Typical Performance of Iterative Reweighted Lasso for Compressed Sensing

Chihiro Teramoto, Kazushi Mimura, and Kazunori Iwata
Faculty of Information Sciences, Hiroshima City University
3-4-1 Ohtsuka-higashi, Asaminami-ku, Hiroshima 731-3194, Japan
Tokyo Institute of Technology, Tokyo
Email: mimura@hiroshima-cu.ac.jp

***Abstract*—The iterative reweighted Lasso is an efficient framework to recover a sparse signal from a noisy linear measurement. Applying the statistical mechanical approach, we investigate the performance of the iterative reweighted Lasso by evaluating the mean squared error of the estimates. Regions where the reweighting becomes effective are revealed. Our analysis allows us to investigate property of functions that define the weight.**

I. Introduction

The compressed sensing is a framework that is to find a sparse solution to underdetermined linear systems [1]. The iterative ℓ_1 reweighting algorithms [2] have been paid attention as methods to refine estimates in the field of compressed sensing [3]-[15]. These algorithms are to solve the ℓ_1 minimization problems while the shape of the ℓ_1 ball is changed, and can outperform the ℓ_1 minimization without reweighting. The reweighting methods were applied to Lasso [16], and its convergence property has been already analyzed [17]. In this paper, we focus on the iterative reweighted Lasso discussed by Fosson [17], and investigate its performance by evaluating the mean squared error of the estimates in the iterative process theoretically. We also discuss the property of functions that determine the weight within the numerical analysis of our theoretical results.

II. Preliminaries

A. Problem Settings

Let $\boldsymbol{x}_0 \in \mathbb{R}^N$ be an unknown sparse vector to be estimated. We consider the following linear measurement system:

$$\boldsymbol{y} = A\boldsymbol{x}_0 + \boldsymbol{n}, \tag{1}$$

where $\boldsymbol{y} \in \mathbb{R}^P$, $A \in \mathbb{R}^{P\times N}$, and $\boldsymbol{n} \in \mathbb{R}^P$ denote a measurement, a measurement matrix, and a noise vector, respectively. In the framework of compressed sensing, the main problem is to infer $\boldsymbol{x}_0$ for a given measurement $\boldsymbol{y}$ and a given measurement matrix A. We here suppose that each entry of the measurement matrix $A = (a_{i,j})$ follows the normal distribution with mean zero and variance $1/N$ independently, i.e., $a_{i,j} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, 1/N)$, and each entry of the noise vector $\boldsymbol{n} = (n_\mu)$ also follows the normal distribution with mean 0 and variance σ_0^2, i.e., $n_\mu \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma_0^2)$.

B. ℓ_1 minimization and Lasso

In the noiseless case, the following ℓ_1 minimization [1] is commonly applied to infer the original vector: $\hat{\boldsymbol{x}} = \mathrm{argmin}_{\boldsymbol{x}\in\mathbb{R}^N} \|\boldsymbol{x}\|_1$ s.t. $\boldsymbol{y} = A\boldsymbol{x}$. On the other hand, when the measurement is noisy, the *least absolute shrinkage and selection operator (Lasso)* [16] is widely employed for this estimation problem:

$$\hat{\boldsymbol{x}} = \underset{\boldsymbol{x}\in\mathbb{R}^N}{\mathrm{argmin}}(\|\boldsymbol{y} - A\boldsymbol{x}\|_2^2 + \lambda\|\boldsymbol{x}\|_1), \tag{2}$$

where $\lambda \in \mathbb{R}_+$ denotes the amplitude of the regularization term.

C. Iterative Reweighted ℓ_1 minimization

We first briefly summarize the *iterative reweighted ℓ_1 minimization (IRL1)* proposed by Candès et al. [2]. In the ℓ_1 minimization, the estimates is given as a vector at which the minimum ℓ_1 ball touches a feasible set $\{\boldsymbol{x} : \boldsymbol{y} = A\boldsymbol{x}\}$. The weighted ℓ_1 minimization is a method to resolve a minimization problem by changing the shape of the ℓ_1 ball based on the estimate $\hat{\boldsymbol{x}}_1 = \mathrm{argmin}_{\boldsymbol{x}} \|\boldsymbol{x}\|_1$ s.t. $\boldsymbol{y} = A\boldsymbol{x}$. The weighted ℓ_1-norm is defined by $\|\boldsymbol{w}^\top\boldsymbol{x}\|_1$ for $\boldsymbol{x} \in \mathbb{R}^N$, where $\boldsymbol{w} \in \mathbb{R}^N$ is a given weight vector. By defining the weight vector $\boldsymbol{w}$ using $\hat{\boldsymbol{x}}_1$ properly, the weighted ℓ_1 ball can be sharply pinched at the original vector to be estimated $\boldsymbol{x}_0$. If the first estimate $\hat{\boldsymbol{x}}_1$ is good enough, the estimate given by the weighted ℓ_1 minimization i.e., $\hat{\boldsymbol{x}}_2 = \mathrm{argmin}_{\boldsymbol{x}} \|\boldsymbol{w}^\top\boldsymbol{x}\|_1$ s.t. $\boldsymbol{y} = A\boldsymbol{x}$ can be expected to improve. The iterative reweighted ℓ_1 minimization is a method that repeats this process.

At the t-th iteration, the weight is determined using a function w_t, which is referred to as the *weight function*. It is commonly chosen as

$$w_t(x) = (|x| + \epsilon)^{-1}, \tag{3}$$

for all t, where $\epsilon \in \mathbb{R}_+$ is a constant to avoid that the weight diverges [2]. The weight function w_t is entrywise applied to a vector, i.e., $w_t(\boldsymbol{x}) = (w_t(x_1), \cdots, w_t(x_N))^\top$ for $\boldsymbol{x} = (x_1, \cdots, x_N)^\top$.

D. Iterative Reweighted Lasso

Next, we explain the *iterative reweighted Lasso (IRLasso)* discussed by Fosson [17]. In this paper, we focus on this IRLasso. The reweighting scheme can be also applied to Lasso, which is for noisy measurements, in the same way as IRL1. The algorithm of the IRLasso is summarized in Algorithm 1. Here, $\mathbf{1}_{m\times n}$ denotes an $m \times n$ all-1 matrix.

III. Analysis

We analyze typical performance of IRLasso using the statistical mechanical approach, which is an asymptotic method that quantites of interest are obtained by evaluating a state that the free energy density that corresponds to a given cost function is minimized [18], [19]. IRL1 has been already analyzed [9],

Algorithm 1 Iterative Reweighted Lasso (IRLasso)

Input: Measurement vector $\boldsymbol{y} \in \mathbb{R}^P$, measurement matrix $A \in \mathbb{R}^{P\times N}$, the amplitude of the regularization term $\lambda \in \mathbb{R}_+$, the number of iterations T, and a sequence of weight functions $w_1, \cdots, w_T$.

Output: Estimate $\hat{\boldsymbol{x}}_T \in \mathbb{R}^N$.

1: Set $w_1(\hat{\boldsymbol{x}}_0) = \mathbf{1}_{N\times 1}$.

2: **for** $t = 1$ **to** T **do**

3: Solve $\hat{\boldsymbol{x}}_t = \underset{\boldsymbol{x}_t\in\mathbb{R}^N}{\operatorname{argmin}}(\|\boldsymbol{y}-A\boldsymbol{x}_t\|_2^2+\lambda\|w_t(\hat{\boldsymbol{x}}_{t-1})^\top\boldsymbol{x}_t\|_1)$.

4: **end for**

5: **return** $\hat{\boldsymbol{x}}_T$.

and we follow their analysis. Since this analysis is one of asymptotics, we take the large system limit where $P, N \to \infty$ while the ratio α is kept finite to keep the problem nontrivial. The ratio $\alpha = P/N$ is referred to as the *compression rate*.

According to the method of the statistical mechanics, we first define a cost function in IRLasso, which is called the *Hamiltonian*:

$$H_t(\boldsymbol{x}_t|\boldsymbol{y}, A, \boldsymbol{x}_{t-1}) = \|\boldsymbol{y} - A\boldsymbol{x}_t\|_2^2 + \lambda\|w_t(\boldsymbol{x}_{t-1})^\top\boldsymbol{x}_t\|_1. \quad (4)$$

The Boltamann distribution of this system is

$$p_t(\boldsymbol{x}_t|\boldsymbol{y}, A, \boldsymbol{x}_{t-1}) = Z_t^{-1}\exp[-\beta_t H_t(\boldsymbol{x}_t|\boldsymbol{y}, A, \boldsymbol{x}_{t-1})], \quad (5)$$

where $\beta_t\,(>0)$ denotes a parameter that is called the *inverse temperature*, and Z_t denotes a normalization constant that is defined by

$$Z_t = \int_{\mathbb{R}^N} \exp[-\beta_t H_t(\boldsymbol{x}_t|\boldsymbol{y}, A, \boldsymbol{x}_{t-1})]d\boldsymbol{x}_t, \quad (6)$$

which is referred to as the *partition function*. In the limit where $\beta_t \to \infty$, the Boltzmann distribution p_t takes a non-zero values only at the points that the Hamiltonian takes its minimum value. It should be noted that the Hamiltonian depends on the estimates of the previous iteration via the weight in the weighted ℓ_1-norm. Therefore, the expectation must be taken over the estimates of all previous iterations $\boldsymbol{x}_1, \cdots, \boldsymbol{x}_{T-1}$ to evaluate f_T that is the *free energy density* of the system at the T-th iteration. When the limits $\beta_1 \to \infty, \cdots, \beta_{T-1} \to \infty$ are taken in this order, the Boltzmann distributions $p_1, \cdots, p_{T-1}$ concentrate. Then, the free energy density f_T can be obtained as

$$\begin{aligned}f_T &= \lim_{\beta_T\to\infty}\cdots\lim_{\beta_1\to\infty}\lim_{N\to\infty}\frac{-1}{\beta_T N}\mathbb{E}_{A,\boldsymbol{x}_0,\boldsymbol{x}_1,\cdots,\boldsymbol{x}_{t-1},\boldsymbol{n}}[\ln Z_T]\\ &= \lim_{\beta_T\to\infty}\cdots\lim_{\beta_1\to\infty}\lim_{N\to\infty}\frac{-1}{\beta_T N}\mathbb{E}_{A,\boldsymbol{x}_0,\boldsymbol{n}}[\ln Z_1^T - \ln Z_1^{T-1}],\end{aligned} \quad (7)$$

where $Z_1^T = \prod_{t=1}^T Z_t$ which is the partition function of the joint system of $\boldsymbol{x}_1, \cdots, \boldsymbol{x}_T$.

Applying the replica method [18], [19], [20], [21], i.e., $\lim_{N\to\infty}\frac{1}{N}\mathbb{E}_{A,\boldsymbol{x}_0,\boldsymbol{n}}[\ln Z_1^T] = \lim_{n\to 0}\frac{1}{n}\lim_{N\to\infty}\frac{1}{N}\ln\mathbb{E}_{A,\boldsymbol{x}_0,\boldsymbol{n}}[(Z_1^T)^n]$, we can evaluate the quantity $(\beta_T N)^{-1}\mathbb{E}_{A,\boldsymbol{x}_0,\boldsymbol{n}}[\ln Z_1^T]$ which appers in (7). This quantity can be written as

$$\begin{aligned}&\frac{1}{\beta_T N}\ln\mathbb{E}_{A,\boldsymbol{x}_0,\boldsymbol{n}}[(Z_1^T)^n]\\ &= \frac{1}{\beta_T N}\ln\mathbb{E}_{\boldsymbol{x}_0}\Big\{\int_{\mathbb{R}^{nNT}}\Big(\prod_{a=1}^n\prod_{t=1}^T d\boldsymbol{x}_t^a \mathrm{e}^{-\beta_t\lambda\|w_t(\boldsymbol{x}_{t-1}^a)^\top\boldsymbol{x}_t\|_1}\Big)\\ &\quad\times\mathbb{E}_{A,\boldsymbol{n}}\Big[\exp\Big(-\sum_{a=1}^n\sum_{t=1}^T\beta_t\|A\boldsymbol{x}_t^a - \boldsymbol{y}\|_2^2\Big)\Big]\Big\}.\end{aligned} \quad (8)$$

To calculate this quantity (8), we introduce the following parameters:

$$r = N^{-1}\boldsymbol{x}_0^\top\boldsymbol{x}_0 \in \mathbb{R}^1, \quad (9)$$
$$\boldsymbol{m}^a = N^{-1}X_T^{a\top}\boldsymbol{x}_0 \in \mathbb{R}^{T\times 1}, \quad (10)$$
$$P^a = N^{-1}X_T^{a\top}X_T^a \in \mathbb{R}^{T\times T}, \quad (11)$$
$$Q^{a,b} = N^{-1}X_T^{a\top}X_T^b \in \mathbb{R}^{T\times T}\ (a \neq b), \quad (12)$$

where we here defined $X_T^a = (\boldsymbol{x}_1^a, \cdots, \boldsymbol{x}_T^a) \in \mathbb{R}^{N\times T}$. For example, for a function g, we can rewrite $g(N^{-1}\boldsymbol{x}_0^\top\boldsymbol{x}_0)$ to $\int_{\mathbb{R}}\delta(\boldsymbol{x}_0^\top\boldsymbol{x}_0 - Nr)g(r)dr$. In (8), all parameters r, $\boldsymbol{m}^a$, P^a, and $Q^{a,b}$ can be used in a same way. Let

$$\boldsymbol{u}_{T,\mu}^a = \sqrt{\boldsymbol{\beta}}\circ\{\boldsymbol{a}_\mu(X_T^a - \boldsymbol{x}_0\mathbf{1}_{1\times T}) + n_\mu\mathbf{1}_{1\times T}\} \in \mathbb{R}^{1\times T}, \quad (13)$$

where we use the notation $\sqrt{\boldsymbol{\beta}} = (\sqrt{\beta_1}, \cdots, \sqrt{\beta_T}) \in \mathbb{R}^{1\times T}$, and $\circ$ denotes the Hadamard product, i.e, an entrywise product. Here, $\boldsymbol{a}_\mu \in \mathbb{R}^{N\times 1}$ is the μ-th row vector of the measurement matrix A. Letting $\boldsymbol{u}_{T,\mu} = (\boldsymbol{u}_{T,\mu}^1, \cdots, \boldsymbol{u}_{T,\mu}^n) \in \mathbb{R}^{1\times nT}$, the ℓ_2-norm term that is appeared in the second exponential function in (8) can be represented as $\sum_{a=1}^n\sum_{t=1}^T\beta_t\|A\boldsymbol{x}_t^a - \boldsymbol{y}\|_2^2 = \sum_{\mu=1}^P\boldsymbol{u}_{T,\mu}\boldsymbol{u}_{T,\mu}^\top$.

All rows $\boldsymbol{a}_\mu$ of the measurement matrix A are independent each other. In the large-system limit where $P, N \to \infty$ while $\alpha = P/N$ is kept finite, the vector $\boldsymbol{u}_{T,\mu}$ follows an nT-dimensional multivariate Normal distribution with mean $\mathbf{0}$ and covariance matrix Σ by the central limit theorem: $\boldsymbol{u}_{T,\mu} \sim \mathcal{N}(\mathbf{0}, \Sigma)$. The covariance matrix $\Sigma \in \mathbb{R}^{nT\times nT}$ consists of $n \times n$ blocks, and its (a,b)-block $\Sigma^{a,b} = \mathbb{E}_{A,\boldsymbol{n}}[\boldsymbol{u}_{T,\mu}^{a\top}\boldsymbol{u}_{T,\mu}^b] \in \mathbb{R}^{T\times T}$ can be obtained by

$$\Sigma^{a,b} = \begin{cases} B\circ(P^a - M^a - M^{a\top} + R + S_0), & a = b,\\ B\circ(Q^{a,b} - M^a - M^{a\top} + R + S_0), & a \neq b,\end{cases} \quad (14)$$

where $B = \sqrt{\boldsymbol{\beta}}^\top\sqrt{\boldsymbol{\beta}}$, $M^a = \boldsymbol{m}^a\mathbf{1}_{1\times T}$, $R = r\mathbf{1}_{T\times T}$, and $S_0 = \sigma_0^2\mathbf{1}_{T\times T}$. For sufficiently large P and N, the second expectation in (8) can be evaluated as

$$\begin{aligned}&\mathbb{E}_{A,\boldsymbol{n}}[\exp(-\textstyle\sum_{a=1}^n\sum_{t=1}^T\beta_t\|A\boldsymbol{x}_t^a - \boldsymbol{y}\|_2^2)]\\ &= \mathbb{E}_{\boldsymbol{u}_{T,1},\cdots,\boldsymbol{u}_{T,P}}[\exp(-\textstyle\sum_{\mu=1}^P\boldsymbol{u}_{T,\mu}\boldsymbol{u}_{T,\mu}^\top)]\\ &= (\mathbb{E}_{\boldsymbol{u}\sim\mathcal{N}(\mathbf{0},\Sigma)}[\mathrm{e}^{-\boldsymbol{u}\boldsymbol{u}^\top}])^P = |I + 2\Sigma|^{-P/2},\end{aligned} \quad (15)$$

by applying the Gaussian integral.

To proceed further, we assume the *replica symmetry* that means that the parameters do not depend on the replica index a, i.e., $\boldsymbol{m}^a = \boldsymbol{m}$, $P^a = P$, $Q^{a,b} = Q$ for $\forall a, b \in \{1, \cdots, n\}$. Using (15), the replica symmetric assumption, and the saddle-point method [22], the quantity (8) becomes $\lim_{N\to\infty}(-\beta_T N)^{-1}\ln\mathbb{E}_{A,\boldsymbol{x}_0,\boldsymbol{n}}[(Z_1^T)^n]=\mathrm{extr}_{\breve{\Theta}}(\mathcal{A}+\mathcal{B})$, with

$$\mathcal{A}=(-\beta_T N)^{-1}\ln|I+2\Sigma|^{-P/2}, \tag{16}$$

$$\begin{aligned}\mathcal{B}=&(-\beta_T N)^{-1}\ln\mathbb{E}_{\boldsymbol{x}_0}\Bigg[\int\Bigg(\prod_a\prod_t d\boldsymbol{x}_t^a \mathrm{e}^{-\beta_t\lambda\|w_t(\boldsymbol{x}_{t-1}^a)^\top\boldsymbol{x}_t\|_1}\Bigg)\\ &\delta(\boldsymbol{x}_0^\top\boldsymbol{x}_0-Nr)\Bigg\{\prod_a\delta(X_T^{a\top}\boldsymbol{x}_0-N\boldsymbol{m}^a)\delta(X_T^{a\top}X_T^a-NP^a)\Bigg\}\\ &\Bigg\{\prod_a\prod_{b\neq a}\delta(X_T^{a\top}X_T^b-NQ^{a,b})\Bigg\}\Bigg],\end{aligned} \tag{17}$$

where extr denotes the extremization operator, and we put $\breve{\Theta}=\{r,\boldsymbol{m},P,Q\}$ for abbreviation. For any matrices $X=(x_{i,j})$ and $C=(c_{i,j})$, we define the Dirac delta function as $\delta(X-C)=\prod_{i,j}\delta(x_{i,j}-c_{i,j})$.

First, we calculate the term $\mathcal{A}$. Using (14) and the replica symmetric assumption, we have

$$\begin{aligned}\mathcal{A}=&\alpha(2\beta_T)^{-1}\ln\{|I+2B\circ(P-Q)|^{n-1}\\ &|I+2B\circ(P-Q)+2nB\circ(Q-M-M^\top+R+S_0)|\},\end{aligned} \tag{18}$$

where $M=\boldsymbol{m}\mathbf{1}_{1\times T}$. We next evaluate the term $\mathcal{B}$. Using the Fourier integral form of the Dirac delta function $\delta(x)=(2\pi i)^{-1}\int_{-\infty i}^{\infty i}d\hat{x}\mathrm{e}^{\hat{x}x}$, the Dirac delta functions in (17) becomes

$$\begin{aligned}&\delta(\boldsymbol{x}_0^\top\boldsymbol{x}_0-Nr)=\textstyle\int\frac{d\hat{r}}{2\pi i}\mathrm{e}^{\hat{r}(\boldsymbol{x}_0^\top\boldsymbol{x}_0-Nr)},\\ &\delta(X_T^{a\top}\boldsymbol{x}_0-N\boldsymbol{m}^a)=\textstyle\prod_t\int\frac{d\hat{m}_t^a}{2\pi i}\mathrm{e}^{\hat{m}_t^a(\boldsymbol{x}_t^{a\top}\boldsymbol{x}_0-Nm_t^a)},\\ &\delta(X_T^{a\top}X_T^a-NP^a)=\textstyle\prod_s\prod_t\int\frac{d\hat{P}_{s,t}^a}{2\pi i}\mathrm{e}^{\hat{P}_{s,t}^a(\boldsymbol{x}_s^{a\top}\boldsymbol{x}_t^a-NP_{s,t}^a)},\\ &\delta(X_T^{a\top}X_T^b-NQ^{a,b})=\textstyle\prod_s\prod_t\int\frac{d\hat{Q}_{s,t}^{a,b}}{2\pi i}\mathrm{e}^{\hat{Q}_{s,t}^{a,b}(\boldsymbol{x}_s^{a\top}\boldsymbol{x}_t^b-NQ_{s,t}^{a,b})},\end{aligned}$$

where $\boldsymbol{m}^a=(m_t^a)$, $P^a=(P_{s,t}^a)$, and $Q^{a,b}=(Q_{s,t}^{a,b})$. Using these equations, the replica symmetric assumption, and the saddle-point method [22], the term $\mathcal{B}$ becomes

$$\begin{aligned}\mathcal{B}=&\tfrac{-1}{\beta_T}\mathrm{extr}_{\tilde{\breve{\Theta}}}(-\tilde{r}r-2n\tilde{\boldsymbol{m}}^\top\boldsymbol{m}-n\mathrm{tr}\tilde{P}^\top P-n(n-1)\mathrm{tr}\tilde{Q}^\top Q\\ &+\ln\mathbb{E}_{x_0}\{(\textstyle\prod_a\int d\bar{\boldsymbol{x}}_T^a)\exp[-\sum_a\sum_t\beta_t|w_t(x_{t-1}^a)x_t^a|\\ &\quad+\tilde{r}x_0^2+2\textstyle\sum_a\bar{\boldsymbol{x}}_T^{a\top}\tilde{\boldsymbol{m}}x_0+\sum_a\bar{\boldsymbol{x}}_T^{a\top}\tilde{P}\bar{\boldsymbol{x}}_T^a\\ &\quad+\textstyle\sum_a\sum_{b\neq a}\bar{\boldsymbol{x}}_T^{a\top}\tilde{Q}\bar{\boldsymbol{x}}_T^b]\}),\end{aligned} \tag{19}$$

where we put $\tilde{\breve{\Theta}}=\{\tilde{r},\tilde{\boldsymbol{m}},\tilde{P},\tilde{Q}\}$ and $\bar{\boldsymbol{x}}_T^a=(x_1^a,\cdots,x_T^a)^\top$.

In the reference [9], it has been shown that the off-diagonal entries of $P-Q$ and $\tilde{Q}-\tilde{P}$ asymptotically vanish relatively to its diagonal ones as $\beta_1\to\infty,\cdots,\beta_T\to\infty$. As a result, this property causes that the order parameter matrices P, $\tilde{P}$, Q, and $\tilde{Q}$ can be regarded as diagonal ones asymptotically, i.e., even if we treat that P, $\tilde{P}$, Q, and $\tilde{Q}$ are diagonal, we can obtain the same result. Although this property must be shown, in this analysis we just assume that the matrices P, $\tilde{P}$, Q, and $\tilde{Q}$ are diagonal.

To evaluate the operator extr, we need to solve the saddle-point equations. Some parameters diverges to infinity as $\beta_t\to\infty$. To avoid this divergence, the terms that includes β_t in $\mathcal{A}$ and $\mathcal{B}$ must be $O(\beta_t)$. Thus, we introduce the variable transformations: $\hat{m}_t=2\tilde{m}_t/\beta_t$, $\hat{P}_{tt}=2(\tilde{Q}_{tt}-\tilde{P}_{tt})/\beta_t$, $\chi_{tt}=\beta_t(P_{tt}-Q_{tt})$, and $\hat{\chi}_{tt}=2\tilde{Q}_{tt}/\beta_t^2$, which gives

$$\begin{aligned}&\lim_{N\to\infty}(-\beta_T N)^{-1}\mathbb{E}_{x_0}[\ln Z_1^T]\\ &=\lim_{n\to 0}n^{-1}\lim_{N\to\infty}(-\beta_T N)^{-1}\ln\mathbb{E}_{x_0}[(Z_1^T)^n]\\ &=\tfrac{1}{\beta_T}\mathrm{extr}_\Theta(\textstyle\sum_{t=1}^T\{\beta_t\alpha\frac{\sigma_0^2+(r-2m_t+P_{tt})}{1+2\chi_{tt}}+\beta_t\hat{m}_t m_t-\beta_t\frac{\hat{P}_{tt}P_{tt}}{2}\\ &\quad+\beta_t\tfrac{\hat{\chi}_{tt}\chi_{tt}}{2}+\tfrac{\alpha}{2}\ln(1+2\chi_{tt})\}-\tfrac{1}{\beta_T}\mathbb{E}_{x_0}[\int_{\mathbb{R}^T}D\boldsymbol{z}\ln\{\int_{\mathbb{R}^T}d\bar{\boldsymbol{x}}_T\\ &\quad\exp[-\tfrac{1}{2}\textstyle\sum_{t=1}^T\beta_t\hat{P}_{tt}x_t^2+\sum_{t=1}^T\beta_t(z_t\sqrt{\hat{\chi}_{tt}}+x_0\hat{m}_t)x_t\\ &\quad-\textstyle\sum_{t=1}^T\beta_t\lambda|w_t(x_{t-1})x_t|]\}]),\end{aligned} \tag{20}$$

where $\Theta=\{\boldsymbol{m},P,\chi,\hat{\boldsymbol{m}},\hat{P},\hat{\chi}\}$ and $\bar{\boldsymbol{x}}_T=(x_1,\cdots,x_T)^\top$.

Substituting (20) into (7), we can calculate the free energy density f_T. Taking the limit $\beta_1\to\infty,\cdots,\beta_T\to\infty$ in this order, the integrals on $x_1,\cdots,x_T$ converges to corresponding minimization problems. All of them except the integrals on x_T are cancelled in the free energy density f_T, and the past estimates affect the free energy density only via the weight of the weighted ℓ_1-norm. Finally, we arrive at the free energy density of the estimate at the T-th iteration: $f_T=\mathrm{extr}_{\Theta_T}\mathcal{F}_T(\Theta_T)$ with

$$\begin{aligned}&\mathcal{F}_T(\Theta_T)\\ &=\alpha\frac{\sigma_0^2+(r-2m_T+P_{TT})}{1+2\chi_{TT}}+\hat{m}_T m_T-\frac{\hat{P}_{TT}P_{TT}}{2}\\ &\quad+\frac{\hat{\chi}_{TT}\chi_{TT}}{2}+\mathbb{E}_{x_0}\Bigg[\int D\boldsymbol{z}\min_{x_T}\Bigg(\frac{1}{2}\hat{P}_{TT}x_T^2-\{z_T\sqrt{\hat{\chi}_{TT}}\\ &\quad+\hat{m}_T x_0\}x_T+\lambda|w_T(x_{T-1}^*(z_{T-1}))x_T|\Bigg)\Bigg],\end{aligned} \tag{21}$$

where $D\boldsymbol{z}=\prod_{t=1}^T Dz_t$, $Dz_t=(2\pi)^{-1/2}e^{-z_t^2/2}dz_t$, and $\Theta_T=\{m_T,P_{TT},\chi_{TT},\hat{m}_T,\hat{P}_{TT},\hat{\chi}_{TT}\}$. The mean-squred error of the estimate at the T-th iteration can be evaluated by $\mathrm{MSE}=r-2m_T+P_{TT}$, where m_T and P_{TT} are solutions to the saddle-point equations, i.e., $\frac{\partial\mathcal{F}_T}{\partial m_T}=\frac{\partial\mathcal{F}_T}{\partial P_{TT}}=\frac{\partial\mathcal{F}_T}{\partial\chi_{TT}}=\frac{\partial\mathcal{F}_T}{\partial\hat{m}_T}=\frac{\partial\mathcal{F}_T}{\partial\hat{P}_{TT}}=\frac{\partial\mathcal{F}_T}{\partial\hat{\chi}_{TT}}=0$, which are the stationary point equations. The function $x_t^*(z_t)$ in (21) corresponds to the estimate at the t-th iteration. It can be obtained by

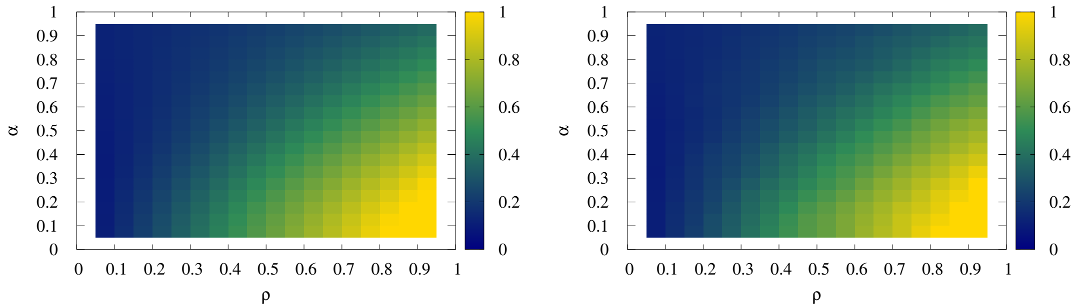
$$\begin{aligned}x_t^*(z_t)=\mathop{\mathrm{argmin}}_{x_t\in\mathbb{R}}\Bigg(&\frac{1}{2}\hat{P}_{tt}x_t^2-\{z_t\sqrt{\hat{\chi}_{tt}}+\hat{m}_t x_0\}x_t\\ &+\lambda|w_t(\hat{x}_{t-1}^*(z_{t-1}))x_t|\Bigg).\end{aligned} \tag{22}$$

Since this equation has a recursive form, we have to solve $x_1^*(z_1),\cdots,x_T^*(z_T)$ in this order. According to Algorithm 1, $w_1(x_0^*(z_0))=1$ must hold. Therefore, we do not need to calculate $x_0^*(z_0)$.

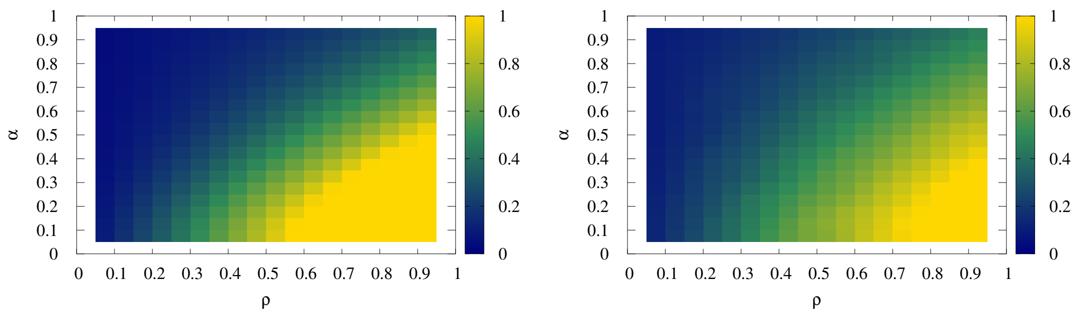
IV. Results

Using (21) and (22), we can evaluate our theoretical MSE of the estimate of IRLasso at the T-th iteration. It is shown that the order parameters r, m_T, and P_{TT} are given as values that minimizes the free energy density averaged over A, $\boldsymbol{x}_0$, and $\boldsymbol{n}$, i.e., $r=\mathbb{E}_{\boldsymbol{x}_0}(\|\boldsymbol{x}_0\|_2^2)$, $m_T=\mathbb{E}_{A,\boldsymbol{x}_0,\boldsymbol{n}}(\boldsymbol{x}_T^\top\boldsymbol{x}_0)$, and $P_{TT}=\mathbb{E}_{A,\boldsymbol{x}_0,\boldsymbol{n}}(\boldsymbol{x}_T^\top\boldsymbol{x}_T)$. We then have MSE at the T-th iteration as $\mathrm{MSE}(T)=\mathbb{E}_{A,\boldsymbol{x}_0,\boldsymbol{n}}(\|\boldsymbol{x}_0-\boldsymbol{x}_T\|_2^2)=r-2m_T+P_{TT}$.

Since we need information of the estimate at the $(t-1)$-th iteration to calculate MSE of the estimate at the t-th iteration,



(a) $t = 1$. Left: theory. Right: simulation. $N = 1000$. Average over 10 trials.



(b) $t = 2$. Left: theory. Right: simulation. $N = 1000$. Average over 10 trials.

Fig. 1. MSE at the first and the second iterations. Horizontal axis: signal density ρ. Vertical axis: compression rate α.

the result has a recursive form. The calculation to obtain MSE of the estimate at the T-th iteration is as follows: 1) Set $t = 0$; 2) $t \leftarrow t + 1$; 3) Calculate m_t, P_{tt}, χ_{tt}, $\hat{m}_t$, $\hat{P}_{tt}$, and $\hat{\chi}_{tt}$ by solving the saddle-point equations; 4) If $t < T$, go back to the step 2; 5) Calculate $\mathrm{MSE}(T) = r - 2m_T + P_{TT}$, which is the result.

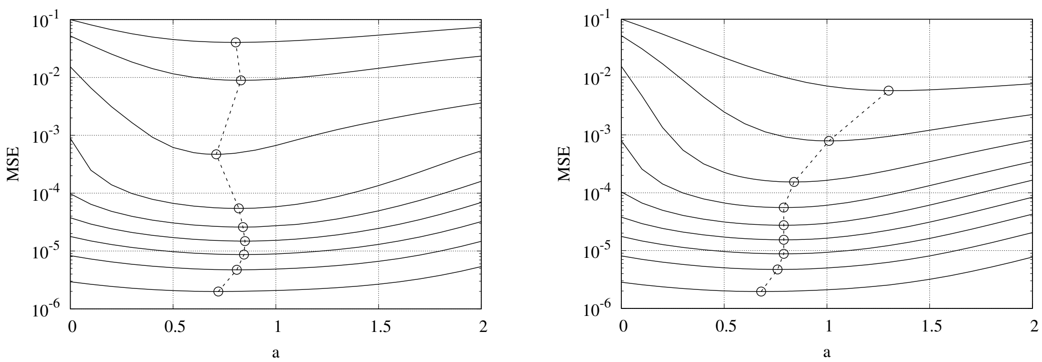
Figure 1 shows comparisons between our theory and numerical simulations. As a distribution of the original vector to be estimated $\boldsymbol{x}_0$, the Bernoulli-Gauss distribution $p_{x_0}(x) = (1 - \rho)\delta(x) + \rho(2\pi)^{-1/2}e^{-x^2/2}$ is applied here. The parameter ρ is referred to as the *signal density*. Note that $r = \mathbb{E}_{x_0}[x_0^2] = \rho$ holds in this case. The horizontal axis is the signal density ρ, and the vertical axis is the compression rate α. We set $w_t(x) = (|x| + 0.1)^{-1}$ for all t as the weight function. The other parameters are $\lambda = 0.1$, $\sigma_0^2 = 0.1$, and $T = 2$. We can confirm that the theoretical results are in good agreement with the numerical simulations.

In the region where MSE is small enough, the estimate at the first iteration is very close to the original vector $\boldsymbol{x}_0$. Therefore, the weight for the second iteration can be appropriately chosen. As a result, the estimate at the second iteration is improved. On the other hand, in the region where MSE is not so small, the weight for the second iteration cannot be appropriately prepared, which causes that the estimate at the second iteration becomes worse than the result of the ℓ_1 minimization without the reweighting.

V. Discussion

Let us consider property of the optimal weight function. Using the result of our analysis, we can consider the weight function that minimizes the typical MSE. Since the weight functions are continuous, it might be hard to optimize a sequence of the weight functions $w_1, \cdots, w_T$. Instead, we treat the case where the form of the weight function is fixed to $w_t(x) = (|x| + \epsilon)^{-a}$ $\forall t$ only, and consider how to optimize the parameters a and ϵ as the first step. The distribution of the original vector is set to the Bernoulli-Gauss distribution.

We first discuss the dependency on a. The other parameters are $\epsilon = 0.1$, $\lambda = 0.001$, and $\sigma_0^2 = 0$. Figure 2 (a) shows the dependency of MSE on a for a fixed ϵ in the case where the estimate at the first iteration is used to determine the weight. It can be confirmed that $a = 1$, which is commonly used, is not optimal, where when the parameter a gives the minimum MSE, it is called optimal. We confirmed that the optimal value of a depends on the parameters such as the compression rate α and the signal density ρ. Figure 2 (b) shows the dependency of MSE on a for a fixed ϵ in the case where the original vector to be estimated $\boldsymbol{x}_0$ is used to determine the weight. The weight defined by the original vector $\boldsymbol{x}_0$ can most sharply pinch the weighted ℓ_1 ball at $\boldsymbol{x}_0$. We can evaluate MSE of this scheme theoretically in the same way to the calculation for $T = 1$. Since the estimate is close to the original vector in the region where MSE is small enough, the optimal a is almost same. However, when MSE is not small, there is a difference between



(a) Based on the estimate at the second iteration.

(b) Based on the original vector to be estimated.

Fig. 2. MSE for $w_t(x) = (|x| + \epsilon)^{-a}$. $\epsilon = 0.1$, $\alpha = 0.9$, and $\rho \in \{0.1, 0.2, \cdots, 0.9\}$ (in order from the bottom). Circles: optimal a.

them.

We have also evaluated the dependency on ϵ and have confirmed that the optimal value of ϵ also depends on the parameters. It should be noted that this analysis must be adopted to other types of measurement matrices, e.g., the rotation invariant matrices [23] and matrices with non-zero-mean entries [24], [25].

VI. Summary

We analyzed the mean squared error of the estimates of the iterative reweighted Lasso by the replica analysis which is one of the statistical mechanical approach. Regions where the reweighting becomes effective were revealed. We showed that the weight function $w_t(x) = (|x| + 0.1)^{-1}$ is not optimal, and the optimal weight function depended on the parameters such as the compression rate and the signal density. Our theory can give property of suitable weight functions. To obtain the optimal weight function is one of our future works.

Acknowledgment

This work was partially supported by JSPS Grant-in-Aid for Challenging Exploratory Research Grant Number 16K12496, JSPS Grant-in-Aid for Scientific Research (B) Grant Numbers 16H02878 and 18H03291, and JST CREST Grant Number JPMJCR15G1, Japan.

References

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

[2] E. J. Candès, M. B. Wakin, and S. P. Boyd, "Enhancing sparsity by reweighted ℓ_1 minimization," *J. Fourier Anal. Appl.*, vol. 14, no. 5–6, pp. 877–905, 2008.

[3] S. Foucartand and M.-J. Laui, "Sparsest solutions of underdetermined linear systems via ℓ_q minimization for $0 < q \leq 1$," *Appl. Comput. Harmon. Anal.*, vol. 26, pp. 395–407, 2009.

[4] D. Needell, "Noisy signal recovery via iterative reweighted l1- minimization," *Proc. of the 43rd Asilomar Conf. on Signals, Systems and Computers*, pp. 113–117, Nov. 2009.

[5] T. Tanaka and J. Raymond, "Optimal incorporation of sparsity information by weighted ℓ_1 optimization," *Proc. of ISIT2010*, pp. 1598–1602, Jun. 2010.

[6] W. Xu, M. Khajehnejad, A. Avestimehr, and B. Hassibi, "Breaking through the thresholds: an analysis for iterative reweighted ℓ_1 minimization via the grassmann angle framework," *Proc. of ICASSP2010*, pp. 5498–5501, Mar. 2010.

[7] M. Khajehnejad, W. Xu, A. Avestimehr, and B. Hassibi, "Improved sparse recovery thresholds with two-step reweighted l1 minimization," *Proc. of ISIT2010*, pp. 1603–1607, Jun. 2010.

[8] 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.

[9] R. Matsushita and T. Tanaka, "Critical compression ratio of iterative reweighted ℓ_1 minimization for compressed sensing," *Proc. of the 2011 IEEE Information Theory Workshop (ITW2011)*, pp. 568–572, Oct. 2011.

[10] W. Xu and A. Tang, "On the scaling law for compressive sensing and its applications," *Proc. of 2011 Annual Conference on Information Sciences and Systems*, Mar. 2011.

[11] Y.-B. Zhao and D. Li, "Reweighted ℓ_1 minimization for sparse solutions to underdetermined linear systems," *SIAM J. Optim.*, vol. 22, no. 3, pp. 1065–1088, 2012.

[12] Q. Lyu, Z. Lin, Y. She, and C. Zhang, "A comparison of typical ℓ_p minimization algorithms," *Neurocomput.*, vol. 119, pp. 413–424, 2013.

[13] M. Asif and J. Romberg, "Fast and accurate algorithms for reweighted ℓ_1-norm minimization," *IEEE Trans. Signal Process.*, vol. 61, no. 23, pp. 5905–5916, Dec. 2013.

[14] X. Chen and W. Zhou, "Convergence of the reweighted ℓ_1 minimization algorithm for ℓ_2-ℓ_p minimization," *Comput. Optim. Appl.*, vol. 59, no. 12, pp. 47–61, 2014.

[15] R. Ahmadand and P. Schniter, "Iteratively reweighted ℓ_1 approaches to sparse composite regularization," *IEEE Trans. Comput. Imag.*, vol. 1, no. 4, pp. 220–235, Dec. 2015.

[16] R. Tibshirani, "Regression Shrinkage and Selection via the lasso," *Journal of the Royal Statistical Society*, vol. 58, no. 1, pp. 267–88, 1996.

[17] S. M. Fosson, "A Biconvex Analysis for Lasso ℓ_1 Reweighting," *IEEE Signal Processing Letters*, vol. 25, no. 12, pp. 1795–1799, Dec. 2018.

[18] M. Mézard, G. Parisi, and M. A. Virasoro, Spin Grass Theory and Beyond, World Scientific Lecture Notes in Physics, World Schientific, 1987.

[19] M. Mézard, and A. Montanari, Information, Physics, and Computation, Oxford University Press, 2006.

[20] Y. Kabashima, T. Wadayama, and T. Tanaka, "A typical reconstruction limit for compressed sensing based on ℓ_p-norm minimization," *Journal of Statistical Mechanics: Theory and Experiment*, L09003, Sep.2009.

[21] F. Krzakala, M. Mézard, F. Sausset, Y. F. Sun, and L. Zdeborová, "Statistical-Physics-Based Reconstruction in Compressed Sensing," *Phys. Rev. X*, vol. 2, 021005, May 2012.

[22] E. T. Copson, *Asymptotic Expansions*, Cambridge Univ. Press, 1965.

[23] G. Parisi and M. Potters, "Mean-field equations for spin models with orthogonal interaction matrices," *J. Phys. A: Math. and Gen.* vol. 28, no. 18, 5267, Sep. 1995.

[24] F. Caltagirone, L. Zdeborová, and F. Krzakala, "On convergence of approximate message passing," *Proc. of ISIT2014*, pp. 1812–1826, Jun. 2014.

[25] T. Tanaka, "Performance analysis of L1-norm minimization for compressed sensing with non-zero-mean matrix elements," *Proc. of ISIT2018*, pp. 401–405, Jun. 2018.