CSE599d: Advanced Query Processing

Lecture 14: The AGM Bound

Dan Suciu

University of Washington

Announcements

- HW3 due: Friday Feb. 27

Where We Are

- Tree decomposition of a graph, tree-width $\textsc{tw}(G)$.
- Hypertree decomposition of a hypergraph/query, hypertree width $\textsc{htw}(Q)$.
- Query evaluation in time $O(N^{\textsc{htw}} + \texttt{OUT})$
- Can we evaluate the bags more efficietly than $N^{\textsc{htw}}$?
  How large are the bags in the first place? Cardinality estimation?

Outline

Cardinality estimation: critical for query optimization.

Alternative: upper bound on the cardinality.

Today: the AGM Bound.

Monday: Worst Case Optimal Join (WCOJ) a.k.a. Generic Join (GJ).

Later: significant extensions of the AGM bound.

Motivation for Upper Bound

Recap: Cardinality Estimation

Given:

- Statistics on the input relations $R_1, R_2, \ldots$
- A selection-join query (same as full conjunctive query) Q

"Estimate":

- The size $|Q(\boldsymbol{D})|$.

Recap: Cardinality Estimation

Assumptions: uniformity, independence, preservation of values

- Selection $\sigma_p(R)$:

$$\boxed{\texttt{Est}(\sigma_p(R)) = \theta_p \cdot |R|} \qquad \boxed{\theta_{p_1 \wedge p_2} = \theta_{p_1} \cdot \theta_{p_2}}$$

- Join:

$$\boxed{\texttt{Est}(R(A,B) \bowtie S(B,C)) = \frac{|R| \cdot |S|}{\max(|\text{Dom}(R.B)|, |\text{Dom}(S.B)|)}}$$

- 1-d histograms address uniformity.
  2-d histograms address independence, but rarely used why???

The Output Bound Problem

Instead of an estimate, compute a guaranteed output bound B:

$$\boxed{|Q(\boldsymbol{D})| \leq B}$$

Challenge: make B tight (as close as possible to $|Q(\boldsymbol{D})|$).

Simple Examples

Assume $|R| \leq N$, $|S| \leq N$, $|T| \leq N$.

- $Q(X,Y,Z) = R(X,Y) \wedge S(Y,Z)$. $\max_{\boldsymbol{D}} |Q(\boldsymbol{D})| = ?$

Simple Examples

Assume $|R| \leq N$, $|S| \leq N$, $|T| \leq N$.

- $Q(X, Y, Z) = R(X, Y) \wedge S(Y, Z)$. $\max_{\boldsymbol{D}} |Q(\boldsymbol{D})| = N^2$

Simple Examples

Assume $|R| \leq N$, $|S| \leq N$, $|T| \leq N$.

- $Q(X, Y, Z) = R(X, Y) \wedge S(Y, Z)$. $\max_{\boldsymbol{D}} |Q(\boldsymbol{D})| = N^2$
  If $S.Y$ is a key: $\max_{\boldsymbol{D}} |Q(\boldsymbol{D})| = N$

Simple Examples

Assume $|R| \leq N$, $|S| \leq N$, $|T| \leq N$.

- $Q(X,Y,Z) = R(X,Y) \wedge S(Y,Z)$. $\max_{\boldsymbol{D}} |Q(\boldsymbol{D})| = N^2$
  If $S.Y$ is a key: $\max_{\boldsymbol{D}} |Q(\boldsymbol{D})| = N$
- $Q(X,Y,Z,U) = R(X,Y) \wedge S(Y,Z) \wedge T(Z,U)$. $\max_{\boldsymbol{D}} |Q(\boldsymbol{D})| = ?$

Simple Examples

Assume $|R| \leq N$, $|S| \leq N$, $|T| \leq N$.

- $Q(X,Y,Z) = R(X,Y) \wedge S(Y,Z)$. $\max_{\mathbf{D}} |Q(\mathbf{D})| = N^2$
  If $S.Y$ is a key: $\max_{\mathbf{D}} |Q(\mathbf{D})| = N$

- $Q(X,Y,Z,U) = R(X,Y) \wedge S(Y,Z) \wedge T(Z,U)$. $\max_{\mathbf{D}} |Q(\mathbf{D})| = N^2$

Simple Examples

Assume $|R| \leq N$, $|S| \leq N$, $|T| \leq N$.

- $Q(X,Y,Z) = R(X,Y) \wedge S(Y,Z)$. $\max_{\boldsymbol{D}} |Q(\boldsymbol{D})| = N^2$
  If $S.Y$ is a key: $\max_{\boldsymbol{D}} |Q(\boldsymbol{D})| = N$
- $Q(X,Y,Z,U) = R(X,Y) \wedge S(Y,Z) \wedge T(Z,U)$. $\max_{\boldsymbol{D}} |Q(\boldsymbol{D})| = N^2$
  Notice the role of an edge cover

Simple Examples

Assume $|R| \leq N$, $|S| \leq N$, $|T| \leq N$.

- $Q(X,Y,Z) = R(X,Y) \wedge S(Y,Z)$. $\max_{\mathbf{D}} |Q(\mathbf{D})| = N^2$
  If $S.Y$ is a key: $\max_{\mathbf{D}} |Q(\mathbf{D})| = N$
- $Q(X,Y,Z,U) = R(X,Y) \wedge S(Y,Z) \wedge T(Z,U)$. $\max_{\mathbf{D}} |Q(\mathbf{D})| = N^2$
  Notice the role of an edge cover
- $Q(X,Y,Z) = R(X,Y) \wedge S(Y,Z) \wedge T(Z,X)$. $\max_{\mathbf{D}} |Q(\mathbf{D})| =?$

Simple Examples

Assume $|R| \leq N$, $|S| \leq N$, $|T| \leq N$.

- $Q(X, Y, Z) = R(X, Y) \wedge S(Y, Z)$. $\max_{\boldsymbol{D}} |Q(\boldsymbol{D})| = N^2$
  If $S.Y$ is a key: $\max_{\boldsymbol{D}} |Q(\boldsymbol{D})| = N$
- $Q(X, Y, Z, U) = R(X, Y) \wedge S(Y, Z) \wedge T(Z, U)$. $\max_{\boldsymbol{D}} |Q(\boldsymbol{D})| = N^2$
  Notice the role of an edge cover
- $Q(X, Y, Z) = R(X, Y) \wedge S(Y, Z) \wedge T(Z, X)$. $\max_{\boldsymbol{D}} |Q(\boldsymbol{D})| = N^2$

Simple Examples

Assume $|R| \leq N$, $|S| \leq N$, $|T| \leq N$.

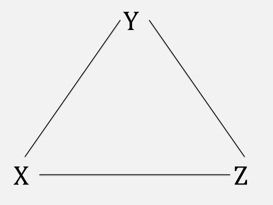
- $Q(X,Y,Z) = R(X,Y) \wedge S(Y,Z)$. $\max_{\mathbf{D}} |Q(\mathbf{D})| = N^2$
  If $S.Y$ is a key: $\max_{\mathbf{D}} |Q(\mathbf{D})| = N$
- $Q(X,Y,Z,U) = R(X,Y) \wedge S(Y,Z) \wedge T(Z,U)$. $\max_{\mathbf{D}} |Q(\mathbf{D})| = N^2$
  Notice the role of an edge cover
- $Q(X,Y,Z) = R(X,Y) \wedge S(Y,Z) \wedge T(Z,X)$. $\max_{\mathbf{D}} |Q(\mathbf{D})| = N^{\frac{3}{2}}$
  Here we use a fractional edge cover

AGM Bound: The Statement

Fractional Edge Covers

Query Q to hypegraph $G = (V, E)$.

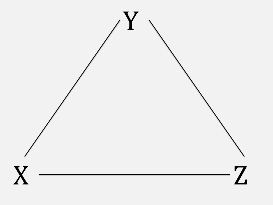
$R(X, Y) \wedge S(Y, Z) \wedge T(Z, X)$



Fractional Edge Covers

Query Q to hypegraph $G = (V, E)$.

$R(X, Y) \wedge S(Y, Z) \wedge T(Z, X)$



A fractional edge cover is $\boldsymbol{w} = (w_e)_{e \in E}, w_e \geq 0$:

$$\forall x \in V, \sum_{e \in E : x \in e} w_e \geq 1.$$

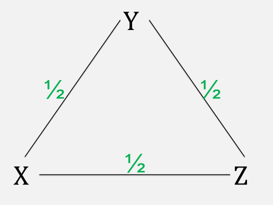
Fractional Edge Covers

Query Q to hypegraph $G = (V, E)$.

$R(X,Y) \wedge S(Y,Z) \wedge T(Z,X)$

A fractional edge cover is $\boldsymbol{w} = (w_e)_{e \in E}$, $w_e \geq 0$:

$$\forall x \in V, \sum_{e \in E : x \in e} w_e \geq 1.$$



The AGM Bound [Atserias et al., 2013]

$$Q(\boldsymbol{X}) = R_1(\boldsymbol{Y}_1) \wedge \cdots \wedge R_m(\boldsymbol{Y}_m)$$

The fractional edge covering number is: $\boxed{\rho^* = \min_{\boldsymbol{w}} (w_1 + \cdots + w_m)}$

The AGM Bound [Atserias et al., 2013]

$$Q(\boldsymbol{X}) = R_1(\boldsymbol{Y}_1) \wedge \cdots \wedge R_m(\boldsymbol{Y}_m)$$

The fractional edge covering number is: $\boxed{\rho^* = \min_{\boldsymbol{w}} (w_1 + \cdots + w_m)}$

Theorem [Upper Bound] If $|R_1|, \cdots, |R_m| \leq N$. Then:

$$\boxed{Q(\boldsymbol{D}) \leq N^{\rho^*}}$$

The AGM Bound [Atserias et al., 2013]

$$Q(\boldsymbol{X}) = R_1(\boldsymbol{Y}_1) \wedge \cdots \wedge R_m(\boldsymbol{Y}_m)$$

The fractional edge covering number is: $\boxed{\rho^* = \min_{\boldsymbol{w}} (w_1 + \cdots + w_m)}$

Theorem [Upper Bound] If $|R_1|, \cdots, |R_m| \leq N$. Then:

$$\boxed{Q(\boldsymbol{D}) \leq N^{\rho^*}}$$

Theorem [Lower Bound] There exists $|R_1|, \cdots, |R_m| \leq N$ and $Q(\boldsymbol{D}) \geq \frac{1}{2^{|\boldsymbol{X}|}} N^{\rho^*}$

$$\boxed{\text{The AGM bound } N^{\rho^*} \text{ is tight}}$$

Examples

What are $\boldsymbol{w}$ and ρ^*? The AGM bound? Assume $|R| = |S| = \cdots = N$.

5-cycle: $R(X, Y) \wedge S(Y, Z) \wedge T(Z, U) \wedge K(U, V) \wedge L(V, X)$

Examples

What are $\boldsymbol{w}$ and ρ^*? The AGM bound? Assume $|R| = |S| = \cdots = N$.

5-cycle: $R(X,Y) \wedge S(Y,Z) \wedge T(Z,U) \wedge K(U,V) \wedge L(V,X)$

$\boldsymbol{w} = (\frac{1}{2}, \frac{1}{2}, \frac{1}{2}, \frac{1}{2}, \frac{1}{2})$, $AGM = N^{5/2}$.

Examples

What are $\boldsymbol{w}$ and ρ^*? The AGM bound? Assume $|R| = |S| = \cdots = N$.

5-cycle: $R(X,Y) \wedge S(Y,Z) \wedge T(Z,U) \wedge K(U,V) \wedge L(V,X)$

$$\boldsymbol{w} = (\tfrac{1}{2}, \tfrac{1}{2}, \tfrac{1}{2}, \tfrac{1}{2}, \tfrac{1}{2}), \qquad AGM = N^{5/2}.$$

Loomis-Whitney: $R(X,Y,Z) \wedge S(X,Y,U) \wedge T(X,Z,U) \wedge K(Y,Z,U)$

Examples

What are $\boldsymbol{w}$ and ρ^*? The AGM bound? Assume $|R| = |S| = \cdots = N$.

5-cycle: $R(X,Y) \wedge S(Y,Z) \wedge T(Z,U) \wedge K(U,V) \wedge L(V,X)$

$$\boldsymbol{w} = (\tfrac{1}{2}, \tfrac{1}{2}, \tfrac{1}{2}, \tfrac{1}{2}, \tfrac{1}{2}), \qquad AGM = N^{5/2}.$$

Loomis-Whitney: $R(X,Y,Z) \wedge S(X,Y,U) \wedge T(X,Z,U) \wedge K(Y,Z,U)$

$$\boldsymbol{w} = (\tfrac{1}{3}, \tfrac{1}{3}, \tfrac{1}{3}, \tfrac{1}{3}), \qquad AGM = N^{4/3}$$

Examples

What are $\boldsymbol{w}$ and ρ^*? The AGM bound? Assume $|R| = |S| = \cdots = N$.

5-cycle: $R(X,Y) \wedge S(Y,Z) \wedge T(Z,U) \wedge K(U,V) \wedge L(V,X)$

$\boldsymbol{w} = (\frac{1}{2}, \frac{1}{2}, \frac{1}{2}, \frac{1}{2}, \frac{1}{2})$, $AGM = N^{5/2}$.

Loomis-Whitney: $R(X,Y,Z) \wedge S(X,Y,U) \wedge T(X,Z,U) \wedge K(Y,Z,U)$

$\boldsymbol{w} = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3}, \frac{1}{3})$, $AGM = N^{4/3}$

Chain of length 4: $R(X,Y) \wedge S(Y,Z) \wedge T(Z,U) \wedge K(U,V)$

Examples

What are $\boldsymbol{w}$ and ρ^*? The AGM bound? Assume $|R| = |S| = \cdots = N$.

5-cycle:

$$R(X,Y) \wedge S(Y,Z) \wedge T(Z,U) \wedge K(U,V) \wedge L(V,X)$$
$$\boldsymbol{w} = (\tfrac{1}{2}, \tfrac{1}{2}, \tfrac{1}{2}, \tfrac{1}{2}, \tfrac{1}{2}), \quad AGM = N^{5/2}.$$

Loomis-Whitney:

$$R(X,Y,Z) \wedge S(X,Y,U) \wedge T(X,Z,U) \wedge K(Y,Z,U)$$
$$\boldsymbol{w} = (\tfrac{1}{3}, \tfrac{1}{3}, \tfrac{1}{3}, \tfrac{1}{3}), \quad AGM = N^{4/3}$$

Chain of length 4:

$$R(X,Y) \wedge S(Y,Z) \wedge T(Z,U) \wedge K(U,V)$$
$\boldsymbol{w} = (1,0,1,1)$, or $(1,1,0,1)$, $\quad AGM = N^3$.

The General AGM Bound

$Q(\boldsymbol{X}) = R_1(\boldsymbol{Y}_1) \wedge \cdots \wedge R_m(\boldsymbol{Y}_m)$

Theorem [Upper Bound] For every fractional edge cover $\boldsymbol{w}$:
$|Q| \leq |R_1|^{w_1} \cdots |R_m|^{w_m}$

The General AGM Bound

$Q(\boldsymbol{X}) = R_1(\boldsymbol{Y}_1) \wedge \cdots \wedge R_m(\boldsymbol{Y}_m)$

Theorem [Upper Bound] For every fractional edge cover $\boldsymbol{w}$:
$|Q| \leq |R_1|^{w_1} \cdots |R_m|^{w_m}$

Theorem [Lower Bound] $\boxed{AGM(Q) \stackrel{\text{def}}{=} \min_{\boldsymbol{w}} |R_1|^{w_1} \cdots |R_m|^{w_m}}$ is "tight".

The General AGM Bound

$$Q(\boldsymbol{X}) = R_1(\boldsymbol{Y}_1) \wedge \cdots \wedge R_m(\boldsymbol{Y}_m)$$

Theorem [Upper Bound] For every fractional edge cover $\boldsymbol{w}$:
$|Q| \leq |R_1|^{w_1} \cdots |R_m|^{w_m}$

Theorem [Lower Bound] $\boxed{AGM(Q) \stackrel{\text{def}}{=} \min_{\boldsymbol{w}} |R_1|^{w_1} \cdots |R_m|^{w_m}}$ is "tight".

$$R(X,Y) \wedge S(Y,Z) \wedge T(Z,X) \qquad AGM(Q) = \min \begin{pmatrix} (|R| \cdot |S| \cdot |T|)^{1/2} \\ |R| \cdot |S| \\ |R| \cdot |T| \\ |S| \cdot |T| \end{pmatrix}$$

Proof Outline

- Proof of upper bound: use an interesting inequality.

- Proof of lower bound: worst-case database using strong duality.

An Inequality

Review: Cauchy-Schwartz, Hölder

Assume non-negative vectors, or matrices $\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}, \ldots$

Cauchy-Schwartz: $\sum_x a_x b_x \leq \left(\sum_x a_x^2\right)^{1/2} \left(\sum_x b_x^2\right)^{1/2}$

Review: Cauchy-Schwartz, Hölder

Assume non-negative vectors, or matrices $\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}, \ldots$

Cauchy-Schwartz: $\sum_x a_x b_x \leq \left(\sum_x a_x^2\right)^{1/2} \left(\sum_x b_x^2\right)^{1/2}$

Hölder: if $\frac{1}{p} + \frac{1}{q} = 1$: $\sum_x a_x b_x \leq \left(\sum_x a_x^p\right)^{1/p} \left(\sum_x b_x^q\right)^{1/q}$

Review: Cauchy-Schwartz, Hölder

Assume non-negative vectors, or matrices $\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}, \ldots$

Cauchy-Schwartz: $\sum_x a_x b_x \leq \left(\sum_x a_x^2\right)^{1/2} \left(\sum_x b_x^2\right)^{1/2}$

Hölder: if $\frac{1}{p} + \frac{1}{q} = 1$: $\sum_x a_x b_x \leq \left(\sum_x a_x^p\right)^{1/p} \left(\sum_x b_x^q\right)^{1/q}$

Equivalently: if $u + v = 1$, $\sum_x a_x b_x \leq \left(\sum_x a_x^{1/u}\right)^u \left(\sum_x b_x^{1/v}\right)^v$

Review: Cauchy-Schwartz, Hölder

Assume non-negative vectors, or matrices $\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}, \ldots$

Cauchy-Schwartz: $\sum_x a_x b_x \leq \left(\sum_x a_x^2\right)^{1/2} \left(\sum_x b_x^2\right)^{1/2}$

Hölder: if $\frac{1}{p} + \frac{1}{q} = 1$: $\sum_x a_x b_x \leq \left(\sum_x a_x^p\right)^{1/p} \left(\sum_x b_x^q\right)^{1/q}$

Equivalently: if $u + v = 1$, $\sum_x a_x b_x \leq \left(\sum_x a_x^{1/u}\right)^u \left(\sum_x b_x^{1/v}\right)^v$

$$\sum_x a_x^u b_x^v \leq \left(\sum_x a_x\right)^u \left(\sum_x b_x\right)^v$$

Review: Cauchy-Schwartz, Hölder

Assume non-negative vectors, or matrices $\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}, \ldots$

Cauchy-Schwartz: $\sum_x a_x b_x \leq \left(\sum_x a_x^2\right)^{1/2} \left(\sum_x b_x^2\right)^{1/2}$

Hölder: if $\frac{1}{p} + \frac{1}{q} = 1$: $\sum_x a_x b_x \leq \left(\sum_x a_x^p\right)^{1/p} \left(\sum_x b_x^q\right)^{1/q}$

Equivalently: if $u + v = 1$, $\sum_x a_x b_x \leq \left(\sum_x a_x^{1/u}\right)^u \left(\sum_x b_x^{1/v}\right)^v$

Also holds when: $u + v \geq 1$. $\sum_x a_x^u b_x^v \leq \left(\sum_x a_x\right)^u \left(\sum_x b_x\right)^v$

Review: Cauchy-Schwartz, Hölder

Assume non-negative vectors, or matrices $\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}, \ldots$

Cauchy-Schwartz: $\sum_x a_x b_x \leq \left(\sum_x a_x^2\right)^{1/2} \left(\sum_x b_x^2\right)^{1/2}$

Hölder: if $\frac{1}{p} + \frac{1}{q} = 1$: $\sum_x a_x b_x \leq \left(\sum_x a_x^p\right)^{1/p} \left(\sum_x b_x^q\right)^{1/q}$

Equivalently: if $u + v = 1$, $\sum_x a_x b_x \leq \left(\sum_x a_x^{1/u}\right)^u \left(\sum_x b_x^{1/v}\right)^v$

Also holds when: $u + v \geq 1$. $\sum_x a_x^u b_x^v \leq \left(\sum_x a_x\right)^u \left(\sum_x b_x\right)^v$

If $u + v + w + \cdots \geq 1$: $\boxed{\sum_x a_x^u b_x^v c_x^w \cdots \leq \left(\sum_x a_x\right)^u \left(\sum_x b_x\right)^v \cdots}$

Friedgut's Inequality [Friedgut, 2004]

Hypergraph $G = (V = \{X_1, \ldots, X_n\}, E = \{e_1, \ldots, e_m\})$.

Friedgut's Inequality [Friedgut, 2004]

Hypergraph $G = (V = \{X_1, \ldots, X_n\}, E = \{e_1, \ldots, e_m\})$.

$N \in \mathbb{N}$, multi-dimensional vectors: $\boldsymbol{a}_1 \in \mathbb{R}_+^{N^{e_1}}, \ldots, \boldsymbol{a}_m \in \mathbb{R}_+^{N^{em}}$.

Friedgut's Inequality [Friedgut, 2004]

Hypergraph $G = (V = \{X_1, \ldots, X_n\}, E = \{e_1, \ldots, e_m\})$.

$N \in \mathbb{N}$, multi-dimensional vectors: $\boldsymbol{a}_1 \in \mathbb{R}_+^{N^{e_1}}, \ldots, \boldsymbol{a}_m \in \mathbb{R}_+^{N^{em}}$.

Fractional edge cover $w_1, \ldots, w_m$.

$$\boxed{\sum_{x_1,\cdots,x_n} \prod_{j=1,m} \left(a_{j,\pi_{e_j}(\boldsymbol{x})}\right)^{w_j} \leq \prod_j \left(\sum_{\boldsymbol{y}_j} a_{j,\boldsymbol{y}_j}\right)^{w_j}}$$

Friedgut's Inequality [Friedgut, 2004]

Hypergraph $G = (V = \{X_1, \ldots, X_n\}, E = \{e_1, \ldots, e_m\})$.

$N \in \mathbb{N}$, multi-dimensional vectors: $\boldsymbol{a}_1 \in \mathbb{R}_+^{N^{e_1}}, \ldots, \boldsymbol{a}_m \in \mathbb{R}_+^{N^{em}}$.

Fractional edge cover $w_1, \ldots, w_m$.

$$\sum_{x_1,\cdots,x_n} \prod_{j=1,m} \left(a_{j,\pi_{e_j}(\boldsymbol{x})}\right)^{w_j} \leq \prod_j \left(\sum_{\boldsymbol{y}_j} a_{j,\boldsymbol{y}_j}\right)^{w_j}$$

$G = (\{X, Y, Z\}, \{(XY), (YZ), (ZX)\})$,

Friedgut's Inequality [Friedgut, 2004]

Hypergraph $G = (V = \{X_1, \ldots, X_n\}, E = \{e_1, \ldots, e_m\})$.

$N \in \mathbb{N}$, multi-dimensional vectors: $\boldsymbol{a}_1 \in \mathbb{R}_+^{N^{e_1}}, \ldots, \boldsymbol{a}_m \in \mathbb{R}_+^{N^{em}}$.

Fractional edge cover $w_1, \ldots, w_m$.

$$\boxed{\sum_{x_1,\cdots,x_n} \prod_{j=1,m} \left(a_{j,\pi_{e_j}(\boldsymbol{x})}\right)^{w_j} \leq \prod_j \left(\sum_{\boldsymbol{y}_j} a_{j,\boldsymbol{y}_j}\right)^{w_j}}$$

$G = (\{X, Y, Z\}, \{(XY), (YZ), (ZX)\}), \boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c} \in \mathbb{R}_+^{N^2}, \boldsymbol{w} = \left(\frac{1}{2}, \frac{1}{2}, \frac{1}{2}\right)$.

Friedgut's Inequality [Friedgut, 2004]

Hypergraph $G = (V = \{X_1, \ldots, X_n\}, E = \{e_1, \ldots, e_m\})$.

$N \in \mathbb{N}$, multi-dimensional vectors: $\boldsymbol{a}_1 \in \mathbb{R}_+^{N^{e_1}}, \ldots, \boldsymbol{a}_m \in \mathbb{R}_+^{N^{em}}$.

Fractional edge cover $w_1, \ldots, w_m$.

$$\sum_{x_1,\cdots,x_n} \prod_{j=1,m} \left(a_{j,\pi_{e_j}(\boldsymbol{x})}\right)^{w_j} \leq \prod_j \left(\sum_{\boldsymbol{y}_j} a_{j,\boldsymbol{y}_j}\right)^{w_j}$$

$G = (\{X, Y, Z\}, \{(XY), (YZ), (ZX)\}), \boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c} \in \mathbb{R}_+^{N^2}, \boldsymbol{w} = \left(\frac{1}{2}, \frac{1}{2}, \frac{1}{2}\right)$.

$$\sum_{xyz} a_{xy}^{\frac{1}{2}} b_{yz}^{\frac{1}{2}} c_{zx}^{\frac{1}{2}} \leq \left(\sum_{xy} a_{xy}\right)^{\frac{1}{2}} \cdot \left(\sum_{yz} b_{yz}\right)^{\frac{1}{2}} \cdot \left(\sum_{zx} c_{zx}\right)^{\frac{1}{2}}$$

Friedgut's Inequality [Friedgut, 2004]

Hypergraph $G = (V = \{X_1, \ldots, X_n\}, E = \{e_1, \ldots, e_m\})$.

$N \in \mathbb{N}$, multi-dimensional vectors: $\boldsymbol{a}_1 \in \mathbb{R}_+^{N^{e_1}}, \ldots, \boldsymbol{a}_m \in \mathbb{R}_+^{N^{e_m}}$.

Fractional edge cover $w_1, \ldots, w_m$.

$$\boxed{\sum_{x_1,\cdots,x_n} \prod_{j=1,m} \left(a_{j,\pi_{e_j}(\boldsymbol{x})}\right)^{w_j} \leq \prod_j \left(\sum_{\boldsymbol{y}_j} a_{j,\boldsymbol{y}_j}\right)^{w_j}}$$

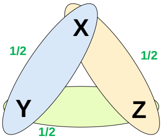
$G = (\{X, Y, Z\}, \{(XY), (YZ), (ZX)\}), \boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c} \in \mathbb{R}_+^{N^2}, \boldsymbol{w} = \left(\frac{1}{2}, \frac{1}{2}, \frac{1}{2}\right)$.

$$\sum_{xyz} a_{xy}^{\frac{1}{2}} b_{yz}^{\frac{1}{2}} c_{zx}^{\frac{1}{2}} \leq \left(\sum_{xy} a_{xy}\right)^{\frac{1}{2}} \cdot \left(\sum_{yz} b_{yz}\right)^{\frac{1}{2}} \cdot \left(\sum_{zx} c_{zx}\right)^{\frac{1}{2}}$$

$$\sum_{xyz} a_{xy} b_{yz} c_{zx} \leq \left(\sum_{xy} a_{xy}^2\right)^{\frac{1}{2}} \cdot \left(\sum_{yz} b_{yz}^2\right)^{\frac{1}{2}} \cdot \left(\sum_{zx} c_{zx}^2\right)^{\frac{1}{2}}$$

Example

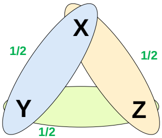
$$\sum_{xyz} a_{xy}^{\frac{1}{2}} b_{yz}^{\frac{1}{2}} c_{zx}^{\frac{1}{2}} \leq \left(\sum_{xy} a_{xy}\right)^{\frac{1}{2}} \cdot \left(\sum_{yz} b_{yz}\right)^{\frac{1}{2}} \cdot \left(\sum_{zx} c_{zx}\right)^{\frac{1}{2}}$$

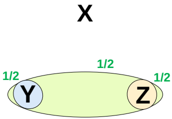


$$\sum_{xyz} a_{xy}^{\frac{1}{2}} b_{yz}^{\frac{1}{2}} c_{zx}^{\frac{1}{2}} = \sum_{x} \left(\sum_{yz} a_{xy}^{\frac{1}{2}} b_{yz}^{\frac{1}{2}} c_{zx}^{\frac{1}{2}} \right)$$

Example

$$\sum_{xyz} a_{xy}^{\frac{1}{2}} b_{yz}^{\frac{1}{2}} c_{zx}^{\frac{1}{2}} \leq \left(\sum_{xy} a_{xy}\right)^{\frac{1}{2}} \cdot \left(\sum_{yz} b_{yz}\right)^{\frac{1}{2}} \cdot \left(\sum_{zx} c_{zx}\right)^{\frac{1}{2}}$$

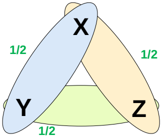


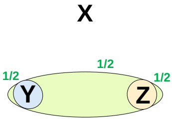


$$\sum_{xyz} a_{xy}^{\frac{1}{2}} b_{yz}^{\frac{1}{2}} c_{zx}^{\frac{1}{2}} = \sum_{x} \left(\sum_{yz} a_{xy}^{\frac{1}{2}} b_{yz}^{\frac{1}{2}} c_{zx}^{\frac{1}{2}} \right)$$

Example

$$\sum_{xyz} a_{xy}^{\frac{1}{2}} b_{yz}^{\frac{1}{2}} c_{zx}^{\frac{1}{2}} \leq \left(\sum_{xy} a_{xy}\right)^{\frac{1}{2}} \cdot \left(\sum_{yz} b_{yz}\right)^{\frac{1}{2}} \cdot \left(\sum_{zx} c_{zx}\right)^{\frac{1}{2}}$$



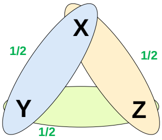


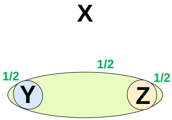
$$\sum_{xyz} a_{xy}^{\frac{1}{2}} b_{yz}^{\frac{1}{2}} c_{zx}^{\frac{1}{2}} = \sum_{x} \left(\sum_{yz} a_{xy}^{\frac{1}{2}} b_{yz}^{\frac{1}{2}} c_{zx}^{\frac{1}{2}} \right)$$

$$\leq \sum_{x} \left(\left(\sum_{y} a_{xy}\right)^{\frac{1}{2}} \cdot \left(\sum_{yz} b_{yz}\right)^{\frac{1}{2}} \cdot \left(\sum_{z} c_{zx}\right)^{\frac{1}{2}} \right)$$

Example

$$\boxed{\sum_{xyz} a_{xy}^{\frac{1}{2}} b_{yz}^{\frac{1}{2}} c_{zx}^{\frac{1}{2}} \leq \left(\sum_{xy} a_{xy}\right)^{\frac{1}{2}} \cdot \left(\sum_{yz} b_{yz}\right)^{\frac{1}{2}} \cdot \left(\sum_{zx} c_{zx}\right)^{\frac{1}{2}}}$$



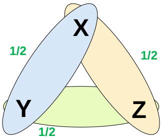


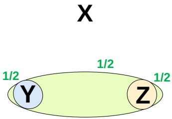
$$\sum_{xyz} a_{xy}^{\frac{1}{2}} b_{yz}^{\frac{1}{2}} c_{zx}^{\frac{1}{2}} = \sum_{x} \left(\sum_{yz} a_{xy}^{\frac{1}{2}} b_{yz}^{\frac{1}{2}} c_{zx}^{\frac{1}{2}} \right)$$

$$\leq \sum_{x} \left(\left(\sum_{y} a_{xy}\right)^{\frac{1}{2}} \cdot \left(\sum_{yz} b_{yz}\right)^{\frac{1}{2}} \cdot \left(\sum_{z} c_{zx}\right)^{\frac{1}{2}} \right) \stackrel{\text{def}}{=} \sum_{x} \left(A_x^{\frac{1}{2}} \cdot B^{\frac{1}{2}} \cdot C_x^{\frac{1}{2}} \right)$$

Example

$$\sum_{xyz} a_{xy}^{\frac{1}{2}} b_{yz}^{\frac{1}{2}} c_{zx}^{\frac{1}{2}} \leq \left(\sum_{xy} a_{xy}\right)^{\frac{1}{2}} \cdot \left(\sum_{yz} b_{yz}\right)^{\frac{1}{2}} \cdot \left(\sum_{zx} c_{zx}\right)^{\frac{1}{2}}$$





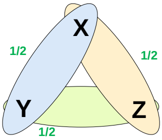
$$\sum_{xyz} a_{xy}^{\frac{1}{2}} b_{yz}^{\frac{1}{2}} c_{zx}^{\frac{1}{2}} = \sum_{x} \left(\sum_{yz} a_{xy}^{\frac{1}{2}} b_{yz}^{\frac{1}{2}} c_{zx}^{\frac{1}{2}} \right)$$

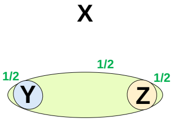
$$\leq \sum_{x} \left(\left(\sum_{y} a_{xy}\right)^{\frac{1}{2}} \cdot \left(\sum_{yz} b_{yz}\right)^{\frac{1}{2}} \cdot \left(\sum_{z} c_{zx}\right)^{\frac{1}{2}} \right) \stackrel{\text{def}}{=} \sum_{x} \left(A_x^{\frac{1}{2}} \cdot B^{\frac{1}{2}} \cdot C_x^{\frac{1}{2}} \right)$$

$$\leq \left(\sum_{x} A_x\right)^{\frac{1}{2}} \cdot B^{\frac{1}{2}} \cdot \left(\sum_{x} C_x\right)^{\frac{1}{2}}$$

Example

$$\boxed{\sum_{xyz} a_{xy}^{\frac{1}{2}} b_{yz}^{\frac{1}{2}} c_{zx}^{\frac{1}{2}} \leq \left(\sum_{xy} a_{xy}\right)^{\frac{1}{2}} \cdot \left(\sum_{yz} b_{yz}\right)^{\frac{1}{2}} \cdot \left(\sum_{zx} c_{zx}\right)^{\frac{1}{2}}}$$





$$\sum_{xyz} a_{xy}^{\frac{1}{2}} b_{yz}^{\frac{1}{2}} c_{zx}^{\frac{1}{2}} = \sum_{x} \left(\sum_{yz} a_{xy}^{\frac{1}{2}} b_{yz}^{\frac{1}{2}} c_{zx}^{\frac{1}{2}} \right)$$

$$\leq \sum_{x} \left(\left(\sum_{y} a_{xy}\right)^{\frac{1}{2}} \cdot \left(\sum_{yz} b_{yz}\right)^{\frac{1}{2}} \cdot \left(\sum_{z} c_{zx}\right)^{\frac{1}{2}} \right) \stackrel{\text{def}}{=} \sum_{x} \left(A_x^{\frac{1}{2}} \cdot B^{\frac{1}{2}} \cdot C_x^{\frac{1}{2}} \right)$$

$$\leq \left(\sum_{x} A_x\right)^{\frac{1}{2}} \cdot B^{\frac{1}{2}} \cdot \left(\sum_{x} C_x\right)^{\frac{1}{2}} = \left(\sum_{xy} a_{xy}\right)^{\frac{1}{2}} \cdot \left(\sum_{yz} b_{yz}\right)^{\frac{1}{2}} \cdot \left(\sum_{zx} c_{zx}\right)^{\frac{1}{2}}$$

Proof of Friedgut's Inequality

$$\sum_{x_1,\cdots,x_n} \prod_{j=1,m} \left(a_{j,\pi_{e_j}(\boldsymbol{x})}\right)^{w_j} \leq \prod_{j=1,m} \left(\sum_{\boldsymbol{y}_j} a_{j,\boldsymbol{y}_j}\right)^{w_j}$$

Induction:

Proof of Friedgut's Inequality

$$\boxed{\sum_{x_1,\cdots,x_n} \prod_{j=1,m} \left(a_{j,\pi_{e_j}(\boldsymbol{x})}\right)^{w_j} \leq \prod_{j=1,m} \left(\sum_{\boldsymbol{y}_j} a_{j,\boldsymbol{y}_j}\right)^{w_j}}$$

Induction: remove X_1 from G to obtain $G - \{X_1\}$, keep $w_1, \ldots, w_m$:

Proof of Friedgut's Inequality

$$\sum_{x_1,\cdots,x_n} \prod_{j=1,m} \left(a_{j,\pi_{e_j}(\boldsymbol{x})}\right)^{w_j} \leq \prod_{j=1,m} \left(\sum_{\boldsymbol{y}_j} a_{j,\boldsymbol{y}_j}\right)^{w_j}$$

Induction: remove X_1 from G to obtain $G - \{X_1\}$, keep $w_1, \ldots, w_m$:

- X_1 is covered in G

Proof of Friedgut's Inequality

$$\sum_{x_1,\cdots,x_n} \prod_{j=1,m} \left(a_{j,\pi_{e_j}(\boldsymbol{x})}\right)^{w_j} \leq \prod_{j=1,m} \left(\sum_{\boldsymbol{y}_j} a_{j,\boldsymbol{y}_j}\right)^{w_j}$$

Induction: remove X_1 from G to obtain $G - \{X_1\}$, keep $w_1, \ldots, w_m$:

- X_1 is covered in G
- Edges in $G - \{X_1\}$ are $e_1 - \{X_1\}, \ldots, e_m - \{X_1\}$; still covered.

Proof of Friedgut's Inequality

$$\sum_{x_1,\cdots,x_n} \prod_{j=1,m} \left(a_{j,\pi_{e_j}(\boldsymbol{x})}\right)^{w_j} \leq \prod_{j=1,m} \left(\sum_{\boldsymbol{y}_j} a_{j,\boldsymbol{y}_j}\right)^{w_j}$$

Induction: remove X_1 from G to obtain $G - \{X_1\}$, keep $w_1, \ldots, w_m$:

- X_1 is covered in G
- Edges in $G - \{X_1\}$ are $e_1 - \{X_1\}, \ldots, e_m - \{X_1\}$; still covered.

$$\sum_{x_1,\cdots,x_n} \prod_{j=1,m} \left(a_{j,\pi_{e_j}(\boldsymbol{x})}\right)^{w_j}$$

Proof of Friedgut's Inequality

$$\sum_{x_1,\cdots,x_n} \prod_{j=1,m} \left(a_{j,\pi_{e_j}(\boldsymbol{x})}\right)^{w_j} \leq \prod_{j=1,m} \left(\sum_{\boldsymbol{y}_j} a_{j,\boldsymbol{y}_j}\right)^{w_j}$$

Induction: remove X_1 from G to obtain $G - \{X_1\}$, keep $w_1, \ldots, w_m$:

- X_1 is covered in G
- Edges in $G - \{X_1\}$ are $e_1 - \{X_1\}, \ldots, e_m - \{X_1\}$; still covered.

$$\sum_{x_1,\cdots,x_n} \prod_{j=1,m} \left(a_{j,\pi_{e_j}(\boldsymbol{x})}\right)^{w_j} = \sum_{x_1} \left(\sum_{x_2,\cdots,x_n} \prod_{j=1,m} \left(a_{j,\pi_{e_j}(\boldsymbol{x})}\right)^{w_j} \right)$$

Proof of Friedgut's Inequality

$$\sum_{x_1,\cdots,x_n}\prod_{j=1,m}\left(a_{j,\pi_{e_j}(\boldsymbol{x})}\right)^{w_j} \leq \prod_{j=1,m}\left(\sum_{\boldsymbol{y}_j} a_{j,\boldsymbol{y}_j}\right)^{w_j}$$

Induction: remove X_1 from G to obtain $G - \{X_1\}$, keep $w_1, \ldots, w_m$:

- X_1 is covered in G
- Edges in $G - \{X_1\}$ are $e_1 - \{X_1\}, \ldots, e_m - \{X_1\}$; still covered.

$$\sum_{x_1,\cdots,x_n}\prod_{j=1,m}\left(a_{j,\pi_{e_j}(\boldsymbol{x})}\right)^{w_j} = \sum_{x_1}\left(\sum_{x_2,\cdots,x_n}\prod_{j=1,m}\left(a_{j,\pi_{e_j}(\boldsymbol{x})}\right)^{w_j}\right)$$

$$\overset{\text{Induction}}{\leq} \sum_{x_1}\left(\prod_{j=1,m}\Big(\sum_{\boldsymbol{y}_j-\{x_1\}} a_{j,\boldsymbol{y}_j}\Big)^{w_j}\right)$$

Proof of Friedgut's Inequality

$$\sum_{x_1,\cdots,x_n} \prod_{j=1,m} \left(a_{j,\pi_{e_j}(\boldsymbol{x})}\right)^{w_j} \leq \prod_{j=1,m} \left(\sum_{\boldsymbol{y}_j} a_{j,\boldsymbol{y}_j}\right)^{w_j}$$

Induction: remove X_1 from G to obtain $G - \{X_1\}$, keep $w_1, \ldots, w_m$:

- X_1 is covered in G
- Edges in $G - \{X_1\}$ are $e_1 - \{X_1\}, \ldots, e_m - \{X_1\}$; still covered.

$$\sum_{x_1,\cdots,x_n} \prod_{j=1,m} \left(a_{j,\pi_{e_j}(\boldsymbol{x})}\right)^{w_j} = \sum_{x_1} \left(\sum_{x_2,\cdots,x_n} \prod_{j=1,m} \left(a_{j,\pi_{e_j}(\boldsymbol{x})}\right)^{w_j} \right)$$

$$\overset{\text{Induction}}{\leq} \sum_{x_1} \left(\prod_{j=1,m} \Big(\sum_{\boldsymbol{y}_j - \{x_1\}} a_{j,\boldsymbol{y}_j} \Big)^{w_j} \right)$$

$$\overset{\text{Hölder}}{\leq} \prod_{j=1,m} \Big(\sum_{\boldsymbol{y}_j} a_{j,\boldsymbol{y}_j} \Big)^{w_j}$$

How To Impress Your Friends Doing ML

Upper bound on the trace of ABC.

Query/hypergraph: $A(X,Y) \wedge B(Y,Z) \wedge C(Z,X)$ $\boldsymbol{w} = (\frac{1}{2}, \frac{1}{2}, \frac{1}{2})$

$$\mathtt{Tr}(\boldsymbol{a} \cdot \boldsymbol{b} \cdot \boldsymbol{c}) \stackrel{\text{def}}{=} \sum_{xyz} a_{xy} b_{yz} c_{zx} \leq \left(\sum_{xy} a_{xy}^2\right)^{\frac{1}{2}} \left(\sum_{yz} b_{yz}^2\right)^{\frac{1}{2}} \left(\sum_{zx} c_{zx}^2\right)^{\frac{1}{2}} = ||\boldsymbol{a}||_2 \cdot ||\boldsymbol{b}||_2 \cdot ||\boldsymbol{c}||_2$$

How To Impress Your Friends Doing ML

Upper bound on the trace of ABC.

Query/hypergraph: $A(X,Y) \wedge B(Y,Z) \wedge C(Z,X)$ $\quad\quad$ $\boldsymbol{w} = (\frac{1}{2}, \frac{1}{2}, \frac{1}{2})$

$$\mathtt{Tr}(\boldsymbol{a}\cdot\boldsymbol{b}\cdot\boldsymbol{c}) \stackrel{\text{def}}{=} \sum_{xyz} a_{xy} b_{yz} c_{zx} \leq \left(\sum_{xy} a_{xy}^2\right)^{\frac{1}{2}} \left(\sum_{yz} b_{yz}^2\right)^{\frac{1}{2}} \left(\sum_{zx} c_{zx}^2\right)^{\frac{1}{2}} = ||\boldsymbol{a}||_2 \cdot ||\boldsymbol{b}||_2 \cdot ||\boldsymbol{c}||_2$$

The tensor kernel for Loomis-Whitney:

$A(X,Y,Z) \wedge B(Y,Z,U) \wedge C(Z,U,X) \wedge D(U,X,Y)$ $\quad\quad$ $\boldsymbol{w} = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3}, \frac{1}{3})$

$$\sum_{xyzu} a_{xyz} b_{yzu} c_{zux} d_{uxy} \leq \left(\sum_{xyz} (a_{xyz})^3\right)^{\frac{1}{3}} \cdots \left(\sum_{uxy} (d_{uxy})^3\right)^{\frac{1}{3}} = ||\boldsymbol{a}||_3 \cdot ||\boldsymbol{b}||_3 \cdot ||\boldsymbol{c}||_3 \cdot ||\boldsymbol{d}||_3$$

How To Impress Your Friends Doing ML

Upper bound on the trace of ABC.

Query/hypergraph: $A(X,Y) \wedge B(Y,Z) \wedge C(Z,X)$ $\quad \boldsymbol{w} = (\frac{1}{2}, \frac{1}{2}, \frac{1}{2})$

$$\texttt{Tr}(\boldsymbol{a} \cdot \boldsymbol{b} \cdot \boldsymbol{c}) \stackrel{\text{def}}{=} \sum_{xyz} a_{xy} b_{yz} c_{zx} \leq \left(\sum_{xy} a_{xy}^2\right)^{\frac{1}{2}} \left(\sum_{yz} b_{yz}^2\right)^{\frac{1}{2}} \left(\sum_{zx} c_{zx}^2\right)^{\frac{1}{2}} = ||\boldsymbol{a}||_2 \cdot ||\boldsymbol{b}||_2 \cdot ||\boldsymbol{c}||_2$$

The tensor kernel for Loomis-Whitney:

$A(X,Y,Z) \wedge B(Y,Z,U) \wedge C(Z,U,X) \wedge D(U,X,Y)$ $\quad \boldsymbol{w} = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3}, \frac{1}{3})$

$$\sum_{xyzu} a_{xyz} b_{yzu} c_{zux} d_{uxy} \leq \left(\sum_{xyz} (a_{xyz})^3\right)^{\frac{1}{3}} \cdots \left(\sum_{uxy} (d_{uxy})^3\right)^{\frac{1}{3}} = ||\boldsymbol{a}||_3 \cdot ||\boldsymbol{b}||_3 \cdot ||\boldsymbol{c}||_3 \cdot ||\boldsymbol{d}||_3$$

And another one

$A(X,U) \wedge B(Y,U) \wedge C(Z,U) \wedge D(X,Y,Z)$ $\quad \boldsymbol{w} = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3}, \frac{2}{3})$

$$\sum_{xyzu} a_{xu} b_{yu} c_{zu} d_{xyz} \leq \left(\sum_{xu} a_{xu}^3\right)^{\frac{1}{3}} \cdot \left(\sum_{yu} b_{yu}^3\right)^{\frac{1}{3}} \cdot \left(\sum_{zu} c_{zu}^3\right)^{\frac{1}{3}} \cdot \left(\sum_{xyz} d_{xyz}^{\frac{3}{2}}\right)^{\frac{2}{3}}$$

Proof of the AGM Upper Bound

$Q(X_1, \ldots, X_n) = R_1(\boldsymbol{Y}_1) \wedge \cdots \wedge R_m(\boldsymbol{Y}_m)$

Theorem For every fractional edge cover $\boldsymbol{w}$: $|Q| \leq |R_1^D|^{w_1} \cdots |R_m^D|^{w_m}$

Proof

$N \stackrel{\text{def}}{=}$ number of constants in the database $\boldsymbol{D}$.

$$\forall \boldsymbol{y}_j \in [N]^{e_j}, \qquad a_{j,\boldsymbol{y}_j} \stackrel{\text{def}}{=} \begin{cases} 1 & \text{if } \boldsymbol{y}_j \in R_j^D \\ 0 & \text{otherwise} \end{cases}$$

$$|Q(\boldsymbol{D})| = \sum_{x_1,\ldots,x_n} a_{1,\pi_{e_1}(\boldsymbol{x})} \cdots a_{m,\pi_{em}(\boldsymbol{x})} \leq \Big(\sum_{\boldsymbol{y}_1} a_{1,\boldsymbol{y}_1}^{\frac{1}{w_1}}\Big)^{w_1} \cdots \Big(\sum_{\boldsymbol{y}_m} a_{m,\boldsymbol{y}_m}^{\frac{1}{w_m}}\Big)^{w_m} = |R_1^D|^{w_1} \cdots |R_m^D|^{w_m}$$

Discussion

- Each fractional edge cover gives us some upper bound on $|Q(\boldsymbol{D})|$.
- Their minimum is also an upper bound.
- Next: will prove that this minimum is tight.

Proof of the Lower Bound

Proof of the AGM Lower Bound

By example:

$R(X,Y) \wedge S(Y,Z) \wedge T(Z,X)$ $\quad\quad AGM(Q) = \min_{\boldsymbol{w}} |R|^{w_R} \cdot |S|^{w_S} \cdot |T|^{w_T}$

Proof of the AGM Lower Bound

By example:

$R(X,Y) \wedge S(Y,Z) \wedge T(Z,X)$

$AGM(Q) = \min_{\boldsymbol{w}} |R|^{w_R} \cdot |S|^{w_S} \cdot |T|^{w_T}$

Primal program:

Minimize

$w_R \log |R| + w_S \log |S| + w_T \log |T|$

where $\boldsymbol{w}$ is frac. edge cover:

$$\begin{array}{lllll} X: & w_R + & & w_T & \geq 1 \\ Y: & w_R + & w_S & & \geq 1 \\ Z: & & w_S + & w_T & \geq 1 \end{array}$$

Proof of the AGM Lower Bound

By example:

$R(X,Y) \wedge S(Y,Z) \wedge T(Z,X)$

$$AGM(Q) = \min_{\boldsymbol{w}} |R|^{w_R} \cdot |S|^{w_S} \cdot |T|^{w_T}$$

Primal program:

Minimize

$w_R \log|R| + w_S \log|S| + w_T \log|T|$

where $\boldsymbol{w}$ is frac. edge cover:

$$\begin{array}{llllr} X: & w_R + & & w_T & \geq 1 \\ Y: & w_R + & w_S & & \geq 1 \\ Z: & & w_S + & w_T & \geq 1 \end{array}$$

Dual program:

Maximize

$$v_X + v_Y + v_Z$$

where $\boldsymbol{v}$ is "frac. vertex packing":

$$\begin{array}{llllr} R: & v_X + & v_Y & & \leq \log|R| \\ S: & & v_Y + & v_Z & \leq \log|S| \\ T: & v_X + & & v_Z & \leq \log|T| \end{array}$$

Proof of the AGM Lower Bound

By example:

$R(X,Y) \wedge S(Y,Z) \wedge T(Z,X)$ $\quad\quad$ $AGM(Q) = \min_{\boldsymbol{w}} |R|^{w_R} \cdot |S|^{w_S} \cdot |T|^{w_T}$

Primal program:

Minimize

$w_R \log|R| + w_S \log|S| + w_T \log|T|$

where $\boldsymbol{w}$ is frac. edge cover:

$$\begin{array}{lllll} X: & w_R + & & w_T & \geq 1 \\ Y: & w_R + & w_S & & \geq 1 \\ Z: & & w_S + & w_T & \geq 1 \end{array}$$

Dual program:

Maximize

$$v_X + v_Y + v_Z$$

where $\boldsymbol{v}$ is "frac. vertex packing":

$$\begin{array}{lllll} R: & v_X + & v_Y & & \leq \log|R| \\ S: & & v_Y + & v_Z & \leq \log|S| \\ T: & v_X + & & v_Z & \leq \log|T| \end{array}$$

Take optimum $\boldsymbol{v}$, define: $\text{Dom}(X) \stackrel{\text{def}}{=} [\lfloor 2^{v_X} \rfloor]$, $\text{Dom}(Y) \stackrel{\text{def}}{=} [\lfloor 2^{v_Y} \rfloor]$, $\text{Dom}(Z) \stackrel{\text{def}}{=} [\lfloor 2^{v_Z} \rfloor]$.

Proof of the AGM Lower Bound

By example:

$R(X,Y) \wedge S(Y,Z) \wedge T(Z,X)$ $\quad$ $AGM(Q) = \min_{\boldsymbol{w}} |R|^{w_R} \cdot |S|^{w_S} \cdot |T|^{w_T}$

Primal program:

Minimize

$w_R \log|R| + w_S \log|S| + w_T \log|T|$

where $\boldsymbol{w}$ is frac. edge cover:

$$\begin{array}{llllll} X: & w_R + & & w_T & \geq 1 \\ Y: & w_R + & w_S & & \geq 1 \\ Z: & & w_S + & w_T & \geq 1 \end{array}$$

Dual program:

Maximize

$$v_X + v_Y + v_Z$$

where $\boldsymbol{v}$ is "frac. vertex packing":

$$\begin{array}{lllll} R: & v_X + & v_Y & & \leq \log|R| \\ S: & & v_Y + & v_Z & \leq \log|S| \\ T: & v_X + & & v_Z & \leq \log|T| \end{array}$$

Take optimum $\boldsymbol{v}$, define: $\text{Dom}(X) \stackrel{\text{def}}{=} [\lfloor 2^{v_X} \rfloor]$, $\text{Dom}(Y) \stackrel{\text{def}}{=} [\lfloor 2^{v_Y} \rfloor]$, $\text{Dom}(Z) \stackrel{\text{def}}{=} [\lfloor 2^{v_Z} \rfloor]$.

Worst-case instance (cartesian products): $R^* \stackrel{\text{def}}{=} \text{Dom}(X) \times \text{Dom}(Y)$, $S^*, T^* \stackrel{\text{def}}{=} \cdots$

Proof of the AGM Lower Bound

By example:

$R(X,Y) \wedge S(Y,Z) \wedge T(Z,X)$ $\qquad AGM(Q) = \min_{\boldsymbol{w}} |R|^{w_R} \cdot |S|^{w_S} \cdot |T|^{w_T}$

Primal program:

Minimize

$w_R \log|R| + w_S \log|S| + w_T \log|T|$

where $\boldsymbol{w}$ is frac. edge cover:

$$\begin{array}{lllll} X: & w_R + & & w_T & \geq 1 \\ Y: & w_R + & w_S & & \geq 1 \\ Z: & & w_S + & w_T & \geq 1 \end{array}$$

Dual program:

Maximize

$$v_X + v_Y + v_Z$$

where $\boldsymbol{v}$ is "frac. vertex packing":

$$\begin{array}{lllll} R: & v_X + & v_Y & & \leq \log|R| \\ S: & & v_Y + & v_Z & \leq \log|S| \\ T: & v_X + & & v_Z & \leq \log|T| \end{array}$$

Take optimum $\boldsymbol{v}$, define: $\text{Dom}(X) \stackrel{\text{def}}{=} [\lfloor 2^{v_X} \rfloor]$, $\text{Dom}(Y) \stackrel{\text{def}}{=} [\lfloor 2^{v_Y} \rfloor]$, $\text{Dom}(Z) \stackrel{\text{def}}{=} [\lfloor 2^{v_Z} \rfloor]$.

Worst-case instance (cartesian products): $R^* \stackrel{\text{def}}{=} \text{Dom}(X) \times \text{Dom}(Y)$, $S^*, T^* \stackrel{\text{def}}{=} \cdots$

$$|Q^*| = \lfloor 2^{v_X} \rfloor \cdot \lfloor 2^{v_Y} \rfloor \cdot \lfloor 2^{v_Z} \rfloor \geq \frac{1}{8} 2^{v_X + v_Y + v_Z}$$

Proof of the AGM Lower Bound

By example:

$R(X,Y) \wedge S(Y,Z) \wedge T(Z,X)$ $\quad AGM(Q) = \min_{\boldsymbol{w}} |R|^{w_R} \cdot |S|^{w_S} \cdot |T|^{w_T}$

Primal program:

Minimize

$w_R \log|R| + w_S \log|S| + w_T \log|T|$

where $\boldsymbol{w}$ is frac. edge cover:

$$\begin{array}{llllr} X: & w_R+ & & w_T & \geq 1 \\ Y: & w_R+ & w_S & & \geq 1 \\ Z: & & w_S+ & w_T & \geq 1 \end{array}$$

Dual program:

Maximize

$$v_X + v_Y + v_Z$$

where $\boldsymbol{v}$ is "frac. vertex packing":

$$\begin{array}{llllr} R: & v_X+ & v_Y & & \leq \log|R| \\ S: & & v_Y+ & v_Z & \leq \log|S| \\ T: & v_X+ & & v_Z & \leq \log|T| \end{array}$$

Take optimum $\boldsymbol{v}$, define: $\text{Dom}(X) \stackrel{\text{def}}{=} [\lfloor 2^{v_X} \rfloor]$, $\text{Dom}(Y) \stackrel{\text{def}}{=} [\lfloor 2^{v_Y} \rfloor]$, $\text{Dom}(Z) \stackrel{\text{def}}{=} [\lfloor 2^{v_Z} \rfloor]$.

Worst-case instance (cartesian products): $R^* \stackrel{\text{def}}{=} \text{Dom}(X) \times \text{Dom}(Y)$, $S^*, T^* \stackrel{\text{def}}{=} \cdots$

$$|Q^*| = \lfloor 2^{v_X} \rfloor \cdot \lfloor 2^{v_Y} \rfloor \cdot \lfloor 2^{v_Z} \rfloor \geq \frac{1}{8} 2^{v_X+v_Y+v_Z} = \frac{1}{8} 2^{w_1^* \log|R| + w_2^* \log|S| + w_3^* \log|T|} = \frac{1}{8} AGM(Q)$$

Special Case: $|R| = |S| = \cdots = N$

Fix a hypergraph (V, E); $(v_X)_{X \in V} \in \mathbb{R}_+^{|V|}$ is a fractional vertex packing if:

$$\forall \boldsymbol{Y} \in E :, \boxed{\sum_{X \in \boldsymbol{Y}} v_X \leq 1}$$

Special Case: $|R| = |S| = \cdots = N$

Fix a hypergraph (V, E); $(v_X)_{X \in V} \in \mathbb{R}_+^{|V|}$ is a fractional vertex packing if:

$$\forall \mathbf{Y} \in E :, \boxed{\sum_{X \in \mathbf{Y}} v_X \leq 1}$$

When $|R| = |S| = \cdots = N$, then replace

$v_R + v_S \leq \log N$ $v_R + v_T \leq \log N$ $\ldots$	with	$v_R + v_S \leq 1$ $v_R + v_T \leq 1$ $\ldots$	times $\log N$.

Special Case: $|R| = |S| = \cdots = N$

Fix a hypergraph (V, E); $(v_X)_{X \in V} \in \mathbb{R}_+^{|V|}$ is a fractional vertex packing if:

$$\forall \boldsymbol{Y} \in E :, \boxed{\sum_{X \in \boldsymbol{Y}} v_X \leq 1}$$

When $|R| = |S| = \cdots = N$, then replace

$$\boxed{\begin{aligned} v_R + v_S &\leq \log N \\ v_R + v_T &\leq \log N \\ &\ldots \end{aligned}} \quad \text{with} \quad \boxed{\begin{aligned} v_R + v_S &\leq 1 \\ v_R + v_T &\leq 1 \\ &\ldots \end{aligned}} \quad \text{times } \log N.$$

Then: $R = [N^{v_X}] \times [N^{v_Y}]$, $S = [N^{v_Y}] \times [N^{v_Z}]$, $T = [N^{v_X}] \times [N^{v_Z}]$.

$$Q = [N^{v_X}] \times [N^{v_Y}] \times [N^{v_Z}]$$

Examples

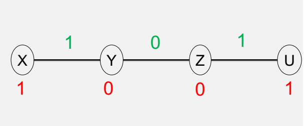
$|R| = |S| = \cdots = N$

$R(X,Y) \wedge S(Y,Z) \wedge T(Z,U)$

Examples

$|R| = |S| = \cdots = N$

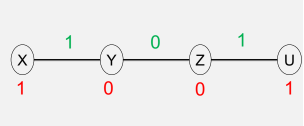
$R(X,Y) \wedge S(Y,Z) \wedge T(Z,U)$



Examples

$|R| = |S| = \cdots = N$

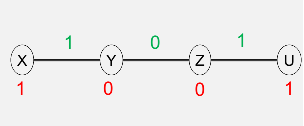
$R(X, Y) \wedge S(Y, Z) \wedge T(Z, U)$
$R = [N] \times [1], S = [1] \times [1], T = [1] \times [N].$



Examples

$|R| = |S| = \cdots = N$

$R(X, Y) \wedge S(Y, Z) \wedge T(Z, U)$
$R = [N] \times [1], S = [1] \times [1], T = [1] \times [N].$

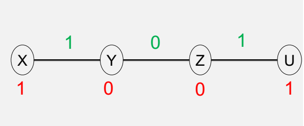


$R(X, Y) \wedge S(Y, Z) \wedge T(Z, U) \wedge K(U, V)$

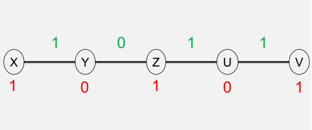
Examples

$|R| = |S| = \cdots = N$

$R(X, Y) \wedge S(Y, Z) \wedge T(Z, U)$
$R = [N] \times [1], S = [1] \times [1], T = [1] \times [N].$



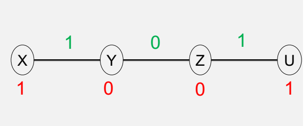
$R(X, Y) \wedge S(Y, Z) \wedge T(Z, U) \wedge K(U, V)$



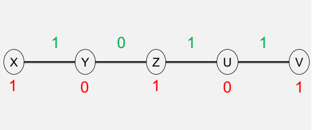
Examples

$|R| = |S| = \cdots = N$

$R(X, Y) \wedge S(Y, Z) \wedge T(Z, U)$
$R = [N] \times [1], S = [1] \times [1], T = [1] \times [N].$



$R(X, Y) \wedge S(Y, Z) \wedge T(Z, U) \wedge K(U, V)$
$R = T = [N] \times [1], S = K = [1] \times [N]$



Summary of the AGM Bound

- Upper / lower bound: fractional edge cover / vertex packing.
- Their equality follows from strong duality.
- The worst-case instance of the AGM bound is a Product Database.
- Full CQs only. Otherwise, ignore non-head variables.

Limitation of AGM: only cardinalities. Next week: extensions to other stats.

Atserias, A., Grohe, M., and Marx, D. (2013).
Size bounds and query plans for relational joins.
SIAM J. Comput., 42(4):1737–1767.

Friedgut, E. (2004).
Hypergraphs, entropy, and inequalities.
Am. Math. Mon., 111(9):749–760.