Exact Reconstruction Analysis of Log-Sum Minimization for Compressed Sensing

Yanning Shen, Jun Fang, *Member, IEEE*, and Hongbin Li, *Senior Member, IEEE*

***Abstract*—The fact that fewer measurements are needed by log-sum minimization for sparse signal recovery than the ℓ_1-minimization has been observed by extensive experiments. Nevertheless, such a benefit brought by the use of the log-sum penalty function has not been rigorously proved. This paper provides a theoretical justification for adopting the log-sum as an alternative sparsity-encouraging function. We prove that minimizing the log-sum penalty function subject to $Az = y$ is able to yield the exact solution, provided that a certain condition is satisfied. Specifically, our analysis suggests that, for a properly chosen regularization parameter, exact reconstruction can be attained when the restricted isometry constant δ_{3K} is smaller than one, which presents a less restrictive isometry condition than that required by the conventional ℓ_1-type methods.**

***Index Terms*—Compressed sensing, iterative reweighted algorithms, log-sum minimization.**

I. Introduction

The problem of compressed sensing involves the recovery of a high dimensional sparse signal from a small number of measurements [1], [2]. The canonical form of this problem can be presented as

$$\min_{\boldsymbol{z}} \quad \|\boldsymbol{z}\|_0 \quad \text{s.t.} \quad \boldsymbol{Az} = \boldsymbol{Ax} \tag{1}$$

where $\boldsymbol{y} = \boldsymbol{Ax}$ denotes the acquired measurements, $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ is the sampling matrix with $m \ll n$, and $\|\boldsymbol{z}\|_0$ stands for the number of the nonzero components of $\boldsymbol{z}$. It is well-known that any K-sparse vector $\boldsymbol{x}$ can be exactly recovered via (1) if $\delta_{2K} < 1$, where δ_{2K} is the restricted isometry constant associated with the measurement matrix $\boldsymbol{A}$, which is defined as the smallest constant such that

$$(1-\delta_{2K})\|\boldsymbol{z}\|_2^2 \le \|\boldsymbol{Az}\|_2^2 \le (1+\delta_{2K})\|\boldsymbol{z}\|_2^2 \tag{2}$$

holds for all $2K$-sparse vectors [1]. The optimization (1), however, is a non-convex and NP-hard problem that has computational complexity growing exponentially with the signal dimension n. Thus, alternative sparsity-promoting functionals which are more computationally efficient in finding the sparse solution are desirable. The most popular alternative is to replace the ℓ_0-norm in (1) with ℓ_1-norm, which leads to a convex optimization problem that can be solved efficiently. Over the past decade, the use of the ℓ_1-norm as a sparsity-promoting functional for sparse signal recovery has been extensively studied [1]–[6]. It has been shown that ℓ_1 minimization allows recovery of sparse signals from only a few measurements. Nevertheless, as compared with (1), ℓ_1-type methods generally require a more restrictive condition for exact signal reconstruction. It is therefore natural to seek an alternative which can bridge the gap between the ℓ_0 and ℓ_1 minimization. One such alternative is the log-sum penalty function. Replacing the ℓ_0-norm with this sparsity-encouraging functional leads to:

$$\min_{\boldsymbol{z}} \quad L(\boldsymbol{z}) = \sum_{i=1}^{n} \log(|z_i| + \epsilon) \quad \text{s.t.} \quad \boldsymbol{Az} = \boldsymbol{Ax} \tag{3}$$

where $\epsilon > 0$ is a positive parameter to ensure that the function is well-defined. Such a log-sum penalty function was originally introduced in [7] for basis selection and has gained increasing attention recently. It was shown in [8], [9] that by resorting to the bound optimization technique, minimizing (3) can be formulated as an iterative reweighted ℓ_1-minimization process which iteratively minimizes a reweighted ℓ_1 function. In a series of experiments [8], the iterative reweighted algorithm presents uniform superiority over the conventional ℓ_1-type methods in the sense that substantially fewer measurements are needed for exact recovery. In fact, is was shown in [10] that when $\epsilon = 0$, the log-sum penalty function is essentially the same as the ℓ_0-norm. Hence it can be expected that the above regularized log-sum penalty function behaves like the ℓ_0-norm when ϵ is small. Nevertheless, such a benefit brought by the use of the regularized log-sum penalty function in sparse signal recovery has not been rigorously proved so far. In the following, we conduct an in-depth investigation of the optimization (3). Our study will provide a rigorous justification for (3) and the iterative reweighted method.

In addition to the log-sum minimization being considered in this paper, another effective sparsity-promoting strategy is to replace the ℓ_0-norm with the ℓ_p-norm ($0 < p < 1$). Theoretical analyses conducted in [11]–[13] prove that ℓ_p-minimization enjoys a nice theoretical guarantee: ℓ_p-minimization requires fewer measurements than ℓ_1-type methods for sparse signal exact reconstruction. We show that a similar theoretical guarantee is available for the minimization of log-sum penalty function as well.

II. Condition for Exact Reconstruction

We provide theoretical analysis concerning a sufficient condition under which the solution to (3) equals to the true signal

Manuscript received June 11, 2013; revised October 02, 2013; accepted October 07, 2013. Date of current version October 15, 2013. This work was supported in part by the National Science Foundation of China under Grant 61172114, and by the National Science Foundation under Grant ECCS-0901066. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Alexander M. Powell.

Y. Shen, and J. Fang are with the National Key Laboratory of Science and Technology on Communications, University of Electronic Science and Technology of China, Chengdu 611731, China (e-mail: 201121260110@std.uestc.edu.cn; JunFang@uestc.edu.cn).

H. Li is with the Department of Electrical and Computer Engineering, Stevens Institute of Technology, Hoboken, NJ 07030 USA (e-mail: Hongbin.Li@stevens.edu).

Digital Object Identifier 10.1109/LSP.2013.2285579

$\boldsymbol{x}$. Our analysis reveals a relation between the restricted isometry constant δ_{3K} and the regularization parameter ϵ. The main result is summarized as follows.

Theorem 1: Let $\boldsymbol{A}$ be an $m \times n$ matrix, and $\boldsymbol{x} \in \mathbb{R}^n$ be a sparse vector with K non-zero entries. Define

$$x_{\min} \triangleq \min_{i \in T_0} |x_i| \qquad \beta \triangleq \frac{\sqrt{1-\delta_{3K}}}{2\sqrt{K}(L-1)} \tag{4}$$

where T_0 is the support of $\boldsymbol{x}$, $L \triangleq \lceil \frac{n-K}{2K} \rceil$, $\lceil u \rceil$ denotes the ceiling operator that gives the smallest integer no smaller than u, and δ_{3K} is the restricted isometry constant of the measurement matrix $\boldsymbol{A}$. If we have

$$\frac{x_{\min}}{\epsilon} > \min_{\{\tau: 0<\tau<1\}} f(\tau) \tag{5}$$

where

$$f(\tau) \triangleq \max\left\{ \frac{(1/\beta)-1}{1-\tau}, \frac{2}{(\beta\tau)^2} \right\} \tag{6}$$

then the global minimizer of (3) is exactly equal to $\boldsymbol{x}$.

Remark 1: The above condition (5) involves a search of τ over the region $(0,1)$. Notice that the first term inside the max (one out of two) operator in (6) tends to infinity when $\tau \to 1$, while the second term becomes arbitrarily large when $\tau \to 0$. Hence the value of τ minimizing $f(\tau)$ should lie somewhere in between. Nevertheless, when we are trying to select an appropriate parameter ϵ to satisfy (5), an arbitrary value of $\tau \in (0,1)$ can be considered. As long as $x_{\min}/\epsilon > f(\tau)$ is satisfied for a particular choice of $\tau_0 \in (0,1)$, we can guarantee that the condition (5) is automatically satisfied since

$$f(\tau_0) \geq \min_{\{\tau: 0<\tau<1\}} f(\tau) \qquad \tau_0 \in (0,1) \tag{7}$$

For example, if we select $\tau = 10/11$, then we only need to choose an ϵ to ensure

$$\frac{x_{\min}}{\epsilon} > \max\left\{ 11\left(\frac{1}{\beta}-1\right), 2\left(\frac{1.1}{\beta}\right)^2 \right\} \tag{8}$$

We would like to emphasize that the condition (5) is a sufficient condition for exact reconstruction. When ϵ becomes arbitrarily large, the condition (5) will not be satisfied. Nevertheless, in this case, exact reconstruction is still possible. Note that when implementing the iterative reweighted ℓ_1-minimization algorithm, all weights are roughly identical. Hence the log-sum penalty function behaves like ℓ_1-minimization for an arbitrarily large ϵ.

Remark 2: Theorem 1 provides a sufficient condition which guarantees an exact reconstruction via solving (3). A close examination of the condition (5) reveals that the regularization parameter ϵ has an inverse relationship with the restricted isometry constant δ_{3K}, that is, a larger δ_{3K} results in a smaller ϵ and vice versa. In particular, when $\delta_{3K} \to 1$, accordingly we should have $\epsilon \to 0$ in order to ensure the condition (5) is met. One the other hand, for any $0 < \delta_{3K} < 1$, we can always find a sufficiently small ϵ such that the condition (5) is satisfied. Hence we can ensure that, when $\delta_{3K} < 1$, any K-sparse signal can be exactly recovered via (3) for a properly chosen ϵ. Recalling that for ℓ_1-minimization methods, the condition for exact reconstruction is given by $\delta_{3k} + 3\delta_{4k} < 2$ (see [3]). Since we have $\delta_{3k} < \delta_{4k}$, the condition $\delta_{3k} + 3\delta_{4k} < 2$ implies $\delta_{3k} < 0.5$. We see that our result presents a less restrictive isometry condition than that of the ℓ_1-minimization methods. This also explains why the use of the log-sum penalty function turns out to be a better alternative to the ℓ_0 than the ℓ_1-norm.

Remark 3: We note that (3) is a non-convex optimization problem and there is no guarantee that the iterative reweighted algorithm will converge to a global minimum of (3). Nevertheless, we can improve the probability of finding a global minimizer by starting from a number of different initialization points and choosing the converged point that achieves the minimum objective function value. Also, empirical studies suggest that the iterative reweighted algorithm is more likely to converge to an undesirable local minimum when $\epsilon \to 0$. To address this issue, similarly to [14], we can use a monotonically decreasing sequence $\{\epsilon^{(t)}\}$ in updating the weighting parameters. For example, at the very first beginning, $\epsilon^{(t)}$ can be set to a relatively large value, say 1, in order to provide a stable coefficient estimate. We then gradually reduce the value of $\epsilon^{(t)}$ in the subsequent iterations until $\epsilon^{(t_k)}$ attains a value such that (5) is met. Such a process actually amounts to solving the following optimization

$$\min_{\boldsymbol{z}} \quad L(\boldsymbol{z}) = \sum_{i=1}^{n} \log(|z_i| + \epsilon^{(t_k)}) \quad \text{s.t.} \quad \boldsymbol{Az} = \boldsymbol{Ax} \tag{9}$$

Previous iterations that use $\epsilon^{(t)} \neq \epsilon^{(t_k)}$ can be considered a procedure looking for a good initialization point. Numerical results demonstrate that this approach significantly improves the ability of avoiding undesirable local minima.

III. Proof of Theorem 1

Suppose that $\boldsymbol{x}^*$ is the solution of (3). Let $\boldsymbol{h} \triangleq \boldsymbol{x}^* - \boldsymbol{x}$. Clearly, the residual vector $\boldsymbol{h}$ lies in the null space of $\boldsymbol{A}$, i.e.

$$\boldsymbol{Ah} = \boldsymbol{0} \tag{10}$$

Meanwhile, since $\boldsymbol{x}^*$ is the global minimizer of (3), we have

$$L(\boldsymbol{x}^*) = L(\boldsymbol{x}+\boldsymbol{h}) \leq L(\boldsymbol{x}) \tag{11}$$

We wish to prove that, given (5), there does not exist a nonzero vector $\boldsymbol{h}$ which satisfies conditions (10)–(11) simultaneously. Otherwise the global minimizer of (3) is unequal to the original sparse signal, i.e. $\boldsymbol{x}^* \neq \boldsymbol{x}$.

Let us first examine (10). Write $\boldsymbol{h} \triangleq [h_1 \quad h_2 \quad \dots \quad h_n]^T$. We decompose the residual vector $\boldsymbol{h}$ into a sum of a set of vectors $\{\boldsymbol{h}_{T_l}\}_{l=0}^{L}$, where $\boldsymbol{h}_{T_l}$ is a vector with its ith entry equal to h_i for $i \in T_l$ and zero otherwise. The index set T_0 is the support of $\boldsymbol{x}$. Also, we use T_0^c to denote the complement of the index set T_0. Without loss of generality, we assume that the index set T_1 contains the indices associated with the $2K$ largest (in magnitude) coefficients of $\boldsymbol{h}_{T_0^c}$, T_2 corresponds to the indices of the next $2K$ largest (in magnitude) coefficients of $\boldsymbol{h}_{T_0^c}$, and so on. Obviously $\{\boldsymbol{h}_{T_l}\}_{l=0}^{L}$ are $2K$-sparse vectors (possibly except $\boldsymbol{h}_{T_L}$). Our analysis will reveal the relation between $\boldsymbol{h}_{T_0}$ and the largest $2K$ entries of $\boldsymbol{h}_{T_0^c}$. The results are summarized as follows.

Lemma 1: Suppose $\boldsymbol{h}$ is a vector in the null space of $\boldsymbol{A}$, then the $2K$ largest (in magnitude) coefficients in $\boldsymbol{h}_{T_0^c}$ are lower bounded by

$$|h_i| \geq \beta \|\boldsymbol{h}_{T_0}\|_2 \qquad \forall i \in T_1 \tag{12}$$

Proof: See Appendix A ■

The above result suggests that for any nonzero vector $\boldsymbol{h}$ in the null space of $\boldsymbol{A}$, the largest $2K$ entries in $\boldsymbol{h}_{T_0^c}$ cannot be made arbitrarily small relative to $\boldsymbol{h}_{T_0}$.

Now let us consider the second condition (11). Decomposing the indices into two sets, the condition can be re-expressed as

$$\sum_{i\in T_0}\log\left(\frac{|x_i+h_i|+\epsilon}{|x_i|+\epsilon}\right)\leq\sum_{i\in T_0^c}\log\left(\frac{\epsilon}{|h_i|+\epsilon}\right)\tag{13}$$

Based on (12) and (13), we now derive a new condition $\boldsymbol{h}$ has to satisfy. Define

$$\alpha_i\triangleq|x_i|/\epsilon\quad\forall i\in T_0\qquad\alpha_{\min}\triangleq\min_{i\in T_0}\alpha_i$$
$$\eta_i\triangleq|h_i|/\epsilon\quad\forall i\in T_0\qquad\eta_{\max}\triangleq\max_{i\in T_0}\eta_i$$

The term on the left-hand side (LHS) of (13) can be lower bounded by

$$\sum_{i\in T_0}\log\left(\frac{|x_i+h_i|+\epsilon}{|x_i|+\epsilon}\right)\geq\sum_{i\in T_0}\log\left(\frac{|\alpha_i-\eta_i|+1}{\alpha_i+1}\right)\tag{14}$$

while the term on the right-hand side (RHS) of (13) can be upper bounded as

$$\begin{aligned}\sum_{i\in T_0^c}\log\left(\frac{\epsilon}{|h_i|+\epsilon}\right)&\overset{(a)}{\leq}\sum_{i\in T_1}\log\left(\frac{\epsilon}{|h_i|+\epsilon}\right)\\&\overset{(b)}{\leq}2K\log\left(\frac{1}{\beta\eta_{\max}+1}\right)\end{aligned}\tag{15}$$

where (a) holds valid as all terms in the summation are non-positive, and (b) follows from Lemma 1:

$$|h_i|\geq\beta\|\boldsymbol{h}_{T_0}\|_2\geq\beta\max_{j\in T_0}|h_j|=\beta\eta_{\max}\epsilon\quad\forall i\in T_1\tag{16}$$

Combining (13)–(15), we arrive at the following inequality

$$\sum_{i\in T_0}\log\left(\frac{|\alpha_i-\eta_i|+1}{\alpha_i+1}\right)\leq2K\log\left(\frac{1}{\beta\eta_{\max}+1}\right)\tag{17}$$

We now prove that given (5), the above inequality (17) holds only when $\boldsymbol{h}_{T_0}=\boldsymbol{0}$ (note that the inequality (17) only involves entries in $\boldsymbol{h}_{T_0}$). To this objective, it suffices to prove that the converse, i.e.

$$\prod_{i\in T_0}\left(\frac{|\alpha_i-\eta_i|+1}{\alpha_i+1}\right)>\left(\frac{1}{\beta\eta_{\max}+1}\right)^{2K}\tag{18}$$

always holds for $\boldsymbol{h}_{T_0}\neq\boldsymbol{0}$ given (5). Define $\gamma_i\triangleq\eta_i/\alpha_i,\forall i\in T_0$ and divide the set $\{\gamma_i\}$ into two subsets:

$$S_1=\{i:\gamma_i\leq\tau\}\qquad S_2=\{i:\gamma_i>\tau\}\tag{19}$$

where τ is a parameter of our own choice. Here we confine τ to be a value between 0 and 1, i.e. $0<\tau<1$, to facilitate our following analysis. Clearly, we have $S_1\subseteq T_0$, $S_2\subseteq T_0$ and $S_1\cup S_2=T_0$. Let k_0 denote the cardinality of the set S_1, and the cardinality of S_2 is $K-k_0$. The term on the LHS of (18) can be decomposed into a product of two terms associated with the two subsets S_1 and S_2:

$$\begin{aligned}&\prod_{i\in T_0}\left(\frac{|\alpha_i-\eta_i|+1}{\alpha_i+1}\right)\\&=\prod_{i\in S_1}\left(\frac{|\alpha_i-\eta_i|+1}{\alpha_i+1}\right)\prod_{i\in S_2}\left(\frac{|\alpha_i-\eta_i|+1}{\alpha_i+1}\right)\end{aligned}\tag{20}$$

Examine the term associated with the first subset S_1. Since $\gamma_i\leq\tau<1$ for $i\in S_1$, we have $\eta_i<\alpha_i$. For any $i\in S_1$, it can be readily verified the following inequality (a) holds

$$\frac{|\alpha_i-\eta_i|+1}{\alpha_i+1}=\frac{\alpha_i-\eta_i+1}{\alpha_i+1}\overset{(a)}{>}\frac{1}{\beta\eta_i+1}\geq\frac{1}{\beta\eta_{\max}+1}\tag{21}$$

if $\eta_i\neq0,\forall i\in T_0$ (that is, $\boldsymbol{h}_{T_0}\neq\boldsymbol{0}$) and

$$\alpha_i>\frac{(1/\beta)-1}{1-\gamma_i}\quad\forall i\in S_1\tag{22}$$

Noting that $1-\gamma_i\geq1-\tau$, the above condition (22) is guaranteed when

$$\alpha_{\min}>\frac{(1/\beta)-1}{1-\tau}\tag{23}$$

From (21)–(23), we see that given the condition (23), the following inequality holds

$$\prod_{i\in S_1}\left(\frac{|\alpha_i-\eta_i|+1}{\alpha_i+1}\right)>\left(\frac{1}{\beta\eta_{\max}+1}\right)^{k_0}\tag{24}$$

Consider the term associated with the subset S_2 in (20). In Appendix B, we proved that the following inequality holds

$$\prod_{i\in S_2}\left(\frac{|\alpha_i-\eta_i|+1}{\alpha_i+1}\right)\geq\left(\frac{1}{\beta\eta_{\max}+1}\right)^{2K-k_0}\tag{25}$$

when the following condition is met

$$\alpha_{\min}>\frac{2}{(\beta\tau)^2}\tag{26}$$

If $\alpha_{\min}$ satisfies both conditions (23) and (26), i.e.

$$\alpha_{\min}>\max\left\{\frac{(1/\beta)-1}{1-\tau},\frac{2}{(\beta\tau)^2}\right\}\tag{27}$$

by combining (24)–(25), we obtain the inequality (18). Therefore we conclude that the inequality (17) holds only when $\boldsymbol{h}_{T_0}=\boldsymbol{0}$ (note that the inequality becomes an equality when $\boldsymbol{h}_{T_0}=\boldsymbol{0}$). Substituting $\boldsymbol{h}_{T_0}=\boldsymbol{0}$ back into (13), we can quickly reach that $\boldsymbol{h}_{T_0^c}=\boldsymbol{0}$ as well. Also, notice that the parameter τ in (27) can take any value in $(0,1)$. Therefore as long as

$$\alpha_{\min}=\frac{x_{\min}}{\epsilon}>\min_{\tau\in(0,1)}\max\left\{\frac{(1/\beta)-1}{1-\tau},\frac{2}{(\beta\tau)^2}\right\}\tag{28}$$

we have $\boldsymbol{h}=\boldsymbol{0}$, which implies the global minimizer of (3), $\boldsymbol{x}^*$, is exactly equal to $\boldsymbol{x}$. The proof is completed here.

IV. Conclusions

We presented a sufficient condition which guarantees that the global minimizer of (3) yields the exact reconstruction.

Analyses show that when $\delta_{3K} < 1$, any K-sparse signal can be exactly reconstructed (3) for a properly chosen ϵ.

APPENDIX I
PROOF OF LEMMA 1

Define a new index set $T_{01} = T_0 \cup T_1$. For any vector $\boldsymbol{h}$ in the null space of $\boldsymbol{A}$, we have

$$\begin{aligned}\boldsymbol{A}\boldsymbol{h} = \boldsymbol{A}\boldsymbol{h}_{T_{01}} + \boldsymbol{A}\boldsymbol{h}_{T_{01}^c} = \boldsymbol{0} \\ \Rightarrow \|\boldsymbol{A}\boldsymbol{h}_{T_{01}^c}\|_2 - \|\boldsymbol{A}\boldsymbol{h}_{T_{01}}\|_2 = 0\end{aligned} \tag{29}$$

By decomposing $\boldsymbol{h}_{T_{01}^c}$, we reach the following inequality

$$\begin{aligned}0 &= \|\boldsymbol{A}\boldsymbol{h}_{T_{01}^c}\|_2 - \|\boldsymbol{A}\boldsymbol{h}_{T_{01}}\|_2 \leq \sum_{l=2}^{L} \|\boldsymbol{A}\boldsymbol{h}_{T_l}\|_2 - \|\boldsymbol{A}\boldsymbol{h}_{T_{01}}\|_2 \\ &\overset{(a)}{\leq} \sqrt{1+\delta_{2k}} \sum_{l=2}^{L} \|\boldsymbol{h}_{T_l}\|_2 - \sqrt{1-\delta_{3K}}\|\boldsymbol{h}_{T_{01}}\|_2 \\ &\overset{(b)}{\leq} \sqrt{2}(L-1)\|\boldsymbol{h}_{T_2}\|_2 - \sqrt{1-\delta_{3K}}\|\boldsymbol{h}_{T_0}\|_2 \\ &\overset{(c)}{\leq} 2\sqrt{K}(L-1)|h_{T_2^1}| - \sqrt{1-\delta_{3K}}\|\boldsymbol{h}_{T_0}\|_2\end{aligned} \tag{30}$$

where (a) is a result of the restricted isometry property of $\boldsymbol{A}$, (b) follows from the fact $\delta_{2K} \leq 1$ and the decreasing order of $\{\boldsymbol{h}_{T_l}\}_{l=1}^{L}$, (c) comes from $\|\boldsymbol{h}_{T_2}\|_2 \leq \sqrt{2K}|h_{T_2^1}|$, $h_{T_2^1}$ is the first (meanwhile the largest in magnitude) entry in $\boldsymbol{h}_{T_2}$. Rearranging (30), we get

$$|h_{T_2^1}| \geq \frac{\sqrt{1-\delta_{3K}}}{2\sqrt{K}(L-1)}\|\boldsymbol{h}_{T_0}\|_2 \triangleq \beta\|\boldsymbol{h}_{T_0}\|_2 \tag{31}$$

Since $\{\boldsymbol{h}_{T_l}\}_{l=1}^{L}$ are arranged in descending order of magnitude, we have

$$|h_i| \geq |h_{T_2^1}| \geq \beta\|\boldsymbol{h}_{T_0}\|_2 \qquad \forall i \in T_1 \tag{32}$$

The proof is completed here.

APPENDIX II
PROOF OF THE INEQUALITY (25)

We relax the terms on both sides of (25) as follows:

$$\prod_{i\in S_2}\left(\frac{|\alpha_i - \eta_i| + 1}{\alpha_i + 1}\right) > \prod_{i\in S_2}\left(\frac{1}{\alpha_i + 1}\right) \tag{33}$$

$$\begin{aligned}\left(\frac{1}{\beta\eta_{\max}+1}\right)^{2K-k_0} &< \left(\frac{1}{\beta\eta_{\max}}\right)^{2K-k_0} \\ &\overset{(a)}{\leq} \frac{1}{\eta_q^K\beta^{2K-k_0}}\prod_{i\in S_2}\frac{1}{\eta_i} \\ &\overset{(b)}{=} \frac{1}{\alpha_q^K\gamma_q^K\beta^{2K-k_0}}\prod_{i\in S_2}\frac{1}{\alpha_i\gamma_i}\end{aligned} \tag{34}$$

where $\eta_q \triangleq \min_{i\in S_2}\eta_i$, (a) comes from the fact: $\eta_{\max} \geq \eta_i, \forall i \in T_0$ and the cardinality of S_2 is equal to $K - k_0$, and in (b), we replace η_i with $\alpha_i\gamma_i$ as we have $\gamma_i = \eta_i/\alpha_i$. To prove (25), it suffices to show the following inequality holds valid

$$\prod_{i\in S_2}\left(\frac{1}{\alpha_i+1}\right) \geq \frac{1}{\alpha_q^K\gamma_q^K\beta^{2K-k_0}}\prod_{i\in S_2}\frac{1}{\alpha_i\gamma_i} \tag{35}$$

Rearranging (35), we get

$$\alpha_q^K > \varphi \triangleq \frac{1}{\gamma_q^K\beta^{2K-k_0}}\prod_{i\in S_2}\frac{\alpha_i+1}{\alpha_i\gamma_i} \tag{36}$$

Note that $\gamma_i > \tau, \forall i \in S_2$ and

$$\frac{\alpha_i+1}{\alpha_i} \leq \frac{\alpha_{\min}+1}{\alpha_{\min}} \triangleq d \tag{37}$$

where the above inequality comes from the fact that $\alpha_i > 0$ and $\alpha_{\min} \leq \alpha_i$. Hence φ defined in (36) is upper bounded by

$$\varphi < \left(\frac{1}{\beta\tau}\right)^{2K-k_0} d^{K-k_0} < \left(\frac{1}{\beta\tau}\right)^{2K} d^K \tag{38}$$

where the last inequality comes by noting that $0 < \beta < \frac{1}{2}$ (c.f. (4)), $0 < \tau < 1$, and $d > 1$. Therefore the following

$$\alpha_{\min}^K > \left(\frac{1}{\beta\tau}\right)^{2K} d^K \Leftrightarrow \alpha_{\min} > d\left(\frac{1}{\beta\tau}\right)^2 \tag{39}$$

is a sufficient condition for the inequality (36) (consequently (25)) to hold valid. Moreover, note that the condition (39) implies that $\alpha_{\min} > 1$. Therefore we have $d = (\alpha_{\min}+1)/\alpha_{\min} < 2$, in which case the condition (39) can be further relaxed as

$$\alpha_{\min} > \frac{2}{(\beta\tau)^2} \tag{40}$$

The proof is completed here.

REFERENCES

[1] E. Candés and T. Tao, "Decoding by linear programming," *IEEE Trans. Inf. Theory*, no. 12, pp. 4203–4215, Dec. 2005.

[2] D. L. Donoho, "Compressive sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.

[3] E. Candés, J. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Commun. Pure Appl. Math.*, vol. 59, no. 8, pp. 1207–1223, 2006.

[4] J. A. Tropp, "Just relax: Convex programming methods for identifying sparse signals in noise," *IEEE Trans. Inf. Theory*, vol. 52, no. 3, pp. 1030–1051, Mar. 2006.

[5] E. Candés, "The restricted isometry property and its implications for compressive sensing," in *Comptes Rendus de l'Academie des Sciences*, ser. I. Paris, France: l'Academie des Sciences, 2008, vol. 346, pp. 589–592.

[6] M. J. Wainwright, "Sharp thresholds for high-dimensional and noisy sparsity recovery using ℓ_1-constrained quadratic programming (lasso)," *IEEE Trans. Inf. Theory*, vol. 55, no. 5, pp. 2183–2202, May 2009.

[7] R. R. Coifman and M. Wickerhauser, "Entropy-based algorithms for best basis selction," *IEEE Trans. Inf. Theory*, vol. IT-38, pp. 713–718, Mar. 1992.

[8] E. Candés, M. Wakin, and S. Boyd, "Enhancing sparsity by reweighted l_1 minimization," *J. Fourier Anal. Applicat.*, vol. 14, pp. 877–905, Dec. 2008.

[9] D. Wipf and S. Nagarajan, "Iterative reweighted ℓ_1 and ℓ_2 methods for finding sparse solutions," *IEEE J. Sel. Topics Signal Process.*, vol. 4, no. 2, pp. 317–329, Apr. 2010.

[10] B. D. Rao and K. Kreutz-Delgado, "An affine scaling methodology for best basis selection," *IEEE Trans. Signal Process.*, vol. 47, no. 1, pp. 187–200, Jan. 1999.

[11] R. Chartrand, "Nonconvex compressed sensing and error correction," in *Int. Conf. Acoustics, Speech, and Signal Processing*, Honolulu, HI, USA, 2007.

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

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

[14] R. Chartrand and W. Yin, "Iterative reweighted algorithm for compressive sensing," in *Int. Conf. Acoustics, Speech, and Signal Processing*, Las Vegas, NV, USA, 2008.