Fast Compressive Phase Retrieval under Bounded Noise

Hongyang Zhang[*1] **Shan You**[*2,3] **Zhouchen Lin**[†2,3] **Chao Xu**[2,3]
[1]Machine Learning Department, Carnegie Mellon University, U.S.A.
[2]Key Lab. of Machine Perception (MOE), School of EECS, Peking University, P. R. China
[3] Cooperative Medianet Innovation Center, Shanghai Jiao Tong University, P. R. China
hongyanz@cs.cmu.edu, youshan@pku.edu.cn, zlin@pku.edu.cn, xuchao@cis.pku.edu.cn

Proof of Measure of Concentration

Lemma 1 ((Shalev-Shwartz and Ben-David 2014))**.** *Let $\mathcal{F}$ be the class of linear predictors with the L_2 norm of the weights bounded by W_2. Assume that the L_2 norm of the instance is bounded by X_2. Then for the ρ-Lipschitz loss ℓ such that $\max_{\langle w,x\rangle\in[-W_2X_2,W_2X_2]}|\ell(w,x,y)| < U$, with probability at least $1-\delta$ over the choice of an i.i.d. sample T of size m,*

$$\forall w\in\{w:\|w\|_2\le W_2\},$$
$$|\mathbb{E}\ell(w,x,y)-\ell(w,T)|\le \frac{2\rho W_2X_2}{\sqrt{m}}+U\sqrt{\frac{2\log(2/\delta)}{m}}.$$

Lemma 2 (Lemma 3 in Main Body)**.** *Let $\mathbf{z}\in\{\mathbf{z}:\|\mathbf{z}\|_2\le 1\}$ and $\{\mathbf{w}_i\}_{i=1}^m$ be random vectors i.i.d. sampled from the standard Gaussian distribution $\mathcal{N}(\mathbf{0},\boldsymbol{I})$. Fix $\mathbf{x}\in\mathbb{R}^n$ and suppose that $m\ge c_0 d\log^4\left(\frac{1}{\delta}\right)\epsilon^{-2}$ with a universal constant c_0, then with probability at least $1-\delta$,*

$$|f_{\mathbf{x}_0}(\mathbf{z})-\mathbb{E}f_{\mathbf{x}_0}(\mathbf{z})|\le\epsilon \tag{1}$$

uniformly holds for all $\mathbf{z}\in\mathbb{R}^n$.

Proof. The proof is basically based on Lemma 1. Note that

$$f_{\mathbf{x}_0}(\mathbf{z})=\frac{1}{m}\sum_{i=1}^m y_i(\mathbf{w}_i^T\mathbf{z})^2=\frac{1}{m}\sum_{i=1}^m((\mathbf{w}_i^T\boldsymbol{\Psi}\mathbf{x}_0)^2+\eta_i)(\mathbf{w}_i^T\mathbf{z})^2. \tag{2}$$

By Lemma 1, we have that

$$\Pr\left[\sup_{\mathbf{z}}|\mathbb{E}f_{\mathbf{x}}(\mathbf{z})-f_{\mathbf{x}}(\mathbf{z})|>\frac{2\rho W_2X_2}{\sqrt{m}}+s\right]\le 2\exp\left(-\frac{ms^2}{2U^2}\right), \tag{3}$$

where the supremum is taken over all $\mathbf{z}\in\{\mathbf{z}:\|\mathbf{z}\|_2\le 1\}$.

To identify the parameters ρ, W_2, X_2, and U above, we exploit the property of standard Gaussian distribution. Specifically, we see that $W_2=1$. By Lemma 4, we have $\|\mathbf{w}_i\|_2\le O(\sqrt{d\log(1/\delta)})\triangleq X_2$ with probability at least $1-\delta$. Let $\ell(\mathbf{w}^T\mathbf{z})=y_i(\mathbf{w}^T\mathbf{z})^2$. Since by Lemma 3 and the fact that η_i is a global constant, $\left|\ell'(\mathbf{w}^T\mathbf{z})\right|\le 2y_i|\mathbf{w}^T\mathbf{z}|\le O(\log^{3/2}(1/\delta))$ for any $\mathbf{w}^T\mathbf{z}$ with high probability. So $\ell(\mathbf{w}^T\mathbf{z})$ is $O(\log^{3/2}(1/\delta))$-Lipschitz, i.e., $\rho=O(\log^{3/2}(1/\delta))$. Furthermore, $|\ell(\mathbf{a}^T\mathbf{z})|\le O(\log^2(1/\delta))\triangleq U$. Plugging in all those parameters, we can see that when $m\ge c_0 d\log^4\left(\frac{1}{\delta}\right)\epsilon^{-2}$, the R.H.S. of (3) is no larger than δ, as desired. □

[*]Equal contributions.

[†]Corresponding author.

Property of Standard Gaussian Distribution

Lemma 3. *Let X be the random variable drawn from standard Gaussian distribution $\mathcal{N}(0,1)$. Then for every $t>0$,*

$$\Pr[|X|>t]\le\exp(-t^2/2). \tag{4}$$

Lemma 4. *Let $\mathcal{P}$ be the isotropic Gaussian distribution in $\mathbb{R}^d$. Then $\Pr_{\mathbf{w}\sim\mathcal{P}}[\|\mathbf{w}\|_2\ge\alpha]\le\left(\frac{e\alpha^2}{d}\right)^{d/2}e^{-\alpha^2/2}$.*

Proof. We have

$$\begin{aligned}\Pr[\|\mathbf{w}\|_2\ge\alpha]&=\Pr[e^{s\|\mathbf{w}\|_2^2}\ge e^{s\alpha^2}]\\&\le\frac{\mathbb{E}e^{s\|\mathbf{w}\|_2^2}}{e^{s\alpha^2}}\\&=e^{-s\alpha^2}(1-2s)^{-d/2},\end{aligned} \tag{5}$$

where the last equality is from the moment generating function of Chi-Square distribution. Setting $s=\frac{\alpha^2-d}{2\alpha^2}$, we obtain the desired result. □

Result on Standard Compressed Sensing

Theorem 5 ((Foucart and Rauhut 2013), Robust Recovery)**.** *Let $\mathbf{x}\in\mathbb{R}^n$ be a t-sparse vector. Suppose that $\mathbf{A}\in\mathbb{R}^{m\times n}$ is a randomly drawn standard Gaussian matrix. Assume that the noisy measurements $\mathbf{b}=\mathbf{A}\mathbf{x}+\mathbf{e}$ are taken with $\|\mathbf{e}\|_2\le\sqrt{\eta/m}$. If*

$$\frac{m^2}{m+1}\ge 2t\left(\sqrt{\log(en/t)}+\sqrt{\log(\delta^{-1})/t}+\tau/\sqrt{t}\right)^2, \tag{6}$$

then with probability at least $1-\delta$, every minimizer $\widetilde{\mathbf{x}}$ to

$$\widetilde{\mathbf{x}}:=\min_{\mathbf{x}}\|\mathbf{x}\|_1,\ \ s.t.\ \ \|\widetilde{\mathbf{z}}-\boldsymbol{\Psi}\mathbf{x}\|_2\le\sqrt{\eta/m}. \tag{7}$$

satisfies

$$\|\widetilde{\mathbf{x}} - \mathbf{x}\|_2 \leq 2\frac{\sqrt{\eta}}{\tau\sqrt{m}}. \tag{8}$$

Property of Bernoulli Model

Lemma 6. *Let n be the number of Bernoulli trials and suppose that $\Omega \sim$ Ber(d/n). Then with probability at least $1-\delta$, $|\Omega| = \Theta(d)$, provided that $d \geq 4\log(1/\delta)$.*

Proof. Take a perturbation ϵ such that $d/n = d_0/n + \epsilon$. By the scalar Chernoff bound which states that

$$\Pr(|\Omega| \leq d_0) \leq e^{-\epsilon^2 n^2/2d_0}, \tag{9}$$

if taking $d_0 = d/2$, $\epsilon = d/2n$ and $d \geq 4\log(1/\delta)$, we have

$$\Pr(|\Omega| \leq d/2) \leq e^{-d/4} \leq \delta. \tag{10}$$

On the other hand, by the scalar Chernoff bound again which states that

$$\Pr(|\Omega| \geq d_0) \leq e^{-\epsilon^2 n^2/3d}, \tag{11}$$

if taking $d_0 = 2d$, $\epsilon = -d/n$ and $d \geq 4\log(1/\delta)$, we obtain

$$\Pr(|\Omega| \geq 2d) \leq e^{-d/3} \leq \delta. \tag{12}$$

Finally, according to (10) and (12), we conclude that $d/2 < |\Omega| < 2d$ with probability at least $1-\delta$. $\square$

References

Foucart, S., and Rauhut, H. 2013. *A Mathematical Introduction to Compressive Sensing*, volume 1. Springer.

Shalev-Shwartz, S., and Ben-David, S. 2014. *Understanding Machine Learning: From Theory to Algorithms*. Cambridge University Press.