Extracting Mergers and Projections of Partitions

Swastik Kopparty, **Vishvajeet N**

Univ. of Toronto, Univ. of Edinburgh

RANDOM 2023

Randomness for algorithms

- Randomized algorithms require access to many *unbiased* and *uncorrelated* random bits

Randomness for algorithms

- Randomized algorithms require access to many *unbiased* and *uncorrelated* random bits
- Sources might be biased and correlated! ("impure/weak sources")

Randomness for algorithms

- Randomized algorithms require access to many *unbiased* and *uncorrelated* random bits
- Sources might be biased and correlated! ("impure/weak sources")
- Algorithms should still work provably

Randomness for algorithms

- Randomized algorithms require access to many *unbiased* and *uncorrelated* random bits
- Sources might be biased and correlated! ("impure/weak sources")
- Algorithms should still work provably

We study objects that purify randomness - condensers and extractors

Randomness condensers

Condenser is an object which “purifies” a “weak source” of randomness, making it “less impure”

Randomness extractors

Extractor is an object which "purifies" a "weak source" of randomness, making it "completely pure"

Weak random source

Measure of purity of source: min-entropy

Definition (k-source)

X supported on $\{0,1\}^n$ such that for all x, $\Pr[X = x] \leq 2^{-k}$

Weak random source

Measure of purity of source: min-entropy

Definition (k-source)

X supported on $\{0,1\}^n$ such that for all x, $\Pr[X = x] \leq 2^{-k}$

Definition (entropy-rate)

k-source on $\{0,1\}^n$ has min-entropy rate $\frac{k}{n}$

Uniform distribution U_n $\Leftrightarrow$ source with min-entropy n $\Leftrightarrow$ entropy-rate 1

Condensers

Input: k-source on $\{0,1\}^n$

Condensers

Input: k-source on $\{0,1\}^n$

Output: distribution on $\{0,1\}^m$ which is ϵ-close to some distribution
with (high) min-entropy k‘

Condensers

Input: k-source on $\{0,1\}^n$

Output: distribution on $\{0,1\}^m$ which is ϵ-close to some distribution with (high) min-entropy k'

we would like $\frac{k'}{m} > \frac{k}{n}$

"increases min-entropy rate of source"

Extractors

Input: k-source on $\{0,1\}^n$

Extractors

Input: k-source on $\{0,1\}^n$

Output: distribution on $\{0,1\}^m$ which is ϵ-close to having entropy rate 1

Seeded extractors

allow access to very few purely-random bits, "seed"

Seeded extractors

allow access to very few purely-random bits, "seed"

Definition (seeded (k, ϵ)-extractor)

$EXT : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ s.t. for every k-source X, distribution of $EXT(X, U_d)$ is ϵ-close to U_m

Seeded extractors

allow access to very few purely-random bits, "seed"

Definition (seeded (k,ϵ)-extractor)

$EXT : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ s.t. for every k-source X, distribution of $EXT(X, U_d)$ is ϵ-close to U_m

function EXT when chosen at random is an extractor:

- seed-length $d = \log(n-k) + 2\log(\frac{1}{\epsilon}) + O(1)$
- output-length $m = k + d - 2\log\frac{1}{\epsilon} - O(1)$

Seeded extractors

allow access to very few purely-random bits, "seed"

Definition (seeded (k, ϵ)-extractor)

$EXT : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ s.t. for every k-source X, distribution of $EXT(X, U_d)$ is ϵ-close to U_m

function EXT when chosen at random is an extractor:

- seed-length $d = \log(n-k) + 2\log(\frac{1}{\epsilon}) + O(1)$
- output-length $m = k + d - 2\log\frac{1}{\epsilon} - O(1)$

optimal! $d \geq \log(n-k) + 2\log(\frac{1}{\epsilon}) + O(1)$ [RTS00],[AGO$^+$20]

Seeded condensers

we must have $k' < k + d$, and usually want improved entropy-rate $\frac{k'}{m} > \frac{k}{n}$

Definition (seeded condenser)

$COND : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ s.t. for every k-source X, distribution of $COND(X, U_d)$ is ϵ-close to some source on $\{0,1\}^m$ with min-entropy k'

Seeded condensers

we must have $k' < k + d$, and usually want improved entropy-rate $\frac{k'}{m} > \frac{k}{n}$

Definition (seeded condenser)

$COND : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ s.t. for every k-source X, distribution of $COND(X, U_d)$ is ϵ-close to some source on $\{0,1\}^m$ with min-entropy k'

function $COND$ when chosen at random is a condenser:

- seed-length $d = \log(n - m + 1) + \log(\frac{1}{\epsilon}) + O(1)$
- output min-entropy $k + d$ for any $k < m - d - \log(\frac{1}{\epsilon}) - O(1)$

Somewhere-random sources

Definition (t-part somewhere-random source)

$X_1, \ldots X_t \in (\{0,1\}^n)^t$ s.t. there exists an (unknown) i, s.t. X_i is uniform over $\{0,1\}^n$

special case of a general k-source with entropy rate $1/t$

Mergers

Definition (Merger)

$M : (\{0,1\}^n)^t \times \{0,1\}^d \to \{0,1\}^m$ that purifies every input t-part somewhere-random source X

- studied in the "condensing regime":
  - output is ϵ-close to some source with min-entropy rate $(1-\delta)$; for $\delta > 0$

Mergers

Definition (Merger)

$M : (\{0,1\}^n)^t \times \{0,1\}^d \to \{0,1\}^m$ that purifies every input t-part somewhere-random source X

- studied in the "condensing regime":
  - output is ϵ-close to some source with min-entropy rate $(1 - \delta)$; for $\delta > 0$
- historically crucial in advancement of new extractors [TS00],[LRVW03],[DKSS09], ...

Mergers

Definition (Merger)

$M : (\{0,1\}^n)^t \times \{0,1\}^d \to \{0,1\}^m$ that purifies every input t-part somewhere-random source X

- studied in the "condensing regime":
  - output is ϵ-close to some source with min-entropy rate $(1-\delta)$; for $\delta > 0$
- historically crucial in advancement of new extractors [TS00],[LRVW03],[DKSS09], ...
- [DKSS09]: explicit (condensing) mergers exist with seed-length $d = \frac{1}{\delta} \cdot \log(\frac{2t}{\epsilon})$
  - builds on prior work on Kakeya Set problem [Dvi08],[DW09]
  - constant seed-length since t usually constant

Mergers

Definition (Merger)

$M : (\{0,1\}^n)^t \times \{0,1\}^d \to \{0,1\}^m$ that purifies every input t-part somewhere-random source X

- studied in the "condensing regime":
  - output is ϵ-close to some source with min-entropy rate $(1-\delta)$; for $\delta > 0$
- historically crucial in advancement of new extractors [TS00],[LRVW03],[DKSS09], . . .
- [DKSS09]: explicit (condensing) mergers exist with seed-length $d = \frac{1}{\delta} \cdot \log(\frac{2t}{\epsilon})$
  - builds on prior work on Kakeya Set problem [Dvi08],[DW09]
  - constant seed-length since t usually constant
- condensing mergers perform much better than condensers, which require $\log n$ seed in this regime

Extracting mergers

We study mergers in the "*extracting* regime":

Definition (seeded extracting merger)

$E : \{\{0,1\}^n\}^t \times \{0,1\}^d \to \{0,1\}^m$ s.t. for every t-part somewhere-random source X, distribution of $E(X, U_d)$ is ϵ-close to U_m

Extracting mergers

We study mergers in the "*extracting* regime":

Definition (seeded extracting merger)

$E : \{\{0,1\}^n\}^t \times \{0,1\}^d \to \{0,1\}^m$ s.t. for every t-part somewhere-random source X, distribution of $E(X, U_d)$ is ϵ-close to U_m

Do there exist extracting mergers requiring small, maybe constant seed?

OR

Are they completely overshadowed by standard extractors for min-entropy rate $\frac{1}{t}$?

Extracting mergers (upper bound)

Theorem

Let n, t be integers and $\epsilon > 0$. Then for any integer $m \leq n$, setting:
$$d = \log m + \log(t-1) + 2\log\frac{1}{\epsilon} + O(1),$$
there exists a function $E : (\{0,1\}^n)^t \times \{0,1\}^d \to \{0,1\}^m$ that is an ϵ-extracting merger

- we extract $\mathrm{poly}(\frac{1}{\epsilon})$ fully-random bits with constant $O(\log t + \log\frac{1}{\epsilon})$ bits of seed

Extracting mergers (upper bound)

Theorem

Let n, t be integers and $\epsilon > 0$. Then for any integer $m \leq n$, setting:

$$d = \log m + \log(t-1) + 2\log\frac{1}{\epsilon} + O(1),$$

there exists a function $E : (\{0,1\}^n)^t \times \{0,1\}^d \to \{0,1\}^m$ that is an ϵ-extracting merger

- we extract $\mathrm{poly}(\frac{1}{\epsilon})$ fully-random bits with constant $O(\log t + \log\frac{1}{\epsilon})$ bits of seed
- standard extractor requires $\Theta(\log n + \log t + \log(\frac{1}{\epsilon}))$ bits to extract a single bit!

Extracting mergers (upper bound)

Theorem

Let n, t be integers and $\epsilon > 0$. Then for any integer $m \leq n$, setting:
$$d = \log m + \log(t-1) + 2\log\frac{1}{\epsilon} + O(1),$$
there exists a function $E : (\{0,1\}^n)^t \times \{0,1\}^d \to \{0,1\}^m$ that is an ϵ-extracting merger

- we extract $\mathrm{poly}(\frac{1}{\epsilon})$ fully-random bits with constant $O(\log t + \log\frac{1}{\epsilon})$ bits of seed
- standard extractor requires $\Theta(\log n + \log t + \log(\frac{1}{\epsilon}))$ bits to extract a single bit!
- function taken at random with constant seed-length is NOT an extracting merger!

Extracting mergers (upper bound)

Theorem

Let n, t be integers and $\epsilon > 0$. Then for any integer $m \leq n$, setting:

$$d = \log m + \log(t-1) + 2\log\frac{1}{\epsilon} + O(1),$$

there exists a function $E : (\{0,1\}^n)^t \times \{0,1\}^d \to \{0,1\}^m$ that is an ϵ-extracting merger

- E extracts $\Omega(n)$ fully-random bits with $\Omega(\log n)$ bits of seed

Extracting mergers (upper bound)

Theorem

Let n, t be integers and $\epsilon > 0$. Then for any integer $m \leq n$, setting:

$$d = \log m + \log(t-1) + 2\log\frac{1}{\epsilon} + O(1),$$

there exists a function $E : (\{0,1\}^n)^t \times \{0,1\}^d \to \{0,1\}^m$ that is an ϵ-extracting merger

- E extracts $\Omega(n)$ fully-random bits with $\Omega(\log n)$ bits of seed
- no better than using standard extractor which requires $\Theta(\log n)$ bits of seed!

Extracting mergers (upper bound)

Theorem

Let n, t be integers and $\epsilon > 0$. Then for any integer $m \leq n$, setting:

$$d = \log m + \log(t-1) + 2\log\frac{1}{\epsilon} + O(1),$$

there exists a function $E : (\{0,1\}^n)^t \times \{0,1\}^d \to \{0,1\}^m$ that is an ϵ-extracting merger

- E extracts $\Omega(n)$ fully-random bits with $\Omega(\log n)$ bits of seed
- no better than using standard extractor which requires $\Theta(\log n)$ bits of seed!

what is the true seed-length requirement?

Extracting mergers (lower bound)

Theorem ($t = 2$)

Let $E : (\{0,1\}^n)^2 \times \{0,1\}^d \to \{0,1\}^m$ be a ϵ-extracting merger.
Then for $\epsilon \geq 2^{-\Omega(m)}$, we have:

$$d \geq \log m + \log \frac{1}{\epsilon} - O(1)$$

and for $\epsilon < 2^{-\Omega(m)}$, we have:

$$d \geq \Omega(m)$$

Extracting mergers (lower bound)

Theorem ($t = 2$)

Let $E : (\{0,1\}^n)^2 \times \{0,1\}^d \to \{0,1\}^m$ be a ϵ-extracting merger. Then for $\epsilon \geq 2^{-\Omega(m)}$, we have:

$$d \geq \log m + \log \frac{1}{\epsilon} - O(1)$$

and for $\epsilon < 2^{-\Omega(m)}$, we have:

$$d \geq \Omega(m)$$

"if you want to extract all bits, there is no advantage in knowing that source is somewhere-random"

Extracting mergers (lower bound)

Proof overview:

$E : [N] \times [N] \times [D] \to [M]$ be an ϵ-extractor for somewhere-random source (X, Y)

Extracting mergers (lower bound)

Proof overview:

$E : [N] \times [N] \times [D] \to [M]$ be an ϵ-extractor for somewhere-random source (X, Y)

- consider uniformly-random $S \subseteq [M]$ s.t. $|S| = 10\ \epsilon \cdot M$

Extracting mergers (lower bound)

Proof overview:

$E : [N] \times [N] \times [D] \to [M]$ be an ϵ-extractor for somewhere-random source (X, Y)

- consider uniformly-random $S \subseteq [M]$ s.t. $|S| = 10\ \epsilon \cdot M$
- attempt to pick $g : [N] \to [N]$ such that $\Pr_{Y,J}[E(g(Y), Y, J) \in S] = 0$

Extracting mergers (lower bound)

Proof overview:

$E : [N] \times [N] \times [D] \to [M]$ be an ϵ-extractor for somewhere-random source (X, Y)

- consider uniformly-random $S \subseteq [M]$ s.t. $|S| = 10\,\epsilon \cdot M$
- attempt to pick $g : [N] \to [N]$ such that $\Pr_{Y,J}[E(g(Y), Y, J) \in S] = 0$
  - attempt must fail, otherwise contradiction to extracting merger property!

Extracting mergers (lower bound)

Proof overview:

$E : [N] \times [N] \times [D] \to [M]$ be an ϵ-extractor for somewhere-random source (X, Y)

- consider uniformly-random $S \subseteq [M]$ s.t. $|S| = 10\ \epsilon \cdot M$
- attempt to pick $g : [N] \to [N]$ such that $\Pr_{Y,J}[E(g(Y), Y, J) \in S] = 0$
  - attempt must fail, otherwise contradiction to extracting merger property!
- reveals strange structure, which we dig more into, prove lower bound on D

Extracting mergers (lower bound)

Proof overview:

$E : [N] \times [N] \times [D] \to [M]$ be an ϵ-extractor for somewhere-random source (X, Y)

- consider uniformly-random $S \subseteq [M]$ s.t. $|S| = 10\ \epsilon \cdot M$
- attempt to pick $g : [N] \to [N]$ such that $\Pr_{Y,J}[E(g(Y), Y, J) \in S] = 0$
  - attempt must fail, otherwise contradiction to extracting merger property!
- reveals strange structure, which we dig more into, prove lower bound on D

Proof uses second-moment strengthening of approach in [RTS00]

t-part s-where random sources

min-entropy rate $\frac{s}{t}$ sources with $s > 1$ uniform and independent blocks

t-part s-where random sources

min-entropy rate $\frac{s}{t}$ sources with $s > 1$ uniform and independent blocks

Definition (t-part s-where-random source)

$X_1, \ldots X_t \in (\{0,1\}^n)^t$ s.t. there exists an (unknown) $J, |J| = s$ s.t. $\times_{i \in J} X_i$ is uniform over $(\{0,1\}^n)^s$

Extracting multimergers

Definition (Seeded extracting multimergers)

$E : \{\{0,1\}^n\}^t \times \{0,1\}^d \to \{0,1\}^m$ s.t. for every t-part s-where random source X, distribution of $E(X, U_d)$ is ϵ-close to U_m

Extracting multimergers

Definition (Seeded extracting multimergers)

$E : \{\{0,1\}^n\}^t \times \{0,1\}^d \to \{0,1\}^m$ s.t. for every t-part s-where random source X, distribution of $E(X, U_d)$ is ϵ-close to U_m

Do there exist seedless extracting multimergers?

3-bit Majority multimerger

Theorem (Majority extracts)

Let $E((X_1, \ldots, X_n) \times (Y_1, \ldots, Y_n) \times (Z_1, \ldots, Z_n)) = MAJ(X_1, Y_1, Z_1)$ then, for every 3-part 2-where random source (X, Y, Z), distribution of $E(X, U_d)$ is $\frac{1}{4}$-close to U_1

3-bit Majority multimerger

Theorem (Majority extracts)

Let $E((X_1, \ldots, X_n) \times (Y_1, \ldots, Y_n) \times (Z_1, \ldots, Z_n)) = MAJ(X_1, Y_1, Z_1)$ *then, for every 3-part 2-where random source* (X, Y, Z)*, distribution of* $E(X, U_d)$ *is* $\frac{1}{4}$*-close to* U_1

can we do better without using any seed?

Projections and seedless extraction

Equivalent to an independent geometric question:

Theorem

There exists a seedless ϵ-extractor *for t-part s*-where random sources

if and only if

There exists a partition of $(\{0,1\}^n)^t$ *into A and B such that every s-dimensional projection of A and B has size between* $\frac{1}{2} - \epsilon$ *and* $\frac{1}{2} + \epsilon$

Projections of partitions

Question: what is the smallest error of a (t, s)-seedless extracting multi-merger with 1-bit output?

Projections of partitions

Question: what is the smallest error of a (t, s)-seedless extracting multi-merger with 1-bit output?

Theorem

For any partition of $[N]^t$ into c parts, there exists a s-dimensional projection of size at least $?? \cdot N^s$

Projections of partitions

Question: what is the smallest error of a (t, s)-seedless extracting multi-merger with 1-bit output?

Theorem

For any partition of $[N]^t$ into c parts, there exists a s-dimensional projection of size at least $?? \cdot N^s$

natural partition analogues of Shearer's Lemma

Projections of partitions

Question: what is the smallest error of a (t, s)-seedless extracting multi-merger with 1-bit output?

Theorem

For any partition of $[N]^t$ into c parts, there exists a s-dimensional projection of size at least $?? \cdot N^s$

natural partition analogues of Shearer's Lemma

focus here on $c = 2$ case

Lower bound on projection size $(3 \to 2)$

we seek:

Theorem

For any bipartition of $[N]^3$, there exists a 2-dimensional projection of size $\geq$??

Lower bound on projection size $(3 \to 2)$

we seek:

Theorem

For any bipartition of $[N]^3$, there exists a 2-dimensional projection of size $\geq ??$

attempt:

- there exists a part with size $\geq \frac{N^3}{2}$

Lower bound on projection size $(3 \to 2)$

we seek:

Theorem

For any bipartition of $[N]^3$, there exists a 2-dimensional projection of size $\geq ??$

attempt:

- there exists a part with size $\geq \frac{N^3}{2}$
- by Shearer's Lemma, this part has a projection of size $\geq (\frac{N^3}{2})^{\frac{2}{3}} \geq 0.629.. \cdot N^2$

Lower bound on projection size $(3 \to 2)$

we seek:

Theorem

For any bipartition of $[N]^3$, there exists a 2-dimensional projection of size $\geq$??

attempt:

- there exists a part with size $\geq \frac{N^3}{2}$
- by Shearer's Lemma, this part has a projection of size $\geq (\frac{N^3}{2})^{\frac{2}{3}} \geq 0.629.. \cdot N^2$

We show there exists a projection of size $\geq \frac{3}{4} N^2 \Rightarrow$

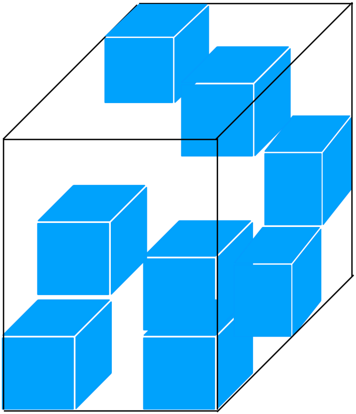
majority multimerger is optimal!

Lower bound on projection size $(3 \to 2)$

Proof overview:

1. fix arbitrary partition A, B (a.k.a "colouring")

Lower bound on projection size $(3 \to 2)$



Lower bound on projection size $(3 \to 2)$

Proof overview:

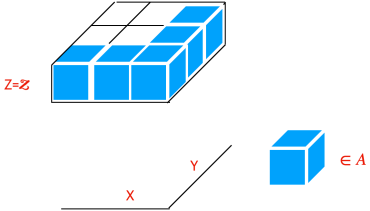
1. fix arbitrary partition A, B (a.k.a "colouring")
2. assume for sake of contradiction "top projections" are small ($\Pi_Z(A), \Pi_Z(B) < \frac{3}{4}$)

Lower bound on projection size $(3 \to 2)$

Proof overview:

1. fix arbitrary partition A, B (a.k.a "colouring")
2. assume for sake of contradiction "top projections" are small ($\Pi_Z(A), \Pi_Z(B) < \frac{3}{4}$)
3. fix a slice, study the sum of projections of A and B on a slice;

Lower bound on projection size



Lower bound on projection size

Proof overview:

1. fix arbitrary partition A, B(a.k.a "colouring")
2. assume for sake of contradiction "top projections" are small ($\Pi_Z(A), \Pi_Z(B) < \frac{3}{4}$)
3. fix a slice, study the sum of projections of A and B on a slice; (miraculously) this sum has lower bound of $1 - \sqrt{1 - 3/4} = \frac{1}{2}$

Lower bound on projection size

Proof overview:

1. fix arbitrary partition A, B(a.k.a "colouring")
2. assume for sake of contradiction "top projections" are small ($\Pi_Z(A), \Pi_Z(B) < \frac{3}{4}$)
3. fix a slice, study the sum of projections of A and B on a slice; (miraculously) this sum has lower bound of $1 - \sqrt{1 - 3/4} = \frac{1}{2}$
4. averaging over slices give a large side projection

Lower bound on projection size

Proof overview:

1. fix arbitrary partition A, B(a.k.a "colouring")
2. assume for sake of contradiction "top projections" are small ($\Pi_Z(A), \Pi_Z(B) < \frac{3}{4}$)
3. fix a slice, study the sum of projections of A and B on a slice; (miraculously) this sum has lower bound of $1 - \sqrt{1 - 3/4} = \frac{1}{2}$
4. averaging over slices give a large side projection

3-bit majority multimerger is optimal!!

Projections of partitions

- We also prove lower bounds for partitioning $[N]^2$ into 3 parts
- Our proofs work even for open covers of the solid cube $[0, 1]^3$, rectangle $[0, 1]^2$
- Our bounds beat those obtained by Shearer's Lemma/Loomis Whitney inequality

Summary

- We study mergers in the extracting regime and our extractors can extract $\mathrm{poly}(\frac{1}{\epsilon})$ using constant seed
- Our lower bound shows that standard extractors overshadow extracting mergers when tasked with extracting $\Omega(n)$ bits from source
- We prove lower bounds on sizes of low-dimensional projections of partitions and our bounds beat those obtained by Shearer's Lemma/Loomis Whitney

Thank you!