Rounding Techniques in Approximation Algorithms

Lecture 18: Isotropic Updates Imply Concentration

Lecturer: Nathan Klein Scribe: Taishan Chen

1 Obtaining Strong Concentration

We pick up where we left off last time and prove that isotropic updates imply concentration. Remember we proved the following lemma:

Lemma 1.1. *Let $0 < \alpha < 1$, $t \geq 0$. Let $Z_0, Z_1, ...$ be random variables with Z_0 deterministic. Let $Y_k = Z_k - Z_{k-1} \leq 1$. Finally, assume*

$$\mathbb{E}\left[Y_k \mid Z_1, ..., Z_{k-1}\right] \leq -\alpha\mathbb{E}\left[Y_k^2 \mid Z_1, ..., Z_{k-1}\right].$$

Then

$$\mathbb{P}\left[Z_k - Z_0 > t\right] \leq e^{-\alpha t}$$

We will now use this to show concentration.

1.1 Concentration in a Direction

We say a distribution μ over binary random variables $X_1, \ldots, X_n$ has strong concentration in direction $c \in \mathbb{R}$ if the random variable $Y = \sum_{i=1}^n c_i X_i$ obeys an exponentially decaying tail bound. Sub-isotropic updates give strong concentration in every direction c, and Bansal proves a slightly weakened form of Bernstein's inequality. However, for simplicity, for this lecture we will focus on the case when $c \in \{0,1\}^n$, which is the Chernoff bound setting.

Since our updates will be sub-isotropic in whichever variables we have $c_i = 1$ in, we will simply do a Chernoff bound on the full set of variables, again for notational ease. In particular, we will show that (where $\mu = \mathbb{E}\left[X\right] = \sum x_i^{(0)}$):

Theorem 1.2. *Let $X = \sum \tilde{x}_i$ where $\tilde{x}$ is the final integral solution sampled from μ, the distribution induced by sub-isotropic rounding. Then,*

$$\mathbb{P}\left[X \geq (1+\delta)\mu\right] \leq \exp\left(-\frac{\delta^2\mu}{2\beta(1+\delta/3)}\right)$$

Note if $\beta = 1$ this recovers the standard upper tail Chernoff bound with a $\frac{2}{3}\delta$ in the denominator instead of a δ. This is equivalent to $\mathbb{P}\left[X - \mu \geq \epsilon\mu\right]$.

1.2 Proof Overview

At first, it's not clear how we should apply Lemma 1.1 as the natural martingale does not have negative expected movement as is required by the first criteria. To address this, we will add the *variance* of our current point to the natural martingale, which will decrease over time.

Formally, we define a function $\mathrm{Var}(x) : \mathbb{R}^n \to \mathbb{R}$ to be $\mathrm{Var}(x) = \sum_{i=1}^n x_i(1-x_i)$. This measure the variance of a distribution of n independent Bernoullis with marginals x. We define

$$Z_k = \sum x_i^{(k)} + \lambda \mathrm{Var}(x^{(k)}).$$

Now we will show the two criteria of Lemma 1.1 are met with $\alpha = \frac{\lambda}{8\eta}$ where $\lambda = \frac{t}{t+2\mathrm{Var}(x^{(0)})}$.

Intuitively, we get these criteria because the variance is decreasing in expectation at every step. For a simple example, imagine we are in one dimension and we move to $x_i + \epsilon$ with probability $1/2$ and $x_i - \epsilon$ otherwise. Then, the expected variance of our new point is

$$\frac{1}{2}((x_i+\epsilon)(1-x_i-\epsilon) + (x_i-\epsilon)(1-x_i+\epsilon)) = x_i(1-x_i) - \epsilon^2$$

For more intuition, notice that after enough steps our solution will reach an integer point at which point variance is 0. So, it makes sense that this quantity decreases in expectation as we move, i.e. that the expected value of Y_k is negative. However, we need to show that it is negative proportional to the expected value of Y_k^2.

1.3 Demonstrating the Two Criteria

Let's show the two criteria.

Lemma 1.3. $Y_k \leq 1 \quad \forall k, 1 \leq k \leq t$

Lemma 1.4. $E[Y_k|Z_1,\ldots,Z_{k-1}] \leq -\alpha \cdot E[Y_k^2|Z_1,\ldots,Z_{k-1}]$

To prove the first lemma, first we define $y^{(k)} = x^{(k)} - x^{(k-1)}$ to be the difference of x between consecutive two steps. In other words, $y^{(k)}$ is the update we make in the step k. By calculation,

$$\begin{aligned}
Y_k = Z_k - Z_{k-1} &= \sum_i y_i^{(k)} + \lambda \sum_i x_i^{(k)}(1-x_i^{(k)}) - \lambda \sum_i x_i^{(k-1)}(1-x_i^{(k-1)}) \\
&= \sum_i y_i^{(k)} + \lambda(\sum_i (x_i^{(k)} - x_i^{(k-1)})(1-(x_i^{(k)} - x_i^{(k-1)}) - 2x_i^{(k-1)}(x_i^{(k)} - x_i^{(k-1)}))) \\
&= \sum_i y_i^{(k)} + \lambda(\sum_i y_i^{(k)}(1-y_i^{(k)}) - 2x_i^{(k-1)}y_i^{(k)}) \\
&= \sum_i y_i^{(k)} + \lambda(\sum_i y_i^{(k)}(1-y_i^{(k)} - 2x_i^{(k-1)}))
\end{aligned}$$

Since $\lambda = \frac{t}{t+2\mathrm{Var}(x^{(0)})} < 1$, $(1-y_i^{(k)} - 2x_i^{(k-1)}) < 1$, we have $Y_k \leq 2\sum_i y_i^{(k)}$. Remind that the update $y^{(k)}$ is chosen from a sub-isotrphic distribution, where $y^{(k)} = \epsilon u^{1/2} r$. Now we prove a helper claim:

Claim 1.5. $\|y^{(k)}\|_2 \leq \epsilon n \leq \frac{1}{2\sqrt{n}}$.

Proof. First ϵ is a scalar, so $\|y^{(k)}\|_2 = \epsilon\|u^{1/2}r\|_2$. We know that u is a positive semi-definitive (PSD) matrix which suffices that $\forall i \in [1,n], u_{ii} \leq 1$. Let c_i be the i-th column of $u^{1/2}$, we have $u_{ii} = \|c_i\|_2 \leq 1$. Use triangle inequality, and because r is a vector with only $+1, -1$ entries,

$$\|u^{1/2}r\|_2 \le \|r\|_1 \le n$$

That immediately gives $\|y^{(k)}\|_2 \le \epsilon n$. ϵ is defined as $\frac{1}{2n^{3/2}}$, so $\|y^{(k)}\|_2 \le \frac{1}{2\sqrt{n}}$. □

We can now finish the first criteria easily:

Claim 1.6. $Y_k \le 1$.

Proof. By Cauchy–Schwarz,

$$Y_k \le 2\sum_i y_i^{(k)} = 2\|y^{(k)}\|_1 \le 2\sqrt{n}\|y^{(k)}\|_2 \le 1$$

as desired. □

We then look at the second criteria. To prove it, we consider two expected values $\mathbb{E}_{k-1}[Y_k]$ and $\mathbb{E}_{k-1}[Y_k^2]$ given previous $k-1$ steps fixed. For the first one,

$$\mathbb{E}_{k-1}[Y_k] = \mathbb{E}[\sum_i y_i^{(k)} + \lambda(\sum_i y_i^{(k)}(1 - y_i^{(k)} - 2x_i^{(k-1)}))]$$

Since for all i, k, $\mathbb{E}[y_i^{(k)}] = 0$ (remember $y^{(k)}$ is from the sub-isotrophic distribution),

$$E_{k-1}[Y_k] = -\lambda\mathbb{E}[\sum_i (y_i^{(k)})^2]$$

Then for $\mathbb{E}_{k-1}[Y_k^2]$, by applying $(a+b)^2 \le 2a^2 + 2b^2$ twice,

$$\begin{aligned}
Y_k^2 &= (\sum_i y_i^{(k)} + \lambda(\sum_i y_i^{(k)}(1 - y_i^{(k)} - 2x_i^{(k-1)})))^2 \\
&\le 2(\sum_i y_i^{(k)})^2 + 2\lambda^2(\sum_i y_i^{(k)}(1 - y_i^{(k)} - 2x_i^{(k-1)}))^2 \\
&\le 2(\sum_i y_i^{(k)})^2 + 4\lambda^2((\sum_i y_i^{(k)}(1 - 2x_i^{(k-1)}))^2 + (\sum_i (y_i^{(k)})^2)^2)
\end{aligned}$$

For the second term, we simply use $1 - 2x_i^{(k-1)} \le 1$, and

$$(\sum_i y_i^{(k)}(1 - 2x_i^{(k-1)}))^2 \le \sum_i (y_i^{(k)})^2$$

As for the third term, remember in Claim 1.3 we proved that $\|y^{(k)}\|_2^2 = \sum_i (y_i^{(k)})^2 \le \frac{1}{2\sqrt{n}} \le \frac{1}{2}$, so

$$(\sum_i (y_i^{(k)})^2)^2 \le \frac{1}{2}\sum_i (y_i^{(k)})^2$$

Take the two inequalities above, we can bound Y_k^2 as

$$Y_k^2 \leq (2+6\lambda^2)(\sum_i y_i^{(k)})^2 \leq 8(\sum_i y_i^{(k)})^2$$

Then the expectation $\mathbb{E}_{k-1}[Y_k^2]$ can be upper bounded as $8E_{k-1}[(\sum_i y_i^{(k)})^2]$. Since our updates are sub-isotropic, $\mathbb{E}[\langle c, v\rangle^2] \leq \eta \sum_i c_i^2 \mathbb{E}[v_i^2]$, so

$$\mathbb{E}[\langle 1, y^{(k)}\rangle] = \mathbb{E}[(\sum_i y_i^{(k)})^2] \leq \eta \sum_i \mathbb{E}[(y_i^{(k)})^2]$$

In all, we have

$$\mathbb{E}_{k-1}[Y_k] \leq -\lambda \mathbb{E}[\sum_i (y_i^{(k)})^2] \leq -\lambda \frac{\mathbb{E}[(\sum_i y_i^{(k)})^2]}{\eta} \leq -\frac{\lambda}{8\eta}\mathbb{E}_{k-1}[Y_k^2]$$

By setting $\alpha = \frac{\lambda}{8\eta}$ and applying the inequality on expectation from step k, $k-1$, ..., 1 recursively, we get the $\mathbb{E}[Y_k] \leq -\alpha\mathbb{E}[Y_k^2]$ as desired.

Since the criteria have been proven, we can now show that Z_k won't be too far from the initial Z_0.

Lemma 1.7. $\mathbb{P}\left[Z_k - Z_0 \geq t\right] \leq \exp(-\frac{t^2/8\eta}{t+2\mathrm{Var}[x^{(0)}]})$

Proof. With the criteria satisfied, we have $\mathbb{P}\left[Z_k - Z_0 \geq t\right] \leq \exp(-\alpha t)$. Then substitute $\alpha = \frac{\lambda}{8\eta}$ and $\lambda = \frac{t}{t+2\mathrm{Var}[x^{(0)}]} \leq 1$. □

Lastly, we prove the original theorem that

Lemma 1.8. $\Pr[X \geq (1+\delta)\mu] \leq \exp\left(-\frac{\delta^2\mu}{2\beta(1+\delta/3)}\right)$

Proof. By definition, $Z_0 = \sum_i x_i^{(0)} + \lambda \mathrm{Var}[x^{(0)}]$. In the end, all x_i have to be integers, so $\mathrm{Var}[x^{(k)}] = 0$, $Z_k = \sum_i x_i^{(k)}$. Let $v = \mathrm{Var}[x^{(0)}]$,

$$Z_k - Z_0 = \sum_i x_i^{(k)} - \sum_i x_i^{(0)} - \lambda v = X - \mu - \lambda v$$

$$\mathbb{P}\left[Z_k - Z_0 \geq t\right] = \mathbb{P}\left[X \geq \mu + t + \lambda v\right]$$

Remind again that $\lambda = \frac{t}{t+2v}$,

$$\mathbb{P}\left[Z_k - Z_0 \geq t\right] = \mathbb{P}\left[X \geq \mu + t(1+\frac{v}{t+2v})\right]$$

As $v \geq 0$, $t(1+\frac{v}{t+2v}) \leq \frac{3}{2}t$, then

$$\mathbb{P}\left[X \geq \mu + \frac{3}{2}t\right] \leq \mathbb{P}\left[Z_k - Z_0 \geq t\right] \leq \exp(-\frac{t^2/8\eta}{t+2v})$$

Now let $t' = \frac{3}{2}t$ to make it standard:

$$\mathbb{P}\left[X \geq \mu + t'\right] \leq \exp(-\frac{t'^2/12\eta}{t'+3v})$$

This is already one form of the desired. To get our standard bound, let $t' = \delta\mu$,

$$\mathbb{P}\left[X \geq (1+\delta)\mu\right] \leq \exp(-\frac{\delta^2\mu^2/12\eta}{\delta\mu + 3v}) = \exp(-\frac{\delta^2\mu}{12\eta(\delta + 3v/\mu)})$$

Since for all k, $x_k \leq 1$, $\mathrm{Var}[X^{(0)}] \leq E[X^{(0)}]$ (in other words, $v \leq \mu$). Therefore,

$$\mathbb{P}\left[X \geq (1+\delta)\mu\right] \leq \exp(-\frac{\delta^2\mu}{12\eta(\delta + 3)})$$

At last, let $\beta = 18\eta$ and it becomes the desired inequality.

□