP distribution are the local maxima of

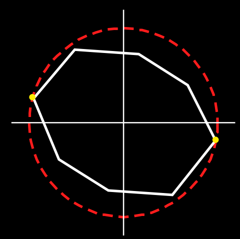
$$\arg\max_{\underline{x}} P(\underline{x}) \quad s.t. \ \|\underline{x}\|_2 = 1$$

- For MAP-Synthesis, the principal signals are in fact a *subset* of the dictionary atoms. However, this issue is rarely observed:

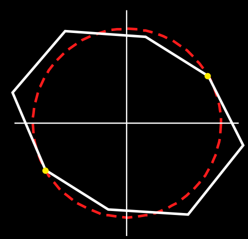
Theorem: The principal signals of a MAP-Synthesis prior ***coincide*** *with the dictionary atoms when the dictionary is normalized to a fixed-length.*

Highly Recoverable Signals

- Not every vertex necessarily defines a principal signal:



Principal signal



Non-principal signal

Principal Signals

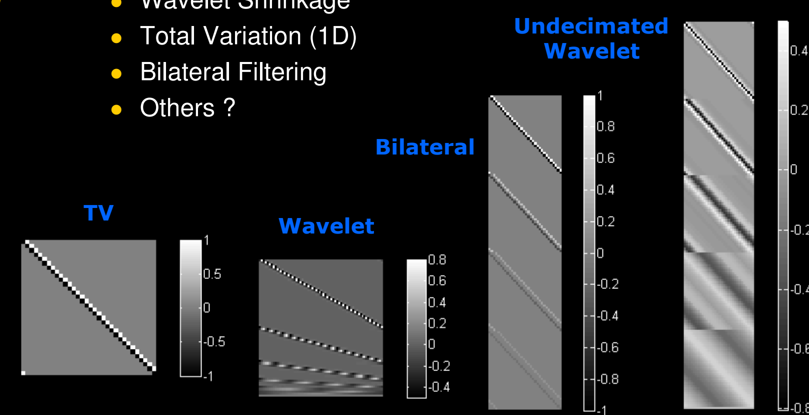
- Unfortunately, in the general case we have no closed-form description for these signals.
- Algorithms have been developed for locating these signals in the general case, for both MAP-Analysis and MAP-Synthesis.
- These algorithms, however, are quite heavy.

Locating Principal Signals

- MAP-Synthesis:
  - Select an atom.
  - Connect it to each of the other atoms and their antipods.
  - Check if maximally distant relative to all these directions.
  - If so, atom is principal; otherwise it is not.
- MAP-Analysis:
  - Select an initial vertex.
  - Determine its incident edge loops.
  - If vertex is locally maximal – stop.
  - Otherwise, choose a more distant vertex from one of its incident edge loops, and repeat.

Analysis Priors ("MAP-Analysis")

- Many existing algorithms take this form:
  - Wavelet Shrinkage
  - Total Variation (1D)
  - Bilateral Filtering
  - Others ?



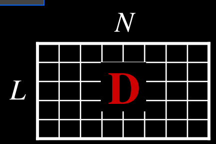
Synthesis Priors ("MAP-Synthesis")

- Synthesis priors stem from the concept of sparse representation in overcomplete dictionaries:

$$\hat{\gamma}(\underline{y}) = \arg\min_{\underline{\gamma}} \left\{ \|\underline{\gamma}\|_p^p \right\} \quad s.t. \quad \underline{y} = \mathbf{D}\underline{\gamma} \qquad \left(\mathbf{D} \in \mathbb{R}^{NxL}\right)$$

$$\underline{x}_{MAP-S} = \mathbf{D} \cdot \arg\min_{\underline{\gamma}} \left\{ \left\| \underline{y} - T\{\mathbf{D}\underline{\gamma}\} \right\|_2^2 + \lambda \cdot \|\underline{\gamma}\|_p^p \right\}$$

- $\mathbf{D}$ is generally overcomplete $(L \geq N)$:



- Typically $0 \leq p \leq 1$
- Can also be explained in terms of MAP estimation.