Handout for lecture: Fast Random Projection

Edo Liberty*

1 About this handout

This handout contains the background information and the main lemmas needed in order to follow the result described in the paper "Fast Dimension Reduction Using Rademacher Series on Dual BCH Codes" which was a joint work with Nir Ailon from Google research (Institute for Advanced Study at the time). Many thanks also to Mark W. Tygert.

2 Notation and preliminaries

In what follows we use d to denote the original dimension and $k < d$ the target (reduced) dimension. The input vector will be $x = (x_1, \ldots, x_d)^T \in \ell_2^d$. Since we only consider linear reductions we will assume without loss of generality that $\|x\|_2 = 1$.

We use ℓ_p^d to denote d dimensional real space equipped with the norm $\|x\| = \|x\|_p = \left(\sum_{i=1}^d |x_i|^p\right)^{1/p}$, where $1 \le p < \infty$ and $\|x\|_\infty = \max\{|x_i|\}$. The dual norm index q is defined by the solution to the equation $1/q + 1/p = 1$. We remind the reader that $\|x\|_p = \sup_{\substack{y \in \ell_q^d \\ \|y\|=1}} x^T y$. For a real $k \times d$ matrix A, the matrix norm $\|A\|_{p_1 \to p}$ is defined as the operator norm of $A : \ell_{p_1}^d \to \ell_p^k$ or:

$$\|A\|_{p_1 \to p} = \sup_{\substack{x \in \ell_{p_1}^d \\ \|x\|=1}} \|Ax\|_p = \sup_{\substack{y \in \ell_q^k \\ \|y\|=1}} \sup_{\substack{x \in \ell_{p_1}^d \\ \|x\|=1}} y^T A x \ .$$

Definition 2.1 *A distribution $\mathcal{D}(d,k)$ on $k \times d$ real matrices ($k \le d$) has the Johnson-Lindenstrauss property (JLP) with respect to a norm index p if for any unit vector $x \in \ell_2^d$ and $0 \le \varepsilon < 1/2$,*

$$\Pr_{A \sim \mathcal{D}_{d,k}} [1 - \varepsilon \le \|Ax\|_p \le 1 + \varepsilon] \ge 1 - c_1 e^{-c_2 k \varepsilon^2} \quad (1)$$

for some global $c_1, c_2 > 0$.

3 Walsh-Hadamard matrix

Recall that the Walsh-Hadamard matrix H_d is a $d \times d$ orthogonal matrix with $H_d(i,j) = 2^{-d/2}(-1)^{\langle i,j \rangle}$ for all $i, j \in [0, d-1]$, where $\langle i, j \rangle$ is the dot product (over $\mathbb{F}_2$) of i, j viewed as $(\log d)$-bit vectors.

*Yale University, New Haven CT (edo.liberty@gmail.com)

The matrix encodes the Fourier transform over the binary hypercube. It is well known that applying the Walsh-Hadamard transform ($x \mapsto H_d x$) can be completed in time $O(d \log d)$ for any $x \in \ell_2^d$, and that the mapping is isomorphic.

We will also use the fact that the Walsh-Hadamard matrix can be recursively constructed as follows

$$H_1 = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}, \quad H_q = \begin{pmatrix} H_{q/2} & H_{q/2} \\ H_{q/2} & -H_{q/2} \end{pmatrix} \tag{2}$$

4 Tools from Banach Spaces

The following is known as an interpolation theorem in the theory of Banach spaces. For a proof, refer to [1].

Theorem 4.1 Riesz-Thorin *Let A be an $m \times d$ real matrix, and assume $\|A\|_{p_1 \to r_1} \le C_1$ and $\|A\|_{p_2 \to r_2} \le C_2$ for some norm indices p_1, r_1, p_2, r_2. Let λ be a real number in the interval $[0, 1]$, and let p, r be such that $1/p = \lambda(1/p_1) + (1 - \lambda)(1/p_2)$ and $1/r = \lambda(1/r_1) + (1 - \lambda)(1/r_2)$. Then $\|A\|_{p \to r} \le C_1^\lambda C_2^{1-\lambda}$.*

Theorem 4.2 [Hausdorff-Young] *For norm index $1 \le p \le 2$, $\|H\|_{p \to q} \le d^{-1/p+1/2}$, where q is the dual norm index of p.*

(The theorem is usually stated with respect to the Fourier operator for functions on the real line or on the circle, and is a simple application of Riesz-Thorin by noticing that $\|H\|_{2 \to 2} = 1$ and $\|H\|_{1 \to \infty} = d^{-1/2}$.)

5 Concentration of *Rademacher* random variables

Let M be a real $m \times d$ matrix, and let $z \in \mathbf{R}^d$ be a random vector with each z_i distributed uniformly and independently over $\{\pm 1\}$. The random vector $Mz \in \ell_p^m$ is known as a *Rademacher* random variable. A nice exposition of concentration bounds for Rademacher variables is provided in Chapter 4.7 of [2] for more general Banach spaces. For our purposes, it suffices to review the result for finite dimensional ℓ_p space. Consider the norm $Z = \|Mz\|_p$ (we say that Z is the norm of a Rademacher random variable in ℓ_p^d corresponding to M). We associate two numbers with Z,

- The deviation σ, defined as $\|M\|_{2 \to p}$, and
- a median μ of Z.

Theorem 5.1 *For any $t \ge 0$,*

$$\Pr[|Z - \mu| > t] \le 4e^{-t^2/(8\sigma^2)} \, .$$

The theorem is a simple consequence of a theorem of Talagrand (Chapter 1, [2]) on measure concentration of functions on $\{-1, +1\}^d$ which can be extended to convex functions on ℓ_2^d with bounded Lipschitz norm.

6 Tools from Error Correcting Codes

Definition 6.1 *A matrix $A \in \mathbf{R}^{m\times d}$ is a* code *matrix if every row of A is equal to some row of H multiplied by $\sqrt{d/m}$.*

The normalization is chosen so that columns have Euclidean norm 1.

Definition A code matrix A of size $m \times d$ is *a-wise independent* if for each $1 \le i_1 < i_2 < \ldots < i_a \le m$ and $(b_1, b_2, \ldots, b_a) \in \{+1, -1\}^a$, the number of columns $A^{(j)}$ for which $(A^{(j)}_{i_1}, A^{(j)}_{i_2}, \ldots, A^{(j)}_{i_a}) = m^{-1/2}(b_1, b_2, \ldots, b_a)$ is exactly $d/2^a$.

Lemma 6.2 *There exists a 4-wise independent code matrix of size $k \times f_{BCH}(k)$, where $f_{BCH}(k) = \Theta(k^2)$.*

The family of matrices is known as binary dual BCH codes of designed distance 5. Details of the construction can be found in [3].

References

[1] J. Bergh and J. Lofstrom. *Interpolation Spaces.* Springer-Verlag, 1976.

[2] M. Ledoux and M. Talagrand. *Probability in Banach Spaces: Isoperimetry and Processes.* Springer-Verlag, 1991.

[3] F. MacWilliams and N. Sloane. *The Theory of Error Correcting Codes.* North-Holland, 1983.