Geometry and homotopy for ℓ_1 sparse representations

Mark D. Plumbley

***Abstract*—We explore the geometry of ℓ_1 sparse representations in both the noiseless (Basis Pursuit) and noisy (Basis Pursuit De-Noising) case using a homotopy method. We will see that the concept of the basis vertex c, which has unit inner product with active basis vectors, is a useful geometric concept, both for visualization and for algorithm construction. We derive an explicit homotopy continuation algorithm and find that this method has interesting parallels with the Polytope Faces Pursuit algorithm for the noiseless case. Numerical results confirm the operation of the algorithm.**

I. Introduction

Consider the problem of representing a vector $\mathbf{y} = [y_1, \ldots, y_d]^T$ as a linear combination of vectors $\mathbf{a}_i$, i.e. $\mathbf{y} = \mathbf{A}\mathbf{x}$ where $\mathbf{A} = [\mathbf{a}_i]$ is the $d \times n$ basis matrix. If $n > d$ there are many possible solutions to this problem, so we might like to find the *sparsest* solution

$$\min_{\mathbf{x}} \|\mathbf{x}\|_0 \quad \text{such that} \quad \mathbf{A}\mathbf{x} = \mathbf{y} \tag{1}$$

where $\|\mathbf{x}\|_0$ is the ℓ_0 norm of $\mathbf{x}$, i.e. the number of nonzero elements of $\mathbf{x}$, or alternatively the solution of the easier *Basis Pursuit* linear program (LP)

$$\min_{\mathbf{x}} \|\mathbf{x}\|_1 \quad \text{such that} \quad \mathbf{A}\mathbf{x} = \mathbf{y} \tag{2}$$

where $\|\mathbf{x}\|_1$ is now the ℓ_1 norm of $\mathbf{x}$.

In many real world cases the observation $\mathbf{y}$ may be noisy and in this case we wish to find an approximate representation $\hat{\mathbf{y}} = \mathbf{A}\mathbf{x} \approx \mathbf{y}$ as a solution to one of the following equivalent least squares problems [1]

$$\min_{\mathbf{x}} \frac{1}{2}\|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 \quad \text{such that} \quad \|\mathbf{x}\|_1 \le t, \quad t > 0 \tag{3}$$

$$\min_{\mathbf{x}} \frac{1}{2}\|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 - \lambda\|\mathbf{x}\|_1, \quad \lambda > 0. \tag{4}$$

These methods have been discussed by different authors in various fields. For example, equation (3) was introduced by Tibshirani [2] as the *Lasso* (least absolute shrinkage and selection operator), while (4) was introduced by Chen, Donoho and Saunders [3] as basis pursuit de-noising (BPDN), with the resulting estimate $\hat{\mathbf{y}} = \mathbf{A}\mathbf{x}$ considered to be a de-noised version of $\mathbf{y}$. In this context (2) emerges from (4) as $\lambda \to 0$, or alternatively from (3) for t which achieves the minimum of (2) itself. For further background and discussions see e.g. [1], [4], [5], [6], [7].

Osborne et al [8] investigated the trajectory of the solution to (3) as t changes, and found that the optimal solution $\mathbf{x}(t)$ follows a piecewise linear path. Based on this observation, they proposed a *homotopy continuation method* to solve (3) based on following this path starting from $\mathbf{x}(t) = \mathbf{0}$ at $t = 0$ and adding (or removing) active variables as required, as t increases. Malioutov et al [9] have recently demonstrated this homotopy continuation approach for sparse representation of signals.

In a recent and closely related approach, Efron et al [10] have introduced *Least Angle Regression* (LARS) as an interesting geometrical approach to solve (3). LARS is based on the concept of following the path $\mathbf{x}(t)$ which is equiangular between all current *predictors*, i.e. the basis vectors $\mathbf{a}_i$ taking part in the current reconstruction $\hat{\mathbf{y}}$ of $\mathbf{y}$. A new basis vector $\mathbf{a}_j$ is added as soon as the angle with $\mathbf{a}_j$ equals the common angle with the currently selected basis vectors. The highest correlation (least angle) basis vectors are selected first, hence *least angle* regression. With a suitable modification to remove basis functions that cannot be part of a solution of (3), the LARS will follow the homotopy continuation path and produce all Lasso solutions [10].

In this paper we will further explore the geometry of the homotopy continuation method, using elements of the geometry of polar polytopes that we introduced previously for the noiseless Basis Pursuit case (2).

II. Noiseless case

Let us write (2) in its standard form [3]

$$\min_{\tilde{\mathbf{x}}} \mathbf{1}^T\tilde{\mathbf{x}} \quad \text{such that} \quad \mathbf{y} = \tilde{\mathbf{A}}\tilde{\mathbf{x}},\ \tilde{\mathbf{x}} \ge \mathbf{0} \tag{5}$$

where $\tilde{\mathbf{A}} = [\mathbf{A}, -\mathbf{A}]$ and $\tilde{\mathbf{x}}$ has $2n$ nonnegative components $\tilde{x}_i = \max(x_i, 0)$ for $1 \le i \le n$ and $\tilde{x}_i = \max(-x_{i-n}, 0)$ for $n+1 \le i \le 2n$, then (5) has the dual linear program [3], [5]

$$\max_{\mathbf{c}} \mathbf{y}^T\mathbf{c} \quad \text{such that} \quad \tilde{\mathbf{A}}^T\mathbf{c} \le \mathbf{1}. \tag{6}$$

An optimum $\tilde{\mathbf{x}}^*$ of (5) has a corresponding optimum $\mathbf{c}^*$ to (6).

This has a simple geometrical interpretation in terms of the set $P^* = \{\mathbf{c} \mid \tilde{\mathbf{a}}_i^T\mathbf{c} \le 1\}$ of feasible solutions to the vector inequality $\tilde{\mathbf{A}}^T\mathbf{c} \le \mathbf{1}$. The set P^* is a *polytope*, a d-dimensional generalization of a bounded polygon/polyhedron [11], [12]. (See also the work of Donoho [13] for sparse representation results arising from the geometry of polytopes). The optimum solution to (6) is achieved when $\mathbf{c}^*$ is a *vertex* of P^* (Fig. 1). In the figure, the scaled vectors $\mathbf{a}_i^\dagger = \mathbf{a}_i/|\mathbf{a}_i|^2$ satisfy $\mathbf{a}_i^T\mathbf{a}_i^\dagger = 1$, so they touch the faces of P^* defined by $\mathbf{a}_i^T\mathbf{c} = 1$ (or the extension of the face if $\mathbf{a}_i^\dagger$ lies outside P^*).

Now any vertex $\mathbf{c}_\mathcal{I}$ of this d-dimensional polytope P^* corresponds to a particular $d \times d$ basis matrix $\tilde{\mathbf{A}}_\mathcal{I} = [\tilde{\mathbf{a}}_i]_{i \in \mathcal{I}}$ formed from a subset of the signed basis vectors $\tilde{\mathbf{a}}_i \in \tilde{\mathbf{A}}$. For example, the optimum point $\mathbf{c}^*$ in Fig. 1 is the vertex defined by the faces $+\mathbf{a}_1^T\mathbf{c} = 1$ and $+\mathbf{a}_2^T\mathbf{c} = 1$ and so given the basis set $\tilde{\mathbf{A}}_{++0} = [+\mathbf{a}_1, +\mathbf{a}_2]$ we have $\tilde{\mathbf{A}}_{++0}^T\mathbf{c} = \mathbf{1}$ so $\mathbf{c}_{++-} = \tilde{\mathbf{A}}_{++0}^{-1}\mathbf{1}$. More generally, if the optimum were achieved for a face of P^* with $m < d$ equalities $\tilde{\mathbf{A}}_\mathcal{I}^T\mathbf{c} = \mathbf{1}$ where $\tilde{\mathbf{A}}_\mathcal{I}$ is now

Department of Electronic Engineering, Queen Mary University of London, Mile End Road, London E1 4NS, UK. Email: mark.plumbley@elec.qmul.ac.uk

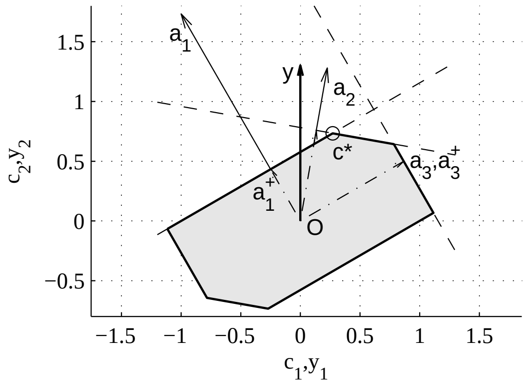


Fig. 1. Polytope P^* in 2-D with optimum basis vertex c^*.

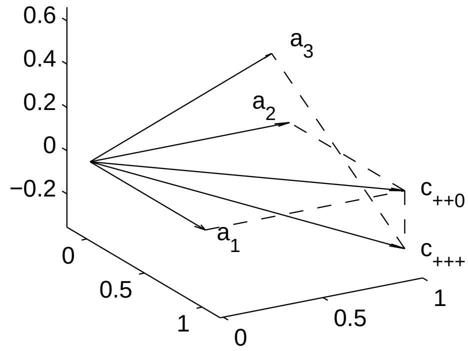
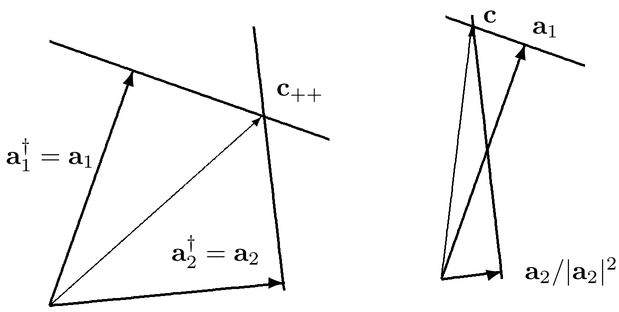


Fig. 3. Example of an external vertex (c_{+++}) in 3-D.



(a) (b)

Fig. 2. Vertex $\mathbf{c}_{++}$ of unit basis set $[+\mathbf{a}_1, +\mathbf{a}_2]$, for (a) internal and (b) external vertex

a $d \times m$ basis matrix, we have $\mathbf{c}_\mathcal{I} = \tilde{\mathbf{A}}_\mathcal{I}^{\dagger T}\mathbf{1}$ where $\tilde{\mathbf{A}}^\dagger$ is the Moore-Penrose pseudoinverse of $\mathbf{A}$, so that $\tilde{\mathbf{A}}_\mathcal{I}^T\mathbf{c}_\mathcal{I} = \mathbf{1}$ as required. Once we have $\mathbf{c}^*$, the optimum solution to (5) can be found from $\tilde{\mathbf{x}}^* = \tilde{\mathbf{A}}_{++0}^\dagger\mathbf{y}$.

In many cases the basis matrix $\mathbf{A}$ consists of unit norm atoms $\mathbf{a}_i$ with $|\mathbf{a}_i| = 1$, in which case the scaled basis vectors $\mathbf{a}_i^\dagger = \mathbf{a}_i/|\mathbf{a}_i|^2$ are equal to the basis vectors themselves, $\mathbf{a}_i^\dagger = \mathbf{a}_i$ (Fig. 2(a)). Unit norm basis sets in two dimensions always have an *internal* vertex $\mathbf{c}$ in the conic hull of the corresponding set of basis vectors, i.e. such that $\mathbf{c}$ is a nonnegative weighted sum $\mathbf{c}_\mathcal{I} = \sum_{i\in\mathcal{I}} z_i\tilde{\mathbf{a}}_i = \tilde{\mathbf{A}}_\mathcal{I}\mathbf{z}$ of the basis vectors. However the weighing vector $\mathbf{z}$ is not guaranteed to be nonnegative in general for non-unit-norm basis vector sets or for $d > 2$ dimensions. For example, $\mathbf{c}^*$ in Fig. 1 is an *external* vertex, outside the conic hull of $[+\mathbf{a}_1, +\mathbf{a}_2]$ and has $z_1 < 0$ (see also Fig. 2(b)). As an example with unit norm basis vectors, consider $\tilde{\mathbf{A}}_{+++} = [+\mathbf{a}_1, +\mathbf{a}_2, +\mathbf{a}_3]$ with $\mathbf{a}_1 = [1, 0, 0]^T$, $\mathbf{a}_2 = [0, 1, 0]^T$, $\mathbf{a}_3 = (1/\sqrt{3})[1, 1, 1]^T$ (Fig. 3). We can see that $\mathbf{c}_{+++}$ is an external vertex, while $\mathbf{c}_{++0} = \tilde{\mathbf{A}}_{++0}^\dagger$ is an internal vertex for the basis set $\tilde{\mathbf{A}}_{++0} = [+\mathbf{a}_1, +\mathbf{a}_2]$.

III. An algorithm for the noiseless case

The linear program (5) can of course be solved using many linear programming tools such as the interior point method [3], [14]. However, in [12] we recently introduced a greedy algorithm, *Polytope Faces Pursuit*, which follows the path in P^* in the direction of $\mathbf{y}$, projecting this search direction onto faces of P^* as they are encountered. The derived algorithm is similar in structure to orthogonal matching pursuits (OMP) [15], but instead of the usual maximum correlation criterion $\tilde{\mathbf{a}}^k = \arg\max_{\tilde{\mathbf{a}}_i} \tilde{\mathbf{a}}_i^T\mathbf{r}^{k-1}$ where $\mathbf{r}^k = \mathbf{y} - \tilde{\mathbf{A}}^k\tilde{\mathbf{x}}^k$ is the residual after step k, we have a scaled correlation criterion

$$\tilde{\mathbf{a}}^k = \arg\max_{\tilde{\mathbf{a}}_i} \frac{\tilde{\mathbf{a}}_i^T\mathbf{r}^{k-1}}{1 - \tilde{\mathbf{a}}_i^T\mathbf{c}^{k-1}} \tag{7}$$

which takes into account the next face encountered when searching along the current polytope face corresponding to $\mathbf{c}^{k-1}$. The algorithm also includes a constraint release criterion to allow us to remove basis vectors and to move away from a restraining face if necessary (Algorithm 1).

Algorithm 1 Polytope Faces Pursuit [12]

Input: $\mathbf{y}$, $\tilde{\mathbf{A}} \triangleq [\tilde{\mathbf{a}}_i] = [\mathbf{A}, -\mathbf{A}]$
Set stopping conditions $k_{\max} > 0$ and $\theta_{\min} \geq 0$
Initialize: $k \leftarrow 0$, $\mathcal{I}^k \leftarrow \emptyset$, $\tilde{\mathbf{A}}^k \leftarrow \emptyset$,
$\mathbf{c}^k \leftarrow \mathbf{0}$, $\tilde{\mathbf{x}}^k \leftarrow \emptyset$, $\hat{\mathbf{y}}^k \leftarrow \mathbf{0}$, $\mathbf{r}^k \leftarrow \mathbf{y}$
while $k < k_{\max}$ and $\max_i \tilde{\mathbf{a}}_i^T\mathbf{r}^{k-1} > \theta_{\min}$ **do**
  $k \leftarrow k + 1$
  {Find next face}
  $\lambda_i \leftarrow (\tilde{\mathbf{a}}_i^T\bar{\mathbf{r}}^{k-1})/(1 - \tilde{\mathbf{a}}_i^T\mathbf{c}^{k-1})$, $1 \leq i \leq 2n$,
  $i^k \leftarrow \arg\max_{i\notin\mathcal{I}^{k-1}}\{\lambda_i \mid \tilde{\mathbf{a}}_i^T\mathbf{r}^{k-1} > 0\}$
  {Add face constraint}
  $\tilde{\mathbf{A}}^k \leftarrow [\tilde{\mathbf{A}}^{k-1}, \tilde{\mathbf{a}}_{i^k}]$, $\mathcal{I}^k \leftarrow \mathcal{I}^{k-1} \cup \{i^k\}$,
  $\mathbf{B}^k \leftarrow (\tilde{\mathbf{A}}^k)^\dagger$, $\tilde{\mathbf{x}}^k \leftarrow \mathbf{B}^k\mathbf{y}$,
  while $\tilde{\mathbf{x}}^k \ngeq \mathbf{0}$ **do** {Release retarding constraints}
    Select some $j \in \mathcal{I}^k$ for which $\tilde{s}_j^k < 0$;
    $\tilde{\mathbf{A}}^k \leftarrow \tilde{\mathbf{A}}^k \setminus \tilde{\mathbf{a}}_j$, $\mathcal{I}^k \leftarrow \mathcal{I}^k \setminus \{j\}$,
    $\mathbf{B}^k \leftarrow (\tilde{\mathbf{A}}^k)^\dagger$, $\tilde{\mathbf{x}}^k \leftarrow \mathbf{B}^k\mathbf{y}$
  end while
  $\mathbf{c}^k \leftarrow (\mathbf{B}^k)^T\mathbf{1}$, $\hat{\mathbf{y}}^k \leftarrow \tilde{\mathbf{A}}^k\tilde{\mathbf{x}}^k$, $\mathbf{r}^k \leftarrow \mathbf{y} - \hat{\mathbf{y}}^k$
end while
$\tilde{\mathbf{x}} \leftarrow \mathbf{0}$ + corresponding elements of $\tilde{\mathbf{x}}^k$
$\mathbf{x} \triangleq [x_i]$ where $x_i \leftarrow (\tilde{x}_i - \tilde{x}_{i+n})$, $1 \leq i \leq n$
Output: $\mathbf{x}$

We can see that the basis vertex $\mathbf{c}$ plays a key role in this algorithm. The quantity $(1 - \tilde{\mathbf{a}}_i^T\mathbf{c}^{k-1})$ modifies the relative admissibility or otherwise of candidate basis vectors: faces determined by basis vectors with $\tilde{\mathbf{a}}_i^T \mathbf{c}^{k-1}$ close to 1 will be more likely to be encountered as the 'next face' than others. In addition, although it is not obvious from the algorithm statements, Algorithm 1 does in fact maintain the feasibility conditions $\tilde{\mathbf{A}}^T \mathbf{c} \leq \mathbf{1}$, i.e. $\mathbf{c} \in P^*$ during the addition and removal of basis vector constraints. For more discussion of this algorithm and experimental results see [12].

IV. The noisy case: quadratic programming

Let us now examine the homotopy method to find solutions for the noisy case (3)-(4). For our purposes we find it helpful to derive the homotopy method based on the work of Fuchs [5]. The quadratic program (4) has the following dual quadratic program (DQP)

$$\min_{\mathbf{x}} \|\mathbf{A}\mathbf{x}\|_2^2 \quad \text{such that} \quad \|\mathbf{A}^T(\mathbf{A}\mathbf{x} - \mathbf{y})\|_\infty \leq \lambda \tag{8}$$

from which it follows that at any feasible point $\mathbf{x}$ with residual $\mathbf{r} = \mathbf{y} - \mathbf{A}\mathbf{x}$ we have $\mathbf{a}_i^T \mathbf{r} \leq \lambda$ for all basis vectors $\mathbf{a}_i$ [5]. To get the standard form we construct $\tilde{\mathbf{A}} = [\mathbf{A}, -\mathbf{A}]$ and the nonnegative vector $\tilde{\mathbf{x}}$ as above, giving the standard dual quadratic program

$$\min_{\mathbf{x}} \|\tilde{\mathbf{A}}\tilde{\mathbf{x}}\|_2^2 \quad \text{such that} \quad \tilde{\mathbf{A}}^T(\mathbf{y} - \tilde{\mathbf{A}}\tilde{\mathbf{x}}) \leq \lambda \mathbf{1}. \tag{9}$$

The feasibility conditions can also be written $\tilde{\mathbf{a}}_i^T \mathbf{r} \leq \lambda, \ 1 \leq i \leq 2n$ where $\mathbf{r} = \mathbf{y} - \tilde{\mathbf{A}}\tilde{\mathbf{x}} = \mathbf{y} - \mathbf{A}\mathbf{x}$ is the residual.

Let $\tilde{\mathbf{x}}_{\text{act}}$ be the vector containing the nonzero ('active') elements of $\tilde{\mathbf{x}}$, with $\tilde{\mathbf{A}}_{\text{act}}$ the matrix containing the corresponding columns of $\tilde{\mathbf{A}}$, so that $\tilde{\mathbf{A}}_{\text{act}}\tilde{\mathbf{x}}_{\text{act}} = \tilde{\mathbf{A}}\tilde{\mathbf{x}}$ $(= \mathbf{A}\mathbf{x})$. Then by adapting the conditions from [5], we can write the necessary and sufficient conditions for $\tilde{\mathbf{x}}^*$ to be a solution of (9) (with active elements $\tilde{\mathbf{x}}^*_{\text{act}}$) as

$$\tilde{\mathbf{A}}_{\text{act}}^T(\mathbf{y} - \tilde{\mathbf{A}}_{\text{act}}\tilde{\mathbf{x}}^*_{\text{act}}) = \lambda \mathbf{1} \tag{10}$$

$$\tilde{\mathbf{a}}_j^T(\mathbf{y} - \tilde{\mathbf{A}}_{\text{act}}\tilde{\mathbf{x}}^*_{\text{act}}) \leq \lambda, \qquad \text{for all } \tilde{\mathbf{a}}_j \notin \tilde{\mathbf{A}}_{\text{act}} \tag{11}$$

where full rank $\tilde{\mathbf{A}}_{\text{act}}$ and strict inequality in (11) are sufficient for $\tilde{\mathbf{x}}^*$ to be unique.

From (10) all active basis functions satisfy $\tilde{\mathbf{a}}_j^T \mathbf{r} = \lambda$. In the case of unit norm basis vectors, this equal-inner-product condition becomes an equal-angle condition between the basis vectors $\tilde{\mathbf{a}}_j$ and the residual $\mathbf{r}$ [10].

Pre-multiplying (10) by $\tilde{\mathbf{A}}_{\text{act}}^{\dagger\,T}$ and using the orthogonal projection $\mathbf{Q}_{\text{act}} = \tilde{\mathbf{A}}_{\text{act}}\tilde{\mathbf{A}}_{\text{act}}^{\dagger} = \tilde{\mathbf{A}}_{\text{act}}^{\dagger\,T}\tilde{\mathbf{A}}_{\text{act}}^T$ for which $\mathbf{Q}_{\text{act}}\tilde{\mathbf{A}}_{\text{act}} = \tilde{\mathbf{A}}_{\text{act}}$ we get

$$\tilde{\mathbf{A}}_{\text{act}}^{\dagger\,T}\tilde{\mathbf{A}}_{\text{act}}^T(\mathbf{y} - \tilde{\mathbf{A}}_{\text{act}}\tilde{\mathbf{x}}^*_{\text{act}}) = \lambda \tilde{\mathbf{A}}_{\text{act}}^{\dagger\,T}\mathbf{1} \tag{12}$$

$$\mathbf{Q}_{\text{act}}\mathbf{y} - \mathbf{Q}_{\text{act}}\tilde{\mathbf{A}}_{\text{act}}\tilde{\mathbf{x}}^*_{\text{act}} = \lambda \mathbf{c}_{\text{act}} \tag{13}$$

$$-(\mathbf{I} - \mathbf{Q}_{\text{act}})\mathbf{y} + (\mathbf{y} - \tilde{\mathbf{A}}_{\text{act}}\tilde{\mathbf{x}}^*_{\text{act}}) = \lambda \mathbf{c}_{\text{act}} \tag{14}$$

$$(\mathbf{y} - \tilde{\mathbf{A}}_{\text{act}}\tilde{\mathbf{x}}^*_{\text{act}}) - (\mathbf{I} - \mathbf{Q}_{\text{act}})\mathbf{y} = +\lambda \mathbf{c}_{\text{act}} \tag{15}$$

$$\mathbf{r} = \bar{\mathbf{r}} + \lambda \mathbf{c}_{\text{act}} \tag{16}$$

where $\bar{\mathbf{r}} = (\mathbf{I} - \mathbf{Q}_{\text{act}})\mathbf{y}$ is the orthogonal residual, determined only by $\mathbf{y}$ and the active basis set $\tilde{\mathbf{A}}_{\text{act}}$. We can also write $\bar{\mathbf{r}} = \mathbf{y} - \bar{\mathbf{y}}$ with $\bar{\mathbf{y}} = \mathbf{Q}_{\text{act}}\mathbf{y}$ denoting the orthogonal projection of $\mathbf{y}$ into the subspace spanned by the active basis set. Since $\mathbf{Q}_{\text{act}} = \tilde{\mathbf{A}}_{\text{act}}\tilde{\mathbf{A}}_{\text{act}}^{\dagger}$ we can also write this as $\bar{\mathbf{y}} = \tilde{\mathbf{A}}_{\text{act}}\bar{\mathbf{x}}_{\text{act}}$ with $\bar{\mathbf{x}}_{\text{act}} = \tilde{\mathbf{A}}_{\text{act}}^{\dagger}\mathbf{y}$. Note that, in contrast to $\tilde{\mathbf{x}}_{\text{act}}$, $\bar{\mathbf{x}}_{\text{act}}$ is not guaranteed to have nonnegative entries. Finally, we also have $\bar{\mathbf{r}} = (\mathbf{I} - \mathbf{Q}_{\text{act}})\mathbf{r}$ since $\mathbf{Q}_{\text{act}}\mathbf{c}_{\text{act}} = \mathbf{c}_{\text{act}}$ and hence $(\mathbf{I} - \mathbf{Q}_{\text{act}})\mathbf{c}_{\text{act}} = \mathbf{0}$, so $\bar{\mathbf{r}}$ is the orthogonal project of the residual $\mathbf{r}$ into the space orthogonal to that spanned by the active basis vectors.

Therefore, for a given valid active basis set $\tilde{\mathbf{A}}_{\text{act}}$, we have $d\mathbf{r}/d\lambda = \mathbf{c}_{\text{act}} = \tilde{\mathbf{A}}_{\text{act}}^{\dagger\,T}\mathbf{1}$ so the residual $\mathbf{r}$ changes linearly with λ, with the direction of change given by the current active basis vertex $\mathbf{c}_{\text{act}}$. Similarly, since the denoised signal $\hat{\mathbf{y}}$ is given by $\hat{\mathbf{y}} \triangleq \mathbf{A}\mathbf{x}^* = \mathbf{y} - \mathbf{r}$, then $d\hat{\mathbf{y}}/d\lambda = -\mathbf{c}_{\text{act}}$, and therefore (in the domain where $\tilde{\mathbf{A}}_{\text{act}}$ is the active basis set) is 'shrunk' linearly in the direction $-\mathbf{c}_{\text{act}}$ as the penalty factor λ increases.

In the special case of unit norm basis vectors $\tilde{\mathbf{a}}_i$, since $\tilde{\mathbf{a}}_i^T \mathbf{c}_{\text{act}} = 1$ for all basis vectors in the active set, they all drop the same angle $\phi = \cos^{-1}(\tilde{\mathbf{a}}_i^T(\mathbf{c}_{\text{act}}/|\mathbf{c}_{\text{act}}|))$ with $\mathbf{c}_{\text{act}}$. Hence, as noted by Efron et al [10], $\mathbf{y}(\lambda)$ and $\mathbf{r}(\lambda)$ follow a path dropping equal angles to all basis vectors. However, for non-unit-norm basis vectors, this equal angle condition no longer holds, although we still have 'unity (equal) inner product' conditions such as $\tilde{\mathbf{a}}_i^T(d\hat{\mathbf{y}}/d\lambda) = 1$ for basis vectors in the active set.

V. Homotopy algorithm for the noisy case

If we consider λ to be a variable, from (10) for us to be just about to switch in a basis vector $\tilde{\mathbf{a}}_i$ we must have $\lambda = \tilde{\mathbf{a}}_i^T(\mathbf{y} - \tilde{\mathbf{A}}_{\text{act}}\tilde{\mathbf{x}}^*_{\text{act}}) = \tilde{\mathbf{a}}_i^T \mathbf{r}$. But since $\mathbf{r} = \bar{\mathbf{r}} + \lambda \mathbf{c}_{\text{act}}$ that means $\lambda = \tilde{\mathbf{a}}_i^T(\bar{\mathbf{r}} + \lambda \mathbf{c}_{\text{act}})$ leading to

$$\lambda = \frac{\tilde{\mathbf{a}}_i^T \bar{\mathbf{r}}}{1 - \tilde{\mathbf{a}}_i^T \mathbf{c}_{\text{act}}} \tag{17}$$

with $\bar{\mathbf{r}} = (\mathbf{I} - \mathbf{Q}_{\text{act}})\mathbf{y}$ the orthogonal residual (NB: not the usual residual $\mathbf{r} = \mathbf{y} - \hat{\mathbf{y}}$). For the homotopy method here we use a decreasing λ: the next basis function $\tilde{\mathbf{a}}^k$ to be switched in is

$$\tilde{\mathbf{a}}^k = \arg \max_{\tilde{\mathbf{a}}_i \notin \tilde{\mathbf{A}}_{\text{act}}} \frac{\tilde{\mathbf{a}}_i^T \bar{\mathbf{r}}}{1 - \tilde{\mathbf{a}}_i^T \mathbf{c}_{\text{act}}} \tag{18}$$

where the maximum is over all basis vectors not currently in the active set. We are immediately struck by the similarity of (18) with (7): with the use of the orthogonal residual $\bar{\mathbf{r}}$ in place of the original noiseless residual $\mathbf{r}$ the condition is identical.

In the full algorithm we also need to consider whether we should switch out a basis function first. A little manipulation gives us that $d\tilde{\mathbf{x}}^*_{\text{act}}/d\lambda = -\tilde{\mathbf{z}}_{\text{act}}$ where $\tilde{\mathbf{z}}_{\text{act}} = \tilde{\mathbf{A}}_{\text{act}}^{\dagger}\mathbf{c}_{\text{act}}$ is the vector of coefficients of $\mathbf{c}_{\text{act}}$ represented in terms of the basis vectors, $\mathbf{c}_{\text{act}} = \tilde{\mathbf{A}}_{\text{act}}\tilde{\mathbf{z}}_{\text{act}}$. Hence if we have an internal basis, for which $\mathbf{z} \geq \mathbf{0}$, the components of $\tilde{\mathbf{x}}^*_{\text{act}}$ can only increase as λ decreases, so no vectors will be switched out. Efron et al [10] call $\mathbf{z} > \mathbf{0}$ the *positive cone condition*. Straightforward analysis confirms that a given basis vector $\tilde{\mathbf{a}}_i$ is only switched in or out when it has a zero coefficient $\tilde{x}_i = 0$, so the coefficient vector $\tilde{\mathbf{x}}$, denoised reconstruction $\hat{\mathbf{y}} = \tilde{\mathbf{A}}\tilde{\mathbf{x}}$, and full residual $\mathbf{r} = \mathbf{y} - \hat{\mathbf{y}}$ are all unchanged at the switching point.

The complete algorithm is given in Algorithm 2, and is essentially an explicit algorithm to implement the decreasing-λ approach described in the recent paper by Malioutov et al [9]. In contrast, the homotopy method of Osborne et al [8] and the LARS algorithm of Efron et al [10] work by increasing t in (3).

Algorithm 2 Homotopy Continuation

Input: $\mathbf{y}$, $\tilde{\mathbf{A}} \triangleq [\tilde{\mathbf{a}}_i] = [\mathbf{A}, -\mathbf{A}]$
Set stopping condition $\lambda_{\min} \geq 0$
Initialize: $k \leftarrow 0,\ \lambda^k \leftarrow \infty,\ \mathcal{I}^k \leftarrow \emptyset,$
$\tilde{\mathbf{A}}^k \leftarrow \emptyset,\ \mathbf{c}^k \leftarrow \mathbf{0},\ \tilde{\mathbf{z}}^k \leftarrow \emptyset,\ \overline{\mathbf{x}}^k \leftarrow \emptyset,\ \overline{\mathbf{r}}^k \leftarrow \mathbf{y}$
while $\lambda^k > \lambda_{\min}$ **do**
  $k \leftarrow k+1$
  {Find update candidate}
  $\lambda_i \leftarrow (\tilde{\mathbf{a}}_i^T \overline{\mathbf{r}}^{k-1})/(1 - \tilde{\mathbf{a}}_i^T \mathbf{c}^{k-1}),\ 1 \leq i \leq 2n,$
  $i_+^k \leftarrow \arg\max_{i \notin \mathcal{I}^{k-1}} \{\lambda_i \mid \tilde{\mathbf{a}}_i^T \overline{\mathbf{r}}^{k-1} > 0\},$
  $\lambda_+^k \leftarrow (\tilde{\mathbf{a}}_{i_+^k}^T \overline{\mathbf{r}}^{k-1})/(1 - \tilde{\mathbf{a}}_{i_+^k}^T \mathbf{c}^{k-1})$
  if $\tilde{\mathbf{z}}^{k-1} \ngeq 0$ **then** {Find downdate candidate}
    $i_-^k \leftarrow \arg\max_{i \in \mathcal{I}^{k-1}} \{\bar{x}_i^{k-1}/\tilde{z}_i^{k-1} \mid z_i^{k-1} < 0\},$
    $\lambda_-^k \leftarrow \bar{x}_{i_-^k}^{k-1}/\tilde{z}_{i_-^k}^{k-1}$
  else
    $i_-^k \leftarrow 0,\ \lambda_-^k \leftarrow 0$
  end if
  if $\max(\lambda_+^k, \lambda_-^k) > \lambda_{\min}$ **then** {Adjust basis}
    if $\lambda_+^k > \lambda_-^k$ **then** {Update}
      $\lambda^k \leftarrow \lambda_+^k, \tilde{\mathbf{A}}^k \leftarrow [\tilde{\mathbf{A}}^{k-1}, \tilde{\mathbf{a}}_{i_+^k}], \mathcal{I}^k \leftarrow \mathcal{I}^{k-1} \cup \{i_+^k\}$
    else {Downdate}
      $\lambda^k \leftarrow \lambda_-^k, \tilde{\mathbf{A}}^k \leftarrow \tilde{\mathbf{A}}^{k-1} \setminus \tilde{\mathbf{a}}_{i_-^k}, \mathcal{I}^k \leftarrow \mathcal{I}^{k-1} \setminus \{i_-^k\}$
    end if
    $\mathbf{B}^k \leftarrow (\tilde{\mathbf{A}}^k)^\dagger,\ \mathbf{c}^k \leftarrow (\mathbf{B}^k)^T \mathbf{1},\ \tilde{\mathbf{z}}^k \leftarrow \mathbf{B}^k \mathbf{c}^k,$
    $\overline{\mathbf{x}}^k \leftarrow \mathbf{B}^k \mathbf{y},\ \overline{\mathbf{y}}^k \leftarrow \tilde{\mathbf{A}}^k \overline{\mathbf{x}}^k,\ \overline{\mathbf{r}}^k \leftarrow \mathbf{y} - \overline{\mathbf{y}}^k$
  else {Finalize}
    $\lambda^k \leftarrow \lambda_{\min}$
  end if
end while
$\tilde{\mathbf{x}}^k \leftarrow \overline{\mathbf{x}}^k - \lambda^k \tilde{\mathbf{z}}^k,$
$\tilde{\mathbf{x}} \leftarrow \mathbf{0} +$ corresponding elements of $\tilde{\mathbf{x}}^k$
$\mathbf{x} = [x_i]$ where $x_i \leftarrow (\tilde{x}_i - \tilde{x}_{i+n}),\ 1 \leq i \leq n$
Output: $\mathbf{x},\ \hat{\mathbf{y}}$

To demonstrate the operation of the homotopy continuation algorithm (Algorithm 2), Fig. 4 shows the signal ‘Gong’, a decaying sinusoid after $t = t_0$, analyzed with a cosine packet dictionary [3]. We can clearly see the successive introduction of active components of the sparse representation as λ decreases from left to right (Fig. 4(c)).

VI. Conclusions

We have explored the geometry of ℓ_1 sparse representations, in both the noiseless (Basis Pursuit) case and the noisy case, and found that the *basis vertex* is a useful geometric concept in both cases. We derived an explicit homotopy continuation algorithm for the noisy case, and we saw that the criterion to add a new basis vector is very similar to that in the greedy Polytope Faces Pursuit algorithm for the noiseless case. Numerical simulations confirm that the algorithm behaves as expected.

Acknowledgements

This work was partially supported by EPSRC grants GR/S82213/01, GR/S75802/01, EP/C005554/1 and EP/D000246/1. Some of the figures were generated using the Multi-Parametric Toolbox (MPT) for Matlab [16], WaveLab802 (`http://www-stat.stanford.edu/~wavelab/`) and Atomizer802 (`http://www-stat.stanford.edu/~atomizer/`).

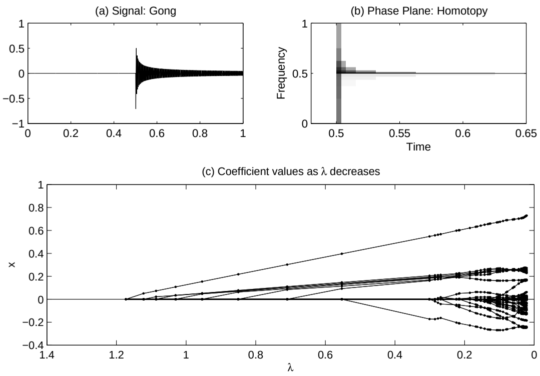


Fig. 4. Analysis of signal ‘Gong’, showing (a) original signal, (b) analysis, and (c) coefficient values x_i as λ decreases.

References

[1] B. A. Turlach, “On algorithms for solving least squares problems under an l_1 penalty or an l_1 constraint,” in *2004 Proceedings of the American Statistical Association*. Statistical Computing Section [CD-ROM], American Statistical Association, Alexandria, VA, 2005, pp. 2572–2577.

[2] R. Tibshirani, “Regression shrinkage and selection via the lasso,” *J. Royal. Statist. Soc B*, vol. 58, no. 1, pp. 267–288, 1996.

[3] S. S. Chen, D. L. Donoho, and M. A. Saunders, “Atomic decomposition by basis pursuit,” *SIAM Journal on Scientific Computing*, vol. 20, no. 1, pp. 33–61, 1998.

[4] D. Donoho, M. Elad, and V. Temlyakov, “Stable recovery of sparse overcomplete representations in the presence of noise,” 2005, to appear in: IEEE Trans. On Information Theory.

[5] J.-J. Fuchs, “On sparse representations in arbitrary redundant bases,” *IEEE Transactions on Information Theory*, vol. 50, no. 6, pp. 1341–1344, 2004.

[6] K. Kreutz-Delgado, J. F. Murray, B. D. Rao, K. Engan, T.-W. Lee, and T. J. Sejnowski, “Dictionary learning algorithms for sparse representation,” *Neural Computation*, vol. 15, pp. 349–396, 2003.

[7] J. A. Tropp, “Just relax: Convex programming methods for subset selection and sparse approximation,” Institute for Computational Engineering and Sciences, University of Texas at Austin, ICES Report 04-04, February 2004.

[8] M. R. Osborne, B. Presnell, and B. A. Turlach, “A new approach to variable selection in least squares problems,” *IMA Journal of Numerical Analysis*, vol. 20, no. 3, pp. 389–403, 2000.

[9] D. M. Malioutov, M. Cetin, and A. S. Willsky, “Homotopy continuation for sparse signal representation,” in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, (ICASSP ’05)*, March 18-23 2005, pp. V: 733 – 736.

[10] B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani, “Least angle regression,” *Annals of Statistics*, vol. 32, no. 2, pp. 407–499, Apr. 2004.

[11] M. D. Plumbley, “Polar polytopes and recovery of sparse representations,” 2005, submitted for publication.

[12] ——, “Recovery of sparse representations by polytope faces pursuit,” 2005, submitted.

[13] D. L. Donoho, “Neighborly polytopes and sparse solutions of underdetermined linear equations,” Statistics Department, Stanford University, Tech. Rep., December 2004.

[14] M. H. Wright, “The interior-point revolution in optimization: History, recent developments, and lasting consequences,” *Bulletin (New Series) of the American Mathematical Society*, vol. 42, no. 1, pp. 39–56, 2004.

[15] Y. C. Pati, R. Rezaiifar, and P. S. Krishnaprasad, “Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition,” in *Conference Record of The Twenty-Seventh Asilomar Conference on Signals, Systems and Computers, Pacific Grove, CA*, 1-3 Nov. 1993, pp. 40–44.

[16] M. Kvasnica, P. Grieder, and M. Baotić, “Multi-Parametric Toolbox (MPT),” 2004. [Online]. Available: http://control.ee.ethz.ch/~mpt/