Research Article

A Subspace Embedding Method in L_2 Norm via Fast Cauchy Transform

Xu Xiang and Li-Zhi Cheng

The State Key Laboratory for High Performance Computation, National University of Defense and Technology, Changsha, Hunan 410073, China

Correspondence should be addressed to Xu Xiang; xiangxu1402@163.com

Received 27 September 2013; Accepted 2 December 2013

Academic Editor: Hector Pomares

Copyright © 2013 X. Xiang and L.-Z. Cheng. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

We propose a subspace embedding method via Fast Cauchy Transform (FCT) in L_2 norm. It is motivated by and complements the work of the subspace embedding method in L_p norm, for all $p \in [1, \infty]$ except $p = 2$, by K. L. Clarkson (ACM-SIAM, 2013). Unlike the traditionally used orthogonal basis in Johnson-Lindenstrauss (JL) embedding, we employ the well-conditioned basis in L_2 norm to obtain concentration property of FCT in L_2 norm.

1. Introduction

Subspace embedding methods have recently proven useful in modern science and engineering. A well-known example is the famous Johnson-Lindenstrauss (JL) lemma [1, 2] which provides low-distortion subspace embeddings in the L_2 norm and has obtained lots of important applications in dimension reduction.

A vital mathematical problem in subspace embedding is to devise a basis U for the embedded subspace such that

$$\forall x \in \mathbb{R}^d, \quad \|x\|_p \leq \|Ux\|_p \leq \text{poly}(d)^2 \|x\|_p, \quad (1)$$

where $\text{poly}(d)$ stands for the distortion; $\text{poly}(d)$ is a polynomial about d . Obviously, this can be done by taking an orthonormal basis for L_2 norm due to the invariant property under an orthonormal transformation. For general L_p norm with $p \neq 2$, the invariant property under an orthonormal transformation disappears. However, works [3, 4] pointed out that it can be overcome by constructing well-conditioned bases for (1). Actually, well-conditioned bases are more practical than orthonormal basis. Recently, Sohler and Woodruff [5] have proved that, using random Cauchy variables, over linear mapping R , with arbitrary large constant probability, for any fixed d -dimensional subspace, for all $x \in L$, $\|x\|_1 \leq \|Rx\|_1 = O(d \log d) \|x\|_1$ holds. Clarkson et al. [6] extended this to L_p norm linear regression, for all $p \in [1, \infty]$ except

$p = 2$. The result of FCT in L_1 norm could be formulated as follows.

Theorem 1 (Fast Cauchy Transform (FCT) in L_1 norm, [6]). *There is a distribution, over matrices $\Pi_1 \in \mathbb{R}^{r_1 \times n}$, with $r_1 = O(d \log(d/\delta))$ and $s = r_1^6$, such that for an arbitrary $A \in \mathbb{R}^{n \times d}$ and for all $x \in \mathbb{R}^d$, the inequality*

$$\|Ax\|_1 \leq \|\Pi_1 Ax\|_1 \leq \kappa \|Ax\|_1 \quad (2)$$

holds with probability $1 - \delta$, where

$$\kappa = O\left(\frac{d\sqrt{s}}{\delta} \log(r_1 d)\right). \quad (3)$$

That is, with large probability, simultaneously for all x , $\|Ax\|_1 \approx \|\Pi_1 Ax\|_1$. Setting $s = r_1^6$, we obtain the distortion $O(d^{2+\eta})$ for an arbitrarily small constant $\eta > 0$. Further, for any $y \in \mathbb{R}^n$, the product $\Pi_1 y$ can be computed in $O(n \log r_1)$ time.

In this paper, we study the Fast Cauchy Transform in the case of L_2 norm and prove the following.

Theorem 2 (FCT in L_2 norm). *There is a distribution, over matrices $\Pi_2 \in \mathbb{R}^{r_1 \times n}$, with $r_1 = O(d \log(d/\delta))$, such that for an arbitrary $A \in \mathbb{R}^{n \times d}$ and for all $x \in \mathbb{R}^d$, the inequality*

$$\|Ax\|_2 \leq \|\Pi_2 Ax\|_2 \leq \kappa \|Ax\|_2 \quad (4)$$

holds with probability $1 - \delta$, where

$$\kappa = O\left(\frac{\sqrt{r_1 d^3}}{\delta}\right). \quad (5)$$

For any $y \in \mathbb{R}^n$, the product $\Pi_2 y$ can be computed in $O(n \log r_1)$ time. Setting δ to a small constant, since $\sqrt{s} = r_1^3$ and $r_1 = O(d \log d)$, it follows that $\kappa = O(d^2 \log^{1/2} d)$. To get concentration, we need to independently bounded γ_j^2 in our proof which required $s = r_1^6$ to obtain the high probability result, and we to obtain the distortion $O(d^{1+\eta})$. From L_1 norm to L_2 norm, in the embedding matrix, Cauchy variable C_i becomes C_i^2 and the Cauchy distribution has density $p(x) = (1/\pi)(1/(1+x^2))$. Aiming to obtain L_2 embedding properties, we take C_i^2 in the tail inequality and obtain the FCT in L_2 norm which is our main result.

2. Preliminaries

Let $A \in \mathbb{R}^{n \times d}$ be an $n \times d$ input matrix, where we assume $n \gg d$ and A has full column rank. For $p \in [1, \infty]$, the L_p norm of a vector x is $\|x\|_p = (\sum_i |x_i|^p)^{1/p}$ and $\max_i |x_i|$ for $p = \infty$. Let $[n]$ denote the set $\{1, 2, \dots, n\}$. For matrices, we use the L_2 operator norm $\|A\|_2 = \sup_{\|x\|_2=1} \|Ax\|_2$ and the entrywise L_p norm $\|X\|_p = (\sum_{i,j} |X_{ij}|^p)^{1/p}$. I_n denotes the $n \times n$ identity matrix and c refers to a generic constant. The Cauchy distribution having density $p(x) = (1/\pi)(1/(1+x^2))$. The Cauchy distribution will factor heavily in our discussion and bounds for sums of Cauchy random variables' square will be used throughout. First, we give Maurer's inequality as follows.

Lemma 3 (see [7]). Let $X_i \geq 0$ be independent random variables, $\sum_i \mathbb{E}[X_i^2] < \infty$, and define $X = \sum_i X_i$. Then, for any $t > 0$,

$$\Pr[X \leq \mathbb{E}[X] - t] \leq \exp\left(\frac{-t^2}{2 \sum_i \mathbb{E}[X_i^2]}\right). \quad (6)$$

Here, we take X to be $X = \sum_{i \in [n]} \xi_i(1 - x_i)$.

Lemma 4 (see [6]). Let x_i be i.i.d. $(0,1)$ random variables with probability p and let $X = \sum_{i \in [n]} \xi_i x_i$, where $\xi_i \geq 0$, with $\sum_{i \in [n]} \xi_i = \xi$ and $\sum_{i \in [n]} \xi_i^2 \leq \xi^2 / \beta^2$. Then, for any $t > 0$,

$$\Pr[X \geq \xi(p + t)] \leq \exp\left(-\frac{\beta^2 t^2}{2(1-p)}\right). \quad (7)$$

3. Main Technical Results: The Fast Cauchy Transform of L_2 Norm

3.1. Two Tail Inequalities. In the section, we present two tail inequalities and the definition of well-conditioned basis in L_2 norm. Then we present the Fast Cauchy Transform in L_2 norm, which is analog of the fast Johnson-Lindenstrauss transform, and prove our main result Theorem 2 in Section 3.2.

Lemma 5 (upper-tail inequality). For $i \in [m]$, let C_i be m i.i.d. Cauchy random variables and $\gamma \geq 0$ with $\gamma = \sum_{i \in [m]} \gamma_i$. Let $X = \sum_{i \in [m]} \gamma_i C_i^2$. Then, for any $t > 0$,

$$\Pr[X > \gamma t] \leq O\left(\sqrt{\frac{m}{t}}\right). \quad (8)$$

Proof. For fixed $M > 0$ and defined $F_i = \{C_i^2 \leq M\}$ and $F = \bigcap_{i \in [m]} F_i$, then we have $F_i = \{C_i \leq \sqrt{M}\}$; note that $F \cap F_i = F$. Because of $\tan^{-1} x \leq x$, we have that

$$\begin{aligned} \Pr[F_i] &= \frac{2}{\pi} \tan^{-1}(\sqrt{M}) \\ &= 1 - \frac{2}{\pi} \tan^{-1}\left(\frac{1}{\sqrt{M}}\right) \geq 1 - \frac{2}{\pi \sqrt{M}}. \end{aligned} \quad (9)$$

By a union bound, $\Pr[F] \geq 1 - (2m/\pi M)$. Further, $\Pr[F | F_i] \Pr[F_i] = \Pr[F \cap F_i] = \Pr[F_i]$; hence, $\Pr[F | F_i] = \Pr[F] / \Pr[F_i]$. First, we need to bound $\mathbb{E}[C_i^2 | F]$,

$$\begin{aligned} \mathbb{E}[C_i^2 | F_i] &= \mathbb{E}[C_i^2 | F_i \cap F] \Pr[F | F_i] + \mathbb{E}[C_i^2 | F_i \cap \bar{F}] \Pr[\bar{F} | F_i] \\ &\geq \mathbb{E}[C_i^2 | F_i \cap F] \Pr[F | F_i]. \end{aligned} \quad (10)$$

Then,

$$\mathbb{E}[C_i^2 | F] \leq \frac{\mathbb{E}[C_i^2 | F_i]}{\Pr[F | F_i]} = \frac{\mathbb{E}[C_i^2 | F_i] \Pr[F_i]}{\Pr[F]}. \quad (11)$$

By using the pdf of a Cauchy variable,

$$\mathbb{E}[C_i^2 | F_i] = \frac{2(\sqrt{M}/\pi - (1/\pi) \tan^{-1}(\sqrt{M}))}{\Pr[F_i]}, \quad (12)$$

so

$$\begin{aligned} \mathbb{E}[C_i^2 | F] &\leq \frac{2((\sqrt{M}/\pi) - (1/\pi) \tan^{-1}(\sqrt{M}))}{\Pr[F]} \\ &\leq \frac{2((\sqrt{M}/\pi) - (1/\pi) \tan^{-1}(\sqrt{M}))}{(1 - 2m/\pi \sqrt{M})}. \end{aligned} \quad (13)$$

We conclude that

$$\begin{aligned} \mathbb{E}[X | F] &= \sum_{i \in [m]} \gamma_i \mathbb{E}[C_i^2 | F] \\ &\leq \frac{2\gamma((\sqrt{M}/\pi) - (1/\pi) \tan^{-1}(\sqrt{M}))}{(1 - 2m/\pi \sqrt{M})}. \end{aligned} \quad (14)$$

Because of Markov's inequality and $\Pr[X \geq \gamma t | \bar{F}] \leq 1$, we have

$$\begin{aligned} \Pr[X \geq \gamma t] &= \Pr[X \geq \gamma t | F] \Pr[F] + \Pr[X \geq \gamma t | \bar{F}] (1 - \Pr[F]) \\ &\leq \frac{(2/t)((\sqrt{M}/\pi) - (1/\pi) \tan^{-1}(\sqrt{M}))}{(1 - 2m/\pi \sqrt{M})} + \frac{2m}{\pi \sqrt{M}}. \end{aligned} \quad (15)$$

We set $M = mt$; then

$$\Pr[X \geq \gamma t] \leq O\left(\sqrt{\frac{m}{t}}\right). \quad (16)$$

□

Lemma 6 (lower-tail inequality). *For $i \in [r]$, let C_i be independent Cauchy random variables and with $\gamma = \sum_{i \in [m]} \gamma_i^2$ and $\sum_{i \in [m]} \gamma_i^4 \leq \gamma^2/\beta^2$. Let $X = \sum_{i \in [r]} \gamma_i^2 C_i^2$. Then, for any $t > 0$,*

$$\Pr[X \leq \gamma(1-t)] \leq \exp\left(-\frac{\beta^2 t^2}{3}\right). \quad (17)$$

Proof. To bound the lower tail, we use Lemma 3. By homogeneity, it suffices to prove the result for $\gamma = 1$. Let $Z_i = \gamma_i^2 \min(C_i^2, M)$. Clearly $Z_i \leq \gamma_i^2 C_i^2$ and defining $Z = \sum_i Z_i$, we have that $Z \leq X$ and $\Pr[X \leq 1-t] \leq \Pr[Z \leq 1-t]$. Thus, we have that

$$\begin{aligned} \Pr[Z \leq 1-t] &= \Pr[Z \leq \mathbb{E}[Z] - (\mathbb{E}[Z] - 1 + t)] \\ &\leq \exp\left(\frac{-(\mathbb{E}[Z] - 1 + t)^2}{2 \sum_i \mathbb{E}[Z_i^2]}\right), \end{aligned} \quad (18)$$

where the last inequality holds by Lemma 3 for $1-t < \mathbb{E}[Z]$. Using the distribution of half-Cauchy, we can verify that by choosing a fixed M to make sure $2 \int_0^{\sqrt{M}} x^2(1/\pi)(1/(1+x^2))dx = 1$, since $1 < M < 2$, $\mathbb{E}[Z_i] = \gamma_i^2$ and $\mathbb{E}[Z_i^2] \leq (3/2)\gamma_i^4$, so $\sum_i \mathbb{E}[Z_i] = 1$ and $\sum_i \mathbb{E}[Z_i^2] \leq (3/2)\sum_i \gamma_i^4 \leq (3/2)\beta^2$. It follows that $\Pr[Z \leq 1-t] \leq \exp(-t^2\beta^2/3)$ and the result follows. □

3.2. Definition of L_2 Well-Conditioned Basis and Construction of FCT in L_2 Norm. Clarkson et al. [6] constructed Π_1 as $\Pi_1 \equiv 4BC\tilde{H}$; let $\delta \in (0, 1]$ be a parameter governing the failure probability of our algorithm; we modified B and obtain Π_2 , where: $B \in \mathbb{R}^{r_1 \times 2n}$ has each column chosen independently and uniformly from the r_1 orthogonal standard basis vector for $\mathbb{R}^{r_1}$ and B_{ij} is chosen independently from a $(0,1)$ -distribution; for α sufficiently large, we will set the parameter $r_1 = \alpha d \log d/\delta$; $C \in \mathbb{R}^{2n \times 2n}$ is a diagonal matrix with diagonal entries chosen independently from a Cauchy distribution; and $\tilde{H} \in \mathbb{R}^{2n \times n}$ is a block-diagonal matrix comprised of n/s blocks along the diagonal. Each block is the $2s \times s$ matrix $G_s = [H_s; I_s]$, where I_s is the $s \times s$ identity matrix and H_s is the normalized Hadamard matrix. We will set $s = r_1^6$. The effects of $\tilde{H}$ here are to spread the weight of a vector, so that $\tilde{H}y$ has most entries that are not too small.

We describe L_2 norm well-conditioned basis and our main aim for constructing an L_2 well-conditioned basis is to make sure that we can tolerate the distortion of subspace embedding. To form an approximate basis for the range of A , we are led to faster algorithms for a range of related problems including low-rank matrix approximation [8, 9].

Definition 7 (L_2 well-conditioned basis, [3, 4]). A basis U for the range of A is L_2 -(α, β)-conditioned if $\|U\|_2 \leq \alpha$ and for all $X \in \mathbb{R}^d$, $\|x\|_2 \leq \beta\|Ux\|_2$.

Next we show the construction of L_2 well-conditioned basis U , which consists two steps: (a) let Π_2 be a matrix satisfying (4) and compute $\Pi_2 A$ and its QR-factorization $\Pi_2 A = QR$, where Q is an orthogonal matrix; (b) output $U = AR^{-1} = A(Q^T \Pi_2 A)^{-1}$. This structure is similar to the algorithm of [10] for computing an L_2 well-conditioned basis.

To obtain Theorem 2, we divide it into two propositions. To prove the upper bound, we use the existence of a $(d, 1)$ -conditioned basis U and apply this basis to show that $\|\Pi_2 Ux\|_2$ cannot expand too much. To prove the lower bound, we show that the inequality holds with high probability for a particular y ; then we use a suitable γ -net to obtain the result for all y .

Proposition 8. *With probability at least $1 - \delta$, for all $x \in \mathbb{R}^d$, $\|\Pi_2 Ax\|_2 \leq \kappa\|Ax\|_2$, where $\kappa = O(\sqrt{r_1 d^3/\delta})$.*

Proof. Let $U \in \mathbb{R}^{n \times d}$ be a L_2 -($d, 1$)-conditioned basis space of A , which implies that for some $Z \in \mathbb{R}^{d \times d}$ we have $A = UZ$. By the construction of U , for any $x \in \mathbb{R}^d$, $\|x\|_2 \leq \|Ux\|_2$ and so

$$\|\Pi_2 Ux\|_2 \leq \|\Pi_2 U\|_2 \|x\|_2 \leq \|\Pi_2 U\|_2 \|Ux\|_2. \quad (19)$$

Thus our main aim is to show that $\|\Pi_2 U\|_2 \leq \kappa$. We have

$$\begin{aligned} \|\Pi_2 U\|_2 &= 4\|BC\tilde{H}U\|_2 \\ &= 4 \sum_{j \in d} \|BC\tilde{H}U^{(j)}\|_2 = 4 \sum_{j \in d} \|BC\tilde{U}^{(j)}\|_2, \end{aligned} \quad (20)$$

where $\tilde{U} = \tilde{H}U$. We need to bound $\|\tilde{U}\|_2^2$. For any vector $y \in \mathbb{R}^n$, we represent y by n/s block of size s , so $z_i \in \mathbb{R}^s$ and $y^T = [z_1^T, z_2^T, \dots, z_{n/s}^T]$. Recall that $G_s = [H_s; I_s]$ and $\|G_s\|_2 = \sqrt{2}$. Then,

$$\|\tilde{H}y\|_2^2 = \sum_{i \in n/s} \|G_s z_i\|_2^2. \quad (21)$$

It follows that

$$\|\tilde{H}y\|_2^2 \leq \sum_{i \in n/s} \|G_s\|_2^2 \|z_i\|_2^2 = 2\|y\|_2^2. \quad (22)$$

Applying this to $y \in U^{(j)}$ for $j \in [d]$,

$$\|\tilde{U}\|_2^2 = \sum_{j \in [d]} \|\tilde{U}^{(j)}\|_2^2 \leq 2\|U\|_2^2 \leq 2d^2. \quad (23)$$

The (i, j) entry of $BC\tilde{U}$ is $\sum_k B_{ik} C_{kk} \tilde{U}_{kj}$; here we take $\gamma_{ij} = \sum_k B_{ik} \tilde{U}_{kj}$. So,

$$\|BC\tilde{U}\|_2^2 = \sum_{i,j} \left(\sum_k B_{ik} C_{kk} \tilde{U}_{kj} \right)^2 = \sum_{i,j} (\gamma_{ij} \tilde{C}_{ij})^2 = \sum_{i,j} \gamma_{ij}^2 \tilde{C}_{ij}^2. \quad (24)$$

Here $\tilde{C}_{i,j}$ are dependent Cauchy random variables and $\gamma = \sum_{i,j} \gamma_{ij}^2$. Using B as a standard orthogonal basis, we obtain

$$\gamma = \sum_{i,j} \gamma_{ij}^2 = \sum_{ij} \left(\sum_k B_{ik} \tilde{U}_{kj} \right)^2 = \|\tilde{U}\|_2^2. \quad (25)$$

Hence, we can apply Lemma 5, and $m = r_1 d$ to obtain,

$$\Pr \left[\|BC\tilde{U}\|_2^2 \geq t\gamma \right] \leq O \left(\sqrt{\frac{m}{t}} \right). \quad (26)$$

Setting $\sqrt{m/t}$ to δ , then we can get $t = O((1/\delta^2)r_1 d)$. Thus, with probability at least $1 - \delta$,

$$\|BC\tilde{H}\|_2 = O \left(\frac{\sqrt{r_1 d^3}}{\delta} \right). \quad (27)$$

To prove the second section, first we will show a result for fixed $y \in \mathbb{R}^n$; we will describe in the next lemma. $\square$

Lemma 9. Let $\Pr[\|\Pi_2 y\|_2 < 2\|y\|_2] \leq \exp(-r_1^3/64) + \exp(-s^{1/2}/16r_1^2 + \log r_1)$.

Proof. We represent any vector $y \in \mathbb{R}^n$ by its n/s blocks of size s , so $z_i \in \mathbb{R}^s$ and $y^T = [z_1^T, z_2^T, \dots, z_{n/s}^T]$. Let $g = \tilde{H}y$, $g = [G_s z_1; G_s z_2, \dots, G_s z_{n/s}]$. We have that $\|g\|_2^2 = \sum_i \|G_s z_i\|_2^2 = 2 \sum_i \|z_i\|_2^2 = 2\|y\|_2^2$ and

$$\begin{aligned} \|g\|_1 &= \sum_i \|G_s z_i\|_1 \geq \frac{1}{2} s^{1/4} \sum_i \|z_i\|_2 \\ &\geq \frac{1}{2} s^{1/4} \left(\sum_i \|z_i\|_2^2 \right)^{1/2} = \frac{1}{2} s^{1/4} \|y\|_2. \end{aligned} \quad (28)$$

We can conclude that $\|g\|_1 \geq (1/2\sqrt{2})s^{1/4}\|g\|_2$, so $\|g\|_2^2 \geq (1/8)s^{1/4}\|g\|_4^2$. To analyze $4\|BCg\|_2^2$ ($\Pi_2 y = BCg$, $g = \tilde{H}y$), we could have $(BCg)_j = \sum_i B_{ji} C_{ii} g_i$. Here we take $\gamma_j = \sum_{i=1}^n B_{ji} g_i$; then $(BCg)_j = \gamma_j \tilde{C}_j$ ($\tilde{C}_j$ is an independent Cauchy random variable), so

$$\|BCg\|_2^2 = \sum_j \left(\sum_i B_{ji} C_{ii} g_i \right)^2 = \sum_j \gamma_j^2 \tilde{C}_j^2. \quad (29)$$

To apply Lemma 6, we need to bound $\sum_j \gamma_j^2$ and $\sum_j \gamma_j^4$. First, because B is standard orthogonal, then

$$\sum_j \gamma_j^4 = \sum_j \left(\sum_i B_{ji} g_i \right)^2 = \sum_j \sum_i B_{ji}^2 g_i^2 = \|g\|_2^2. \quad (30)$$

To bound $\sum_j \gamma_j^4$, we will show that γ_j^2 is nearly uniform. Because B_{ji} and B_{jk} are independent for $i \neq k$, we can use Lemma 4 with $1 - p = 1 - 1/r_1$, with $\xi_i = g_i^2$ and $\sum_i \xi_i = \|g\|_2^2$; setting $t = 1/r_1$ in Lemma 4,

$$\begin{aligned} \Pr \left[\gamma_j^2 \geq (p+t) \cdot \left(\sum_i \xi_i \right) \right] \\ \leq \exp \left(-\frac{t^2}{2} \cdot \frac{\|g\|_2^4}{\|g\|_4^4} \right) \leq \exp \left(-\frac{s^{1/2}}{16r_1^2} \right). \end{aligned} \quad (31)$$

By a union bound, none of the γ_j^2 exceed $2\|g\|_2^2/r_1^2$ with probability at most $r_1 \exp(-s^{1/2}/16r_1^2)$. We assume this high probability event, so we get $\sum_j \gamma_j^4 \leq 4\|g\|_2^4/r_1^3$. We apply Lemma 4 with $t = 3/4$ and $\beta^2 = r_1^3/4$ to obtain $\Pr[\|BCg\|_2 \leq \|g\|_2 \cdot (1/2)] \leq \exp(-r_1^3/64)$. By a union bound, with probability at least $1 - \exp(-r_1^3/64) - \exp((-s^{1/2}/16r_1^2) + \log r_1)$. $\square$

Proposition 10. Assume Proposition 8 holds. Then, for all $x \in \mathbb{R}_d$, $\|\Pi_2 Ax\|_2 \geq \|Ax\|_2$ holds with probability at least

$$\begin{aligned} 1 - \exp(d \log(2d\kappa)) \\ \times \left(\exp \left(-\frac{r_1^3}{64} \right) + \exp \left(-\frac{s^{1/2}}{16r_1^2} + \log r_1 \right) \right). \end{aligned} \quad (32)$$

Proof. The proposition follows by putting a γ -net Γ on the range of A (observe that the range of A has dimension at most d). Specifically, let L be any fixed d dimensional subspace of $\mathbb{R}^n$ (L is the range of A). Consider the γ -net on L with cubes of side γ/d ; there are $(2\gamma/d)^d$ such cubes required to cover the hypercube $\|y\|_\infty$; and for any two points y_1, y_2 inside the same cube, $\|y_1 - y_2\|_2 \leq \gamma/\sqrt{d}$. From each of the γ/d -cubes, select a fixed generic point y^* and $\|y^*\|_2 = 1$. In order to make sure that y^* belongs to $(2\gamma/d)^d$, by a union bound and Lemma 9,

$$\begin{aligned} \Pr \left[\frac{\min_{y^*} \|\Pi_2 y^*\|_2}{\|y^*\|_2} < \frac{1}{2} \right] \\ \leq \left(\frac{2d}{\gamma} \right)^d \left(\exp \left(-\frac{r_1^3}{64} \right) + \exp \left(-\frac{s^{1/2}}{16r_1^2} + \log r_1 \right) \right). \end{aligned} \quad (33)$$

We will condition on the high probability event that $\|\Pi_2 y\|_2 \geq \|y^*\|_2/2$ for all y^* . For any $y \in L$ with $\|y\|_2 = 1$, let y^* denote the representative point which is in the same cube of y ($\|y^*\|_2 = 1$). Then $\|y - y^*\|_2 \leq \gamma/\sqrt{d}$. Considering

$$\begin{aligned} \|\Pi_2 y\|_2 \\ = \|\Pi_2 y^* + \Pi_2(y - y^*)\|_2 \\ \geq \|\Pi_2 y^*\|_2 - \|\Pi_2(y - y^*)\|_2 \geq 2\|y^*\|_2 - \kappa\|y - y^*\|_2, \end{aligned} \quad (34)$$

by choosing $\gamma = 1/\kappa$, we have $\|\Pi_2 y\|_2 \geq \|y\|_2$, with probability at least

$$\begin{aligned} 1 - \exp(d \log(2\kappa d)) \\ \times \left(\exp \left(-\frac{r_1^3}{64} \right) + \exp \left(-\frac{s^{1/2}}{16r_1^2} + \log r_1 \right) \right). \end{aligned} \quad (35)$$

Since $s^{1/2} = r_1^3$, by choosing $r_1 = \alpha d \log(d/\delta)$ for large enough α , the final probability of failure is at most 2δ . And also since $\kappa = O(\sqrt{r_1 d^3}/\delta)$ and $r_1 = \alpha d \log(d/\delta)$, by the inequality (1), we could get the distortion $O(d^{1+\eta})$ for an arbitrarily small constant $\eta > 0$. $\square$

4. Conclusion

In the paper, we extend selections of the embedding matrix and apply the well-conditioned basis in L_2 norm. Based on Maurer's inequality, we obtain the tail inequality about Cauchy variables. By the specified matrix Π_2 , we present Fast Cauchy Transform in L_2 norm.

Acknowledgments

The work is supported in part by the National Science Foundation of China under Grants no. 61271014, no. 61072118, and no. 61170159; by Science Project of National University of Defense Technology JC120201, also by National Natural Science Foundation of Hunan Province (China) 13JJ2001. The authors would like to thank anonymous referees for helpful suggestions and they would also like to thank Hui Zhang and Tao Sun for many thoughtful discussions.

References

- [1] W. B. Johnson and J. Lindenstrauss, "Extensions of Lipschitz mappings into a Hilbert space," *Contemporary Mathematics*, vol. 26, pp. 189–206, 1984.
- [2] H. Zhang and L. Z. Cheng, "New bounds for Circulant Johnson-Lindenstrauss embeddings," Accepted by *Communications in Mathematical Sciences*, <http://arxiv.org/abs/1308.6339>.
- [3] K. L. Clarkson, "Subgradient arid sampling algorithms for ℓ_1 regression," in *Proceedings of the 16th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA '05)*, pp. 257–266, January 2005.
- [4] A. Dasgupta, P. Drineas, B. Harb, R. Kumar, and M. W. Mahoney, "Sampling algorithms and coresets for ℓ_p regression," *SIAM Journal on Computing*, vol. 38, no. 5, pp. 2060–2078, 2009.
- [5] C. Sohler and D. P. Woodruff, "Subspace embeddings for the ℓ_1 -norm with applications," in *Proceedings of the 43rd ACM Symposium on Theory of Computing (STOC '11)*, pp. 755–764, June 2011.
- [6] K. L. Clarkson, P. Drineas, M. Magdon-Ismail, X. R. Meng, M. W. Mahoney, and D. P. Woodruff, "The fast cauchy transform and faster robust linear regression," in *Proceeding of the ACM-SIAM Symposium on Discrete Algorithms*, 2013.
- [7] A. Maurer, "A bound on the deviation probability for sums of non-negative random variables," *Journal of Inequalities in Pure and Applied Mathematics*, vol. 4, no. 1, article 15, 2003.
- [8] P. Drineas, M. W. Mahoney, S. Muthukrishnan, and T. Sarlós, "Faster least squares approximation," *Numerische Mathematik*, vol. 117, no. 2, pp. 219–249, 2010.
- [9] M. W. Mahoney, *Randomized Algorithms for Matrices and Data. Foundations and Trends in Machine Learning*, NOW Publishers, Boston, Mass, USA, 2011.
- [10] P. Drineas, M. Magdon-Ismail, M. W. Mahoney, and D. P. Woodruff, "Fast approximation of matrix coherence and statistical leverage," in *Proceedings of the 29th International Conference on Machine Learning*, 2012.