Quantile Regression

Ruibin Xi

Computation of Quantile Regression (1)

- We have

$$\begin{aligned}\rho_\tau(y_i - x_i^T b) &= \rho_\tau(e_i) \\ &= \rho_\tau(u_i - v_i) \\ &= \tau u_i + (\tau - 1)(-v_i) \\ &= \tau u_i + (1-\tau) v_i\end{aligned}$$

- Minimizing $\sum_{i=1}^{n} \rho_\tau(y_i - x_i^T b)$ is equivalent to minimizing

$$\sum_{i=1}^{n} [\tau u_i + (1-\tau) v_i]$$

with constraints

$$u_i \geq 0, \quad v_i \geq 0, \quad y_i - x_i^T b = u_i - v_i$$

Computation of Quantile Regression (2)

- If we write

$$u = (u_1, \cdots, u_n)^T \quad v = (v_1, \cdots, v_n)^T$$

- the above problem can be written as

$$\min \left\{ \tau e_n^\top u + (1-\tau) e_n^\top v \,|\, y - Xb = u - v,\ b \in \mathbb{R}^p,\ (u, v) \in \mathbb{R}_+^{2n} \right\}$$

- Note that this is a linear programming (LP) problem

Linear Programming (1)

- Standard maximum problem
  - Optimize a linear function subject to linear inequalities

(P) Maximize $\boldsymbol{c}^T\boldsymbol{x} = c_1x_1 + \cdots + c_nx_n$

Subject to

$$
\begin{aligned}
a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &\le b_1 \\
a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &\le b_2 \\
&\vdots \\
a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &\le b_m
\end{aligned}
\qquad (\text{or } \boldsymbol{A}\boldsymbol{x} \le \boldsymbol{b})
$$

$$x_1 \ge 0, x_2 \ge 0, \ldots, x_n \ge 0 \qquad (\text{or } \boldsymbol{x} \ge \boldsymbol{0}).$$

Linear Programming (2)

- Standard minimum problem
  - Optimize a linear function subject to linear inequalities

(P) Minimize $\boldsymbol{y}^T\boldsymbol{b} = y_1b_1 + \cdots + y_mb_m$

Subject to

$$
\begin{aligned}
y_1a_{11} + y_2a_{21} + \cdots + y_ma_{m1} &\geq c_1 \\
y_1a_{12} + y_2a_{22} + \cdots + y_ma_{m2} &\geq c_2 \\
&\vdots \\
y_1a_{1n} + y_2a_{2n} + \cdots + y_ma_{mn} &\geq c_n
\end{aligned}
\qquad (\text{or } \boldsymbol{y}^T\boldsymbol{A} \geq \boldsymbol{c}^T)
$$

$$y_1 \geq 0, y_2 \geq 0, \ldots, y_m \geq 0 \qquad (\text{or } \boldsymbol{y} \geq \boldsymbol{0})$$

Brewery Example (1)

Small brewery produces ale and beer.

- Production limited by scarce resources: corn, hops, barley malt.
- Recipes for ale and beer require different proportions of resources.

Beverage	Corn (pounds)	Hops (ounces)	Malt (pounds)	Profit ($)
Ale (barrel)	5	4	35	13
Beer (barrel)	15	4	20	23
constraint	480	160	1190	

How can brewer maximize profits?

- Devote all resources to ale: 34 barrels of ale ⇒ $442
- Devote all resources to beer: 32 barrels of beer ⇒ $736
- 7.5 barrels of ale, 29.5 barrels of beer ⇒ $776
- 12 barrels of ale, 28 barrels of beer ⇒ $800

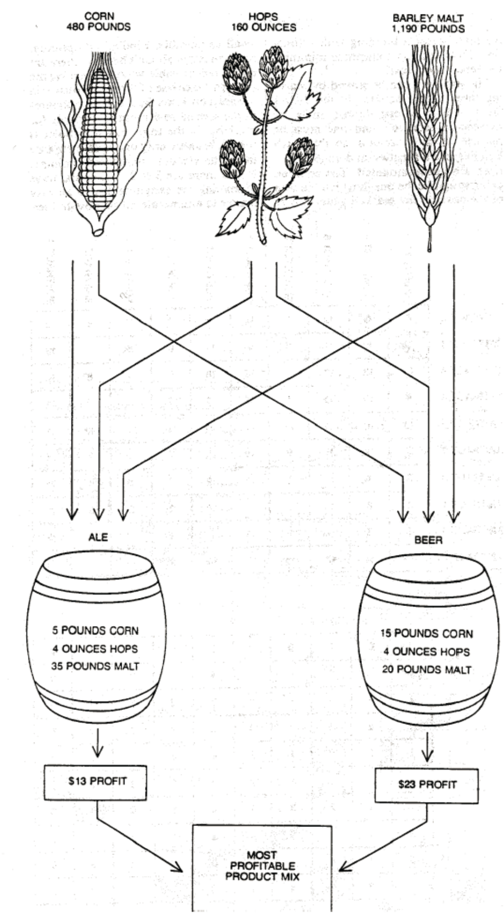
Brewery Example (2)

objective function

	Ale	Beer			
max	$13A$	$+\ 23B$			Profit
s. t.	$5A$	$+\ 15B$	$\leq$	480	Corn
	$4A$	$+\ 4B$	$\leq$	160	Hops
	$35A$	$+\ 20B$	$\leq$	1190	Malt
	A	$,\ B$	$\geq$	0	

constraint

decision variable



Standard LP forms

- Standard maximum (equality) form

$$\max\{c^T x : \; Ax = b, x \geq 0\}$$

- Standard minimum (equality) form

$$\min\{b^T y : \; A^T y = c, y \geq 0\}$$

Brewery Example: converting to standard LP form

Original input.

$$
\begin{array}{lrcrcr}
\max & 13A & + & 23B & & \\
\text{s. t.} & 5A & + & 15B & \leq & 480 \\
 & 4A & + & 4B & \leq & 160 \\
 & 35A & + & 20B & \leq & 1190 \\
 & A & , & B & \geq & 0
\end{array}
$$

Standard form.

- Add slack variable for each inequality.
- Now a 5-dimensional problem.

$$
\begin{array}{lrcrcccccccr}
\max & 13A & + & 23B & & & & & & & & \\
\text{s. t.} & 5A & + & 15B & + & S_C & & & & & = & 480 \\
 & 4A & + & 4B & & & + & S_H & & & = & 160 \\
 & 35A & + & 20B & & & & & + & S_M & = & 1190 \\
 & A & , & B & , & S_C & , & S_H & , & S_M & \geq & 0
\end{array}
$$

Equivalent form

Easy to convert variants to standard form.

$$
\begin{array}{llrcl}
\text{(P)} & \max & c^T x & & \\
& \text{s. t.} & Ax & = & b \\
& & x & \geq & 0
\end{array}
$$

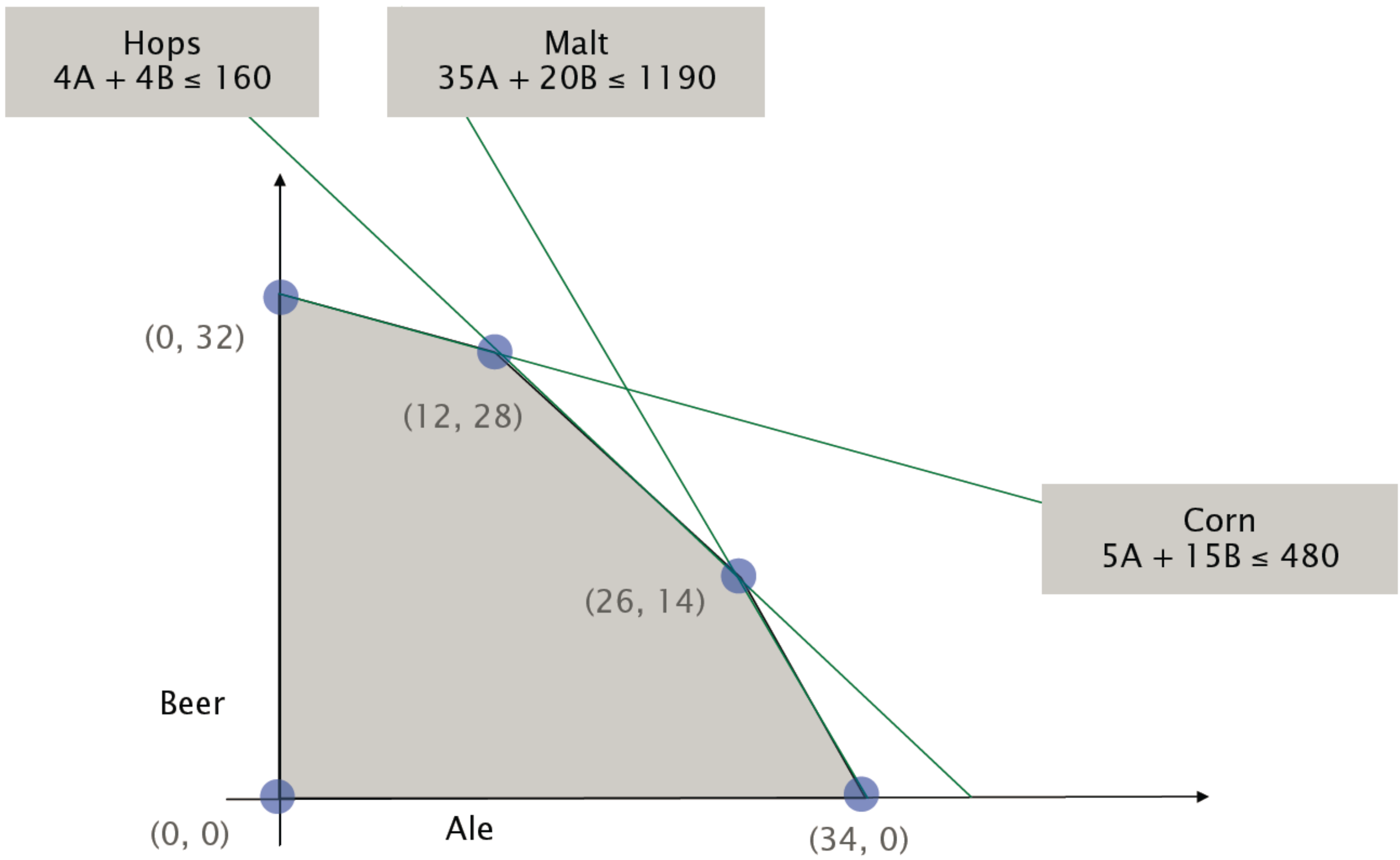
Less than to equality. $x + 2y - 3z \leq 17 \;\Rightarrow\; x + 2y - 3z + s = 17,\ s \geq 0$

Greater than to equality. $x + 2y - 3z \geq 17 \;\Rightarrow\; x + 2y - 3z - s = 17,\ s \geq 0$

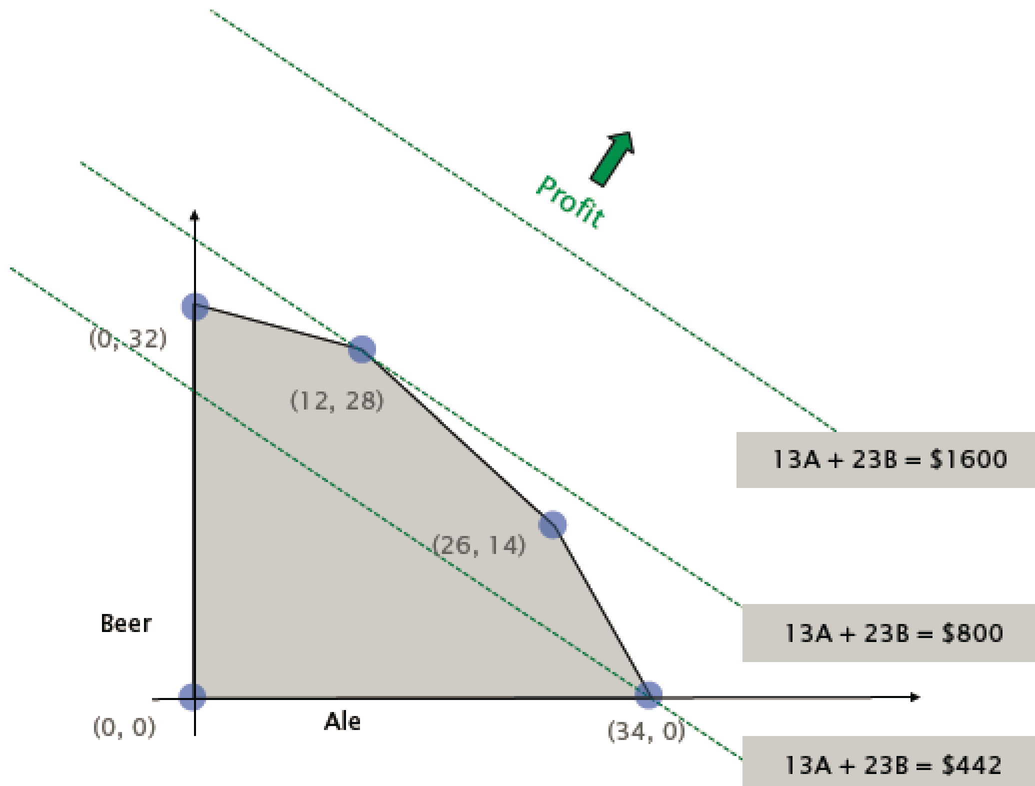
Min to max. $\min\ x + 2y - 3z \;\Rightarrow\; \max\ -x - 2y + 3z$

Unrestricted to nonnegative. x unrestricted $\;\Rightarrow\; x = x^+ - x^-,\ x^+ \geq 0, x^- \geq 0$

Brewery problem: feasible region

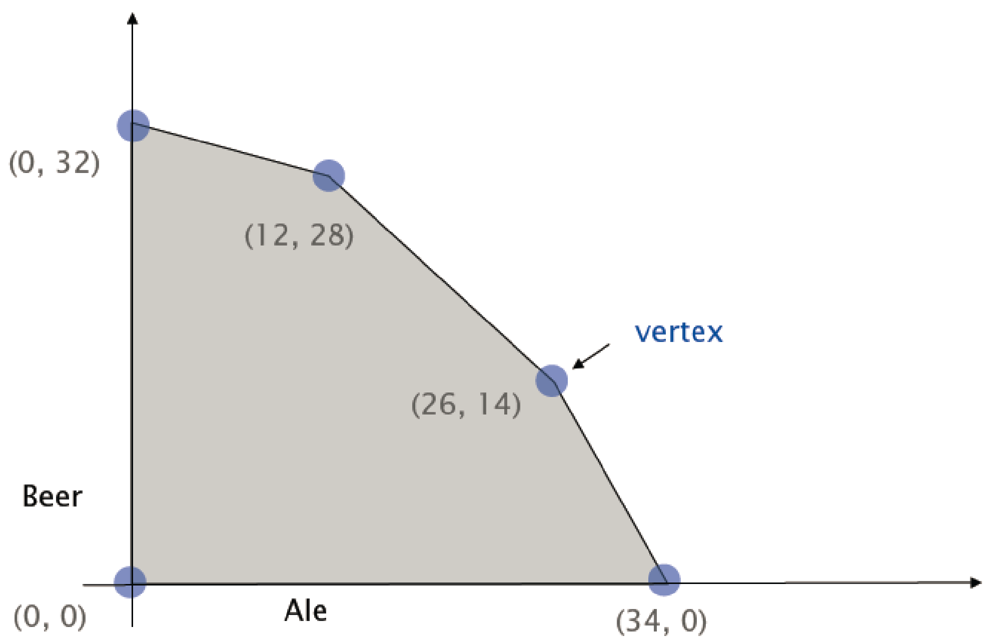


Brewery problem: Objective function



Brewery problem: Geometry

Brewery problem observation. Regardless of objective function coefficients, an optimal solution occurs at a vertex.



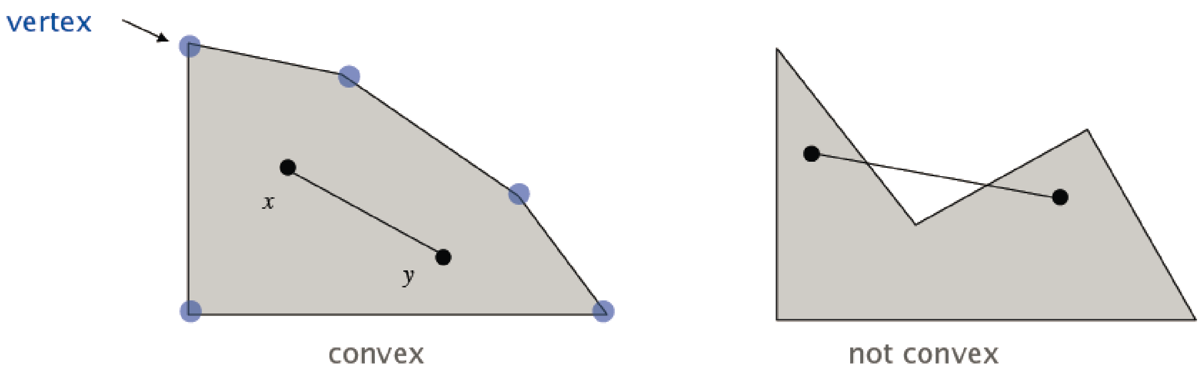
Convexity

Convex set. If two points x and y are in the set, then so is $\lambda x + (1-\lambda) y$ for $0 \leq \lambda \leq 1$.

convex combination

not a vertex iff $\exists\, d \neq 0$ s.t. $x \pm d$ in set

Vertex. A point x in the set that can't be written as a strict convex combination of two distinct points in the set.



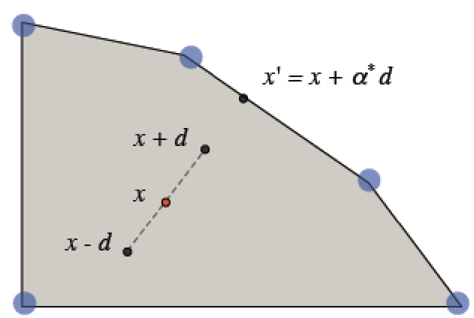
Observation. LP feasible region is a convex set.

Vertices (1)

Theorem. If there exists an optimal solution to (P), then there exists one that is a vertex.

$$\text{(P)} \quad \max \; c^T x \quad \text{s. t.} \; Ax = b, \; x \geq 0$$

Intuition. If x is not a vertex, move in a non-decreasing direction until you reach a boundary. Repeat.



Vertices (2)

Pf.

- Suppose x is an optimal solution that is not a vertex.
- There exist direction $d \neq 0$ such that $x \pm d \in P$.
- $A\,d = 0$ because $A(x \pm d) = b$.
- Assume $c^T d \leq 0$ (by taking either d or $-d$).
- Consider $x + \lambda\, d,\ \lambda > 0$:

Case 1. [there exists j such that $d_j < 0$]

- Increase λ to λ^* until first new component of $x + \lambda\, d$ hits 0.
- $x + \lambda^* d$ is feasible since $A(x + \lambda^* d) = Ax = b$ and $x + \lambda^* y \geq 0$.
- $x + \lambda^* d$ has one more zero component than x.
- $c^T x' = c^T (x + \lambda^* d) = c^T x + \lambda^* c^T d \leq c^T x$.

$d_k = 0$ whenever $x_k = 0$ because $x \pm d \in P$

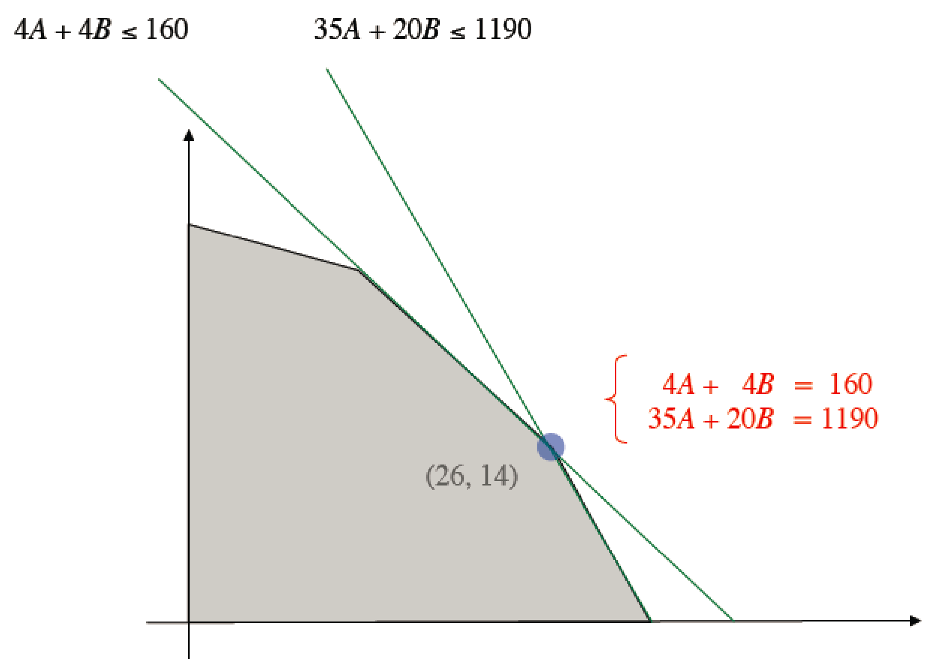
Case 2. [$d_j \geq 0$ for all j]

- $x + \lambda d$ is feasible for all $\lambda \geq 0$ since $A(x + \lambda d) = b$ and $x + \lambda d \geq x \geq 0$.
- As $\lambda \to \infty$, $c^T(x + \lambda d) \to -\infty$ because $c^T d < 0$. ▪

if $c^T d = 0$, choose d so that case 1 applies

Basic feasible solutions(1)

Intuition. A vertex in $\Re^m$ is uniquely specified by m linearly independent equations.



Basic feasible solutions(2)

Theorem. Let $P = \{ x : Ax = b, x \geq 0 \}$. For $x \in P$, define $B = \{ j : x_j > 0 \}$. Then x is a vertex iff A_B has linearly independent columns.

Notation. Let $B =$ set of column indices. Define A_B to be the subset of columns of A indexed by B.

Ex.

$$A = \begin{bmatrix} 2 & 1 & 3 & 0 \\ 7 & 3 & 2 & 1 \\ 0 & 0 & 0 & 5 \end{bmatrix}, \quad b = \begin{bmatrix} 7 \\ 16 \\ 0 \end{bmatrix}$$

$$x = \begin{bmatrix} 2 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \quad B = \{1, 3\}, \quad A_B = \begin{bmatrix} 2 & 3 \\ 7 & 2 \\ 0 & 0 \end{bmatrix}$$

Basic feasible solutions(3)

Theorem. Let $P = \{ x : Ax = b, x \geq 0 \}$. For $x \in P$, define $B = \{ j : x_j > 0 \}$. Then x is a vertex iff A_B has linearly independent columns.

Notation. Let B = set of column indices. Define A_B to be the subset of columns of A indexed by B.

Ex.

$$A = \begin{bmatrix} 2 & 1 & 3 & 0 \\ 7 & 3 & 2 & 1 \\ 0 & 0 & 0 & 5 \end{bmatrix}, \quad b = \begin{bmatrix} 7 \\ 16 \\ 0 \end{bmatrix}$$

$$x = \begin{bmatrix} 2 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \quad B = \{1, 3\}, \quad A_B = \begin{bmatrix} 2 & 3 \\ 7 & 2 \\ 0 & 0 \end{bmatrix}$$

Basic feasible solutions(4)

Theorem. Let $P = \{ x : Ax = b, x \geq 0 \}$. For $x \in P$, define $B = \{ j : x_j > 0 \}$. Then x is a vertex iff A_B has linearly independent columns.

Pf. $\Leftarrow$

- Assume x is not a vertex.
- There exist direction $d \neq 0$ such that $x \pm d \in P$.
- $A\,d = 0$ because $A(x \pm d) = b$.
- Define $B' = \{ j : d_j \neq 0 \}$.
- $A_{B'}$ has linearly dependent columns since $d \neq 0$.
- Moreover, $d_j = 0$ whenever $x_j = 0$ because $x \pm d \geq 0$.
- Thus $B' \subseteq B$, so $A_{B'}$ is a submatrix of A_B.
- Therefore, A_B has linearly dependent columns.

Basic feasible solutions(5)

Theorem. Let $P = \{ x : Ax = b, x \geq 0 \}$. For $x \in P$, define $B = \{ j : x_j > 0 \}$. Then x is a vertex iff A_B has linearly independent columns.

Pf. $\Rightarrow$

- Assume A_B has linearly dependent columns.
- There exist $d \neq 0$ such that $A_B\, d = 0$.
- Extend d to $\Re^n$ by adding 0 components.
- Now, $A\, d = 0$ and $d_j = 0$ whenever $x_j = 0$.
- For sufficiently small λ, $x \pm \lambda\, d \in P \Rightarrow x$ is not a vertex. ▪

Basic feasible solutions(6)

Theorem. Given $P = \{ x : Ax = b, x \geq 0 \}$, x is a vertex iff there exists $B \subseteq \{ 1, \ldots, n \}$ such $|B| = m$ and:

- A_B is nonsingular.
- $x_B = A_B^{-1} b \geq 0$.
- $x_N = 0$.

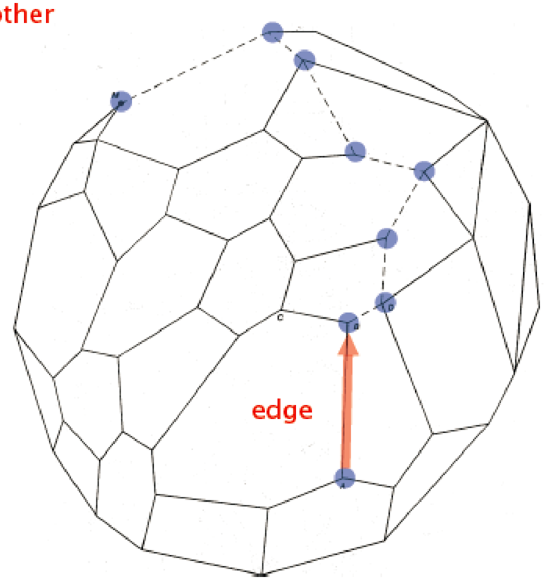
basic feasible solution

Assumption. $A \in \Re^{m \times n}$ has full row rank.

Simplex algorithm

Simplex algorithm. [George Dantzig 1947] Move from BFS to adjacent BFS, without decreasing objective function.

replace one basic variable with another



Greedy property. BFS optimal iff no adjacent BFS is better.
Challenge. Number of BFS can be exponential!