

# ANALYSIS VERSUS SYNTHESIS IN SIGNAL PRIORS<sup>†</sup>

Michael Elad\*, Peyman Milanfar<sup>+</sup>, and Ron Rubinfeld\*

\*Computer Science Department  
Technion IIT, Haifa, Israel

<sup>+</sup>Electrical Engineering Department  
University of California, Santa Cruz, CA

## ABSTRACT

The concept of prior probability for signals plays a key role in the successful solution of many inverse problems. Much of the literature on this topic can be divided between analysis-based and synthesis-based priors. Analysis-based priors assign probability to a signal through various forward measurements of it, while synthesis-based priors seek a reconstruction of the signal as a combination of atom signals. In this paper we describe these two prior classes, focusing on the distinction between them. We show that although when reducing to the complete and under-complete formulations the two become equivalent, in their more interesting overcomplete formulation the two types depart. Focusing on the  $\ell^1$  denoising case, we present several ways of comparing the two types of priors, establishing the existence of an unbridgeable gap between them.

## 1. INTRODUCTION

The general inverse problem seeks the recovery of an unknown signal  $\underline{X} \in \mathbb{R}^N$  (a vector of dimension  $N$  over the real numbers) based on indirect measurements of it given in the vector  $\underline{Y} \in \mathbb{R}^M$ . A typical model for describing the relation between  $\underline{X}$  and  $\underline{Y}$  is

$$\underline{Y} = \mathbf{T}\{\underline{X}\} + \underline{V}, \quad (1)$$

where  $\mathbf{T} : \mathbb{R}^N \rightarrow \mathbb{R}^M$  is a (possibly non-linear) known operator, and  $\underline{V} \in \mathbb{R}^M$  is a zero-mean white Gaussian additive noise vector (other models for the noise can also be considered, but here we restrict our discussion to the assumptions made above for simplicity). The structure (1) represents many important problems in signal and image processing; in this paper we focus on the denoising problem, corresponding to the choice  $\mathbf{T}\{\underline{X}\} = \underline{X}$ . Many of our conclusions for this simplified case will remain relevant when considering the more general inverse problem.

Inverting the above process in (1) can be done in many different ways. When lacking any a-priori knowledge about the unknown, Maximum Likelihood (ML) estimation suggests finding the  $\underline{X}$  that leads to the most probable set of measurements  $\underline{Y}$ . However, this option is often problematic, as most inverse problems are ill-posed. A stabilized solution to the inverse problem posed above comes from the Maximum-A-posteriori Probability (MAP) estimator, which regularizes the estimation process using an assumed *prior distribution* on the signal space. Indeed, such signal priors are implicitly used in many other signal processing applications such as compression, signal decomposition, recognition, and more.

**MAP-Analysis Approach.** When studying the variety of published work in the field, two main prior types emerge. The first utilizes an analysis-based approach, deriving the probability of a signal from a set of forward transforms applied to it. Such priors are the backbone of many classic as well as more recent algorithms, and most commonly appear as regularizing elements in optimization problems or PDE methods.

In this paper, we focus on a robust Gibbs-like distribution, of the form

$$P\{\underline{X}\} = \text{Const} \cdot \exp\{-\alpha \cdot \|\Omega \underline{X}\|_p^p\},$$

where  $\Omega \in M^{L \times N}$  is some pre-specified matrix, and  $\|\cdot\|_p^p$  is the  $\ell^p$  norm. The term  $\|\Omega \underline{X}\|_p^p$  is an energy functional that is supposed to

be low for highly probable signals, and higher as the signal is less probable. We refer to  $\Omega$  as the *analyzing operator*. Merged with the Gaussianity assumption on the additive noise, this poses the MAP denoising process as the minimization problem

$$\hat{\underline{X}}_{MAP-A} = \underset{\underline{X}}{\text{Argmin}} \|\underline{Y} - \underline{X}\|_2^2 + \lambda \cdot \|\Omega \underline{X}\|_p^p. \quad (2)$$

If robust norms are used ( $p < 2$  or some robust M-function [1]), an iterative algorithm is typically employed for the minimization of this penalty function. Preference should be given to  $p \geq 1$  so that the overall penalty function is convex, thus guaranteeing a unique solution. We name this method the *MAP-Analysis* approach since the prior is based on a sequence of linear filters applied to the signal, essentially analyzing its behavior.

**MAP-Synthesis Approach.** The second type of prior arises from employing a synthesis-based approach. Synthesis-based methods are a more recent contribution, and stem in a large part from the Basis Pursuit method pioneered by Chen, Donoho & Saunders [2].

Suppose that a signal  $\underline{X} \in \mathbb{R}^N$  is to be represented as a linear combination of “building-block” atoms taken as the columns of a full-rank matrix  $\mathbf{D} \in M^{N \times L}$ , with  $L \geq N$  (notice the different size compared to  $\Omega$ ); we refer to the columns of  $\mathbf{D}$  as the *atom* signals. This leads to the linear under-determined equation set  $\underline{X} = \mathbf{D}\underline{\gamma}$ , where  $\underline{\gamma} \in \mathbb{R}^L$  is overcomplete. We assume for the idealized signal  $\underline{X}$  that its representation  $\underline{\gamma}$  is *sparse*, implying that only a few atoms are involved in its construction. Assuming  $\underline{Y}$  is a noisy version of this signal, then the following is the *MAP-Synthesis* option for the recovery of  $\underline{X}$ :

$$\hat{\underline{X}}_{MAP-S} = \mathbf{D} \cdot \underset{\underline{\gamma}}{\text{Argmin}} \|\underline{Y} - \mathbf{D}\underline{\gamma}\|_2^2 + \lambda \cdot \|\underline{\gamma}\|_p^p. \quad (3)$$

In this expression, the  $\ell^p$ -norm with  $p < 2$  seeks the sparsest representation vector  $\underline{\gamma}$  that explains  $\underline{Y}$  in terms of the dictionary columns. Note that if the solution of the optimization problem is denoted as  $\hat{\underline{\gamma}}$ , the estimated output signal is given by  $\hat{\underline{X}}_{MAP-S} = \mathbf{D}\hat{\underline{\gamma}}$ .

Through the MAP framework, this approach may be generalized to incomplete dictionaries. We let  $\Gamma_{\underline{X}} = \{\underline{\gamma} | \underline{X} = \mathbf{D}\underline{\gamma}\}$  denote the set of representations of  $\underline{X}$  in  $\mathbf{D}$ , where  $\Gamma_{\underline{X}}$  may be infinite, empty, or a singleton. The a-priori probability assumed for  $\underline{X}$  depends on its sparsest representation in  $\mathbf{D}$ ; in this setting, signals not spanned by the columns of  $\mathbf{D}$  are assigned a-priori probability 0.

The MAP-Synthesis prior is therefore given as a Gibbs distribution on the optimal representations:

$$P\{\underline{X}\} = \begin{cases} \text{Const} \cdot \exp\{-\alpha \cdot \|\hat{\underline{\gamma}}(\underline{X})\|_p^p\} & \text{if } \Gamma_{\underline{X}} \neq \emptyset \\ 0 & \text{otherwise} \end{cases} \quad (4)$$

where

$$\hat{\underline{\gamma}}(\underline{X}) = \underset{\underline{\gamma} \in \Gamma_{\underline{X}}}{\text{Argmin}} \|\underline{\gamma}\|_p^p.$$

This prior, when plugged into the MAP formulation, leads precisely to the process described in (3). From a practical point of view, an iterative algorithm is required for the solution of (3), and there are many methods to do so effectively. For  $p \geq 1$ , we are guaranteed to have a unique solution.

**Analysis versus Synthesis.** Comparing the two recovery processes in (2) and (3), we see that the two describe very similar structures.

<sup>†</sup>This research was supported by the USA-Israel binational science foundation (BSF) grant #2004199

The heuristic behind each remains sparsifying the representation of the signal — be this its forward projection on the basis elements, or its reconstruction as their linear combination.

Analysis-based methods, specifically in their robust form, are a very common structure in image processing and computer vision applications. MAP-Analysis leads to a simple optimization problem, which in the overcomplete case is considerably easier to solve (due to the smaller unknown) compared to similar-sized MAP-Synthesis. Nonetheless, a growing number of works are utilizing the synthesis approach for inverse problem regularization. Synthesis-based priors are an attractive choice due to their more intuitive and versatile structure. This trend is strengthened by a wealth of theoretical and practical advancements, making this approach both more appealing and computationally tractable [3, 4, 5].

Despite these achievements, MAP-Synthesis remains a prohibitive option in many cases. This has led several works to seek alternative approaches over direct minimization. One specific option is to approximate the synthesis-based method by an analysis-based one, as is done in [6] where the analysis operator is taken as the pseudo-inverse of the synthesis dictionary. This approach is only partially justified there; the algebraic viewpoint of this transform is discussed in section 3.

In light of these recent developments, it is our goal in this paper to clarify the distinction between the two approaches, and shed some light on the conceptual and technical gaps between them. We show that for specific cases, the two approaches are equivalent, utilizing a pseudo-inverse relation between the analysis operator and synthesis dictionary. Such is the case for the square and under-complete formulations, as well as for the  $\ell^2$  (i.e.  $p = 2$ ) choice. However, as we go to the general overcomplete formulation ( $L > N$ ), the equivalence between the two MAP options breaks. In this paper we characterize this gap, and show that in the general case the two methods can behave very differently. We concentrate on the  $p = 1$  case, which is often favored due to its convexity and robustness, and provide theoretical as well as numerical results indicating that the two methods are fundamentally distinct. A more complete discussion of this work can be found in [7].

## 2. THE SQUARE AND UNDER-DETERMINED CASES

We begin our discussion by showing that in the (under-)determined case (i.e.,  $L \leq N$ ), the two methods remain equivalent.

**Theorem 1. Square Non-Singular Case – Complete Equivalence.** *MAP-Analysis and MAP-Synthesis are equivalent if MAP-Analysis utilizes a square and non-singular analyzing operator  $\Omega$ . The equivalent MAP-Synthesis method is obtained for the dictionary  $\mathbf{D} = \Omega^{-1}$ .*

*Proof.* We start with the MAP-Analysis approach as posed in equation (2). Since  $\Omega$  is square and non-singular, defining  $\Omega \underline{X} = \underline{\gamma}$  leads to  $\underline{X} = \Omega^{-1} \underline{\gamma}$ . Putting this into (2), we get an alternative optimization problem with  $\underline{\gamma}$  replacing  $\underline{X}$  as unknown,

$$\hat{\underline{X}} = \Omega^{-1} \cdot \underset{\underline{\gamma}}{\text{Argmin}} \|\underline{Y} - \Omega^{-1} \underline{\gamma}\|_2^2 + \lambda \cdot \|\underline{\gamma}\|_p^p,$$

and the equivalence to the MAP-Synthesis method in (3) is evident. Likewise, starting from the MAP-Synthesis formulation and using the same argument, we can obtain a MAP-Analysis one — and thus the two methods are equivalent.  $\square$

An immediate consequence of this theorem is that the two prior assumptions behind the two methods are essentially the same for full-rank and  $L = N$  analyzing operator/dictionary. Choosing a sparse representation over a square-non-singular dictionary amounts to the same probability density as that obtained from using the corresponding (inverse) analyzing operator and requiring sparsity of the filtered coefficients.

The generalization of Theorem 1 for the  $L \leq N$  case requires a more delicate analysis, but is based on similar arguments. We point out that complete equivalence cannot be guaranteed in this

case, due to the property of MAP-Synthesis to only produce results in the column-span of  $\mathbf{D}$ , whereas MAP-Analysis poses no such restriction. We arrive at the following result (stated without proof):

**Theorem 2. Under-Complete Case – Near-Equivalence.** *MAP-Analysis with a full-rank operator  $\Omega \in M^{L \times N}$  ( $L \leq N$ ) is nearly-equivalent to MAP-Synthesis with  $\mathbf{D} = \Omega^+$ . This is expressed by the relation  $\hat{\underline{X}}_{\text{MAP-A}} = \hat{\underline{X}}_{\text{MAP-S}} + \underline{Y}^{\mathbf{D}^\perp}$ , with  $\underline{Y}^{\mathbf{D}^\perp}$  representing the component of the input orthogonal to the column-span of  $\mathbf{D}$ .*

Theorem 2 represents, both conceptually and computationally, a complete equivalence between the two approaches, as knowing the solution to either one immediately fixes the solution to the other. We also see that when the input is in the column-span of  $\mathbf{D}$  (as in the square non-singular case), we have  $\hat{\underline{X}}_{\text{MAP-A}} = \hat{\underline{X}}_{\text{MAP-S}}$ .

## 3. THE OVER-DETERMINED CASE

We have seen that the two methods are equivalent for the  $L \leq N$  case. Our main interest however is in the overcomplete ( $L > N$ ) case, advocated strongly by the Basis Pursuit approach. To get an idea of the difficulties arising when going to this overcomplete case, we begin by considering the natural pseudo-inverse relation, which has thus far successfully achieved equivalence in the (under-)complete case. We assume  $\Omega$  has full column rank, and hence  $\Omega^+ \Omega = I$ . Beginning with the MAP-Analysis formulation in (2), we let  $\Omega \underline{X} = \underline{\gamma}$ . Since  $\Omega^+ \Omega = I$ , recovering  $\underline{X}$  from  $\underline{\gamma}$  is done by  $\underline{X} = \Omega^+ \underline{\gamma}$ . However, in replacing the unknown from  $\underline{X}$  to  $\underline{\gamma}$  we must add the constraint that  $\underline{\gamma}$  is spanned by the columns of  $\Omega$ , due to its definition; this can be represented by the constraint  $\Omega \Omega^+ \underline{\gamma} = \underline{\gamma}$ . Thus we obtain the following equivalent MAP-Analysis form:

$$\hat{\underline{X}}_{\text{MAP-A}} = \Omega^+ \cdot \underset{\underline{\gamma}: \Omega \Omega^+ \underline{\gamma} = \underline{\gamma}}{\text{Argmin}} \|\underline{Y} - \Omega^+ \underline{\gamma}\|_2^2 + \lambda \cdot \|\underline{\gamma}\|_p^p. \quad (5)$$

If the MAP-Synthesis solution (with  $\mathbf{D} = \Omega^+$ ) satisfies the constraint  $\Omega \Omega^+ \underline{\gamma} = \underline{\gamma}$ , then omitting it in (5) has no effect, and both approaches arrive at the same solution. However, in the general case this constraint is not satisfied, and thus the two methods lead to different results. An interesting observation is that while the representation results could differ vastly, the final estimators  $\hat{\underline{X}} = \Omega^+ \hat{\underline{\gamma}}$  in both could be very similar; this is because in multiplying by  $\Omega^+$  we null-out content not in the column-span of  $\Omega$ , and essentially satisfy the constraint. However, as we will see, this does not turn out to close the gap between the two methods. An exception is the non-robust  $\ell^2$  case, in which equivalence still holds:

**Theorem 3. Over-Complete Case – Equivalence for  $p = 2$ .** *MAP-Analysis with a full-rank analyzing operator  $\Omega \in M^{L \times N}$  ( $L > N$ ) is equivalent to MAP-Synthesis with  $\mathbf{D} = \Omega^+$  for  $p = 2$ .*

*Proof.* From (5) the proof is trivial. When  $p = 2$ , the unknown  $\underline{\gamma}$  can be assumed to be the sum of two parts,  $\underline{\gamma} = \underline{\gamma}^\Omega + \underline{\gamma}^{\Omega^\perp}$ , where  $\underline{\gamma}^\Omega$  comes from the column-span of  $\Omega$ , and  $\underline{\gamma}^{\Omega^\perp}$  from the orthogonal subspace. The second penalty term ( $\|\underline{\gamma}\|_2^2$ ) clearly prefers  $\underline{\gamma}^{\Omega^\perp}$  to be zero; as to the first term ( $\|\underline{Y} - \Omega^+ \underline{\gamma}\|_2^2$ ),  $\underline{\gamma}^{\Omega^\perp}$  has no impact on it as it is nulled down by  $\Omega^+$ . Thus,  $\underline{\gamma}^{\Omega^\perp}$  that violates the constraint in  $\underline{\gamma}$  is chosen as zero, and the two methods coincide.  $\square$

### 3.1 MAP-Analysis and MAP-Synthesis in $\ell^1$

In the remainder of this paper we consider the two MAP approaches with  $p = 1$ . The use of the  $\ell^1$  norm for signal and image recovery has received considerable attention beginning at the late 1980's, with the development of robust statistics; probably most notable of these are the Total Variation -based methods [8] which are analysis-based, and the later Basis Pursuit method which is synthesis-based. The  $\ell^1$  option is a favorable choice due to its combination of convexity, robustness, and proximity to  $\ell^0$  in the synthesis case [3, 5].

Looking at the MAP formulations in (2) and (3), we see that both depend on a weighting parameter  $\lambda$  to control the regularization element; for  $\lambda = 0$  both reproduce the input as the solution, and as  $\lambda \rightarrow \infty$  they deviate from the input until finally converging to 0. However, the rate at which this occurs may vary substantially between the two methods, and thus this parameterization is inconvenient for our purposes. Alternatively, we propose the following reformulations of the two problems:

$$\begin{aligned}\hat{\mathbf{X}}_{MAP-A}(a) &= \underset{\mathbf{X}}{\text{Argmin}} \|\Omega\mathbf{X}\|_1 \quad \text{s.t.} \quad \|\mathbf{Y} - \mathbf{X}\|_2 \leq a \\ \hat{\mathbf{X}}_{MAP-S}(a) &= \mathbf{D} \cdot \underset{\boldsymbol{\gamma}}{\text{Argmin}} \|\boldsymbol{\gamma}\|_1 \quad \text{s.t.} \quad \|\mathbf{Y} - \mathbf{D}\boldsymbol{\gamma}\|_2 \leq a.\end{aligned}$$

These formulations are conceptually simpler, with  $a$  directly controlling the deviation from the input. The original MAP target functions are essentially the Lagrangian functionals of these constrained versions (with  $\lambda$  representing the inverse of the Lagrange multiplier), and thus the two formulations are equivalent.

### 3.2 Geometry of MAP

In the modified formulations, the two MAP methods take a clear geometric structure. The solutions of both are obviously confined to a ball of radius  $a$  about  $\mathbf{Y}$  (this is true as we assume  $\mathbf{D}$  is full rank). We also assume this ball does not include the origin, otherwise the solution is trivially zero. Considering MAP-Analysis first, the level-sets of its target function  $f_A(\mathbf{X}) = \|\Omega\mathbf{X}\|_1$  are a collection of concentric, centro-symmetric polytopes  $\{\mathbf{X} \mid \|\Omega\mathbf{X}\|_1 \leq c\}$ . Graphically, the solution can be obtained by taking a small level-set  $\{\|\Omega\mathbf{X}\|_1 \leq c\}$  about the origin, and gradually inflating it (by increasing  $c$ ) until it first encounters the ball. The point of intersection constitutes the MAP-Analysis solution, as there cannot be a point in the ball with a smaller value of  $\|\Omega\mathbf{X}\|_1$ .

As to MAP-Synthesis, a similar process may be described using the collection of concentric and centro-symmetric polytopes  $\mathbf{D} \cdot \{\boldsymbol{\gamma} \mid \|\boldsymbol{\gamma}\|_1 \leq c\}$ <sup>1</sup>. A fact to note for both MAP methods is that these “inflations” we visualize are performed via simple scaling: we have  $\{\|\Omega\mathbf{X}\|_1 \leq c\} = c \cdot \{\|\Omega\mathbf{X}\|_1 \leq 1\}$ , and  $\mathbf{D} \cdot \{\|\boldsymbol{\gamma}\|_1 \leq c\} = c \cdot \mathbf{D} \cdot \{\|\boldsymbol{\gamma}\|_1 \leq 1\}$ . This implies that given the canonical MAP defining polytopes  $\Psi_\Omega = \{\|\Omega\mathbf{X}\|_1 \leq 1\}$  and  $\Phi_{\mathbf{D}} = \mathbf{D} \cdot \{\|\boldsymbol{\gamma}\|_1 \leq 1\}$ , the inflation processes are fully defined, and so are the MAP solutions; in fact, specifying these polytopes is completely equivalent to specifying  $\Omega$  or  $\mathbf{D}$ , respectively. We find that the behavior of each of the methods is governed exclusively by the geometry of a single high-dimensional polytope, and this provides us with the necessary basis for comparing the two methods. We therefore continue by characterizing the geometry of these polytopes; for the discussion, we recall that a *face* of an  $N$ -dimensional polytope is any intersection of this polytope with a tangent hyperplane (e.g. a vertex, edge, ridge etc.), and a *facet* is an  $N - 1$ -dimensional face.

**MAP-Analysis Defining Polytope.** The MAP-Analysis defining polytope is a level set of the target function  $f_A(\mathbf{X}) = \|\Omega\mathbf{X}\|_1$ . Applying the gradient operator to  $f_A$ , we find that the normal to this surface satisfies  $\underline{n}(\mathbf{X}) \propto \Omega^T \text{sign}(\Omega\mathbf{X})$ . Evidently  $\underline{n}(\mathbf{X})$  is defined for any  $\mathbf{X}$  in which all coordinates of  $\Omega\mathbf{X}$  are non-zero; where one or more of these vanishes,  $\underline{n}(\mathbf{X})$  exhibits a discontinuity arbitrarily filled-in by the sign function. Intuitively, consider the signals  $\mathbf{X}$  on the boundary of the defining polytope, then the facets correspond to the locations where  $\underline{n}(\mathbf{X})$  is smooth, whereas the other faces correspond to where  $\underline{n}(\mathbf{X})$  is discontinuous. This formalizes as

**Lemma 4.** *Let  $\mathbf{X} \in \partial\Psi_\Omega$  (the boundary of the defining polytope), and let  $k$  denote the rank of the rows in  $\Omega$  to which  $\mathbf{X}$  is orthogonal to. Then  $\mathbf{X}$  resides strictly within<sup>2</sup> a face of dimension  $(N - k - 1)$  of the MAP-Analysis defining polytope.*

<sup>1</sup>Note that these sets exist in *signal space*, and have the explicit form  $\{\mathbf{X} \mid \exists \boldsymbol{\gamma}, \mathbf{X} = \mathbf{D}\boldsymbol{\gamma} \wedge \|\boldsymbol{\gamma}\|_1 \leq c\}$

<sup>2</sup>We use the term *strictly within a face* to indicate a signal located in the interior of a face, in the sense that there exists a finite  $\varepsilon$ -ball about it — of the same dimension as the face — entirely contained within this face (note that

The lemma implies that to obtain a vertex of  $\Psi_\Omega$ , we choose  $N - 1$  linearly-independent rows in  $\Omega$ , determine their 1D null-space  $\underline{v}$  and normalize such that  $\|\Omega\underline{v}\|_1 = 1$  (this defines two antipodal vertices). Edges are similarly obtained, by choosing  $N - 2$  linearly-independent rows, and taking any properly normalized signal in their 2D null-space. This leads to an immediate conclusion concerning the *complexity* of the MAP-Analysis defining polytope: evidently its vertex count is equal to the number of choices of  $N - 1$  linearly-independent rows in  $\Omega$ , and in the worst-case, this may reach an exponential  $\binom{L}{N-1}$ . In fact, there are examples where this bound is met, and thus the bound is *tight* (for the worst-case). This is also the expected number of vertices when the directions of the rows in  $\Omega$  are random and uniformly distributed on the unit sphere.

Several conclusions can be drawn from Lemma 4 concerning the structure of the MAP-Analysis defining polytope. For instance, each set of edges, stemming from the same 2D null-space of some  $N - 2$  rows in  $\Omega$ , forms a planar *edge-loop* of consecutive edges all existing on this common plane. We thus find that the edges of  $\Psi_\Omega$  are arranged in loops about the origin. Similar arguments lead to higher dimensional regularities, corresponding to the choices of  $N - k$  independent rows from  $\Omega$ , for  $k > 2$ . We also find that this polytope has a strict *neighborliness pattern*, where each of its vertices has *precisely*  $2(N - 1)$  neighbors.

**MAP-Synthesis Defining Polytope.** The MAP-Synthesis defining polytope is given by  $\Phi_{\mathbf{D}} = \mathbf{D} \cdot \{\boldsymbol{\gamma} \mid \|\boldsymbol{\gamma}\|_1 \leq 1\}$ . It is a known result (and quite simple to show) that this polytope is obtained as the convex hull of the columns of  $\mathbf{D}$  and  $-\mathbf{D}$ :

**Lemma 5.** *The MAP-Synthesis defining polytope  $\Phi_{\mathbf{D}} = \mathbf{D} \cdot \{\|\boldsymbol{\gamma}\|_1 \leq 1\}$  is obtained as the convex hull of  $\{\pm \underline{d}_i\}_{i=1\dots L}$ , where  $\{\underline{d}_i\}$  are the columns of  $\mathbf{D}$ .*

This lemma simply states that the vertices of  $\Phi_{\mathbf{D}}$  are those columns of  $\pm\mathbf{D}$  which cannot be represented as a convex combination of any other columns (and their antipodes); the other faces are the convex combinations of neighboring vertices. A vertex can therefore be represented as  $\underline{v} = \mathbf{D}\boldsymbol{\gamma}$  where  $\boldsymbol{\gamma}$  has a single non-zero element  $\gamma_i = \pm 1$ , and a point on an edge can be represented similarly with  $\boldsymbol{\gamma}$  having two non-vanishing elements  $\gamma_i, \gamma_j$  satisfying  $|\gamma_i| + |\gamma_j| = 1$ . In general, a point on a  $k$ -dimensional face will have a representation  $\mathbf{X} = \mathbf{D}\boldsymbol{\gamma}$  with  $\boldsymbol{\gamma}$  having  $k + 1$  non-vanishing elements, and  $\|\boldsymbol{\gamma}\|_1 = 1$ . We emphasize that this is *not* a sufficient condition, so a signal  $\mathbf{X} = \mathbf{D}\boldsymbol{\gamma}$  synthesized from a sparse representation  $\boldsymbol{\gamma}$  might not reside on a low-dimensional face if the corresponding columns of  $\pm\mathbf{D}$  are not neighbors, or do not constitute vertices.

### 3.3 MAP-Synthesis as a Superset of MAP-Analysis

The geometrical viewpoint leads to an important consequence concerning the relationship between the two MAP formulations. From the geometrical description, it is clear that any  $\ell^1$  MAP-Analysis estimator may be reformulated as an equivalent MAP-Synthesis one; this is accomplished by simply taking all the MAP-Analysis defining polytope vertices — one of each antipodal pair — as the MAP-Synthesis dictionary atoms. Since both methods will have the same defining polytope, they will be completely equivalent. This establishes the generality of MAP-Synthesis over MAP-Analysis in  $\ell^1$ :

**Theorem 6. Over-Complete  $\ell^1$  Case – Generality of MAP-Synthesis.** *For any  $\ell^1$  MAP-Analysis form with full-rank analyzing operator  $\Omega$  ( $L \geq N$ ), there exists a dictionary  $\mathbf{D}(\Omega)$  describing an equivalent  $\ell^1$  MAP-Synthesis form. The reverse is not true.*

The reverse direction fails due to the strict regularity imposed on the MAP-Analysis defining polytopes. Since this regularity does

this covers signals that are vertices, who reside strictly within themselves). Also, as opposed to standard residence, strict residence is *unique*, as the faces are considered open rather than closed and thus do not overlap.

not apply to MAP-Synthesis, it may clearly describe structures not represented in MAP-Analysis form.

We emphasize that the actual equivalence transform is of little practical value; except for the special case of  $N = 2$ , where the size of  $\mathbf{D}(\Omega)$  will be equal to (or even smaller than) that of  $\Omega^T$ , the size of  $\mathbf{D}(\Omega)$  will generally grow exponentially. Nonetheless, the theorem describes a definite one-way relationship between the two formulations: the synthesis formulation is clearly more general than the analysis one, with a *vast collection* of MAP-Synthesis priors unrepresented by the stricter MAP-Analysis form.

The fact that the equivalence transform is impractical stems directly from the incompatibility between the two polytopal structures in their vertex counts; this incompatibility is complemented by a parallel one in the neighborliness properties of these polytopes. Our observation for MAP-Analysis that every vertex has a linear number of neighbors, while their total number is exponential, implies that the probability of any two vertices to be neighbors approaches 0 as  $N \rightarrow \infty$ . In contrast, Donoho has recently presented opposite results [9] for MAP-Synthesis polytopes; as it turns out, for these polytopes the probability of any 2 (non-antipodal) vertices to be neighbors approaches 1 as  $N \rightarrow \infty$ .<sup>3</sup> We therefore find that while MAP-Analysis polytopes feature very *large* numbers of vertices with very *low* neighborliness, MAP-Synthesis polytopes exhibit *low* vertex counts and very *high* neighborliness. We see that the two structures are in fact *fundamentally different*.

### 3.4 Experiencing the Gap

The geometrical inconsistencies suggest a large gap between the two formulations in the over-determined  $\ell^1$  case. In this section we provide some feeling about this gap, through an actual example.

**Recoverability and Principal Signal.** To compare the two methods, we will seek those signals which are favored in some sense by each of the priors. We confine ourselves to a fixed-energy sphere, w.l.o.g. the unit sphere, and on this sphere search for the signals most effectively recovered by the specific MAP method. We use the term *recoverability* of a signal in a qualitative manner, referring to the ability of the MAP method to recover this signal successfully given noisy versions of it. More specifically, assume an energy-conserving variant of the MAP estimator where the denoised solution is post-processed by re-normalizing it to the magnitude of the input (thus eliminating its decay to zero caused by the low-energy preference of the prior). Under these conditions, the MAP estimation essentially searches the neighborhood of the input on the fixed-energy sphere, outputting a higher-probability (and presumably less noisy) signal near the input. We thus consider a signal as *highly recoverable* when its prior probability is maximal relative to a significant enough part of the directions about it on the fixed-energy sphere; indeed, such signals are generally recovered most successfully by the MAP method. The local maxima of the distribution on the unit sphere, which are clearly the most highly recoverable signals, will be referred to as the *principal signals* of the distribution.

The geometry of the MAP defining polytope directly dictates the behavior of the distribution on the unit sphere, and consequently the recoverability of the signals on it. The relation may be described as follows: For both MAP priors, the boundaries of the defining polytopes define iso-surfaces of signals with equal a-priori probability. On the fixed-energy sphere, those signals whose projections on the defining polytope are farther from the origin correspond to the more probable signals. It follows that the highly recoverable signals must be projected to the *extreme points* – the low-dimensional faces – of the polytope. Specifically, the principal signals of the distribution project to vertices of the MAP defining polytope.

It is important to point out that projection onto a low-dimensional face is only a *necessary* condition for high recoverability. For instance, a vertex will not always lead to a principal signal, as demonstrated in figure (1). Simulation results show a dra-

<sup>3</sup>The dictionary is assumed to be of linear size in  $N$ , and to fulfill certain randomness conditions; see Theorem 1 in [9].



Figure 1: Principal signals and the MAP defining polytope. The dotted circles denote the unit sphere in 2D signal space. The two polygons are different scales of the same MAP defining polytope. (a) A principal signal, obtained at a vertex of the defining polytope. (b) A vertex which does not represent a principal signal.

matic difference in the recovery performance of principal vs. non-principal vertices.

**Locating Principal Signals.** The principal signals of a MAP prior are tightly related to the vertices of its defining polytope. However, not every vertex necessarily implies a principal signal; for this, it must be maximally distant from the origin relative to all the directions about it on the boundary of the defining polytope. We begin with MAP-Synthesis, whose defining polytope is the convex hull of the columns of  $\pm\mathbf{D}$ . We assume the vertices of this polytope to be known — this can be achieved by solving a set of LP problems, or alternatively by normalizing the dictionary atoms to a fixed-length, in which case they all become vertices. Still, identifying those who represent principal signals is in general very difficult. We therefore limit the discussion to dictionaries with normalized columns — which are very common in practice — and in this case it can be shown that indeed *all* atoms constitute principal signals, hence locating these becomes trivial.

As to MAP-Analysis, the geometry of its defining polytope is described in Lemma 4, which also provides a practical method for generating its vertices. To locate those which represent principal signals, we use a simple traversal algorithm: Beginning with some initial vertex, we examine its incident edge-loops. If it is found to be maximal relative to all its incident edges, then (and only then) does it constitute a principal signal. Otherwise, it is not maximal relative to some incident edge. In this case we replace it with a vertex with larger  $\ell^2$ -norm from the violating edge loop, and continue the traversal. This swapping continues until a local maximum is encountered, providing one MAP-Analysis principal signal. The entire process can now be repeated using a new vertex as a starting point. The traversal approach is obviously time-consuming relative to the immediacy in the MAP-Synthesis case; on the other hand, it poses no assumptions on the structure of the operator  $\Omega$ .

**Simulation Results.** We provide some simulation results demonstrating the theoretical gaps we have described. For the experiment, we selected the pseudo-inverse relation between the dictionary and analysis operator; this is a natural choice for bridging the two methods, however in reality, it may lead to very different behaviors of the two methods. We selected the  $128 \times 256$  *Identity-Hadamard* dictionary  $\mathbf{D} = \frac{1}{\sqrt{2}} [\mathbf{I} \ \mathbf{H}]$  and its pseudo-inverse  $\Omega = \mathbf{D}^T = \frac{1}{\sqrt{2}} \begin{bmatrix} \mathbf{I} \\ \mathbf{H}^T \end{bmatrix}$  as the synthesis dictionary and analysis operator. This is an interesting choice as the two feature the same two-ortho structure, and furthermore  $\mathbf{D}$  is a near-optimal Grassmanian frame making it favorable for MAP-Synthesis methods [10].

The dictionary size immediately limits the number of distinct MAP-Synthesis principal signals to a mere 256. In contrast, MAP-Analysis boasts an enormous amount of principal signals: our traversal algorithm easily produced 10,000 such signals. Moreover, the program was designed to reject new signals if these resided in a radius of  $< .1$  from any existing principal signal; however, after 10,000 generated signals, the rejection rate remained negligible, suggesting that the true number of such signals is much greater (with an only known upper bound of order  $\binom{L}{N-1} = \binom{256}{127} \approx 10^{75}$ ).

We note that in our case the MAP-Synthesis principal signals all double as MAP-Analysis principal signals. To sharpen the comparison, we therefore produced additional sets of highly recoverable MAP-Synthesis signals, which were obtained on low-dimensional faces of the MAP-Synthesis defining polytope. Of course, we have

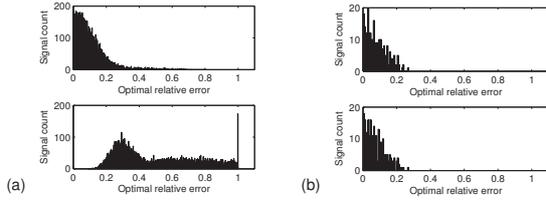


Figure 2: Denoising MAP principal signals. (a) Results for MAP-Analysis principal signals (10,000 examples). Distributions of optimal errors obtained using MAP-Analysis (above) and MAP-Synthesis (below). (b) The same for MAP-Synthesis principal signals (256 examples).

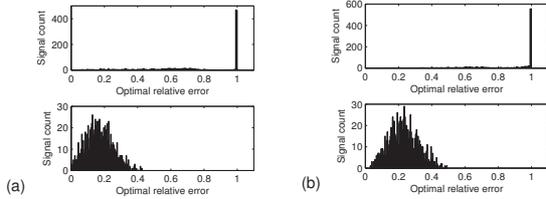


Figure 3: Denoising signals on low-dimensional MAP-Synthesis faces. (a) Results for signals on 2D faces (1,000 examples). Distributions of optimal errors obtained using MAP-Analysis (above) and MAP-Synthesis (below). (b) The same for signals on 3D faces (1,000 examples).

mentioned earlier that this is only a *necessary* condition for high-recoverability; however in many cases (specifically when the dictionary is normalized) this indeed turns out to produce well-recovered signals. For this experiment, we generated 1,000 signals on 2D faces, 1,000 on 3D faces, and so on up to 12D faces.

To quantify the performance of a specific method on a set of signals, we applied that method (in its energy-preserving form), with varying  $a$  values, to each of the signals in the set. We then selected, for each signal individually, the optimal  $a$  value  $a_{opt}$  and its associated relative error  $err_{opt} = \|\widehat{X}_{MAP}(a_{opt}) - X\|_2 / \|Y - X\|_2$  to represent the performance of the method on this signal. We collected the optimal errors for all signals in the set, and these were used to characterize the performance of the method on the entire set.

Figures (2)-(4) summarize the results. The first two present histograms of the optimal errors obtained on the principal signal sets and the MAP-Synthesis 2D and 3D signal sets. The final figure summarizes the results for all 12 sets of MAP-Synthesis signals.

Examining these results, we find that as anticipated, each method is successful in recovering its own sets of favorable signals; this therefore acts to support the geometrical models we have presented. It is also interesting to note that the two methods exhibit *comparable performance* when evaluated each on their own set of principal signals. This observation is particularly evident in the MAP-Synthesis principal signal test – figure (2(b)) – as the signals in this test are also principal to MAP-Analysis.

On the other hand, the results also depict a clear disparity between the two methods. We see that MAP-Analysis completely fails in recovering the MAP-Synthesis favorable signals, while MAP-Synthesis performs notably poorly compared to MAP-Analysis on its massive number of principal signals. These acute inconsistencies lead to the inevitable conclusion that the pseudo-inverse relation *does not* bridge between the two methods, as the theoretical models have anticipated.

#### 4. BEYOND DENOISING

Our discussion thus far was focused (for simplicity) on the denoising problem. However, many of the points made in this paper are in fact more general statements. Theorem 1, which established equivalence between the two prior types in the square case, is a general result which applies to the two prior structures. Most of our statements for the overcomplete case are also general, as they stem directly from the geometry of the distribution level-sets, and as such are independent of the specific choice of problem. Specifically, the

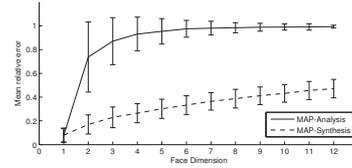


Figure 4: Denoising MAP-Synthesis highly recoverable signals. The graphs show the mean optimal errors obtained versus the MAP-Synthesis face dimension; error bars correspond to the standard deviation of the errors.

results provided in Theorem 6, as well as all conclusions which follow, still hold when considering the more general case. Finally, the equivalence for the  $\ell^2$  case is a general result as well.

The notable exception is the equivalence in the strict under-complete case, which *does not* reproduce in general. As it turns out, the fact that the MAP-Synthesis solution is constrained to the column-span of  $\mathbf{D}$  may become crucial in general, preventing the possibility of an equivalence.

#### 5. CONCLUSIONS

In this paper we have discussed and compared two popular MAP-based methods for inverse problems — the MAP-Analysis and the MAP-Synthesis formulations. We have shown that the two are essentially identical in the (under-)complete case, utilizing a pseudo-inverse relation between the dictionary and analysis operator. However, in the overcomplete case the two methods were shown to depart. We concentrated on the overcomplete  $\ell^1$  case, and found that the geometrical structures underlying the two approaches exhibit very different properties. This perspective has led to a generality relation of MAP-Synthesis over MAP-Analysis, though the actual transform remains impractical. Our theoretical results were demonstrated for the pseudo-inverse relation, where the two methods performed dramatically differently on large families of signals. We conclude that the two widely-used MAP-based methods retain a respectable distance between them. Whether any one may be considered superior, however, remains an open question.

**Acknowledgements** The authors would like to thank Prof. David L. Donoho for the enlightening discussions and ideas which helped in developing the presented work.

#### REFERENCES

- [1] Elad, M. (2002) On the bilateral filter and ways to improve it, *IEEE Trans Image Proc*, Vol. 11, No. 10, pp. 1141–1151.
- [2] Chen, S.S., Donoho, D.L. & Saunders, M.A. (2001) Atomic decomposition by basis pursuit, *SIAM Review*, Vol. 43, No. 1, pp. 129–59.
- [3] Donoho, D.L. & Elad, M. (2003) Optimally sparse representation in general (non-orthogonal) dictionaries via  $\ell^1$  minimization, *Proc Nat Acad Sci*, Vol. 100, No. 5, pp. 2197–2202.
- [4] Tropp, J.A. (2004) Greed is good: Algorithmic results for sparse approximation, *IEEE Trans Inf Theo*, Vol. 50, No. 10, pp. 2231–2242.
- [5] Donoho, D.L., Elad, M. & Temlyakov, V. (2004) Stable recovery of sparse overcomplete representations in the presence of noise, to appear, *IEEE Trans Inf Theo*.
- [6] Elad, M., Starck, J.L., Querre, P. & Donoho, D.L. (2005) Simultaneous cartoon and texture image inpainting using morphological component analysis, *J Appl Comput Harmon Anal*, Vol. 19, pp. 340–358.
- [7] Elad, M., Milanfar, P. & Rubinstein, R. (2005) Analysis versus synthesis in signal priors, *Technical Report – CS, Technion IIT*.
- [8] Rudin, L., Osher, S., & Fatemi, E. (1992) Nonlinear total variation based noise removal algorithms, *Physica D*, Vol. 60, pp. 259–268.
- [9] Donoho, D.L. (2005) High-dimensional centrally-symmetric polytopes with neighborliness proportional to dimension, *Technical Report – Statistics, Stanford*.
- [10] Strohmer, T. & Heath, R. (2003) Grassmannian frames with applications to coding and communication, *Arxiv preprint math.FA/0301135*.