Lecture 8

Rounding for MAXQP; Dimension Reduction

8.1 Rounding MAXQP

Recall the setup from last time: we want to show that there exists a universal constant $K > 0$ such that for all $A \in \mathbb{R}^{m \times n}$,

$$\begin{aligned} \text{MAXQP}(A) &= \max_{x_i, y_j \in \{\pm 1\}} \sum A_{ij} x_i y_j \\ &\leq \sup_{u_i, v_j \in \mathbb{R}^{n+m},\, \|u_i\|, \|v_j\| \leq 1} \sum A_{ij} u_i \cdot v_j \;=: \text{SDP}(A) \\ &\leq \frac{1}{K} \text{MAXQP}(A). \end{aligned}$$

Our technique for this will be to start with an optimal vector solution to SDP(A), perform a (deterministic!) embedding of the vectors in a higher-dimensional space, and finally find a "good coordinate" to perform the rounding in the higher-dimensional space.

The embedding is defined as follows. For any $\vec{u} \in \mathbb{R}^d$, define $h(\vec{u}) \in \mathbb{R}^t$, such that $h(\vec{u})_i = \frac{\vec{g}_i \cdot \vec{u}}{\sqrt{t}} = \left(\frac{G\vec{u}}{\sqrt{t}}\right)_i$ for $i = 1, \ldots, t$. For $M > 0$ define

$$h^M(\vec{u}) = \begin{cases} h(\vec{u})_i & \text{if } |h(\vec{u})_i| \leq M/\sqrt{t}, \\ M/\sqrt{t} & \text{if } h(\vec{u})_i > M/\sqrt{t}, \\ -M/\sqrt{t} & \text{if } h(\vec{u})_i < -M/\sqrt{t}. \end{cases}$$

Then we have the following lemma:

Lemma 8.1. *For any $\vec{u}, \vec{v} \in \mathbb{R}^d$, with $\|\vec{u}\| = \|\vec{v}\| = 1$,*

(1) $h(\vec{u}) \cdot h(\vec{v}) = \vec{u} \cdot \vec{v}$.

(2) $\|h(\vec{u})\| = 1$.

(3) $\|h^M(\vec{u})\| \le 1$.

(4) $\|h(\vec{u}) - h^M(\vec{u})\| \le \frac{\sqrt{3}}{M}$.

We'll prove the lemma later, first let's use it to finish the proof of the theorem. The last property gives a precise trade-off between the size of M and the quality of the approximation of the vector $h(\vec{u})$ by its truncation $h^M(\vec{u})$. Using the lemma, we can write

$$
\begin{aligned}
&\mathrm{SDP}(A)\\
&= \sum_{i,j} A_{ij} \vec{u}_i \cdot \vec{v}_j\\
&\overset{(1)}{=} \sum_{i,j} A_{ij} h(\vec{u}_i) \cdot h(\vec{v}_j)\\
&= \sum_{i,j} A_{ij} h^M(\vec{u}_i) \cdot h^M(\vec{v}_j) + \sum_{i,j} A_{ij} h^M(\vec{u}_i) \cdot (h(\vec{v}_j) - h^M(\vec{v}_j)) + \sum_{i,j} A_{ij} (h(\vec{u}_i) - h^M(\vec{u}_i)) h^M(\vec{v}_j)\\
&\le \sum_{i,j} A_{ij} h^M(\vec{u}_i) \cdot h^M(\vec{v}_j) + \underbrace{\Big| \sum_{i,j} A_{ij} h^M(\vec{u}_i) \cdot (h(\vec{v}_j) - h^M(\vec{v}_j)) \Big|}_{\overset{(4)}{\le} \frac{\sqrt{3}}{M} \cdot \mathrm{SDP}(A)} + \underbrace{\Big| \sum_{i,j} A_{ij} (h(\vec{u}_i) - h^M(\vec{u}_i)) h^M(\vec{v}_j) \Big|}_{\overset{(4)}{\le} \frac{\sqrt{3}}{M} \cdot \mathrm{SDP}(A)}\\
&\le \frac{M^2}{t} \sum_{k=1}^{t} \sum_{i,j} A_{ij} \Big((h^M(\vec{u}_i))_k \cdot \frac{\sqrt{t}}{M} \Big) \Big((h^M(\vec{v}_j))_k \cdot \frac{\sqrt{t}}{M} \Big) + \frac{2\sqrt{3}}{M} \cdot \mathrm{SDP}(A).
\end{aligned}
$$

Thus we can find a k such that

$$
\sum_{i,j} A_{ij} \Big((h^M(\vec{u}_i))_k \cdot \frac{\sqrt{t}}{M} \Big) \Big((h^M(\vec{v}_j))_k \cdot \frac{\sqrt{t}}{M} \Big) \ge \frac{1}{M^2} \Big(\mathrm{SDP}(A) - \frac{2\sqrt{3}}{M} \mathrm{SDP}(A) \Big).
$$

To conclude, we set $M = 4$, $x_i = (h^M(\vec{u}_i))_k \cdot \frac{\sqrt{t}}{M}$, $y_j = (h^M(\vec{v}_j))_k \cdot \frac{\sqrt{t}}{M}$ and get

$$
\sum_{i,j} A_{ij} x_{ik} y_{jk} \ge \mathrm{SDP}(A) \cdot \Big(\frac{M - 2\sqrt{3}}{M^3} \Big) \approx 0.01 \cdot \mathrm{SDP}(A).
$$

Here $x_i, y_j \in [-1, 1]$, and we can always find values in $\{\pm 1\}$ that are at least as good (to see how, fix all the x_i and observe that there is always an optimal setting of each individual y_j that is either $+1$ or -1).

Thus the theorem is proved, conditioned on the lemma, whose proof we now turn to.

Proof of Lemma 8.1. The proof of the lemma relies on the 4-wise independence condition on the vectors $\vec{g}_i$.

(1) We check that

$$
\begin{aligned}
h(\vec{u}) \cdot h(\vec{v}) &= \sum_{k=1}^{t} \frac{\vec{g}_k \cdot \vec{u}}{\sqrt{t}} \frac{\vec{g}_k \cdot \vec{v}}{\sqrt{t}} \\
&= \sum_{k=1}^{t} \sum_{i,j=1}^{d} \frac{(\vec{g}_k)_i \cdot \vec{u}_i}{\sqrt{t}} \frac{(\vec{g}_k)_j \cdot \vec{v}_j}{\sqrt{t}} \\
&= \sum_{i,j=1}^{d} \underbrace{\left(\frac{1}{t} \sum_{k=1}^{t} (\vec{g}_k)_i (\vec{g}_k)_j \right)}_{=\delta_{ij} \text{ by 2-independence}} \vec{u}_i \vec{v}_j \\
&= \sum_{i=1}^{d} \vec{u}_i \vec{v}_i \\
&= \vec{u} \cdot \vec{v}.
\end{aligned}
$$

(2) Using (1),

$$
\|h(\vec{u})\|^2 = h(\vec{u}) \cdot h(\vec{u}) = \vec{u} \cdot \vec{u} = \|\vec{u}\|^2 = 1.
$$

(3) Using (2),

$$
\|h^M(\vec{u})\|^2 = \sum_k (h^M(\vec{u}))_k^2 \leq \sum_k (h(\vec{u}))_k^2 = 1.
$$

(4) From the definition of $h^M(\cdot)$, we have

$$
\begin{aligned}
\|h(\vec{u}) - h^M(\vec{u})\|^2 &= \sum_{k : |h(\vec{u})_k| > \frac{M}{\sqrt{t}}} |h(\vec{u})_k|^2 \\
&= \frac{1}{t} \sum_{k : |\vec{g}_k \cdot \vec{u}| > M} (\vec{g}_k \cdot \vec{u})^2 \\
&\leq \sqrt{\frac{1}{t} \sum_{k : |\vec{g}_k \cdot \vec{u}| > M} (\vec{g}_k \cdot \vec{u})^4} \sqrt{\frac{1}{t} \sum_{k : |\vec{g}_k \cdot \vec{u}| > M} 1} \\
&\leq \sqrt{3} \frac{\sqrt{3}}{M^2} \\
&= \frac{3}{M^2},
\end{aligned}
$$

where the second line is by the Cauchy-Schwarz inequality and the third uses

$$\begin{aligned}\frac{1}{t}\sum_{k:|\vec{g}_k\cdot\vec{u}|>M}(\vec{g}_k\cdot\vec{u})^4 &= \frac{1}{t}\sum_k\sum_{i,j,l,m}(\vec{g}_k)_i(\vec{g}_k)_j(\vec{g}_k)_l(\vec{g}_k)_m\vec{u}_i\vec{u}_j\vec{u}_l\vec{u}_m\\ &= \frac{1}{t}\left(\sum_i(\vec{u}_i)^4+3\sum_{i\neq l}(\vec{u}_i)^2(\vec{u}_l)^2\right)\\ &\leq \frac{3}{t}\left(\|\vec{u}_i\|\right)^2\\ &= 3,\end{aligned}$$

where the second line follows by observing that $\sum_k(\vec{g}_k)_i(\vec{g}_k)_j(\vec{g}_k)_l(\vec{g}_k)_m = 0$ unless $i = j$ and $l = m$, or $i = l$ and $j = m$, or $i = m$ and $j = l$. From this bound we also gets $\#\{k : |\vec{g}_k \cdot \vec{u}| > M\} \leq \frac{3}{M^4}t$, as required above.

□

Remark 8.2. The same guarantees for the rounding procedure can be obtained by taking random projections on Gaussian vectors. We will see this later in class.

8.2 Dimension Reduction and the Johnson-Lindenstrauss Lemma

In this lecture we introduce a very powerful technique for solving high-dimensional problems: the Johnson-Lindenstrauss (JL) dimensionality reduction lemma. This technique is naturally useful for geometric problems, but not only; other applications include:

- Proximity problems: nearest neighbor, closest/furthest pair, Euclidean minimum spanning tree;
- Clustering, information retrieval;
- Learning an unknown mixture of Gaussians;
- Dimensionality reduction for online settings, such as sketching for streaming with limited storage;

and many more; in fact there is even a whole book devoted to the topic: "The Random Projection Method", by Santosh Vempala. Before stating and proving the JL lemma we introduce a little background on Gaussian random variables.

8.2.1 Gaussian random variables

A real continuous random variable X is defined by a nonnegative, integrable density function $\gamma : \mathbb{R} \to \mathbb{R}_+$ such that

$$\mathbf{Pr}\left(X \in [a, b]\right) = \int_a^b \gamma(x)\, dx.$$

Definition 8.3. We say X is *Gaussian*, or normally distributed, with mean μ and variance σ^2 when

$$\gamma(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2}, \tag{8.1}$$

in which case we write $X \sim \mathcal{N}(\mu, \sigma^2)$. We further say X is *standard Gaussian* if $\mu = 0$ and $\sigma^2 = 1$.

Here are some standard but useful properties of the Gaussian distribution.

Lemma 8.4.

(1) The density γ defined in (8.1) is normalized, i.e.

$$\int_{-\infty}^{\infty} \gamma(x)\, dx = 1.$$

(2) Let $Z = c_1 X_1 + c_2 X_2$ where $X_1, X_2 \sim \mathcal{N}(0, 1)$ are independent random variables. Then $Z \sim \mathcal{N}(0, c_1^2 + c_2^2)$.

(3) Let $X \sim \mathcal{N}(0, 1)$ and $t < 1/2$. Then

$$\mathbf{E}[e^{tX^2}] = \frac{1}{\sqrt{1 - 2t}}.$$

Exercise 1. Prove the lemma. For 1., show that it is enough to check normalization for $\mu = 0, \sigma^2 = 1$. If $I = \int_{-\infty}^{\infty} e^{-x^2/2}\, dx$, first verify that I is well-defined (it should be bounded) and then show that $I^2 = 2\pi$ by performing a change of variables. For 2., write the integral for the cumulative distribution function $\mathbf{Pr}(Z \leq t)$ and again use rotation symmetry.

8.2.2 The Johnson-Lindenstrauss lemma

The JL lemma addresses the question of low-dimensional embeddings of points in Euclidean space that approximately preserve distances. The construction we give not only shows existence, but also yields a randomized algorithm for finding a linear embedding satisfying the desired properties with high probability. The lemma can be stated as follows.

Lemma 8.5 (Johnson-Lindenstrauss)**.** *For any integer $d > 0$ and $0 < \varepsilon, \delta < 1/2$ there exists a distribution on $k \times d$ real matrices G for $k > 4\ln(2/\delta)/(\varepsilon^2 - \varepsilon^3)$ such that for any $x \in \mathbb{R}^d$,*

$$\mathbf{Pr}\left((1 - \varepsilon)\|x\|^2 \leq \|Gx\|^2 \leq (1 + \varepsilon)\|x\|^2\right) > 1 - \delta.$$

Proof. The distribution we use is simple: $G \in \mathbb{R}^{k \times d}$ is formed by choosing coefficients $G_{ij} \sim \mathcal{N}(0, 1/k)$ i.i.d. Let

$$Z = \frac{\|Gx\|^2}{\|x\|^2} = \sum_{i=1}^{k} \frac{(Gx)_i^2}{\|x\|^2},$$

so that the desired bounds are $(1-\varepsilon) \leq Z \leq (1+\varepsilon)$ w.h.p. Observe that, for all i, $(Gx)_i = \sum_{j=1}^{d} G_{ij} x_j \sim \mathcal{N}(0, \|x\|^2/k)$, by the second item from Lemma 8.4 and our choice of G_{ij}, so

$$\mathbf{E}[(Gx)_i^2] = \mathbf{Var}[(Gx)_i] = \frac{\|x\|^2}{k},$$

and $\mathbf{E}[Z] = \frac{1}{\|x\|^2} \sum_{i=1}^{k} \frac{\|x\|^2}{k} = 1$.

Now we want a high probability concentration bound on Z. Let $X_i = \frac{(Gx)_i}{\|x\|}$, so that $Z = \sum_{i=1}^{k} X_i^2$. Since the rows of G are independent it follows that the X_i are independent, so Z is a sum of i.i.d. random variables. Unfortunately these random variables are not bounded, so we can't directly apply the Chernoff bound, or generalizations such as Bernstein's inequality. Instead we go back to the basics and apply the Laplace transform method from scracth. This will also give us a sharper bound.

We bound $\mathbf{Pr}(Z \geq (1+\varepsilon))$, the other tail being analogous.

$$\begin{aligned} \mathbf{Pr}(Z \geq (1+\varepsilon)) &= \mathbf{Pr}\left(e^{tkZ} \geq e^{tk(1+\varepsilon)}\right) \\ &\leq \frac{\mathbf{E}[e^{tkZ}]}{e^{tk(1+\varepsilon)}} \\ &= \frac{\prod_{i=1}^{k} \mathbf{E}[e^{tkX_i^2}]}{e^{tk(1+\varepsilon)}} \\ &= \frac{1}{(1-2t)^{k/2} e^{tk(1+\varepsilon)}}, \end{aligned}$$

where the second equality uses independence of the X_i and the third equality follows from item 3. in Lemma 8.4 (using $\sqrt{k} X_i \sim \mathcal{N}(0,1)$ and assuming $t < 1/2$). Now we pick $t = \frac{\varepsilon}{2(1+\varepsilon)} < 1/2$, so that this upper bound simplifies to

$$\begin{aligned} \left((1+\varepsilon)e^{-\varepsilon}\right)^{k/2} &\leq ((1+\varepsilon)(1-\varepsilon+\varepsilon^2/2))^{k/2} \\ &= (1-\varepsilon^2/2+\varepsilon^3/2)^{k/2} \\ &\leq e^{-(\varepsilon^2-\varepsilon^3)k/4}, \end{aligned}$$

where the inequalities follow by Taylor expansion. Choosing $k > 4\ln(2/\delta)/(\varepsilon^2 - \varepsilon^3)$ yields a tail bound of at most $\delta/2$. Adding the bound for the lower tail yields an overall probability of failure of at most δ. □

The following is a simple but important consequence:

Theorem 8.6. *Let P be a set of n points in $\mathbb{R}^d$ and $0 < \varepsilon < 1$. For dimension $k > \frac{8 \ln n}{\varepsilon^2 - \varepsilon^3}$, there exists a linear map $\varphi : \mathbb{R}^d \to \mathbb{R}^k$, such that, for all $u, v \in P$,*

$$(1-\varepsilon)\|u-v\|^2 \leq \|\varphi(u) - \varphi(v)\|^2 \leq (1+\varepsilon)\|u-v\|^2,$$

that is, the mapping φ does not distort distances too much.

Proof. The construction for φ is simply to choose G randomly as in Lemma 8.5, and set $\varphi(u) = Gu$. Then for every pair (u, v), $\|\varphi(u) - \varphi(v)\|^2 = \|G(u-v)\|^2$. If we choose $\delta = O(1/n^2)$ in Lemma 8.5 we get that for any pair (u, v) the probability that $(1-\varepsilon)\|u-v\|^2 < \|G(u-v)\|^2 < (1+\varepsilon)\|G(u+v)\|^2$ is at least $1 - 1/(2n^2)$. Taking a union bound over all possible pairs, a random G will work with probability at least $1/2$, and in particular there exists a G, hence a φ, that works. □

Remark 8.7. We can tweak the constant factor in the threshold for k to get a high probability bound that φ works, e.g., choosing $k > \frac{12 \ln n}{\varepsilon^2 - \varepsilon^3}$ yields failure probability at most $\binom{n}{2}\frac{2}{n^3} < \frac{1}{n}$. Such a bound yields a Las Vegas algorithm for constructing a working φ (repeatedly sample a φ and check the distortion of φ until success).

Remark 8.8. A somewhat cleaner threshold for k of the form $k = \Omega(\varepsilon^{-2} \ln n)$ suffices if we have an upper bound on ε, e.g., if $\varepsilon < 1/2$, then $\varepsilon^2 - \varepsilon^3 > \varepsilon^2/2$ so

$$k > \frac{16 \ln n}{\varepsilon^2} \implies k > \frac{8 \ln n}{\varepsilon^2 - \varepsilon^3}.$$