NP/CMP Equivalence: A Phenomenon Hidden Among Sparsity Models l_0 Minimization and l_p Minimization for Information Processing

Jigen Peng, Shigang Yue, *Member, IEEE*, and Haiyang Li

***Abstract*— In this paper, we have proved that in every underdetermined linear system $Ax = b$, there corresponds a constant $p^*(A, b) > 0$ such that every solution to the l_p-norm minimization problem also solves the l_0-norm minimization problem whenever $0 < p < p^*(A, b)$. This phenomenon is named NP/CMP equivalence.**

***Index Terms*— Information processing, sparse representation, sparse recovery, l_p minimization, underdetermined linear system.**

I. Introduction

IN SPARSE information processing, the following minimization is commonly employed to model basic sparse problems such as sparse representation and sparse recovery,

$$(P_0) \qquad \min_x \|x\|_0 \quad \text{subject to} \quad Ax = b \tag{1}$$

where A is a real matrix of order $m \times n$ with $m < n$, b is a nonzero real vector of m-dimension, and $\|x\|_0$ is the so-called l_0-norm of real vector x, which counts the number of the non-zero entries in x [3], [13], [26]. Unfortunately, although the l_0-norm characterizes the sparsity of the vector x, the optimization problem (P_0) is actually NP-Hard because of the discrete and discontinuous nature of the l_0-norm. This has resulted in many substitution models for (P_0), where $\|x\|_0$ is replaced with functions that evaluate the desirability of a would-be solution to $Ax = b$ (see [4], [11], [17], [20], [26], and references therein). Because of the relationship

$$\|x\|_0 = \lim_{p\to 0^+} \sum_{i=1}^{n} |x_i|^p = \lim_{p\to 0^+} \|x\|_p^p, \quad \forall x = (x_1, x_2, \cdots, x_n)^T, \tag{2}$$

Manuscript received April 1, 2014; revised February 3, 2015; accepted April 7, 2015. Date of publication May 5, 2015; date of current version June 12, 2015. This work was supported in part by NSFC under Contact 11131006 and Contact 11271297, in part by EU FP7 EYE2E Project under Grant 269118 and LIVCODE Project under Grant 295151, and in part by the National Basic Research Program of China under Contact 2013CB329404. *(Corresponding author: Shigang Yue.)*

J. Peng is with the School of Mathematics and Statistics, Xi'an Jiaotong University, Xi'an 710049, China, and also with the Beijing Center for Mathematics and Information Interdisciplinary Sciences, Beijing 100048, China (e-mail: jgpeng@mail.xjtu.edu.cn).

S. Yue is with the School of Computer Science, University of Lincoln, Lincoln LN6 7TS, UK (e-mail: syue@lincoln.ac.uk).

H. Li is with the School of Science, Xi'an Polytechnic University, Xi'an 710048, China (e-mail: fplihaiyang@126.com).

Communicated by O. Milenkovic, Associate Editor for Coding Theory.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIT.2015.2429611

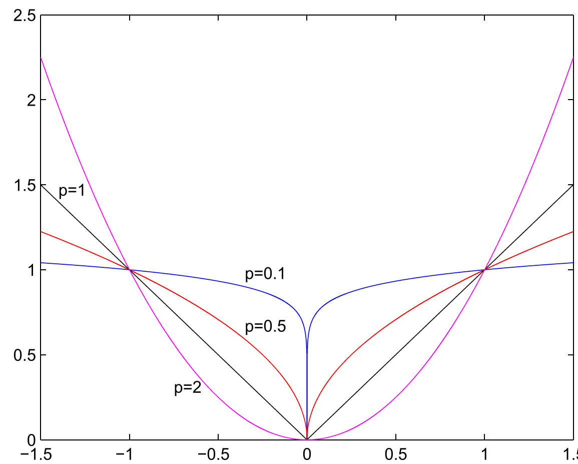

Fig. 1. The behavior of l_p-norm for various values of p. See [3, p. 38].

the following minimizations seem to be the most natural choice,

$$(P_p) \qquad \min_x \|x\|_p^p \quad \text{subject to} \quad Ax = b \tag{3}$$

where $0 < p \leq 1$. Key work by Dohoho and Huo [10], and Candes and Tao [5] on $p = 1$, and by Gribonval and Nielsen [19] on $0 < p < 1$, has resulted in the optimization models described above gaining in popularity in the literature (see [7], [9], [14], [21], [23], [25], [27], [28]). However, there remains a key issue with respect to these choices: to what extent the minimizations (P_p) can achieve the same result as the initial minimization (P_0). A lot of excellent theoretical work (see [4], [10], [12], [19]), together with some empirical evidence (see [6]), has shown that, provided some conditions are met, such as assuming the restricted isometric property (RIP), the l_1-norm minimization (P_1) can really make an exact recovery. The original notion of RIP has received much attention and has already been tailored to a more general case where $0 < p < 1$ (see [7], [9], [20]). Work undertaken by Donoho and Tanner in [11] using convex geometry demonstrated a surprising phenomenon that for any real matrix A, whenever the nonnegative solution to (P_0) is sufficiently sparse, it is also a unique solution to (P_1). That is, there exists a certain equivalence between (P_0) and (P_1). As the former is discrete and so NP-Hard, and the latter is continuous and equivalent to a linear programming (LP), this phenomenon was called NP/LP equivalence. This relationship (3) together with its geometric illustration shown in Figure 1 appears to indicate a

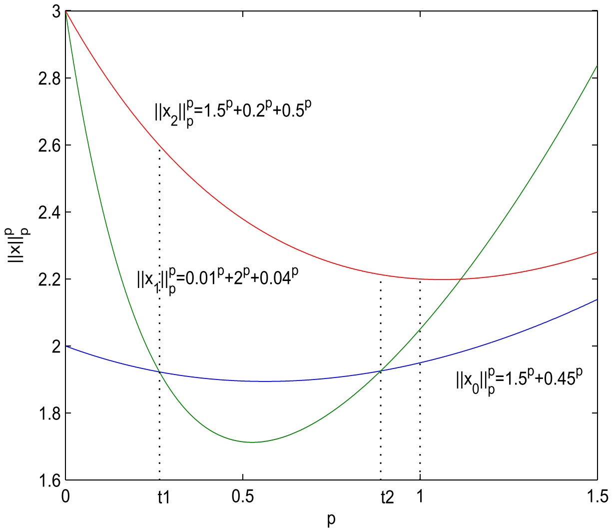

Fig. 2. The optimal solution to l_p-minimization varies with decreasing p.

more aggressive tendency of (P_p) to drive its solution to be sparse as p decreases. However, this is not true. For example, consider the underdetermined system $Ax = b$, where

$$A = \begin{pmatrix} 2 & \frac{151}{765} & \frac{4477}{306} & -\frac{40}{9} \\ 1 & \frac{119}{170} & \frac{153}{68} & 0 \\ -1 & -\frac{629}{1530} & -\frac{2873}{612} & \frac{10}{9} \end{pmatrix}, \quad b = (1, 1.5, -1)^T. \tag{4}$$

It is easy to verify that the sparsity solutions to $Ax = b$ are $x_0 = (1.5, 0, 0, 0.45), x_1 = (0.01, 2, 0.04, 0)$ and $x_2 = (0, 1.5, 0.2, 0.5)$. It can be seen from Figure 2 that when $t_2 < p \leq 1$, $x_0 = (1.5, 0, 0, 0.45)$ is the solution to l_p-norm minimization; when $t_1 < p < t_2$, $x_1 = (0.01, 2, 0.04, 0)$ is the solution to l_p-norm minimization and when $0 \leq p < t_1$, $x_0 = (1.5, 0, 0, 0.45)$ is the solution to l_p-norm minimization.

Nevertheless, based on the fact that (P_0) and (P_1) are just the extremes of (P_p) with respect to p in the interval $(0, 1)$, we still believe that there exists a more general equivalence between (P_0) and (P_p). The aim of this paper is to demonstrate this equivalence. Our paper is organized as follows. In Section 2 we first derive the constructions and locations of solutions to (P_p) based on the classical Bauer maximum principle using a decomposition of the n-dimension real space $\mathbb{R}^n$ with respect to the system $Ax = b$. In section 3, we focus on proving the main theorem, which establishes the equivalence between (P_p) and (P_0), named NP/CMP equivalence, where CMP stands for continuous minimization problem. Finally, in section 4 of the paper we summarize our findings.

For convenience, in this paper we denote by $\mathbb{R}^n$ the n-dimension real space, and for a vector $x \in \mathbb{R}^n$ by x_i its i^{th} component and by $|x|$ its module vector (i.e., $|x| = (|x_1|, |x_2|, \cdots, |x_n|)^T$). We also use $\mathbb{R}^n_+$ to represent the positive cone $\{x \in \mathbb{R}^n : x_i \geq 0, i = 1, 2, \cdots, n\}$.

II. Preliminaries: Constructions and Locations of Solutions to (P_p)

This preliminary section focuses on exploring how to construct the solutions to the problems (P_p) and where the solutions locate. For the minimization problems (P_p), which are subject to underdetermined linear systems, we first analyze the constructions of solutions to the linear equation $Ax = b$, where A is an $m \times n$ real matrix with $m < n$ and b is an n-dimension real vector. Without loss of generality, we assume throughout this paper that A has full row rank. Clearly, by the nature of the geometric aspect of A, we have construction decomposition of solutions to $Ax = b$ as shown below.

Lemma 1: Denote by $N(A)$ the null space of A, and $R(A^T)$ the range of A^T, which is the transposition of A. Then it follows that the following space decomposition

$$\mathbb{R}^n = N(A) \oplus R(A^T), \tag{5}$$

which means that, for every $x \in \mathbb{R}^n$ there uniquely exist $x_N \in N(A)$ and $x_R \in R(A^T)$ such that $x = x_N + x_R$. Therefore, every solution to $Ax = b$ can be explicitly expressed as $x = x_N + A^T(AA^T)^{-1}b$ with $x_N \in N(A)$.

For the sake of convenience, we will adopt the notation $\langle x, y \rangle$ to represent the inner product of real vectors x and y hereafter.

Remark 1: Space decomposition (5) implies that each $x_N \in N(A)$ can be written into the form $x_N = h - A^T(AA^T)^{-1}Ah$ with $h = x_N + A^T b$, and that the null of A has such a parameterized expression as $N(A) = \{h - A^T(AA^T)^{-1}Ah : h \in \mathbb{R}^n\}$. As a result, the constrained minimizations (P_p) for all $p \geq 0$ can be equivalently transformed into the global minimizations as follows

$$\min_{h \in \mathbb{R}^n} \|h + A^T(AA^T)^{-1}(b - Ah)\|_p^p. \tag{6}$$

On the other hand, according to Lemma 1 for all $p > 0$ the optimal values to (P_p) are upper bounded by $\|A^T(AA^T)^{-1}b\|_p^p$, i.e., $\|x\|_p^p \leq \|A^T(AA^T)^{-1}b\|_p^p$ for every solution x to (P_p). This means that the (P_p)s are actually constrained within a bounded subset, for example, the l^∞ ball $B_\infty(r) := \{x \in \mathbb{R}^n : |x_i| \leq r, i = 1, 2, \cdots, n\}$ with $r = n \cdot \sup_{1 \leq i \leq n} |(A^T(AA^T)^{-1}b)_i|$. That is, (P_p) with $p > 0$ is equivalent to the following minimization problem

$$(P'_p) \qquad \min_{Ax=b, x \in B_\infty(r)} \|x\|_p^p. \tag{7}$$

It is important to note that the optimization problem (7) is subject to a special constraint set, which is described as a typical polytope in terms of the following definition.

Definition 1 [16]: A polyhedron G in $\mathbb{R}^n$ is a subset of the intersection of many finite halfspaces, where a halfspace refers to a set of the form $\{x \in \mathbb{R}^n : \langle h, x \rangle \leq \gamma\}$ for some vector $h \in \mathbb{R}^n$ and real number $\gamma \in \mathbb{R}$. Moreover, a polyhedron is called polytope if it is bounded.

By definition, a polyhedron can be compactly expressed as $G = \{x \in \mathbb{R}^n : Hx \leq g\}$ for matrix H and vector g of a certain dimension, where $Hx \leq g$ means that the corresponding inequalities hold for all scale components (i.e., $\langle H_i, x \rangle \leq g_i$, where H_i is the i^{th} row of H. Below we always adopt this notation). Moreover, it is clear that a polytope is closed and convex, and the intersection of many finite polyhedrons is a polytope as long as some of those polyhedrons are bounded. Obviously, the l_∞ ball $B_\infty(r)$ (see Remark 1), as well as the set of solutions to $Ax = b$, is a polyhedron, and so its intersection $\Omega = \{x \in \mathbb{R}^n : Ax = b, |x_i| \leq r, i = 1, 2, \cdots, n\}$ is a polytope.

According to convex polytope theory [18], we know that polytopes possess an important characteristic: that they can be generated by convexifying a finite number of points, all of which are extreme points. In convexity analysis, extreme points always play vital roles. For example, the optimal solution to linear programming, whose constraint set is generally a polytope (or polyhedron), can be achieved at some extreme point (see [18], [24] and references therein). Below we recall the definition of an extreme point.

Definition 2 [24, Ch. 8]: Let Ω be a set of a vector space, and $x^* \in \Omega$. Then, x^* is called an extreme point of Ω if it does not lie in the interior of any line-segment entirely contained in Ω (i.e., x^* necessarily coincides with x_1 or x_2 whenever $x^* = \alpha x_1 + (1-\alpha)x_2$ with $x_1, x_2 \in \Omega$ and $\alpha \in [0, 1]$).

The famous Klein-Milman theorem [1] states that it is necessary for every compact convex subset of a locally convex topological vector space to possesses extreme point(s) and that the vector space is just the closed convex hull of those extreme points. As a particular case, the Minkowski-Caratheodory theorem [24, p. 126, Ch. 8] states that every point from a compact convex subset of $\mathbb{R}^n$ is a convex combination of at most $n + 1$ extreme points. It is well-known that a polytope possesses at most a finite number of extreme points (indeed, if $G = \{x \in \mathbb{R}^n : Hx \leq g\}$ is a polytope, then it has at most 2^m extreme points, where m is the dimension of g (see [22]). We previously mentioned that linear programming can attain its optimal value at some extreme point of the constraint set. In fact, Bauer [2] proved this assertion early in 1960 for a general convex maximization problem with a compact constraint set. We state it as a lemma below.

Lemma 2 Bauer's Maximum Principle [1, p. 298, Ch. 7]: If G is a compact convex subset of locally convex Hausdorff vector space, then every continuous convex function on G has a maximizer that is an extreme point of G.

Recall that the function f on a convex set G is said to be concave if there holds the inequality $\alpha f(x) + (1-\alpha) f(y) \leq f(\alpha x + (1 - \alpha y))$ for all $x, y \in G$ and $\alpha \in (0, 1)$. Lemma 2 means that every continuous concave real function defined on a compact convex subset of locally convex Hausdorff vector space, achieves its minimum value at some extreme points. By definition, a linear function is not only convex but also can be considered as being concave. This is why a linear programming problem always reaches its optimal value at some extreme point, whether it is a minimization or maximization problem. For example, consider the minimization problem (P_1) under the additional nonnegative constrains $x \geq 0$, or equivalently the following linear programming

$$(LP) \quad \min_x \langle \mathbf{1}, x \rangle \quad \text{subject to} \quad Ax = b, x \geq 0, \tag{8}$$

where $\mathbf{1} \in \mathbb{R}^n$ is the vector whose components are all one. Due to the nonnegativity assumption on the variable x, it is easy to show that if the linear programming above has a optimal solution, the optimal solution exists in a bounded set, which would be a polytope. This is true whether the constraint set of linear programming is bounded or not. So by Lemma 2 and Minkowski-Caratheodory theorem it is clear that the minimization problem (P_1) is solvable by searching for the extreme points of the constraint set. However, the extreme point set possesses up to 2^m members, so searching throughout all extreme points is an overwhelming task for large m, which may be an implicit reason why (P_0) is equivalent to (P_1) in some cases.

Before closing this section, we present a further remark on Lemma 2, which is important in substantiating the proof of our main result.

Remark 2: Recall that a concave function f is said to be strict if the equality in $\alpha f(x)+(1-\alpha) f(y) \leq f(\alpha x+(1-\alpha y))$ is available only for $x = y$. From this, it is easy to check that it is necessary for every minimizer of a strictly concave function on a compact convex set G to exist at a certain extreme point of G. Obviously, the function $f_p(x) = \|x\|_p^p$ with $0 < p < 1$ is strictly concave in the positive cone $\mathbb{R}^n_+$.

III. Main Result: Equivalence Between Minimizations (P_0) and (P_p)

In this section, using the preparations provided in the previous section, we will establish the equivalence between (P_0) and (P_p). In order to achieve this, we first consider the minimization problems (P_p) for $0 < p \leq 1$ and present the following lemma.

Lemma 3: Let A be an $m \times n$ real matrix of full row rank, and $b \in \mathbb{R}^n$. Suppose r is a sufficiently large positive real number and define a subset of $\mathbb{R}^n$ as

$$\begin{aligned} G(r) &= \{z \in \mathbb{R}^n : \forall i = 1, 2, \cdots, n, 0 \leq z_i \leq r, \text{and} \\ &\qquad \text{there is a solution } x \text{ to } Ax = b \text{ such that } |x| \leq z\}, \end{aligned} \tag{9}$$

where $|x|$ stands for the module vector of x. Then, $G(r)$ is a polytope in $\mathbb{R}^n$.

Proof: Obviously, $G(r)$ is bounded and closed. Below we will prove that $G(r)$ is a polyhedron. According to Lemma 1 (combined with Remark 1), we know that $x \in \mathbb{R}^n$ solves the system $Ax = b$ if, and only if, it bears the form $x = h - A^T(AA^T)^{-1}Ah + A^T(AA^T)^{-1}b$ for some $h \in \mathbb{R}^n$. Denote by Λ, the set of those vectors $(z^T, h^T)^T$ of $\mathbb{R}^n \times \mathbb{R}^n$ satisfying the following two inequalities

$$-z - (I - A^T(AA^T)^{-1}A)h \leq A^T(AA^T)^{-1}b, \tag{10}$$

and

$$-z + (I - A^T(AA^T)^{-1}A)h \leq -A^T(AA^T)^{-1}b. \tag{11}$$

It is obvious that Λ is a polyhedron of the product space $\mathbb{R}^n \times \mathbb{R}^n$. Let P be the projection from $\mathbb{R}^n \times \mathbb{R}^n$ to the first part, i.e., $P(z, y) = z$. Then, by the projection property of a polyhedron [16] we know that $P(\Lambda)$ is also a polyhedron of $\mathbb{R}^n$. Therefore, to close the proof, we only need to show that $G(r) = P(\Lambda) \cap B^+_\infty(r)$, where $B^+_\infty(r)$ stands for the subset of nonnegative elements of $B_\infty(r)$.

In light of the discussion above, it is clear that $G(r) \subseteq P(\Lambda) \cap B^+_\infty(r)$. For the converse containing relation, let $z \in P(\Lambda) \cap B^+_\infty(r)$. Then, $0 \leq z_i \leq r$ for all

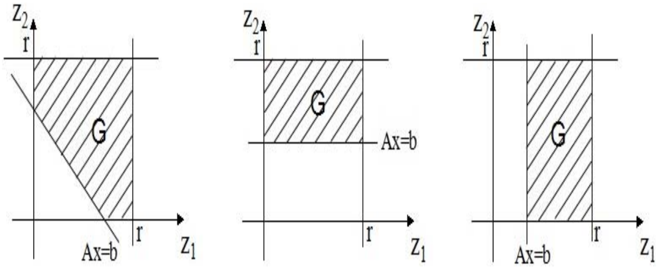

Fig. 3. Shapes of the polytope $G(r)$ defined as in (3) in the plane.

$i = 1, 2, \cdots, n$, and there corresponds a $h \in \mathbb{R}^n$ such that $(z^T, h^T)^T \in \Lambda$, that is, the pair (z, h) satisfies both the inequalities (10) and (11). Let $x = h + A^T(AA^T)^{-1}(b - Ah)$. Then, it is possible to show that x solves $Ax = b$, and it follows from the inequalities (10) and (11) that $|x| \le z$. So $z \in G(r)$. The proof is therefore completed. $\square$

Figure 3 displays basic shapes of $G(r)$ in the plane $\mathbb{R}^2$, except for several degenerate cases (e.g., the segment between these two points $(0, r)^T$ and $(r, r)^T$).

In order to prove the following theorem, which is the main result of this paper, we will remark on the solutions to (P_0). Due to the integer-value virture of l_0-norm, the optimal value of (P_0) is achieved in a bounded set. That is, there exists a constant $r_0 > 0$ such that

$$\min_{Ax=b} \|x\|_0 = \min_{Ax=b, x\in B_\infty(r_0)} \|x\|_0. \tag{12}$$

(Geometrically, it appears to be true that the optimal solutions to (P_0) locate at those intersections of the plane $Ax = b$ with coordinate axis or coordinate planes).

Theorem 1: There exists a constant $p(A, b) > 0$ such that, whenever $0 < p < p(A, b)$, every solution to (P_p) also solves (P_0).

Proof: Let $G(r_1)$ be defined as in (9) with $r_1 = n \cdot \sup_{1\le i\le n} |(A(AA^T)^{-1}b)_i|$. Then, by Lemma 3 we know that $G(r_1)$ is a polytope and hence has a finite number of extreme points. Denote by $E(G(r_1))$, the set of extreme points of $G(r_1)$, and define a constant $r_m(A, b)$ as follows

$$r_m(A, b) = \min_{z\in E(G(r_1)), z_i \ne 0, 1\le i\le n} z_i. \tag{13}$$

Clearly, the defined constant $r_m(A, b)$ is finite and positive (i.e., $0 < r_m(A, b) < \infty$), due to the finiteness of $E(G(r_1))$. (Here we pay attention to the fact that $r_m(A, b)$ only depends on A and b. However, for simplicity we may draw out A, b from $r_m(A, b)$ in the following discussion).

Let r_0 be given as in (12), and $r = \max\{r_0, r_1\}$. Then, we have $G(r_1) \subset G(r)$, and similar to (7) we have

$$\min_{Ax=b} \|x\|_p^p = \min_{Ax=b, x\in B_\infty(r_1)} \|x\|_p^p = \min_{Ax=b, x\in B_\infty(r)} \|x\|_p^p. \tag{14}$$

and

$$\min_{Ax=b} \|x\|_0 = \min_{Ax=b, x\in B_\infty(r_0)} \|x\|_0 = \min_{Ax=b, x\in B_\infty(r)} \|x\|_0. \tag{15}$$

For any given solution x_p to (P_p), by Remark 1 we have $x_p \in B_\infty(r_1)$. Now, let $z_p = |x_p|$. Then, by Lemma 3 it is clear to that $z_p \in G(r_1)$ and solves the following minimization problem

$$(P_p^+) \qquad \min_{z\in G(r_1)} \|z\|_p^p = \sum_{i=1}^n z_i^p. \tag{16}$$

Moreover, since $f(z) = \|z\|_p^p$ is strictly concave in $G(r_1)$, it follows by Lemma 2 and its Remark 2 that $z_p \in E(G(r_1))$. Hence, by noting the fact that the mapping $p \to x^p$ is decreasing for $x \in (0, 1)$ while increasing for $x > 1$, we have

$$\begin{aligned}
\|x_p\|_0 &= \|z_p\|_0 \\
&= \lim_{q\downarrow 0} \left[z_{p1}^q + z_{p2}^q + \cdots + z_{pn}^q\right] \\
&= \lim_{q\downarrow 0} \left[\left(\frac{z_{p1}}{r_m}\right)^q + \left(\frac{z_{p2}}{r_m}\right)^q + \cdots + \left(\frac{z_{pn}}{r_m}\right)^q\right] \\
&\le \left(\frac{z_{p1}}{r_m}\right)^p + \left(\frac{z_{p2}}{r_m}\right)^p + \cdots + \left(\frac{z_{pn}}{r_m}\right)^p \\
&= r_m^{-p} \min_{z\in G(r_1)} \|z\|_p^p = r_m^{-p} \min_{Ax=b, x\in B_\infty(r_1)} \|x\|_p^p \\
&= r_m^{-p} \min_{Ax=b, x\in B_\infty(r)} \|x\|_p^p \\
&= \left(\frac{r}{r_m}\right)^p \min_{Ax=b, x\in B_\infty(r)} \|r^{-1}x\|_p^p \\
&\le \left(\frac{r}{r_m}\right)^p \min_{Ax=b, x\in B_\infty(r)} \|r^{-1}x\|_0 \\
&= \left(\frac{r}{r_m}\right)^p \min_{Ax=b, x\in B_\infty(r)} \|x\|_0 = \left(\frac{r}{r_m}\right)^p \min_{Ax=b} \|x\|_0,
\end{aligned}$$

where the last equality follows from (15). Because $\|x_p\|_0$ is an integer number, from the inequality above, it follows that $\|x_p\|_0 = \min_{Ax=b} \|x\|_0$ (that is, x_p solves (P_0)) when

$$\left(\frac{r}{r_m}\right)^p \min_{Ax=b} \|x\|_0 < \min_{Ax=b} \|x\|_0 + 1 \tag{17}$$

Obviously, the inequality above is true whenever

$$p < \frac{\ln\left(\min_{Ax=b} \|x\|_0 + 1\right) - \ln\left(\min_{Ax=b} \|x\|_0\right)}{\ln r - \ln r_m}. \tag{18}$$

Therefore, with $p(A, b)$ denoting the right side of the inequality above, we conclude that when $0 < p < p(A, b)$, every solution x_p to (P_p) also solves (P_0). The proof is thus completed. $\square$

Remark 3: From the proof it is clear to see that the parameters r_0 and r_1 are only used to bound the ranges of solutions to (P_0) and (P_1) respectively. So, to obtain a better $p(A, b)$ we can replace the number r in the inequality (18) with another number that bounds the solutions to (P_p) for all $0 \le p \le 1$.

As is well known, (P_0) is combinatorial and NP-hard in general, while (P_p) for $p > 0$ is continuous and may be polynomially computable. In [8], the authors proved that (P_p) minimization could be completed by the iteration reweighted least squares minimization algorithm (the IRLS algorithm in brief), that the rate of local convergence of the algorithm was superlinear and that the rate was faster for smaller p and increased towards quadratic as $p \to 0$. And, in [28], the authors demonstrated that $l_{0.5}$ regularization could be fast solved by the iterative half thresholding algorithm (the half algorithm in brief) and that the algorithm was convergent when applied to the k-sparsity problem. In addition, for the IRLS algorithm, at each iteration, the solution of a least squares problem is required, of which the computational complexity is $\mathcal{O}(mN^2)$. Moreover, at each iteration step of the half algorithm, some productions between matrix and vector are required, and thus the computational complexity is $\mathcal{O}(mN)$. In this regard, the significance of the theorem lies in that it really bridges the gap between a combinatorial problem and a continuous one. To highlight the NP nature of (P_0) and the continuity feature of (P_p), we name the phenomenon stated by Theorem 1 "NP/CMP equivalence." Correspondingly, we call the maximal $p(A,b)$ "NP/CMP equivalence constant," and denote it by $p^*(A,b)$.

Obviously, it is important to evaluate the NP/CMP equivalence constant $p^*(A,b)$ for us to choose an appropriate model (P_p) substituting for (P_0). However, this is hard and difficult work even though the inequality (18) can be used to derive a rudimentary estimation. With the relationship (2), it maybe assumed that the constant $p(A,b)$ is determined by some single value (that is, if (P_{p_1}) is equivalent to (P_0), so are all (P_p)s for $p \le p_1$). However, the following example shows that it is not true.

Example 1: Consider the minimization (P_p) with respect to the underdetermined linear system $Ax = b$ with

$$A = \begin{pmatrix} -\frac{20}{29} & 1 & \frac{31}{87} & 0 \\ 0 & 1 & \frac{8}{15} & 1 \\ \frac{60}{29} & 0 & \frac{463}{435} & -1 \end{pmatrix}, \quad b = (1,2,3)^T. \tag{19}$$

It is easy to show that the solutions $x = (x_1, x_2, x_3, x_4)^T$ have the following parameterized form

$$x_1 = t, \quad x_2 = -\frac{4}{27} + \frac{40}{27}t,$$
$$x_3 = \frac{29}{9}\left(1 - \frac{20}{29}t\right), \quad x_4 = \frac{58}{135}\left(1 - \frac{20}{29}t\right). \tag{20}$$

where the parameter t varies in $\mathbb{R}$. Hence, l_p-norm can be computed from the following function of t,

$$\|x\|_p^p = |t|^p + \left|-\frac{4}{27} + \frac{40}{27}t\right|^p + \left|\frac{29}{9}(1 - \frac{20}{29}t)\right|^p + \left|\frac{58}{135}(1 - \frac{20}{29}t)\right|^p. \tag{21}$$

Obviously, the unique solution to (P_0) is $x_0 = (1.45, 2, 0, 0)^T$ (with respect to $t = 1.45$), and all the solutions to (P_p) for $0 \le p \le 1$ exist in the set $B_\infty(2)$. It is simple to confirm that the constant $r_m(A,b)$ defined as in (13) equals 0.1. Hence, from the inequality (18) we can derive that $p^*(A,b) \ge 0.135$.

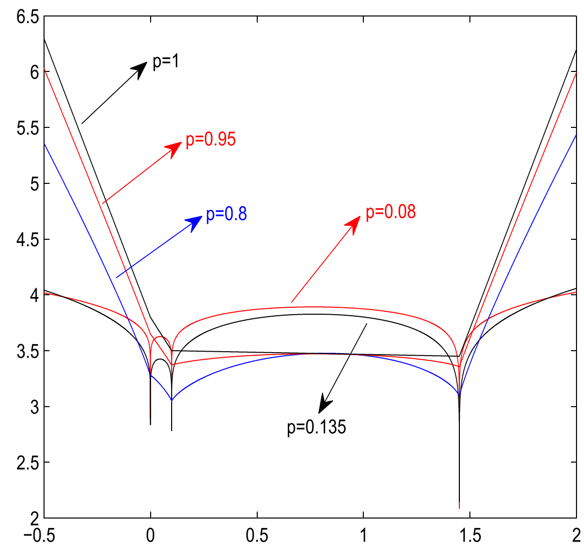

Fig. 4. Behaviors of the l_p-norms as the functions of t for $p = 0.08, 0.135, 0.8, 0.95$ and 1, respectively.

So, from Theorem 1 we know that $x_0 = (1.45, 2, 0, 0)^T$ is the unique solution to (P_p) for $0 < p < 0.135$. Figure 4 demonstrates that the $l_{0.08}$-norm and $l_{0.135}$-norm reach their minimums at $t = 1.45$, which corresponds to the unique solution $x_0 = (1.45, 2, 0, 0)^T$.

Now we consider three cases where $p = 0.8, 0.95, 1$, respectively. The behaviors of $\|x\|_p^p$ as the functions of t are displayed in Figure 4 for those ps. By the formula (21) it is easy to test that $x_{0.8} = (0.1, 0, 3, 0.4)^T$ solves $(P_{0.8})$, while $x_{0.95} = x_1 = (1.45, 2, 0, 0)^T$ solves both $(P_{0.95})$ and (P_1), respectively. But $\|x_{0.8}\|_0 = 3 > \|x_{0.95}\| = 2 = \|x_0\|_0$. This shows that $p^*(A,b) < 0.8$ in spite of the fact that $(P_{0.95})$ possesses the same unique solution as (P_0).

It is worthwhile to note that the example above indicates the same result as the example in section 2, i.e. that in the whole interval $(0,1)$ of p it is not true that the smaller p is, the sparser the solution to (P_p) is.

IV. Conclusion

Among the numerous substitution models for the l_0 minimization problem (P_0), the l_p-norm minimizations (P_p) with $0 < p \le 1$ have been considered as the most natural choice. However, the question "to what extent these models (P_p) can replace (P_0)" has never been answered. In this paper, we have clearly demonstrated the equivalence between (P_0) and (P_p), and in doing so have answered this question. The established equivalence means that solving (P_0) can be completely overcome by solving the continuous minimization (P_p) for some small p, while the latter is computable by some commonly used means at least for some special p. However, it should be pointed out that the main result obtained in this paper is qualitative, and so has not given quantitative characterization to the NP/CMP equivalence constant. The authors think this is important for model choice and subsequently for algorithm design. In conclusion, the authors hope that in publishing this paper, a brick will be thrown out and be replaced with a gem.

ACKNOWLEDGMENT

The authors are grateful to the editor and the referees for their comments which led to substantial improvements in the manuscript.

REFERENCES

[1] C. D. Aliprantis and K. C. Border, *Infinite Dimensional Analysis: A Hitchhiker's Guide*. New York, NY, USA: Springer-Verlag, 2006.

[2] H. Bauer, "Minimalstellen von funktionen und extremalpunkte," *Arch. Math.*, vol. 11, no. 1, pp. 200–205, 1960.

[3] A. M. Bruckstein, D. L. Donoho, and M. Elad, "From sparse solutions of systems of equations to sparse modeling of signals and images," *SIAM Rev.*, vol. 51, no. 1, pp. 34–81, 2009.

[4] E. J. Candes and T. Tao, "Decoding by linear programming," *IEEE Trans. Inf. Theory*, vol. 51, no. 12, pp. 4203–4215, Dec. 2005.

[5] E. J. Candes and T. Tao, "Near-optimal signal recovery from random projections: Universal encoding strategies?" *IEEE Trans. Inf. Theory*, vol. 52, no. 12, pp. 5406–5425, Dec. 2006.

[6] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM J. Sci. Comput.*, vol. 20, no. 1, pp. 33–61, 1999.

[7] R. Chartrand, "Exact reconstruction of sparse signals via nonconvex minimization," *IEEE Signal Process.*, vol. 14, no. 10, pp. 707–710, Oct. 2007.

[8] I. Daubechies, R. DeVore, M. Fornasier, and C. S. Güntürk, "Iteratively reweighted least squares minimization for sparse recovery," *Commun. Pure Appl. Math.*, vol. 63, no. 1, pp. 1–38, 2010.

[9] M. E. Davies and R. Gribonval, "Restricted isometry constants where ℓ^p sparse recovery can fail for $0 < p \leq 1$," *IEEE Trans. Inf. Theory*, vol. 55, no. 5, pp. 2203–2214, May 2009.

[10] D. L. Donoho and X. Huo, "Uncertainty principles and ideal atomic decomposition," *IEEE Trans. Inf. Theory*, vol. 47, no. 7, pp. 2845–2862, Nov. 2001.

[11] D. L. Donoho and J. Tanner, "Sparse nonnegative solution of underdetermined linear equations by linear programming," *Proc. Nat. Acad. Sci. USA*, vol. 102, no. 27, pp. 9446–9451, 2005.

[12] D. L. Donoho and M. Elad, "Optimally sparse representation in general (nonorthogonal) dictionaries via ℓ^1 minimization," *Proc. Nat. Acad. Sci. USA*, vol. 100, no. 5, pp. 2197–2202, 2003.

[13] M. Elad, *Sparse and Redundant Representations: From Theory to Applications in Signal and Image Processing*. New York, NY, USA: Springer-Verlag, 2010.

[14] S. Foucart and M.-J. Lai, "Sparsest solutions of underdetermined linear systems via ℓ_q-minimization for $0 < q \leq 1$," *Appl. Comput. Harmonic Anal.*, vol. 26, no. 3, pp. 395–407, 2009.

[15] G. M. Fung and O. L. Mangasarian, "Equivalence of minimal ℓ_0- and ℓ_p-norm solutions of linear equalities, inequalities and linear programs for sufficiently small p," *J. Optim. Theory Appl.*, vol. 151, no. 1, pp. 1–10, 2011.

[16] M. X. Goemans. (2009, Feb. 20). *Linear Programming and Polyhedral Combinatorics*. [Online]. Available: http://www-math.mit.edu/~goemans/18433S09/polyhedral.pdf

[17] I. F. Gorodnitsky and B. D. Rao, "Sparse signal reconstruction from limited data using FOCUSS: A re-weighted minimum norm algorithm," *IEEE Trans. Signal Process.*, vol. 45, no. 3, pp. 600–616, Mar. 1997.

[18] B. Grünbaum, *Convex Polytopes* (Graduate Texts in Mathematics). New York, NY, USA: Springer-Verlag, 2003.

[19] R. Gribonval and M. Nielsen, "Sparse representations in unions of bases," *IEEE Trans. Inf. Theory*, vol. 49, no. 12, pp. 3320–3325, Dec. 2003.

[20] C. Herzet, C. Soussen, J. Idier, and R. Gribonval, "Exact recovery conditions for sparse representations with partial support information," *IEEE Trans. Inf. Theory*, vol. 59, no. 11, pp. 7509–7524, Nov. 2013.

[21] M.-J. Lai and J. Wang, "An unconstrained ℓ_p minimization with $0 < q \leq 1$ for sparse solution of underdetermined linear systems," *SIAM J. Optim.*, vol. 21, no. 1, pp. 82–101, 2011.

[22] K. G. Murty, "A problem in enumerating extreme points, and an efficient algorithm for one class of polytopes," *Optim. Lett.*, vol. 3, no. 2, pp. 211–237, 2009.

[23] R. Saab, R. Chartrand, and O. Yilmaz, "Stable sparse approximations via nonconvex optimization," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, Mar./Apr. 2008, pp. 3885–3888.

[24] B. Simon, *Convexity: An Analytic Viewpoint* (Cambridge Tracts in Mathematics). Cambridge, U.K.: Cambridge Univ. Press, 2011.

[25] Q. Sun, "Recovery of sparsest signals via ℓ_p minimization," *Appl. Comput. Harmonic Anal.*, vol. 32, no. 3, pp. 329–341, 2012.

[26] S. Theodoridis, Y. Kopsinis, and K. Slavakis. (2012, Nov.) "Sparsity-aware learning and compressed sensing: An overview." [Online]. Available: http://arxiv.org/abs/1211.5231.

[27] M. Wang, W. Xu, and A. Tang, "On the performance of sparse recovery via ℓ_p-minimization ($0 \leq p \leq 1$)," *IEEE Trans. Inf. Theory*, vol. 57, no. 11, pp. 7255–7278, Nov. 2011.

[28] Z. Xu, X. Chang, F. Xu, and H. Zhang, "$L_{1/2}$ regularization: A thresholding representation theory and a fast solver," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 23, no. 7, pp. 1013–1027, Jul. 2012.

Jigen Peng received the B.S. degree in mathematics from Jingxi University, Jinagxi, China, in 1989, and the M.Sc. and Ph.D. degrees in applied mathematics and computing mathematics from Xi'an Jiaotong University, Xi'an, China, in 1992 and 1998, respectively. He is currently a professor with the School of Mathematics and Statistics, Xi'an Jiaotong University. His current research interests include nonlinear functional analysis and applications, data set matching theory, the machine learning theory, and sparse information processing.

Shigang Yue (M'05) is a Professor of Computer Science in the School of Computer Science, University of Lincoln, United Kingdom. He received his Ph.D and MSc degrees from the Beijing University of Technology (BJUT) in 1996 and 1993, and his BEng degree from the Qingdao Technological University (1988). He worked in BJUT as a Lecturer (1996- 1998) and an Associate Professor (1998-1999). He was an Alexander von Humboldt Research Fellow (2000, 2001) at the University of Kaiserslautern, Germany. Before joining the University of Lincoln as a Senior Lecturer (2007) and promoted to Reader (2010) and full Professor (2012), he held research positions in the University of Cambridge, Newcastle University and the University College London (UCL) respectively. His research interests are mainly within the field of artificial intelligence, computer vision, robotics, information processing, brains and neuroscience. He is particularly interested in biological visual neural systems, evolution of neuronal systems and applications - e.g., in collision detection for vehicles, interactive systems and robots. He is the founding director of Computational Intelligence Laboratory (CIL) in Lincoln. He is the coordinator for several EU FP7 projects. He is a member of IEEE, INNS, ISAL and ISBE.

Haiyang Li received the B.S. degree in mathematics from Yan'an University, Shaanxi, China, in 1996, and the M.Sc. and Ph.D. degrees in fundamental mathematics from Shaanxi Normal University, Xi'an, China, in 2005 and 2008, respectively. He is currently a professor with the School of Science, Xi'an Polytechnic University. His current research interests include quantum logic theory, the machine learning theory, and sparse information processing.