بسم الله الرحمن الرحیم

جلسه هفتم

خلاصه‌سازی برای مه‌داده

خلاصه‌سازی خطی

شرایط مسئله

- آرایه با عملیات جمع/منها:
  - نگه‌دارنده بردار $x \in R^n$
  - تغییرات: (i, Δ) —> معنی: $xi = xi + \Delta$
  - پرسش: i و پاسخ: xi

شرایط مسئله

- آرایه با عملیات جمع/منها:
  - نگه‌دارنده بردار $x \in R^n$
  - تغییرات: (i, Δ) —> معنی: $xi = xi + \Delta$
  - پرسش: i و پاسخ: xi

- سه حالت مسئله
  - فقط اضافه: $\Delta = ۱$
  - شمارنده مثبت: $\Delta = \pm۱$ ولی $x_i \geq 0$
  - شمارنده آزاد: $\Delta = \pm۱$

آرایه و پرسش با خطای L1

- پرسش نقطه l1 (l1 point query):
  - پرسش (k, ℓ_1)ـ نقطه‌ای
    - ورودی i
    - جواب: $x_i \pm \|x\|_1/k$

کمتر از k <---

آرایه و پرسش با خطای L1

- پرسش نقطه l1 (l1 point query):
  - پرسش (k, ℓ_1)_ نقطه‌ای
    - ورودی i
    - جواب: $x_i \pm \|x\|_1/k$

- پرسش وزین‌ها (Heavy Hitters):
  - پرسش (k, ℓ_1)_ وزین‌ها
  - پاسخ: $L \subseteq [n]$
    - ۱) |L| = O(k)
    - ۲) $|x_i| > \|x\|_1/k \Longrightarrow i \in L$ <— کمتر از k

نقطه‌ای ‎—‎> وزین‌ها

Lemma 4.1.2. *Suppose there is an algorithm* $\mathcal{A}$ *solving* $(3k, \ell_1)$*-point query with failure probability at most* δ/n *and using* S *words of memory. Then there is an algorithm* $\mathcal{A}'$ *solving* (k, ℓ_1)*-heavy hitters with failure probability at most* δ *and using space at most* $S + 3k$*.*

نقطه‌ای —> وزین‌ها

Lemma 4.1.2. *Suppose there is an algorithm $\mathcal{A}$ solving $(3k, \ell_1)$-point query with failure probability at most δ/n and using S words of memory. Then there is an algorithm $\mathcal{A}'$ solving (k, ℓ_1)-heavy hitters with failure probability at most δ and using space at most $S + 3k$.*

الگوریتم A

الگوریتم A’

- همه را می‌پرسیم
  - با احتمال $1-\boldsymbol{\delta}$ همه درستند
- ۳k بزرگ‌ترین‌ها را در T نگه می‌داریم

نقطه‌ای —> وزین‌ها

Lemma 4.1.2. *Suppose there is an algorithm $\mathcal{A}$ solving $(3k, \ell_1)$-point query with failure probability at most δ/n and using S words of memory. Then there is an algorithm $\mathcal{A}'$ solving (k, ℓ_1)-heavy hitters with failure probability at most δ and using space at most $S + 3k$.*

الگوریتم A’

- همه را می‌پرسیم
  - با احتمال $1-\boldsymbol{\delta}$ همه درستند
- ۳k بزرگ‌ترین‌ها را در T نگه می‌داریم

الگوریتم A

باقی اثبات: A کار می‌کند —> بزرگتر از $||x||_1/k$ در T هستند

$(2/3)\|x\|_1/k$

نقطه‌ای ‌—‌‌‌> وزن‌ها

Lemma 4.1.2. *Suppose there is an algorithm $\mathcal{A}$ solving $(3k, \ell_1)$-point query with failure probability at most δ/n and using S words of memory. Then there is an algorithm $\mathcal{A}'$ solving (k, ℓ_1)-heavy hitters with failure probability at most δ and using space at most $S + 3k$.*

الگوریتم 'A

- همه را می‌پرسیم

الگوریتم
A

- با احتمال $1-\delta$ همه درستند
- ۳k بزرگ‌ترین‌ها را در T نگه می‌داریم

باقی اثبات: A کار می‌کند —> بزرگتر از $||x||_1/k$ در T هستند

$(2/3)\|x\|_1/k$

- اندیس i که $|x_i| < \|x\|_1/(3k)$
  پاسخ A به آن‌ها < $(2/3)\|x\|_1/k$

نقطه‌ای —> وزین‌ها

Lemma 4.1.2. *Suppose there is an algorithm $\mathcal{A}$ solving $(3k, \ell_1)$-point query with failure probability at most δ/n and using S words of memory. Then there is an algorithm $\mathcal{A}'$ solving (k, ℓ_1)-heavy hitters with failure probability at most δ and using space at most $S + 3k$.*

الگوریتم A

الگوریتم 'A

- همه را می‌پرسیم
    - با احتمال $1-\delta$ همه درستند
- ۳k بزرگ‌ترین‌ها را در T نگه می‌داریم

باقی اثبات: A کار می‌کند —> بزرگتر از $||x||_1/k$ در T هستند

$(2/3)\|x\|_1/k$

- اندیس i خوب، پاسخ A بزرگ‌تر از
  $(2/3)\|x\|_1/k = \|x\|_1/k - \|x\|_1/(3k)$
- اندیس i که $|x_i| < \|x\|_1/(3k)$
  پاسخ A به آن‌ها $< (2/3)\|x\|_1/k$

نقطه‌ای ——> وزین‌ها

Lemma 4.1.2. *Suppose there is an algorithm $\mathcal{A}$ solving $(3k, \ell_1)$-point query with failure probability at most δ/n and using S words of memory. Then there is an algorithm $\mathcal{A}'$ solving (k, ℓ_1)-heavy hitters with failure probability at most δ and using space at most $S + 3k$.*

الگوریتم 'A

- همه را می‌پرسیم

الگوریتم
A

- با احتمال $1-\delta$ همه درستند
- ۳k بزرگ‌ترین‌ها را در T نگه می‌داریم

باقی اثبات: A کار می‌کند ——> بزرگتر از $||x||_1/k$ در T هستند

$(2/3)\|x\|_1/k$

- اندیس i خوب، پاسخ A بزرگ‌تر از

$(2/3)\|x\|_1/k = \|x\|_1/k - \|x\|_1/(3k)$

- تعداد اندیس i که $|x_i| \geq \|x\|_1/(3k)$

کمتر مساوی 3k تا

- اندیس i که $|x_i| < \|x\|_1/(3k)$

پاسخ A به آن‌ها $< (2/3)\|x\|_1/k$

کران بالا برای خطای N پرسش

- اتفاق Ai: i_امین پرسش غلط باشد
- اینکه حداقل یکی از Aiها غلط باشد = اجتماع Aiها

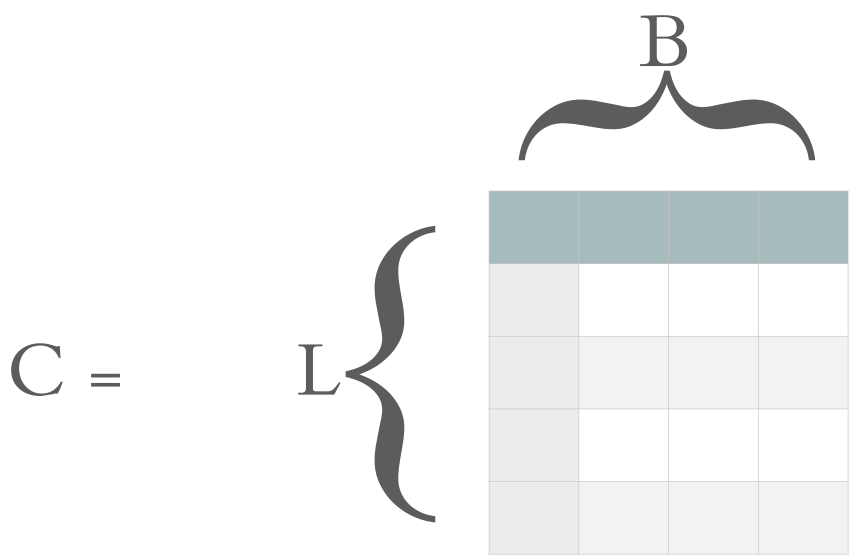
احتمال خطا = $P[\cup A_i] \leq \sum P[A_i]$

- برای مسئله قبل: $P[Ai] < \delta/n$
  - ==> احتمال اینکه حداقل یکی از خروجی‌ها غلط باشد $\leq \delta$

خلاصه‌سازی COUNTMIN

برای مسئله پرسش نقطه‌ای l۱
و $x_i \geq 0$

- توابع درهم‌سازی ۲ ـ طرفه $h_1, \ldots h_L : [n] \rightarrow [\mathrm{B}]$

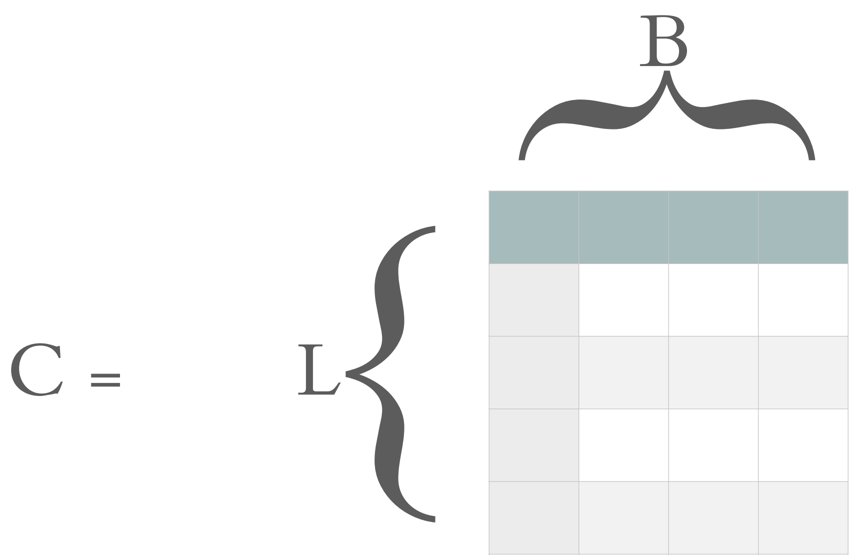


- $B = 2k,\ L = \lceil \log_2(1/\delta) \rceil$.

خلاصه‌سازی COUNTMIN

برای مسئله پرسش نقطه‌ای l_1
و $x_i \geq 0$

- توابع درهم‌سازی ۲ـ طرفه $h_1, \ldots h_L : [n] \to [\mathrm{B}]$



- $B = 2k,\ L = \lceil \log_2(1/\delta) \rceil$.

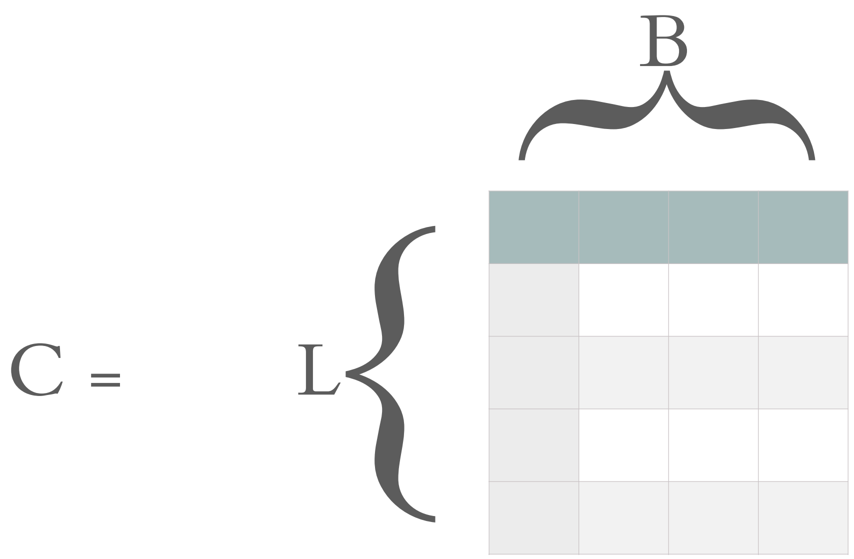
- تغییر (i, Δ):

$$C_{a,h_a(i)} \mathrel{+}= \Delta \qquad \text{for } a = 1, \ldots, L.$$

خلاصه‌سازی COUNTMIN

برای مسئله پرسش نقطه‌ای l1
و $x_i \geq 0$

- توابع درهم‌سازی ۲ـ طرفه $h_1, \ldots h_L : [n] \to [B]$



- $B = 2k,\ L = \lceil \log_2(1/\delta) \rceil$.

- تغییر (i, Δ):

$$C_{a,h_a(i)} \mathrel{+}= \Delta \qquad \text{for } a = 1, \ldots, L.$$

- پاسخ به پرسش i:

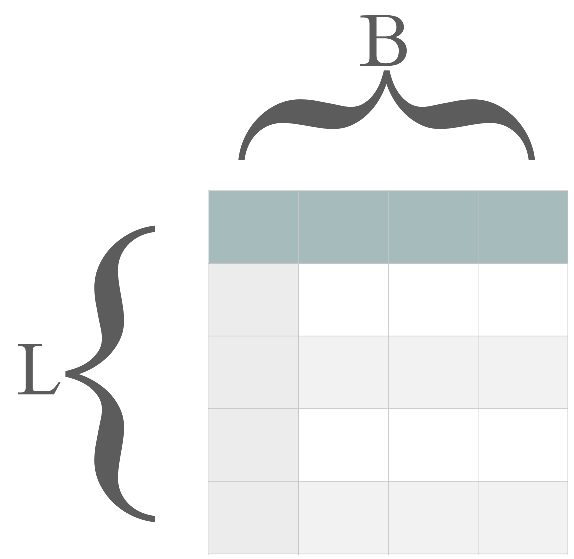
$$\min_{1 \leq a \leq L} C_{a,h_a(i)}$$

خلاصه‌سازی COUNTMIN

برای مسئله پرسش نقطه‌ای l1 و $x_i \geq 0$

- توابع درهم‌سازی ۲ـ طرفه $h_1, \dots h_L : [n] \rightarrow [\mathrm{B}]$

C =



- $B = 2k,\ L = \lceil \log_2(1/\delta) \rceil$

- تغییر (i, Δ):

$$C_{a,h_a(i)} \mathrel{+}= \Delta \qquad \text{for } a = 1, \dots, L.$$

- پاسخ به پرسش i:

$$\min_{1 \leq a \leq L} C_{a,h_a(i)}$$

حافظه: O(LB) تا عدد

تحلیل الگوریتم COUNTMIN

Lemma 4.1.3. CountMin.query(i) *returns* $x_i \pm \|x\|_1/k$ *w.p.* $\geq 1 - \delta$.

تحلیل الگوریتم COUNTMIN

Lemma 4.1.3. CountMin.query(i) *returns* $x_i \pm \|x\|_1/k$ *w.p.* $\geq 1 - \delta$.

اتفاق بد در سطر r: xjهای دیگر روی i بیافتند $h_r(j) = h_r(i)$

Zj: برای r و i ثابت، $h_r(j) = h_r(i)$

تحلیل الگوریتم COUNTMIN

Lemma 4.1.3. CountMin.query(i) *returns* $x_i \pm \|x\|_1/k$ *w.p.* $\geq 1-\delta$.

اتفاق بد در سطر r: xjهای دیگر روی i بیافتند $h_r(j) = h_r(i)$

Zj: برای r و i ثابت، $h_r(j) = h_r(i)$

$$C_{r,h_r(i)} = x_i + \underbrace{\sum_{j\neq i} x_j Z_j}_{E :=}$$

تحلیل الگوریتم COUNTMIN

Lemma 4.1.3. CountMin.query(i) *returns* $x_i \pm \|x\|_1/k$ *w.p.* $\geq 1-\delta$.

اتفاق بد در سطر r: xjهای دیگر روی i بیافتند $h_r(j) = h_r(i)$

Zj: برای r و i ثابت، $h_r(j) = h_r(i)$

$$C_{r,h_r(i)} = x_i + \underbrace{\sum_{j\neq i} x_j Z_j}_{E :=}$$

پس

$$\mathbb{E}E = \sum_{j\neq i} |x_j|/B \leq \|x\|_1/(2k)$$

تحلیل الگوریتم COUNTMIN

Lemma 4.1.3. CountMin.query(i) *returns* $x_i \pm \|x\|_1/k$ *w.p.* $\geq 1-\delta$.

اتفاق بد در سطر r: xjهای دیگر روی i بیافتند $h_r(j) = h_r(i)$

Zj: برای r و i ثابت، $h_r(j) = h_r(i)$

$$C_{r,h_r(i)} = x_i + \underbrace{\sum_{j\neq i} x_j Z_j}_{E :=}$$

پس

EE = $\sum_{j\neq i} |x_j|/B \leq \|x\|_1/(2k)$

پس

$\mathbb{P}(E > \|x\|_1/k) < 1/2$.

تحلیل الگوریتم COUNTMIN

Lemma 4.1.3. CountMin.query(i) *returns* $x_i \pm \|x\|_1/k$ *w.p.* $\geq 1-\delta$.

اتفاق بد در سطر r: xjهای دیگر روی i بیافتند $h_r(j) = h_r(i)$

Zj: برای r و i ثابت، $h_r(j) = h_r(i)$

$$C_{r,h_r(i)} = x_i + \underbrace{\textstyle\sum_{j\neq i} x_j Z_j}_{E :=}$$

پس

$$\mathbb{E}E = \textstyle\sum_{j\neq i} |x_j|/B \leq \|x\|_1/(2k)$$

پس

$$\mathbb{P}(E > \|x\|_1/k) < 1/2.$$

پس

$$\mathbb{P}(\min_r C_{r,h_r(i)} > x_i + \|x\|_1/k) < 1/2^L \leq \delta$$

Lemma 4.1.3. CountMin.query(i) *returns* $x_i \pm \|x\|_1/k$ *w.p.* $\geq 1-\delta$.

پس

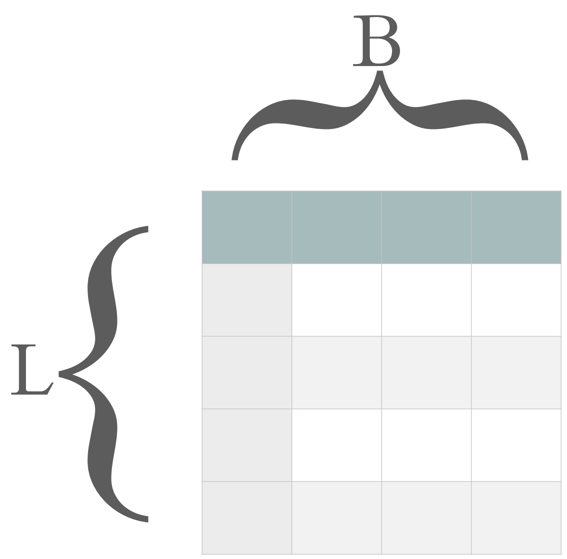
Theorem 4.1.4. *There is an algorithm solving the* ℓ_1 k*-heavy hitter problem in the strict turnstile model with failure probability* δ*, space* $O(k\log(n/\delta))$*, update time* $O(\log(n/\delta))$*, and query time* $O(n\log(n/\delta))$.

تا عدد

همه را می‌پرسیم

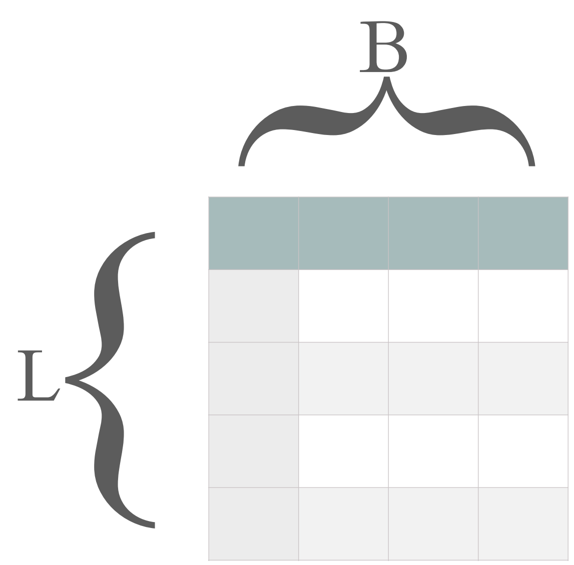
خلاصه‌سازی خطی

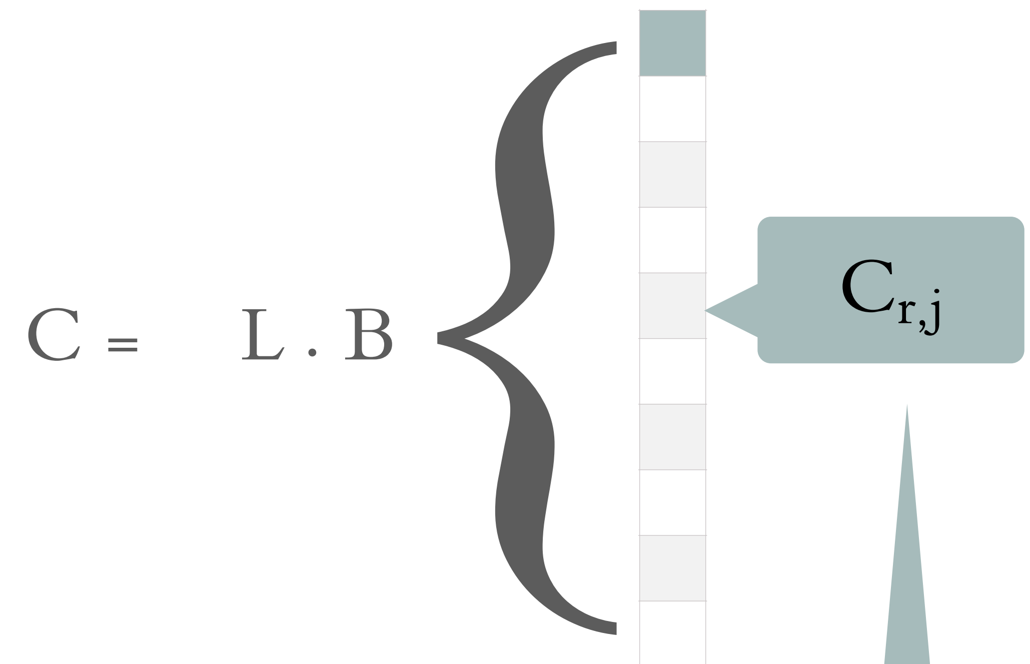
C =



خلاصه‌سازی خطی

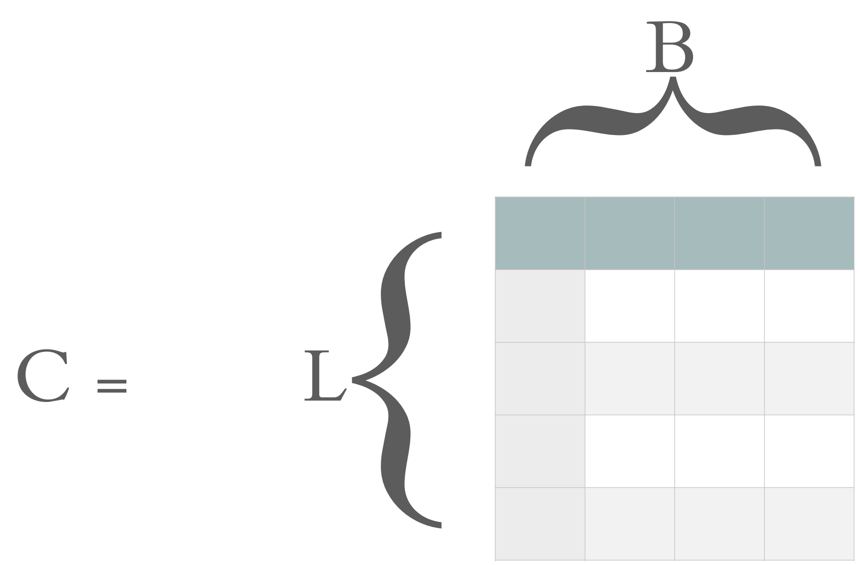
$C =$

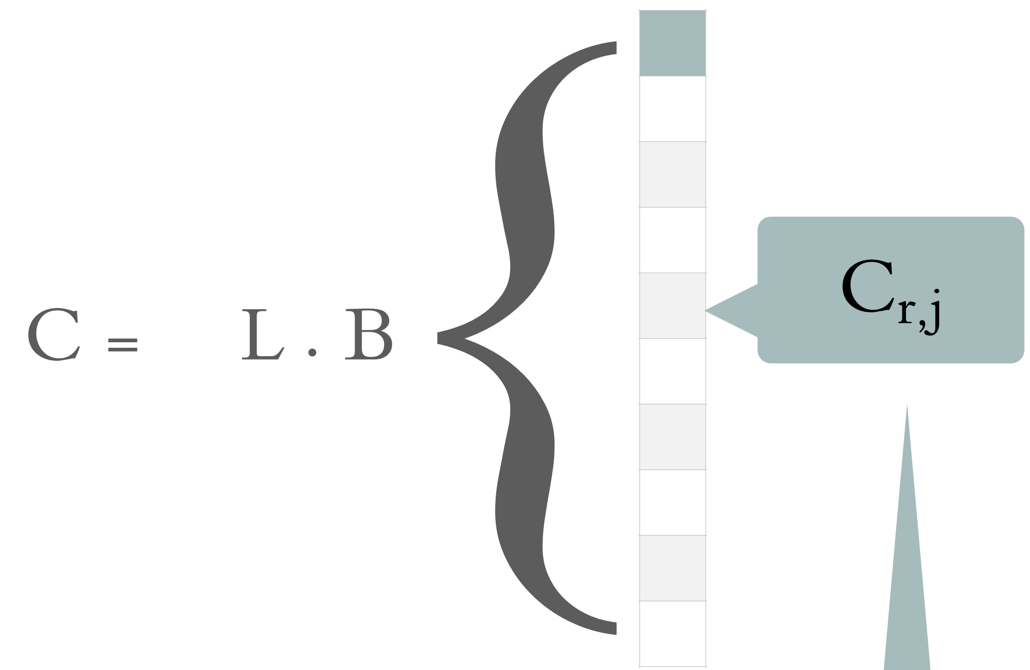




جمع x_i اون‌هایی که در سطر r به j درهم می‌شوند

خلاصه‌سازی خطی





جمع x_i اون‌هایی که در سطر r به j درهم می‌شوند

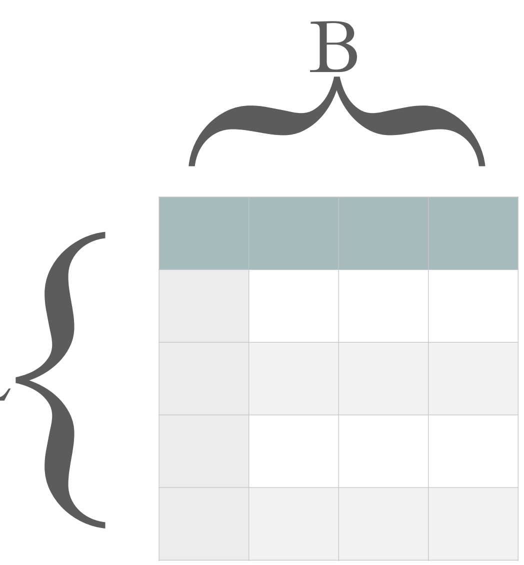
به‌روزرسانی

$$C = \Pi x \xrightarrow{\boxed{(i, \Delta)}} C = \Pi (x + \Delta e_i)$$

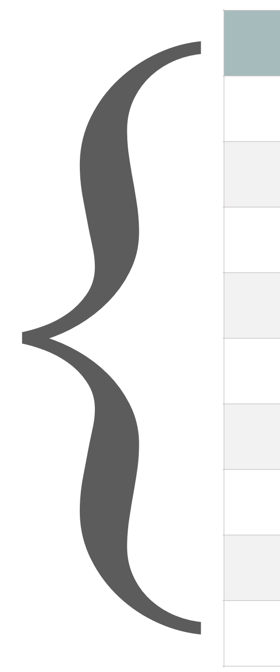
$$= \Pi x + \Delta \Pi e_i$$

خلاصه‌سازی خطی

$C =$ L



$C = L \cdot B$



$C_{r,j}$

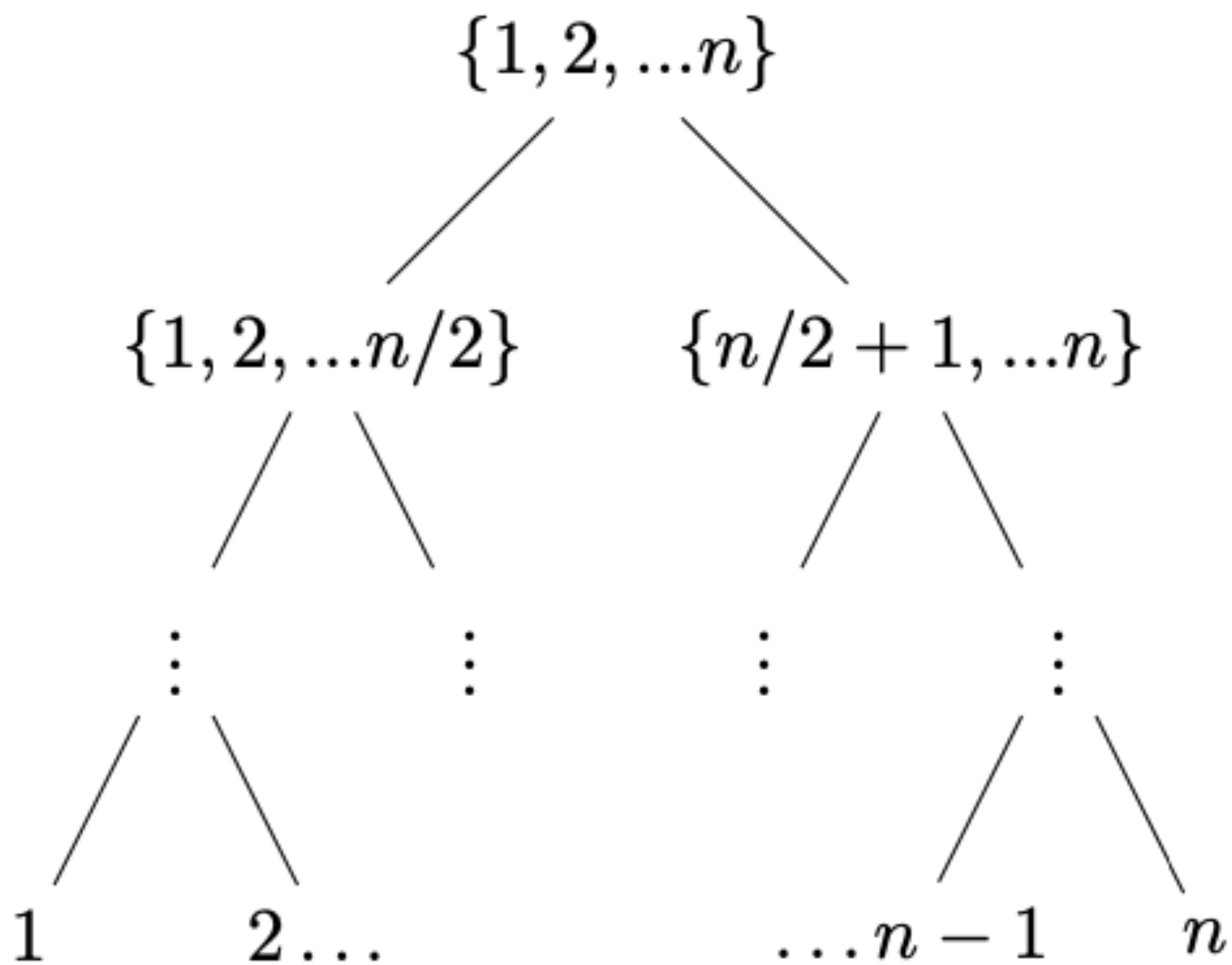
جمع x_i اون‌هایی که در سطر r به j درهم می‌شوند

به‌روزرسانی

$$C = \Pi x \xrightarrow{(i, \Delta)} C = \Pi (x + \Delta e_i)$$

$$= \Pi x + \Delta \Pi e_i$$

تسریع K_وزین‌ها



CM_0 آرایه ۱ تایی

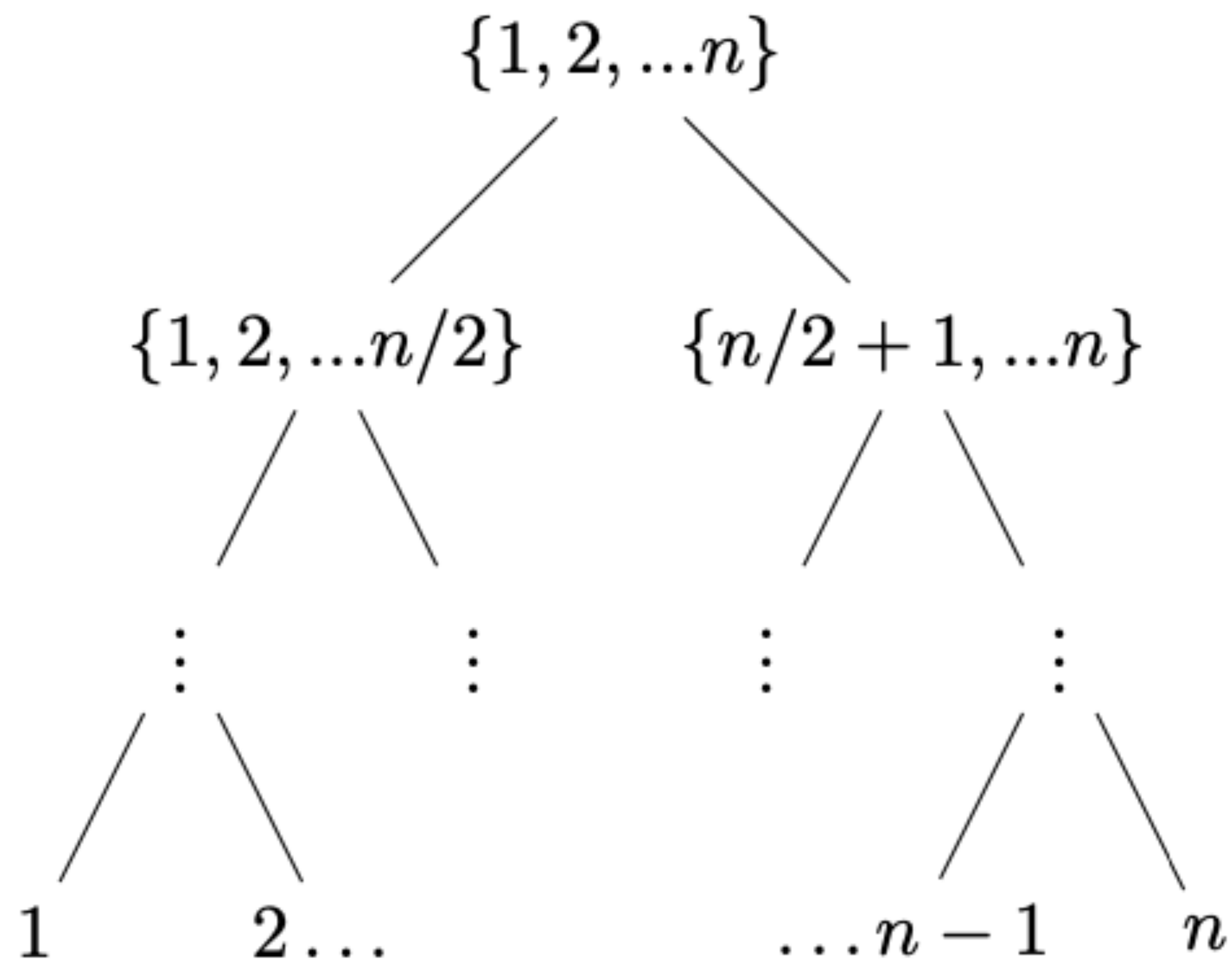
CM_1 آرایه ۲ تایی

CM_lg n آرایه n تایی

۴k_وزین‌ها

احتمال خطا: $\eta = \delta/(4k \log n)$

تسریع K_وزین‌ها



CM_0 آرایه ۱ تایی

CM_1 آرایه ۲ تایی

CM_lg n آرایه n تایی

۴k_وزین‌ها

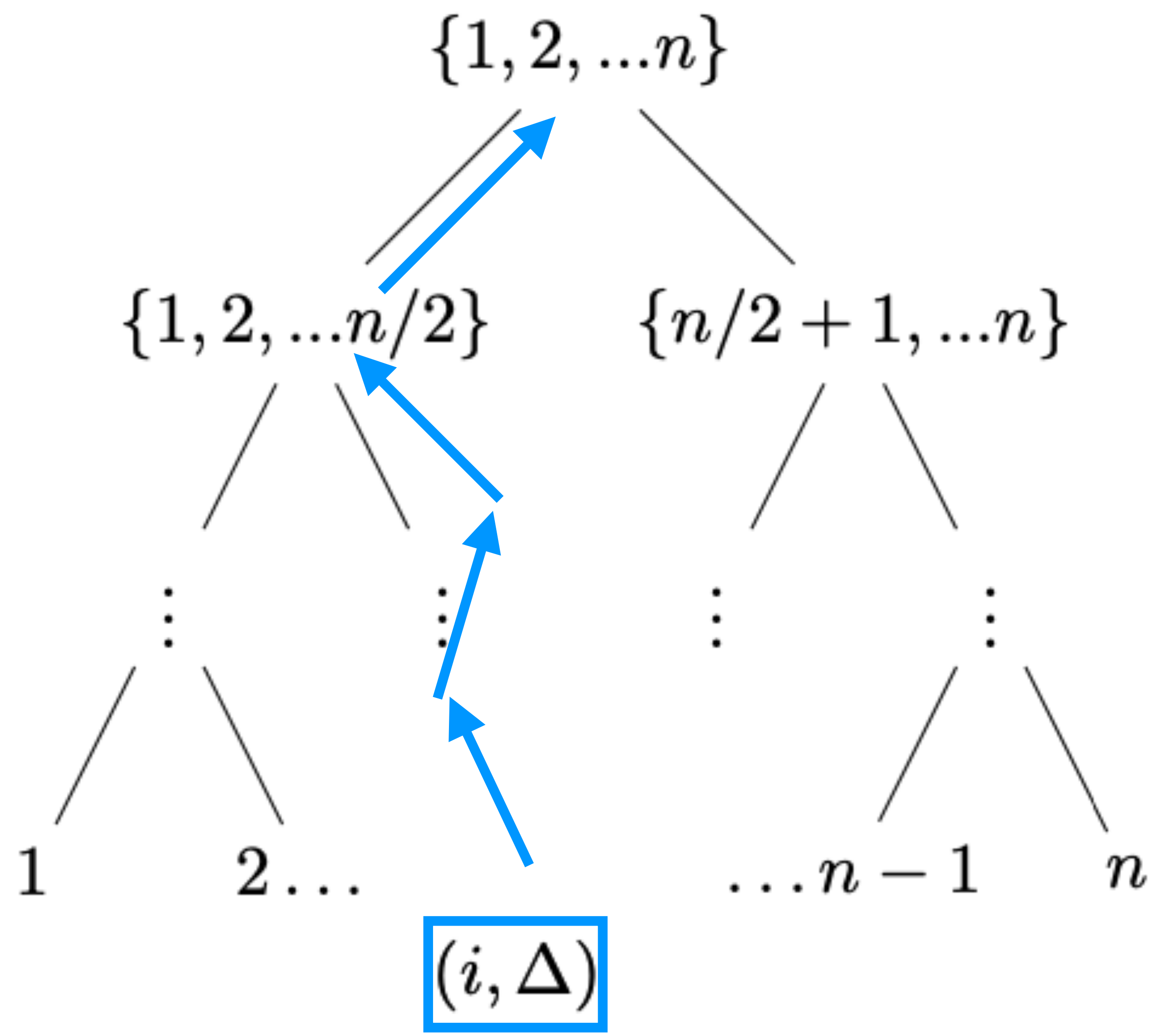
احتمال خطا: $\eta = \delta/(4k \log n)$

حافظه:

$O(k \log(1/\eta) \log n)$

$O(k \log((k \log n)/\delta) \log n)$

تسریع K_وزین‌ها



CM_0 آرایه ۱ تایی

CM_1 آرایه ۲ تایی

CM_lg n آرایه n تایی

۴k_وزین‌ها

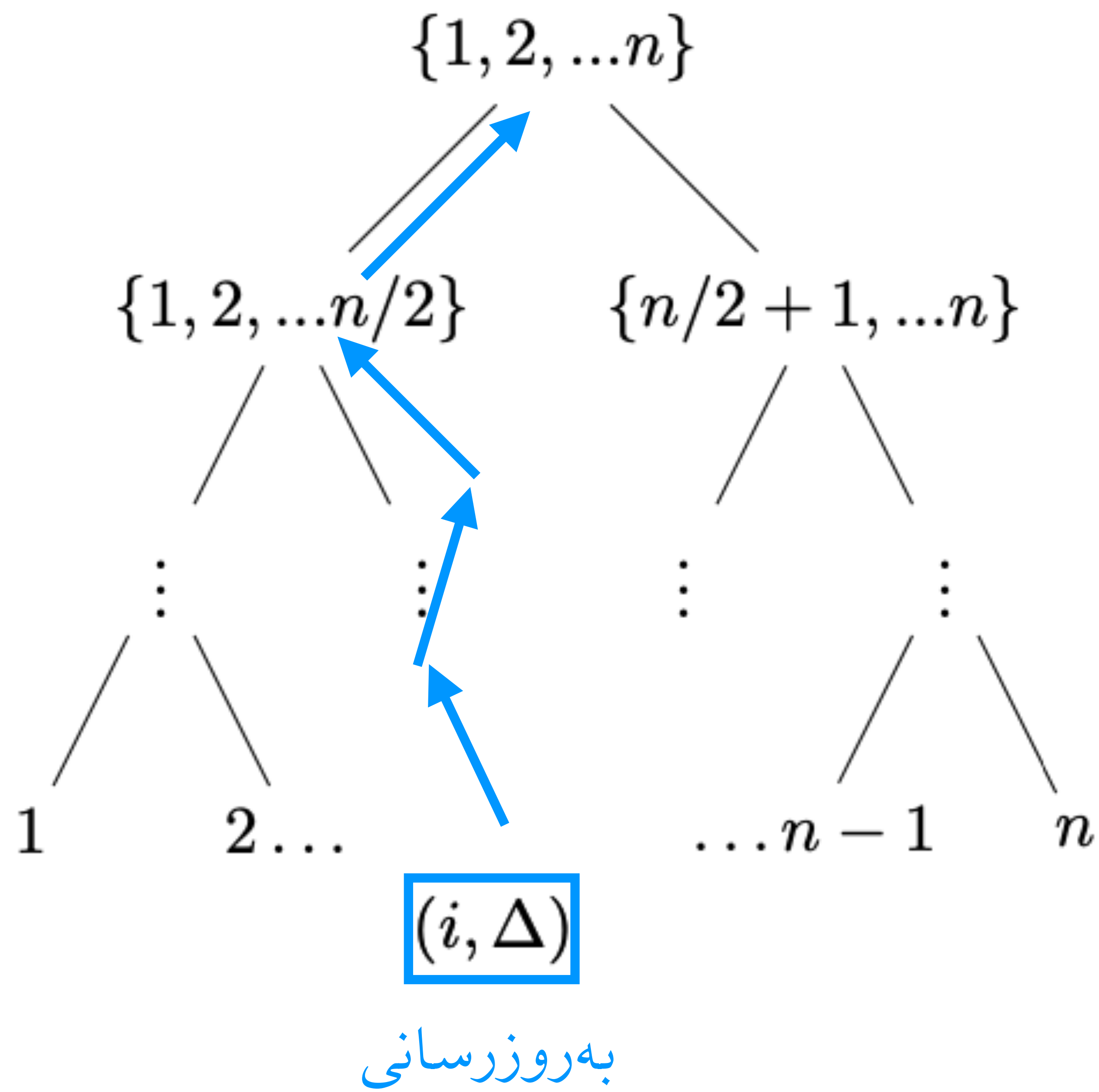
احتمال خطا: $\eta = \delta/(4k \log n)$

حافظه:

$O(k \log(1/\eta) \log n)$

$O(k \log((k \log n)/\delta) \log n)$

تسریع K_وزین‌ها



CM_0 آرایه ۱ تایی

CM_1 آرایه ۲ تایی

CM_lg n آرایه n تایی

۴k_وزین‌ها

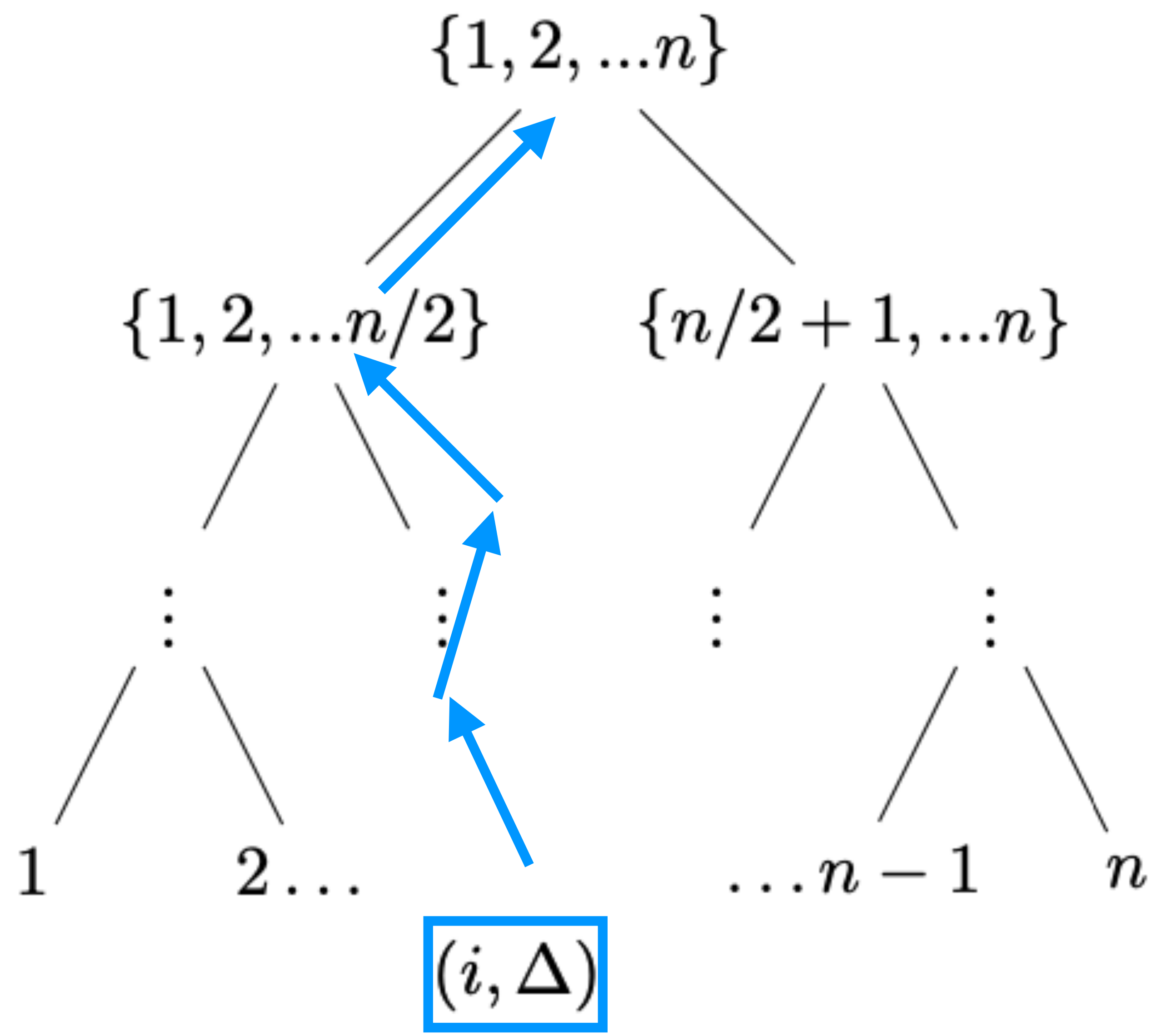
احتمال خطا: $\eta = \delta/(4k \log n)$

حافظه:

$O(k \log(1/\eta) \log n)$

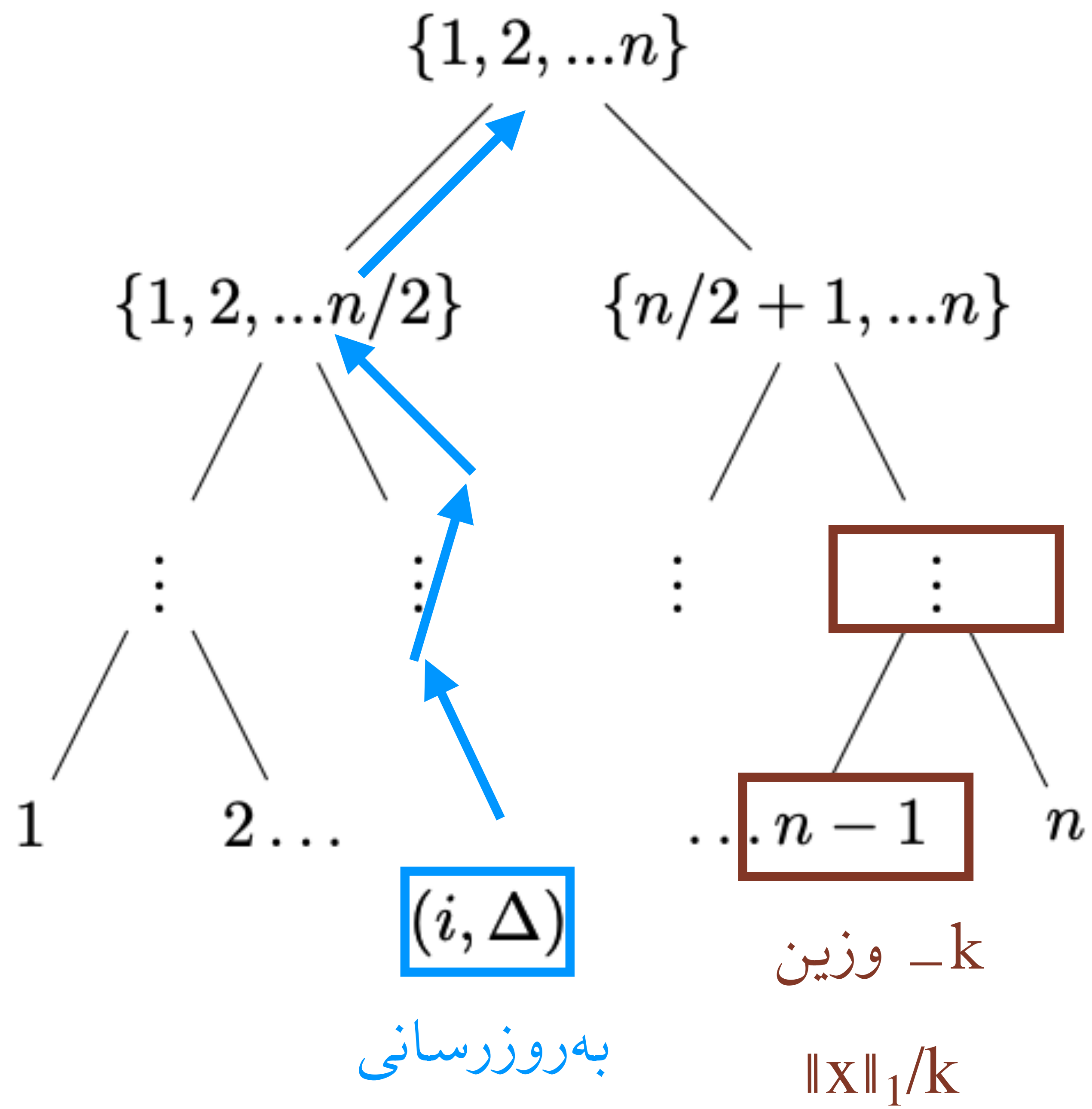
$O(k \log((k \log n)/\delta) \log n)$

پرسش + تحلیل تسریع شده

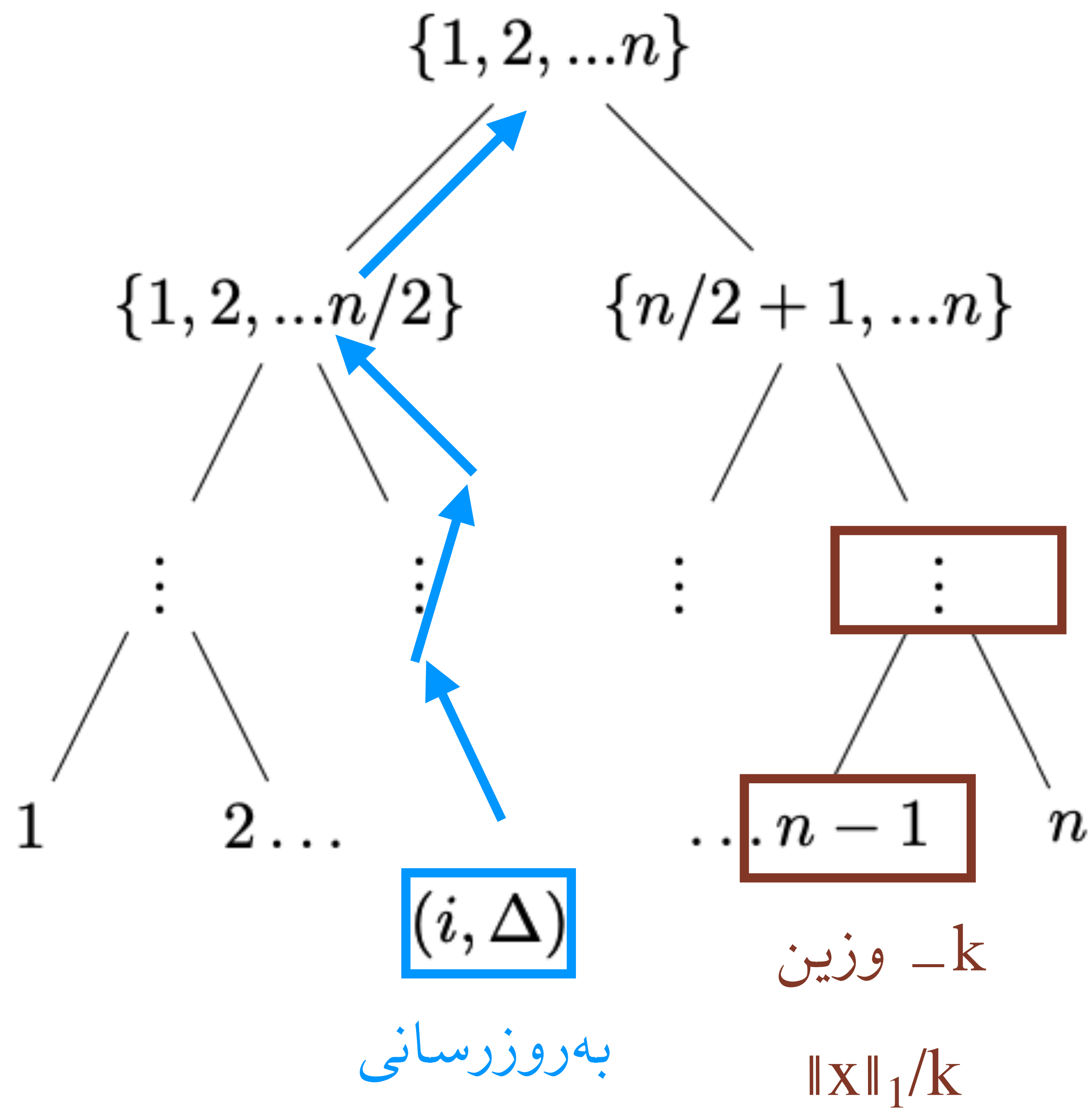


به‌روزرسانی

پرسش + تحلیل تسریع شده



پرسش + تحلیل تسریع شده



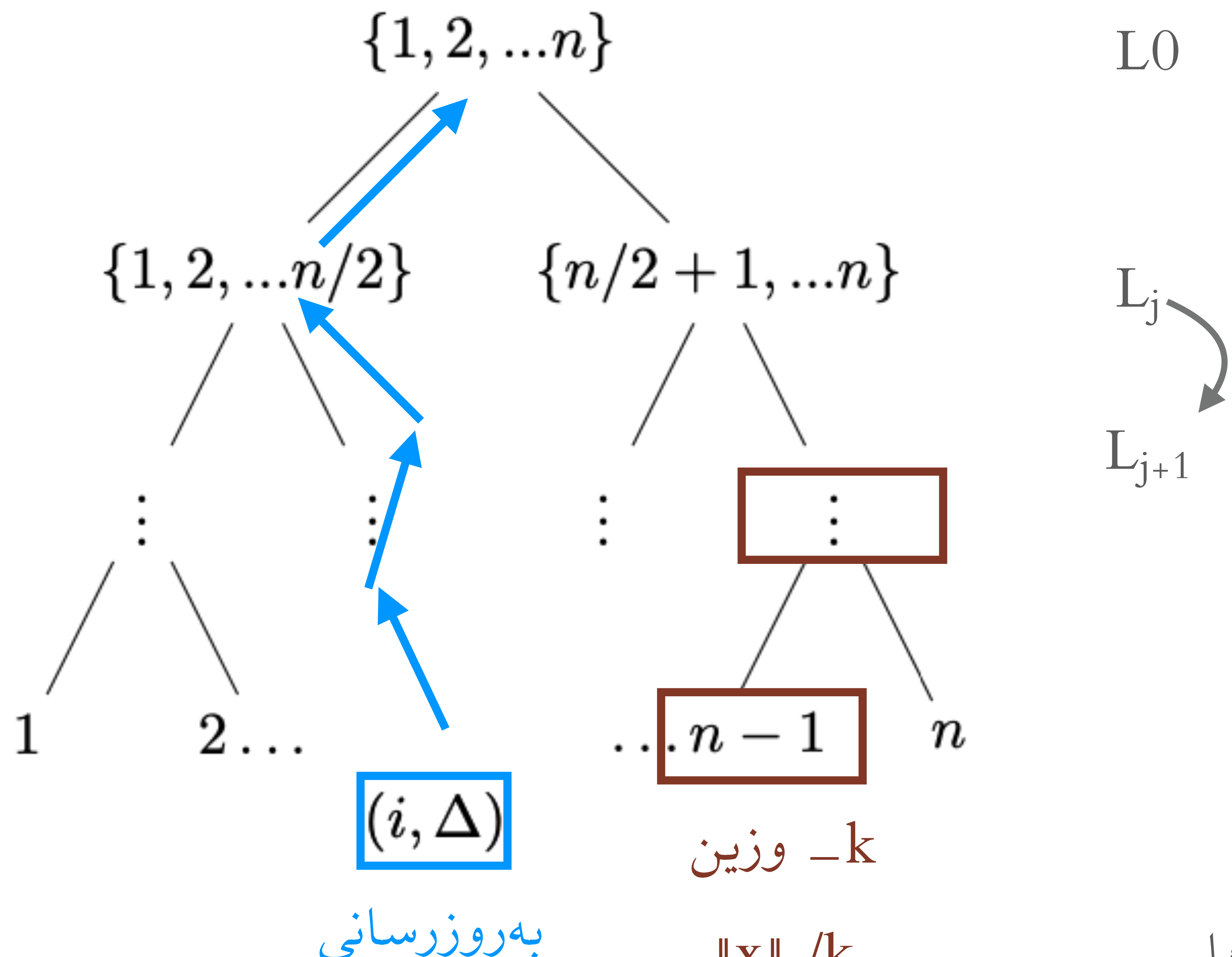
L0

L_j

L_{j+1}

فرزندان اعضای L_j:
اگر بزرگترمساوی
$3/4||x||_1/k$

پرسش + تحلیل تسریع شده



فرزندان اعضای L_j:
اگر بزرگترمساوی
$3/4||x||_1/k$

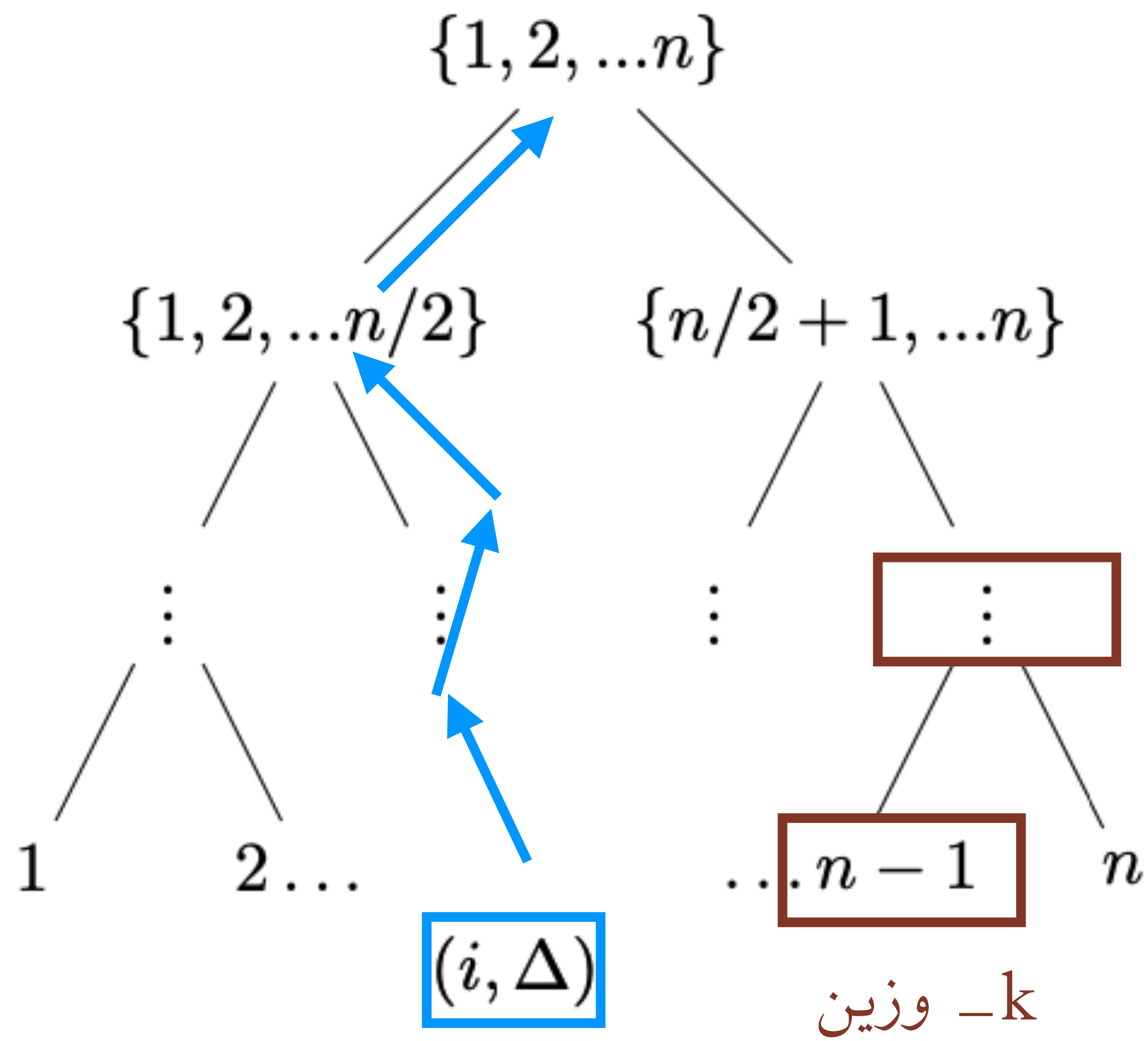
تحلیل

خوب‌ها ($||x||_1/k$): در Ljها

بدهایی که خیلی کوچک ($\frac{2}{4}||x||_1/k$): نیستند

در Lj: کمترمساوی 2k تا

پرسش + تحلیل تسریع شده



L0

L_j

L_{j+1}

فرزندان اعضای L_j:
اگر بزرگترمساوی
$3/4||x||_1/k$

تحلیل

خوب‌ها ($||x||_1/k$): در Ljها

بدهایی که خیلی کوچک ($\frac{2}{4}||x||_1/k$): نیستند

در Lj: کمترمساوی 2k تا

احتمال همه پرسش‌ها درست:
حداقل $1-(4k \log n) * \delta/(4k \log n)$

تعداد پرسش

احتمال خطا

آرایه و پرسش با خطای L۲

- پرسش نقطه l۲ (l2 point query):
  - پرسش (k, ℓ_2)ـ نقطه‌ای
  - سوال: i
  - جواب: $x_i \pm \|x\|_2/\sqrt{k}$

آرایه و پرسش با خطای L۲

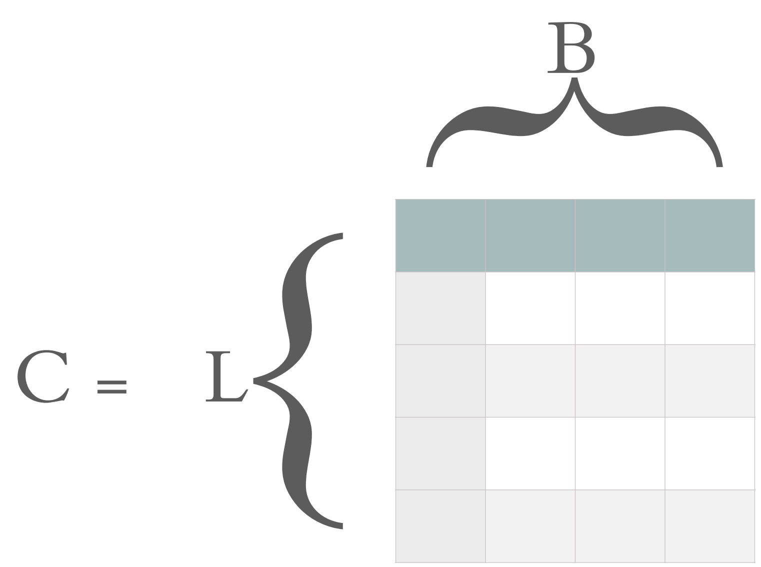
- پرسش نقطه l۲ (l2 point query):
  - پرسش (k, ℓ_2)_ نقطه‌ای
  - سوال: i
  - جواب: $x_i \pm \|x\|_2/\sqrt{k}$

- پرسش وزین‌ها:
  - پرسش (k, ℓ_2)_ وزین‌ها
  - پاسخ: $L \subseteq [n]$
    - ۱) $|L| = O(k)$
    - ۲) $|x_i| > \|x\|_2/\sqrt{k} \implies i \in L$

الگوریتم COUNTSKETCH

برای مسئله پرسش نقطه‌ای l۲ و x_i آزاد

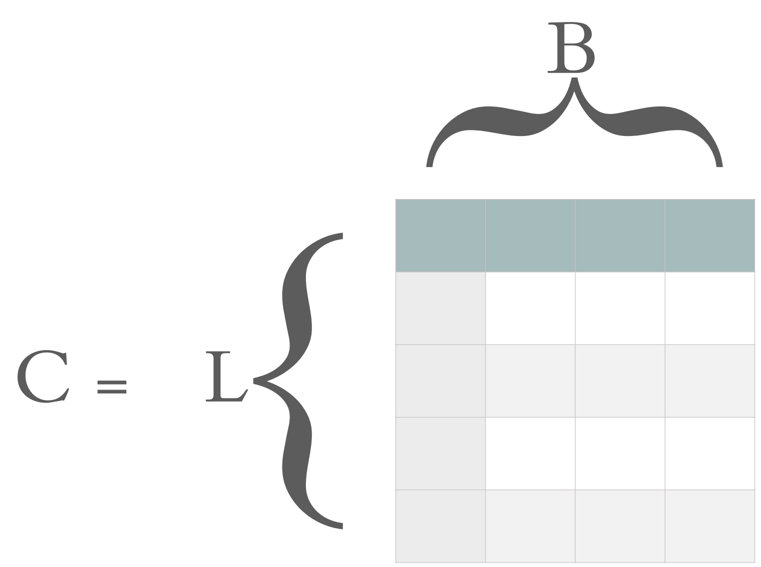
- توابع درهم‌سازی ۲ـ طرفه $h_1, \dots h_L : [n] \to [B]$
- توابع درهم‌سازی ۲ـ طرفه $\sigma_1, \sigma_2, \dots, \sigma_L : [n] \to \{-1, +1\}$
- $B = 9k \quad L = C \log(1/\delta)$



الگوریتم COUNTSKETCH

برای مسئله پرسش نقطه‌ای l۲
و x_i آزاد

- توابع درهم‌سازی ۲ـطرفه $h_1, \dots h_L : [n] \to [\mathrm{B}]$
  - توابع درهم‌سازی ۲ـطرفه $\sigma_1, \sigma_2, \dots, \sigma_L : [n] \to \{-1, +1\}$
- $B = 9k \quad L = C\log(1/\delta)$



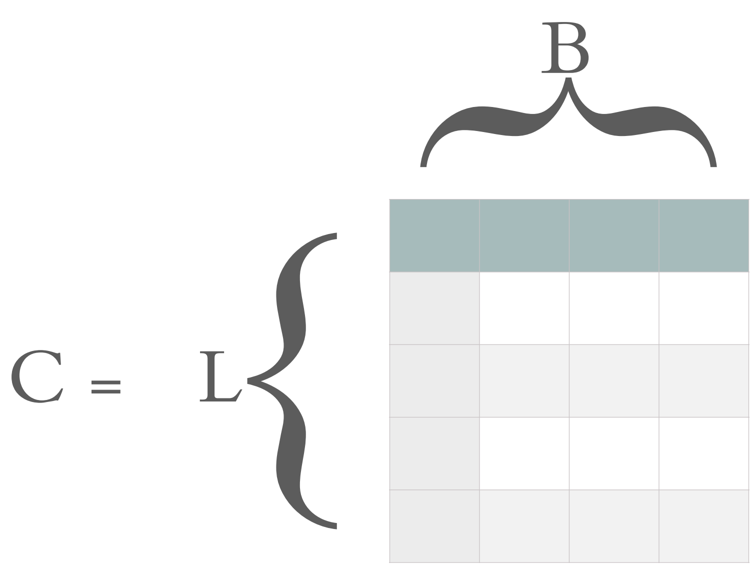
- تغییر (i, Δ):

$$C_{a,h_a(i)} \mathrel{+}= \sigma_r(i)\,\Delta \quad \text{for } a = 1, \dots, L.$$

برای مسئله پرسش نقطه‌ای l۲
و x_i آزاد

الگوریتم COUNTSKETCH

- توابع درهم‌سازی ۲ـ طرفه $h_1, \ldots h_L : [n] \rightarrow [\mathrm{B}]$
- توابع درهم‌سازی ۲ـ طرفه $\sigma_1, \sigma_2, \ldots, \sigma_L : [n] \rightarrow \{-1, +1\}$
- $B = 9k \quad L = C \log(1/\delta)$



- تغییر (i, Δ):

$$C_{a,h_a(i)} \mathrel{+}= \sigma_r(i)\,\Delta \qquad \text{for } a = 1, \ldots, L.$$

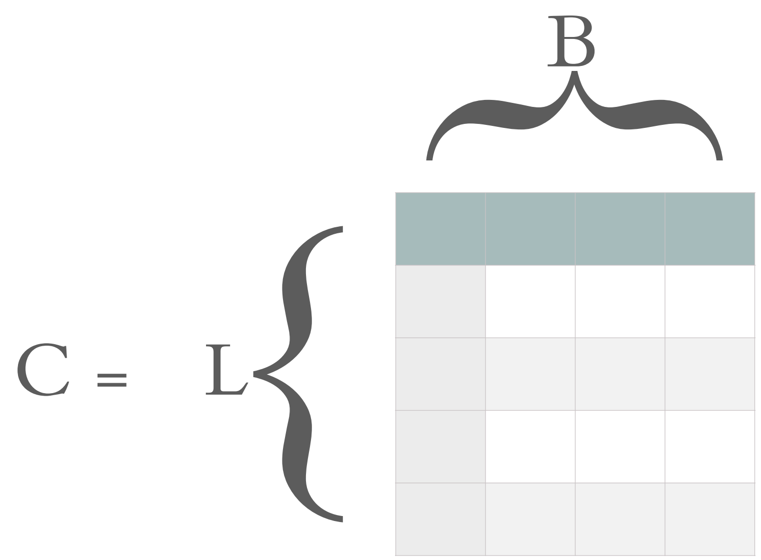
- پاسخ به پرسش i: میانه (روی r):

$$\sigma_r(i) C_{r,h_r(i)}$$

الگوریتم COUNTSKETCH

برای مسئله پرسش نقطه‌ای ۱۲ و آزاد x_i

- توابع درهم‌سازی ۲ـ طرفه $h_1, ...h_L : [n] \to [B]$
- توابع درهم‌سازی ۲ـ طرفه $\sigma_1, \sigma_2, \ldots, \sigma_L : [n] \to \{-1, +1\}$
- $B = 9k \quad L = C \log(1/\delta)$



- تغییر (i, Δ):

$$C_{a,h_a(i)} \mathrel{+}= \sigma_r(i)\,\Delta \quad \text{for } a = 1, \ldots, L.$$

- پاسخ به پرسش i: میانه (روی r):

$$\sigma_r(i) C_{r,h_r(i)}$$

حافظه: O(LB)

تحلیل الگوریتم COUNTSKETCH

Lemma 4.1.8. CountSketch.query(i) *returns* $x_i \pm \|x\|_2/\sqrt{k}$ *w.p.* $\geq 1 - \delta$.

اتفاق بد در سطر r: xjهای دیگر روی i بیافتند $h_r(j) = h_r(i)$

Zj: برای r و i ثابت، $h_r(j) = h_r(i)$

تحلیل الگوریتم COUNTSKETCH

Lemma 4.1.8. CountSketch.query(i) *returns* $x_i \pm \|x\|_2/\sqrt{k}$ *w.p.* $\geq 1-\delta$.

اتفاق بد در سطر r: xjهای دیگر روی i بیافتند $h_r(j) = h_r(i)$

Zj: برای r و i ثابت، $h_r(j) = h_r(i)$

$$E_r := \sigma_r(i) C_{r,h_r(i)} - x_i = \sum_{j \neq i} \sigma_r(i)\sigma_r(j) Z_j x_j$$

جواب ما

تحلیل الگوریتم COUNTSKETCH

Lemma 4.1.8. CountSketch.query(i) *returns* $x_i \pm \|x\|_2/\sqrt{k}$ *w.p.* $\geq 1-\delta$.

اتفاق بد در سطر r: xjهای دیگر روی i بیافتند $h_r(j) = h_r(i)$

Zj: برای r و i ثابت، $h_r(j) = h_r(i)$

$$E_r := \sigma_r(i)C_{r,h_r(i)} - x_i = \sum_{j\neq i} \sigma_r(i)\sigma_r(j)Z_jx_j$$

جواب ما

حکم جدید

میانه از L اجرای موازی

$$\mathbb{P}(|E_r| > \|x\|_2/\sqrt{k}) < 1/3$$

تحلیل الگوریتم COUNTSKETCH

Lemma 4.1.8. CountSketch.query(i) *returns* $x_i \pm \|x\|_2/\sqrt{k}$ *w.p.* $\geq 1-\delta$.

اتفاق بد در سطر r: xjهای دیگر روی i بیافتند $h_r(j) = h_r(i)$

Zj: برای r و i ثابت، $h_r(j) = h_r(i)$

$$E_r := \sigma_r(i)C_{r,h_r(i)} - x_i = \sum_{j\neq i} \sigma_r(i)\sigma_r(j)Z_j x_j$$

جواب ما

حکم جدید

میانه از L اجرای موازی

$$\mathbb{P}(|E_r| > \|x\|_2/\sqrt{k}) < 1/3$$

مارکوف:

$$P[\,|E_r| > 3 \cdot E\,|E_r|\,] < 1/3$$

حکم جدید

$$\mathbb{E}\,|E_r| \leq \frac{1}{3} \cdot \frac{\|x\|_2}{\sqrt{k}}$$

$$\mathbb{E}\,|E_r| \leq \sqrt{\mathbb{E}\,E_r^2}$$

$$\sum p_i x_i \leq \sqrt{\sum p_i x_i^2}$$

$$\left(\sum v_i u_i \leq \sqrt{v_i^2}\sqrt{u_i^2} \right)$$

$$v_i = \sqrt{p_i}\,x_i, u_i = \sqrt{p_i}$$

$$\mathbb{E}\,|E_r| \leq \sqrt{\mathbb{E}\,E_r^2} = \left(\mathbb{E}\left[\sum_{j,j'\neq i} \sigma_r(j)\sigma_r(j')Z_jZ'_jx_jx_{j'}\right]\right)^{1/2}$$

$E_r =$

$\sum_{j\neq i} \sigma_r(i)\sigma_r(j)Z_jx_j$

$\sum p_ix_i \leq \sqrt{\sum p_ix_i^2}$

$(\sum v_iu_i \leq \sqrt{v_i^2}\sqrt{u_i^2})$

$v_i = \sqrt{p_i}x_i, u_i = \sqrt{p_i}$

$$\mathbb{E}\,|E_r| \leq \sqrt{\mathbb{E}\,E_r^2} = \left(\mathbb{E}\left[\sum_{j,j'\neq i}\sigma_r(j)\sigma_r(j')Z_jZ'_jx_jx_{j'}\right]\right)^{1/2}$$

$E_r =$

$\sum_{j\neq i}\sigma_r(i)\sigma_r(j)Z_jx_j$

$$= \left(\mathbb{E}\left[\sum_{j\neq i}Z_jx_j^2 + \sum_{\substack{j,j'\neq i\\ j\neq j'}}\sigma_r(j)\sigma_r(j')Z_jZ'_jx_jx_{j'}\right]\right)^{1/2}$$

$\sum p_i x_i \leq \sqrt{\sum p_i x_i^2}$

$\left(\sum v_i u_i \leq \sqrt{v_i^2}\sqrt{u_i^2}\right)$

$v_i = \sqrt{p_i}x_i, u_i = \sqrt{p_i}$

$$\mathbb{E}\,|E_r| \le \sqrt{\mathbb{E}\,E_r^2} = \left(\mathbb{E}\left[\sum_{j,j'\neq i} \sigma_r(j)\sigma_r(j')Z_jZ_j'x_jx_{j'}\right]\right)^{1/2}$$

$E_r =$ $\sum_{j\neq i}\sigma_r(i)\sigma_r(j)Z_jx_j$

$$= \left(\mathbb{E}\left[\sum_{j\neq i} Z_jx_j^2 + \sum_{\substack{j,j'\neq i\\ j\neq j'}} \sigma_r(j)\sigma_r(j')Z_jZ_j'x_jx_{j'}\right]\right)^{1/2}$$

$$= \left(\sum_{j\neq i}(\mathbb{E}\,Z_j)x_j^2 + \sum_{\substack{j,j'\neq i\\ j\neq j'}} (\mathbb{E}\,\sigma_r(j)\sigma_r(j'))(\mathbb{E}\,Z_jZ_j')x_jx_{j'}\right)^{1/2}$$

$\sum p_ix_i \le \sqrt{\sum p_ix_i^2}$
($\sum v_iu_i \le \sqrt{v_i^2}\sqrt{u_i^2}$)

$v_i = \sqrt{p_i}x_i, u_i = \sqrt{p_i}$

$$\mathbb{E}\,|E_r| \le \sqrt{\mathbb{E}\,E_r^2} = \left(\mathbb{E}\left[\sum_{j,j'\neq i}\sigma_r(j)\sigma_r(j')Z_jZ'_jx_jx_{j'}\right]\right)^{1/2}$$

$E_r = \sum_{j\neq i}\sigma_r(i)\sigma_r(j)Z_jx_j$

$\sum p_ix_i \le \sqrt{\sum p_ix_i^2}$

$(\ \sum v_iu_i \le \sqrt{v_i^2}\sqrt{u_i^2}\ )$

$v_i = \sqrt{p_i}x_i, u_i = \sqrt{p_i}$

$$= \left(\mathbb{E}\left[\sum_{j\neq i}Z_jx_j^2 + \sum_{\substack{j,j'\neq i\\ j\neq j'}}\sigma_r(j)\sigma_r(j')Z_jZ'_jx_jx_{j'}\right]\right)^{1/2}$$

$$= \left(\sum_{j\neq i}(\mathbb{E}\,Z_j)x_j^2 + \sum_{\substack{j,j'\neq i\\ j\neq j'}}(\mathbb{E}\,\sigma_r(j)\sigma_r(j'))(\mathbb{E}\,Z_jZ'_j)x_jx_{j'}\right)^{1/2}$$

استقلال ۲_طرفه

$$= \left(\sum_{j\neq i}(\mathbb{E}\,Z_j)x_j^2 + \sum_{\substack{j,j'\neq i\\ j\neq j'}}(\mathbb{E}\,\sigma_r(j))(\mathbb{E}\,\sigma_r(j'))(\mathbb{E}\,Z_jZ'_j)x_jx_{j'}\right)^{1/2}$$

$$\mathbb{E}\,|E_r| \le \sqrt{\mathbb{E}\,E_r^2} = \left(\mathbb{E}\left[\sum_{j,j'\neq i} \sigma_r(j)\sigma_r(j')Z_jZ'_jx_jx_{j'}\right]\right)^{1/2}$$

$E_r = \sum_{j\neq i} \sigma_r(i)\sigma_r(j)Z_jx_j$

$\sum p_ix_i \le \sqrt{\sum p_ix_i^2}$

$(\sum v_iu_i \le \sqrt{v_i^2}\sqrt{u_i^2})$

$v_i = \sqrt{p_i}x_i, u_i = \sqrt{p_i}$

$$= \left(\mathbb{E}\left[\sum_{j\neq i} Z_jx_j^2 + \sum_{\substack{j,j'\neq i\\ j\neq j'}} \sigma_r(j)\sigma_r(j')Z_jZ'_jx_jx_{j'}\right]\right)^{1/2}$$

$$= \left(\sum_{j\neq i}(\mathbb{E}\,Z_j)x_j^2 + \sum_{\substack{j,j'\neq i\\ j\neq j'}} (\mathbb{E}\,\sigma_r(j)\sigma_r(j'))(\mathbb{E}\,Z_jZ'_j)x_jx_{j'}\right)^{1/2}$$

استقلال ۲_طرفه

$$= \left(\sum_{j\neq i}(\mathbb{E}\,Z_j)x_j^2 + \sum_{\substack{j,j'\neq i\\ j\neq j'}} (\mathbb{E}\,\sigma_r(j))(\mathbb{E}\,\sigma_r(j'))(\mathbb{E}\,Z_jZ'_j)x_jx_{j'}\right)^{1/2}$$

=0

$$= \left(\sum_{j\neq i}(\mathbb{E}\,Z_j)x_j^2\right)^{1/2}$$

$$
\begin{aligned}
\mathbb{E}\,|E_r| \le \sqrt{\mathbb{E}\,E_r^2} &= \left(\mathbb{E}\left[\sum_{j,j'\neq i}\sigma_r(j)\sigma_r(j')Z_jZ_j'x_jx_{j'}\right]\right)^{1/2} \\
&= \left(\mathbb{E}\left[\sum_{j\neq i}Z_jx_j^2 + \sum_{\substack{j,j'\neq i\\ j\neq j'}}\sigma_r(j)\sigma_r(j')Z_jZ_j'x_jx_{j'}\right]\right)^{1/2} \\
&= \left(\sum_{j\neq i}(\mathbb{E}\,Z_j)x_j^2 + \sum_{\substack{j,j'\neq i\\ j\neq j'}}(\mathbb{E}\,\sigma_r(j)\sigma_r(j'))(\mathbb{E}\,Z_jZ_j')x_jx_{j'}\right)^{1/2} \\
&= \left(\sum_{j\neq i}(\mathbb{E}\,Z_j)x_j^2 + \sum_{\substack{j,j'\neq i\\ j\neq j'}}\underbrace{(\mathbb{E}\,\sigma_r(j))}_{=0}(\mathbb{E}\,\sigma_r(j'))(\mathbb{E}\,Z_jZ_j')x_jx_{j'}\right)^{1/2} \\
&= \left(\sum_{j\neq i}\underbrace{(\mathbb{E}\,Z_j)}_{=1/B}x_j^2\right)^{1/2} \le \frac{\|x\|_2}{\sqrt{B}} = \frac{1}{3}\cdot\frac{\|x\|_2}{\sqrt{k}}
\end{aligned}
$$

$E_r = \sum_{j\neq i}\sigma_r(i)\sigma_r(j)Z_jx_j$

$\sum p_i x_i \le \sqrt{\sum p_i x_i^2}$

$\left(\sum v_i u_i \le \sqrt{v_i^2}\sqrt{u_i^2}\right)$

$v_i = \sqrt{p_i}\,x_i,\ u_i = \sqrt{p_i}$

استقلال ۲_طرفه

تحلیل الگوریتم COUNTSKETCH

Lemma 4.1.8. CountSketch.query(i) *returns* $x_i \pm \|x\|_2/\sqrt{k}$ *w.p.* $\geq 1-\delta$.

اتفاق بد در سطر r: xjهای دیگر روی i بیافتند $h_r(j) = h_r(i)$

Zj: برای r و i ثابت، $h_r(j) = h_r(i)$

$$E_r := \sigma_r(i)C_{r,h_r(i)} - x_i = \sum_{j\neq i} \sigma_r(i)\sigma_r(j)Z_j x_j$$

جواب ما

میانه از L اجرای موازی

$$\mathbb{P}(|E_r| > \|x\|_2/\sqrt{k}) < 1/3$$

حکم جدید

مارکوف:

$$P[|E_r| > 3 \cdot E|E_r|] < 1/3$$

$$\mathbb{E}|E_r| \leq \frac{1}{3} \cdot \frac{\|x\|_2}{\sqrt{k}}$$

حکم جدید