upp e entary Mater a Ef c ent A or t s for obust One b t Co press ve ens n

L un Z an ZHANGLJ@LAMDA.NJU.EDU.CN
National Key Laboratory for Novel Software Technology, Nanjing University, Nanjing 210023, China

J nfen Y JINFENGY@US.IBM.COM
IBM Thomas J. Watson Research Center, Yorktown Heights, NY 10598, USA

on J n RONGJIN@CSE.MSU.EDU
Department of Computer Science and Engineering, Michigan State University, East Lansing, MI 48824, USA

A. Proof of Le a 7

We consider the following general optimization problem

$$\min_{\|\mathbf{x}\|_2 \leq 1} -\mathbf{x}^\top \mathbf{y} + \gamma \|\mathbf{x}\|_1. \tag{15}$$

Before we proceed, we need the following lemma.

Le a 6 *The solution to the optimization problem*

$$\min_x \frac{1}{2}(x - y)^2 + \gamma |x|$$

is given by

$$P_\gamma(y) = \begin{cases} 0, & \text{if } |y| \leq \gamma; \\ sign(y)(|y| - \gamma), & \text{otherwise.} \end{cases}$$

where $P_\gamma(\cdot)$ is the soft-thresholding operator defined in (7) (Donoho, 1995).

The proof of Lemma 6 can be found in (Duchi & Singer, 2009). Based on the above lemma, it is easy to verify that

$$\min_x \frac{1}{2}(x - y)^2 + \gamma |x| = \begin{cases} \frac{y^2}{2}, & \text{if } |y| \leq \gamma; \\ \gamma |y| - \frac{\gamma^2}{2}, & \text{otherwise.} \end{cases} \tag{16}$$

First, we consider the case $\|y\|_\infty \leq \gamma$. Then, it is easy to verify that

$$\mathbf{0} \in \operatorname*{argmin}_{\mathbf{x}} -\mathbf{x}^\top \mathbf{y} + \gamma \|\mathbf{x}\|_1.$$

Since $\|\mathbf{0}\|_2 \leq 1$, $\mathbf{0}$ is also an optimal solution to (15).

Next, we consider the case $\|y\|_\infty > \gamma$. Following the standard analysis of convex optimization (Boyd & Vandenberghe, 2004), the Lagrange dual function $g(\mu)$ of (15) is given by

$$\begin{aligned} & g(\mu) \\ = & \min_{\mathbf{x}} -\mathbf{x}^\top \mathbf{y} + \gamma \|\mathbf{x}\|_1 + \mu(\|\mathbf{x}\|_2^2 - 1) \\ = & \min_{\mathbf{x}} 2\mu \left(\frac{1}{2} \left\| \mathbf{x} - \frac{\mathbf{y}}{2\mu} \right\|_2^2 + \frac{\gamma}{2\mu} \|\mathbf{x}\|_1 \right) - \frac{\|\mathbf{y}\|_2^2}{4\mu} - \mu \\ = & 2\mu \left(\sum_i \min_{x_i} \frac{1}{2} \left(x_i - \frac{y_i}{2\mu} \right)^2 + \frac{\gamma}{2\mu} |x_i| \right) - \frac{\|\mathbf{y}\|_2^2}{4\mu} - \mu \\ \overset{(16)}{=} & 2\mu \left(\sum_{i:|y_i| \leq \gamma} \frac{y_i^2}{8\mu^2} + \sum_{i:|y_i| > \gamma} \left(\frac{\gamma |y_i|}{4\mu^2} - \frac{\gamma^2}{8\mu^2} \right) \right) \\ & - \frac{\|\mathbf{y}\|_2^2}{4\mu} - \mu \\ = & \sum_{i:|y_i| > \gamma} \left(\frac{\gamma |y_i|}{2\mu} - \frac{\gamma^2}{4\mu} - \frac{y_i^2}{4\mu} \right) - \mu \\ = & -\frac{\sum_{i:|y_i|>\gamma} (|y_i| - \gamma)^2}{4\mu} - \mu = -\frac{\|P_\gamma(\mathbf{y})\|_2^2}{4\mu} - \mu. \end{aligned}$$

So, the Lagrange dual problem is

$$\max_{\mu \geq 0} -\frac{\|P_\gamma(\mathbf{y})\|_2^2}{4\mu} - \mu$$

and the optimal dual solution is

$$\mu_* = \frac{\|P_\gamma(\mathbf{y})\|_2}{2}.$$

Following the standard analysis (Boyd & Vandenberghe, 2004, Section 5.5.5), the optimal primal solution is

$$\begin{aligned} \mathbf{x}_* = & \operatorname*{argmin}_{\mathbf{x}} \frac{1}{2} \left\| \mathbf{x} - \frac{\mathbf{y}}{2\mu_*} \right\|_2^2 + \frac{\gamma}{2\mu_*} \|\mathbf{x}\|_1 \\ \overset{\text{Lemma 6}}{=} & \frac{1}{\|P_\gamma(\mathbf{y})\|_2} P_\gamma(\mathbf{y}). \end{aligned}$$

B. Proof of Le a

We first consider the case $\text{sign}(\mathbf{x}_k^\top \mathbf{u}) = 1$, i.e.,

$$\mathbf{x}_k^\top \frac{\mathbf{u}}{\|\mathbf{u}\|_2} > \delta_k.$$

Then, we have

$$\begin{aligned}\mathbf{x}_*^\top \frac{\mathbf{u}}{\|\mathbf{u}\|_2} &= \mathbf{x}_k^\top \frac{\mathbf{u}}{\|\mathbf{u}\|_2} + (\mathbf{x}_* - \mathbf{x}_k)^\top \frac{\mathbf{u}}{\|\mathbf{u}\|_2} \\ &> \delta_k - \|\mathbf{x}_* - \mathbf{x}_k\|_2 \overset{(10)}{\geq} 0.\end{aligned}$$

Thus,

$$\text{sign}(\mathbf{x}_*^\top \mathbf{u}) = \text{sign}\left(\mathbf{x}_*^\top \frac{\mathbf{u}}{\|\mathbf{u}\|_2}\right) = 1 = \text{sign}(\mathbf{x}_k^\top \mathbf{u}).$$

The case that $\text{sign}(\mathbf{x}_k^\top \mathbf{u}_i^k) = -1$ can be proved in a similar way.

C. Proof of Le a

First, we have

$$\mathbf{x}_*^\top \mathrm{E}[\mathbf{u}_i \mathbf{y}_i] = \mathrm{E}\left[\mathbf{y}_i \mathbf{x}_*^\top \mathbf{u}_i\right] \overset{(4)}{=} \mathrm{E}\left[\theta(\mathbf{x}_*^\top \mathbf{u}_i)\mathbf{x}_*^\top \mathbf{u}_i\right] \overset{(5)}{=} \lambda,$$

where we use the fact that for a fixed $\mathbf{x}_*$, $\mathbf{x}_*^\top \mathbf{u}_i$ can be treated as a standard Gaussian random variable.

Consider any vector $\mathbf{x} \perp \mathbf{x}_*$. Since $\mathbf{x}_*^\top \mathbf{u}_i$ and $\mathbf{x}^\top \mathbf{u}_i$ are two independent Gaussian random variable, $\mathbf{y}_i$ is independent from $\mathbf{x}^\top \mathbf{u}_i$. Thus, we have

$$\mathbf{x}^\top \mathrm{E}[\mathbf{u}_i \mathbf{y}_i] = \mathrm{E}\left[\mathbf{y}_i \mathbf{x}^\top \mathbf{u}_i\right] = 0.$$

Then, it is easy to prove Lemma 4 by contradiction.

D. Proof of ore

The proof of Theorem 3 is almost identical to that of Theorem 2. The only difference is that in this case, we have

$$\delta_k = \frac{1}{2^{(k-1)/4}},$$

and the total number of calls to the Oracle is upper bounded by

$$\begin{aligned}&\mathbf{m}_1 + 2(\mathbf{K}-1)\mathbf{t} + 2\sqrt{\mathbf{n}}\sum_{k=2}^{K} \delta_k \mathbf{m}_k \\ =&\mathbf{m}_1 + 2(\mathbf{K}-1)\mathbf{t} + 2\sqrt{\mathbf{n}}\mathbf{m}_1 \sum_{k=2}^{K} 2^{3(k-1)/4} \\ \leq& 2(\mathbf{K}-1)\mathbf{t} + (3\sqrt{\mathbf{n}}2^{3K/4} + 1)\mathbf{m}_1.\end{aligned}$$

E. Proof of Coro ary 1

We first consider the case that

$$\mathbf{m} \leq 2(\mathbf{K}-1)\mathbf{t} + (5\sqrt{\mathbf{n}}2^{K/2} + 1)\mathbf{m}_1,$$

which implies

$$\mathbf{m} = \mathbf{O}(2^{K/2}\sqrt{\mathbf{n}}\mathbf{m}_1) = \mathbf{O}(2^{K/2}\mathbf{s}\sqrt{\mathbf{n}}\log \mathbf{n}).$$

Thus,

$$\|\mathbf{x}_{K+1} - \widehat{\mathbf{x}}\|_2 = \frac{1}{2^{K/2}} = \mathbf{O}\left(\frac{\mathbf{s}\sqrt{\mathbf{n}}\log \mathbf{n}}{\mathbf{m}}\right).$$

In the case that

$$\mathbf{m} \leq \mathbf{m}_1 2^K,$$

we have

$$\mathbf{m} = \mathbf{O}(2^K \mathbf{m}_1) = \mathbf{O}(2^K \mathbf{s} \log \mathbf{n}),$$

and thus,

$$\|\mathbf{x}_{K+1} - \widehat{\mathbf{x}}\|_2 = \frac{1}{2^{K/2}} = \mathbf{O}\left(\sqrt{\frac{\mathbf{s}\log \mathbf{n}}{\mathbf{m}}}\right).$$

F. Proof of Coro ary

The proof is the same as that for Corollary 1.

G. Mu t p cat ve C ernoff Bound

ore . *Let $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$ be independent binary random variables with* $\Pr[\mathbf{X}_i = 1] = \mathbf{p}_i$. *Denote $\mathbf{S} = \sum_{i=1}^n \mathbf{X}_i$ and $\mu = \mathrm{E}[\mathbf{S}] = \sum_{i=1}^n \mathbf{p}_i$. We have (Angluin & Valiant, 1979)*

$$\begin{aligned}&\Pr[\mathbf{S} \leq (1-\epsilon)\mu] \leq \exp\left(-\frac{\epsilon^2}{2}\mu\right), \text{ for } 0 < \epsilon < 1, \\ &\Pr[\mathbf{S} \geq (1+\epsilon)\mu] \leq \exp\left(-\frac{\epsilon^2}{2+\epsilon}\mu\right), \text{ for } \epsilon > 0.\end{aligned}$$

For the second bound, let $\mathbf{t} = \frac{\epsilon^2}{2+\epsilon}\mu$, which implies $\epsilon = \frac{t+\sqrt{t^2+8\mu t}}{2\mu}$. Then, with a probability at least $\mathbf{e}^{-t}$, we have

$$\mathbf{S} \leq \left(1 + \frac{\mathbf{t} + \sqrt{\mathbf{t}^2 + 8\mu\mathbf{t}}}{2\mu}\right)\mu \leq 2\mu + 2\mathbf{t}.$$