CSE599d: Advanced Query Processing

Lecture 15: The AGM Bound

Dan Suciu

University of Washington

Announcements

- HW3 is not yet graded

- HW4 is posted and due on Friday, March 13

Today’s Agenda

- AGM: the upper bound (review)
- AGM: the lower bound
- Extensions: adding keys, adding projections

AGM Upper Bound

Fractional Edge Covers

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

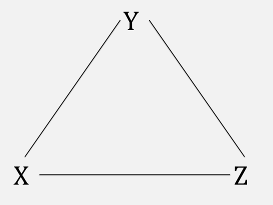
A fractional edge cover is $\boldsymbol{w} = (w_e)_{e \in E}$, $w_e \geq 0$:

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

Fractional Edge Covers

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

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



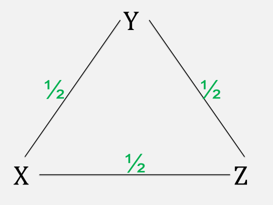
A fractional edge cover is $\boldsymbol{w} = (w_e)_{e \in E}, w_e \geq 0$:

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

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 Fractional Edge Covering Number

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

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

The AGM Bound: Uniform Cardinalities

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

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

[Atserias et al., 2013]

Simple Examples

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

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

Simple Examples

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

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

$$\boldsymbol{w} = (1,0,1,0,1), \qquad |Q| \leq N^3.$$

Simple Examples

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

5-cycle:

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

$$\boldsymbol{w} = (1,0,1,0,1), \qquad |Q| \leq N^3.$$

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

Simple Examples

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

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

$$\boldsymbol{w} = (1,0,1,0,1), \qquad |Q| \leq N^3.$$

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

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

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

Simple Examples

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

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

$$\boldsymbol{w} = (1,0,1,0,1), \qquad |Q| \leq N^3.$$

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

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

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

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 $(1,1,0,1)$, $\qquad AGM = N^3.$

The AGM Bound: Non-uniform Cardinalities

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

Theorem [Upper Bound] For every fractional edge cover $\boldsymbol{w}$:

$$\boxed{|Q| \leq |R_1|^{w_1} \cdots |R_m|^{w_m}}$$

[Atserias et al., 2013]

The AGM Bound: Non-uniform Cardinalities

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

Theorem [Upper Bound] For every fractional edge cover $\boldsymbol{w}$:

$$\boxed{|Q| \leq |R_1|^{w_1} \cdots |R_m|^{w_m}}$$

The AGM bound is the minimum over all fractional edge covers $\boldsymbol{w}$:

$$\boxed{AGM(Q) := \min_{\boldsymbol{w}} |R_1|^{w_1} \cdots |R_m|^{w_m}}$$

[Atserias et al., 2013]

The AGM Bound: Non-uniform Cardinalities

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

Theorem [Upper Bound] For every fractional edge cover $\boldsymbol{w}$:

$$\boxed{|Q| \leq |R_1|^{w_1} \cdots |R_m|^{w_m}}$$

The AGM bound is the minimum over all fractional edge covers $\boldsymbol{w}$:

$$\boxed{AGM(Q) := \min_{\boldsymbol{w}} |R_1|^{w_1} \cdots |R_m|^{w_m}}$$

When $|R_1| = |R_2| = \cdots = N$ then $AGM(Q) = N^{\min_{\boldsymbol{w}} \sum w_i} = N^{\rho^*}$

[Atserias et al., 2013]

Example

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

$$AGM(Q) = \min\begin{pmatrix} (|R|\cdot|S|\cdot|T|)^{1/2} \\ |R|\cdot|S| \\ |R|\cdot|T| \\ |S|\cdot|T| \end{pmatrix}$$

Example

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

Assuming T is the largest among R, S, T:

$$AGM(Q) = \begin{cases} (|R| \cdot |S| \cdot |T|)^{1/2} & \text{when } |T| \leq |R| \cdot |S| \\ |R| \cdot |S| & \text{otherwise} \end{cases}$$

Proof of the Upper Bound

There are two ways to prove the upper bound:

- Using a class of inequalities introduced by Friedgut. Reviewed next.
- Using entropic inequalities: Next week.

Review: Generalized Hölder's Inequality

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

If $u + v + w + \cdots \geq 1$ then:

$$\boxed{\sum_x a_x^u b_x^v c_x^w \cdots \leq \left(\sum_x a_x\right)^u \left(\sum_x b_x\right)^v \left(\sum_x c_x\right)^w \cdots}$$

Review: Generalized Hölder's Inequality

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

If $u + v + w + \cdots \geq 1$ then:

$$\boxed{\sum_x a_x^u b_x^v c_x^w \cdots \leq \left(\sum_x a_x\right)^u \left(\sum_x b_x\right)^v \left(\sum_x c_x\right)^w \cdots}$$

Condition $u + v + w + \cdots \geq 1$ says that $(u, v, \cdots)$ is a fractional edge cover of the hypergraph (V, E), where:

$$V := \{X\}, E := \{\{X\}, \{X\}, \ldots\}$$

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

For any fractional edge cover $w_1, \ldots, w_m$ of G, the following holds:

$$\boxed{\sum_{\boldsymbol{x} \in [N]^n} \left(\prod_{j=1,m} a_{j,\pi_{e_j}(\boldsymbol{x})}^{w_j} \right) \leq \prod_{j=1,m} \left(\sum_{\boldsymbol{y}_j \in [N]^{e_j}} a_{j,\boldsymbol{y}_j} \right)^{w_j}}$$

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

For any fractional edge cover $w_1, \ldots, w_m$ of G, the following holds:

$$\boxed{\sum_{\boldsymbol{x} \in [N]^n} \left(\prod_{j=1,m} a_{j,\pi_{e_j}(\boldsymbol{x})}^{w_j} \right) \leq \prod_{j=1,m} \left(\sum_{\boldsymbol{y}_j \in [N]^{e_j}} a_{j,\boldsymbol{y}_j} \right)^{w_j}}$$

E.g. $\sum_{x,y,z \in [N]} a_{xy}^{\frac{1}{2}} b_{yz}^{\frac{1}{2}} c_{zx}^{\frac{1}{2}} \leq \left(\sum_{x,y \in [N]} a_{xy}\right)^{\frac{1}{2}} \cdot \left(\sum_{y,z \in [N]} b_{yz}\right)^{\frac{1}{2}} \cdot \left(\sum_{z,x \in [N]} c_{zx}\right)^{\frac{1}{2}}$

Proof of Friedgut's Inequality

$$\boxed{\sum_{\boldsymbol{x}\in[N]^n}\left(\prod_{j=1,m} a^{w_j}_{j,\pi_{e_j}(\boldsymbol{x})}\right) \leq \prod_{j=1,m}\left(\sum_{\boldsymbol{y}_j\in[N]^{e_j}} a_{j,\boldsymbol{y}_j}\right)^{w_j}}$$

Induction:

Proof of Friedgut's Inequality

$$\boxed{\sum_{\boldsymbol{x}\in[N]^n}\left(\prod_{j=1,m} a_{j,\pi_{e_j}(\boldsymbol{x})}^{w_j}\right) \leq \prod_{j=1,m}\left(\sum_{\boldsymbol{y}_j\in[N]^{e_j}} a_{j,\boldsymbol{y}_j}\right)^{w_j}}$$

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

Proof of Friedgut's Inequality

$$\boxed{\sum_{\boldsymbol{x}\in[N]^n}\left(\prod_{j=1,m} a^{w_j}_{j,\pi_{e_j}(\boldsymbol{x})}\right) \leq \prod_{j=1,m}\left(\sum_{\boldsymbol{y}_j\in[N]^{e_j}} a_{j,\boldsymbol{y}_j}\right)^{w_j}}$$

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

$$\boxed{\sum_{\boldsymbol{x}\in[N]^n}\left(\prod_{j=1,m} a_{j,\pi_{e_j}(\boldsymbol{x})}^{w_j}\right) \leq \prod_{j=1,m}\left(\sum_{\boldsymbol{y}_j\in[N]^{e_j}} a_{j,\boldsymbol{y}_j}\right)^{w_j}}$$

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 $e_1 - \{X_1\}, \ldots, e_m - \{X_1\}$; $\boldsymbol{w}$ still a cover.

Proof of Friedgut's Inequality

$$\boxed{\sum_{\boldsymbol{x}\in[N]^n}\left(\prod_{j=1,m} a_{j,\pi_{e_j}(\boldsymbol{x})}^{w_j}\right) \leq \prod_{j=1,m}\left(\sum_{\boldsymbol{y}_j\in[N]^{e_j}} a_{j,\boldsymbol{y}_j}\right)^{w_j}}$$

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 $e_1 - \{X_1\}, \ldots, e_m - \{X_1\}$; $\boldsymbol{w}$ still a cover.

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

Proof of Friedgut's Inequality

$$\boxed{\sum_{\mathbf{x}\in[N]^n}\left(\prod_{j=1,m} a_{j,\pi_{e_j}(\mathbf{x})}^{w_j}\right) \leq \prod_{j=1,m}\left(\sum_{\mathbf{y}_j\in[N]^{e_j}} a_{j,\mathbf{y}_j}\right)^{w_j}}$$

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 $e_1 - \{X_1\}, \ldots, e_m - \{X_1\}$; $\mathbf{w}$ still a cover.

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

Proof of Friedgut's Inequality

$$\boxed{\sum_{\boldsymbol{x}\in[N]^n}\left(\prod_{j=1,m} a_{j,\pi_{e_j}(\boldsymbol{x})}^{w_j}\right) \leq \prod_{j=1,m}\left(\sum_{\boldsymbol{y}_j\in[N]^{e_j}} a_{j,\boldsymbol{y}_j}\right)^{w_j}}$$

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 $e_1 - \{X_1\}, \ldots, e_m - \{X_1\}$; $\boldsymbol{w}$ still a cover.

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

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

Proof of Friedgut's Inequality

$$\boxed{\sum_{\boldsymbol{x}\in[N]^n}\left(\prod_{j=1,m} a_{j,\pi_{e_j}(\boldsymbol{x})}^{w_j}\right) \leq \prod_{j=1,m}\left(\sum_{\boldsymbol{y}_j\in[N]^{e_j}} a_{j,\boldsymbol{y}_j}\right)^{w_j}}$$

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 $e_1 - \{X_1\}, \ldots, e_m - \{X_1\}$; $\boldsymbol{w}$ still a cover.

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

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

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

Moving Exponents from Left to Right

Cauchy-Schwartz becomes a special case:

$$\sum_x a_x^{\frac{1}{2}} b_x^{\frac{1}{2}} \leq \Big(\sum_x a_x\Big)^{\frac{1}{2}} \Big(\sum_x b_x\Big)^{\frac{1}{2}}$$

$$\sum_x a_x b_x \leq \Big(\sum_x a_x^2\Big)^{\frac{1}{2}} \Big(\sum_x b_x^2\Big)^{\frac{1}{2}}$$

Moving Exponents from Left to Right

Cauchy-Schwartz becomes a special case:

$$\sum_x a_x^{\frac{1}{2}} b_x^{\frac{1}{2}} \leq \Big(\sum_x a_x\Big)^{\frac{1}{2}} \Big(\sum_x b_x\Big)^{\frac{1}{2}}$$

$$\sum_x a_x b_x \leq \Big(\sum_x a_x^2\Big)^{\frac{1}{2}} \Big(\sum_x b_x^2\Big)^{\frac{1}{2}}$$

For a triangle, we obtain:

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

$$\mathtt{Tr}(\boldsymbol{a} \cdot \boldsymbol{b} \cdot \boldsymbol{c}) = \sum_{xyz} a_{xy} b_{yz} c_{zx} \leq \Big(\sum_{xy} a_{xy}^2\Big)^{\frac{1}{2}} \cdot \Big(\sum_{yz} b_{yz}^2\Big)^{\frac{1}{2}} \cdot \Big(\sum_{zx} c_{zx}^2\Big)^{\frac{1}{2}}$$

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 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 $\boldsymbol{D}$.

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

Statement of the Lower Bound: Uniform Cardinalities

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

Theorem [Lower Bound]
For every N, there exists an instance $\boldsymbol{D} = (R_1^D, \ldots, R_m^D)$ s.t.

$$|R_1^D|, \cdots, |R_m^D| \leq N \qquad \text{and} \qquad Q(\boldsymbol{D}) \geq \frac{1}{2^{|\boldsymbol{X}|}} N^{\rho^*}$$

Statement of the Lower Bound: Uniform Cardinalities

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

Theorem [Lower Bound]
For every N, there exists an instance $\boldsymbol{D} = (R_1^D, \ldots, R_m^D)$ s.t.

$$|R_1^D|, \cdots, |R_m^D| \leq N \qquad \text{and} \qquad Q(\boldsymbol{D}) \geq \frac{1}{2^{|\boldsymbol{X}|}} N^{\rho^*}$$

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

The case of non-uniform cardinalities is similar; will discuss shortly.

Background: The Duality Theorem for Linear Programming

Primal Linear Program:

Minimize $\mathbf{c}^T\mathbf{x}$

where $\mathbf{x} \geq 0$ satisfies:

$$\mathbf{A}\mathbf{x} \geq \mathbf{b}$$

Background: The Duality Theorem for Linear Programming

Primal Linear Program:

Minimize $\boldsymbol{c}^T\boldsymbol{x}$

where $\boldsymbol{x} \geq 0$ satisfies:

$$\boldsymbol{A}\boldsymbol{x} \geq \boldsymbol{b}$$

Dual Linear Program:

Maximize $\boldsymbol{y}^T\boldsymbol{b}$

where $\boldsymbol{y} \geq \mathbf{0}$ satisifes:

$$\boldsymbol{y}^T\boldsymbol{A} \leq \boldsymbol{c}^T$$

Background: The Duality Theorem for Linear Programming

Primal Linear Program:
Minimize $\boldsymbol{c}^T\boldsymbol{x}$
where $\boldsymbol{x} \geq 0$ satisfies:
$$\boldsymbol{A}\boldsymbol{x} \geq \boldsymbol{b}$$

Dual Linear Program:
Maximize $\boldsymbol{y}^T\boldsymbol{b}$
where $\boldsymbol{y} \geq \boldsymbol{0}$ satisifes:
$$\boldsymbol{y}^T\boldsymbol{A} \leq \boldsymbol{c}^T$$

Weak Duality Theorem If $\boldsymbol{x}, \boldsymbol{y}$ are feasible solutions then $\boldsymbol{y}^T\boldsymbol{b} \leq \boldsymbol{c}^T\boldsymbol{x}$

proof in class

Background: The Duality Theorem for Linear Programming

Primal Linear Program:
Minimize $\boldsymbol{c}^T\boldsymbol{x}$
where $\boldsymbol{x} \geq 0$ satisfies:
$$\boldsymbol{A}\boldsymbol{x} \geq \boldsymbol{b}$$

Dual Linear Program:
Maximize $\boldsymbol{y}^T\boldsymbol{b}$
where $\boldsymbol{y} \geq \boldsymbol{0}$ satisifes:
$$\boldsymbol{y}^T\boldsymbol{A} \leq \boldsymbol{c}^T$$

Weak Duality Theorem If $\boldsymbol{x}, \boldsymbol{y}$ are feasible solutions then $\boldsymbol{y}^T\boldsymbol{b} \leq \boldsymbol{c}^T\boldsymbol{x}$

proof in class

Strong Duality Theorem If $\boldsymbol{x}^*, \boldsymbol{y}^*$ are optimal solutions then $(\boldsymbol{y}^*)^T\boldsymbol{b} = \boldsymbol{c}^T\boldsymbol{x}^*$

Fractional Edge Cover v.s. Fractional Vertex Packing

$G = (V, E)$

Fractional Edge Cover:

Minimize $\sum_{e \in E} w_e$

where $\boldsymbol{w} = (w_e)_{e \in E} \geq \mathbf{0}$ satisfies:

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

Fractional Edge Cover v.s. Fractional Vertex Packing

$G = (V, E)$

Fractional Edge Cover:

Minimize $\sum_{e\in E} w_e$

where $\boldsymbol{w} = (w_e)_{e\in E} \geq \mathbf{0}$ satisfies:

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

Fractional Vertex Packing:

Maximize $\sum_{x\in V} v_x$

where $\boldsymbol{v} = (v_x)_{x\in V} \geq \mathbf{0}$ satisifes:

$$\forall e \in E : \sum_{x:x\in e} v_x \leq 1$$

Fractional Edge Cover v.s. Fractional Vertex Packing

$G = (V, E)$

Fractional Edge Cover:

Minimize $\sum_{e \in E} w_e$

where $\boldsymbol{w} = (w_e)_{e \in E} \geq \mathbf{0}$ satisfies:

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

Fractional Vertex Packing:

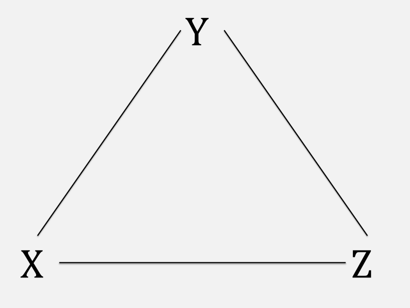
Maximize $\sum_{x \in V} v_x$

where $\boldsymbol{v} = (v_x)_{x \in V} \geq \mathbf{0}$ satisifes:

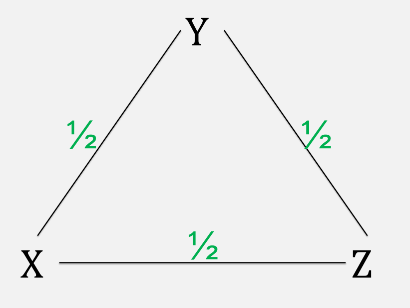
$$\forall e \in E : \sum_{x : x \in e} v_x \leq 1$$

The optimal values are equal: $\sum_e w_e^* = \sum_x v_x^*$.

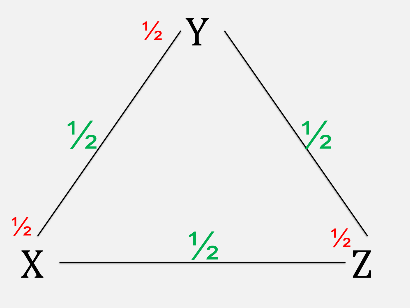
Example



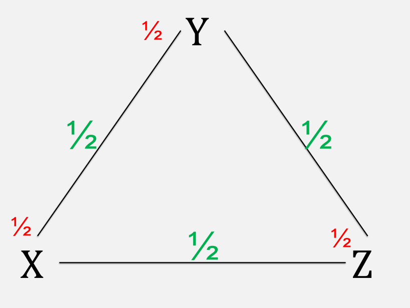
Example



Example



Example



The optimal values are equal: $\sum_e w_e^* = \sum_x v_x^* = 3/2$.

We use $\boldsymbol{v}^*$ to construct a worst-case instance for the query Q.

The Worst-Case Instance By Example

$R(X,Y) \wedge S(Y,Z) \wedge T(Z,X), \qquad |R|, |S|, |T| \leq N, \qquad AGM(Q) = N^{3/2}$

The Worst-Case Instance By Example

$R(X,Y) \wedge S(Y,Z) \wedge T(Z,X)$, $|R|,|S|,|T| \leq N$, $AGM(Q) = N^{3/2}$

Fractional Edge Cover:
Minimize $w_R + w_S + w_T$

where $\boldsymbol{w} \geq \mathbf{0}$ satisfies:

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

The Worst-Case Instance By Example

$R(X,Y) \wedge S(Y,Z) \wedge T(Z,X)$, $|R|,|S|,|T| \leq N$, $AGM(Q) = N^{3/2}$

Fractional Edge Cover:
Minimize $w_R + w_S + w_T$

where $\boldsymbol{w} \geq \mathbf{0}$ satisfies:

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

Fractional Vertex Packing:
Maximize $v_X + v_Y + v_Z$

where $\boldsymbol{v} \geq 0$ satisfies:

$$\begin{array}{lllll} R: & v_X+ & v_Y & & \leq 1 \\ S: & & v_Y+ & v_Z & \leq 1 \\ T: & v_X+ & & v_Z & \leq 1 \end{array}$$

The Worst-Case Instance By Example

$$R(X,Y) \wedge S(Y,Z) \wedge T(Z,X), \qquad |R|,|S|,|T| \leq N, \qquad AGM(Q) = N^{3/2}$$

Fractional Edge Cover:

Minimize $w_R + w_S + w_T$

where $\mathbf{w} \geq \mathbf{0}$ satisfies:

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

Fractional Vertex Packing:

Maximize $v_X + v_Y + v_Z$

where $\mathbf{v} \geq 0$ satisfies:

$$\begin{array}{lllll} R: & v_X+ & v_Y & & \leq 1 \\ S: & & v_Y+ & v_Z & \leq 1 \\ T: & v_X+ & & v_Z & \leq 1 \end{array}$$

By strong duality: $\rho^* = w_R^* + w_S^* + w_T^* = v_x^* + v_y^* + v_z^*$.

The Worst-Case Instance By Example

$R(X,Y) \wedge S(Y,Z) \wedge T(Z,X), \qquad |R|, |S|, |T| \leq N, \qquad AGM(Q) = N^{3/2}$

Fractional Edge Cover:
Minimize $w_R + w_S + w_T$

where $\boldsymbol{w} \geq \mathbf{0}$ satisfies:

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

Fractional Vertex Packing:
Maximize $v_X + v_Y + v_Z$

where $\boldsymbol{v} \geq 0$ satisfies:

$$\begin{array}{lllll} R: & v_X + & v_Y & & \leq 1 \\ S: & & v_Y + & v_Z & \leq 1 \\ T: & v_X + & & v_Z & \leq 1 \end{array}$$

By strong duality: $\rho^* = w_R^* + w_S^* + w_T^* = v_x^* + v_y^* + v_z^*$.

Set $s_x := \lfloor N^{v_x^*} \rfloor$, $s_y := \lfloor N^{v_y^*} \rfloor$, $s_z := \lfloor N^{v_z^*} \rfloor$.

The Worst-Case Instance By Example

$R(X, Y) \wedge S(Y, Z) \wedge T(Z, X), \qquad |R|, |S|, |T| \leq N, \qquad AGM(Q) = N^{3/2}$

Fractional Edge Cover:
Minimize $w_R + w_S + w_T$

where $\boldsymbol{w} \geq \mathbf{0}$ satisfies:

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

Fractional Vertex Packing:
Maximize $v_X + v_Y + v_Z$

where $\boldsymbol{v} \geq 0$ satisfies:

$$\begin{array}{lllll} R: & v_X + & v_Y & & \leq 1 \\ S: & & v_Y + & v_Z & \leq 1 \\ T: & v_X + & & v_Z & \leq 1 \end{array}$$

By strong duality: $\rho^* = w_R^* + w_S^* + w_T^* = v_x^* + v_y^* + v_z^*$.

Set $s_x := \lfloor N^{v_x^*} \rfloor$, $s_y := \lfloor N^{v_y^*} \rfloor$, $s_z := \lfloor N^{v_z^*} \rfloor$.

$\boldsymbol{D}$ is a cross product: $\boxed{R^D := [s_x] \times [s_y],\ S^D := [s_y] \times [s_z],\ T^D := [s_z] \times [s_x]}$

The Worst-Case Instance By Example

$$R(X,Y) \wedge S(Y,Z) \wedge T(Z,X), \qquad |R|,|S|,|T| \leq N, \qquad AGM(Q) = N^{3/2}$$

Fractional Edge Cover:
Minimize $w_R + w_S + w_T$

where $\boldsymbol{w} \geq \mathbf{0}$ satisfies:

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

Fractional Vertex Packing:
Maximize $v_X + v_Y + v_Z$

where $\boldsymbol{v} \geq 0$ satisfies:

$$\begin{array}{llllll} R: & v_X + & v_Y & & \leq 1 \\ S: & & v_Y + & v_Z & \leq 1 \\ T: & v_X + & & v_Z & \leq 1 \end{array}$$

By strong duality: $\rho^* = w_R^* + w_S^* + w_T^* = v_x^* + v_y^* + v_z^*$.

Set $s_x := \lfloor N^{v_x^*} \rfloor$, $s_y := \lfloor N^{v_y^*} \rfloor$, $s_z := \lfloor N^{v_z^*} \rfloor$.

$\boldsymbol{D}$ is a cross product: $\boxed{R^D := [s_x] \times [s_y],\ S^D := [s_y] \times [s_z],\ T^D := [s_z] \times [s_x]}$

$$|Q(\boldsymbol{D})| = s_x \cdot s_y \cdot s_z \geq \frac{1}{8} N^{v_x^* + v_y^* + v_z^*} = \frac{1}{8} N^{\rho^*}$$

Proof of the Lower Bound

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

Theorem [Lower Bound]
For every N, there exists an instance $\boldsymbol{D} = (R_1^D, \ldots, R_m^D)$ s.t.

$$|R_1^D|, \cdots, |R_m^D| \leq N \qquad \text{and} \qquad Q(\boldsymbol{D}) \geq \frac{1}{2^{|\boldsymbol{X}|}} N^{\rho^*}$$

Proof of the Lower Bound

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

Theorem [Lower Bound]
For every N, there exists an instance $\boldsymbol{D} = (R_1^D, \ldots, R_m^D)$ s.t.

$$|R_1^D|, \cdots, |R_m^D| \le N \qquad \text{and} \qquad Q(\boldsymbol{D}) \ge \frac{1}{2^{|\boldsymbol{X}|}} N^{\rho^*}$$

Proof By strong duality: $\rho^* = \sum_j w_j^* = \sum_i v_i^*$.

Proof of the Lower Bound

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

Theorem [Lower Bound]
For every N, there exists an instance $\boldsymbol{D} = (R_1^D, \ldots, R_m^D)$ s.t.

$$|R_1^D|, \cdots, |R_m^D| \leq N \qquad \text{and} \qquad Q(\boldsymbol{D}) \geq \frac{1}{2^{|\boldsymbol{X}|}} N^{\rho^*}$$

Proof By strong duality: $\rho^* = \sum_j w_j^* = \sum_i v_i^*$.

For every variable X_i set $s_i := \lfloor N^{v_i^*} \rfloor$

Proof of the Lower Bound

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

Theorem [Lower Bound]
For every N, there exists an instance $\boldsymbol{D} = (R_1^D, \ldots, R_m^D)$ s.t.

$$|R_1^D|, \cdots, |R_m^D| \leq N \qquad \text{and} \qquad Q(\boldsymbol{D}) \geq \frac{1}{2^{|\boldsymbol{X}|}} N^{\rho^*}$$

Proof By strong duality: $\rho^* = \sum_j w_j^* = \sum_i v_i^*$.

For every variable X_i set $s_i := \lfloor N^{v_i^*} \rfloor$

Define $\boldsymbol{D}$ the cross product database: $\boxed{R_j^D := \times_{i : X_i \in \texttt{Vars}(R_j)} [s_i]}$

$$|Q(\boldsymbol{D})| = \prod_{i=1,n} s_i \;\; \geq \;\; \frac{1}{2^n} \prod_i N^{v_i^*} = \frac{1}{2^n} N^{\rho^*}$$

Statement of the Lower Bound: Non-Uniform Cardinalities

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

Theorem [Lower Bound]
For all numbers $N_1, \ldots, N_m$ there exists $\boldsymbol{D} = (R_1^D, \ldots, R_m^D)$ s.t:

$$|R_1^D| \leq N_1, \cdots, |R_m^D| \leq N_m \qquad \text{and} \qquad Q(\boldsymbol{D}) \geq \frac{1}{2^{|\boldsymbol{X}|}} \min_{\boldsymbol{w}} N_j{}^{w_j}$$

Statement of the Lower Bound: Non-Uniform Cardinalities

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

Theorem [Lower Bound]
For all numbers $N_1, \ldots, N_m$ there exists $\boldsymbol{D} = (R_1^D, \ldots, R_m^D)$ s.t:

$$|R_1^D| \leq N_1, \cdots, |R_m^D| \leq N_m \qquad \text{and} \qquad Q(\boldsymbol{D}) \geq \frac{1}{2^{|\boldsymbol{X}|}} \min_{\boldsymbol{w}} N_j^{w_j}$$

$$\boxed{\text{The AGM bound } \min_{\boldsymbol{w}} N_j^{w_j} \text{ is tight}}$$

Proof of the AGM Lower Bound: Non-uniform Cardinalities

By example:

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

Proof of the AGM Lower Bound: Non-uniform Cardinalities

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: Non-uniform Cardinalities

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:

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

Dual program:

Maximize

$$v_X + v_Y + v_Z$$

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

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

Proof of the AGM Lower Bound: Non-uniform Cardinalities

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:

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

Dual program:
Maximize

$$v_X + v_Y + v_Z$$

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

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

By strong duality: $w_R^* + w_S^* + w_T^* = v_x^* + v_y^* + v_z^*$.
Worst-case instance is the product database:
$s_x \overset{\text{def}}{=} \lfloor 2^{v_X^*} \rfloor, s_y \overset{\text{def}}{=} \lfloor 2^{v_Y^*} \rfloor, s_z \overset{\text{def}}{=} \lfloor 2^{v_Z^*} \rfloor, R^D := [s_x] \times [s_y], S^D := [s_y] \times [s_z], T^D := [s_z] \times [s_x]$

Proof of the AGM Lower Bound: Non-uniform Cardinalities

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:

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

Dual program:

Maximize

$$v_X + v_Y + v_Z$$

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

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

By strong duality: $w_R^* + w_S^* + w_T^* = v_x^* + v_y^* + v_z^*$.

Worst-case instance is the product database:

$s_x \overset{\text{def}}{=} \lfloor 2^{v_X^*} \rfloor, s_y \overset{\text{def}}{=} \lfloor 2^{v_Y^*} \rfloor, s_z \overset{\text{def}}{=} \lfloor 2^{v_Z^*} \rfloor, R^D := [s_x] \times [s_y], S^D := [s_y] \times [s_z], T^D := [s_z] \times [s_x]$

$$|Q(\boldsymbol{D})| = s_x s_y s_z \geq \frac{1}{8} 2^{v_X^* + v_Y^* + v_Z^*}$$

Proof of the AGM Lower Bound: Non-uniform Cardinalities

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:

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

Dual program:

Maximize

$$v_X + v_Y + v_Z$$

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

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

By strong duality: $w_R^* + w_S^* + w_T^* = v_x^* + v_y^* + v_z^*$.

Worst-case instance is the product database:

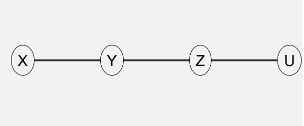
$s_x \stackrel{\text{def}}{=} \lfloor 2^{v_X^*} \rfloor, s_y \stackrel{\text{def}}{=} \lfloor 2^{v_Y^*} \rfloor, s_z \stackrel{\text{def}}{=} \lfloor 2^{v_Z^*} \rfloor, R^D := [s_x] \times [s_y], S^D := [s_y] \times [s_z], T^D := [s_z] \times [s_x]$

$$|Q(\boldsymbol{D})| = s_x s_y s_z \geq \frac{1}{8} 2^{v_X^* + v_Y^* + v_Z^*} = \frac{1}{8} 2^{w_R^* + w_S^* + w_T^*} = \frac{1}{8} AGM(Q)$$

Examples

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

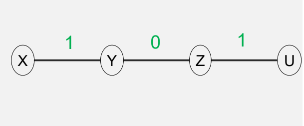
$R(X,Y) \wedge S(Y,Z) \wedge T(Z,U)$



Examples

$|R| = |S| = \cdots = {\color{red}N}$

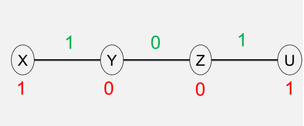
$R(X, Y) \wedge S(Y, Z) \wedge T(Z, U)$



Examples

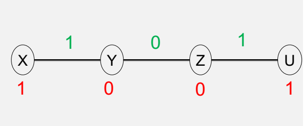
$|R| = |S| = \cdots = N$

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

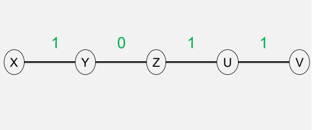


Examples

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



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

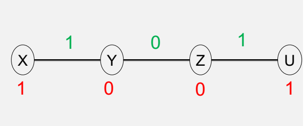


$R(X, Y) \wedge S(Y, Z) \wedge T(Z, U) \wedge K(U, V)$

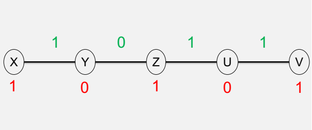
Examples

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

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



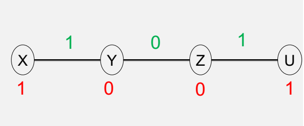
$R(X, Y) \wedge S(Y, Z) \wedge T(Z, U) \wedge K(U, V)$



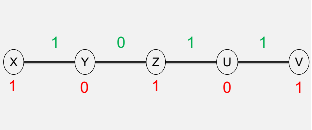
Examples

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

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



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



Extensions of the AGM Bound

Extensions

- Use more stats: functional dependencies, domain cardinalities, and others.

- Beyond full queries: projections (same as Group-By).

Domain Sizes

If $|\mathsf{Dom}(X)|$ is known, simply add it to the hypergraph.

Domain Sizes

If $|\mathsf{Dom}(X)|$ is known, simply add it to the hypergraph.

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

We know $|R|$, $|S|$, $|T|$, $|\mathsf{Dom}(X)|$.

Domain Sizes

If $|\mathsf{Dom}(X)|$ is known, simply add it to the hypergraph.

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

We know $|R|, |S|, |T|, |\mathsf{Dom}(X)|$.

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

Domain Sizes

If $|\mathrm{Dom}(X)|$ is known, simply add it to the hypergraph.

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

We know $|R|, |S|, |T|, |\mathrm{Dom}(X)|$.

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

Fractional edge covers: $(0, \frac{1}{2}, \frac{1}{2}, \frac{1}{2})$ and $(1, 0, 1, 0)$

$$\boxed{AGM(Q) = \min((|R| \cdot |S| \cdot |T|)^{1/2}, |A| \cdot |S|)}$$

Simple Functional Dependencies

Given FDs, $|Q| \ll AGM(Q)$.

E.g. $R(X,Y) \wedge S(Y,Z)$: N^2 becomes N when $Y \rightarrow Z$.

Simple Functional Dependencies

Given FDs, $|Q| \ll AGM(Q)$.

E.g. $R(X,Y) \wedge S(Y,Z)$: N^2 becomes N when $Y \rightarrow Z$.

$\boldsymbol{U} \rightarrow \boldsymbol{V}$ is simple if $|\boldsymbol{U}| = 1$.

Simple Functional Dependencies

Given FDs, $|Q| \ll AGM(Q)$.

E.g. $R(X, Y) \wedge S(Y, Z)$: N^2 becomes N when $Y \rightarrow Z$.

$\boldsymbol{U} \rightarrow \boldsymbol{V}$ is simple if $|\boldsymbol{U}| = 1$.

Method [Khamis et al., 2016]:

- Expand Q to Q^+ by replacing each atom $R(\boldsymbol{Y})$ with $R'(\boldsymbol{Y}^+)$.
- Return $AGM(Q^+)$.
- This bound is tight. Proof: very useful exercise.

Example

$Q(X,Y,Z) = R(X,Y) \wedge S(Y,Z) \wedge T(Z,X)$
Fractional edge covers: $(1,1,0), (1,0,1), (0,1,1), (1/2,1/2,1/2)$

$$\boxed{|Q| \leq \min(|R| \cdot |S|, |R| \cdot |T|, |S| \cdot |T|, \sqrt{|R| \cdot |S| \cdot |T|})}$$

Example

$Q(X, Y, Z) = R(X, Y) \wedge S(Y, Z) \wedge T(Z, X)$
Fractional edge covers: $(1, 1, 0), (1, 0, 1), (0, 1, 1), (1/2, 1/2, 1/2)$

$$\boxed{|Q| \leq \min(|R| \cdot |S|, |R| \cdot |T|, |S| \cdot |T|, \sqrt{|R| \cdot |S| \cdot |T|})}$$

Assume that $S.Y$ is a key: $Y \rightarrow Z$

Example

$Q(X, Y, Z) = R(X, Y) \wedge S(Y, Z) \wedge T(Z, X)$
Fractional edge covers: $(1, 1, 0), (1, 0, 1), (0, 1, 1), (1/2, 1/2, 1/2)$

$$\boxed{|Q| \leq \min(|R| \cdot |S|, |R| \cdot |T|, |S| \cdot |T|, \sqrt{|R| \cdot |S| \cdot |T|})}$$

Assume that $S.Y$ is a key: $Y \rightarrow Z$
$Q^+(X, Y, Z) = R'(X, Y, Z) \wedge S(Y, Z) \wedge T(Z, X)$

Example

$Q(X, Y, Z) = R(X, Y) \wedge S(Y, Z) \wedge T(Z, X)$
Fractional edge covers: $(1, 1, 0), (1, 0, 1), (0, 1, 1), (1/2, 1/2, 1/2)$

$$\boxed{|Q| \leq \min(|R| \cdot |S|, |R| \cdot |T|, |S| \cdot |T|, \sqrt{|R| \cdot |S| \cdot |T|})}$$

Assume that $S.Y$ is a key: $Y \rightarrow Z$
$Q^+(X, Y, Z) = R'(X, Y, Z) \wedge S(Y, Z) \wedge T(Z, X)$
Fractional edge covers: $(1, 0, 0), (0, 1, 1)$

$$\boxed{|Q| \leq \min(|R|, |S| \cdot |T|)}$$

Discussion

The expansion procedure is very easy, but limited only to simple FDs:

$AGM(Q^+)$ is always an upper bound on Q's output, but may not be tight.

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. Later: 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.

Khamis, M. A., Ngo, H. Q., and Suciu, D. (2016).
Computing join queries with functional dependencies.
In Milo, T. and Tan, W., editors, *Proceedings of the 35th ACM SIGMOD-SIGACT-SIGAI Symposium on Principles of Database Systems, PODS 2016, San Francisco, CA, USA, June 26 - July 01, 2016*, pages 327–342. ACM.