On Polar Polytopes and the Recovery of Sparse Representations

Mark D. Plumbley, *Member, IEEE*

Abstract—Suppose we have a signal $\mathbf{y}$ which we wish to represent using a linear combination of a number of basis atoms $\mathbf{a}_i$, $\mathbf{y} = \sum_i x_i \mathbf{a}_i = \mathbf{A}\mathbf{x}$. The problem of finding the minimum ℓ_0 norm representation for $\mathbf{y}$ is a hard problem. The Basis Pursuit (BP) approach proposes to find the minimum ℓ_1 norm representation instead, which corresponds to a linear program (LP) that can be solved using modern LP techniques, and several recent authors have given conditions for the BP (minimum ℓ_1 norm) and sparse (minimum ℓ_0 norm) representations to be identical. In this paper, we explore this *sparse representation problem* using the geometry of convex polytopes, as recently introduced into the field by Donoho. By considering the dual LP we find that the so-called polar polytope P^* of the centrally-symmetric polytope P whose vertices are the atom pairs $\pm\mathbf{a}_i$ is particularly helpful in providing us with geometrical insight into optimality conditions given by Fuchs and Tropp for non-unit-norm atom sets. In exploring this geometry we are able to tighten some of these earlier results, showing for example that a condition due to Fuchs is both necessary and sufficient for ℓ_1-unique-optimality, and there are cases where Orthogonal Matching Pursuit can eventually find all ℓ_1-unique-optimal solutions with m nonzeros even if the Exact Recover Condition fails for m.

Index Terms—Basis Pursuit (BP), Orthogonal Matching Pursuit (OMP), linear programming, polytopes, sparse representations.

I. INTRODUCTION

Suppose we have a real vector $\mathbf{y} = [y_1, \ldots, y_d]^T$ which we wish to represent using a linear combination from $n > d$ nonzero d-dimensional real basis atoms $\mathbf{a}_i$, $\mathbf{y} = \sum_i x_i \mathbf{a}_i$. In other words, we wish to find an n-vector $\mathbf{x} = [x_1, \ldots, x_n]^T$ such that $\mathbf{y} = \mathbf{A}\mathbf{x}$, where $\mathbf{A} = [\mathbf{a}_i]$ is the $d \times n$ matrix whose ith column is $\mathbf{a}_i$. Since $n > d$ there are many possible representations $\mathbf{A}\mathbf{x} = \mathbf{y}$ for a given $\mathbf{A}$ and $\mathbf{y}$. The *sparse representation problem* is then to find the representation $\mathbf{x}$ with the fewest possible non-zero components,

$$\min_{\mathbf{x}} \|\mathbf{x}\|_0 \quad \text{such that} \quad \mathbf{A}\mathbf{x} = \mathbf{y} \qquad \text{(P0)}$$

where $\|\mathbf{x}\|_0$ is the ℓ_0 norm of $\mathbf{x}$, i.e. the number of non-zero elements. In the special case where the atoms $\mathbf{a}_i$ have unit norm $\|\mathbf{a}_i\|_2 = 1$, we may call $\mathbf{A}$ a *dictionary* [1], although this unit-norm requirement is not necessary for most of the results in the present paper.

Problem (P0) is known to be NP-hard [2], so instead we can consider the 'relaxed' ℓ_1 problem

$$\min_{\mathbf{x}} \|\mathbf{x}\|_1 \quad \text{such that} \quad \mathbf{A}\mathbf{x} = \mathbf{y} \qquad \text{(P1)}$$

M. D. Plumbley is with the Department of Electronic Engineering, Queen Mary, University of London, London E1 4NS, U.K. (e-mail: mark.plumbley@elec.qmul.ac.uk). This work is partially supported by EPSRC grants GR/S82213/01, GR/S75802/01, EP/C005554/1 and EP/D000246/1.

where $\|\mathbf{x}\|_1 = \sum_i |x_i|$ is the ℓ_1 norm of $\mathbf{x}$. Problem (P1), which is known as *Basis Pursuit* (BP) in the signal processing community [3], can be formulated as a linear programming problem, and can be solved using well known optimization methods such as the simplex method or interior point methods [4]. A number of authors have explored the conditions under which the minimum of (P1) is unique and identical to a unique minimum of (P0), sometimes called *exact recovery* or ℓ_1/ℓ_0 *equivalence* [5]–[9].

In fact this property of ℓ_1/ℓ_0 equivalence for a particular solution $\mathbf{x}_0$ requires showing both that $\mathbf{x}_0$ is is the unique minimum of (P0), known as ℓ_0-*unique-optimality*, and also that it the unique minimum of (P1), known as ℓ_1-*unique-optimality* [7]. In this paper we mostly concentrate on the problem of ℓ_1-unique-optimality, and use the result of Donoho and Elad [7] for ℓ_0-unique-optimality.

In an interesting new direction, Donoho [10], [11] has explored the link between the recovery of sparse representations and the geometry of *polytopes*, convex sets defined by a finite set of vertices or inequalities. Donoho showed that ℓ_1/ℓ_0 equivalence of certain representations $\mathbf{x}_0$ can be linked to the existence of particular faces of a polytope P whose vertices are the atom pairs $\pm\mathbf{a}_i$ with $\mathbf{a}_i \in \mathbf{A}$. If the atom pairs $\pm\mathbf{a}_i$ are in general position, and P is k-*neighbourly*, that is if each subset of k signed atoms forms the vertices of a true face of P, then ℓ_1/ℓ_0 equivalence holds for all representations $\mathbf{x}_0$ with at most k nonzeros. This powerful new approach means that results from the field of polytopes can be brought across to the sparse representations problem, and vice versa. For example, using the classic work of McMullen and Shephard [12] on centrally symmetric polytopes, Donoho showed [10, Corollary 1.3] the surprising result that for $n-2 \geq d > 2$, the condition $k \leq \lfloor (d+1)/3 \rfloor$ must hold for ℓ_1/ℓ_0 equivalence of all representations $\mathbf{x}_0$ having at most k nonzeros.

For linear programming problems we can construct a dual problem, which can sometimes lead to a simpler solution than the original, primal problem. In a related way, the polytope P with vertices $\pm\mathbf{a}_i$ also has a dual, $P^* = \{\mathbf{y} \mid \pm\mathbf{a}_i^T\mathbf{y} \leq 1\}$, called the *polar polytope* of P. In this paper we will use the dual problem and the polar polytope P^* to give us new insight into the sparse representation problem. We will investigate two results due to Fuchs, which we call the *Fuchs Condition* [13] and *Fuchs Corollary* [14]. We will show that the Fuchs Condition is both necessary and sufficient for ℓ_1-unique-optimality, is equivalent to existence of a face on the polar polytope P^*, and equivalent to a previous result of Donoho [10]. We shall also see that the Fuchs Corollary and the Exact Recovery Condition of Tropp [9] are corollaries of the Fuchs

Condition, and give interpretations of these in terms of our polytope geometry. We end with some graphical examples, which illustrate the differences between these conditions for both non-unit-norm atom sets and unit-norm dictionaries.

II. FUCHS CONDITION: NECESSITY AND SUFFICIENCY

Let us write the linear program (P1) in its *standard form* [15]

$$\min_{\tilde{\mathbf{x}}} \mathbf{1}^T \tilde{\mathbf{x}} \quad \text{such that} \quad \tilde{\mathbf{A}}\tilde{\mathbf{x}} = \mathbf{y} \quad \text{and} \quad \tilde{\mathbf{x}} \geq \mathbf{0} \quad \text{(LP)}$$

where $\mathbf{1}$ is a vector of 1s, $\tilde{\mathbf{A}} = [\mathbf{A}, -\mathbf{A}]$ and $\tilde{\mathbf{x}}^T = [\mathbf{x}_+^T, \mathbf{x}_-^T]$ with $\mathbf{x} = \mathbf{x}_+ - \mathbf{x}_-$ such that $\mathbf{x}_+, \mathbf{x}_- \geq \mathbf{0}$, i.e. $[\mathbf{x}_+]_i = \max(x_i, 0)$ and $[\mathbf{x}_-]_i = \max(-x_i, 0)$. Thus any solution to $\mathbf{A}\mathbf{x} = \mathbf{y}$ can be written in the form $\tilde{\mathbf{A}}\tilde{\mathbf{x}} = \mathbf{y}$ with nonnegative $\tilde{\mathbf{x}}$.

The primal linear program (LP) has a corresponding dual problem

$$\max_{\mathbf{c}} \mathbf{y}^T \mathbf{c} \quad \text{such that} \quad \tilde{\mathbf{A}}^T \mathbf{c} \leq \mathbf{1}. \quad \text{(DLP)}$$

If a solution to (LP) exists, i.e. $\mathbf{y} \in \mathrm{colspan}(\mathbf{A})$, then we say the primal problem (LP) is *feasible*. For a linear program, if the primal problem is feasible, then *strong duality* holds between (LP) and (DLP). Hence (DLP) must also be feasible, and the optimum values of (LP) and (DLP) are equal [15]. Therefore we have the following optimality conditions for this system:

Lemma 1: Suppose that problem (LP) is feasible. Then the pair $\tilde{\mathbf{x}}^*, \mathbf{c}^*$ is an optimum point of (LP), (DLP) *if and only if* the following conditions hold:

1. $\tilde{\mathbf{A}}\tilde{\mathbf{x}}^* = \mathbf{y}, \quad \tilde{\mathbf{x}}^* \geq \mathbf{0}$ (Primal feasibility)

2. $\tilde{\mathbf{A}}^T \mathbf{c}^* \leq \mathbf{1}$ (Dual feasibility)

3. $(\tilde{\mathbf{a}}_j^T \mathbf{c}^* - 1)\tilde{x}_j^* = 0$

    for all $\quad j = 1, \ldots, 2n$ (Complementary slackness)

Proof: This follows immediately by writing down the Karush-Kuhn-Tucker (KKT) conditions (see e.g. [15]) for this system and eliminating free variables. ∎

Suppose now that $\tilde{\mathbf{x}}_0 \geq \mathbf{0}$ is a solution of $\tilde{\mathbf{A}}\tilde{\mathbf{x}} = \mathbf{y}$ with $m = \|\tilde{\mathbf{x}}_0\|_0$ nonzeros. Let $\tilde{\mathbf{x}}_{\mathrm{opt}}$ be the m-dimensional vector built from the nonzero components of $\tilde{\mathbf{x}}_0$, with $\tilde{\mathbf{A}}_{\mathrm{opt}}$ the $d \times m$ matrix built from the corresponding columns of $\tilde{\mathbf{A}}$, such that $\mathbf{y} = \tilde{\mathbf{A}}_{\mathrm{opt}}\tilde{\mathbf{x}}_{\mathrm{opt}} = \tilde{\mathbf{A}}\tilde{\mathbf{x}}_0$. Then the following result gives the necessary and sufficient conditions for $\tilde{\mathbf{x}}_0$ to be the unique optimum of (LP):

Theorem 1 (Fuchs Condition: Standard Form): $\tilde{\mathbf{x}}_0$ is the unique optimum point of (LP) if and only if (a) $\tilde{\mathbf{A}}_{\mathrm{opt}}$ has full rank and (b) there exists some $\mathbf{c} \in \mathbb{R}^d$ such that

$$\tilde{\mathbf{a}}_j^T \mathbf{c} = 1 \quad \tilde{\mathbf{a}}_j \in \tilde{\mathbf{A}}_{\mathrm{opt}} \quad (1)$$

$$\tilde{\mathbf{a}}_j^T \mathbf{c} < 1 \quad \tilde{\mathbf{a}}_j \notin \tilde{\mathbf{A}}_{\mathrm{opt}} \quad (2)$$

where $\tilde{\mathbf{a}}_j$ ranges over the columns of $\tilde{\mathbf{A}}$.

Proof: For the 'if' direction, we note that $\tilde{\mathbf{x}}_0$ satisfies the the primal feasibility condition of Lemma 1, and conditions (1) and (2) ensure that $\mathbf{c}$ satisfies the dual feasibility condition. Furthermore, the nonzero components of $\tilde{\mathbf{x}}_0$ correspond to the basis vectors $\tilde{\mathbf{a}}_j \in \tilde{\mathbf{A}}_{\mathrm{opt}}$, which therefore have $\tilde{\mathbf{a}}_j^T \mathbf{c} = 1$,

i.e. $\tilde{\mathbf{a}}_j^T \mathbf{c} - 1 = 0$, so $\tilde{\mathbf{x}}_0$ and $\mathbf{c}$ satisfy the complementary slackness condition, and $\tilde{\mathbf{x}}_0$ must be an optimal point of (LP). Since $\tilde{\mathbf{A}}_{\mathrm{opt}}$ is full rank, the optimal solution is unique and is given by $\tilde{\mathbf{x}}_{\mathrm{opt}} = \tilde{\mathbf{A}}_{\mathrm{opt}}^\dagger \mathbf{y}$, where $\tilde{\mathbf{A}}_{\mathrm{opt}}^\dagger$ is the Moore-Penrose pseudoinverse of $\tilde{\mathbf{A}}_{\mathrm{opt}}$.

For the 'only if' direction, first let us consider condition (a). Suppose tentatively that $\tilde{\mathbf{A}}_{\mathrm{opt}}$ does not have full rank. Then let us choose a small nonzero vector $\tilde{\mathbf{x}}_{\mathrm{opt}}^N \neq 0$ in the null space of $\tilde{\mathbf{A}}_{\mathrm{opt}}$, i.e. $\tilde{\mathbf{A}}_{\mathrm{opt}}\tilde{\mathbf{x}}_{\mathrm{opt}}^N = 0$. We choose $\tilde{\mathbf{x}}_{\mathrm{opt}}^N$ to be sufficiently small so that its largest absolute element value $\|\tilde{\mathbf{x}}_{\mathrm{opt}}^N\|_\infty$ is strictly less than the smallest element of $\tilde{\mathbf{x}}_{\mathrm{opt}}$. Therefore the three m-vectors $\tilde{\mathbf{x}}_{\mathrm{opt}} - \tilde{\mathbf{x}}_{\mathrm{opt}}^N$, $\tilde{\mathbf{x}}_{\mathrm{opt}}$ and $\tilde{\mathbf{x}}_{\mathrm{opt}} + \tilde{\mathbf{x}}_{\mathrm{opt}}^N$ have all strictly positive elements and satisfy $\tilde{\mathbf{A}}_{\mathrm{opt}}(\tilde{\mathbf{x}}_{\mathrm{opt}} - \tilde{\mathbf{x}}_{\mathrm{opt}}^N) = \tilde{\mathbf{A}}_{\mathrm{opt}}\tilde{\mathbf{x}}_{\mathrm{opt}} = \tilde{\mathbf{A}}_{\mathrm{opt}}(\tilde{\mathbf{x}}_{\mathrm{opt}} + \tilde{\mathbf{x}}_{\mathrm{opt}}^N) = \mathbf{y}$, so these three vectors all correspond to feasible points of (LP). Now if $\mathbf{1}^T \tilde{\mathbf{x}}_{\mathrm{opt}}^N > 0$, we have $\mathbf{1}^T(\tilde{\mathbf{x}}_{\mathrm{opt}} - \tilde{\mathbf{x}}_{\mathrm{opt}}^N) = \mathbf{1}^T \tilde{\mathbf{x}}_{\mathrm{opt}} - \mathbf{1}^T \tilde{\mathbf{x}}_{\mathrm{opt}}^N < \mathbf{1}^T \tilde{\mathbf{x}}_{\mathrm{opt}}$, so $\tilde{\mathbf{x}}_{\mathrm{opt}} - \tilde{\mathbf{x}}_{\mathrm{opt}}^N$ corresponds to some $\tilde{\mathbf{x}}'$ with a smaller cost $\mathbf{1}^T \tilde{\mathbf{x}}' < \mathbf{1}^T \tilde{\mathbf{x}}_0$, and so $\tilde{\mathbf{x}}_0$ cannot be the optimum of (LP), yielding a contradiction. Similarly if $\mathbf{1}^T \tilde{\mathbf{x}}_{\mathrm{opt}}^N < 0$ we have $\mathbf{1}^T(\tilde{\mathbf{x}}_{\mathrm{opt}} + \tilde{\mathbf{x}}_{\mathrm{opt}}^N) < \mathbf{1}^T \tilde{\mathbf{x}}_{\mathrm{opt}}$, so again $\tilde{\mathbf{x}}_0$ cannot be the optimum of (LP). Finally if $\mathbf{1}^T \tilde{\mathbf{x}}_{\mathrm{opt}}^N = 0$, then $\mathbf{1}^T(\tilde{\mathbf{x}}_{\mathrm{opt}} + \tilde{\mathbf{x}}_{\mathrm{opt}}^N) = \mathbf{1}^T \tilde{\mathbf{x}}_{\mathrm{opt}}$, so the distinct vectors $\tilde{\mathbf{x}}_{\mathrm{opt}} + \tilde{\mathbf{x}}_{\mathrm{opt}}^N$ and $\tilde{\mathbf{x}}_{\mathrm{opt}}$ have the same cost, contradicting the requirement for $\tilde{\mathbf{x}}_0$ to be the *unique* minimum of (LP). Therefore any $\tilde{\mathbf{A}}_{\mathrm{opt}}$ must have full rank for $\tilde{\mathbf{x}}_0$ to be the unique optimum of (LP).

Now consider equation (1) of condition (b). Since $\tilde{\mathbf{x}}_0$ is a minimum of (LP) the problem is feasible, and by strong duality and Lemma 1 there must also be at least one optimum $\mathbf{c}^*$ to the dual problem (DLP) satisfying the dual feasibility condition $\tilde{\mathbf{A}}^T \mathbf{c}^* \leq \mathbf{1}$, i.e. $\tilde{\mathbf{a}}_i^T \mathbf{c}^* \leq 1$ for all $i = 1, \ldots, 2n$. Furthermore any optimum $\mathbf{c}^*$ must satisfy complementary slackness, so for any i with $\tilde{x}_i > 0$, i.e. $\tilde{\mathbf{a}}_i \in \tilde{\mathbf{A}}_{\mathrm{opt}}$, we must have $\tilde{\mathbf{a}}_i^T \mathbf{c}^* = 1$.

Finally, consider the remaining inequality (2) of condition (b). Since (LP) has a solution, it has a *strictly complementary* solution [16], i.e. one satisfying $(1 - \tilde{\mathbf{a}}_j^T \mathbf{c}^*) + \tilde{x}_j^* > 0$ for all $j = 1, \ldots, 2n$. Therefore using this $\mathbf{c}^*$, for any j such that $\tilde{\mathbf{a}}_j \notin \tilde{\mathbf{A}}_{\mathrm{opt}}$ and hence $\tilde{x}_j^* = 0$, we must have $(1 - \tilde{\mathbf{a}}_j^T \mathbf{c}^*) > 0$, which is the strict inequality required for equation (2). ∎

This theorem gives us necessary and sufficient conditions for $\tilde{\mathbf{x}}_0$ to be ℓ_1-unique-optimal. Let us now find the equivalent conditions for the original problem (P1), the sufficiency ('if' direction) of which was shown by Fuchs [13, Theorem 4], so we refer to this as the *Fuchs Condition*. Suppose that $\mathbf{x}_0$ is a solution of $\mathbf{A}\mathbf{x} = \mathbf{y}$ with $m = \|\mathbf{x}_0\|_0$ nonzeros. Let $\mathbf{x}_{\mathrm{opt}}$ be the m-dimensional vector built from the nonzero components of $\mathbf{x}_0$, with $\mathbf{A}_{\mathrm{opt}}$ the $d \times m$ matrix built from the corresponding columns of $\mathbf{A}$ such that $\mathbf{y} = \mathbf{A}_{\mathrm{opt}}\mathbf{x}_{\mathrm{opt}} = \mathbf{A}\mathbf{x}_0$. Then we have the following result:

Theorem 2 (Fuchs Condition: Original Form): Let $\mathbf{x}_0$, $\mathbf{A}_{\mathrm{opt}}$ be defined as above. Then $\mathbf{x}_0$ is the unique solution to (P1) if and only if (a) $\mathbf{A}_{\mathrm{opt}}$ is full rank, and (b) there exists some $\mathbf{c} \in \mathbb{R}^d$ satisfying

$$\mathbf{A}_{\mathrm{opt}}^T \mathbf{c} = \mathrm{sign}\, \mathbf{x}_{\mathrm{opt}} \quad (3)$$

$$|\mathbf{a}_j^T \mathbf{c}| < 1 \quad \text{for any } \mathbf{a}_j \in \mathbf{A}, \mathbf{a}_j \notin \mathbf{A}_{\mathrm{opt}} \quad (4)$$

The Fuchs Condition (Theorem 2, or Theorem 1 in standard

form) is therefore the weakest possible condition for ℓ_1-unique-optimality.

The proof of Theorem 2 (necessity and sufficiency of the Fuchs Condition) follows immediately from the following Lemma.

Lemma 2: A vector $\mathbf{c}$ satisfies condition (b) in Theorem 2 if and only if it satisfies the condition (b) in Theorem 1.

Proof: First we note that $\mathbf{A}_{\text{opt}}$ and $\tilde{\mathbf{A}}_{\text{opt}}$ contain identical columns except for sign changes, so the full rank condition on each is equivalent.

For the other conditions in Theorem 2, for $\mathbf{a}_j \in \mathbf{A}_{\text{opt}}$, for which $[\mathbf{x}_{\text{opt}}]_j \neq 0$, we have $\mathbf{a}_j^T \mathbf{c} = \text{sign}([\mathbf{x}_{\text{opt}}]_j)$. If $[\mathbf{x}_{\text{opt}}]_j > 0$ we get $\mathbf{a}_j^T \mathbf{c} = 1$ and $-\mathbf{a}_j^T \mathbf{c} = -1 < 1$ so $\tilde{\mathbf{a}}_j = \mathbf{a}_j \in \tilde{\mathbf{A}}_{\text{opt}}$, $\tilde{\mathbf{a}}_{n+j} = -\mathbf{a}_j \notin \tilde{\mathbf{A}}_{\text{opt}}$. Alternatively if $[\mathbf{x}_{\text{opt}}]_j < 0$ we get $\mathbf{a}_j^T \mathbf{c} = -1 < 1$ and $-\mathbf{a}_j^T \mathbf{c} = 1$ so $\tilde{\mathbf{a}}_j = \mathbf{a}_j \notin \tilde{\mathbf{A}}_{\text{opt}}$, $\tilde{\mathbf{a}}_{n+j} = -\mathbf{a}_j \in \tilde{\mathbf{A}}_{\text{opt}}$. For $\mathbf{a}_j \notin \mathbf{A}_{\text{opt}}$, we have $|\mathbf{a}_j^T \mathbf{c}| < 1$ so $-1 < \mathbf{a}_j^T \mathbf{c} < 1$, i.e. $-\mathbf{a}_j^T \mathbf{c} < 1$ and $+\mathbf{a}_j^T \mathbf{c} < 1$, thus $\tilde{\mathbf{a}}_{n+j}^T \mathbf{c} < 1$ and $\tilde{\mathbf{a}}_j^T \mathbf{c} < 1$, so $\tilde{\mathbf{a}}_{n+j} \notin \tilde{\mathbf{A}}_{\text{opt}}$ and $\tilde{\mathbf{a}}_j \notin \tilde{\mathbf{A}}_{\text{opt}}$.

Showing the converse is similarly straightforward, noting that $\mathbf{a}_j^T \mathbf{c} = 1$ and $-\mathbf{a}_j^T \mathbf{c} = 1$ can never both be satisfied at once. ∎

III. GEOMETRY OF THE FUCHS CONDITION

In its original form, the Fuchs Condition (Theorem 2) is somewhat difficult to interpret (see e.g. comments in [17], [18]). However, in its standard form (Theorem 1) we find that we can relate it to the geometry of *polytopes* [19].

In the context of sparse representations, Donoho [10] introduced the polytope $P = \text{conv}\{\pm\mathbf{a}_i | \mathbf{a}_i \in \mathbf{A}\}$, the convex hull of the positive and negative versions of the basis atoms $\mathbf{a}_i$. In our standard-form notation with $\tilde{\mathbf{A}} = [\mathbf{A}, -\mathbf{A}]$, we can equivalently express this polytope as $P = \text{conv}\{\tilde{\mathbf{a}}_i | \tilde{\mathbf{a}}_i \in \tilde{\mathbf{A}}\}$. This type of polytope, with mirror-image vertices, is called a *centrally-symmetric* polytope, since $\mathbf{u} \in P \implies -\mathbf{u} \in P$ so it is symmetric about the origin. In general the vertices of P will be a subset of the atom pairs, although for the unit norm case $\|\mathbf{a}_i\|_2 = 1$, uniqueness of the atom pairs $\pm\mathbf{a}_i$ is sufficient for them all to be vertices.

Now, for the polytope P let us introduce its associated *polar polytope*

$$P^* = \{\mathbf{c} \mid \tilde{\mathbf{a}}_i^T \mathbf{c} \leq 1, \tilde{\mathbf{a}}_i \in \tilde{\mathbf{A}}\} = \{\mathbf{c} \mid \tilde{\mathbf{A}}^T \mathbf{c} \leq \mathbf{1}\}. \quad (5)$$

In this context, the original polytope P is known as the *primal polytope*. We notice that P^* is exactly the set of feasible vectors $\mathbf{c}$ for the dual linear program (DLP). This allows us to express the Fuchs Condition in an equivalent geometrical form.

Theorem 3: The solution $\mathbf{x}_0$ with $m = \|\mathbf{x}_0\|_0$ nonzeros in Theorem 1 is the unique optimum point of (LP) if and only if the polar polytope P^* has a $(d-m)$-dimensional face $F_{\text{opt}}^* = \{\mathbf{c} \in P^* | \tilde{\mathbf{A}}_{\text{opt}}^T \mathbf{c} = \mathbf{1}\}$.

Proof: For $0 \leq m < d$, the conditions in Theorem 1 are equivalent to the requirement for F_{opt}^* to exist and to be nondegenerate, i.e. to have exactly $d - m$ dimensions, so that a point $\mathbf{c}$ can exist in the relative interior of F_{opt}^*, $\mathbf{c} \in \text{relint } F_{\text{opt}}^*$. For $m = d$ the conditions are equivalent to $\mathbf{c}$ being exactly the vertex (0-face) $\mathbf{c} = (\tilde{\mathbf{A}}_{\text{opt}}^{-1})^T \mathbf{1}$. For the

converse, we additionally note that the requirement for the face F_{opt}^* to have exactly $d - m$ dimensions, m less than the dimensionality of its polytope P^*, means that $\tilde{\mathbf{A}}_{\text{opt}}$ must have full rank m. ∎

Now the $(d-m)$-face $F_{\text{opt}}^* = \{\mathbf{c} \in P^* | \tilde{\mathbf{A}}_{\text{opt}} \mathbf{c} = \mathbf{1}\}$ of the polar polytope P^* corresponds to the $(m-1)$-face $F_{\text{opt}} = P \cap \text{conv}\{\tilde{\mathbf{a}}_j \in \tilde{\mathbf{A}}_{\text{opt}}\}$ of the primal polytope P. We will call F_{opt}^* the *dual face*, and F_{opt} the corresponding *primal face*. Using this correspondence we have the following result, echoing a result of Donoho [10]:

Theorem 4: Let $\tilde{\mathbf{x}}_0 \geq \mathbf{0}$ with m nonzeros be a solution of $\tilde{\mathbf{A}}\tilde{\mathbf{x}} = \mathbf{y}$, and let $\tilde{\mathbf{x}}_{\text{opt}}$ and $\tilde{\mathbf{A}}_{\text{opt}}$ be constructed as before. Then $\tilde{\mathbf{x}}_0$ is the unique optimum point of (LP) if and only if $F_{\text{opt}} = \text{conv}\{\tilde{\mathbf{a}}_j \in \tilde{\mathbf{A}}_{\text{opt}}\}$ is an $(m-1)$-face of P and is a simplex.

Proof: It is a standard result for polar polytopes that a dual face F_{opt}^* exists and is nondegenerate if and only if the corresponding primal face F_{opt} exists and is a simplex [19]. Therefore Theorem 4 follows immediately from Theorem 3. ∎

Starting with the Fuchs condition in standard form (Theorem 1) we have found a number of alternative conditions for ℓ_1-unique optimality. Let us summarize these as follows.

Theorem 5: Suppose that we have solution $\mathbf{x}_0$ to $\mathbf{Ax} = \mathbf{y}$ with $m = \|\mathbf{x}_0\|_0$ nonzeros. Or equivalently: suppose we have a nonegative solution $\tilde{\mathbf{x}}_0 \geq \mathbf{0}$ to $\tilde{\mathbf{A}}\tilde{\mathbf{x}} = \mathbf{y}$ with $\tilde{\mathbf{A}} = [\mathbf{A}, -\mathbf{A}]$. Then the following conditions are equivalent:

1) $\mathbf{x}_0$ is the unique minimum of (P1);
2) $\tilde{\mathbf{x}}_0$ is the unique minimum of (LP);
3) the Fuchs Condition holds in the standard form (Theorem 1);
4) the Fuchs Condition holds in the original form (Theorem 2);
5) the dual face $F_{\text{opt}}^* = \{\mathbf{c} \in P^* \mid \tilde{\mathbf{A}}_{\text{opt}}^T \mathbf{c} = \mathbf{1}\} = \{\mathbf{c} \in P^* \mid \pm\mathbf{a}_i^T \mathbf{c} \leq 1, \mathbf{a}_i \in \mathbf{A}_{\text{opt}}\}$ exists and has $d - m$ dimensions (Theorem 3);
6) the primal face $F_{\text{opt}} = P \cap \text{conv}\{\tilde{\mathbf{a}}_j \in \tilde{\mathbf{A}}_{\text{opt}}\}$ exists and is an $(m-1)$-simplex (Theorem 4).

IV. GLOBAL OPTIMALITY AND EQUIVALENCE

We have so far considered the case of a single solution $\mathbf{x}_0$ to (P1), or equivalently a single solution $\tilde{\mathbf{x}}_0$ to (LP). Now let us consider equivalence for sets of points.

A. Discreteness and Covering of Sign and Support

In its standard form, we see that the Fuchs Condition (Theorem 1) only depends on $\tilde{\mathbf{A}}_{\text{opt}}$, or in original form (Theorem 2) on $\mathbf{A}_{\text{opt}}$ and the signs of $\mathbf{x}_{\text{opt}}$. Thus the following follows immediately.

Lemma 3 (Discreteness): The condition for $\tilde{\mathbf{x}}_0$ to be the unique minimum of (LP) depends only on the support of $\tilde{\mathbf{x}}_0$. Or equivalently: The condition for $\mathbf{x}_0$ to be the unique minimum of (P1) depends only on the support of $\mathbf{x}_0$ and signs of $\mathbf{x}_0$ on its support.

Proof: The support of $\tilde{\mathbf{x}}_0$ determines $\tilde{\mathbf{A}}_{\text{opt}}$ and hence both the rank of $\tilde{\mathbf{A}}_{\text{opt}}$ and existence of $\mathbf{c}$ in the Fuchs condition in

the standard form (Theorem 1). The support and signs of $\mathbf{x}_0$ determines the support of $\tilde{\mathbf{x}}_0$. ∎

As noted by Donoho, who also showed a similar result [10, Lemma 6.5], this 'discreteness of individual equivalence' has been observed by previous authors [5], [20]. It means for instance that if a particular $\mathbf{x}_0$ is the unique optimal solution to (P1) with $\mathbf{y} = \mathbf{A}\mathbf{x}_0$, then all solutions $\mathbf{x}_0'$ to (P1) with a different $\mathbf{y}' = \mathbf{A}\mathbf{x}_0'$ but the same support and signs as $\mathbf{x}_0$ and the same matrix $\mathbf{A}$ must also be the unique optimal solution of their respective problems.

Now let us say that a representation $\tilde{\mathbf{x}}$ *covers* $\tilde{\mathbf{x}}'$ if the support of $\tilde{\mathbf{x}}$ covers the support of $\tilde{\mathbf{x}}'$, i.e. that $\tilde{x}_i = 0 \implies \tilde{x}_i' = 0$. Then we have the following useful result about about 'sparser' representations.

Lemma 4 (Covering): Suppose $\tilde{\mathbf{x}}_0 \geq \mathbf{0}$ is the unique minimum of (LP). Then any solution $\tilde{\mathbf{x}}' \geq \mathbf{0}$ to $\tilde{\mathbf{A}}\tilde{\mathbf{x}}' = \mathbf{y}'$ is also ℓ_1-unique-optimal if $\tilde{\mathbf{x}}_0$ covers $\tilde{\mathbf{x}}'$. Or equivalently: suppose $\mathbf{x}_0$ is the unique minimum of (P1). Then any solution $\mathbf{x}'$ to $\mathbf{A}\mathbf{x}' = \mathbf{y}'$ is also ℓ_1-unique-optimal if $\mathbf{x}_0$ covers $\mathbf{x}'$ and the nonzero elements of $\mathbf{x}'$ have the same signs as the corresponding elements of $\mathbf{x}_0$.

Proof: The faces of a k-simplex are themselves simplical. Therefore the primal face F' of P corresponding to any $\tilde{\mathbf{x}}'$ must be a simplical face of F_{opt}, so by Theorem 4 $\tilde{\mathbf{x}}'$ is itself ℓ_1-unique-optimal. The equivalent version follows from the construction of $\tilde{\mathbf{x}}$ from $\mathbf{x}$. ∎

B. Global ℓ_1-Unique-Optimality

Now let us consider ℓ_1-unique-optimality of all representations $\mathbf{x}_0$ with at most k nonzeros. For this we will need the following definitions. The centrally-symmetric polytope $P = \mathrm{conv}\{\pm\mathbf{a}_i | \mathbf{a}_i \in \mathbf{A}\}$ is called k-*neighbourly* if every subset of k vertices of P, which does not contain two opposite vertices of P, are the vertices of a $(k-1)$-simplex which is a face of P. We also define the k-*rank* (Kruskal Rank) of a matrix as follows [21, p162]: A matrix $\mathbf{A}$ has k-rank of k_a if the columns are linearly independent in *every* set of k_a columns from $\mathbf{A}$, and if there is at least one set of $(k_a + 1)$ columns from $\mathbf{A}$ that includes linearly dependent columns. (For the *Spark* of a matrix, introduced by Donoho and Elad [7], we have $\mathrm{Spark}(\mathbf{A}) = k_a + 1$.) We are now in a position to give our version of the main result of Dohono [10, Theorem 1].

Theorem 6 (Donoho): Suppose we have an $n \times d$ matrix $\mathbf{A}$ with k-rank k_a, i.e. all sets of k_a columns of $\mathbf{A}$ are linearly independent. Then the primal polytope P has $2n$ vertices and is k-neighbourly, if and only if (a) $k_a \geq k$, and (b) every solution $\mathbf{x}_0$ to $\mathbf{y} = \mathbf{A}\mathbf{x}_0$ with at most k nonzeros is the unique solution to (P1).

Proof: For the 'if' direction, $k_a \geq k$ requires that each subset of k columns from $\mathbf{A}$ has full rank k. Therefore for each support pattern of k nonzero elements, solutions $\mathbf{x}_0$ to $\mathbf{y} = \mathbf{A}\mathbf{x}_0$ exist for each of the 2^k sign combinations. If all of the solutions are ℓ_1-unique-optimal, then each of the $\binom{n}{k} \times 2^k$ faces of P corresponding to the patterns of support and signs must be a $(k-1)$-simplex, so P is k-neighbourly and has n vertex pairs.

For the 'only if' direction, the $2n$ vertices of P correspond to the basis vectors pairs $\pm\mathbf{a}_j$. Since P is k-neighbourly, all $\binom{n}{k} \times 2^k$ ways we can choose a set of k basis vectors and signs $\sigma_j \in \{-1, +1\}$ correspond to a simplical face of P, so every solution $\mathbf{x}_0$ with k nonzeros must be ℓ_1-unique optimal. Consequently each subset of k columns from $\mathbf{A}$ has full rank k, so $k_a \geq k$. Finally, if P is k-neighbourly then it is also m-neighbourly for all $m = 0, \ldots, k$ [19] so this also holds for every solution $\mathbf{x}_0$ with $0 \leq m \leq k$ nonzeros. ∎

Remark This is expressed in a slightly stronger form than the original [10, Theorem 1], which included an assumption that the columns of $\mathbf{A}$ are in *general position*. Instead we have the somewhat weaker condition that $k_a \geq k$, i.e. $\mathrm{Spark}(\mathbf{A}) > k$, in one direction. Without this condition it might be possible for all possible solutions to be ℓ_1-unique-optimal without P having all $2n$ vertices, or without P having all $\binom{n}{k} \times 2^k$ primal faces. It would be interesting to see if it is possible to weaken this correspondence further, to polytopes P without all $2n$ vertices.

C. ℓ_1/ℓ_0 Equivalence

To show ℓ_1/ℓ_0 equivalence for a particular representation, the following result will be useful:

Lemma 5 (Donoho and Elad [7]: ℓ_0-Uniqueness): A representation $\mathbf{y} = \mathbf{A}\mathbf{x}_0$ with $m = \|\mathbf{x}_0\|_0$ nonzeros is ℓ_0-unique-optimal (i.e. the sparsest possible) if $m \leq k_a/2$, where k_a is the k-rank of $\mathbf{A}$.

Proof: See [7, Corollary 1] using $k_a = \mathrm{Spark}(\mathbf{A}) - 1$. ∎

We can therefore state the following.

Corollary 1: Suppose we have a solution $\mathbf{x}_0$ to $\mathbf{y} = \mathbf{A}\mathbf{x}_0$ with $m = \|\mathbf{x}_0\|_0$ nonzeros, and for which one of the conditions of Theorem 5 holds. Then $\mathbf{x}_0$ is both ℓ_0-unique-optimal and ℓ_1-unique-optimal if $m \leq k_a/2$, where k_a is the k-rank of $\mathbf{A}$.

Proof: This simply combines the conditions on m from Lemma 5 and Theorem 5. ∎

To show further that ℓ_1/ℓ_0 equivalence holds for a set of representations, it is sufficient for both ℓ_0-unique-optimality and ℓ_1-unique-optimality to hold for all representations $\mathbf{x}_0$ in that set.

Theorem 7: Suppose that P is k-neighbourly. Then ℓ_1/ℓ_0-equivalence holds for all solutions $\mathbf{x}_0$ to (P1) with $m = \|\mathbf{x}_0\|_0$ nonzeros, provided $m \leq \min(k, k_a/2)$.

Proof: This simply combines the conditions on m from Lemma 5 and Theorem 6. ∎

V. FUCHS COROLLARY

The Fuchs Condition (Theorems 1 and 2) concerns the existence of a point $\mathbf{c}$, but may not be easy to test. However, suppose we choose a specific vector $\mathbf{c}_{\mathrm{opt}} = \tilde{\mathbf{A}}_{\mathrm{opt}}^{\dagger}{}^T \mathbf{1} = \mathbf{A}_{\mathrm{opt}}^{\dagger}{}^T \mathrm{sign}\,\mathbf{x}_{\mathrm{opt}}$, which we call the *vertex* of our (signed) basis set $\tilde{\mathbf{A}}_{\mathrm{opt}}$ or $(\mathbf{A}_{\mathrm{opt}}^{\dagger}, \mathrm{sign}\,\mathbf{x}_{\mathrm{opt}})$. Then we obtain the following result:

Corollary 2 (Fuchs Corollary: Standard Form): For a desired solution $\tilde{\mathbf{x}}_0$ to $\tilde{\mathbf{A}}\tilde{\mathbf{x}} = \mathbf{y}$, let us construct $\tilde{\mathbf{x}}_{\mathrm{opt}}$ and $\tilde{\mathbf{A}}_{\mathrm{opt}}$ as in Theorem 1. If $\tilde{\mathbf{A}}_{\mathrm{opt}}$ has full rank and

$$\tilde{\mathbf{a}}_j^T \mathbf{c}_{\mathrm{opt}} < 1 \qquad \text{for all} \quad \tilde{\mathbf{a}}_j \in \tilde{\mathbf{A}}, \tilde{\mathbf{a}}_j \notin \tilde{\mathbf{A}}_{\mathrm{opt}} \qquad (6)$$

is satisfied with the specific dual vector $\mathbf{c}_{\mathrm{opt}} = \tilde{\mathbf{A}}_{\mathrm{opt}}^{\dagger}{}^T \mathbf{1}$, then $\tilde{\mathbf{x}}_0$ is the unique optimum to (DLP).

This result, a stronger condition than the Fuchs Condition (Theorem 1), was originally introduced by Fuchs [14] in the following form:

Corollary 3 (Fuchs Corollary: Original Form): Let $\mathbf{x}_{\mathrm{opt}}$ and $\mathbf{A}_{\mathrm{opt}}$ be given as in Theorem 2. If $\mathbf{A}_{\mathrm{opt}}$ is full rank, and

$$|\mathbf{a}_j^T \mathbf{A}_{\mathrm{opt}}^{\dagger}{}^T \operatorname{sign} \mathbf{x}_{\mathrm{opt}}| < 1 \qquad \text{for any } \mathbf{a}_j \in \mathbf{A}, \, \mathbf{a}_j \notin \mathbf{A}_{\mathrm{opt}} \tag{7}$$

then $\mathbf{x}_0$ is the unique solution to (P1).

While the Fuchs Corollary is a stronger condition than necessary, in that there are unique optima for which the conditions on $\mathbf{c}_{\mathrm{opt}}$ are not satisfied (see [13]) its advantage over the Fuchs Condition is that it is easier to test. While testing the Fuchs Condition would require a search for the relevant face of P^* to be conducted, the probe point (basis vertex) $\mathbf{c}_{\mathrm{opt}}$ can be constructed directly from $\mathbf{x}_0$ and $\mathbf{A}$.

In terms of the polar polytope geometry, the Fuchs Corollary requires both that the dual face $F_{\mathrm{opt}}^* = \{\mathbf{c} \in P^* | \tilde{\mathbf{A}}_{\mathrm{opt}} \mathbf{c} = \mathbf{1}\}$ exists and has dimension $d - m$, as for the Fuchs Condition, and also that the basis vertex $\mathbf{c}_{\mathrm{opt}} = \tilde{\mathbf{A}}_{\mathrm{opt}}^{\dagger}{}^T \mathbf{1}$ is contained in its relative interior, $\mathbf{c}_{\mathrm{opt}} \in \operatorname{relint} F_{\mathrm{opt}}^*$.

VI. EXACT RECOVERY CONDITION

Perhaps more well known than the Fuchs Condition is the *Exact Recovery Condition* (ERC) introduced by Tropp [9].

Theorem 8 (Tropp [9]: Exact Recovery Condition): Let us have $\mathbf{x}_0$ and $\mathbf{A}_{\mathrm{opt}}$ as in Theorem 2. If

$$\max_{\mathbf{a}_j \notin \mathbf{A}_{\mathrm{opt}}} \left\| \mathbf{A}_{\mathrm{opt}}^{\dagger} \mathbf{a}_j \right\|_1 < 1 \tag{8}$$

where $\mathbf{a}_j$ ranges over the atoms in $\mathbf{A}$ which are not in the m-term representation of $\mathbf{y}$, then $\mathbf{x}_0$ is the unique solution to (P1).

The quantity $\max_{\mathbf{a}_j \notin \mathbf{A}_{\mathrm{opt}}} \left\| \mathbf{A}_{\mathrm{opt}}^{\dagger} \mathbf{a}_j \right\|_1$ is referred to as the *exact recovery coefficient*.

Although the approaches of Fuchs [13] and Tropp [9] are very different, Gribonval and Nielsen [17] pointed that they are closely linked. Specifically we have

$$\max_{\mathbf{x}_{\mathrm{opt}}} \max_{\mathbf{a}_j \notin \mathbf{A}_{\mathrm{opt}}} | \operatorname{sign}(\mathbf{x}_{\mathrm{opt}}^T) \mathbf{A}_{\mathrm{opt}}^{\dagger} \mathbf{a}_j |$$
$$= \max_{\mathbf{x}_{\mathrm{opt}}} \max_{\mathbf{a}_j \notin \mathbf{A}_{\mathrm{opt}}} | \langle \operatorname{sign}(\mathbf{x}_{\mathrm{opt}}), \mathbf{A}_{\mathrm{opt}}^{\dagger} \mathbf{a}_j \rangle |$$
$$= \max_{\mathbf{a}_j \notin \mathbf{A}_{\mathrm{opt}}} \left\| \mathbf{A}_{\mathrm{opt}}^{\dagger} \mathbf{a}_j \right\|_1 \tag{9}$$

so the Exact Recovery Condition (Theorem 8) is itself a corollary of the Fuchs Corollary (Corollary 3). Thus ERC is a stronger condition than the Fuchs Condition (Theorem 2), and it is also a stronger condition than that in the Fuchs Corollary (Corollary 3), which we can see as follows:

Lemma 6: Suppose we have a desired solution $\mathbf{x}_0$ to $\mathbf{y} = \mathbf{A}\mathbf{x}_0$. Then the Exact Recovery Condition (Theorem 8) is satisfied if the Fuchs Corollary (Corollary 3) is satisfied for all $\mathbf{x}_0'$ with the same support as $\mathbf{x}_0$, including solutions $\mathbf{x}_0'$ with the same support but different signs.

Proof: This follows from (9) (see [17]). ∎

We shall now develop an interpretation of ERC in terms of our polytope geometry. Let us construct the polytope $P_{\mathrm{opt}} = \operatorname{conv}\{\pm\mathbf{a}_i \mid \mathbf{a}_i \in \mathbf{A}_{\mathrm{opt}}\}$ which we shall call the *primal basis polytope*, and its corresponding relative polar polytope $P_{\mathrm{opt}}^* = \{\mathbf{c} \in \operatorname{aff} P_{\mathrm{opt}} \mid \pm\mathbf{c}^T\mathbf{a}_i \leq 1, \mathbf{a}_i \in \mathbf{A}_{\mathrm{opt}}\}$ where $\operatorname{aff} P_{\mathrm{opt}}$ is the affine hull of P_{opt}; we call P_{opt}^* the *dual basis polytope*. We then obtain the following result:

Theorem 9: Suppose that $\mathbf{A}_{\mathrm{opt}}$ has full rank m. Then the Exact Recovery Condition, Equation (8), is satisfied if and only if (a) $P_{\mathrm{opt}}^* \subset P^*$, i.e. the dual basis polytope P_{opt}^* is contained within the complete polar polytope P^*, and (b) P_{opt}^* does not touch any face of P^* for which $\pm\mathbf{a}_j^T\mathbf{c} = 1$ for some $\mathbf{a}_j \notin \mathbf{A}_{\mathrm{opt}}$.

Proof: We construct the set of m-dimensional sign vectors $\sigma = [\sigma_1, \ldots, \sigma_m]^T \in \{+1, -1\}^m$ and the set of basis vertices $V_{\mathrm{opt}}^* = \{\mathbf{c} = \mathbf{A}_{\mathrm{opt}}^{\dagger}{}^T \sigma\}$. Equation (8) is equivalent to the condition $\mathbf{a}_j^T\mathbf{c} < 1$ for all $\mathbf{c} \in V_{\mathrm{opt}}^*, \mathbf{a}_j \notin \mathbf{A}_{\mathrm{opt}}$. Now $\mathbf{c} = \mathbf{A}_{\mathrm{opt}}^{\dagger}{}^T \sigma$ is the vector in the span of the columns of $\mathbf{A}_{\mathrm{opt}}$ which satisfies $\mathbf{A}_{\mathrm{opt}}^T\mathbf{c} = \sigma$ i.e. $\operatorname{diag}(\sigma)\mathbf{A}_{\mathrm{opt}}^T\mathbf{c} = \mathbf{1}$, or in other words $\pm_i\mathbf{a}_i^T\mathbf{c} = 1$ for $\mathbf{a}_i \in \mathbf{A}_{\mathrm{opt}}$ and some combination of signs $\pm_i$. Hence V_{opt}^* is actually the set of 2^m vertices of the dual basis polytope P_{opt}^*. Thus the condition $P_{\mathrm{opt}}^* \subset P^*$ corresponds to a non-strict inequality in Equation (8). The remaining condition for P_{opt}^* not to touch any face of P^* for which $\pm\mathbf{a}_j^T\mathbf{c} = 1$ for some $\mathbf{a}_j \notin \mathbf{A}_{\mathrm{opt}}$ corresponds to exclusion of the equality, to leave a strict inequality. ∎

In some ways, the Exact Recovery Condition is not as well behaved as the Fuchs Condition. For example, we have the following somewhat surprising result:

Theorem 10 (Non-nestedness of ERC): Satisfying ERC for all solutions $\mathbf{x}_0$ with m nonzeros is not sufficient to ensure that all solutions $\mathbf{x}_0$ with $k < m$ nonzeros also satisfy ERC.

Proof: Let $\mathbf{A} = [\mathbf{a}_1, \mathbf{a}_2]$ with $\mathbf{a}_1 = [1,0]^T$ and $\mathbf{a}_2 = [\sqrt{2}, \sqrt{2}]^T$. ERC is trivially satisfied for $m = 2$ since $\mathbf{A} = \mathbf{A}_{\mathrm{opt}}$ so there are no $\mathbf{A} \ni \mathbf{a}_j \notin \mathbf{A}_{\mathrm{opt}}$. Now for $\mathbf{x}_0 = [\beta, 0]^T$ with $k = 1$, we have $\mathbf{A}_{\mathrm{opt}} = \mathbf{a}_1 = [1,0]^T$ and $\mathbf{A}_{\mathrm{opt}}^{\dagger} = \mathbf{a}_1^T = [1,0]$, so $\mathbf{A}_{\mathrm{opt}}^{\dagger}\mathbf{a}_2 = \sqrt{2} > 1$ and hence ERC fails for $k = 1$. Examples can be also be constructed for unit-norm dictionaries (e.g. see Section VIII-B). ∎

This result is in contrast to the Fuchs Condition, where Lemma 4 tells us that if the Fuchs Condition is satisfied for all $\mathbf{x}_0$ with m nonzeros, then it will be satisfied for any $\mathbf{x}_0$ with $k < m$ nonzeros [10].

VII. ERC AND MATCHING PURSUIT ALGORITHMS

We have seen that Tropp's ERC is sufficient but not necessary for ℓ_1-unique-optimality. However, Tropp [9] also showed that the Exact Recovery Condition (8) is necessary and sufficient for the Orthogonal Matching Pursuit (OMP) algorithm [22] to find the ℓ_1-unique-optimal solution in the following sense:

Theorem 11 (Tropp [9]: Exact Recovery for OMP): Suppose we have a desired solution $\mathbf{x}_0$ for $\mathbf{y} = \mathbf{A}\mathbf{x}_0$ with full rank $\mathbf{A}_{\mathrm{opt}}$ as in Theorem 8. Then Orthogonal Matching Pursuit (OMP) will recover $\mathbf{x}_0$ in m steps if the Exact Recovery Condition (8) holds. Conversely, suppose ERC fails for some $\mathbf{y} = \mathbf{A}\mathbf{x}_0$ with optimal synthesis matrix $\mathbf{A}_{\mathrm{opt}}$.

Then there are signals in the column span of $\mathbf{A}_{\mathrm{opt}}$ which Orthogonal Matching Pursuit cannot recover *in m steps*.

Proof: See [9]. ∎

The condition for recovery 'in m steps' is implicit in Tropp's statement of this theorem, but is used in the proof of the converse direction. To show that the 'm steps' condition is necessary we can give the following counter-result:

Lemma 7: There exist problems (P1) with solutions $\mathbf{x}_0$ with m nonzeros for which ERC fails, but which OMP can recover in $k > m$ steps, provided that OMP is eventually allowed to drop any zeros in the final representation.

Proof: Let us use the same example as in the proof of Theorem 10, and suppose we wish to recover the signal $\mathbf{x}_0 = [1, 0]^T$ from $\mathbf{y} = \mathbf{A}\mathbf{x}_0 = \mathbf{a}_1 = [1, 0]^T$, for which $\mathbf{A}_{\mathrm{opt}} = [\mathbf{a}_1]$. From the proof of Theorem 10 we know that ERC fails for this $\mathbf{x}_0$. Now let us run OMP [22]. In step 1 we have $\mathbf{a}_2^T\mathbf{y} = \sqrt{2} > 1 = \mathbf{a}_1^T\mathbf{y}$, so OMP chooses the wrong atom $\mathbf{a}_2$, yielding a basis set after 1 step of $\mathbf{A}^{(1)} = [\mathbf{a}_2]$. Choosing x_2 to minimize the mean squared error we get $\mathbf{x}^{(1)} = \mathbf{a}_2^+\mathbf{y} = [1/(2\sqrt{2})]$ producing a reconstruction $\hat{\mathbf{y}}^{(1)} = \mathbf{x}^{(1)}\mathbf{a}_2 = (1/(2\sqrt{2})) \times [\sqrt{2}, \sqrt{2}]^T = [0.5, 0.5]$ and residual $\mathbf{r}^{(1)} = \mathbf{y} - \hat{\mathbf{y}}^{(1)} = [0.5, -0.5]^T \neq \mathbf{0}$. So as expected, OMP has not recovered $x_0 = [1, 0]^T$ in $m = 1$ steps.

But if we allow OMP to run for a second step, we find $\mathbf{a}_1^T\mathbf{r}^{(1)} = 0.5$ while $\mathbf{a}_2^T\mathbf{r}^{(1)} = 0$ as we would expect for OMP. Hence in step 2, OMP chooses the remaining basis $\mathbf{a}_1$ so $\mathbf{A}^{(2)} = [\mathbf{a}_1, \mathbf{a}_2]$ (reordering the atoms for convenience). Now choosing $\mathbf{x} = [x_1, x_2]$ to minimize the mean squared error we get $\mathbf{x}^{(2)} = [x_1^{(2)}, x_2^{(2)}] = (\mathbf{A}^\dagger)^{(2)}\mathbf{y} = [1, 0]^T$ producing a reconstruction $\hat{\mathbf{y}}^{(2)} = \mathbf{x}^{(2)}\mathbf{A}^{(2)} = \mathbf{y}_0$ and $\mathbf{r} = 0$. Since $x_2^{(2)} = 0$, OMP has found the correct 1-term reconstruction of $\mathbf{y}$, albeit taking 2 steps to do so. The same result holds for scaled and negated versions of $\mathbf{A}_{\mathrm{opt}} = [\mathbf{a}_1]$. ∎

Thus failure of ERC does not require that OMP will fail, only that there are cases for which it cannot succeed in m steps. We can therefore state the following condition, weaker than the Exact Recovery Condon, for possible *delayed recovery* by OMP.

Corollary 4: Suppose that $\mathbf{x}_0$ with m_0 nonzeros is a desired solution of $\mathbf{y}_0 = \mathbf{A}\mathbf{x}_0$ which fails ERC. Suppose further that there exists a different solution $\mathbf{y}_1 = \mathbf{A}\mathbf{x}_1$ for which ERC is satisfied, and which *covers* $\mathbf{x}_0$ in the sense that the support of $\mathbf{x}_1$ is a superset of the support of $\mathbf{x}_0$. Then OMP will 'eventually' recover $\mathbf{x}_0$ in m_1 steps, where $m_1 > m_0$ is the number of nonzeros in $\mathbf{x}_1$

Proof: This follows from the proof of Theorem 11, but allowing some of the elements of $\mathbf{x}_1$ to be zero to match the zeros of the desired solution $\mathbf{x}_0$. ∎

At present it is unclear whether it is common for ERC to fail at one level m_0 but be satisfied at higher levels $m_1 > m_0$, so it remains to be seen whether this concept of *delayed recovery* by OMP will turn out to be useful.

VIII. ILLUSTRATIVE GRAPHICAL EXAMPLES

One of the advantages of the polytope geometry we have considered here is that it can give us some insight into the various optimality conditions we are interested in, and the
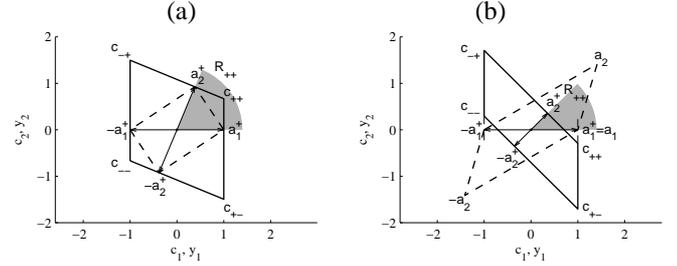

Fig. 1. Primal polytope P (dotted) and polar polytope P^* (solid) for (a) a unit-norm dictionary and (b) a non-unit-norm atom set. The shaded region R_{++} denotes a cone in $\mathbf{y}$-space represented by nonnegative amounts $x_1, x_2 \geq 0$ of the basis vectors $+\mathbf{a}_1, +\mathbf{a}_2$.

differences between them. In this section we will give 2-dimensional examples for a unit-norm dictionary and non-unit-norm atom set, and a 3-dimensional example for a unit-norm dictionary.

A. Two Dimensional Examples

Fig. 1 shows an example of (a) a unit-norm dictionary and (b) a basis set with one unit-norm atom and one non-unit-norm atom. The vectors $\pm\mathbf{a}_i^+$ are scaled versions of the atoms defined by $\pm\mathbf{a}_i^+ = \pm\mathbf{a}_i/\|\mathbf{a}_i\|_2^2$. We notice that the $\mathbf{a}_i^+$ touch the supporting hyperplanes of the dual polytope P^* since $\mathbf{a}_i^T\mathbf{a}_i^+ = \mathbf{a}_i^T\mathbf{a}_i/\|\mathbf{a}_i\|_2^2 = 1$.

Suppose that we have a 'sparse' representation $\mathbf{x}_0 = [\beta, 0]^T$ for some $\beta > 0$, with $m = \|\mathbf{x}_0\|_0 = 1$. In each of Figs. 1(a) and (b) the face $F_{\mathrm{opt}}^* = \{\mathbf{c} \in P^*|\tilde{\mathbf{A}}_{\mathrm{opt}}^T\mathbf{c} = \mathbf{1}\} \subset P^*$ exists: specifically F_{opt}^* is the line joining $\mathbf{c}_{++}$ to $\mathbf{c}_{+-}$, and we can see this is a face with $d - m = 2 - 1 = 1$ dimension. Any point $\mathbf{c}$ on the line joining $\mathbf{c}_{++}$ to $\mathbf{c}_{+-}$, excluding the end points $\mathbf{c}_{++}$ and $\mathbf{c}_{+-}$ themselves, will correspond to $\mathbf{c} \in \mathrm{relint}\, F_{\mathrm{opt}}^*$. Therefore this point will satisfy condition (b) in the Fuchs Condition (Theorems 1 and 2). Thus in each of Figs. 1(a) and (b), if the point $\mathbf{x}_0 = [\beta, 0]^T$ is a feasible solution to (P1), then it is ℓ_1-unique-optimal. To confirm this, for the primal face F_{opt}, in each of Figs. 1(a) and (b) we have the single-point set $F_{\mathrm{opt}} = \mathrm{conv}\{\mathbf{a}_1\} = \{\mathbf{a}_1\} \subset P$, which also exists and is has $(m - 1) = 0$ dimensions, as expected.

For the Fuchs Corollary, in each of Figs. 1(a) and (b) we have $\tilde{\mathbf{A}}_{\mathrm{opt}} = [\mathbf{a}_1]$ and hence $\tilde{\mathbf{A}}_{\mathrm{opt}}^{\dagger T} = [+\mathbf{a}_1^+]$ so our basis vertex is given by $\mathbf{c}_{\mathrm{opt}} = \tilde{\mathbf{A}}_{\mathrm{opt}}^{\dagger T}\mathbf{1} = +\mathbf{a}_1^+ \cdot 1 = \mathbf{a}_1^+$. Since $F_{\mathrm{opt}}^* = \mathrm{conv}\{\mathbf{c}_{++}, \mathbf{c}_{+-}\}$ which is the line segment joining $\mathbf{c}_{++}$ to $\mathbf{c}_{+-}$, clearly $\mathbf{c}_{\mathrm{opt}} \in \mathrm{relint}\, F_{\mathrm{opt}}^*$ in Fig. 1(a), but $\mathbf{c}_{\mathrm{opt}} \notin \mathrm{relint}\, F_{\mathrm{opt}}^*$ in Fig. 1(b). Therefore, while the Fuchs Condition (Theorems 1 and 2) is satisfied for $\mathbf{x}_0 = [\beta, 0]^T$ in both Fig. 1(a) and (b), the Fuchs Corollary is only satisfied for this $\mathbf{x}_0$ in Fig. 1(a), but not in Fig. 1(b). This confirms that the Fuchs Corollary is indeed strictly stronger than the Fuchs Condition (see also [13]).

For the Exact Recovery Condition, our primal basis polytope is $P_{\mathrm{opt}} = \mathrm{conv}\{-\mathbf{a}_1, +\mathbf{a}_1\}$. In this case we get $P_{\mathrm{opt}}^* = \mathrm{conv}\{-\mathbf{a}_1^+, +\mathbf{a}_1^+\}$ so P_{opt}^* is the line segment joining $-\mathbf{a}_1^+$ and $+\mathbf{a}_1^+$. In Fig. 1(a) we can see that $P_{\mathrm{opt}}^* \subset P^*$ and P_{opt}^* is well away from the faces along $+\mathbf{a}_2^T\mathbf{c} = 1$ (joining $\mathbf{c}_{-+}$ to $\mathbf{c}_{++}$) and $-\mathbf{a}_2^T\mathbf{c} = 1$ (joining $\mathbf{c}_{--}$ to $\mathbf{c}_{+-}$). Hence
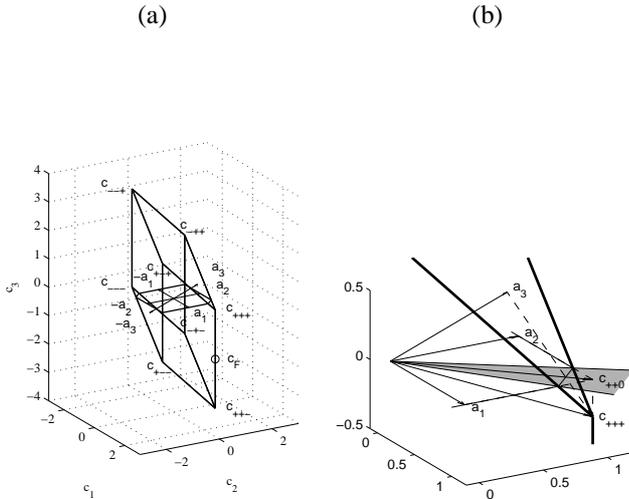
(a)   (b)

Fig. 2. Failure of ERC for unit norm vectors in $d = 3$ dimensions, showing (a) the complete polar polytope, with section line through the 'waist' at $y_3 = 0$, and (b) the magnified section showing the atoms $\mathbf{a}_1$, $\mathbf{a}_2$ and $\mathbf{a}_3$ ($= \mathbf{a}_1^+, \mathbf{a}_2^+$, and $\mathbf{a}_3^+$ since $|a_i| = 1$).

ERC is satisfied in Fig. 1(a). However, in Fig. 1(b) we can see that $P_{\text{opt}}^* \not\subset P^*$ so ERC is not satisfied. If we repeat this analysis for some $\mathbf{x}_0$ with $\mathbf{A}_{\text{opt}} = [\mathbf{a}_2]$, we see that $P_{\text{opt}}^* = \text{conv}\{-\mathbf{a}_2^+, +\mathbf{a}_2^+\}$ so $P_{\text{opt}}^* \subset P^*$, and P_{opt}^* is away from the other faces, in both Fig. 1(a) and (b), and hence ERC is satisfied for both. Similarly for some $\mathbf{x}_0$ with $\mathbf{A}_{\text{opt}} = [\mathbf{a}_1, \mathbf{a}_2]$, we now have $P_{\text{opt}}^* = P^*$ so clearly $P_{\text{opt}}^* \subset P^*$, and there are no $\mathbf{a}_j \notin \mathbf{A}_{\text{opt}}$. Hence ERC is satisfied for $m = 2$ in both Fig. 1(a) and (b).

B. Unit-norm Dictionaries: A Three-dimensional Example

Many of the equivalence results of previous authors are for dictionaries of unit norm atoms $|\mathbf{a}_i| = 1$. The special properties of unit-norm dictionaries mean that it is more awkward to find low-dimensional examples to illustrate the distinction between Fuchs Condition, Fuchs Corollary and ERC. Nevertheless, we can illustrate many of these issues using the dictionary matrix $\mathbf{A} = [\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3]$ with the unit norm atoms given by $\mathbf{a}_1 = [1, 0, 0]^T$, $\mathbf{a}_2 = [0, 1, 0]^T$, $\mathbf{a}_3 = (1/\sqrt{3})[1, 1, 1]^T$. In particular, suppose that our desired vector to recover is $\mathbf{x}_0 = [1, 1, 0]^T$ so that $\mathbf{y} = \mathbf{A}\mathbf{x}_0 = \mathbf{a}_1 + \mathbf{a}_2$. Therefore the optimal basis set that we would like to recover given $\mathbf{y}$ is $\mathbf{A}_{\text{opt}} = [\mathbf{a}_1, \mathbf{a}_2]$, which has vertex $\mathbf{c}_{\text{opt}} = \mathbf{A}_{\text{opt}}^{\dagger}{}^T \mathbf{1} = [1, 1, 0]^T$.

For the Fuchs Condition, consider the point $\mathbf{c}_F = [1, 1, -2]^T$ marked in Fig. 2(a). We can verify that $\mathbf{c}_F^T \mathbf{a}_1 = \mathbf{c}_F^T \mathbf{a}_2 = 1$, and $|\mathbf{c}_F^T \mathbf{a}_3| = |(1 + 1 - 2)/\sqrt{3}| = 0 < 1$ therefore the Fuchs Condition is satisfied. In fact the relevant dual face is $F_{\text{opt}}^* = \text{conv}\{\mathbf{c}_{+++}, \mathbf{c}_{++-}\}$ so any $\mathbf{c} \in \text{relint } F_{\text{opt}}^*$, i.e. anywhere along the line segment strictly between $\mathbf{c}_{+++}$ and $\mathbf{c}_{++-}$, will be suitable to satisfy the Fuchs Condition.

Considering the Fuchs Corollary, this requires $\mathbf{c}_{\text{opt}} = \mathbf{A}_{\text{opt}}^{\dagger}{}^T \mathbf{1} = \mathbf{c}_{++0}$ to be contained in F_{opt}^*. However, Fig. 2(b) shows that $\mathbf{c}_{++0} \notin P^*$, so $\mathbf{c}_{++0} \notin F_{\text{opt}}^*$ since $F_{\text{opt}}^* \subset P^*$ is itself a face of P^*. Therefore the Fuchs Corollary is not satisfied.

For the Exact Recovery Condition, we require that $\|\mathbf{A}_{\text{opt}}^{\dagger} \mathbf{a}_3\|_1 < 1$ so we must have e.g. $\mathbf{c}_{\text{opt}}^T \mathbf{a}_3 < 1$. However, calculation gives $\mathbf{c}_{\text{opt}}^T \mathbf{a}_3 = 2/\sqrt{3} > 1$ so ERC fails for this basis. The shaded cone in Fig. 2(b) shows the segment of the plane spanned by $\{\mathbf{a}_1, \mathbf{a}_2\}$ for which $\mathbf{a}_3^T \mathbf{c} > \max_{i=1,2} \mathbf{a}_i^T \mathbf{c}$. Here we see that the vertex $\mathbf{c}_{\text{opt}} = \mathbf{c}_{++0}$ is in this shaded region (Fig. 2(b)), and has been 'cut off' by the halfspace $\mathbf{a}_3^T \mathbf{c} \leq 1$. As confirmation of this, the dual basis polytope P_{opt}^* is the square in the plane $x_3 = 0$ with vertices at $[\pm 1, \pm 1, 0]$. We can see that the corner containing $[1, 1, 0]$ ($= \mathbf{c}_{++0}$) is not contained within the full dual polytope, so $P_{\text{opt}}^* \not\subset P^*$, and hence ERC is not satisfied.

To summarize, since the Fuchs Condition is satisfied, any desired solution $x_0 = [\beta_1, \beta_2, 0]^T$ with $\beta_1, \beta_2 > 0$ in this example will be recovered by Basis Pursuit, even though the Fuchs Corollary and ERC both fail. Note however that the solution $x_0 = [\beta_1, -\beta_2, 0]$ with $\beta_1, \beta_2 > 0$, with same support but different signs would satisfy both the Fuchs Condition and the Fuchs Corollary, while ERC must still fail since the support of the desired solution is unchanged.

IX. CONCLUSIONS

We have explored the geometry of the sparse representation problem using centrally-symmetric polytopes and polar (dual) polytopes. We have seen that polytopes can give us a useful insight into the optimality conditions introduced by Fuchs, for example, which had previously been considered to be difficult to interpret.

In exploring this geometry we have also been able to tighten some of these previous results, and link these to the polytope-based results of Donoho for the primal polytope. For example, we showed that the Fuchs Condition is both necessary and sufficient for ℓ_1-unique-optimality, and that there are situations where Orthogonal Matching Pursuit (OMP) can find all ℓ_1-unique-optimal solutions with m nonzeros, even if the Exact Recovery Condition (ERC) fails for m, if it is allowed to run for additional steps.

X. ACKNOWLEDGEMENTS

REFERENCES

[1] S. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Transactions on Signal Processing*, vol. 41, no. 12, pp. 3397–3415, 1993.

[2] X. Huo and J. Chen, "Hardness on numerical realization of some penalized likelihood estimators," Industrial and Technology Statistics Group, Georgia Institute of Technology, Technical Paper 08/2006, 2006.

[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] 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.

[5] D. L. Donoho and X. Huo, "Uncertainty principles and ideal atomic decomposition," *IEEE Transactions on Information Theory*, vol. 47, no. 7, pp. 2845–2862, November 2001.

[6] M. Elad and A. M. Bruckstein, "A generalized uncertainty principle and sparse representation in pairs of bases," *IEEE Transactions on Information Theory*, vol. 48, no. 9, pp. 2558–2567, September 2002.

[7] D. L. Donoho and M. Elad, "Optimally sparse representation in general (nonorthogonal) dictionaries via ℓ^1 minimization," *Proc. Nat. Aca. Sci.*, vol. 100, pp. 2197–2202, March 2003.

[8] R. Gribonval and M. Nielsen, "Sparse representations in unions of bases," *IEEE Transactions on Information Theory*, vol. 49, no. 12, pp. 3320–3325, December 2003.

[9] J. A. Tropp, "Greed is good: Algorithmic results for sparse approximation," *IEEE Transactions on Information Theory*, vol. 50, no. 10, pp. 2231–2242, Oct. 2004.

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

[11] ——, "High-dimensional centrosymmetric polytopes with neighborliness proportional to dimension," Statistics Department, Stanford University, Tech. Rep., January 2005.

[12] P. McMullen and G. C. Shephard, "Diagrams for centrally symmetric polytopes," *Mathematika*, vol. 15, pp. 123–138, 1968.

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

[14] ——, "Detection and estimation of superimposed signals," in *Proceedings of the 1998 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '98)*, vol. 3, 12-15 May 1998, pp. 1649 – 1652 vol.3.

[15] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.

[16] A. Goldman and A. Tucker, "Linear inequalities and related systems," in *Theory of Linear Programming*, H. Kuhn and A. Tucker, Eds. Princeton, New Jersey: Princeton University Press, 1956, pp. 53–97.

[17] R. Gribonval and M. Nielsen, "Approximation with highly redundant dictionaries," in *Wavelets: Applications in Signal and Image Processing, Proc. SPIE'03, San Diego, USA*, August 2003, pp. 216–227.

[18] J. A. Tropp, "Recovery of short, complex linear combinations via ℓ_1 minimization," *IEEE Transactions on Information Theory*, vol. 51, no. 4, pp. 1568–1570, April 2005.

[19] B. Grünbaum, *Convex Polytopes*, 2nd ed., ser. Graduate Texts in Mathematics 221. New York: Springer-Verlag, 2003.

[20] D. M. Malioutov, M. Cetin, and A. Willsky, "Optimal sparse representations in general overcomplete bases," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '04)*, vol. 2, 17-21 May 2004, pp. II–793–796.

[21] R. A. Harshman and M. E. Lundy, "The PARAFAC model for three-way factor analysis and multidimensional scaling," in *Research Methods for Multimode Data Analysis*, H. G. Law, J. C. W. Snyder, J. Hattie, and R. P. McDonald, Eds. New York: Praeger, 1984, pp. 122–215.

[22] 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.

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

Mark D. Plumbley (S'88–M'90) received the Ph.D. degree in neural networks as a Research Student in the Engineering Department, Cambridge University, Cambridge, U.K., in 1991. Following the Ph.D. degree, he joined Kings College London, London, U.K., in 1991, and in 2002 moved to Queen Mary University, London, to help establish the new Centre for Digital Music. His research interests include the analysis of musical audio, such as automatic music transcription, beat tracking, and audio source separation, using techniques such as independent component analysis (ICA) and sparse representations.