IOP PUBLISHING INVERSE PROBLEMS

Inverse Problems **23** (2007) 947–968 doi:10.1088/0266-5611/23/3/007

Analysis versus synthesis in signal priors

Michael Elad[1], **Peyman Milanfar**[2] **and Ron Rubinstein**[1]

[1] Computer Science Department, Technion—Israel Institute of Technology, Haifa 32000, Israel
[2] Electrical Engineering Department, Baskin School of Engineering, University of California, Santa Cruz 95064, CA, USA

E-mail: ronrubin@cs.technion.ac.il

Received 2 December 2006, in final form 19 February 2007
Published 10 April 2007
Online at stacks.iop.org/IP/23/947

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. The algebraic similarity between the two suggests that they could be strongly related; however, in the absence of a detailed study, contradicting approaches have emerged. While the computationally intensive synthesis approach is receiving ever-increasing attention and is notably preferred, other works hypothesize that the two might actually be much closer, going as far as to suggest that one can approximate the other. In this paper we describe the two prior classes in detail, focusing on the distinction between them. We show that although in the simpler complete and undercomplete formulations the two approaches are equivalent, in their overcomplete formulation they depart. Focusing on the ℓ^1 case, we present a novel approach for comparing the two types of priors based on high-dimensional polytopal geometry. We arrive at a series of theoretical and numerical results establishing the existence of an unbridgeable gap between the two.

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

$$\underline{Y} = \mathbf{T}\{\underline{X}\} + \underline{V}, \tag{1}$$

where $\mathbf{T} : \mathbb{R}^N \to \mathbb{R}^M$ is a (possibly nonlinear) known operator, and $\underline{V} \in \mathbb{R}^M$ is a zero-mean white Gaussian additive noise vector (other models for the noise could also be considered, but here we restrict the discussion to the assumptions made above for simplicity). Many important problems in signal and image processing are represented using this structure: these include denoising, interpolation, scaling, super-resolution, inverse Radon transform, reconstruction from projections in general, and motion estimation, to name just a few. In all these problems, the general task is an inversion of the operator $\mathbf{T}$.

Inverting the above process 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 $\underline{Y}$. We get a solution of the form

$$\begin{aligned}\underline{\hat{X}}_{\mathrm{ML}} &= \operatorname*{Argmax}_{\underline{X}} \operatorname{Prob}\{\underline{Y}|\underline{X}\} \\ &= \operatorname*{Argmax}_{\underline{X}} \exp\left\{-\frac{1}{2\sigma_{\mathrm{v}}^2}\|\underline{Y} - \mathbf{T}\{\underline{X}\}\|^2\right\} \\ &= \operatorname*{Argmin}_{\underline{X}} \|\underline{Y} - \mathbf{T}\{\underline{X}\}\|_2^2.\end{aligned}$$

As an example, if $\mathbf{T}\{\underline{X}\} = \mathbf{H}\underline{X}$, where $\mathbf{H}$ is a known degradation operator represented as a full rank matrix with more columns than rows, the ML solution amounts to the pseudo-inverse of the degrading operator, thus $\underline{\hat{X}}_{\mathrm{ML}} = \mathbf{H}^{+}\underline{Y}$. For the denoising problem ($\mathbf{H} = \mathbf{I}$), ML suggests the solution $\underline{\hat{X}}_{\mathrm{ML}} = \underline{Y}$, which clearly demonstrates the weakness of ML.

Generally speaking, the literature today offers, through the Bayesian approach, a stabilized solution to the inverse problem posed above. We concentrate on the use of 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.

1.1. 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 form 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

$$\operatorname{Prob}\{\underline{X}\} = \operatorname{Const} \cdot \exp\left\{-\alpha \cdot \|\boldsymbol{\Omega}\underline{X}\|_p^p\right\},$$

where $\boldsymbol{\Omega} \in M^{[L\times N]}$ is some pre-specified matrix and $\|\cdot\|_p^p$ is the ℓ^p norm. The term $\|\boldsymbol{\Omega}\underline{X}\|_p^p$ is an energy functional that is supposed to be low for highly probable signals, and higher if the signal is less probable. We refer to $\boldsymbol{\Omega}$ as the *analysing operator*. Merged with the Gaussianity assumption on the additive noise, this poses the MAP recovery process as the minimization problem

$$\begin{aligned}\underline{\hat{X}}_{\text{MAP-A}} &= \operatorname*{Argmax}_{\underline{X}} \operatorname{Prob}\{\underline{X}|\underline{Y}\} \\ &= \operatorname*{Argmax}_{\underline{X}} \mathrm{P}\{\underline{Y}|\underline{X}\}\mathrm{P}\{\underline{X}\}/\mathrm{P}\{\underline{Y}\} \\ &= \operatorname*{Argmin}_{\underline{X}} -\log \mathrm{P}\{\underline{Y}|\underline{X}\} - \log \mathrm{P}\{\underline{X}\} \\ &= \operatorname*{Argmin}_{\underline{X}} \|\underline{Y} - \mathbf{T}\{\underline{X}\}\|_2^2 + \lambda \cdot \|\boldsymbol{\Omega}\underline{X}\|_p^p \qquad (2)\end{aligned}$$

where $\lambda = 2\alpha\sigma_v^2$. When robust norms are used ($p < 2$ or some robust M-function [1]), an iterative algorithm is typically employed for the minimization of (2). Preference is generally given to $p \geqslant 1$ 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 analysing its behaviour.

The analysis structure is quite common in inverse problems in signal processing, image processing and computer vision. In a typical image processing application where an image is an unknown, $\mathbf{\Omega}$ is chosen as some sort of derivative operator, promoting spatial smoothness in the image $\underline{X}$. As to the choice of p, choosing the ℓ^2 norm is known to lead to a simplified analytic treatment, but also known to give non-robust results (i.e. smoothing of discontinuities). Thus, recent contributions concentrate on robustness by using ℓ^p norms with $p < 2$, leading to nonlinear filtering algorithms [1–9].

1.2. 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 and Saunders [10].

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 $L \geqslant N$ (notice the different size compared to $\mathbf{\Omega}$). This matrix has N rows and L columns, and 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 $\underline{X}$:

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

In this expression, the ℓ^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 $\hat{\underline{X}}_{\text{MAP-S}} = \mathbf{D}\hat{\underline{\gamma}}$.

Synthesis-based methods have evolved rapidly over the past decade. Significant progress has been seen in the development of modern dictionaries for sparse image representation, such as the Ridgelet, Curvelet and Contourlet dictionaries [11–13]; training from example sets has also been successfully explored [14]. Parallel advancements, many of them theoretical in nature, have been achieved in the areas of sparse coding (i.e. finding sparse representations) and sparsity-based signal recovery [15, 16].

Through the MAP framework, the synthesis approach may be generalized to incomplete dictionaries. We let $\mathbf{\Gamma}_{\underline{X}} = \{\underline{\gamma} | \underline{X} = \mathbf{D}\underline{\gamma}\}$ denote the set of representations of $\underline{X}$ in $\mathbf{D}$, where $\mathbf{\Gamma}_{\underline{X}}$ 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:

$$\text{Prob}\{\underline{X}\} = \begin{cases} \text{Const} \cdot \exp\left\{-\alpha \cdot \|\hat{\underline{\gamma}}(\underline{X})\|_p^p\right\} & \text{if} \quad \mathbf{\Gamma}_{\underline{X}} \neq \emptyset \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

where

$$\hat{\underline{\gamma}}(\underline{X}) = \operatorname{Arg}\min_{\underline{\gamma}\in\Gamma_{\underline{X}}} \|\underline{\gamma}\|_p^p.$$

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 $p \geqslant 1$, we are guaranteed to have a unique solution.

1.3. Analysis versus synthesis

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

How do the two methods compare? The conjecture that natural images can be effectively described as sparse combinations of atomic elements has found empirical support [17] which the analysis-based approach lacks. The concept also has clear advantages in applications such as image compression, feature extraction, content-based image retrieval and others. Furthermore, as opposed to the analysis approach, the synthesis approach has a constructive form providing an explicit description of the signals it represents, and as such, is more intuitive to interpret and design.

A different concern about the analysis approach is its capacity to benefit from the increased redundancy. As this approach requires a signal to simultaneously agree with *all* the rows of $\mathbf{\Omega}$, this might become impossible with a highly redundant operator, rendering the prior useless. The synthesis approach, in contrast, seems to benefit from higher redundancy, as this enriches the prior, enabling it to describe more complex types of signals.

On the other hand, the compactness promoted by the synthesis approach might also come as a weakness. In such a framework where only a small number of atoms are used to represent each signal, the significance of every atom grows enormously; any wrong choice—in a denoising scenario for instance—could potentially lead to a 'domino effect' where additional erroneous atoms are selected as compensation, deviating further from the desired description. In the analysis formulation, however, all atoms take an equal part in describing the signal, thus minimizing the dependence on each individual one, and stabilizing the recovery process.

Analysis-based methods, specifically in their robust form ($p < 2$), are a very common structure in image processing and computer vision applications. In a large part, this is because MAP-analysis leads to a simple optimization problem, which (in the overcomplete case) is considerably easier to solve—due to the smaller dimension of the unknown—compared to a similar-sized MAP-synthesis form. At the same time, however, a growing number of works are employing the synthesis approach for inverse problem regularization. The synthesis-based approach is attractive due to its intuitive and versatile structure, and, informally, is widely considered to provide superior results. This recent trend is strengthened by a wealth of theoretical and practical advancements, making the synthesis approach both more appealing and computationally tractable [15, 16, 18].

Nonetheless, MAP-synthesis remains a prohibitive option in many cases. This has led several works to seek alternative approaches over direct minimization. One option which has been proposed is the use of an analysis-based method to approximate the synthesis-based one, as is done in [19] where the analysis operator is taken as the pseudo-inverse of the synthesis dictionary. This approach has only been partially justified, however, leaving the question of its generality much unattended.

1.4. This paper's contribution

As can be seen from the discussion, the literature to date is highly ambivalent with respect to the two regularization approaches. The extensive research of the synthesis-based methods implicitly suggests that MAP-synthesis is superior to MAP-analysis—especially considering the huge gap in complexity between the two structures. At the same time, other works, building on the algebraic similarity presented in the next section, hypothesize that the two are actually much closer, in fact close enough to approximate one another [19].

In light of these 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 indeed 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 undercomplete formulations, as well as for the ℓ^2 (i.e. $p = 2$) choice. However, as we go towards the general overcomplete formulation ($L > N$), we find that the equivalence between the two MAP options breaks. Concentrating on the $p = 1$ case, often favoured due to its convexity and robustness, we provide theoretical as well as numerical results indicating that the two methods are fundamentally distinct. Our results break, in fact, *both* of the above common assumptions: first in establishing the gap between the two approaches, and second by presenting simulations where the analysis approach actually supersedes its synthesis counterpart.

This paper is organized as follows. Section 2 describes the square and under-determined cases, where the two methods exhibit almost complete equivalence. In section 3 we turn to discuss the overcomplete case, focusing on the ℓ^1 choice. Taking a geometrical viewpoint, we construct the theoretical model describing the gap between the two methods, and discuss some consequences of this model. Simulation results are provided in section 4, and section 5 concludes with a summary of the claims made in the paper.

2. The square and under-determined cases

We begin by showing that in the (under-)determined case (i.e., $L \leqslant N$), the two methods are practically 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 analysing operator* $\mathbf{\Omega}$. *The equivalent MAP-synthesis method is obtained for the dictionary* $\mathbf{D} = \mathbf{\Omega}^{-1}$.

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

$$\hat{\underline{X}} = \mathbf{\Omega}^{-1} \cdot \operatorname*{Argmin}_{\underline{\gamma}} \|\underline{Y} - \mathbf{T}\{\mathbf{\Omega}^{-1}\underline{\gamma}\}\|_2^2 + \lambda \cdot \|\underline{\gamma}\|_p^p,$$

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. □

The generalization of theorem 1 for the $L \leqslant N$ case requires more care, and is only true for the denoising ($\mathbf{T} = \mathbf{I}$) case. Before stating the theorem, 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 $\mathbf{D}$, while MAP-analysis poses no such restriction. Nevertheless, the following theorem represents both conceptually and computationally a complete equivalence

between the two, as knowing the solution to either one immediately fixes the solution to the other. We arrive at the following result, whose proof is postponed until the appendix.

Theorem 2 (undercomplete denoising case—near-equivalence). *MAP-analysis denoising with a full-rank analysing operator* $\mathbf{\Omega} \in M^{[L\times N]}$ $(L \leqslant N)$ *is nearly equivalent to MAP-synthesis with the dictionary* $\mathbf{D} = \mathbf{\Omega}^+$. *This is expressed by the relation* $\underline{\hat{X}}_{\text{MAP-A}} = \underline{\hat{X}}_{\text{MAP-S}} + \underline{Y}^{\mathbf{D}\perp}$, *with* $\underline{Y}^{\mathbf{D}\perp}$ *representing the component of the input orthogonal to the columns of* $\mathbf{D}$.

The proof is given in appendix A.

We also see that when the input is in the column-span of $\mathbf{D}$ (as in the square non-singular case), we obtain $\underline{\hat{X}}_{\text{MAP-A}} = \underline{\hat{X}}_{\text{MAP-S}}$.

3. The over-determined case

We have seen that the two methods are practically equivalent for the $L \leqslant N$ case. Our main interest however is in the overcomplete $(L > N)$ case, advocated strongly by the basis pursuit approach. A natural starting point for analysing the overcomplete case is the pseudo-inverse relation, which, as we have just seen, successfully achieves equivalence in the (under-)complete case. We assume hereon that $\mathbf{\Omega}$ has full column rank, and hence $\mathbf{\Omega}^+\mathbf{\Omega} = I$. Beginning with the MAP-analysis formulation in (2), we let $\mathbf{\Omega}\underline{X} = \underline{\gamma}$. Since $\mathbf{\Omega}^+\mathbf{\Omega} = I$, recovering $\underline{X}$ from $\underline{\gamma}$ is done by $\underline{X} = \mathbf{\Omega}^+\underline{\gamma}$. 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 $\mathbf{\Omega}$, due to its definition (this can be represented by the constraint $\mathbf{\Omega}\mathbf{\Omega}^+\underline{\gamma} = \underline{\gamma}$). Thus we obtain the following equivalent MAP-analysis form:

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

Comparing this to (3), we see that if the MAP-synthesis solution (with $\mathbf{D} = \mathbf{\Omega}^+$) satisfies the constraint $\mathbf{\Omega}\mathbf{\Omega}^+\underline{\gamma} = \underline{\gamma}$, 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 solutions could differ vastly, the final estimators $\underline{\hat{X}} = \mathbf{\Omega}^+\underline{\hat{\gamma}}$ in both might be very similar; this is because in multiplying by $\mathbf{\Omega}^+$ we null out content not in the column-span of $\mathbf{\Omega}$, essentially satisfying the constraint. However, as we will see, this does not turn out to close the gap between the two methods. The exception to this is the non-robust ℓ^2 case, in which equivalence still holds.

Theorem 3 (overcomplete case—equivalence for $p = 2$). *MAP-analysis with a full-rank analysing operator* $\mathbf{\Omega} \in M^{L\times N}$ $(L > N)$ *is equivalent to MAP-synthesis with* $\mathbf{D} = \mathbf{\Omega}^+$ *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 $\mathbf{\Omega}$, and $\underline{\gamma}^{\Omega\perp}$ from the orthogonal subspace. The second penalty term $\left(\|\underline{\gamma}\|_2^2\right)$ clearly prefers $\underline{\gamma}^{\Omega\perp}$ to be zero; as to the first term $\left(\|\underline{Y} - \mathbf{T}\{\mathbf{\Omega}^+\underline{\gamma}\}\|_2^2\right)$, $\underline{\gamma}^{\Omega\perp}$ has no impact on it as it is nulled out by $\mathbf{\Omega}^+$. Thus, $\underline{\gamma}^{\Omega\perp}$ that violates the constraint in $\underline{\gamma}$ is chosen as zero, and the two methods coincide. □

3.1. MAP-analysis and MAP-synthesis in ℓ^1

From this point on we consider the two MAP methods with $p = 1$. The ℓ^1 choice is essentially the 'meeting point' between the analysis and synthesis approaches, which prefer $p \geqslant 1$ and

$0 \leqslant p \leqslant 1$, respectively. The use of the ℓ^1 norm in signal and image recovery has received considerable attention beginning at the late 1980s, with the adoption of robust statistics by the signal processing community. Probably most notable of the analysis-based methods is the total-variation approach [2][3], with some additional examples including [6–9]. Classical synthesis-based methods include the basis pursuit method [10] and the Lasso [20].

For the ℓ^1 choice, we have the following forms of the two recovery processes:

$$\underline{\hat{X}}_{\text{MAP-A}} = \underset{\underline{X}}{\text{Argmin}} \|\underline{Y} - \mathbf{T}\{\underline{X}\}\|_2^2 + \lambda \cdot \|\mathbf{\Omega}\underline{X}\|_1$$

$$\underline{\hat{X}}_{\text{MAP-S}} = \mathbf{D} \cdot \underset{\underline{\gamma}}{\text{Argmin}} \|\underline{Y} - \mathbf{T}\{\mathbf{D}\underline{\gamma}\}\|_2^2 + \lambda \cdot \|\underline{\gamma}\|_1.$$

The ℓ^1 option is a favourable choice for these methods due to its combination of convexity, robustness, as well as proximity to ℓ^0 in the synthesis case [15, 18].

Looking at the two MAP formulations, we see that both depend on a weighting parameter λ to control the regularizing element; for $\lambda = 0$ both reproduce the ML estimator, and as $\lambda \to \infty$ they deviate from it until finally converging to 0. However, the rate at which this occurs may vary substantially between the two methods, and hence this parametrization is inconvenient for our purposes. To overcome this, we propose the following reformulations of the two problems:

$$\underline{\hat{X}}_{\text{MAP-A}}(a) = \underset{\underline{X}}{\text{Argmin}} \|\mathbf{\Omega}\underline{X}\|_1 \qquad \text{Subject To} \quad \|\underline{Y} - \mathbf{T}\{\underline{X}\}\|_2 \leqslant a$$

$$\underline{\hat{X}}_{\text{MAP-S}}(a) = \mathbf{D} \cdot \underset{\underline{\gamma}}{\text{Argmin}} \|\underline{\gamma}\|_1 \qquad \text{Subject To} \quad \|\underline{Y} - \mathbf{T}\{\mathbf{D}\underline{\gamma}\}\|_2 \leqslant a.$$

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

3.2. A geometrical viewpoint

The above formulations have a simple geometrical interpretation, which provides an interesting way of comparing the two MAP approaches. The solutions of both problems are obviously confined to the same region of 'radius' a about $\underline{Y}$ (this is true as we assume $\mathbf{D}$ to be full-rank); we also assume this region does not include the origin, otherwise the solution is trivially zero. Considering MAP-analysis first, the level-sets of its target function $f_{\text{A}}(\underline{X}) = \|\mathbf{\Omega}\underline{X}\|_1$ are a collection of concentric, centro-symmetric polytopes $\{\underline{X} | \|\mathbf{\Omega}\underline{X}\|_1 \leqslant c\}$. Graphically, the solution can be obtained by taking a small level-set $\{\|\mathbf{\Omega}\underline{X}\|_1 \leqslant c\}$ about the origin, and gradually inflating it (by increasing c) until it first encounters the region $\{\|\underline{Y} - \mathbf{T}\{\underline{X}\}\|_2 \leqslant a\}$. The point of intersection constitutes the solution to the MAP-analysis problem, as there cannot be a point in this region having a smaller value of $\|\mathbf{\Omega}\underline{X}\|_1$.

As to MAP-synthesis, a similar process may be described using the collection of concentric, centro-symmetric polytopes $\mathbf{D} \cdot \{\underline{\gamma} | \|\underline{\gamma}\|_1 \leqslant c\}$[4]. This is reasoned as follows: consider the set $\mathbf{D} \cdot \{\|\underline{\gamma}\|_1 \leqslant c\}$ where c is small enough such that this set does not intersect the region $\{\|\underline{Y} - \mathbf{T}\{\underline{X}\}\|_2 \leqslant a\}$. Then for any $\underline{X}$ in this region, there does not exist a representation $\underline{\gamma}$ satisfying $\|\underline{\gamma}\|_1 \leqslant c$, or in other words, any representation as $\underline{X} = \mathbf{D}\underline{\gamma}$ must satisfy $\|\underline{\gamma}\|_1 > c$. This, of course, is true for any c which is small enough; therefore if we inflate this set (by enlarging c) until it first touches the region at the value $\hat{c}$, then for the

[3] Total variation takes a 'true' MAP-analysis form only in the 1D case.

[4] Note that these sets exist in *signal space*, and have the explicit form $\{\underline{X} | \exists \underline{\gamma}, \underline{X} = \mathbf{D}\underline{\gamma} \wedge \|\underline{\gamma}\|_1 \leqslant c\}$.

intersection point $\hat{\underline{X}} = \mathbf{D}\hat{\underline{\gamma}}$ we know it has a representation satisfying $\|\hat{\underline{\gamma}}\|_1 = \hat{c}$, whereas for any $c < \hat{c}$ the signals within the region have no such representation, and hence $\hat{\underline{X}}$ must be the MAP-synthesis solution.

Conveniently, for both MAP methods these 'inflations' are performed via simple scaling: we have $\{\|\mathbf{\Omega}\underline{X}\|_1 \leqslant c\} = c \cdot \{\|\mathbf{\Omega}\underline{X}\|_1 \leqslant 1\}$ and $\mathbf{D}\{\|\underline{\gamma}\|_1 \leqslant c\} = c \cdot \mathbf{D}\{\|\underline{\gamma}\|_1 \leqslant 1\}$. This implies that given the canonical *MAP defining polytopes* $\Psi_{\mathbf{\Omega}} := \{\|\mathbf{\Omega}\underline{X}\|_1 \leqslant 1\}$ and $\Phi_{\mathbf{D}} := \mathbf{D} \cdot \{\|\underline{\gamma}\|_1 \leqslant 1\}$, the inflation processes are fully defined, and so are the MAP solutions; in fact, specifying these polytopes is completely equivalent to specifying $\mathbf{\Omega}$ or $\mathbf{D}$, respectively. We find that the behaviour of each of the methods is governed exclusively by the geometry of a single high-dimensional polytope, providing us with the basis for comparing the two methods[5]. We therefore continue by characterizing the geometry of these two polytopes.

Before continuing, we briefly review some elementary polytope terminology. Given an N-dimensional polytope, its boundary is an $(N-1)$-dimensional manifold; each of the polytope's *facets* is an $(N-1)$-dimensional surface constituting one segment of this manifold. A facet may also be referred to as an $(N-1)$-*dimensional face*. Similarly, the boundary of each facet consists of $(N-2)$-*dimensional faces*, and so on. A polytope's vertices, edges and ridges are its faces of dimensions 0, 1 and 2, respectively.

3.2.1. The MAP-analysis defining polytope. The MAP-analysis defining polytope is a level set of the MAP-analysis target function, $f_{\mathrm{A}}(\underline{X}) = \|\mathbf{\Omega}\underline{X}\|_1$:

$$\Psi_{\mathbf{\Omega}} = \{\underline{X} | \|\mathbf{\Omega}\underline{X}\|_1 \leqslant 1\}.$$

Applying the gradient operator to f_{A}, we find that the normal to this surface satisfies

$$\underline{n}(\underline{X}) \propto \nabla f_{\mathrm{A}}(\underline{X}) = \mathbf{\Omega}^T \operatorname{sign}(\mathbf{\Omega}\underline{X}).$$

Evidently $\underline{n}(\underline{X})$ is defined for any $\underline{X}$ in which all coordinates of $\mathbf{\Omega}\underline{X}$ are non-zero; where one or more of these vanishes, $\underline{n}(\underline{X})$ exhibits a discontinuity arbitrarily filled in by the sign function. $\underline{n}(\underline{X})$ is therefore (as expected) piecewise-smooth. Intuitively, consider the signals $\underline{X}$ on the boundary of the defining polytope; then the facets correspond to the locations where $\underline{n}(\underline{X})$ is smooth, whereas the other faces correspond to where $\underline{n}(\underline{X})$ is discontinuous. The discontinuities in $\underline{n}(\underline{X})$ obviously result from $\underline{X}$ being orthogonal to rows in $\mathbf{\Omega}$; the following claim, whose proof is provided in the appendix, relates the face dimension to the rank of these rows:

Claim 1. *Let $\underline{X} \in \partial\Psi_{\mathbf{\Omega}}$ (the boundary of the defining polytope), and let k denote the rank of the rows in $\mathbf{\Omega}$ to which $\underline{X}$ is orthogonal to. Then $\underline{X}$ resides strictly within a face of dimension $(N-k-1)$ of the MAP-analysis defining polytope.*

The proof is given in appendix B.

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 ϵ-ball about it—of the same dimension as the face—entirely contained within this face (note that this also 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.

The claim implies that to obtain a vertex of $\Psi_{\mathbf{\Omega}}$, we choose $N-1$ linearly independent rows in $\mathbf{\Omega}$, determine their 1D null-space $\underline{v}$ and normalize such that $\|\mathbf{\Omega}\underline{v}\|_1 = 1$ (note that

[5] In fact, the same arguments hold for any ℓ^p formulation, replacing the ℓ^1-norms in the definitions of $\Psi_{\mathbf{\Omega}}$ and $\Phi_{\mathbf{D}}$ with the proper ℓ^p-norms. However, analysing these defining shapes for a general p is a difficult task, and thus we restrict ourselves to the ℓ^1 case.

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 *vertex complexity* of the MAP-analysis defining polytope, as its vertex count is equal to the number of possible choices of $N-1$ linearly independent rows in $\mathbf{\Omega}$. In the worst case, this may reach an exponential $\binom{L}{N-1}$, and in fact this is a *tight bound* for the worst case. As an example, assume the rows of $\mathbf{\Omega}$ are chosen such that their directions $\{\underline{\hat{w}}_i\}$ are uniformly distributed on the unit sphere. Under these conditions, the probability of any set of $N-1$ rows to be dependent *vanishes* for all practical purposes, and thus we obtain that for this randomized case the expected number of MAP-analysis vertices achieves $\Theta\binom{L}{N-1}$. Obviously this is also the tight bound for the worst-case vertex count.

An interesting observation is that the MAP-analysis defining polytope exhibits a highly regular structure. For instance, consider the set of edges associated with some choice of $N-2$ independent rows from $\mathbf{\Omega}$. Letting $\{\underline{u}, \underline{v}\}$ span their 2D null-space, these edges are obtained as any linear combination of the two (for instance, of the form $\underline{X} = \cos(\theta)\underline{u} + \sin(\theta)\underline{v}$), properly normalized to ensure $\|\mathbf{\Omega}\underline{X}\|_1 = 1$. It follows that this set of edges forms a closed *edge-loop* of the polytope; the planar edge loop consists of consecutive edges, all existing on a common plane. We conclude that the edges of $\Psi_{\mathbf{\Omega}}$ are arranged in 'loops' about the origin, each loop associated with a choice of $N-2$ independent rows from $\mathbf{\Omega}$. Similar arguments generalize to higher-dimensional regularities, corresponding to the choices of $N-k$ independent rows from $\mathbf{\Omega}$ for $k > 2$.

Finally, the organized structure is also found in a highly regular *neighbourliness pattern*. Since every vertex is obtained as the null-space of some $N-1$ rows from $\mathbf{\Omega}$, and each choice of $N-2$ of these defines an edge loop passing through this vertex, we have that each vertex of $\Psi_{\mathbf{\Omega}}$ is incident to *exactly* $N-1$ edge loops, and consequently, every vertex of $\Psi_{\mathbf{\Omega}}$ has precisely $2(N-1)$ neighbours.

3.2.2. The MAP-synthesis defining polytope. The MAP-synthesis defining polytope is given by

$$\Phi_{\mathbf{D}} = \mathbf{D} \cdot \{\underline{\gamma} | \|\underline{\gamma}\|_1 \leqslant 1\}.$$

It is a known result that this polytope is obtained as the convex hull of the columns of $\mathbf{D}$ and $-\mathbf{D}$; a proof is brought in the appendix for completeness:

Claim 2. *The MAP-synthesis defining polytope* $\Phi_{\mathbf{D}} = \mathbf{D} \cdot \{\|\underline{\gamma}\|_1 \leqslant 1\}$ *is obtained as the convex hull of* $\{\pm\underline{d}_i\}_{i=1,\ldots,L}$, *where* $\{\underline{d}_i\}$ *are the columns of* $\mathbf{D}$.

The proof is given in appendix C.

The claim simply states that the vertices of the MAP-synthesis defining polytope 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 neighbouring vertices. A vertex can therefore be represented as $\underline{V} = \mathbf{D}\underline{\gamma}$, where $\underline{\gamma}$ has a single non-zero element $\gamma_i = \pm 1$, and a point on an edge can be represented similarly with $\underline{\gamma}$ having two non-vanishing elements γ_i, γ_j satisfying $|\gamma_i| + |\gamma_j| = 1$. In general, a point on a k-dimensional face will have a representation $\underline{X} = \mathbf{D}\underline{\gamma}$ with $\underline{\gamma}$ having $k+1$ non-vanishing elements, and $\|\underline{\gamma}\|_1 = 1$. We emphasize that this is *not* a sufficient condition, so a signal $\underline{X} = \mathbf{D}\underline{\gamma}$ synthesized from a sparse representation $\underline{\gamma}$ might not reside on a low-dimensional face if the corresponding columns of $\pm\mathbf{D}$ are not neighbours or do not constitute polytope vertices.

An immediate implication of claim 2 concerns the redundancy of certain atom signals in $\mathbf{D}$. From the claim, it is clear that any column of $\mathbf{D}$ residing strictly within the convex hull of

the remaining columns has absolutely no effect on the MAP-synthesis defining polytope and thus can be removed.

Corollary 1. *Let* $\underline{d}_k$ *be a column of* **D** *which is obtained as a convex combination of the remaining columns and their antipodes,* $\{\pm\underline{d}_i\}_{i=1..\hat{k}..L}$. *Then the MAP-synthesis problem obtained by removing* $\underline{d}_k$ *from* **D** *is equivalent to the original one.*

Redundant columns in **D** can be safely removed without altering the MAP-synthesis solution, and by locating these we may be able to prune the dictionary, generally obtaining a simpler formulation. The problem of determining whether some vector $\underline{X}$ is a convex combination of the set $\{\underline{Y}_i\}$ can be formulated as a linear-programming (LP) problem, and thus locating all redundant columns in **D** requires L executions of LP. As an alternative to removal, we may choose to elongate the redundant atom such that it becomes a vertex of the MAP-synthesis defining polytope, and thus expressed by the prior. However, increasing a dictionary atom may have the effect of assimilating a different one into the convex hull. One simple method to ensure none of the columns in **D** are redundant is to normalize them to a *fixed length* (see section 3.3.3 and specifically claim 3).

3.3. Consequences of the geometrical viewpoint

The geometrical analysis leads to some important consequences concerning the two MAP methods. In this section we describe a few of these conclusions.

3.3.1. The analysis-synthesis gap. From the geometrical viewpoint, we find that in contrast to the algebraic similarity, the analysis and synthesis structures are actually very different. As we have seen, the two polytopal structures asymptotically differ in their vertex counts. A parallel difference exists in the neighbourliness properties of these polytopes; since every vertex has a linear number of neighbours in the MAP-analysis case (while their total number is exponential) it follows that the probability of any two vertices to be neighbours approaches 0 as $N \to \infty$. In contrast, Donoho [21] has recently shown that, for MAP-synthesis polytopes, the probability of any two (non-antipodal) vertices to be neighbours approaches 1 as $N \to \infty$.[6] We find that while MAP-analysis polytopes feature very *large* numbers of vertices with very *low* neighbourliness, MAP-synthesis polytopes exhibit *low* vertex counts and very *high* neighbourliness. Combined with the high regularity of the MAP-analysis polytopes, we see that the two approaches actually describe very different structures. These theoretical gaps indeed translate to very concrete behavioural differences between the two methods, and this will be shown in the experiments section.

3.3.2. MAP-synthesis as a superset of MAP-analysis. An interesting consequence of the geometrical description is that any ℓ^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—and setting them 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 ℓ^1:

[6] The dictionary is assumed to be of linear size in N, as well as to fulfil certain randomness conditions; see theorem 1 in [21].

Theorem 4 (overcomplete ℓ^1 case—generality of MAP-synthesis). *For any ℓ^1 MAP-analysis form with full-rank analysing operator $\mathbf{\Omega}$ ($L \geqslant N$), there exists a dictionary $\mathbf{D}(\mathbf{\Omega})$ describing an equivalent ℓ^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 not apply to MAP-synthesis, it may clearly describe structures not represented in the MAP-analysis form.

The actual equivalence transform presented here has little practical value; except for the special case of $N = 2$, where the size of $\mathbf{D}(\mathbf{\Omega})$ will be equal to (or even smaller than) that of $\mathbf{\Omega}^T$, the size of $\mathbf{D}(\mathbf{\Omega})$ 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 indeed a *vast collection* of MAP-synthesis priors unrepresented by the stricter MAP-analysis form.

3.3.3. MAP principal signals. The constructive nature of MAP-synthesis provides a good understanding of the signals which are most 'favoured' by this prior; in essence, these are the dictionary atoms and their sparse combinations. The parallel entities for the MAP-analysis prior, however, are difficult to derive using algebraic tools. The geometric interpretation enables us to define these qualitative terms in a precise manner, and give a description of the MAP-analysis counterparts of the synthesis atoms.

Roughly speaking, we consider a signal to be favoured by some prior when this prior is capable of recovering the signal well given deteriorated versions of it; intuitively, these should be the signals with maximal *a priori* probability. However, we observe that both MAP structures are energy dependent; therefore, the most probable signals for both are simply the zero signal and its immediate neighbourhood. Moreover, the intuition itself here is not entirely accurate: a highly probable signal will not be well recovered if there exists a near-by signal with even higher probability.

To resolve this, we confine ourselves to a *fixed-energy* sphere; on this sphere we seek the most effectively recovered signals by the specific MAP method. Since the recovery is a local process, we will further be interested in the *local* maxima of the distribution on this sphere rather than the global ones. Our line of thought can be described as follows. Consider an energy-preserving denoising process, 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 neighbourhood of the input on the fixed-energy sphere, outputting a higher-probability (and presumably less noisy) signal near the input. A signal will therefore be well recovered when its prior probability is maximal relative to a significant enough part of its neighbourhood on the fixed-energy sphere. Specifically, the *local maxima* of the distribution will be the most effectively recovered signals on the sphere.

Reducing without loss of generality to the unit sphere, we refer to the local maxima of the distribution as the *principal signals* of the distribution. Formally,

Definition 1. *Let* $\text{Prob}\{\underline{X}\}$ *be any MAP-analysis or MAP-synthesis distribution. Then the principal signals of this distribution are defined as the* {*local*} *maxima of the optimization problem*

$$\underset{\underline{X}}{\text{Argmax}}\ \text{Prob}\{\underline{X}\} \qquad \text{Subject To} \quad \|\underline{X}\|_2 = 1.$$

As we will soon see, in the synthesis case these signals are tightly related to the MAP-synthesis dictionary atoms.

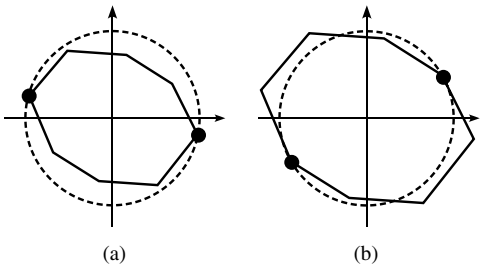

Figure 1. Principal signals and the MAP defining polytope. The dotted circles denote the unit sphere in 2D signal space. The two polygons are different scales of the same MAP defining polytope. (a) A principal signal, intersected by a vertex of the defining polytope. (b) A vertex which is not a principal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

The geometry of the MAP defining polytope directly dictates the behaviour of the distribution on the unit sphere, and consequently the locations of the principal signals on it. For both priors, the boundaries of the defining polytopes define iso-surfaces of signals with equal *a priori* probability; these have the form $r \cdot \partial\Psi_{\mathbf{\Omega}}$ or $r \cdot \partial\Phi_{\mathbf{D}}$—where $r \in \mathbb{R}^+$ is a non-negative scaling factor—and, for increasing r, represent decreasingly probable signals. Beginning with such an iso-surface $r \cdot \partial\Psi_{\mathbf{\Omega}}$ or $r \cdot \partial\Phi_{\mathbf{D}}$, with small enough r such that it is entirely bounded by the unit sphere, then as r is increased, the surface intersects the sphere at decreasingly probably locations, until finally completely enclosing it. Clearly, to be a local maximum a signal must be intersected by the inflating iso-surfaces before its surrounding neighbourhood. Consequently, such a local maximum is intersected by an extreme point—a *vertex*—of the polytope. We conclude that the MAP principal signals project to *vertices* of the MAP defining polytope.

We immediately point out, however, that projection onto a vertex is only a *necessary* condition for principality, as demonstrated in figure 1. Simulation results show a dramatic difference in the recovery performance of principal versus non-principal polytope vertices.

For a vertex to be principal, it must be maximally distant from the origin relative to all the directions about it on the boundary of the defining polytope. Luckily, determining this only requires examining those directions from the vertex to its one-dimensional incident edges (this follows from the fact that for any scalar function, the convex combination of a set of descent directions is also a descent direction).

In the case of MAP-synthesis, its defining polytope vertices are a subset of the dictionary atoms; hence the principal signals are a subset of these atoms. However, not all atoms constitute polytope vertices, and only a few of these are actually principal. Furthermore, determining which of the atoms are vertices is a difficult task, and so is the task of determining the incident edges of each vertex. However, given an atom $\underline{d}$, a simple work-around to determine its principality is to examine *all* line segments connecting $\underline{d}$ with the remaining atoms and their antipodes. If $\underline{d}$ is found to be maximally distant relative to all these line segments, clearly it is a vertex as well as a principal signal; on the other hand, if $\underline{d}$ is found not to be maximal relative to some segment, it immediately follows that it is not principal.

In practice, many MAP-synthesis dictionaries have their atoms *normalized* to a fixed length. As we mentioned earlier (without proof), this ensures that all the atoms constitute defining polytope vertices. However, for such dictionaries, a stronger claim can be made: indeed, when the atoms are normalized, they all constitute principal signals of the MAP distribution. We have the following result, whose proof is provided in the appendix.

Claim 3 (principal signals of MAP-synthesis with a normalized dictionary). *Let* $\mathbf{D}$ *be a MAP-synthesis dictionary with fixed-energy columns. Then the dictionary atoms coincide with the principal signals of the MAP-synthesis prior.*

The proof is given in appendix D.

In the general case, however, the MAP-synthesis principal signals remain a subset of the dictionary atoms. Since dictionaries in practice are commonly normalized, this distinction is not usually made. Nevertheless, when the dictionary atoms are not normalized, the difference in recovery performance can be substantial; while the principal signals are truly 'favoured' by the prior, other atoms might not be at all.

In the MAP-analysis case, the distinction becomes more significant. The number of MAP-analysis vertices is exponentially large, and empirical evidence suggests that most of these are non-principal and not well recovered. Unfortunately, we are not currently aware of any simple analytical method for characterizing the MAP-analysis principal signals. Nonetheless, these signals can be generated by a computer. For the simulations in this paper we used a simple traversal algorithm for locating these signals; this enabled us to produce large sets of MAP-analysis principal signals and study their behaviour.

Our traversal algorithm locates one principal signal at a time. Beginning with some initial vertex $\underline{v}$, we examine its incident edge-loops, and for each loop, we determine $\underline{u}$ such that $\{\underline{v}, \underline{u}\}$ orthogonally span the plane in which the loop exists. Assuming a small enough ϵ, $\underline{v}$'s infinitesimal neighbours on this edge loop can be approximated by $\underline{v}_+ = (\underline{v} + \epsilon\underline{u})/\|\mathbf{\Omega}(\underline{v} + \epsilon\underline{u})\|_1$ and $\underline{v}_- = (\underline{v} - \epsilon\underline{u})/\|\mathbf{\Omega}(\underline{v} - \epsilon\underline{u})\|_1$, where the normalization is applied to ensure $\|\mathbf{\Omega}\underline{v}_+\|_1 = \|\mathbf{\Omega}\underline{v}_-\|_1 = 1$. By comparing the ℓ^2 norms of $\underline{v}$, $\underline{v}_+$ and $\underline{v}_-$, we determine whether $\underline{v}$ is maximal relative to its two incident edges on this edge loop. Now, if $\underline{v}$ is found to be maximal relative to all its incident edges, it is a principal signal. Otherwise, it is not maximal relative to some incident edge. In this case we replace it with a vertex with larger ℓ^2-norm from the violating edge loop (in our implementation, we choose the one with largest ℓ^2-norm in the loop), and continue the traversal. This swapping continues until a local maximum is encountered, providing one MAP-analysis principal signal. The entire process is then repeated using a new vertex as a starting point.

4. Numerical results

The geometrical viewpoint reveals a large gap between the two formulations in the over-determined ℓ^1 case. In this section we provide some simulation results, demonstrating this theoretical gap.

4.1. Synthetic experiments

The following synthetic experiments demonstrate how the gap can be easily brought to an extreme even in a simple case. To obtain these results we compared the two methods on their most favourable signals: their principal signals.

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 behaviours of the two methods. We selected the 128×256 *identity-Hadamard* dictionary $\mathbf{D} = \frac{1}{\sqrt{2}}[\mathbf{I}\ \mathbf{H}]$ and its pseudo-inverse $\mathbf{\Omega} = \mathbf{D}^T = \frac{1}{\sqrt{2}}[\mathbf{I}\ \mathbf{H}]^T$ 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 favourable for MAP-synthesis methods [22, 23].

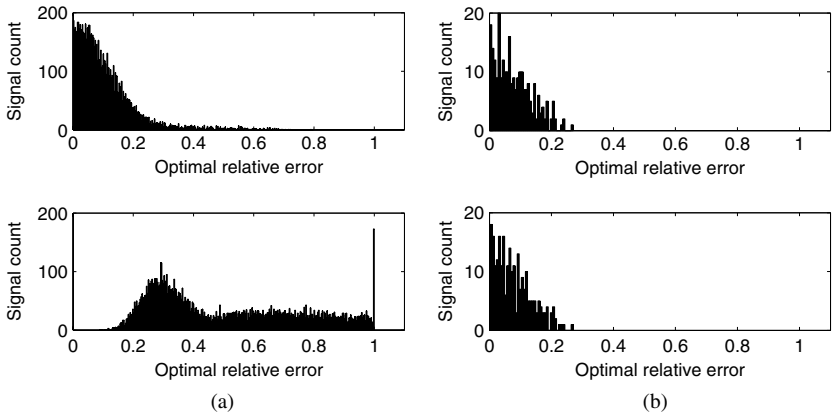

Figure 2. Denoising MAP principal signals. (a) Results for MAP-analysis principal signal (10 000 examples): distributions of optimal errors obtained using MAP-analysis (above) and MAP-synthesis (below). (b) The same for MAP-synthesis principal signals (256 examples).

The dictionary size immediately limits the number of distinct MAP-synthesis principal signals to a mere 256. In contrast, MAP-analysis boasts an enormous number of them: our traversal algorithm easily produced 10 000 such signals. What is more, our program was designed to reject new signals if these resided in a radius of <0.1 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 $(\frac{L}{N-1}) = (\frac{256}{127}) \approx 10^{75}$). These are obviously impressive numbers compared to the modest number of MAP-synthesis principal signals.

An interesting point in this experiment is that the MAP-synthesis principal signals in our case all double as MAP-analysis principal signals. To sharpen the comparison, we therefore generated additional sets of preferable MAP-synthesis signals, which we obtained on low-dimensional faces of the MAP-synthesis defining polytope (i.e. sparse combinations of atoms). For the experiment, we generated 1000 signals on 2D faces, 1000 on 3D faces, and so on up to 12D faces.

To quantify the performance of a specific method on a set of signals, we generated noisy versions of the signals in the set, and applied the method (in its energy-preserving form), with varying a values, to each of the contaminated signals. We then selected, for each signal individually, the optimal a value a_{opt} and its associated relative error $\text{err}_{\text{opt}} = \|\widehat{\underline{X}}_{\text{MAP}}(a_{\text{opt}}) - \underline{X}\|_2 / \|\underline{Y} - \underline{X}\|_2$ 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.

The results demonstrate several points. First, we see that each method is indeed successful in recovering its own sets of principal signals; this agrees with the predictions of the geometrical model. Also interesting is the fact that the two methods exhibit comparable performance when evaluated each on their own set of principal signals; this observation is particularly evident from figure 2(b), where the signals are simultaneously principal to both MAP-analysis and MAP-synthesis.

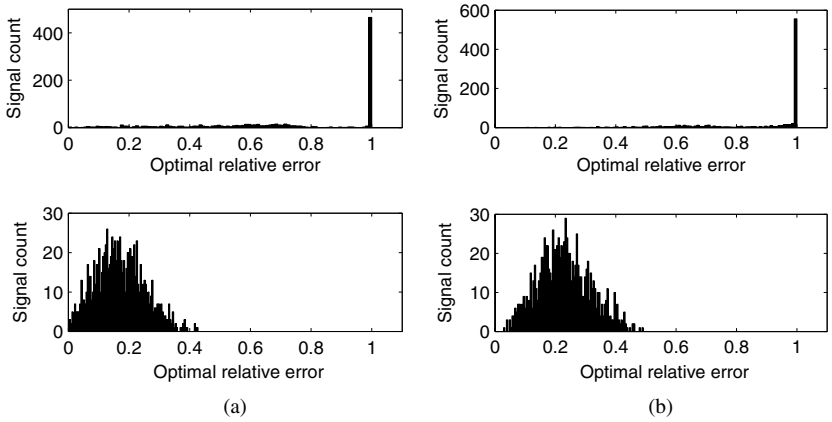

Figure 3. Denoising signals on low-dimensional MAP-synthesis faces. (a) Results for signals on 2D faces (1000 examples): distributions of optimal errors obtained using MAP-analysis (above) and MAP-synthesis (below). (b) The same for signals on 3D faces (1000 examples).

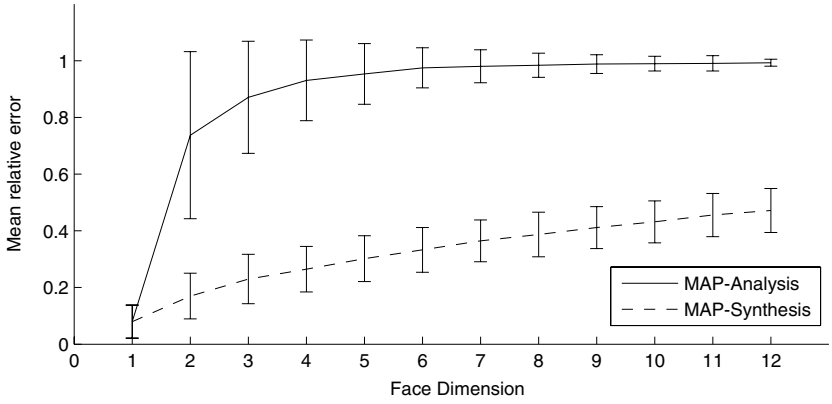

Figure 4. Denoising MAP-synthesis highly recoverable signals. The graphs show the mean optimal errors obtained versus the MAP-synthesis face dimension; error bars correspond to the standard deviation of the errors.

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 favourable signals, while MAP-synthesis performs notably poorly compared to MAP-analysis on its massive number of principal signals. The results also illustrate the asymptotical nature of the gap between the two approaches in the number of principal signals each one accepts.

The acute inconsistencies lead to the inevitable conclusion that the pseudo-inverse relation does not bridge between the two methods. Moreover, we see here that the difference in complexity between the two structures has a strong expression in practice, indicative of an inherent gap between the two formulations. Though the experiment specifically utilizes the pseudo-inverse relation, the gap depicted here cannot be associated with this specific choice; indeed, any reasonably sized MAP-synthesis dictionary will be limited in the number of favourable signals it can accommodate, and consequently in its ability to handle the large number of MAP-analysis principal signals. In the other direction, any attempt to adapt a

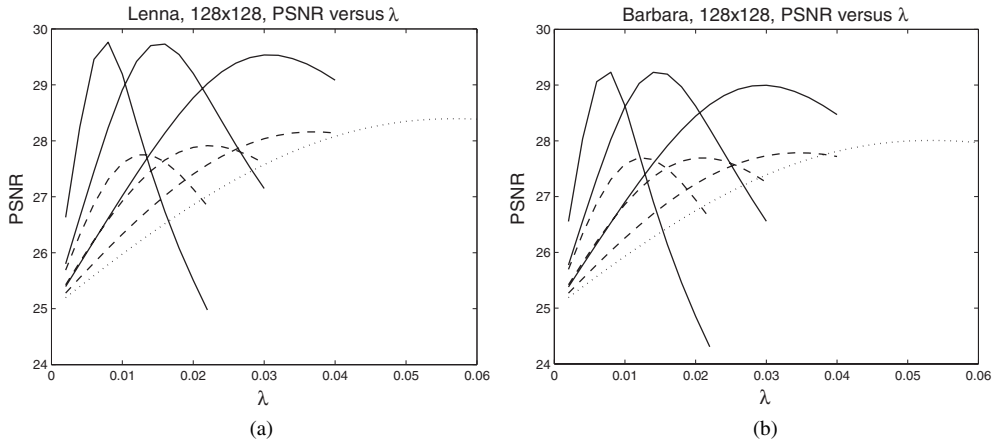

Figure 5. Image denoising using the redundant DCT transform. Solid lines (left to right): MAP-analysis with block shifts of 1, 2 and 4 pixels; dashed lines (left to right): MAP-synthesis with block shifts of 1, 2 and 4 pixels; dotted line: MAP-analysis/MAP-synthesis with a block shift of 8 pixels (unitary transform). Images are of size 128 × 128. (a) Results for *Lenna*; (b) results for *Barbara*. Images downloaded from http://www.wikipedia.com, and downscaled using bilinear interpolation.

MAP-analysis prior to a given set of MAP-synthesis signals is bound to give rise to an enormous number of additional (unwanted) favourable signals.

4.2. Real-world experiments

In this section we present some comparative denoising results obtained for actual image data. For these experiments we selected the *overcomplete DCT* transform; this transform partitions the image into overlapping blocks, and applies to each block a unitary DCT transform. The overcomplete DCT transform constitutes a tight frame when all image pixels are covered by an equal number of blocks. Our experiments used 8 × 8 blocks, with a shift of either 1, 2 or 4 pixels between neighbouring blocks. We also used shifts of 8 pixels (i.e. no overlap, leading to a unitary transform) as reference. Boundary cases were handled by assuming periodicity, ensuring the tight frame condition.

Since the transform is tight, the synthesis dictionary was simply taken as the transpose of the analysis operator, leading to a dictionary constructed of 8 × 8 DCT bases in all possible shifts over the image domain. Motivations for choosing this transform include: (1) the transform is widely used in image processing, and has been employed in both analysis and synthesis frameworks; (2) it is a tight frame, and has an efficient implementation; and (3) it is highly redundant, whilst offering a convenient way for controlling its redundancy (specifically, 4× for a shift size of 4, 16× for a shift size of 2 and 64× for a shift size of 1).

We ran the experiments on a collection of standard test images, including *Lenna*, *Barbara* and *Mandrill*. Each of these was downscaled to a size of 128 × 128 to reduce computation costs. We added white Gaussian noise to each source image, producing 25 dB PSNR inputs. Each input was denoised using both MAP-analysis and MAP-synthesis with varying λ values, and the output PSNR was determined for each value.

The results for *Lenna* and *Barbara* are shown in figure 5. The results for *Mandrill* were similar. As can be seen in the figures, the results are quite surprising: MAP-analysis actually beats MAP-synthesis—in a convincing way—in every test. Compared to the baseline unitary transform (dotted line), where both methods coincide, MAP-analysis

(solid) shows a significant gain when introducing overcompleteness, which slightly improves as the redundancy increases; in contrast, MAP-synthesis (dashed) shows slightly *degraded* performance as the overcompleteness is increased. As a consequence, the distance between the two methods grows with the redundancy.

The experiments presented here were also carried out using the Contourlet transform [13], which has a 4:3 redundancy factor. In these experiments the two methods led to almost identical outputs, an outcome which conforms with the low redundancy of the transform. Interestingly, however, the picture remained the same: in all tests, MAP-analysis actually showed a small edge over MAP-synthesis.

The reasons for the superiority of MAP-analysis in the denoising scenario require further study; however, in our context we see that the gap indeed exists, and can become dramatic even in practical situations. One possible explanation for this could be the advantage of MAP-analysis discussed in section 1.3: since MAP-analysis utilizes all its filters simultaneously to support the recovery process, it may be more robust in the presence of noise compared to MAP-synthesis, whose compact representation may be unstable when noise is introduced, leading to recovery errors. A different possibility is that the high overcompleteness in MAP-synthesis, rather than positively enriching its descriptiveness, leads to a reverse effect where the dictionary becomes 'too descriptive', representing a wide range of undesirable signals. This effect does not apply to MAP-analysis where increasing the number of filters still requires the signal to agree with all existing ones.

5. Conclusions: analysis versus synthesis revisited

We began our discussion presenting two popular MAP-based methods for inverse problem regularization—the MAP-analysis and the MAP-synthesis approaches—and showing the algebraical similarity between the two. We saw that the two are equivalent in the square non-singular case as well as in the undercomplete denoising case; however, in the overcomplete case the two methods were shown to depart. We concentrated on the interesting ℓ^1 case, and found that the geometrical structures underlying the two exhibited very different properties. This perspective has led to a generality relation of MAP-synthesis over MAP-analysis, as well as to the characterization of the MAP-analysis parallels of the MAP-synthesis atoms.

The geometrical model does not provide a definite answer to the question of *which is better*. It does, however, shed some light on the real gap that exists between the two approaches, a gap which is not evident from the algebra alone. We have used the geometrical model to locate those signals where the gap is expected to be the largest, leading us to the results of the synthetic experiments; we saw that for these signals the gap indeed becomes large. The experiments also demonstrated the asymptotical nature of the difference between the two structures in their number of principal signals. Our real-world experiments showed that this gap exists not only in theory, and, no less important, that MAP-synthesis should not be *a priori* considered to be superior to MAP-analysis.

Our results are not to be interpreted as a recommendation for this method or another. The synthetic experiments indicate that each of the methods is successful, only on *different sets of signals*. The real-world experiments, which demonstrated a significant advantage to MAP-analysis, should be regarded as a sample case rather than a conclusion. MAP-synthesis remains advantageous in its simplicity of dictionary design, and we further emphasize that the interesting ℓ^0 MAP-synthesis case, though generally close to the ℓ^1 case, has not been treated. Nonetheless, as MAP-analysis is *significantly* simpler to solve, our results come to emphasize that despite the recent blossom of MAP-synthesis methods both approaches are still worthy

candidates for inverse problem regularization. The question of which will actually be better for a specific application and family of signals remains open.

Acknowledgments

The authors would like to thank Professor David L Donoho for the enlightening discussions and ideas which helped in developing the presented work. This research was supported by the USA–Israel binational science foundation (BSF) grant no 2004199.

Appendix A. Equivalence in the undercomplete case

Theorem 2 (undercomplete denoising case—near-equivalence). *MAP-analysis denoising with a full-rank analysing operator* $\mathbf{\Omega} \in M^{[L\times N]}$ $(L \leqslant N)$ *is nearly equivalent to MAP-synthesis with the dictionary* $\mathbf{D} = \mathbf{\Omega}^{+}$. *This is expressed by the relation* $\underline{\hat{X}}_{\text{MAP-A}} = \underline{\hat{X}}_{\text{MAP-S}} + \underline{Y}^{\mathbf{D}^{\perp}}$, *with* $\underline{Y}^{\mathbf{D}^{\perp}}$ *representing the component of the input orthogonal to the columns of* $\mathbf{D}$.

Proof. In the following, we assume the relation $\mathbf{D} = \mathbf{\Omega}^{+}$; we additionally assume that $\mathbf{\Omega}$ has full row-rank (equivalently, that $\mathbf{D}$ has full column-rank), and thus $\mathbf{D} = \mathbf{\Omega}^{T}(\mathbf{\Omega}\mathbf{\Omega}^{T})^{-1}$ and $\mathbf{\Omega}\mathbf{D} = I$. We introduce the notation $\underline{Z} = \underline{Z}^{\mathbf{D}} + \underline{Z}^{\mathbf{D}^{\perp}}$ to denote the (single) decomposition of a signal $\underline{Z}$ to the part $\underline{Z}^{\mathbf{D}}$ in the column-span of $\mathbf{D}$ and the part $\underline{Z}^{\mathbf{D}^{\perp}}$ in the orthogonal subspace.

We begin with the MAP-analysis formulation in (2):

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

Decomposing in respect to the column-span of $\mathbf{D}$, we obtain

$$\begin{aligned}\underline{\hat{X}}_{\text{MAP-A}} &= \underset{\underline{X}^{\mathbf{D}},\underline{X}^{\mathbf{D}^{\perp}}}{\text{Argmin}} \|\underline{Y}^{\mathbf{D}} + \underline{Y}^{\mathbf{D}^{\perp}} - \underline{X}^{\mathbf{D}} - \underline{X}^{\mathbf{D}^{\perp}}\|_2^2 + \lambda \cdot \|\mathbf{\Omega}(\underline{X}^{\mathbf{D}} + \underline{X}^{\mathbf{D}^{\perp}})\|_p^p \\ &= \underset{\underline{X}^{\mathbf{D}},\underline{X}^{\mathbf{D}^{\perp}}}{\text{Argmin}} \|\underline{Y}^{\mathbf{D}} - \underline{X}^{\mathbf{D}}\|_2^2 + \|\underline{Y}^{\mathbf{D}^{\perp}} - \underline{X}^{\mathbf{D}^{\perp}}\|_2^2 + \lambda \cdot \|\mathbf{\Omega}\underline{X}^{\mathbf{D}} + \mathbf{\Omega}\underline{X}^{\mathbf{D}^{\perp}}\|_p^p.\end{aligned}$$

We note that $\underline{Z}$ is orthogonal to the columns of $\mathbf{D}$ iff it is orthogonal to the rows of $\mathbf{\Omega}$: since $\mathbf{\Omega}\mathbf{\Omega}^{T}$ is invertible, we have $0 = \mathbf{D}^{T}\underline{Z} \Longleftrightarrow 0 = (\mathbf{\Omega}\mathbf{\Omega}^{T})\mathbf{D}^{T}\underline{Z} = (\mathbf{\Omega}\mathbf{\Omega}^{T})(\mathbf{\Omega}\mathbf{\Omega}^{T})^{-1}\mathbf{\Omega}\underline{Z} = \mathbf{\Omega}\underline{Z}$. This implies $\mathbf{\Omega}\underline{X}^{\mathbf{D}^{\perp}} = 0$, leading to

$$\underline{\hat{X}}_{\text{MAP-A}} = \underset{\underline{X}^{\mathbf{D}},\underline{X}^{\mathbf{D}^{\perp}}}{\text{Argmin}} \|\underline{Y}^{\mathbf{D}} - \underline{X}^{\mathbf{D}}\|_2^2 + \|\underline{Y}^{\mathbf{D}^{\perp}} - \underline{X}^{\mathbf{D}^{\perp}}\|_2^2 + \lambda \cdot \|\mathbf{\Omega}\underline{X}^{\mathbf{D}}\|_p^p.$$

Obviously any solution to this problem will satisfy $\underline{\hat{X}}^{\mathbf{D}^{\perp}} = \underline{Y}^{\mathbf{D}^{\perp}}$, as there is no additional penalty term for $\underline{X}^{\mathbf{D}^{\perp}}$. Therefore the MAP-analysis problem reduces to an optimization problem for $\underline{\hat{X}}^{\mathbf{D}}_{\text{MAP-A}}$:

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

Signals $\underline{X}^{\mathbf{D}}$ spanned by the columns of $\mathbf{D}$ have a representation as $\underline{X}^{\mathbf{D}} = \mathbf{D}\underline{\gamma}$. We can thus reformulate the above as an optimization on $\underline{\gamma}$, leading to

$$\begin{aligned}\underline{\hat{X}}^{\mathbf{D}}_{\text{MAP-A}} &= \mathbf{D} \cdot \underset{\underline{\gamma}}{\text{Argmin}} \|\underline{Y}^{\mathbf{D}} - \mathbf{D}\underline{\gamma}\|_2^2 + \lambda \cdot \|\mathbf{\Omega}\mathbf{D}\underline{\gamma}\|_p^p \\ &= \mathbf{D} \cdot \underset{\underline{\gamma}}{\text{Argmin}} \|\underline{Y}^{\mathbf{D}} - \mathbf{D}\underline{\gamma}\|_2^2 + \lambda \cdot \|\underline{\gamma}\|_p^p.\end{aligned}$$

We see that the solution to $\underline{\hat{X}}^{\mathbf{D}}_{\text{MAP-A}}$ comes from a MAP-synthesis structure with $\mathbf{D} = \mathbf{\Omega}^{+}$, and applied to $\underline{Y}^{\mathbf{D}}$. We conclude by showing that $\underline{Y}^{\mathbf{D}}$ in this formulation may be replaced with $\underline{Y}$. We do this using similar arguments to those applied above, in a reverse manner:

$$\begin{aligned}
\underline{\hat{X}}^{\mathbf{D}}_{\text{MAP-A}} &= \mathbf{D} \cdot \underset{\underline{\gamma}}{\text{Argmin}} \, \|\underline{Y}^{\mathbf{D}} - \mathbf{D}\underline{\gamma}\|_2^2 + \lambda \cdot \|\underline{\gamma}\|_p^p \\
&= \mathbf{D} \cdot \underset{\underline{\gamma}}{\text{Argmin}} \, \|\underline{Y}^{\mathbf{D}} - \mathbf{D}\underline{\gamma}\|_2^2 + \|\underline{Y}^{\mathbf{D}^{\perp}}\|_2^2 + \lambda \cdot \|\underline{\gamma}\|_p^p \\
&= \mathbf{D} \cdot \underset{\underline{\gamma}}{\text{Argmin}} \, \|\underline{Y}^{\mathbf{D}} + \underline{Y}^{\mathbf{D}^{\perp}} - \mathbf{D}\underline{\gamma}\|_2^2 + \lambda \cdot \|\underline{\gamma}\|_p^p \\
&= \mathbf{D} \cdot \underset{\underline{\gamma}}{\text{Argmin}} \, \|\underline{Y} - \mathbf{D}\underline{\gamma}\|_2^2 + \lambda \cdot \|\underline{\gamma}\|_p^p.
\end{aligned}$$

Summing up, for the (under-)determined case, and with the relation $\mathbf{D} = \mathbf{\Omega}^{+}$, we have shown that given a signal $\underline{Y} = \underline{Y}^{\mathbf{D}} + \underline{Y}^{\mathbf{D}^{\perp}}$, the MAP-analysis solution and the MAP-synthesis solution are related by $\underline{\hat{X}}_{\text{MAP-A}} = \underline{\hat{X}}_{\text{MAP-S}} + \underline{Y}^{\mathbf{D}^{\perp}}$, as claimed. □

Appendix B. MAP-analysis defining polytope

Lemma 1 (facets of the MAP-analysis defining polytope). *Let $\underline{X} \in \partial\Psi_{\mathbf{\Omega}}$, where $\Psi_{\mathbf{\Omega}}$ is the MAP-analysis defining polytope $\{\underline{X} | \|\mathbf{\Omega}\underline{X}\|_1 \leqslant 1\}$. If $\mathbf{\Omega}\underline{X}$ has no vanishing elements, then $\underline{X}$ resides strictly within a facet ($(N-1)$-dimensional face) of the MAP-analysis defining polytope.*

Proof. Let $f_{\text{A}}(\underline{X}) = \|\mathbf{\Omega}\underline{X}\|_1$ (the MAP-analysis target function), and assume $\mathbf{\Omega}\underline{X}$ has no vanishing elements; then $\nabla f_{\text{A}}(\underline{X}) = \mathbf{\Omega}^T \text{sign}(\mathbf{\Omega}\underline{X})$, and is defined at $\underline{X}$. Also, since all elements of $\mathbf{\Omega}\underline{X}$ are finite and non-zero, there exists a ball $\mathcal{B}_\epsilon(\underline{X})$ around $\underline{X}$ such that, for all $\underline{x} \in \mathcal{B}_\epsilon(\underline{X})$, $\mathbf{\Omega}\underline{x}$ has no vanishing elements. Now consider the intersection $\partial\Psi_{\mathbf{\Omega}} \cap \mathcal{B}_\epsilon(\underline{X})$: this is a neighbourhood of $\underline{X}$ on the boundary of the defining polytope, and for all $\underline{x}$ in it, $\mathbf{\Omega}\underline{x}$ has no zero coordinates. From continuity of $\mathbf{\Omega}\underline{x}$, we conclude that none of its coordinates change sign within this neighbourhood, so for all $\underline{x}$ in it, $\text{sign}(\mathbf{\Omega}\underline{x}) = \text{sign}(\mathbf{\Omega}\underline{X})$ and also $\nabla f_{\text{A}}(\underline{x}) = \nabla f_{\text{A}}(\underline{X})$. As the defining polytope is a level-set of f_{A}, ∇f_{A} (where defined) designates the direction of the normal to this polytope. We therefore have a finite neighbourhood of $\underline{X}$ on the boundary of the polytope where the normal is fixed, and thus $\underline{X}$ must reside strictly within a facet of this polytope. □

We now bring the proof of claim 1, generalizing the above lemma.

Claim 1 (faces of the MAP-analysis defining polytope). *Let $\underline{X} \in \partial\Psi_{\mathbf{\Omega}}$, and let k denote the rank of the rows in $\mathbf{\Omega}$ to which $\underline{X}$ is orthogonal to. Then $\underline{X}$ resides strictly within a face of dimension $(N-k-1)$ of the MAP-analysis defining polytope.*

Proof. Assume a signal $\underline{X} \in \partial\Psi_{\mathbf{\Omega}}$. Let $\{\underline{w}_1, \ldots, \underline{w}_k\}$ orthonormally span the rows in $\mathbf{\Omega}$ to which $\underline{X}$ is orthogonal to, and let $\{\underline{u}_1, \ldots, \underline{u}_{N-k}\}$ span their complementary space. We denote $\mathcal{U} = \text{Span}\{\underline{u}_i\}$ and $\mathcal{W} = \text{Span}\{\underline{w}_j\}$. Clearly $\underline{X} \in \mathcal{U}$, from orthogonality to $\{\underline{w}_j\}$.

First, we consider the space $\mathcal{U}$. Any vector $\underline{v} \in \mathcal{U}$ may be written as $\underline{v} = \mathbf{U}\underline{\alpha}(\underline{v})$, where $\mathbf{U} = [\underline{u}_1 | \ldots | \underline{u}_{N-k}]$ is an $N \times (N-k)$ matrix, and $\underline{\alpha}(\underline{v}) = \mathbf{U}^T\underline{v}$. Since $\underline{v} \in \mathcal{U}$, it is orthogonal to all the rows in $\mathbf{\Omega}$ to which $\underline{X}$ is orthogonal to; therefore, letting $\widehat{\mathbf{\Omega}}$ be the matrix obtained by discarding these rows from $\mathbf{\Omega}$, then for any $\underline{v} \in \mathcal{U}$, we have $\|\mathbf{\Omega}\underline{v}\|_1 = \|\widehat{\mathbf{\Omega}}\underline{v}\|_1$. Note that since we assume $\mathbf{\Omega}$ is full rank, then after removing from it the rows whose span is $\mathcal{W}$, the remaining rows of $\widehat{\mathbf{\Omega}}$ still span *at least* the complement space $\mathcal{U}$.

Now, define $\omega = \widehat{\Omega}\mathbf{U}$; we have $\|\Omega\underline{v}\|_1 = \|\widehat{\Omega}\underline{v}\|_1 = \|\widehat{\Omega}\mathbf{U}\underline{\alpha}(\underline{v})\|_1 = \|\omega\underline{\alpha}(\underline{v})\|_1$ for any $\underline{v} \in \mathcal{U}$. Multiplying $\widehat{\Omega}$ to the left of $\mathbf{U}$ is essentially an orthogonal projection of its rows on the subspace $\mathcal{U}$; since the rows of $\widehat{\Omega}$ span $\mathcal{U}$, the rank of the result must be equal to that of $\mathbf{U}$. Therefore the rank of ω is $(N-k)$, so it must have at least this number of rows, and is thus an overcomplete analysis operator on the $\underline{\alpha}$-space.

Since $\underline{X} \in \mathcal{U}$, all the equalities above hold for $\underline{X}$. Specifically, $\|\omega\underline{\alpha}(\underline{X})\|_1 = \|\Omega\underline{X}\|_1 = 1$, so by definition $\underline{\alpha}(\underline{X}) \in \partial\Psi_\omega$. In other words, $\underline{\alpha}(\underline{X})$ must reside on the boundary of the defining polytope corresponding to the $(N-k)$-dimensional MAP-analysis problem for the $\underline{\alpha}$-space with operator ω. We further know that $\widehat{\Omega}\underline{X}$ has no vanishing elements, since all such elements have been removed, so $\omega\underline{\alpha}(\underline{X}) = \widehat{\Omega}\underline{X}$ has no vanishing elements. We have thus established all the conditions of lemma 1 for $\underline{\alpha}(\underline{X})$, and it follows that $\underline{\alpha}(\underline{X})$ resides strictly within a facet of the $(N-k)$-dimensional polytope Ψ_ω.

Given this, we know there exists an $(N-k-1)$-dimensional ball about $\underline{\alpha}(\underline{X})$ such that this ball is entirely contained within the boundary of Ψ_ω. By applying $\mathbf{U}$ to the points of this ball, we orthonormally inject it to the N-dimensional signal space, obtaining an $(N-k-1)$-dimensional ball about $\underline{X} = \mathbf{U}\underline{\alpha}(\underline{X})$. This ball resides entirely on the boundary of Ψ_Ω, since for any signal $\underline{x} = \mathbf{U}\underline{\alpha}(\underline{x})$ in this ball, $\underline{x} \in \mathcal{U}$ and so $\|\Omega\underline{x}\|_1 = \|\omega\underline{\alpha}(\underline{x})\|_1 = 1$. Evidently, we have an $(N-k-1)$-dimensional ball about $\underline{X}$, residing entirely on the boundary of the defining polytope; therefore $\underline{X}$ must reside on a face of dimension *at least* $(N-k-1)$ of this polytope. To conclude the proof, we show this residence is strict; in other words, we prove that there does not exist a ball of higher dimension about $\underline{X}$ residing entirely within the polytope's boundary.

Consider a d-dimensional ball about $\underline{X}$, contained entirely within the boundary of the defining polytope; then for any point $\underline{X} + \underline{\epsilon}$ in this ball, the point $\underline{X} - \underline{\epsilon}$ is also in the ball. Now, write $\underline{\epsilon}$ as

$$\underline{\epsilon} = \sum_i a_i \underline{u}_i + \sum_j b_j \underline{w}_j,$$

where $\{\underline{u}_i\}$ and $\{\underline{w}_j\}$ are the orthonormal bases as defined above. Since both points are on the boundary of Ψ_Ω, we have $\|\Omega\underline{X}\|_1 = \|\Omega(\underline{X}+\underline{\epsilon})\|_1 = \|\Omega(\underline{X}-\underline{\epsilon})\|_1 = 1$. Written explicitly, these expand to

$$\Omega(\underline{X} \pm \underline{\epsilon}) = \Omega\left[\underline{X} \pm \left(\sum a_i \underline{u}_i + \sum b_j \underline{w}_j\right)\right].$$

Since $\left(\underline{X} \pm \sum a_i \underline{u}_i\right) \in \mathcal{U}$, all vanishing coefficients in $\Omega\underline{X}$ also vanish in $\Omega\left(\underline{X} \pm \sum a_i \underline{u}_i\right)$. As to the second part, assume by contradiction that $\sum b_j \underline{w}_j \in \mathcal{W}$ is non-zero. Clearly the same coefficients cannot all vanish in $\Omega \sum b_j \underline{w}_j$, as the corresponding rows in Ω span $\mathcal{W}$. Therefore adding or subtracting $\Omega\underline{\epsilon}$ to $\Omega\underline{X}$ necessarily *increases* the absolute-value-sum of these coefficients. On the other hand, the entire ℓ^1 norm of $\Omega(\underline{X} \pm \underline{\epsilon})$ remains fixed; so, for the remainder of the coefficients, the addition or subtraction of $\Omega\underline{\epsilon}$ must strictly *reduce* their absolute-value-sum. However, this may not occur simultaneously for both addition *and* subtraction. Therefore, the only resolution to this is to require $b_j \equiv 0$ for all j, implying that necessarily $\underline{\epsilon} \in \mathcal{U}$. Thus, we have limited the dimension of the ball about $\underline{X}$ to $N-k$ (the dimension of $\mathcal{U}$). Finally, $\underline{X} \in \mathcal{U}$, but clearly $\|\Omega(\underline{X} + \delta\underline{X})\|_1 \neq \|\Omega\underline{X}\|_1$ for any $\delta \neq 0$. So $\underline{\epsilon}$ cannot be proportional to $\underline{X}$, and hence the ball about $\underline{X}$ must be of dimension less than $\mathcal{U}$. We conclude that $d \leqslant (N-k-1)$, so $\underline{X}$ can reside strictly within a face of dimension no more than $(N-k-1)$. Since we have already shown the existence of such a face, we conclude that $\underline{X}$ resides strictly within an $(N-k-1)$-dimensional face of the MAP-analysis defining polytope, as claimed. □

Appendix C. MAP-synthesis defining polytope

Claim 2 (geometry of the MAP-synthesis defining polytope). *The MAP-synthesis defining polytope* $\Phi_{\mathbf{D}} = \mathbf{D}\{\|\underline{\gamma}\|_1 \leqslant 1\}$ *is obtained as the convex hull of* $\{\pm\underline{d}_i\}_{i=1,\ldots,L}$, *where* $\{\underline{d}_i\}$ *are the columns of* $\mathbf{D}$.

Proof. For the proof we note that $\underline{d}_i = \mathbf{D}\underline{e}_i$, where $\{\underline{e}_i\}$ is the standard basis of $\mathbb{R}^L$. We introduce the notation $\mathcal{CH}\{\underline{v}_i\}$ to denote the convex hull of the set $\{\underline{v}_i\}$.

$\mathcal{CH}\{\underline{\mathbf{d}}_\mathbf{i}\} \subseteq \Phi_{\mathbf{D}}$: We have $\pm\underline{e}_i \in \{\|\underline{\gamma}\|_1 \leqslant 1\}$ for all i, and therefore $\pm\underline{d}_i = \mathbf{D}(\pm\underline{e}_i) \in \mathbf{D}\{\|\underline{\gamma}\|_1 \leqslant 1\} = \Phi_{\mathbf{D}}$. Since $\Phi_{\mathbf{D}}$ is convex, it must also contain the convex hull of $\{\pm\underline{d}_i\}$.

$\Phi_{\mathbf{D}} \subseteq \mathcal{CH}\{\underline{\mathbf{d}}_\mathbf{i}\}$: Let $\underline{X} \in \Phi_{\mathbf{D}}$, then there exists a representation $\underline{\gamma}$ such that $\underline{X} = \mathbf{D}\underline{\gamma}$, where $\|\underline{\gamma}\|_1 \leqslant 1$. Since $\underline{\gamma} \in \{\|\underline{\gamma}\|_1 \leqslant 1\}$, it is a convex combination of $\{\pm\underline{e}_i\}$, and can be written as $\underline{\gamma} = \sum_i \{a_i\underline{e}_i + b_i(-\underline{e}_i)\}$. This implies $\underline{X} = \mathbf{D}\underline{\gamma} = \sum_i \{a_i\underline{d}_i + b_i(-\underline{d}_i)\}$, so $\underline{X}$ is a convex combination of $\{\pm\underline{d}_i\}$, and as such exists in their convex hull. □

Appendix D. MAP-synthesis with a normalized dictionary

Lemma 2. *Let* $\mathcal{P}$ *be a polytope with fixed-length vertices, i.e., for all vertices* $\underline{v}$ *of* $\mathcal{P}$, $\|\underline{v}\|_2 = c$ *for some constant* c. *Then for every non-vertex point* $\underline{p}$ *on the boundary of the polytope,* $\|\underline{p}\|_2 < c$.

Proof. Consider a facet φ of $\mathcal{P}$, defined by the vertices $\{\underline{v}_1, \ldots, \underline{v}_n\}$. This facet constitutes the intersection of some $(n-1)$-dimensional hyperplane with the polytope. Now, consider the ℓ^2-norm function $f(\underline{X}) = \|\underline{X}\|_2$, constrained to this plane. The iso-surfaces of f on this plane are a set of concentric ellipsoids about some central point of minimal ℓ^2-norm. Since $\{\underline{v}_1, \ldots, \underline{v}_n\}$ are of a fixed length, they all reside on the same ellipsoid. The facet φ, which is the convex hull of $\{\underline{v}_1, \ldots, \underline{v}_n\}$, must thus exist entirely within this ellipsoid by definition of the convex hull as the minimal convex set containing $\{\underline{v}_1, \ldots, \underline{v}_n\}$. This implies that for every $\underline{p} \in \varphi$, $\|\underline{p}\|_2 \leqslant c$.

To obtain sharp inequality, we assume by contradiction that $\|\underline{p}\|_2 = c$ while $\underline{p}$ is not a vertex. Since $\underline{p}$ is not a vertex, there exist two points $\underline{p}_1, \underline{p}_2 \in \varphi$ such that $\underline{p}$ resides on the line connecting $\underline{p}_1$ and $\underline{p}_2$. However, examining the function f, we have the following observation: for any point in space, advancing from it in two opposite directions will always lead to *at least* one direction of increase in f; this is due to the fact that when constrained to an infinite line, f always achieves a single minimum and no maximum on the line. This implies that at least one of $\underline{p}_1$ and $\underline{p}_2$ will have ℓ^2-norm larger than c, leading to a contradiction. Hence necessarily $\|\underline{p}\|_2 < c$. □

Claim 3 (principal signals of MAP-synthesis with a normalized dictionary). *Let* $\mathbf{D}$ *be a MAP-synthesis dictionary with fixed-energy columns. Then the dictionary atoms coincide with the principal signals of the MAP-synthesis prior.*

Proof. From lemma 2, the proof is trivial. Let us denote the length of the dictionary atoms by c. Then for any atom $\underline{d}$, it follows that it must be a vertex as $\|\underline{d}\|_2 = c$. Now, assume by contradiction that $\underline{d}$ is non-principal; therefore there exists a direction from $\underline{d}$ on the boundary of the defining polytope such that the distance from the origin increases. However, this means that if we advance from $\underline{d}$ in this direction a short enough distance, we will obtain a non-vertex point on the polytope boundary whose length is larger than c, contradicting the previous lemma. We conclude that $\underline{d}$ must be a principal signal. □

References

[1] Elad M 2002 On the origin of the bilateral filter and ways to improve it *IEEE Trans. Image Process.* **11** 1141–51

[2] Rudin L, Osher S and Fatemi E 1992 Nonlinear total variation based noise removal algorithms *Physica* D **60** 259–68

[3] Blomgren P and Chan T F 1998 Color TV: total variation methods for restoration of vector-valued images *IEEE Trans. Image Process.* **7** 304–9

[4] Simoncelli E P 1999 Bayesian denoising of visual images in the wavelet domain *Bayesian Inference in Wavelet Based Models* (*Lecture Notes in Statistics* vol 141) ed P Muller and B Vidakovic (New York: Springer) pp 291–308

[5] Schultz R R and Stevenson R L 1994 A Bayesian approach to image expansion for improved definition *IEEE Trans. Image Process.* **3** 233–42

[6] Farsiu S, Elad M and Milanfar P 2006 Multi-frame demosaicing and super-resolution of color images *IEEE Trans. Image Process.* **15** 141–59

[7] Bouman C and Sauer K 1993 A generalized Gaussian image model for edge-preserving MAP estimation *IEEE Trans. Image Process.* **2** 296–310

[8] Chambolle A, DeVore R E, Lee N and Lucier B J 1998 Nonlinear wavelet image processing: variational problems, compression, and noise removal through wavelet shrinkage *IEEE Trans. Image Process.* **7** 320–35

[9] Higdon D M, Bowsher J E, Johnson V E, Turkington T G, Gilland D R and Jaszczak R J 1997 Fully Bayesian estimation of Gibbs hyperparameters for emission computed tomography data *IEEE Trans. Med. Imaging* **16** 516–26

[10] Chen S S, Donoho D L and Saunders M A 2001 Atomic decomposition by basis pursuit *SIAM Rev.* **43** 129–59

[11] Candes E J and Donoho D L 1999 Ridgelets: A key to higher-dimensional intermittency? *Phil. Trans. R. Soc.* A **357** 2495–509

[12] Candes E J and Donoho D L 1999 Curvelets—a surprisingly effective nonadaptive representation for objects with edges *Curves and Surfaces* ed L L Schumaker *et al* (Nashville, TN: Vanderbilt University Press)

[13] Do M N and Vetterli M 2005 The contourlet transform: an efficient directional multiresolution image representation *IEEE Trans. Image Process.* **14** 2091–106

[14] Aharon M, Elad M and Bruckstein A M 2006 The K-SVD: an algorithm for designing of overcomplete dictionaries for sparse representation *IEEE Trans. Signal Process.* **54** 4311–22

[15] Donoho D L and Elad M 2003 Optimally sparse representation in general (non-orthogonal) dictionaries via ℓ^1 minimization *Proc. Natl Acad. Sci. USA* **100** 2197–202

[16] Tropp J A 2004 Greed is good: algorithmic results for sparse approximation *IEEE Trans. Inform. Theory* **50** 2231–42

[17] Olshausen B A and Field D J 1995 Emergence of simple-cell receptive field properties by learning a sparse code for natural images *Nature* **381** 607–9

[18] Donoho D L, Elad M and Temlyakov V 2006 Stable recovery of sparse overcomplete representations in the presence of noise *IEEE Trans. Inform. Theory* **52** 6–18

[19] Elad M, Starck J L, Querre P and Donoho D L 2005 Simultaneous cartoon and texture image inpainting using morphological component analysis *Appl. Comput. Harmon. Anal.* **19** 340–58

[20] Tibshirani R 1996 Regression shrinkage and selection via the Lasso *J. R. Stat. Soc.* B **58** 267–88

[21] Donoho D L 2005 High-dimensional centrally-symmetric polytopes with neighborliness proportional to dimension *Technical Report—Statistics, Stanford*

[22] Elad M and Bruckstein A M 2002 A generalized uncertainty principle and sparse representation in pairs of $\Re^N$ bases *IEEE Trans. Inform. Theory* **48** 2558–67

[23] Strohmer T and Heath R 2003 Grassmannian frames with applications to coding and communication *Preprint* math.FA/0301135