Efficient error reduction in probabilistic algorithms

Let $\mathcal{L} \subseteq \{0,1\}^*$ be a language.

Given $x \in \{0,1\}^*$, want to decide whether $x \in \mathcal{L}$ using poly-time algorithm with randomness.

$\mathcal{L} \in RP$ if $\exists$ algorithm s.t.

$$\text{If } x \in \mathcal{L} \text{ then } A(x) = 1$$
$$\text{If } x \notin \mathcal{L} \text{ then } \Pr[A(x) = 1] \leq 1/3$$

$\mathcal{L} \in BPP$ if $\exists$ algorithm s.t.

$$\text{If } x \in \mathcal{L} \text{ then } \Pr[A(x) = 1] \geq 2/3$$
$$\text{If } x \notin \mathcal{L} \text{ then } \Pr[A(x) = 1] \leq 1/3$$

All probabilities over internal randomness of A

View A as taking two inputs

- x
- $r \in \{0,1\}^{kn}$ random string

$\cancel{x \in}$ Focus on RP for simplicity

$$x \in \mathcal{L} \implies \Pr_r[A(x,r)=1] = 1$$
$$x \in \mathcal{L} \implies \Pr_r[A(x,r)=1] \leq 1/3$$

Want to make error prob exponentially small

Run A t times with independent randomness

Error prob $\leq (1/3)^t$

Randomness used $t \cdot kn$ bits

Can we get (almost) same exponential decay with less randomness? Yes!

Suppose B is set of bad random strings for which A fails

Set $M = 2^m$ $\quad |B| = \beta \cdot M$ for $\beta \le 1/3$

Pick an (M, d, λ) - spectral expander $G = (V, E)$ that is strongly explicit.

Let A' be following algorithm:

(i) Pick $v_1 \in V = [M]$ uniformly at random
(ii) Do a length $(t-1)$ random walk on G, say $(v_1, v_2, \ldots, v_t)$
(iii) Return 1 if $\forall v_i$ $A(x, v_i) = 1$ and 0 otherwise.

Then
- If $x \in \mathcal{L}$ then $\Pr[A'(x) = 1] = 1$
- If $x \notin \mathcal{L}$ then $\Pr[A'(x) = 1] \le (\beta + \lambda)^{t-1}$

And A' uses only

$$\boxed{m + (t-1)\log d = m + O(t)}$$

bits of randomness

Can get similar result for two-sided errors in BPP, but we will skip this

Random walks on expanders resemble independent sampling

Abstract setting

Universe $[n]$

Bad set $B \subseteq [n]$

~~Wantto~~ to

- get random samples avoiding B
- minimize # random bits used

<u>Solution</u>

- Build expander graph on $[n]$
- Sample according to random walk
- Statistics of hitting B with this highly dependent sampling almost the same as with completely independent sample!

Suppose G is (n, d, λ)-spectral expander

All we know about B is $|B| = \beta n$

Want to sample at least one vertex outside of B

Uniformly sample t vertices $v_1, v_2, \ldots, v_t$

$$\Pr\left[\ \forall i \ \ v_i \in B\ \right] \leq \beta^t$$

Uses $t \log n$ random bits.

As a warmup for expander random walk, consider $t = 2$.

Recall Expander Mixing Lemma

$$\left| |E(S,T)| - \frac{d}{n}|S||T| \right| \le \lambda d \sqrt{|S||T|}$$
$$\le \lambda d n$$

for $S, T \subseteq V(G)$

Rewrite as

$$\left| \frac{|E(S,T)|}{dn} - \frac{|S||T|}{n^2} \right| \le \lambda$$

Sample ordered pair of vertices (i, j)
Consider it a success if $i \in S$, $j \in T$

(i) <u>Uniform sampling</u>
success probability $\frac{|S|}{n} \cdot \frac{|T|}{n}$

(i) <u>Pick i randomly and then j as random neighbour</u>
Success probability $\frac{|E(S,T)|}{dn}$

double-counting #edges because we care about order

Expander Mixing Lemma says that although sample spaces very different success probabilities differ by ~~at~~ most λ

Now let us look at longer walks

Pick $v_1 \in V$ uniformly at random

Pick v_i as random neighbour of v_{i-1}, for $i = 2, \dots, t$

Let (B, t) denote event that all $v_i \in B$

<u>THM 1</u> [Ajtai, Komlós, Szemerédi '87]
[Alon, Feige, Wigderson, Zuckerman '95]

Let G (n, d, λ) - spectral expander and $B \subseteq [n]$, $|B| = \beta n$. Then

$$\Pr[(B, t)] \leq (\beta + \lambda)^{t-1}$$

Let P be projection matrix on coordinates of B, i.e.,

$$(P)_{ij} = \begin{cases} 1 & \text{if } i = j \in B \\ 0 & \text{otherwise} \end{cases}$$

$\forall i \; q_i \geq 0$
$\|q\|_1 = \sum_i q_i = 1$

For any probability distribution q on $[n]$

$$\|Pq\|_1 = \text{probability to sample } i \in B$$

Also note that P is a contraction wrt L_2-norm (and L_1-norm)

$$\|Pq\|_2 \leq \|q\|_2$$

$A = (a_{ij})_{i,j \in [n]}$

Let as before $A = \widetilde{A}_G$ be the normalized adjacency matrix. Then:

<u>LEM 2</u>
$$\Pr[(B, t)] = \|(PA)^{t-1} P \mathbb{1}/n\|_1$$

Proof Let $q_1 = (P \mathbb{1}/n)$

Then $(q_1)_i$ measures $Pr[(\text{starting in } i) \text{ and } (i \in B)]$
Take one step and let distribution be p_2

Then $(p_2)_i = Pr[(\text{in vertex } i \text{ after on step}) \text{ and } (\text{start in } B)]$

$$= \sum_{j=1}^{n} a_{ij}(q_1)_j$$

And $p_2 = AP\mathbb{1}/n$

If we are now outside of B the walk should be ignored, so the probability of being in B first two steps is measured by

$$\| PAP \mathbb{1}/n \|_1$$

Repeating for $t-1$ steps we obtain

$$Pr[(B,t)] = \| (PA)^{t-1} P \mathbb{1}/n \|_1 \qquad \boxed{E_1}$$

If we move from L_1-norm to L_2-norm we can prove:

LEM 3 For any $v \in \mathbb{R}^n$ it holds that

$$\| PAPv \|_2 \le (\beta + \lambda) \|v\|_2$$

From Lem 3 we get

COR 4

$$\| (PA)^{t-1} Pv \|_2 \le (\beta + \lambda)^{t-1} \|v\|_2$$

Proof of Cor 4

By induction. Note that $P^2 v = Pv$, because P is a projection

$$\|(PA)^{t-1} Pv\|_2 = \|(PA)(PA)^{t-2} Pv\|_2$$
$$= \|(PAP)(PA)^{t-2} Pv\|_2$$
$$\le (\beta + \lambda) \|(PA)^{t-2} Pv\|_2$$
$$\le (\beta + \lambda)^{t-1} \|v\|_2 . \qquad \boxed{\not{z}}$$

Proof of Lem 3

Start with three observations

(i) W.l.o.g. assume v is supported on B, i.e. $\{ i \mid v_i \neq 0 \} \subseteq B$ or $Pv = v$

If not, prove Lem 3 for Pv instead

Ok since $\|Pv\|_2 \le \|v_2\|$ as noted above

(ii) W.l.o.g. assume $v_i \ge 0 \quad \forall i$

Doesn't change RHS and can only increase LHS

(iii) Since $\|\alpha \cdot v\| = |\alpha| \|v\| \quad (\alpha \in \mathbb{R})$

we can rescale so that $\sum_{i=1}^n v_i = 1$

(unless $v = 0$, but then Lem 3 is also true :-))

So we can write

$$Pv = v = \mathbb{1}/n + z$$

where $z \perp \mathbb{1}$

Hence

$$PAPv = PAv$$
$$= PA\,\mathbb{1}/n + PAz$$
$$= P\,\mathbb{1}/n + PAz$$

and by the triangle inequality

$$\|PAPv\|_2 \leq \|P\,\mathbb{1}/n\|_2 + \|PAz\|_2$$

The lemma follows if we can show

Claim A $\quad \|P\,\mathbb{1}/n\|_2 \leq \beta \|v\|_2$

Claim B $\quad \|PAz\|_2 \leq \lambda \|v\|_2$

Proof of Claim A:

Since $|B| \leq \beta n$, for any vector w

Pw is supported on at most βn coordinates.

Hence

$$\|P\,\mathbb{1}/n\|_2 \leq \sqrt{\beta n \cdot \tfrac{1}{n^2}} = \sqrt{\beta/n} \qquad (i)$$

Using Cauchy-Schwarz we obtain

$$1 = \sum v_i = \langle \mathbb{1}_B, v \rangle$$
$$\leq \|\mathbb{1}_B\|_2 \; \|v\|_2$$
$$\leq \sqrt{\beta n} \; \|v\|_2$$

and hence

$$\sqrt{\frac{\beta}{n}} \leq \beta \|v\|_2 \qquad (ii)$$

IX

Combining (i) and (ii) yields

$$\| P \mathbb{1}/n \|_2 \leq \beta \| v \|_2$$

proving Claim A.

The proof of Claim B is pretty much immediate given what we have done so far.

We have

$$\| P A z \|_2 \leq \| A z \|_2 \qquad (iii)$$

since P contracts vectors

Since $z \perp \mathbb{1}$, z must shrink by at least a factor $\lambda_2 = \lambda(G)$

and

$$\| A z \|_2 \leq \lambda \| z \|_2 \qquad (iv)$$

Since $z \perp 1$ we have

$$\| v \|_2^2 = \| 1/n \|_2^2 + \| z \|_2^2 > \| z \|_2^2$$

so

$$\lambda \| z \|_2 < \lambda \| v \|_2 \qquad (v)$$

and combining (iii) – (v) yields Claim B. $\boxed{\lrcorner}$

Now we can easily prove Thm 1

Proof of Thm 1

Recall that for any $w \in \mathbb{R}^n$

$$\|w\|_1 \leq \sqrt{n}\, \|w\|_2$$

Thus

$$\|(PA)^{t-1} P \mathbb{1}/n\|_1 \leq \sqrt{n}\, \|(PA)^{t-1} P \mathbb{1}/n\|_2$$

$$[\text{Cor } 4] \qquad \leq \sqrt{n}\, (\beta+\lambda)^{t-1}\, \|\mathbb{1}/n\|_2$$

$$= \sqrt{n}\, (\beta+\lambda)^{t-1}\, \frac{1}{\sqrt{n}}$$

$$= (\beta+\lambda)^{t-1} \qquad \blacksquare$$

Note that we lose an additive 1 in exponent.
If the expander is really good,
so that

$$\lambda < \beta/6$$

compared to $\frac{|B|}{n}$

then [AFKZ 95] showed
you can sort of recover this to get

$$\Pr[(B,t)] \leq \beta(\beta+2\lambda)^{t-1}$$

(and also than probability cannot scale much better:

$$\Pr[(B,t)] \geq \beta(\beta - 2\lambda)^{t-1}\text{)}$$

If one wants efficient error for BPP, one can instead prove bounds on the fraction of steps of the random walk inside the bad set B

<u>THM 5</u> (Expander Chernoff Bound)

Let G (n, d, λ)-spectral expander

Let $B \subseteq [n]$, $|B| = \beta n$

Let $B_i = \begin{cases} 1 & \text{if } i\text{th random vertex } v_i \in B \\ 0 & \text{otherwise} \end{cases}$

Then for every $\delta > 0$ it holds that

$$\Pr\left[\left| \frac{\sum_{i=1}^{t} B_i}{t} - \beta \right| > \delta \right] < 2 \exp\left(-\frac{(1-\lambda)\delta^2 t}{4} \right)$$

Explicit construction of expander graphs

Exist several highly practical and efficient constructions of expander graphs.
However, analysis typically uses deep facts in number theory

We will see more complicated (but strongly explicit) construction but with elementary analysis
Based on [Reingold, Vadhan & Wigderson '00]

Uses several types of graph products
$$G \circ G' \longrightarrow H$$
Need to analyze how properties of G, G' carry over to H.
Interested in parameters:

- \# vertices n
- degree d
- 2nd largest (absolute) eigenvalue λ

Graphs will be <u>multigraphs</u>
Multiple copies of edges and self loops OK
(Mostly technical convenience)

Three different products

(1) Matrix product $G_1 \cdot G_2$

# vertices n	:	same
degree d	:	increases
expansion λ	:	improves

(2) Tensor product $G_1 \otimes G_2$

n	:	increases
d	:	increases
λ	:	preserved

(3) Replacement product $G_1 \circledR G_2$

n	:	increases a little
d	:	decreases a lot
λ	:	only dereases a little

Game plan

(i) Start with good constant size expander (can find using brute force)

(ii) Increase graph size with tensor product

(iii) Get supergreat expansion by matrix product

(iv) Reduce degree by replacement product

(v) To get next graph, repeat from (ii)

Graph representations

(A) Normalized adjacency matrix $A = \tilde{A}_G$

(B) Rotation map

$$\hat{G} : [n] \times [d] \to [n] \times [d]$$

Many possible choices

Number neighbours of vertex $1, 2, \ldots, d$

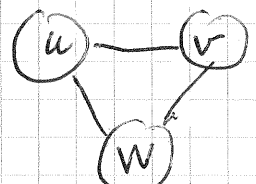
$\hat{G}(v, i) = (u, j)$ if u ith neighbour of v and v jth neighbour of u

$$\hat{G}(\hat{G}(v, i)) = (v, i)$$

$\hat{G}$ is a permutation on $[n] \times [d]$
(onto/surjective and one-to-one/injective)

Will switch between representations as convenient.

Example 6



$= G$

Rotation map

$$\hat{G}(u, 1) = (v, 2)$$
$$\hat{G}(u, 2) = (w, 2)$$
$$\hat{G}(v, 1) = (w, 1)$$
$$\hat{G}(v, 2) = (u, 1)$$
$$\hat{G}(w, 1) = (v, 1)$$
$$\hat{G}(w, 2) = (u, 2)$$

$$\begin{pmatrix} 0 & 1/2 & 1/2 \\ 1/2 & 0 & 1/2 \\ 1/2 & 1/2 & 0 \end{pmatrix} = \tilde{A}_G$$

MATRIX PRODUCT

G (n, d, λ) - spectral expander
G' (n, d', λ') - spectral expander

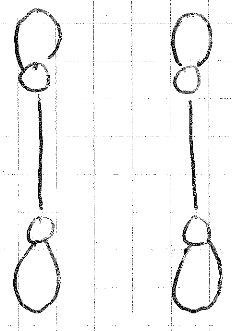
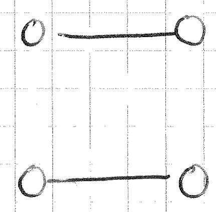
$G \cdot G'$ is the graph whose normalized adjacency matrix is

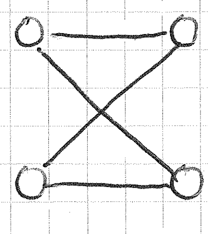
$$\tilde{A}_{G'} \tilde{A}_{G} = A' \cdot A$$

$G \cdot G'$ has edge (u, v) for every path $u \to w \to v$ s.t. $(u, w) \in E(G)$
$(w, v) \in E(G')$

Still n vertices
Degree $d \cdot d'$

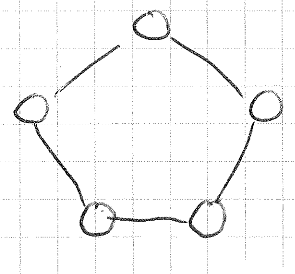
Example 7

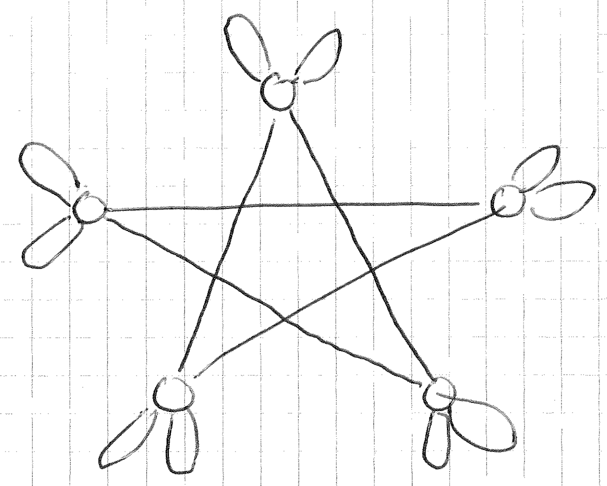
$G =$  $G' =$ 

$G G'$ 

Of particular interest for us is $G' = G$
GRAPH SQUARING

Example 8

$G =$ 

G^2 

LEMMA 9 $\quad \lambda(G^2) = \lambda(G)^2$

Proof: Follows since eigenvalues of $(\tilde{A}_G)^2$ are those of $\tilde{A}_G$ squared (which in turn is left as an exercise.

To compute $\widehat{G^2}(v,(i,i'))$ $\cancel{\widehat{G^2}(v,i)}$, first find $(u,j) = \widehat{G}(v,i)$ and then $\widehat{G}(u,i') = (w,j')$

Hence can do matrix product efficiently also in rotation map representation.

TENSOR PRODUCT

for brevity

G (n, d, λ) - graph

G' (n', d', λ') - graph

$G \otimes G'$ has nn' vertices

degree dd'

Index vertices $(u,v) \in [n] \times [n']$

and neighbours/edges $(i,j) \in [d] \times [d']$

Then $\widehat{G \otimes G'}((u,v),(i,j)) = ((u',v'),(i',j'))$

wher $\hat{G}(u,i) = (u',i')$

$\widehat{G'}(v,j) = (v',j')$

That is, can move from (u,v) to (u',v') by taking step (u,u') in G and (v,v') in G'

Let $A = \tilde{\hat{A}}_G$ and $A' = \tilde{\hat{A}}_{G'}$

Then

tensor product

$A = (a_{ij})$

$$\tilde{\hat{A}}_{G \otimes G'} = A \otimes A' =$$

$$\begin{pmatrix} a_{11} A' & a_{12} A' & \cdots & a_{1n} A' \\ a_{21} A' & a_{22} A' & \cdots & a_{2n} A' \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} A' & a_{n2} A' & \cdots & a_{nn} A' \end{pmatrix}$$

n^2 copies of A', with (i,j)th copy scaled by a_{ij}.

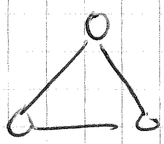
Yet another description

Create cluster C_u with n' vertices for each $u \in V(G)$

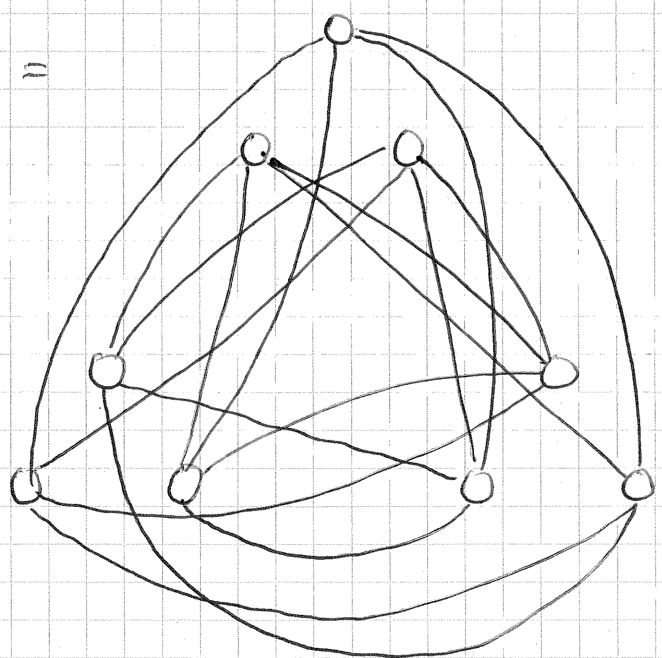
For neighbours $u, u' \in V(G)$, clusters C_u and $C_{u'}$ form bipartite double cover of G'

That is, $(u,v) - (u',v')$ edge iff $(v,v') \in E(G')$

EXAMPLE 10

Let $G =$  and look at $G \otimes G$

$G \otimes G =$



LEMMA 11

If G is an (n, d, λ)-graph and
G' is an (n', d', λ')-graph, then
$G \otimes G'$ is an $(nn', dd', \max(\lambda, \lambda'))$-graph

Intuition: Take a T-step random walk on $G \otimes G'$.
This is similar to taking two independent random walks on G and G'. If both random walks get close to uniform within T steps, then so will walk on $G \otimes G'$.

Proof: Follows immediately from juggling eigenvalues and -vectors of tensored matrices.

PROPOSITION 12

If A has eigenvalues $\mu_1, \ldots, \mu_n$ with eigenvectors $u_1, \ldots, u_n$ and A' has eigenvalues $\mu'_1, \ldots, \mu'_m$ and eigenvectors $u'_1, \ldots, u'_m$, then the eigenvalues of $A \otimes A'$ are $\mu_i \cdot \mu'_j$, $i \in [n]$, $j \in [m]$,
with eigenvectors $u_i \otimes u'_j$, $i \in [n]$, $j \in [m]$.

For $v \in \mathbb{R}^n$, $w \in \mathbb{R}^m$, $v \otimes w \in \mathbb{R}^{nm}$
is the vector s.t. $(v \otimes w)_{(i,j)} = v_i \cdot w_j$

or in block notation
$$\begin{pmatrix} v_1 \cdot w \\ v_2 \cdot w \\ \vdots \\ v_n \, w \end{pmatrix}$$

Proof of Prop 12

Suppose $Av = \mu v$; $A'v' = \mu' v'$

$$(A \otimes A')(v \otimes v') = \begin{pmatrix} a_{11}A' & a_{12}A' & & \\ a_{21}A' & a_{22}A' & & \\ & & \ddots & \\ & & & a_{nn}A' \end{pmatrix} \begin{pmatrix} v_1 v' \\ v_2 v' \\ \vdots \\ v_n v' \end{pmatrix}$$

Look at the ith block in this product

$$\left(\sum_j a_{ij} A'\right)(v_j v') =$$

$$\left(\sum_j a_{ij} v_j\right)(A'v') =$$

$$\left(\sum_j a_{ij} v_j\right)(\cancel{A'}\mu' v') =$$

$$(Av)_i\,(\cancel{A'}\mu' v') =$$

$$\mu v_i\, \mu' v' = (\mu\mu')\, v_i v'$$

so $(A \otimes A')(v \otimes v') = (Av) \otimes (A'v')$
$\qquad\qquad\qquad\qquad\qquad = \mu\mu'(v \otimes v')$

Furthermore if $v_1 \neq v_2$ or $v_1' \neq v_2'$ for eigenvectors v_i of A and v_i' of A', then $(v_1 \otimes v_1')$ and $(v_2 \otimes v_2')$ are orthogonal. Hence these vectors give all eigenvectors, and $\mu_i \mu_j'$ are all eigenvalues.

Proof of Lem 11 XXCT

Largest eigenvalue of $A \otimes A'$ is $\mu_1 \cdot \mu_1' = 1$
Next largest must be one of $\mu_1 \cdot \mu_2' = \mu_2'$ or $\mu_2 \cdot \mu_1' = \mu_2$, and smallest must be one of

$\mu_1 \cdot \mu_n' = \mu_n'$ or $\mu_n \cdot \mu_1' = \mu_n$. Hence

$\lambda(G \otimes G') \leq \max(\lambda(G), \lambda(G'))$
as claimed.

So, what have we seen so far?

By tensoring can make graph larger
- degree increases
- expansion stays the same

Then can take matrix product
- expansion improves a lot
- but also degree increases a lot
- # vertices stays the same

By now expansion is great, but degree is terrible. We want graphs of
constant degree

Reduce degree by third graph product

Replacement product