For use later on, we record another version of the maximum principle that does not require connectivity of the graph.

Lemma 5.4 (Maximum principle)**.** *Let* $u \in \ell^2(\mathcal{X})$ *such that* $\mathcal{L}u(x) > 0$ *for all* $x \in \mathcal{X} \setminus \Gamma$*. Then*

$$\max_{x\in\mathcal{X}} u(x) = \max_{x\in\Gamma} u(x). \tag{5.25}$$

Proof. Let $x_0 \in \mathcal{X}$ such that $u(x_0) = \max_{x\in\mathcal{X}} u(x)$. Since $u(x_0) \geq u(y)$ for all $y \in \mathcal{X}$, we have

$$\mathcal{L}u(x_0) = \sum_{y\in\mathcal{X}} w_{xy}(u(y) - u(x_0)) \leq 0.$$

Since $\mathcal{L}u(x) > 0$ for all $x \in \mathcal{X} \setminus \Gamma$, we must have $x_0 \in \Gamma$, which completes the proof. □

5.2 Concentration of measure

As we will be working with random geometric graphs, we will require some basic probabilistic estimates, referred to as concentration of measure, to control the random behavior of the graph. In this section, we review some basic, and very useful, concentration of measure results. It is a good idea to review the Section A.9 for a review of basic probability before reading this section.

Let $X_1, X_2, \ldots, X_n$ be a sequence of n independent and identically distributed real-valued random variables and let $S_n = \frac{1}{n}\sum_{i=1}^n X_i$. In Section A.9.4 we saw how to use Chebyshev's inequality to obtain bounds of the form

$$\mathbb{P}(|S_n - \mu| \geq t) \leq \frac{\sigma^2}{nt^2} \tag{5.26}$$

for any $t > 0$, where $\mu = \mathbb{E}[X_i]$ and $\sigma^2 = \mathrm{Var}\,(X_i)$. Without further assumptions on the random variables X_i, these estimates are essentially tight. However, if the random variables X_i are almost surely bounded (i.e., $\mathbb{P}(|X_i| \leq b) = 1$ for some $b > 0$), which is often the case in practical applications, then we can obtain far sharper exponential bounds.

To see what to expect, we note that the Central Limit Theorem says (roughly) that

$$S_n = \mu + \frac{1}{\sqrt{n}}N(0,\sigma^2) + o\left(\frac{1}{\sqrt{n}}\right) \quad \text{as } n \to \infty$$

where $N(0,\sigma^2)$ represents a normally distributed random variable with mean zero and variance σ^2. Ignoring error terms, this says that $Y_n := \sqrt{n}(S_n - \mu)$ is approximately $N(0,\sigma^2)$, and so we may expect Gaussian-like estimates of the form

$$\mathbb{P}(|Y_n| \geq x) \leq C \exp\left(-\frac{x^2}{2\sigma^2}\right)$$

for $x > 0$. Setting $x = \sqrt{n}t$ we can rewrite this as

$$\mathbb{P}(|S_n - \mu| \geq t) \leq C \exp\left(-\frac{nt^2}{2\sigma^2}\right) \tag{5.27}$$

for any $t > 0$. Bounds of the form (5.26) and (5.27) are called *concentration inequalities*, or *concentration of measure*. In this section we describe the main ideas for proving exponential bounds of the form (5.27), and prove the Hoeffding and Bernstein inequalities, which are some of the most useful concentration inequalities. For more details we refer the reader to [8].

One normally proves exponential bounds like (5.27) with the Chernoff bounding trick. Let $s > 0$ and note that

$$\mathbb{P}(S_n - \mu \geq t) = \mathbb{P}(s(S_n - \mu) \geq st) = \mathbb{P}\left(e^{s(S_n-\mu)} \geq e^{st}\right).$$

The random variable $Y = e^{s(S_n-\mu)}$ is nonnegative, so we can apply Markov's inequality (see Proposition A.39) to obtain

proof 1

$$\begin{aligned}\mathbb{P}(S_n - \mu \geq t) &\leq e^{-st}\mathbb{E}\left[e^{s(S_n-\mu)}\right] \\ &= e^{-st}\mathbb{E}\left[e^{\frac{s}{n}\sum_{i=1}^n (X_i-\mu)}\right] \\ &= e^{-st}\mathbb{E}\left[\prod_{i=1}^n e^{\frac{s}{n}(X_i-\mu)}\right].\end{aligned}$$

Applying (A.30) yields

$$\mathbb{P}(S_n - \mu \geq t) \leq e^{-st}\prod_{i=1}^n \mathbb{E}\left[e^{\frac{s}{n}(X_i-\mu)}\right] = e^{-st}\mathbb{E}\left[e^{\frac{s}{n}(X_1-\mu)}\right]^n. \tag{5.28}$$

This bound is the main result of the Chernoff bounding trick. The key now is to obtain bounds on the *moment generating function*

$$M_X(\lambda) := \mathbb{E}[e^{\lambda(X-\mu)}],$$

where $X = X_1$.

In the case where the X_i are Bernoulli random variables, we can compute the moment generating function explicitly, and this leads to the Chernoff bounds. Before giving them, we present some preliminary technical propositions regarding the function

$$h(\delta) = (1+\delta)\log(1+\delta) - \delta, \tag{5.29}$$

which appears in many concentration inequalities.

We now choose $C + 2 = 60$, or $C = 58$. Then we check that $\frac{8}{9}(C+3) \leq C$ and so

$$\mathbb{P}(|P| \geq 61 \log n) \leq \mathbb{P}(Y \geq 58 \log n) \leq \frac{1}{n^2}.$$

Since here are n paths from leaves to the root, we union bound over all paths to find that

$$\mathbb{P}(Z \geq 61 \log n) \leq \frac{1}{n}.$$

Therefore, with probability at least $1 - \frac{1}{n}$, quicksort takes at most $O(n \log n)$ operations to complete. △

In general, we cannot compute the moment generating function explicitly, and are left to derive upper bounds. The first bound is due to Hoeffding.

Lemma 5.8 (Hoeffding Lemma)**.** *Let X be a real-valued random variable for which $|X - \mu| \leq b$ almost surely for some $b > 0$, where $\mu = \mathbb{E}[X]$. Then we have*

$$M_X(\lambda) \leq e^{\frac{\lambda^2 b^2}{2}}. \tag{5.35}$$

Proof. Since $x \mapsto e^{sx}$ is a convex function, we have

$$e^{\lambda x} \leq e^{-\lambda b} + \frac{x+b}{b}\sinh(\lambda b)$$

provided $|x| \leq b$ (the right hand side is the secant line from $(-b, e^{-\lambda b})$ to $(b, e^{\lambda b})$; recall $\sinh(t) = (e^t - e^{-t})/2$ and $\cosh(t) = (e^t + e^{-t})/2$). Therefore we have

$$\begin{aligned} M_X(\lambda) = \mathbb{E}\left[e^{\lambda(X-\mu)}\right] &\leq \mathbb{E}\left[e^{-\lambda b} + \frac{X - \mu + b}{b}\sinh(\lambda b)\right] \\ &= e^{-\lambda b} + \frac{\mathbb{E}[X] - \mu + b}{b}\sinh(\lambda b) \\ &= e^{-\lambda b} + \sinh(\lambda b) = \cosh(\lambda b). \end{aligned}$$

The proof is completed by the elementary inequality $\cosh(x) \leq e^{\frac{x^2}{2}}$ (compare Taylor series). □

Combining the Hoeffding Lemma with (5.28) yields

$$\mathbb{P}(S_n - \mu \geq t) \leq e^{-st}\mathbb{E}\left[e^{\frac{s}{n}(X_1 - \mu)}\right]^n = \exp\left(-st + \frac{s^2 b^2}{2n}\right),$$

provided $|X_i - \mu| \leq b$ almost surely. Optimizing over $s > 0$ we find that $s = nt/b^2$, which yields the following result.

Theorem 5.9 (Hoeffding inequality)**.** *Let $X_1, X_2 \ldots, X_n$ be a sequence of* i.i.d. *real-valued random variables with finite expectation $\mu = \mathbb{E}[X_i]$, and write $S_n = \frac{1}{n}\sum_{i=1}^n X_i$. Assume there exists $b > 0$ such that $|X - \mu| \leq b$ almost surely. Then for any $t > 0$ we have*

$$\mathbb{P}(S_n - \mu \geq t) \leq \exp\left(-\frac{nt^2}{2b^2}\right). \tag{5.36}$$

Remark 5.10. Of course, the opposite inequality

$$\mathbb{P}(S_n - \mu \leq -t) \leq \exp\left(-\frac{nt^2}{2b^2}\right)$$

holds by a similar argument. Thus, by the union bound we have

$$\mathbb{P}(|S_n - \mu| \geq t) \leq 2\exp\left(-\frac{nt^2}{2b^2}\right).$$

The Hoeffding inequality is tight if $\sigma^2 \approx b^2$, so that the right hand side looks like the Gaussian distribution in (5.27), up to constants. For example, if X_i are uniformly distributed on $[-b, b]$ then

$$\sigma^2 = \frac{1}{2b}\int_{-b}^{b} x^2\, dx = \frac{b^2}{3}.$$

However, if $\sigma^2 \ll b^2$, then one would expect to see σ^2 in place of b^2 as in (5.27), and the presence of b^2 leads to a suboptimal bound.

Example 5.2. Let

$$Y = \max\left\{1 - \frac{|X|}{\varepsilon}, 0\right\}.$$

where X is uniformly distributed on $[-1, 1]$, as above, and $\varepsilon \ll b$. Then $|Y| \leq 1$, so $b = 1$, but we compute

$$\sigma^2 \leq \frac{1}{2}\int_{-\varepsilon}^{\varepsilon} dx = \varepsilon.$$

Hence, $\sigma^2 \ll 1$ when ε is small, and we expect to get sharper concentration bounds than are provided by the Hoeffding inequality. This example is similar to what we will see later in consistency of graph Laplacians. △

The Bernstein inequality gives the sharper bounds that we desire, and follows from Bernstein's Lemma.

Lemma 5.11 (Bernstein Lemma)**.** *Let X be a real-valued random variable with finite mean $\mu = \mathbb{E}[X]$ and variance $\sigma^2 = Var(X)$, and assume that $|X - \mu| \leq b$ almost surely for some $b > 0$. Then we have*

$$M_X(\lambda) \leq \exp\left(\frac{\sigma^2}{b^2}(e^{\lambda b} - 1 - \lambda b)\right). \tag{5.37}$$

The reader should contrast this with the case where X_i form a uniform grid over $[0,1]^d$. In this case, for Lipschitz functions the numerical integration error is $O(\Delta x)$, where Δx is the grid spacing. For n points on a uniform grid in $[0,1]^d$ the grid spacing is $\Delta x \sim n^{-1/d}$, which suffers from the *curse of dimensionality* when d is large. The Monte Carlo error estimate (5.46), on the other hand, is remarkable in that it is independent of dimension d! Thus, Monte Carlo integration overcomes the *curse of dimensionality* by simply replacing a uniform discretization grid by random variables. Monte Carlo based techniques have been used to solve PDEs in high dimensions via sampling random walks or Brownian motions.

Proof of Theorem 5.16. Let $Y_i = u(X_i)$. We apply Bernstein's inequality with $S_n = \frac{1}{n}\sum_{i=1}^n Y_i = I_n(u)$, $\sigma^2 = \mathrm{Var}(Y_i)$ and $b = 2\|u\|_{L^\infty([0,1]^d)}$ to find that

$$|I(u) - I_n(u)| \leq t$$

with probability at least $1 - 2\exp\left(-\frac{nt^2}{2\left(\sigma^2 + \frac{1}{3}bt\right)}\right)$ for any $t > 0$. Set $t = \lambda\sigma/\sqrt{n}$ for $\lambda > 0$ to find that

$$|I(u) - I_n(u)| \leq \frac{\lambda\sigma}{\sqrt{n}}$$

with probability at least $1 - 2\exp\left(-\frac{\sigma^2\lambda^2}{2\left(\sigma^2 + \frac{b\lambda\sigma}{3\sqrt{n}}\right)}\right)$. Restricting $\lambda \leq 3\sigma\sqrt{n}/b$ completes the proof. $\square$

We conclude this section with the Azuma/McDiarmid inequality. This is slightly more advanced and is not used in the rest of these notes, so the reader may skip ahead without any loss. It turns out that the Chernoff bounding method used to prove the Chernoff, Hoeffding, and Bernstein inequalities does not use in any essential way the linearity of the sum defining S_n. Indeed, what matters is that S_n does not depend too much on any particular random variable X_i. Using these ideas leads to far more general (and more useful) concentration inequalities for functions of the form

$$Y_n = f(X_1, X_2, \ldots, X_n) \tag{5.47}$$

that may depend *nonlinearly* on the X_i. To express that Y_n does not depend too much on any of the X_i, we assume that f satisfies the following bounded differences condition: There exists $b > 0$ such that

$$|f(x_1, \ldots, x_i, \ldots, x_n) - f(x_1, \ldots, \widetilde{x}_i, \ldots, x_n)| \leq b \tag{5.48}$$

for all $x_1, \ldots, x_n$ and $\widetilde{x}_i$. In this case we have the following concentration inequality.

Theorem 5.18 (Azuma/McDiarmid inequality). *Define Y_n by (5.47), where $X_1, \ldots, X_n$ are* i.i.d. *random variables satisfying $|X_i| \leq M$ almost surely, and assume f satisfies (5.48). Then for any $t > 0$*

$$\mathbb{P}(Y_n - \mathbb{E}[Y_n] \geq t) \leq \exp\left(-\frac{t^2}{2nb^2}\right). \tag{5.49}$$

Skip proof first time

Proof. The proof uses conditional probability, which we have not developed in these notes, so we give a sketch of the proof. For $2 \le k \le n$ we define

$$Z_k = \mathbb{E}[Y_n \,|\, X_1, \dots, X_k] - \mathbb{E}[Y_n \,|\, X_1, \dots, X_{k-1}].$$

and set $Z_1 = \mathbb{E}[Y_n \,|\, X_1] - \mathbb{E}[Y_n]$. Since

$$Y_n = \mathbb{E}[Y_n \,|\, X_1, \dots, X_n],$$

we have the telescoping sum

$$Y_n - \mathbb{E}[Y_n] = \sum_{k=1}^{n} Z_k.$$

The random variables Z_k record how much the conditional expectation changes when we add information about X_k. While the Z_k are not independent, they form a *martingale difference sequence*, which allows us to essentially treat them as independent and use a similar proof to Hoeffding's inequality. The useful martingale difference property is the identity

$$\mathbb{E}[Z_k \,|\, X_1, \dots, X_{k-1}] = 0$$

for $k \ge 2$, and $\mathbb{E}[Z_1] = 0$, which follow from the law of iterated expectations.

We now follow the Chernoff bounding method and law of iterated expectations to obtain

$$\begin{aligned}
\mathbb{P}(Y_n - \mathbb{E}[Y_n] \ge t) &= \mathbb{P}(e^{s\sum_{k=1}^n Z_k} \ge e^{st}) \\
&\le e^{-st}\mathbb{E}[e^{s\sum_{k=1}^n Z_k}] \\
&= e^{-st}\mathbb{E}[\mathbb{E}[e^{s\sum_{k=1}^n Z_k} \,|\, X_1, \dots, X_{n-1}]] \\
&= e^{-st}\mathbb{E}[e^{s\sum_{k=1}^{n-1} Z_k}\mathbb{E}[e^{sZ_n} \,|\, X_1, \dots, X_{n-1}]].
\end{aligned}$$

Define

$$U_k = \sup_{|x| \le M} \mathbb{E}[f(X_1, \dots, X_{k-1}, x, X_{k+1}, \dots X_n) - Y_n \,|\, X_1, \dots, X_{k-1}],$$

and

$$L_k = \inf_{|x| \le M} \mathbb{E}[f(X_1, \dots, X_{k-1}, x, X_{k+1}, \dots X_n) - Y_n \,|\, X_1, \dots, X_{k-1}].$$

Then $L_k \le Z_k \le U_k$. By (5.48) we have $U_k \le b$ and $L_k \ge -b$, and so $|Z_k| \le b$. Following a very similar argument as in the proof of Lemma 5.8 we have

$$\begin{aligned}
\mathbb{E}[e^{sZ_k} \,|\, X_1, \dots, X_{k-1}] &\le \mathbb{E}\left[e^{-sb} + \frac{Z_k + b}{b}\sinh(sb) \,|\, X_1, \dots, X_{k-1}\right] \\
&= e^{-st} + \sinh(sb) = \cosh(sb) \le e^{\frac{s^2b^2}{2}},
\end{aligned}$$