Rounding Techniques in Approximation Algorithms

Lecture 16: Sub-Isotropic Rounding

Lecturer: Nathan Klein

1 Sub-isotropy

In the last class, we said a distribution μ over vectors was isotropic if the variance did not depend on the direction, i.e. if for all $c \in \mathbb{R}^n$ we had:

$$\mathbb{E}_{v\sim\mu}\left[\langle c, v\rangle^2\right] = \|c\|^2$$

We showed this was equivalent to the covariance matrix of μ being the identity.

We care about isotropy because if we make isotropic updates, the output should obey strong concentration bounds since we never biased the distribution towards any particular direction.[1] However, we saw last class that achieving sub-isotropic updates is impossible if we want to maintain the invariants of iterative relaxation. Indeed, if we have $c \in W^{(k)}$ (in particular, c is in the rowspace of the matrix $W^{(k)}$), then since $v \in \ker(W^{(k)})$, we will have $\langle c, v\rangle = 0$ with probability 1, so $\mathbb{E}_{v\sim\mu}\left[\langle c, v\rangle^2\right] = 0 \neq \|c\|^2$.

But, maybe this can be fixed. In fact, notice that even randomized pipage rounding did not achieve isotropic updates. There, our updates were of the form $(-1, 1, 0, \dots, 0)$ or $(1, -1, 0, \dots, 0)$ (perhaps after rearranging the coordinates). These are not isotropic. In particular, clearly taking $c = (0, 0, 1, 0, \dots, 0)$ we are again going to have the variance of $\langle c, v\rangle$ equal to 0. This should give us hope since we were nevertheless able to prove strong concentration bounds for pipage rounding.

This motivates the definition of sub-isotropic distributions. In these distributions (i) we are happy with an *upper bound* in terms of $\|c\|^2$ and (ii) we only measure isotropy with respect to the *directions we are moving in*. In addition, we will allow some slack on this upper bound: we can have some mild correlation with certain directions.

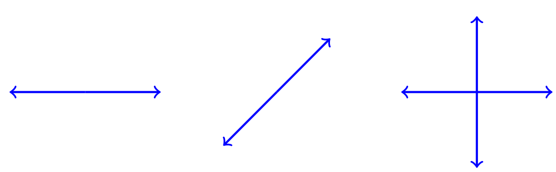

Figure 1: Consider distributions where each vector is output with equal probability. The leftmost distribution is not isotropic, but it is sub-isotropic with $\eta = 1$. The second distribution is not isotropic and is not sub-isotropic for $\eta < 2$, as we either output $(1, 1)$ or $(-1, -1)$ so for $c = (1, 1)$, $\mathbb{E}\left[\langle c, v\rangle^2\right] = 4$, and using that $\mathbb{E}\left[v_i^2 = 1\right]$ here $\sum c_i^2 \mathbb{E}\left[v_i^2\right] = \|c_i\|^2 = 2 < 4$. The third distribution is both isotropic and sub-isotropic.

Definition 1.1 ((a, η) Sub-Isotropic Distribution)**.** *Let μ be a distribution over vectors in $\mathbb{R}^n$. We say it is (a, η) sub-isotropic if:*

[1] At this point, the reader should treat this is a guess! It is not obvious why isotropic updates lead to concentration, and indeed it is necessary that the isotropic updates are not too large, as there are pairwise independent distributions that do not obey Chernoff bounds.

1. $\mathbb{E}_{v\sim\mu}[v] = \mathbf{0}$,

2. *The diagonal entries of the covariance matrix U are all at most 1 and the trace is at least an,*

3. *For all $c \in \mathbb{R}^n$, we have*
$$\mathbb{E}_{v\sim\mu}\left[\langle c, v\rangle^2\right] \le \eta \sum_{i=1}^{n} c_i^2 \mathbb{E}\left[v_i^2\right]$$

We will ultimately show that sub-isotropic updates imply concentration. But for now, some things to notice, to help gain some intuition about this definition:

1. If $\mathbb{E}\left[v_i^2\right]$ was 1 for all i, i.e. the diagonals of the covariance matrix were all 1s as in the definition of isotropy, the last term would be simply $\eta\|c\|^2$, like an upper bound on the correlation with slack η. But now we allow the variance to be, say, 0 along some directions, and as mentioned we do not care about maintaining isotropy for those directions.

2. If the entries v_i were all independent from one another (or even pairwise independent), we would obtain $\mathbb{E}_{v\sim\mu}\left[\langle c, v\rangle^2\right] = \sum_{i=1}^{n} c_i^2 \mathbb{E}\left[v_i^2\right]$, i.e. $\eta = 1$. In a pipage rounding update, we would have $\mathbb{E}_{v\sim\mu}\left[\langle c, v\rangle^2\right] < \sum_{i=1}^{n} c_i^2 \mathbb{E}\left[v_i^2\right]$.

3. The second criteria of sub-isotropy says that our distribution should not just output the 0 vector, and that it is appropriately scaled. The specific condition that the trace is at least *an* is only needed to show the running time of the algorithm is not too large.

1.1 Positive Semi-Definite Matrices

How might we construct such updates? It turns out to be enough to construct a positive semi-definite matrix with certain properties. Remember that a real symmetric matrix U is positive semi-definite (PSD) if any of the three equivalent conditions hold:

1. Its eigenvalues are all non-negative.

2. There exists a matrix $U^{1/2}$ such that $(U^{1/2})^T U^{1/2} = U$, i.e. if U has a square root.

3. $x^T U x \ge 0$ for all $x \in \mathbb{R}^n$.

A common piece of notation which is convenient here is we say $A \preceq B$ if $B - A$ is PSD. So, sometimes $A \succeq 0$ is used to denote that A is PSD. In the second definition, there may be many matrices $U^{1/2}$, but there is a unique one that is also PSD which we will refer to when we use $U^{1/2}$.

A very nice fact is as follows:

Fact 1.2. *Let $U \in \mathbb{R}^{n\times n}$ be a PSD matrix. Then, there exists a distribution μ with covariance matrix U.*

Proof. Let $r \in \mathbb{R}^n$ be a random vector where every entry is independently equal to -1 or 1, each with probability 1/2, call this distribution ν. Now, consider the random process in which we sample a vector r and then output $U^{1/2}r$. This results in a distribution μ over vectors with $\mathbb{E}_{v\sim\mu}[v] = 0$.

The covariance matrix of μ is
$$\mathbb{E}_{v\sim\mu}\left[vv^T\right] = \mathbb{E}_{r\sim\nu}\left[U^{1/2}r(U^{1/2}r)^T\right] = U^{1/2}\mathbb{E}_{r\sim\nu}\left[rr^T\right]U^{1/2} = U$$

where we used that $\mathbb{E}_{r\sim\nu}\left[rr^T\right] = I_n$ since every diagonal is 1 with probability 1, and every off-diagonal is equally likely to be -1 or 1. ☐

Note that it is also possible to use a Gaussian instead, but [Ban19] uses random ± 1 entries to help ensure the updates are bounded.

The first condition of Definition 1.1 is met by such a distribution. The second is met as long as we choose U so that the diagonals are at most 1 and the trace is at least an for whatever constant $a > 0$ is needed. The third condition looks a bit unusual, but it turns out it is also easy to put in terms of the covariance matrix:

Fact 1.3. *$\mathbb{E}_{v\sim\mu}\left[\langle c, v\rangle^2\right] \le \eta \sum_{i=1}^n c_i^2 \mathbb{E}\left[v_i^2\right]$ (condition (3) of Definition 1.1) holds for μ with covariance matrix U if and only if $U \preceq \eta \cdot diag(U)$.*

Proof. $U \preceq \eta \cdot \mathrm{diag}(U)$ is the same as saying $\eta \cdot \mathrm{diag}(U) - U$ is PSD, or using the equivalencies above,

$$c^T(\eta \cdot \mathrm{diag}(U) - U)c \ge 0$$

for all $c \in \mathbb{R}^n$. This implies $c^T U c \le \eta \cdot c^T \mathrm{diag}(U) c$. But, remember from last lecture,

$$\mathbb{E}_{v\sim\mu}\left[\langle c, v\rangle^2\right] = c^T \mathbb{E}_{v\sim\mu}\left[vv^T\right] c = c^T U c$$

So, chaining everything together,

$$\mathbb{E}_{v\sim\mu}\left[\langle c, v\rangle^2\right] = c^T U c \le \eta \cdot c^T \mathrm{diag}(U) c = \eta \sum_{i=1}^n c_i^2 \mathbb{E}\left[v_i^2\right]$$

as desired. ☐

In the next class, we will prove the following theorem of Bansal and Garg [BG17]. By the above discussion, this theorem is sufficient to obtain (a, η) sub-isotropic updates.

Theorem 1.4. *Let $W \subset \mathbb{R}^n$ be a subspace of dimension $(1-\delta)n$. Then there is a PSD matrix $U \in \mathbb{R}^{n\times n}$ satisfying:*

1. *$\langle ww^T, U\rangle = 0$ for all $w \in W$,*[2]
2. *The diagonal entries are at most 1 and the trace is at least an,*
3. *$U \preceq \eta \cdot diag(U)$, or equivalently $U - \eta \cdot diag(U)$ is PSD.*

The first condition implies that we stay in the kernel of W. This is because for any $w \in W$ it gives

$$0 = \langle ww^T, U\rangle = \mathrm{Tr}(ww^T U) = \mathrm{Tr}(w^T U w) = w^T U w = \|U^{1/2} w\|^2,$$

where we used the cyclicity of trace. This implies that $U^{1/2}w$ and its transpose $w^T U^{1/2}$ are the all 0s vector. But now setting an update to be $U^{1/2}r$, for any $w \in W$ we obtain $w^T U^{1/2} r = \mathbf{0}$, so our updates are orthogonal to W.

[2]This is known as the Frobenius inner product.

1.2 Rounding Algorithm

We can finally define the sub-isotropic rounding algorithm. Given an iterative relaxation procedure, we will initialize $x^{(0)}$ as the initial solution to the LP of interest. Then, until we reach an integral point, continuously:

1. Ask the iterative relaxation procedure for a subspace $W^{(k)}$, and call Theorem 1.4 to obtain a PSD matrix U with the necessary conditions. Use Fact 1.2 to construct a (a, η) sub-isotropic distribution $\mu^{(k)}$ over vectors in $\ker(W^{(k)})$.
2. Sample v from $\mu^{(k)}$, and set $y^{(k)} = \epsilon v$ where ϵ is the largest constant so that $x^{(k)} \pm \epsilon v \in [0, 1]^n$ and $\epsilon \leq \frac{1}{2n^{3/2}}$. Then set $x^{(k+1)} = x^{(k)} + y^{(k)}$.

The output obeys whatever guarantee the iterative relaxation algorithm had since we always move in the kernel of $W^{(k)}$. Secondly, since $\mathbb{E}\left[y^{(k)}\right] = 0$ for all updates, $\mathbb{E}[x] = x_0$. So, it remains to show Theorem 1.4 and to show that sub-isotropic updates imply concentration, which we will do in the next lecture.[3]

Note that we will not analyze the running time of the algorithm, but it is not difficult to prove that it runs in expected polynomial time. See [Ban19] for details.

References

[Ban19] Nikhil Bansal. "On a generalization of iterated and randomized rounding". In: *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing*. STOC 2019. Phoenix, AZ, USA: Association for Computing Machinery, 2019, 1125–1135. ISBN: 9781450367059. DOI: 10.1145/3313276.3316313 (cit. on pp. 3, 4).

[BG17] Nikhil Bansal and Shashwat Garg. "Algorithmic discrepancy beyond partial coloring". In: *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing*. STOC 2017. Montreal, Canada: Association for Computing Machinery, 2017, 914–926. ISBN: 9781450345286. DOI: 10.1145/3055399.3055490 (cit. on p. 3).

[3] While we started discussing martingales in this lecture, we will move this to the next lecture notes.