Lean Walsh Transform

Edo Liberty

15th March 2007

1 informal intro

We show an orthogonal matrix A of size $d^{\log_4 3} \times d$ ($\alpha = \log_4 3$) which is applicable in time $O(d)$. By applying a random sign change matrix S to the incoming vectors we show that if $\mathbf{y} = d^{-\alpha/2} AD\mathbf{x}$ then $E\|\mathbf{y}\|^2 = \|\mathbf{x}\|^2$.

Alas if $\mathbf{x}$ is very sparse (only two non zeros) the distortion of $\mathbf{y}$ relative to $\mathbf{x}$ is large with polynomial probability in d. Namely $\Pr\left(\|\mathbf{y}\|^2 \geq \|\mathbf{x}\|^2(1+1/3)\right) \geq \frac{\log(d)}{d}$

If x the largest entree in x is not too large i.e $\|\mathbf{x}\|_\infty \leq \frac{\|\mathbf{x}\|}{\sqrt{m}}$ for some integer m (x has at least m non zero entrees) we seem to avoid the sparse vectors and in practice perform very well.

At this point it seems that in theory the best we can hope for is $m = O(log(n) + log(d))$.

2 Lean Walsh transform

In this section we describe a matrix A which we termed the Lean Walsh transform which has some nice properties. The most interesting ones are that a) A is dense $\forall i, j\ |A(i,j)| = 1$. b) A is of size $d^\alpha \times d$. c) The rows of A are orthogonal, and d) it can be applied to a vector of length d in $O(d)$ operations. This shows that it is possible to project a vector of size d on $d^\alpha = d^{log_4 3}$ *dense* vectors in $O(d)$ time. We also show some statistical behavior of the mutual projections of the columns of A which might be important for proving the result.

We define a lean Walsh transform of order k recursively by

$$A_1 = \begin{pmatrix} 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & -1 & -1 & 1 \end{pmatrix} \quad (1)$$

$$A_k = A_1 \otimes A_{k-1} \quad (2)$$

Where $\otimes$ is the Kronecker product.

The lean Walsh transform A_k is of size $3^k \times 4^k$. since we need to multiply A_k from the right by $\mathbf{x}$, $4^k = d$ (we assume w.l.o.g that d is a power of 4.) $A_{\log_4 d} \in$

$\mathbb{R}^{d^\alpha \times d}$. From this point on $A \equiv A_{\log_4 d}$, $k = \log_4 d$, and $\alpha = \log_4 3 \approx 0.8$. Also w.l.o.g we set $\|\mathbf{x}\|^2 = 1$.

Lemma 2.1 *The time complexity of applying A to any vector $\mathbf{x} \in \mathbb{R}^d$ is $O(d)$.*

Proof Denote $T(d)$ as the time to apply A (of the right size) to d.

$$T(d) = 4T(d/4) + 12d^{\log_4 3} \rightarrow T \in O(d). \tag{3}$$

Lemma 2.2 *The rows of A are orthogonal.*

Proof By induction on the construction of A.

Lemma 2.3 *For any column $\mathbf{c}_i$ of A, and for $0 \leq t \leq k$ there are $\binom{k}{t}3^t$ columns $\mathbf{c}_j$ such that $\langle \mathbf{c}_i, \mathbf{c}_j \rangle = (-1)^t 3^{k-t}$.*

Proof Since A_k has 4^k columns, each one of them is uniquely defined by a sequence $s_i \in \{0,1,2,3\}^k$ which corresponds to the representation of the number i in base 4. Let us denote by s_i and s_j sequences for $\mathbf{c}_i$ and $\mathbf{c}_j$. Moreover we define as $a_{1,2,3,4}$ the columns of A_1, i.e $a_1 = (1,1,1)^T$, $a_2 = (1,-1,-1)^T$, $a_3 = (-1,1,-1)^T$, $a_4 = (-1,-1,1)^T$. Notice that $\mathbf{c}_i = a_{s_i(1)} \otimes a_{s_i(2)} \otimes \ldots \otimes a_{s_i(k)}$. The dot product of the columns $\mathbf{c}_i$ and $\mathbf{c}_j$ is then the product of the dot product of the corresponding columns of A_1. Since the inner product of two columns of A_1 is either 3 (if they are the same) or -1 (if they are different) we have that $\langle \mathbf{c}_i, \mathbf{c}_j \rangle = (-1)^{\|s_i,s_j\|_H} 3^{k-\|s_i,s_j\|_H}$ where $\|\cdot,\cdot\|_H$ denotes the Hamming distance. We define $\mathbf{c}_i$ and $\mathbf{c}_j$ as neighbors of degree t if $\|s_i, s_j\|_H = t$. Let us count the number of degree t neighbors for each vector. We choose t positions out of the k possible ones and for each we have three options for a change. We therefor have $\binom{k}{t}3^t$ neighbors of degree t for $0 \leq t \leq k$.

Lemma 2.4 *The expected inner product square for two columns is small.*

$$E_{i \neq j} \langle \mathbf{c}_i, \mathbf{c}_j \rangle^2 < d^\alpha \tag{4}$$

Proof Due to lemma 2.3 we have that:

$$\begin{aligned} E_{i \neq j} \langle \mathbf{c}_i, \mathbf{c}_j \rangle^2 &< \frac{1}{4^k} \sum_{t=0}^{k} \binom{k}{t} 3^t 2^{2k-2t} && (5) \\ &= \frac{3^{2k}}{4^k} \sum_{t=0}^{k} \binom{k}{t} (1/3)^t && (6) \\ &= 3^k = d^\alpha && (7) \\ & && (8) \end{aligned}$$

Notice that $E_{i \neq j} \frac{\langle \mathbf{c}_i, \mathbf{c}_j \rangle^2}{\|\mathbf{c}_i\|^2 \|\mathbf{c}_j\|^2} \leq d^{-\alpha}$ We will use this result in section 4.

3 unbiased linear mapping

In this section we apply A to a vector $\mathbf{x}$ through a diagonal random sign changing matrix $S = diag(\mathbf{s})$, $\mathbf{s}(i) = \pm 1$ with probability $1/2$. And normalize appropriately, $\mathbf{y} = \frac{1}{d^{\alpha/2}} AD\mathbf{x}$. We show that in expectancy the length of $\mathbf{y}$ is equal to the length of $\mathbf{x}$. Alas since A is deterministic we show that an adversary can arrange a sparse vector that will w.h.p be unmodified by the sign change and create a large distortion. This problem is NOT fixed by applying a permutation to the incoming vector.

Lemma 3.1 *Given* $\mathbf{y} = \frac{1}{d^{\alpha/2}} AD\mathbf{x}$ *then* $E\|\mathbf{y}\|^2 = \|\mathbf{x}\|^2$.

Proof We denote the by $\mathbf{c}_i$ and $\mathbf{c}_j$ the i'th and j'th columns of A.

$$\begin{aligned}
\mathbf{y} &= \frac{1}{d^{\alpha/2}} \sum_{i=1}^{d} \mathbf{c}_i s(i) \mathbf{x}_i && (9)\\
\|\mathbf{y}\|^2 &= \frac{1}{d^{\alpha}} \left(\sum_{i=1}^{d} \mathbf{c}_i^T \mathbf{s}(i)\mathbf{x}(i) \right) \left(\sum_{j=1}^{d} \mathbf{c}_j \mathbf{s}(j)\mathbf{x}(j) \right) && (10)\\
&= \frac{1}{d^{\alpha}} \sum_{i=1}^{d} \mathbf{c}_i^T \mathbf{c}_i \mathbf{x}(i)^2 + \frac{1}{d^{\alpha}} \sum_{i=1}^{d} \sum_{j\neq i}^{d} \mathbf{c}_i^T \mathbf{c}_j \mathbf{s}(i)\mathbf{s}(j)\mathbf{x}(i)\mathbf{x}(j) && (11)\\
&= \sum_{i=1}^{d} \mathbf{x}(i)^2 + \sum_{i=1}^{d} \sum_{j\neq i}^{d} \frac{\mathbf{c}_i^T \mathbf{c}_j}{\|c_i\|\|c_j\|} \mathbf{s}(i)\mathbf{s}(j)\mathbf{x}(i)\mathbf{x}(j) && (12)\\
&= \|\mathbf{x}\|^2 \left(1 + \sum_{i=1}^{d} \sum_{j\neq i}^{d} \frac{\mathbf{c}_i^T \mathbf{c}_j}{\|c_i\|\|c_j\|} \mathbf{s}(i)\mathbf{s}(j) \frac{\mathbf{x}(i)\mathbf{x}(j)}{\|\mathbf{x}\|^2} \right) && (13)\\
E\left(\|\mathbf{y}\|^2\right) &= \|\mathbf{x}\|^2. && (14)
\end{aligned}$$

By linearity of expectation and $E(\mathbf{s}(i)) = E(\mathbf{s}(j)) = 0$.

4 Bounding the distortion if x is not too sparse

this section is currently proved under some independence conditions which do not hold. Proofs need to be modified.

In this section we will see that the distortion for n points is less then ϵ with probability at least $1/4$ if each of the n vectors is not too sparse. By that we mean that $\|\mathbf{x}\|_\infty \leq 1/\sqrt{m}$. We show that this holds if $m \geq \frac{7\log^{3/2}(n)}{\epsilon^2 d^{\alpha/2}}$. Notice that in case that $d = O(\log^{3/\alpha}(n))$ we have that $m = \frac{c}{\epsilon^2}$ for some c.

Lemma 4.1 *The worst case (most prone to distortion) vector x such that $\|\mathbf{x}\|_\infty \leq 1/\sqrt{m}$ is a vector with m entrees of size $1/\sqrt{m}$ and the rest zero.*

Proof Don't know how to proof this yet.

We turn to bound the distortion probability for the worst case vectors. Let $w_{i,j} \equiv \frac{c_i^T c_j}{\|c_i\| \cdot \|c_j\|}$ and $\mathbf{s}(i)\mathbf{s}(j) \equiv s_{i,j}$. The sum in equation 13 for these vectors becomes $\sum_{i=1}^{m} \sum_{j \neq i}^{m} \frac{1}{m} w_{i,j} \mathbf{s}(i,j)$.

Lemma 4.2 *Let $w_{i,j} \equiv \frac{c_i^T c_j}{\|c_i\| \cdot \|c_j\|}$. Assuming that $w_{i,j}$ are drown i.i.d (which is false) with probability $1/2$ for n vectors we have that:*

$$\sum_{i=1}^{m} \sum_{j \neq i}^{m} w_{i,j}^2 < \frac{\sqrt{12 \log(2n)} m}{d^{\alpha/2}} \tag{15}$$

Proof We define $\mu = E(w_{i,j}^2) \leq d^{-\alpha}$ using Chernoff's bound we have that:

$$\Pr\left(\sum_{i=1}^{m} \sum_{j \neq i}^{m} w_{i,j} > (1+t) m^2 \mu \right) < e^{-\mu m^2 t^2 / 3} \tag{16}$$

We look for the value of t that would make this equation hold for n points with probability $1/2$.

$$n e^{-\mu m^2 t^2 / 3} \leq 1/2 \tag{17}$$

$$t \geq \frac{\sqrt{3 log(2n)}}{m \sqrt{\mu}} \tag{18}$$

$$\tag{19}$$

Since $t \geq 1$ we have that with probability $1/2$ for n points.

$$\begin{aligned} \sum_{i=1}^{m} \sum_{j \neq i}^{m} w_{i,j} &< 2 t m^2 \mu && (20) \\ &\leq 2 \frac{\sqrt{3 log(2n)}}{m \sqrt{\mu}} m^2 \mu && (21) \\ &\leq \frac{\sqrt{12 \log(2n)} m}{d^{\alpha/2}} && (22) \end{aligned}$$

Lemma 4.3 *With probability at least $1/4$*

$$\sum_{i=1}^{d} \sum_{j \neq i}^{d} \frac{1}{m} w_{i,j} s_{i,j} \leq \epsilon \tag{23}$$

$$\text{for } m \geq \frac{7 \log^{3/2}(2n)}{\epsilon^2 d^{\alpha/2}} \tag{24}$$

Proof Using the Hoeffding-Chernoff inequality we have that

$$Pr\left(\sum_{i=1}^{d}\sum_{j\neq i}^{d}\frac{1}{m}w_{i,j}s_{i,j}\geq\epsilon\right) \leq e^{-\frac{\epsilon^2}{2\sum_{i=1}^{d}\sum_{j\neq i}^{d}\frac{1}{m^2}w_{i,j}^2}} \quad (25)$$

$$\leq e^{-\frac{\epsilon^2}{2\frac{1}{m^2}\frac{\sqrt{12\log(2n)}m}{d^{\alpha/2}}}} \text{ w.p } 1/2 \quad (26)$$

$$= e^{-\frac{\epsilon^2 m d^{\alpha/2}}{\sqrt{48\log(2n)}}} \quad (27)$$

We now union bound the probability of failure for n point to get the bound on m.

$$ne^{-\frac{\epsilon^2 m d^{\alpha/2}}{\sqrt{48\log(2n)}}} \leq 1/2 \quad (28)$$

$$\frac{\epsilon^2 m d^{\alpha/2}}{\sqrt{48\log(2n)}} \geq \log(2n) \quad (29)$$

$$m \geq \frac{7\log^{3/2}(2n)}{\epsilon^2 d^{\alpha/2}} \quad (30)$$

Since we succeed with probabilty 1/2 twice the overall success probability is 1/4.

Notice that for large dimensions the term $\frac{\log^{3/2}(2n)}{d^{\alpha/2}}$ is smaller then 1. Therefore the assertion of lemma 4.3 is that we succeed with probability 1/4 for a constant $m = O(\epsilon^{-2})$.

5 Bounding the distortion take 2

Let us try to bound the sum

$$\sum_{i=1}^{m}\sum_{j\neq i}^{m}\tfrac{1}{m}w_{i,j}\mathbf{s}(i)\mathbf{s}(j) \quad (31)$$

$$= \tfrac{2}{m}\sum_{i=1}^{m}s(i)\left(\sum_{j=1}^{i-1}w_{i,j}\mathbf{s}(j)\right) \quad (32)$$

In this point we apply Hoeffding's inequality.

Lemma 5.1 *With probability* $1/2$ *we have for every* $1 \leq i \leq d$ *and* n *points that:*

$$\sum_{j=1}^{i-1}w_{i,j}^2 \leq i\sqrt{12\log(2n+1)\mu} \quad (33)$$

Proof By Chernoff and the fact that the variables w_{i,j_1} and w_{i,j_2} are independent we have that:

$$\Pr\left(\sum_{j=1}^{i-1}w_{i,j}^2 > (1+t)\mu i\right) \leq e^{-\mu i t^2/3} \quad (34)$$

We sum up for all i and multiply by n the failure probability to union bound for all i and n vectors.

$$\begin{aligned}
n\sum_{i=1}^{d} e^{-\mu i t^2/3} &= n\sum_{i=1}^{d}\left(e^{-\mu t^2/3}\right)^i && (35)\\
&< n\frac{e^{-\mu t^2/3}}{1-e^{-\mu t^2/3}} && (36)\\
&= \frac{n}{e^{\mu t^2/3}-1} < \frac{1}{2} && (37)\\
\mu t^2/3 &> \log(2n+1) && (38)\\
t &> \sqrt{\frac{3\log(2n+1)}{\mu}} > 1 && (39)
\end{aligned}$$

Since $2t > 1+t$ we have that with probability at least $1/2$

$$\begin{aligned}
\sum_{j=1}^{i-1} w_{i,j}^2 &< 2t\mu i && (40)\\
&\le 2\sqrt{\frac{3\log(2n+1)}{\mu}}\mu i && (41)\\
&= i\sqrt{12\log(2n+1)\mu} && (42)\\
& && (43)
\end{aligned}$$

Lemma 5.2 *For all i and n points with probability $1/2$ we have that*

$$\left(\sum_{j=1}^{i-1} w_{i,j}s(j)\right)^2 \le 7i\sqrt{\mu\log(2n+1)}(log(4n)) \quad (44)$$

Proof Using Hoeffding and lemma 5.1 we have that

$$\begin{aligned}
\Pr\left[(\sum_{j=1}^{i-1} w_{i,j}s(j))^2 > t_i^2\right] &\le 2e^{-\frac{t^2}{2\sum_{j=1}^{i-1} w_{i,j}^2}} && (45)\\
&\le 2e^{-\frac{t^2}{2i\sqrt{12\log(2n+1)\mu}}} && (46)
\end{aligned}$$

We need to bound this probability for all i and n points to be less then $1/2$

$$\begin{aligned}
2nme^{-\frac{t_i^2}{2i\sqrt{12\log(2n+1)\mu}}} &\le \frac{1}{2} && (47)\\
\frac{t_i^2}{2i\sqrt{12\log(2n+1)\mu}} &\ge \log(4nm) && (48)\\
t_i^2 &\ge 7i\sqrt{\mu\log(2n+1)}log(4nm) && (49)
\end{aligned}$$

Lemma 5.3

$$\sum_{i=1}^{m}\sum_{j\neq i}^{m}\frac{1}{m}w_{i,j}\mathbf{s}(i)\mathbf{s}(j) < \epsilon \tag{50}$$

Proof

$$\begin{aligned} & \textstyle\sum_{i=1}^{m}\sum_{j\neq i}^{m}\frac{1}{m}w_{i,j}\mathbf{s}(i)\mathbf{s}(j) && (51)\\ = \;& \textstyle\sum_{i=1}^{m} s(i)\left(\frac{2}{m}\sum_{j=1}^{i-1}w_{i,j}\mathbf{s}(j)\right) && (52)\end{aligned}$$

By the Hoeffding bound lemma 5.2

$$\begin{aligned} \Pr\left[\sum_{i=1}^{m} s(i)\left(\frac{2}{m}\sum_{j=1}^{i-1}w_{i,j}\mathbf{s}(j)\right) > \epsilon\right] &\leq e^{-\frac{\epsilon^2}{\frac{4}{m^2}\sum_{i=1}^{m}\left(\sum_{j=1}^{i-1}w_{i,j}\mathbf{s}(j)\right)^2}} && (53)\\ &\leq e^{-\frac{\epsilon^2}{\frac{4}{m^2}\sum_{i=1}^{m}7i\sqrt{\mu\log(2n+1)}log(4nm)}} && (54)\\ &\leq e^{-\frac{\epsilon^2}{28\sqrt{\mu\log(2n+1)}log(4nm)}} && (55)\end{aligned}$$

Bounding this probability for n points and substituting for $\mu = d^{-\alpha}$

$$\begin{aligned} e^{-\frac{\epsilon^2}{28\sqrt{\mu\log(2n+1)}log(4nm)}} &\leq \frac{1}{2} && (56)\\ \frac{\epsilon^2}{28\sqrt{\mu\log(2n+1)}log(4nm)} &\geq \log(2n) && (57)\\ d^{\alpha/2} &\geq \frac{28\log^3(4nm)}{\epsilon^2} && (58)\end{aligned}$$

6 appendix

6.1 Hoeffding-Chernoff inequality

Let $a_1, ..., a_n \in \mathbb{R}$ and let $s_1, ..., s_n$ be i.i.d. Rademacher random variables $\Pr(s_i = 1) = \Pr(s_i = -1) = 0.5$.

$$\Pr\left(\sum_{i=1}^{n} a_i s_i > t\right) < e^{-\frac{t^2}{2\sum_{i=1}^{n} a_i^2}} \tag{59}$$

6.2 Chernoff bound

Let $X_1, \ldots, X_n$ be i.i.d random variables in $[-1, 1]$ with $E(X_i) = \mu$.

$$\Pr\left(\sum_{i=1}^{n} X_i > (1+t)n\mu\right) \leq e^{-n\mu t^2/3} \tag{60}$$