Augmented l_1 Minimization with Weibull Matrix

Tailong Li[1], Qihao Zhang[2]
[1]School of Economics & Management, Zhejiang Sci-Tech University, West District, Xiasha Higher Education Zone, Hangzhou 310018, P.R. China
[2]Department of Mathematics, University of Science & Technology of China, Hefei 230026, P.R. China
[1] m9845@163.com

Abstract - The linearized Bregman iteration was successful used to find the sparse signal from the its noise measurements. It was proved that the iteration algorithm converges to the augmented l_1 minimization problem $\|\mathbf{x}\|_1 + \frac{1}{2\alpha}\|\mathbf{x}\|_2^2$ [4]. This paper mainly considers the measurement matrix A which is generated by the Weibull random distribution. With the optimal number of the measurements, the stability of the augmented l_1 minimization model is given.

Index Terms - Compressed sensing, sparsity, robust null space property, Weibull random variable, linearized Bregman iteration

1. Introduction

Compressed sensing theory shows that it is high possible to reconstruct sparse signals with fewer measurements than what is classically accepted. It is drawn lots of attentions from different research fields recently. A vector $\mathbf{x} \in \mathbb{R}^n$ is k sparse if the number of its nonzero coefficients is k that is far less n. The linearized Bregman iteration was proposed in [11] to find the true signal $\mathbf{x}$ from the observation

$$\mathbf{y} = A\mathbf{x},$$

where $A \in \mathbb{R}^{m\times n}$ $(m < n)$ is the measurement matrix. The linearized Bregman iteration converges to the solution of [3,4,11]

$$\min_{\mathbf{x}} \|\mathbf{x}\|_1 + \frac{1}{\alpha}\|\mathbf{x}\|_2^2 \quad \text{subject to} \quad \mathbf{y} = A\mathbf{x}, \qquad (1.1)$$

where α is a positive scalar. Here, $\|\mathbf{x}\|_p$, $0 < p \leq \infty$, is the "l_p-norm" on the Euclidean space. Since the model (1.1) smoothes the l_1 norm $\|\mathbf{x}\|_1$ by adding $\frac{1}{2\alpha}\|\mathbf{x}\|_2$, it is named as the augmented l_1 model. A natural question is when the solution of (1.1) is the true sparse signal. To answer this question, the measurement matrix A should satisfy some conditions. A well known property is the restricted isometry property [5]. A given matrix A satisfies the restricted isometry property of order k if there exists a $\delta_k \in (0,1)$ such that $(1-\delta_k)\|\mathbf{x}\|_2^2 \leq \|A\mathbf{x}\|_2^2 \leq (1+\delta_k)\|\mathbf{x}\|_2^2 \quad \text{for all} \quad \|\mathbf{x}\|_0 \leq k$. Suppose that $\mathbf{x}^0$ is k-sparse and A satisfies the restricted isometry property with $\delta_{2k} \leq 0.4404$ and $\alpha \geq 10\|\mathbf{x}^0\|_\infty$, then $\mathbf{x}^0$ is the unique minimizer of (1.1) [7]. Many types of random matrices satisfy the restricted isometry property with high probability, such as subgaussian random matrix [2] and random partial Fourier matrix [7]. For subgaussian random matrix, the condition $\delta_{2k} \leq 0.4404$ implies that the optimal number of the measurements is [2]

$$m = O(k\ln(n/k)).$$

The restricted isometry property also can be used to characterize the stable recovery of near sparse signals from the noisy measurements $\mathbf{y} = A\mathbf{x} + \zeta$, where ζ is the noise vector. The k-sparse approximation error of $\mathbf{x}$ in l_1 is defined by

$$\sigma_k(\mathbf{x}) = \inf_{\|\mathbf{z}\|_0 \leq k} \|\mathbf{x} - \mathbf{z}\|_1.$$

For $p = 2$, let $\mathbf{x}^*$ be the solution of

$$\min_{\mathbf{x}} \|\mathbf{x}\|_1 + \frac{1}{\alpha}\|\mathbf{x}\|_2^2 \quad \text{subject to} \quad \|\mathbf{y} - A\mathbf{x}\|_p \leq \varepsilon. \qquad (1.2)$$

Suppose that $\delta_{2k} \leq 0.3814$ and $\alpha \geq 10\|\mathbf{x}^0\|_\infty$, it was showed in [7] that

$$\|\mathbf{x}^* - \mathbf{x}^0\|_2 \leq C_1\varepsilon + C_2\frac{\sigma_k(\mathbf{x})}{\sqrt{k}},$$

where C_1 and C_2 are constants.

In this paper, we consider the random matrix whose entries $a_{i,j}$ are independent symmetric Weibll random variable with exponent $r \geq 1$ i.e., the tail probabilities satisfy

$$\Pr(|a_{i,j}| \geq t) = \exp\left(-\left(\frac{\sqrt{\Gamma(1+2/r)}t}{\sigma_r}\right)^r\right), \quad t \geq 0.$$

In this paper, the symmetric Weibull random variables are assumed to have mean zero and variance $\sigma_r^2 = 1/m$. Therefore, the squared l_2 norm of each column has mean equal to one. Different with the subgaussian random matrix, the Weibull random matrix does not satisfy the restricted isometry property with the optimal number of measurements. For $r = 1$, it is the Laplace random variables. Then the Laplace random measurement matrices satisfies the restricted isometry property with the number of measurements [1]

$$m = O(k\ln^2(n/k)).$$

For $r = \infty$, it is the Rademacher random variables. To reduce the number of the measurements, we use the robust null space property which was also named as the sparse approximation property to characterize the measurement matrices [6,9]. The main results of this paper are stated as follows:

Theorem 1.1. Let $p = 2$. Let $A \in \mathbb{R}^{m\times n}$ be a Weibull random matrix with exponent $r \geq 1$. Let $\mathbf{y} = A\mathbf{x}^0 + \zeta$, where ζ is an arbitrary noise vector with $\|\zeta\|_2 < \varepsilon$. Then with probability at least $1 - 2\exp(-c_1 m)$, the solution $\mathbf{x}^*$ of (2) with any $\alpha \geq 8\|\mathbf{x}^0\|_\infty$ satisfies

$$\|\mathbf{x}^* - \mathbf{x}^0\|_1 \le C_3\sqrt{k}\varepsilon + C_4\sigma_k(\mathbf{x}^0),$$

and

$$\|\mathbf{x}^* - \mathbf{x}^0\|_2 \le C_3\varepsilon + C_4\frac{\sigma_k(\mathbf{x}^0)}{\sqrt{k}},$$

as long as $m = O(k\ln(n/k))$. The constants C_3 and C_4 are given in (2.16) and (2.17) respectively.

Theorem 1.2. Let $p = 1$. Let $A \in \mathbb{R}^{m\times n}$ be a Weibull random matrix with exponent$r \ge 1$. Let $\mathbf{y} = A\mathbf{x}^0 + \zeta$, where ζ is an arbitrary noise vector with $\|\zeta\|_1 < \varepsilon$. Then with probability at least$C_5 1 - 2\exp(-c_1 m)$, the solution $\mathbf{x}^*$ of (1.2) with any $\alpha \ge 8\|\mathbf{x}^0\|_\infty$ satisfies

$$\|\mathbf{x}^* - \mathbf{x}^0\|_1 \le C_5\varepsilon + C_6\sigma_k(\mathbf{x}^0),$$

as long as $m = O(k\ln(n/k))$.The constants C_5 and C_6 are given in (2.21).

Remark 1.3. It is natural to extend the model (1.1) and model (1.2) to the low rank matrix recovery problem or the signals which are sparse under some powerful transforms. The related problems can be found in [6,7,12] and many references there in.

2. Robust Null Space Property

In this section, the stability of the argument l_1 model is proved via the robust null space property. Let $\mathbf{x}_k$ be the best k-sparse approximation of $\mathbf{x}$ in the l_p. The measurement matrix A satisfies the l_p-robust null space property of order k with constant β_1 and β_2 if

$$\|\mathbf{x}_k\|_p \le \beta_1\|A\mathbf{x}\|_p + \beta_2\frac{\sigma_k(\mathbf{x})}{k^{1-1/p}}, \quad \text{for all } \mathbf{x} \in \mathbb{R}^n. \quad (2.1)$$

The definition of the l_p-robust null space property is slight different with the original one given in [6]. But it is a special case of the more general version of [9]. The connection between the robust null space property and the restricted isometry property was discussed in [6,9].

Lemma 2.1 [6] Let $A \in \mathbb{R}^{m\times n}$ be a Weibull random matrix with exponent $r \ge 1$. Then with probability at least $1 - 2\exp(-c_1 m)$, A satisfies

$$\|\mathbf{x}_k\|_2 \le \tau\frac{\|A\mathbf{x}\|_1}{\sqrt{m}} + \rho\frac{\|\mathbf{x}\|_1}{\sqrt{k}} \quad \text{for all} \quad \mathbf{x} \in \mathbb{R}^n, \quad (2.2)$$

With $\rho \approx \frac{1}{3}$ and $\tau = 4.5$ as long as $m = O(k\ln(n/k))$.

Now we prove that the measurement matrix satisfies the robust null space property.

Lemma 2.2. Suppose that the matrix $A \in \mathbb{R}^{m\times n}$ satisfies (3). Assume that $k < m$. Then we have

$$\|\mathbf{x}_k\|_1 \le \beta_1\|A\mathbf{x}\|_1 + \beta_2\sigma_k(\mathbf{x}) \quad (2.3)$$

and

$$\|\mathbf{x}_k\|_2 \le \beta_1\|A\mathbf{x}\|_2 + \beta_2\frac{\sigma_k(\mathbf{x})}{\sqrt{k}} \quad \text{for all} \quad \mathbf{x} \in \mathbb{R}^n, \quad (2.4)$$

where $\beta_1 = \frac{\tau}{1-\rho}$ and $\beta_2 = \frac{\rho}{1-\rho}$.

Proof. For all $\mathbf{x} \in \mathbb{R}^n$, we have

$$\frac{\|\mathbf{x}\|_1}{\sqrt{n}} \le \|\mathbf{x}\|_2. \quad (2.5)$$

It follows from (2.2) and (2.5) that

$$\frac{\|\mathbf{x}_k\|_1}{\sqrt{k}} \le \|\mathbf{x}_k\|_2 \le \tau\frac{\|A\mathbf{x}\|_1}{\sqrt{m}} + \rho\frac{\|\mathbf{x}_k\|_1 + \sigma_k(\mathbf{x})}{\sqrt{k}} \quad (2.6)$$

which implies that (2.2) holds. It follows from (2.2) and (2.5) that

$$\begin{aligned}\|\mathbf{x}_k\|_2 &\le \tau\frac{\|A\mathbf{x}\|_1}{\sqrt{m}} + \rho\frac{\|\mathbf{x}_k\|_1}{\sqrt{k}} + \rho\frac{\sigma_k(\mathbf{x})}{\sqrt{k}} \\ &\le \tau\|A\mathbf{x}\|_2 + \rho\|\mathbf{x}_k\|_2 + \rho\frac{\sigma_k(\mathbf{x})}{\sqrt{k}} \quad (2.7)\end{aligned}$$

(2.7) implies that (2.4) holds.

The following property given in [7] is also needed.

Proposition 2.3. Let $\mathbf{x}^0 \in \mathbb{R}^n$ be an arbitrary vector, T be the coordinate set of its k largest components in magnitude. Let $\mathbf{x}^*$ be the solution of (1.2). $\mathbf{x}_T^0$ is the vector whose components on T are the as those of the vector $\mathbf{x}$ and vanish on the complement T^c. The error vector $= \mathbf{x}^* - \mathbf{x}^0$ satisfies

$$\sigma_k() \le C_7\|_k\|_1 + C_8\sigma_k(\mathbf{x}^0) \quad (2.8)$$

where $C_7 = \frac{\alpha + \|\mathbf{x}_T^0\|_\infty}{\alpha - \|\mathbf{x}_{T^c}^0\|_\infty}$ and $C_8 = \frac{2\alpha}{\alpha - \|\mathbf{x}_{T^c}^0\|_\infty}$.

Proof of Theorem 1.1. Let $\mathbf{h} = \mathbf{x}^* - \mathbf{x}^0$. Then

$$\|A\mathbf{h}\|_2 \le \|A\mathbf{x}^*\|_2 + \|A\mathbf{x}^0\|_2 \le 2\varepsilon. \quad (2.9)$$

The set of indices $T = \{1, 2, \ldots, k\}$ is assumed to be the locations of the k largest coefficients in absolute value of $\mathbf{x}$. $\mathbf{x}_T$ is the vector equal to $\mathbf{x}$ on the set T and zero elsewhere. For $\mathbf{h} = (h_1, h_2, \ldots, h_n)$, we also assume that

$$|h_{k+1}| \ge |h_{k+2}| \ge \ldots \ge |h_n|.$$

Denote

$$\mathbf{h}_{T_1} = (0, \ldots, h_{k+1}, \ldots, h_{2k}, 0, \ldots, 0)$$

where $T_1 = \{k+1, \ldots, 2k\}$. The definitions of T_j and T_j, $j \ge 2$, are similar. Then

$$\|\mathbf{h}_{T_j}\|_2 \le \frac{\|\mathbf{h}_{T_{j-1}}\|_1}{\sqrt{k}}, \quad j \ge 2,$$

which implies

$$\sum_{j\le 2}\|\mathbf{h}_{T_j}\|_2 \le \frac{\|\mathbf{h}_{T_0^c}\|_1}{\sqrt{k}}. \quad (2.10)$$

It follows from (2.4) and (2.8) that

$$\begin{aligned}\|\mathbf{h}_{T_0}\|_2 &\leq \beta_1\|A\mathbf{h}\|_2+\beta_2\frac{\sigma_k(\mathbf{h})}{\sqrt{k}}\\ &\leq 2\beta_1\varepsilon+\beta_2\frac{C_7\|\mathbf{h}_{T_0}\|_1+C_8\sigma_k(\mathbf{x}^0)}{\sqrt{k}}\\ &\leq 2\beta_1\varepsilon+\beta_2C_7\|\mathbf{h}_{T_0}\|_2+\frac{\beta_2C_8}{\sqrt{k}}\sigma_k(\mathbf{x}^0). \quad (2.11)\end{aligned}$$

Rearranging (2.11) gives

$$\|\mathbf{h}_{T_0}\|_2 \leq \frac{2\beta_1}{1-\beta_2C_7}\varepsilon+\frac{C_8\beta_2}{1-\beta_2C_7}\frac{\sigma_k(\mathbf{x}^0)}{\sqrt{k}}. \quad (2.12)$$

(2.12) together with (2.8) leads to

$$\begin{aligned}\sigma_k(\mathbf{h}) &\leq C_7\|\mathbf{h}_{T_0}\|_1+C_8\sigma_k(\mathbf{x}^0)\\ &\leq C_7\sqrt{k}\|\mathbf{h}_{T_0}\|_2+C_8\sigma_k(\mathbf{x}^0)\\ &\leq \frac{2\beta_1C_7}{1-\beta_2C_7}\sqrt{k}\varepsilon+\left(\frac{C_8\beta_2}{1-\beta_2C_7}+C_8\right)\sigma_k(\mathbf{x}^0) \quad (2.13)\end{aligned}$$

It also follows from (2.4) and (2.8) that

$$\begin{aligned}\|\mathbf{h}_{T_1}\|_2 &\leq 2\beta_1\varepsilon+\beta_2\frac{\|\mathbf{h}_{T_1^c}\|_1}{\sqrt{k}}\\ &\leq 2\beta_1\varepsilon+\beta_2\frac{\|\mathbf{h}_{T_0}\|_1+\|\mathbf{h}_{(T_0\cup T_1)^c}\|_1}{\sqrt{k}}\\ &\leq 2\beta_1\varepsilon+\beta_2\|\mathbf{h}_{T_0}\|_2+\beta_2\frac{\|\mathbf{h}_{(T_0\cup T_1)^c}\|_1}{\sqrt{k}}\\ &\leq 2\beta_1\varepsilon+\beta_2\|\mathbf{h}_{T_0}\|_2+\beta_2\frac{\sigma_k(\mathbf{h})}{\sqrt{k}}\\ &\leq \left(\beta_1+\frac{\beta_1\beta_2(1+C_7)}{1-\beta_2C_7}\right)2\varepsilon\\ &\quad+\left(C_8\beta_2+\frac{2C_8\beta_2^2}{1-\beta_2C_7}\right)\frac{\sigma_k(\mathbf{x}^0)}{\sqrt{k}}. \quad (2.14)\end{aligned}$$

The last inequality (2.14) holds by (2.12) and (2.13). Let $r=1$, or 2. It follows from (2.3), (2.4), (2.10) and (2.14) that

$$\begin{aligned}\|\mathbf{h}\|_r &\leq \|\mathbf{h}_{T_0}\|_r+\|\mathbf{h}_{T_1}\|_r+\sum_{j\geq 2}\|\mathbf{h}_{T_j}\|_r\\ &\leq k^{\frac{1}{r}-\frac{1}{2}}\|\mathbf{h}_{T_0}\|_2+k^{\frac{1}{r}-\frac{1}{2}}\|\mathbf{h}_{T_1}\|_2+k^{\frac{1}{r}-1}\|\mathbf{h}_{T_0^c}\|_1\\ &\leq k^{\frac{1}{r}-\frac{1}{2}}C_3\varepsilon+k^{\frac{1}{r}-1}C_4\sigma_k(\mathbf{x}^0), \quad (2.15)\end{aligned}$$

where

$$C_3=2\beta_1\left(1+\frac{(1+C_7)(1+\beta_2)}{1-\beta_2C_7}\right) \quad (2.16)$$

and

$$C_4=C_8(1+\beta_2)\left(1+\frac{2\beta_2C_8}{1-\beta_2C_7}\right). \quad (2.17)$$

A critical necessary condition for the inequality (2.15) is $1-\beta_2C_7>0$ which is equal to

$$1-\frac{\rho}{1-\rho}\frac{\alpha+\|\mathbf{x}^0_{T_0}\|_\infty}{\alpha-\|\mathbf{x}^0_{T_0^c}\|_\infty}>0 \quad (2.18)$$

Simple calculation shows that (2.18) holds by

$$\alpha>\left[\frac{2}{1-2\rho}\right]\|\mathbf{x}^0\|_\infty=8\|\mathbf{x}\|^0_\infty,$$

where $[\cdot]$ is the integer part of a real number.

Proof of Theorem 1.2. Similarly, let $\mathbf{h}=\mathbf{x}^*-\mathbf{x}^0$. Then we have $\|A\mathbf{h}\|_1\leq 2\varepsilon$ and

$$\|\mathbf{h}_{T_0}\|_1 \leq \frac{2\beta_1}{1-\beta_2C_7}\varepsilon+\frac{C_8\beta_2}{1-\beta_2C_7}\sigma_k(\mathbf{x}^0). \quad (2.19)$$

It follows that

$$\begin{aligned}\sigma_k(\mathbf{h}) &\leq C_7\|\mathbf{h}_{T_0}\|_1+C_8\sigma_k(\mathbf{x}^0)\\ &= \frac{2\beta_1}{1-\beta_2C_7}\varepsilon+\left(\frac{C_8\beta_2}{1-\beta_2C_7}+C_8\right)\sigma_k(\mathbf{x}^0). \quad (2.20)\end{aligned}$$

Combining (2.19) and (2.20) leads to

$$\|\mathbf{h}\|_1=\|\mathbf{x}^*-\mathbf{x}^0\|_1=\|\mathbf{h}_{T_0}\|_1+\sigma_k(\mathbf{h})\leq C_5\varepsilon+C_6\sigma_k(\mathbf{x}^0),$$

where

$$C_5=\frac{4\beta_1}{1-\beta_2C_7} \quad \text{and} \quad C_6=\frac{2C_8\beta_2}{1-\beta_2C_7}+C_8. \quad (2.21)$$

References

[1] R. Adamczak, A. E Litvak, A. Pajor, and N. Tomczak-Jaegermann, “Restricted isometry property of matrices with independent columns andneighborly polytopes by random sampling.” *Constructive Approximation*, 34(1):61-88, 2011.

[2] R. Baraniuk, M. Davenport, R. DeVore, and M. Wakin, “A simple proof of the restricted isometry property for random matrices,” *Constructive Approximation*, 28(3):253-263, 2008.

[3] J. Cai, S. Osher, and Z. Shen, “Convergence of the linearized bregman iteration for l_1-norm minimization,” *Mathematics of Computation*, 78(268):2127-2136, 2009.

[4] J. Cai, S. Osher, and Z. Shen. "Linearized Bregman iterations for compressed sensing," *Mathematics of Computation*, 78(267): 1515-1536, 2009.

[5] E. J Candes and T. Tao. “Decoding by linear programming,” *Information Theory, IEEE Transactions on*, 51(12):4203–4215, 2005.

[6] S. Foucart. “Stability and robustness of l_1-minimizations with Weibull matrices and redundant dictionaries,” *Linear Algebra and its Applications*, 2012.

[7] M. Lai and W. Yin, “Augmented l1and nuclear-norm models with a globally linearly convergent algorithm,” *SIAM Journal on Imaging Sciences*, 6(2):1059–1091, 2013.

[8] H. Rauhut, “Compressive sensing and structured random matrices,” *Theoretical foundations and numerical methods for sparse recovery*, 9:1–92, 2010.

[9] Q. Sun, “Sparse approximation property and stable recovery of sparse signals from noisy measurements,” *Signal Processing, IEEE Transactions on*, 59(10):5086–5090, 2011.

[10] W. Yin, “Analysis and generalizations of the linearized bregman method,” *SIAM Journal on Imaging Sciences*, 3(4):856–877, 2010.

[11] W. Yin, S. Osher, D. Goldfarb, and J. Darbon, “Bregman iterative algorithms for l1-minimization with applications to compressed sensing,” *SIAM Journal on Imaging Sciences*, 1(1):143–168, 2008.

[12] X. Zhang and S. Li, “Compressed sensing via dual frame based $l_1$analysis with weibull matrices,” *IEEE Signal Processing Letters*, 20:265–268, 2013.