NEAR-OPTIMAL MAP INFERENCE FOR DETERMINANTAL POINT PROCESSES

Jennifer Gillenwater Alex Kulesza Ben Taskar
University of Pennsylvania

IMAGE SEARCH: "JAGUAR"

Relevance only:

...

Relevance + diversity:

...

TASK: SUBSET SELECTION

two goals

quality

diversity

MATCHED SUMMARIZATION

Task:

Given a set of documents, select a set of doc pairs such that:
1) the pairs are high-quality (docs within a pair are similar), and
2) the overall set of pairs is diverse.

MATCHED SUMMARIZATION

Ground set: All possible (old, new) pairs.

- **Old (topic = bailout): Let Detroit go bankrupt.**
- **New (topic = bailout) : I'm not willing to sit back and say 'Too bad for Michigan'.**

- **Old (topic = bailout): Let Detroit go bankrupt.**
- **New (topic = abortion): The right next step ... is to see Roe vs Wade overturned.**

Quality only:

- **Old (topic = bailout): Let Detroit go bankrupt.**
- **New (topic = bailout): I'm not willing to sit back and say 'Too bad for Michigan'.**

- **Old (topic = bailout): I think there is need for economic stimulus.**
- **Old (topic = bailout): I have never supported the President's recovery act.**

- **Old (topic = bailout): TARP ought to be ended.**
- **New (topic = bailout): TARP got paid back and it kept the financial system from collapsing.**

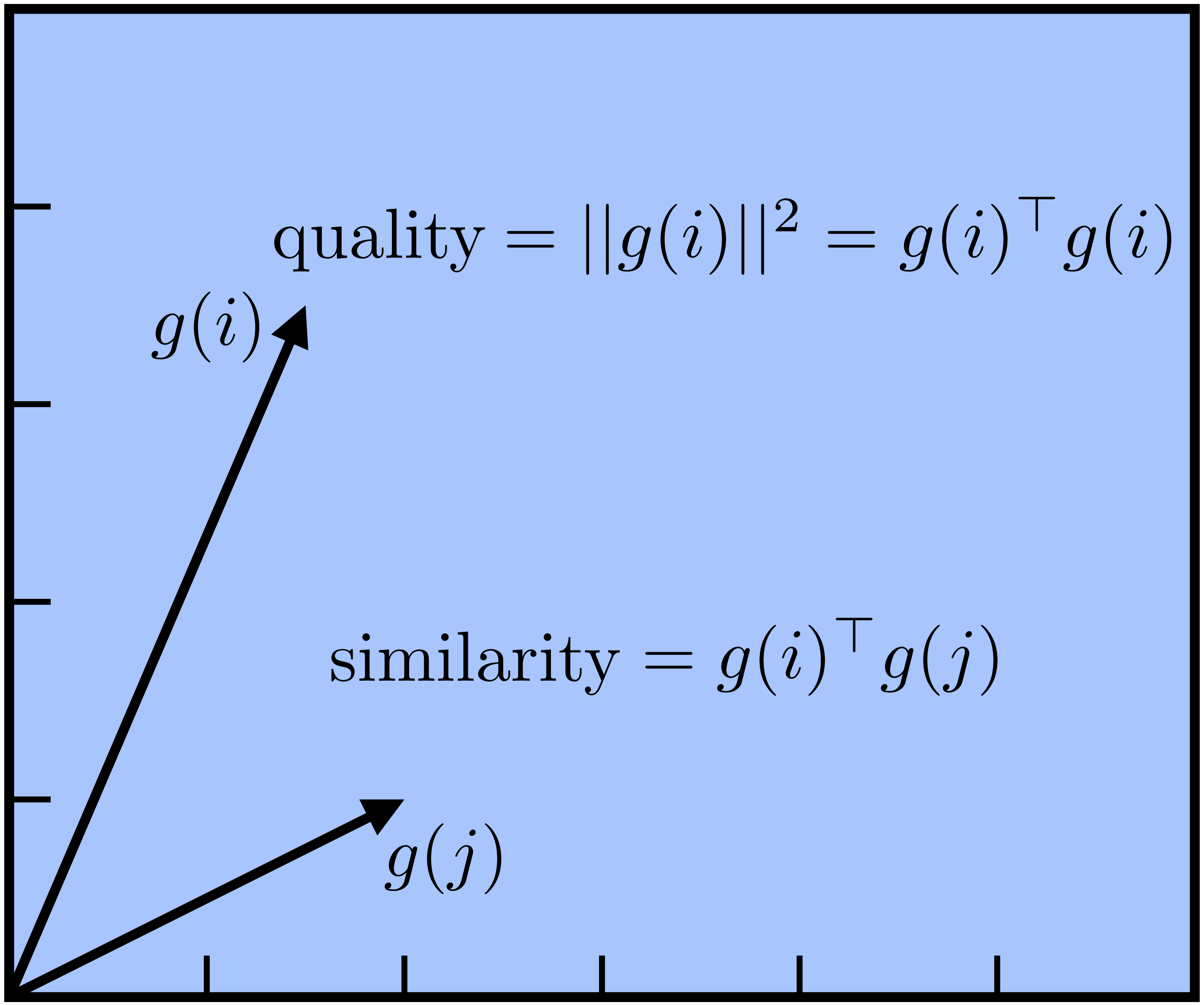
Quality + diversity:

- **Old (topic = bailout): Let Detroit go bankrupt.**
- **New (topic = bailout): I'm not willing to sit back and say 'Too bad for Michigan'.**

- **Old (topic = abortion): I will preserve and protect a woman's right to choose.**
- **New (topic = abortion): The right next step ... is to see Roe vs Wade overturned.**

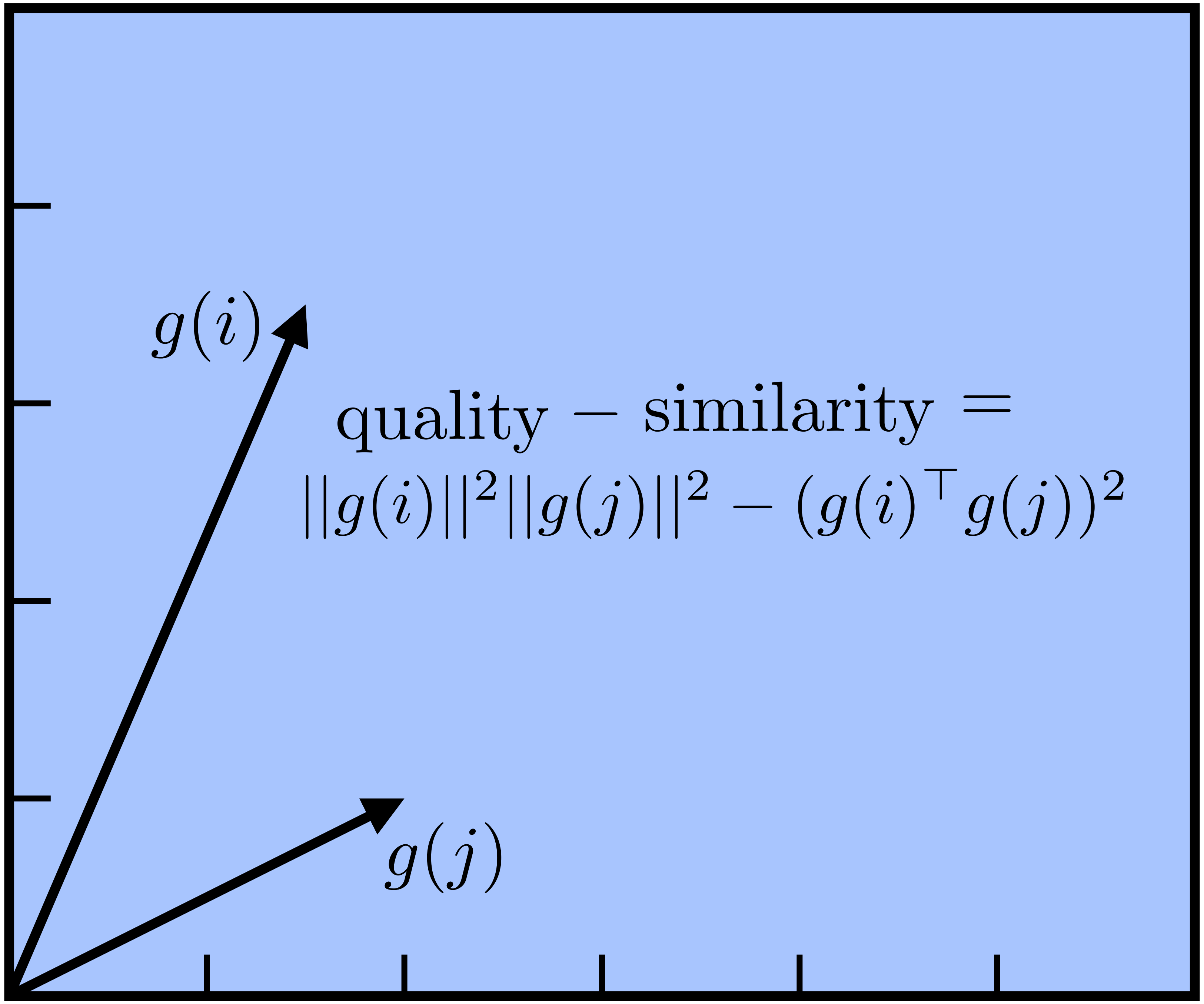
- **Old (topic = gun control): I just signed a major piece of legislation extending the ban on certain assault weapons.**
- **New (topic = gun control): I do not support any new legislation of an assault weapon ban nature.**

FORMALIZING



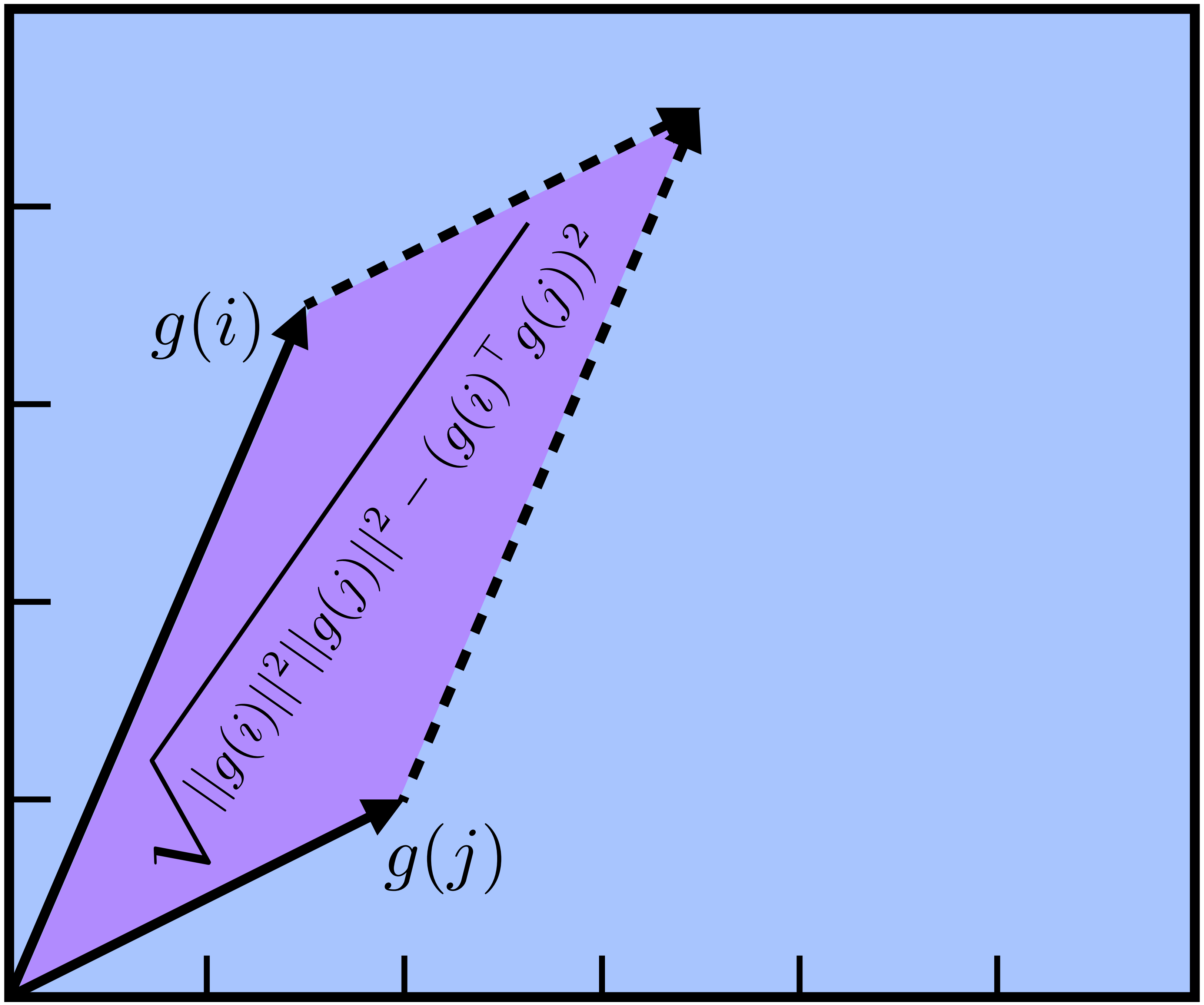
feature space

FORMALIZING



feature space

FORMALIZING



feature space

AREA AS A DET

$$||g(i)||^2 ||g(j)||^2 - (g(i)^\top g(j))^2$$

$$= \det \begin{pmatrix} ||g(i)||^2 & g(i)^\top g(j) \\ g(i)^\top g(j) & ||g(j)||^2 \end{pmatrix}$$

$$\begin{matrix} \text{——} g(1) \text{——} \\ \text{——} g(2) \text{——} \\ \vdots \\ \text{——} g(N) \text{——} \end{matrix}$$

AREA AS A DET

$$||g(i)||^2 ||g(j)||^2 - (g(i)^\top g(j))^2$$

$$= \det \begin{pmatrix} ||g(i)||^2 & g(i)^\top g(j) \\ g(i)^\top g(j) & ||g(j)||^2 \end{pmatrix}$$

$$= \det \left(\begin{bmatrix} \text{---} \; g(i) \; \text{---} \\ \text{---} \; g(j) \; \text{---} \end{bmatrix} \begin{bmatrix} | & | \\ g(i) & g(j) \\ | & | \end{bmatrix} \right)$$

VOLUME AS A DET

$$\text{“goodness” of } \{i,j\} = \text{quality \& diversity of } \{i,j\}$$

$$\propto \ \text{area}(\{i,j\})^2$$

$$\text{volume}_1 = \text{length}$$
$$\text{volume}_2 = \text{area}$$
$$\text{volume}_3 = \text{3D-volume}$$

$$|Y| = d \rightarrow \text{volume}_d(Y)^2 \ \propto \ \det((GG^\top)_Y)$$

$$\text{for positive semi-definite } L = GG^\top$$

$$\mathcal{P}(Y) \propto \det(L_Y)$$

VOLUME AS A DET

$$\text{“goodness” of } \{i,j\} = \text{quality \& diversity of } \{i,j\}$$
$$\propto \ \text{area}(\{i,j\})^2$$

Determinantal Point Process

$$|Y| = d \ \rightarrow \text{volume}_d(Y)^2 \ \propto \ \det((GG^\top)_Y)$$

$$\text{for positive semi-definite } L = GG^\top$$
$$\mathcal{P}(Y) \propto \det(L_Y)$$

VOLUME AS A DET

$$\text{“goodness” of } \{i,j\} = \text{quality \& diversity of } \{i,j\}$$
$$\propto \ \text{area}(\{i,j\})^2$$

DPP

$$|Y| = d \ \rightarrow \text{volume}_d(Y)^2 \ \propto \ \det((GG^\top)_Y)$$

$$\text{for positive semi-definite } L = GG^\top$$
$$\mathcal{P}(Y) \propto \det(L_Y)$$

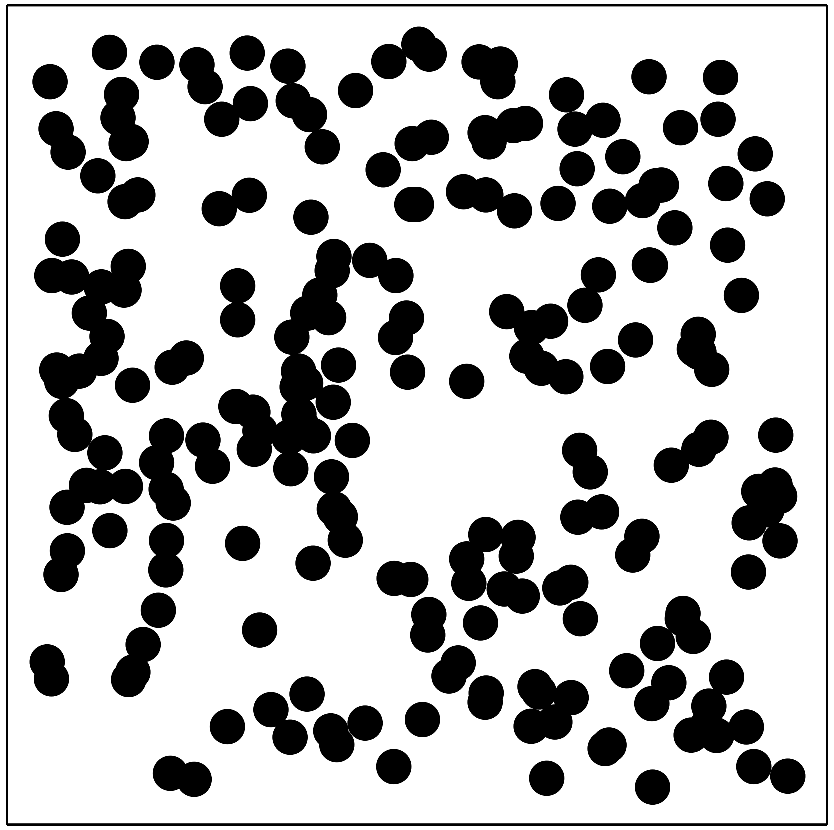
DPP INFERENCE

- Exact and efficient $O(N^3)$
  - normalization: $\sum_Y \det(L_Y) = \det(L + I)$
  - marginalization: $\mathcal{P}(A \subseteq Y)$
  - conditioning: $\mathcal{P}(A \mid B \subseteq Y)$
  - sampling: $Y \sim \mathcal{P}(Y) \propto \det(L_Y)$

DPP INFERENCE

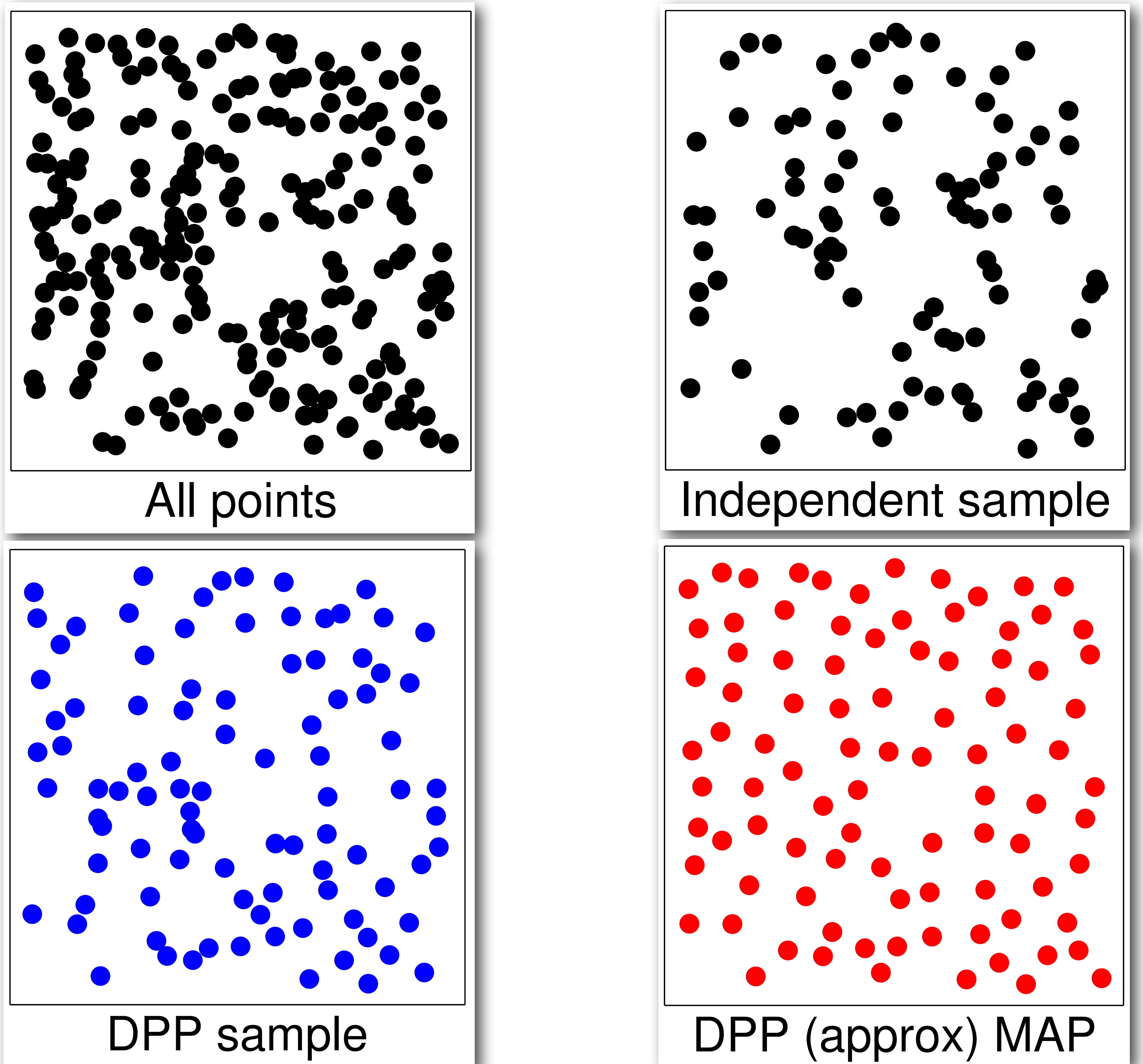
What about DPP MAP?

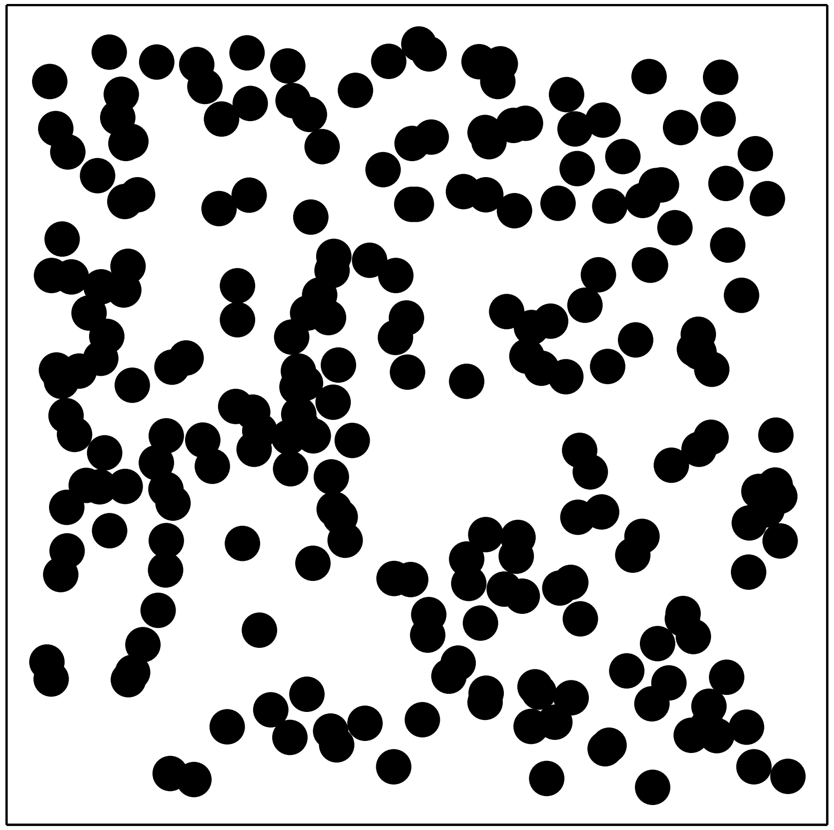
$$\arg\max_Y \det(L_Y)$$



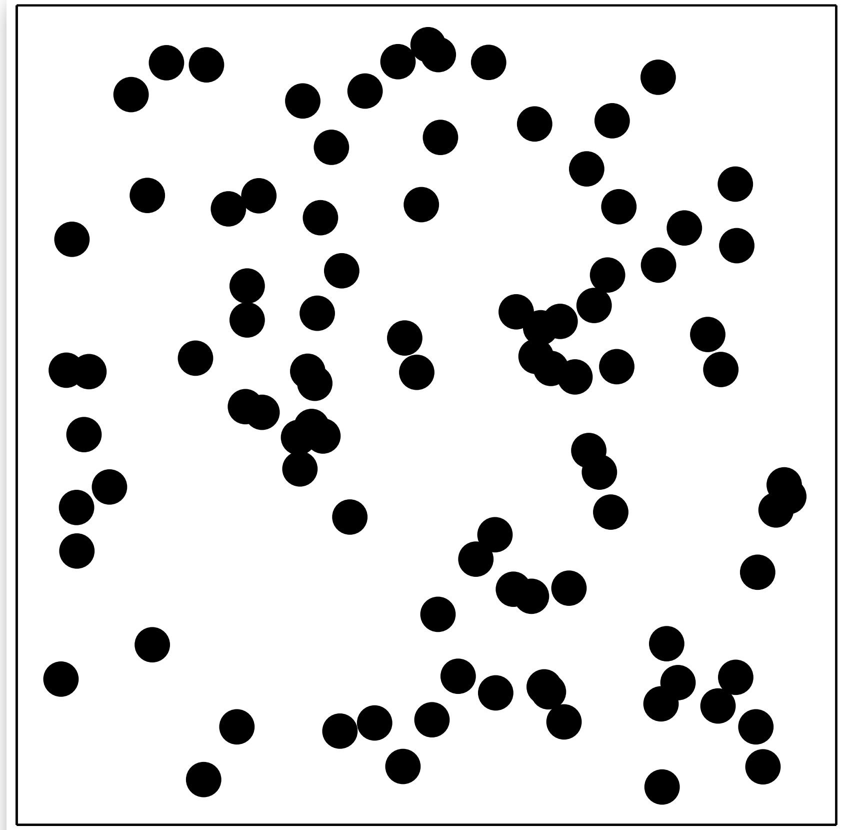
All points

$$g(i)^\top g(j) = L_{ij} = \exp(-||p_i - p_j||^2)$$

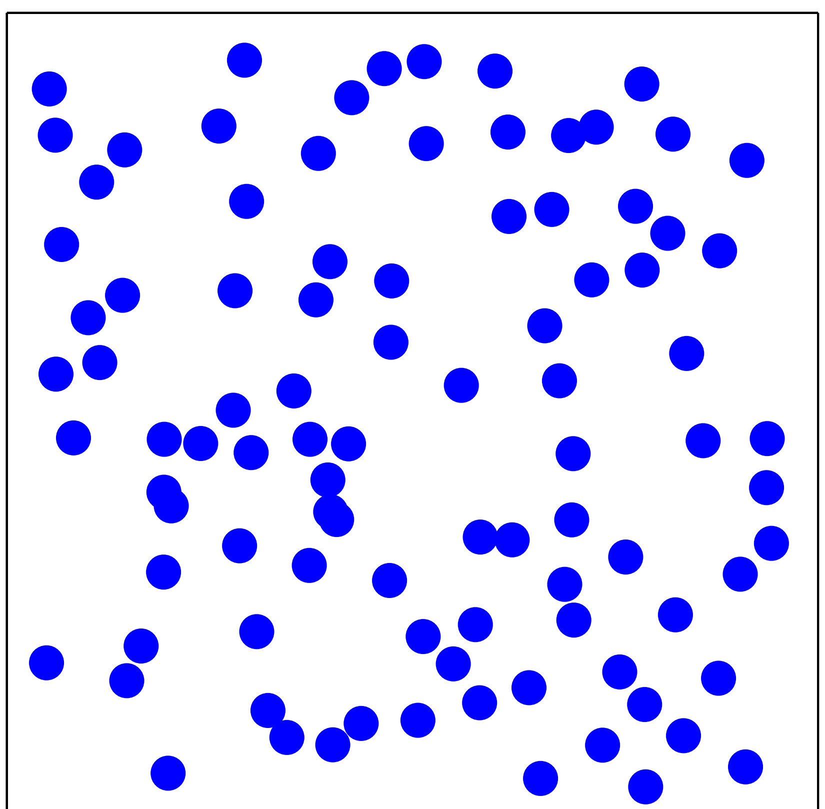

All points
Independent sample
DPP sample
DPP (approx) MAP



All points



Independent sample



DPP sample

DPP (approx) MAP

SUBMODULARITY TO THE RESCUE

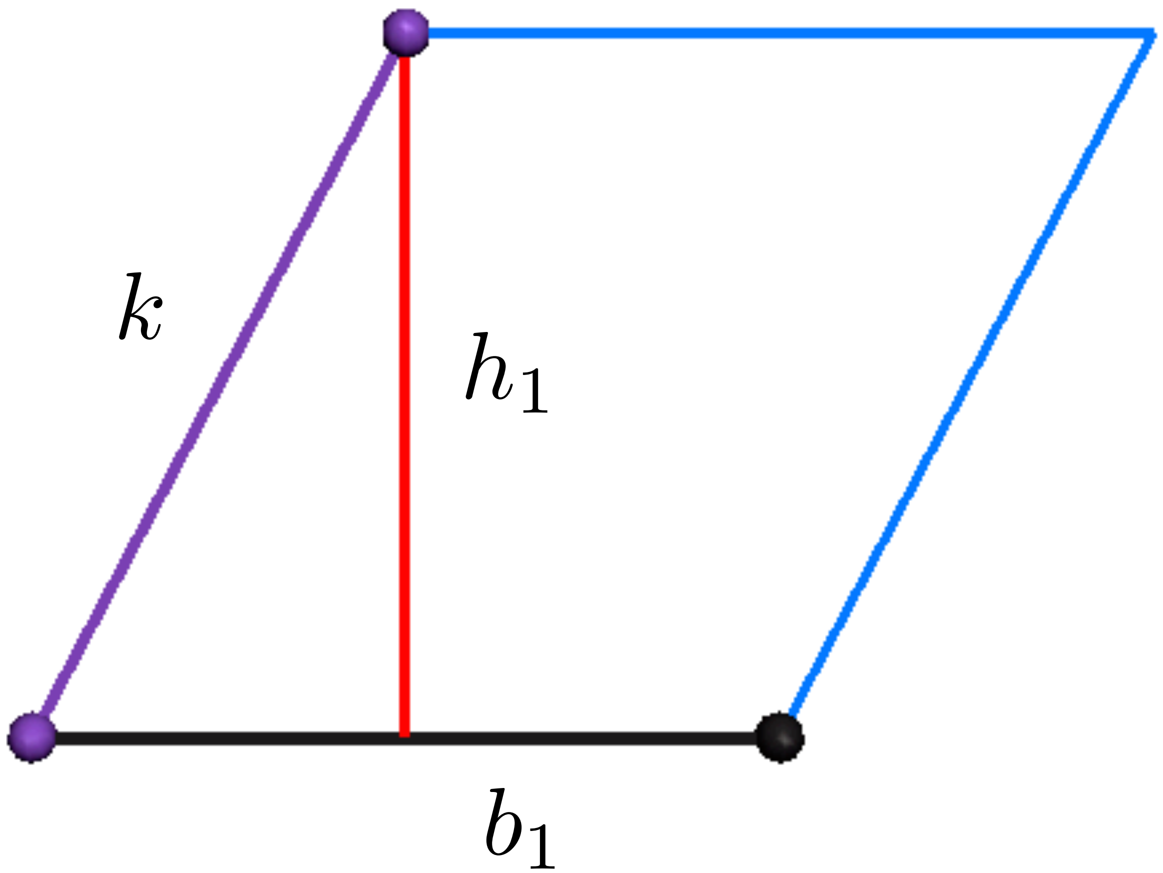
$$f(Y) = \det(L_Y) \text{ is log-submodular}$$

Diminishing returns:

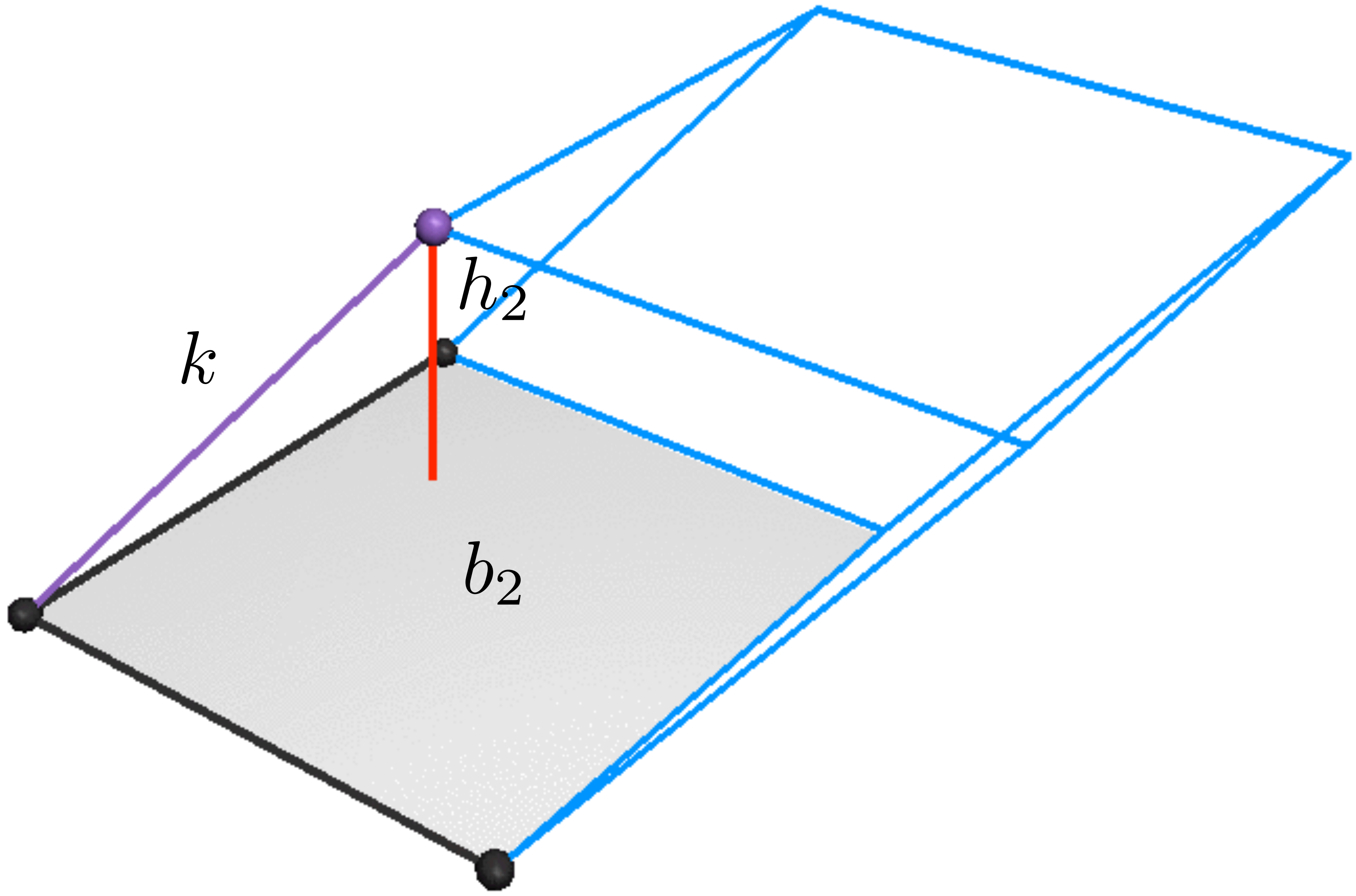
$$\frac{f(Y \cup \{k\})}{f(Y)} \leq \frac{f(X \cup \{k\})}{f(X)}$$

$$X \subseteq Y,\ k \notin Y$$

$$\frac{\mathrm{vol}(X \cup \{k\})}{\mathrm{vol}(X)} = \frac{b_1 h_1}{b_1} = h_1$$

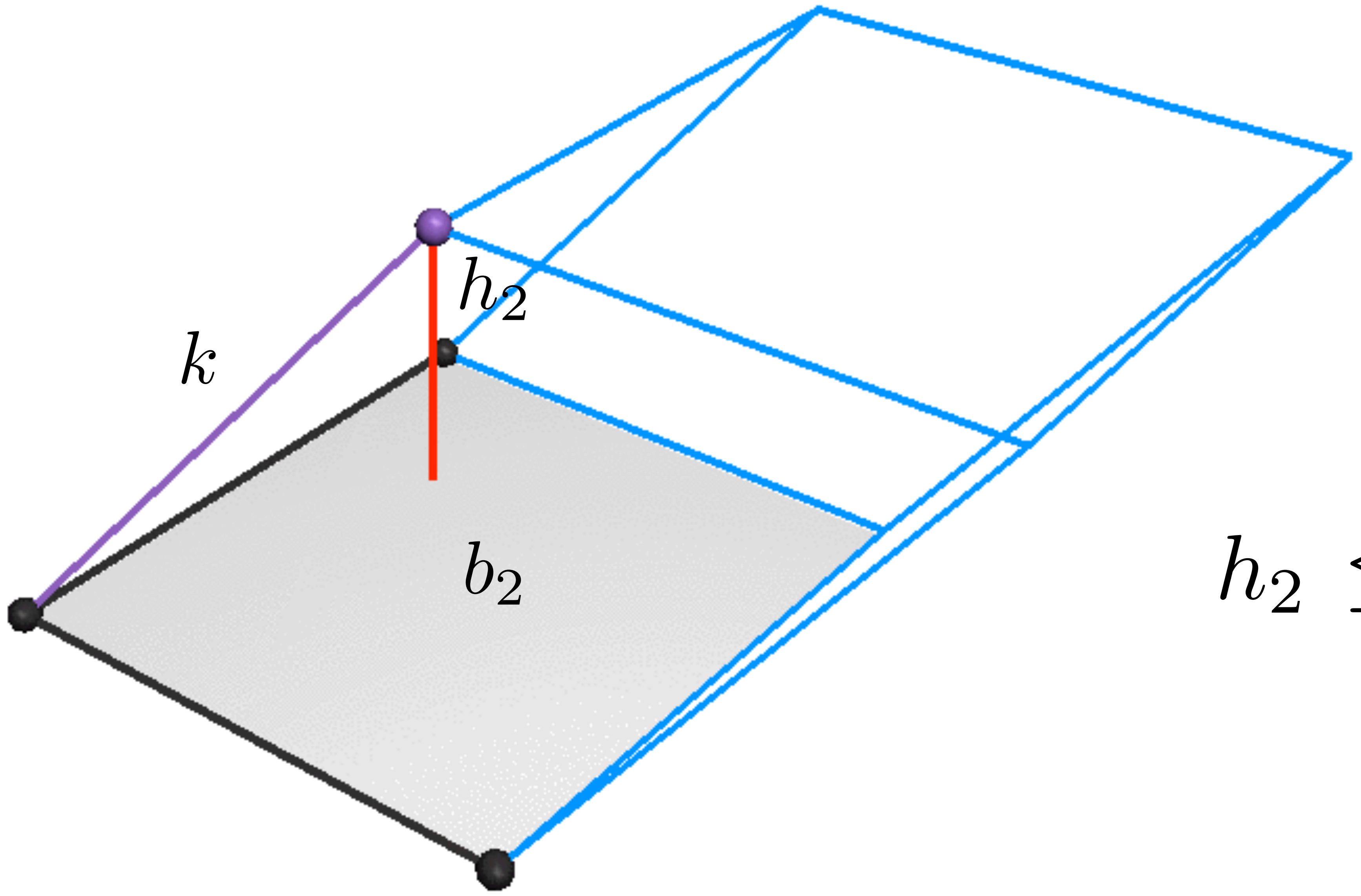


$$\frac{\mathrm{vol}(X \cup \{k\})}{\mathrm{vol}(X)} = \frac{b_1 h_1}{b_1} = h_1$$



$$\frac{\mathrm{vol}(Y \cup \{k\})}{\mathrm{vol}(Y)} = \frac{b_2 h_2}{b_2} = h_2$$

$$\frac{\mathrm{vol}(X \cup \{k\})}{\mathrm{vol}(X)}$$



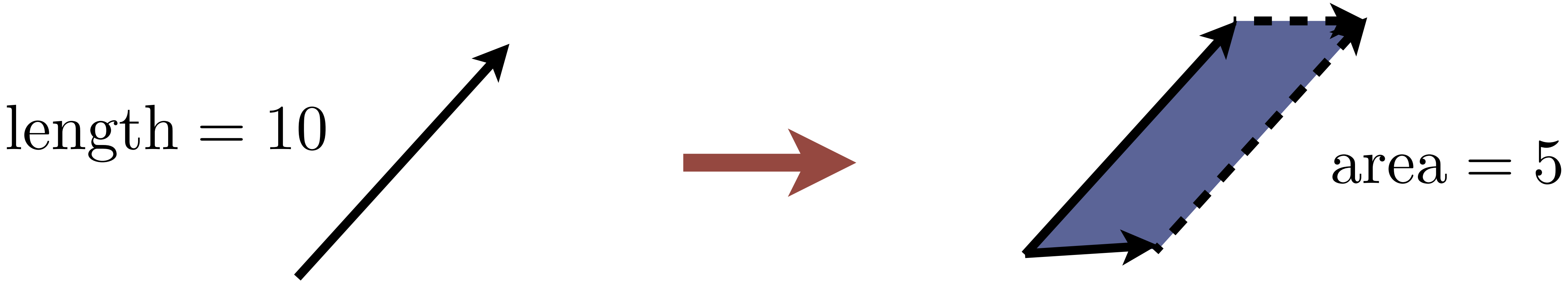
$$h_2 \leq h_1$$

$$\frac{\mathrm{vol}(Y \cup \{k\})}{\mathrm{vol}(Y)}$$

MONOTONICITY

$$X \subseteq Y \implies f(X) \leq f(Y)$$

Det is non-monotone: $\det(L_X) > \det(L_Y)$ for some X, Y



PRIOR WORK

Monotone:
"greedy" (1 - 1/e)-approx
Nemhauser and Wolsey (1978)

Non-monotone:
"symmetric greedy" 1/2-approx
Buchbinder et al. (2012)

Performs poorly in practice

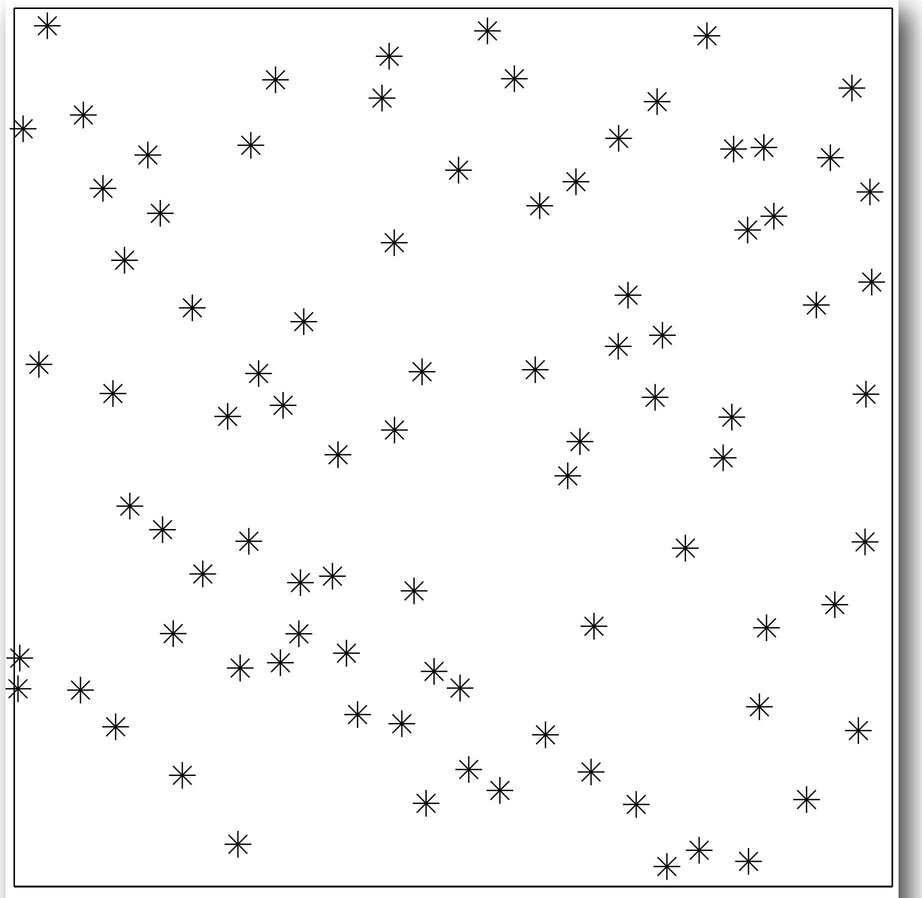
Non-monotone + constraints:
"multilinear" 1/4-approx sans constraints,
various (lesser) guarantees dependent on constraint type
Chekuri et al. (2011)

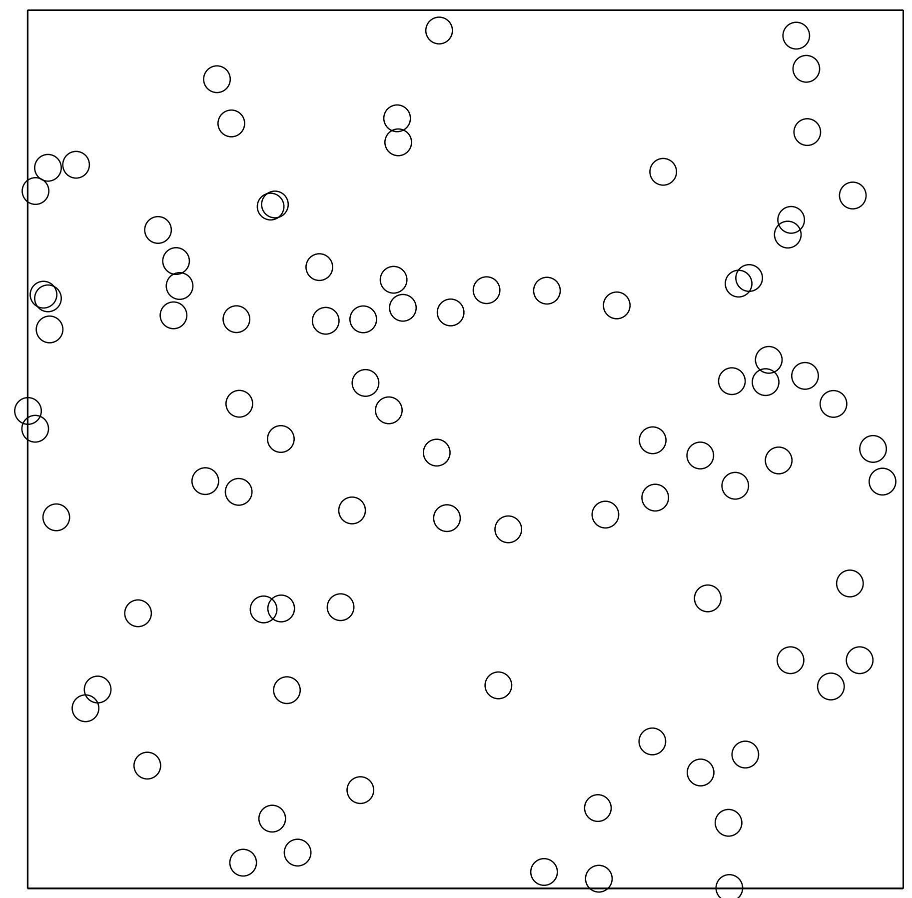
PRIOR WORK

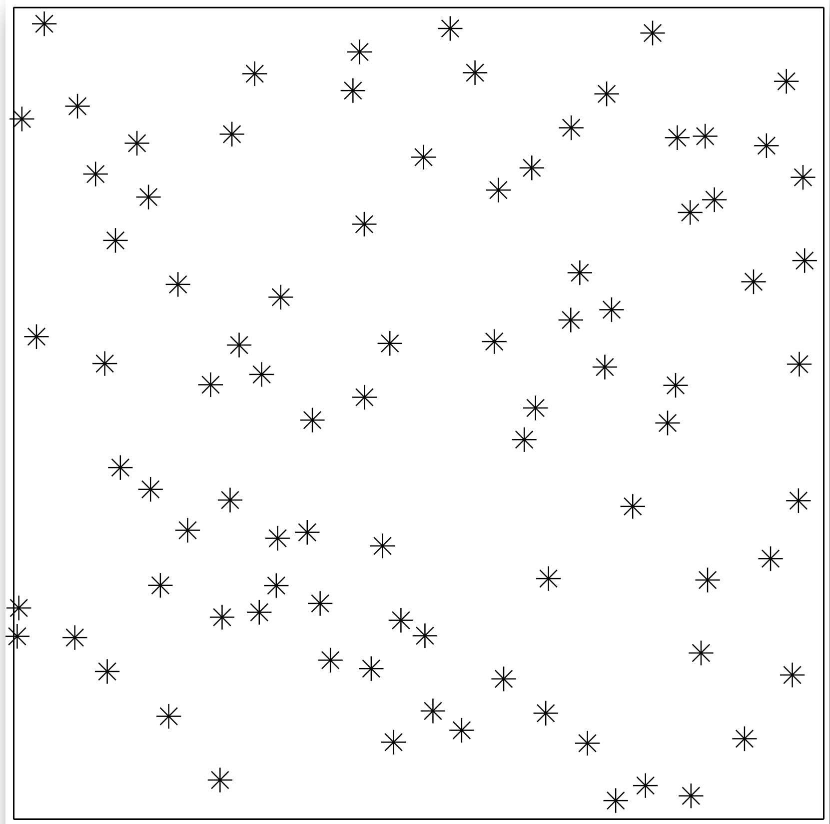
Non-monotone + constraints:
"multilinear" 1/4-approx sans constraints,
various (lesser) guarantees dependent on constraint type
Chekuri et al. (2011)

$\arg\max_Y \det(L_Y)$ → $\arg\max_{Y \in S} \det(L_Y)$
where S is a solvable polytope

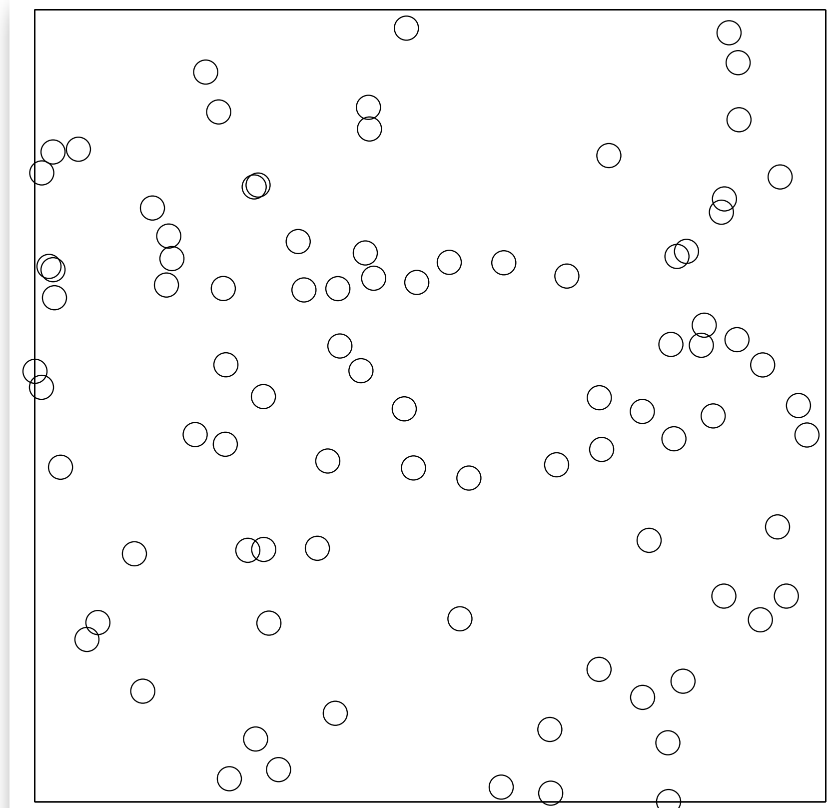
IMAGE COMPARISON WITH CONSTRAINTS


Point set 1

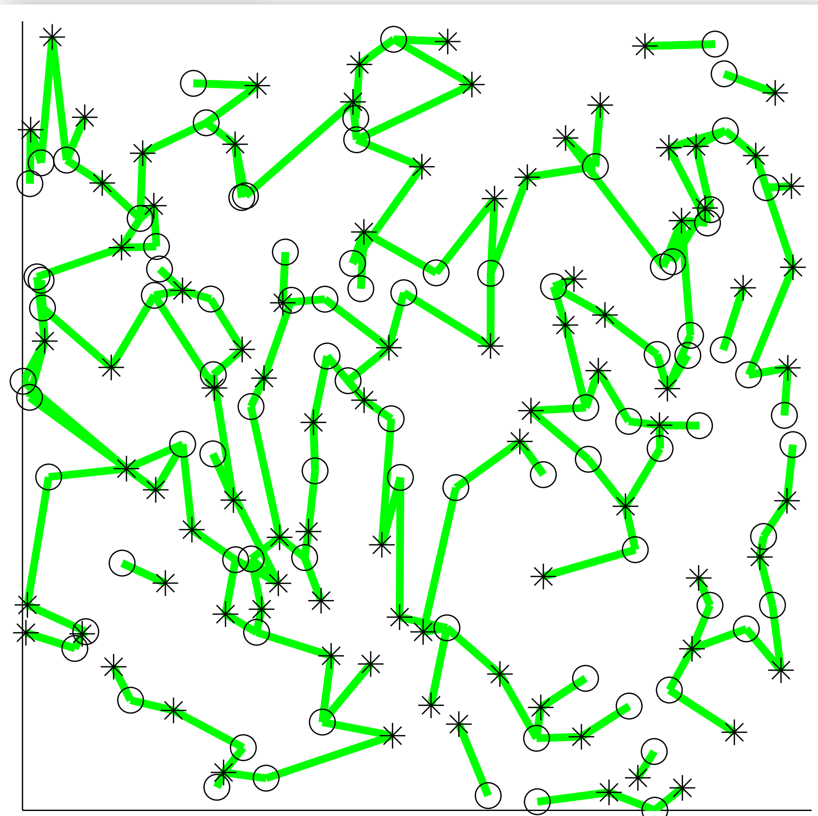

Point set 2



Point set 1

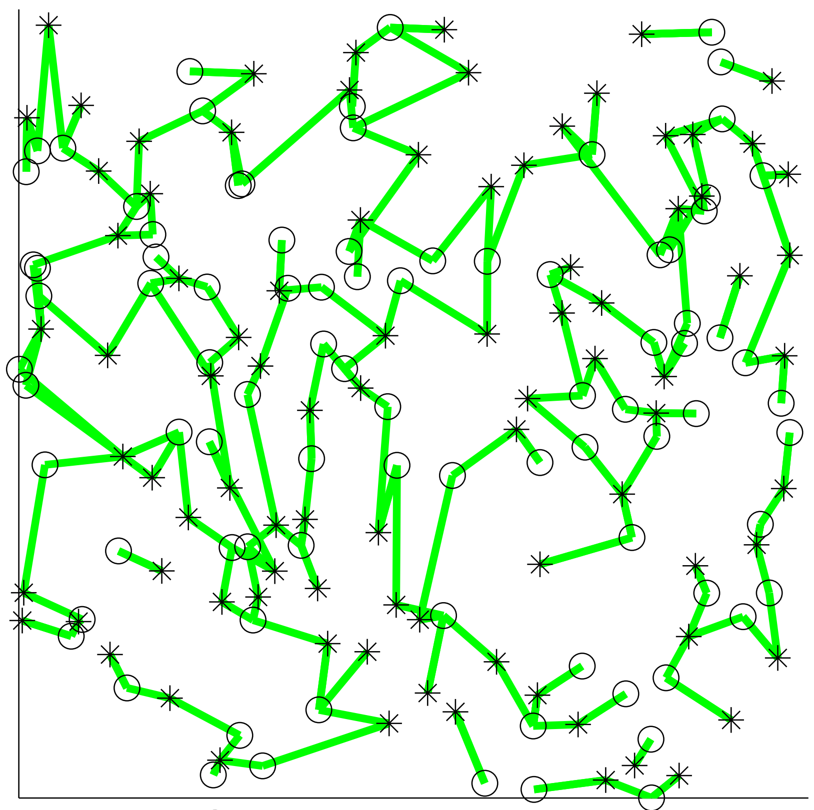


Point set 2

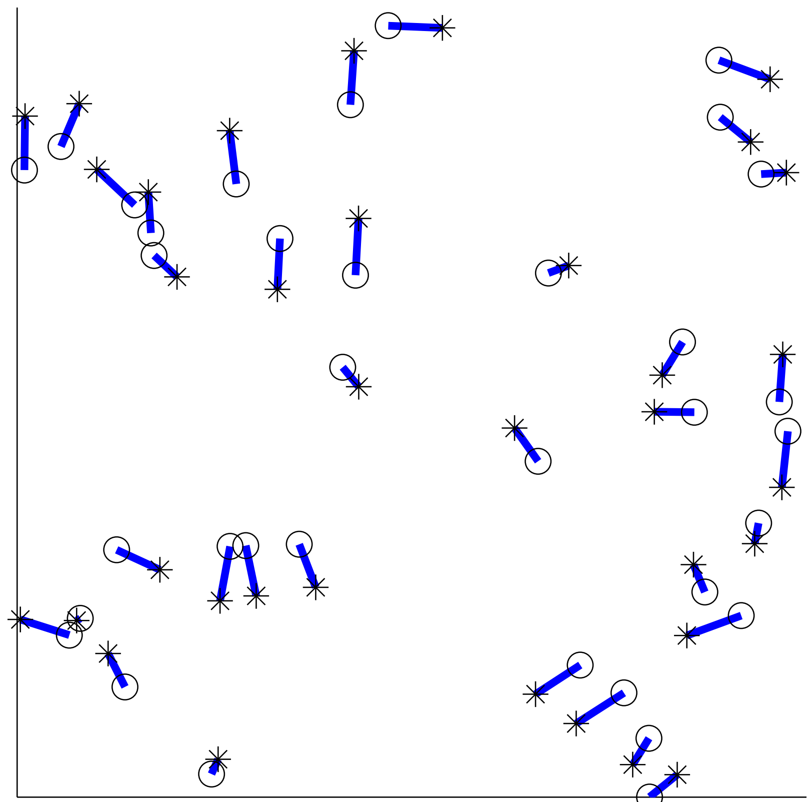


Ground set of pairs

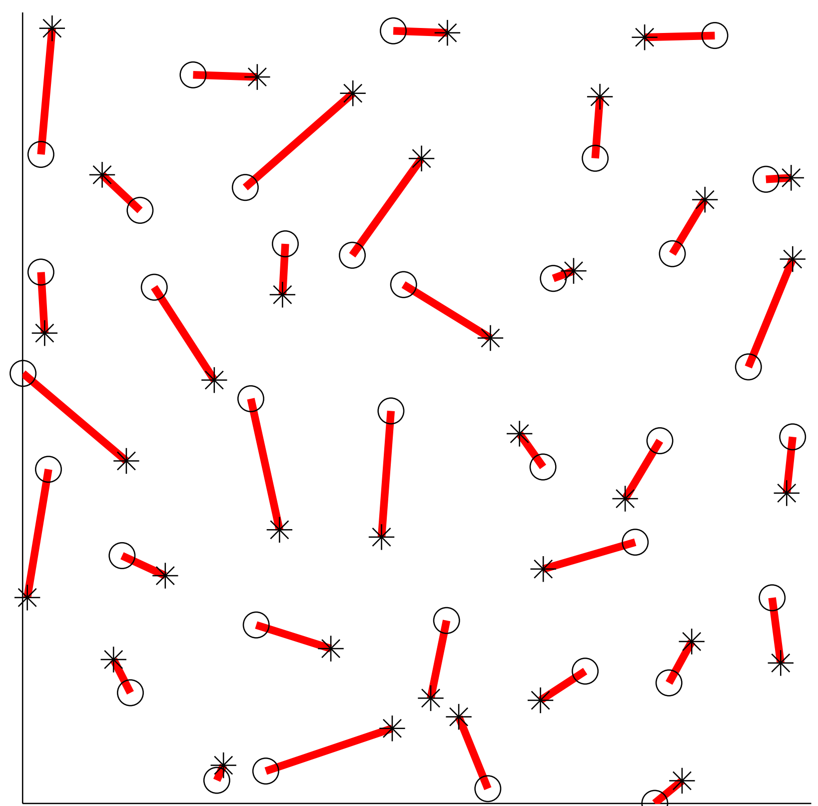
$Y \in$ matching polytope



Ground set of pairs



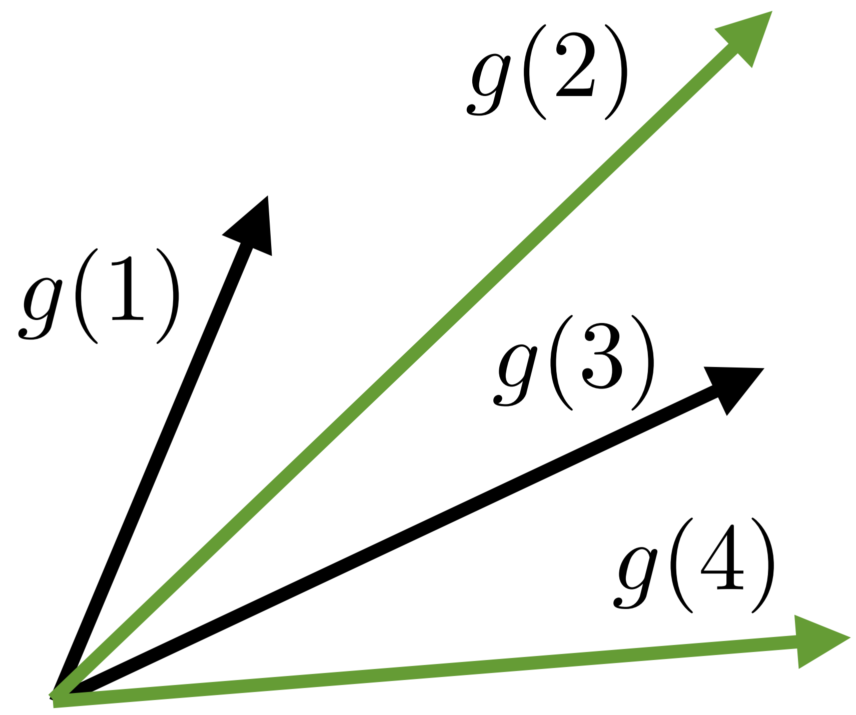
Highest quality 1:1 matching



DPP MAP 1:1 matching

CHEKURI ET AL. 2011

Step 1: Relax inclusion-exclusion



$$Y = \{2, 4\}$$

$$\mathbf{x} = [0, 1, 0, 1]$$

$$x_i \in [0, 1]$$

CHEKURI ET AL. 2011

Step 2: Extend objective

$$F(\mathbf{x}) = E_{\mathbf{x}}[\log f(Y)]$$ multilinear extension

log-submodular, like $\det(L_Y)$

CHEKURI ET AL. 2011

Step 2: Extend objective

$$F(\mathbf{x}) = E_{\mathbf{x}}[\log f(Y)]$$ multilinear extension

$$= \sum_Y \boxed{\prod_{i \in Y} x_i \prod_{i \notin Y} (1 - x_i)} \log f(Y)$$

$\downarrow$

$$p_Y$$

CHEKURI ET AL. 2011

Step 2: Extend objective

$$F(\mathbf{x}) = E_{\mathbf{x}}[\log f(Y)]$$

multilinear extension

$$= \sum_Y \prod_{i \in Y} x_i \prod_{i \notin Y} (1 - x_i) \log f(Y)$$

2^N subsets

CHEKURI ET AL. 2011

Step 2: Extend objective

$$F(\mathbf{x}) = E_{\mathbf{x}}[\log f(Y)]$$ multilinear extension

$$= \sum_Y \prod_{i \in Y} x_i \prod_{i \notin Y} (1 - x_i) \log f(Y)$$

2^N subsets

Step 3: Optimize using gradient-based methods $\frac{\partial F(\mathbf{x})}{\partial \mathbf{x}}$

Step 4: If unconstrained, solution will already be integer; else, round solution: $x_i \in [0,1] \to x_i \in \{0,1\}$

CHEKURI ET AL. 2011

Step 2: Extend objective

$$F(\mathbf{x}) = E_{\mathbf{x}}[\log f(Y)]$$ multilinear extension

$$= \sum_Y \prod_{i \in Y} x_i \prod_{i \notin Y} (1 - x_i) \log f(Y)$$

2^N subsets $\implies$ Monte Carlo required

Step 3: Optimize using gradient-based methods $\frac{\partial F(\mathbf{x})}{\partial \mathbf{x}}$

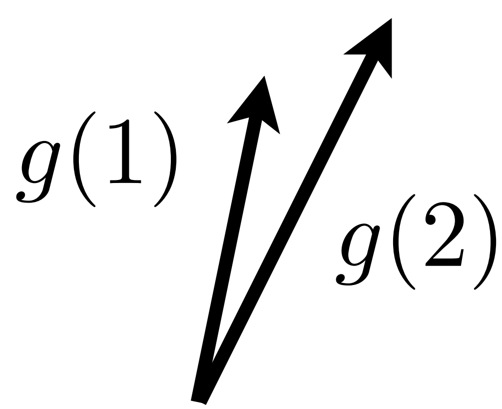
Step 4: If unconstrained, solution will already be integer; else, round solution: $x_i \in [0, 1] \to x_i \in \{0, 1\}$

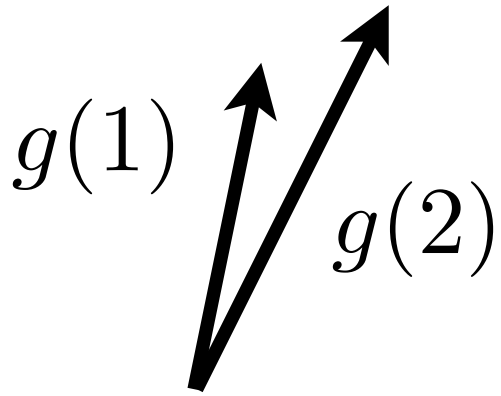
SOFTMAX EXTENSION

Multilinear: $F(\mathbf{x}) = \sum_Y \prod_{i \in Y} x_i \prod_{i \notin Y} (1 - x_i) \log f(Y)$

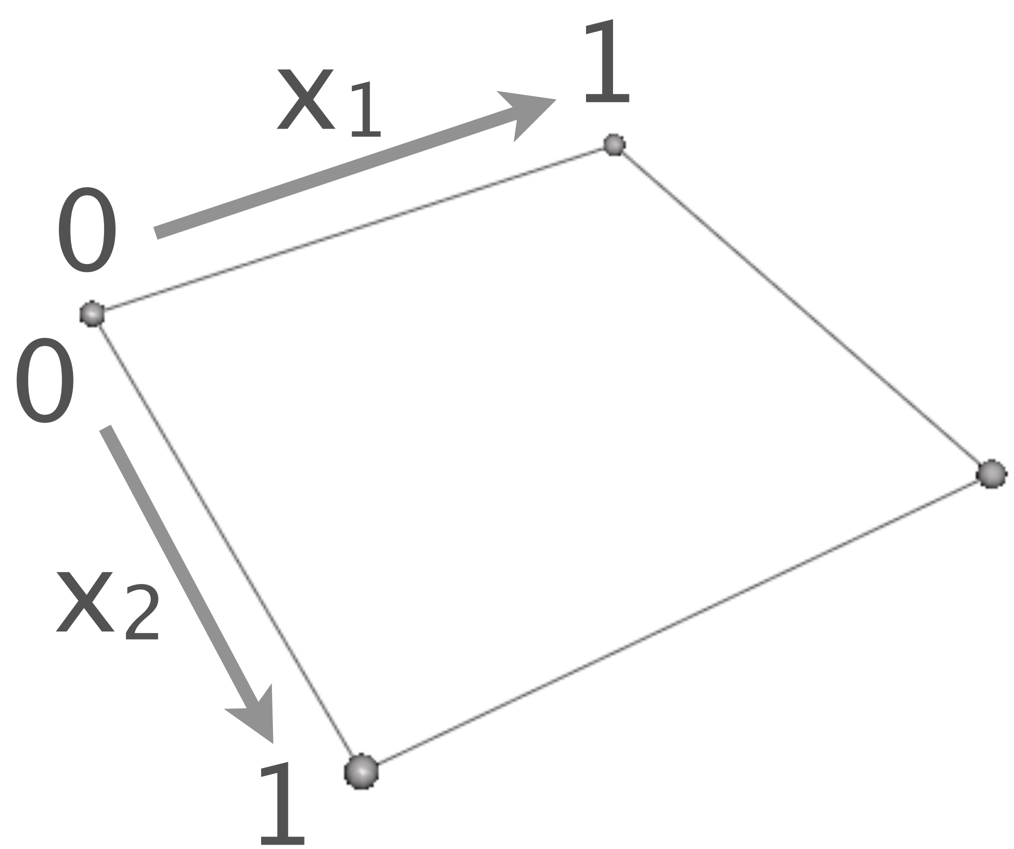
Softmax: $\tilde{F}(\mathbf{x}) = \log \sum_Y \prod_{i \in Y} x_i \prod_{i \notin Y} (1 - x_i) f(Y)$

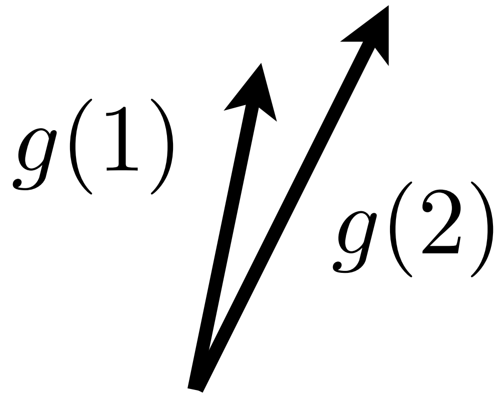
$N = 2$





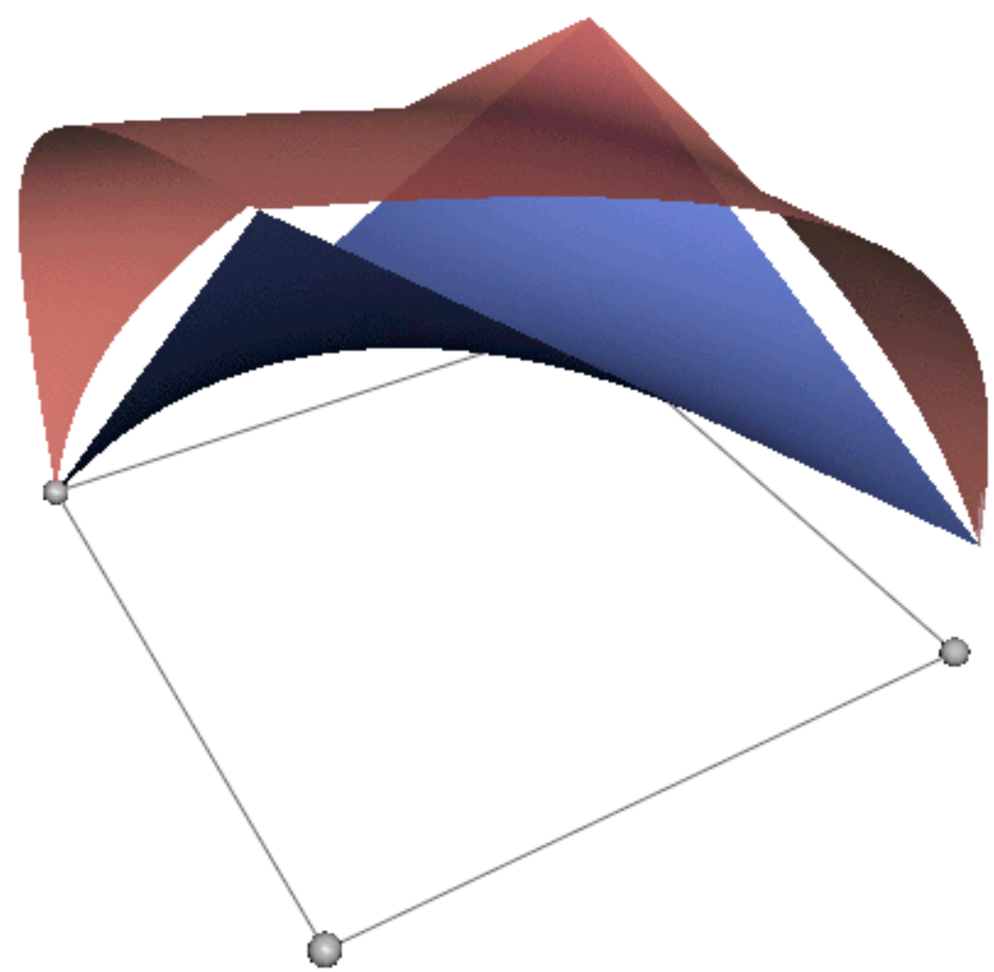
Relaxed domain

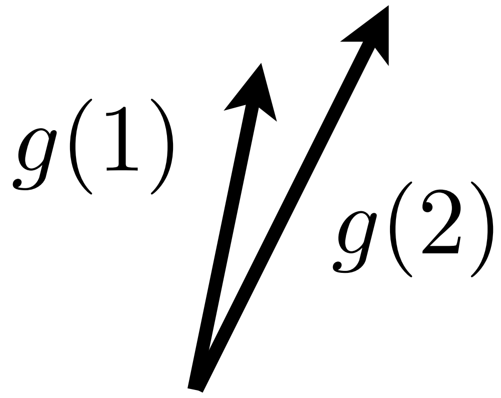




Softmax extension

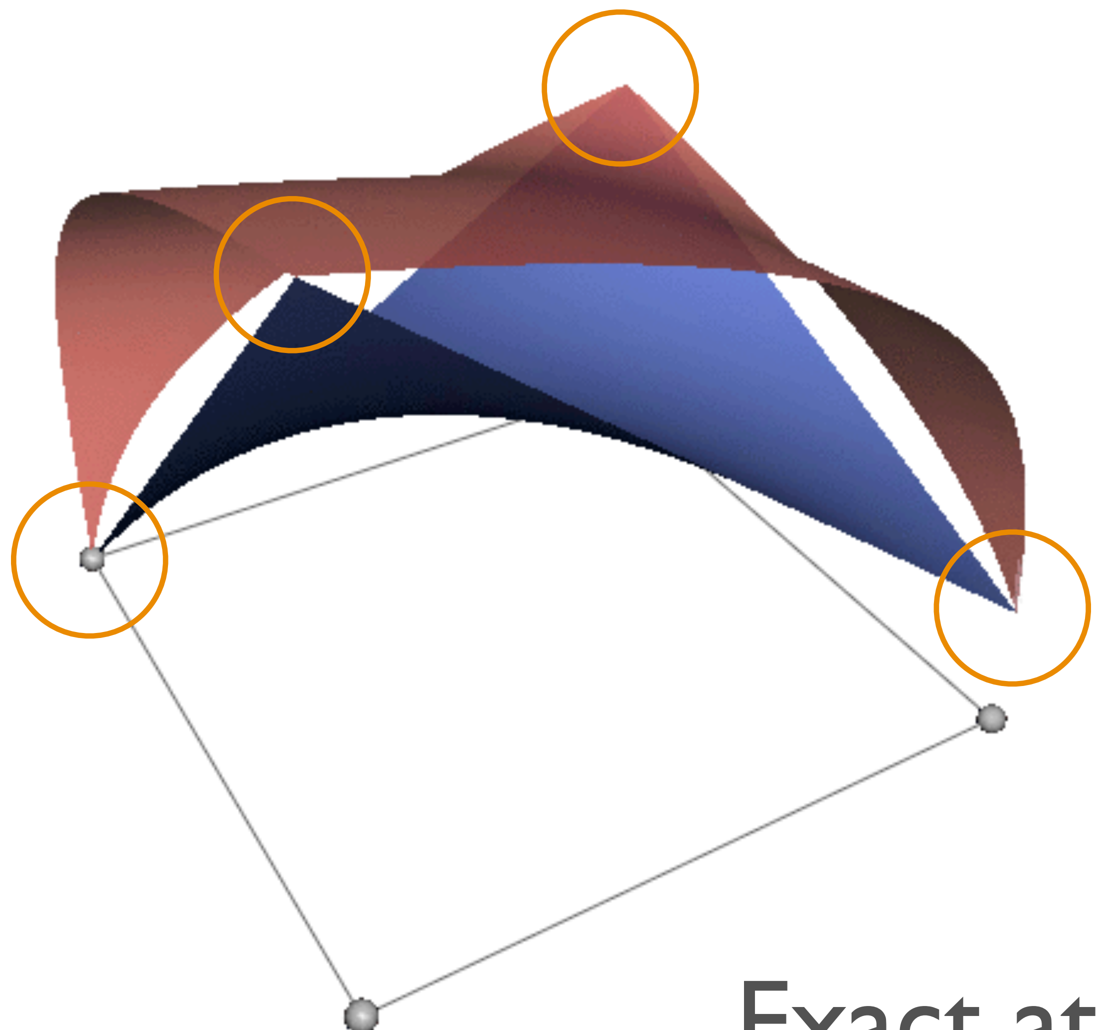
Multilinear extension



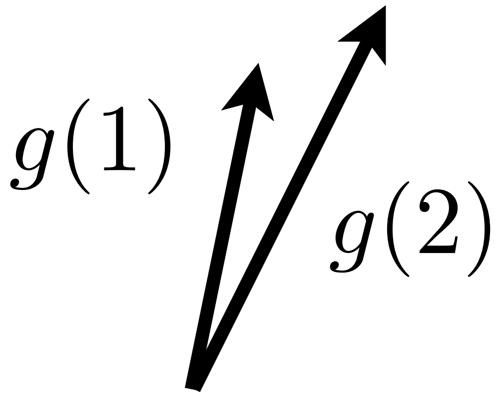


Softmax extension

Multilinear extension

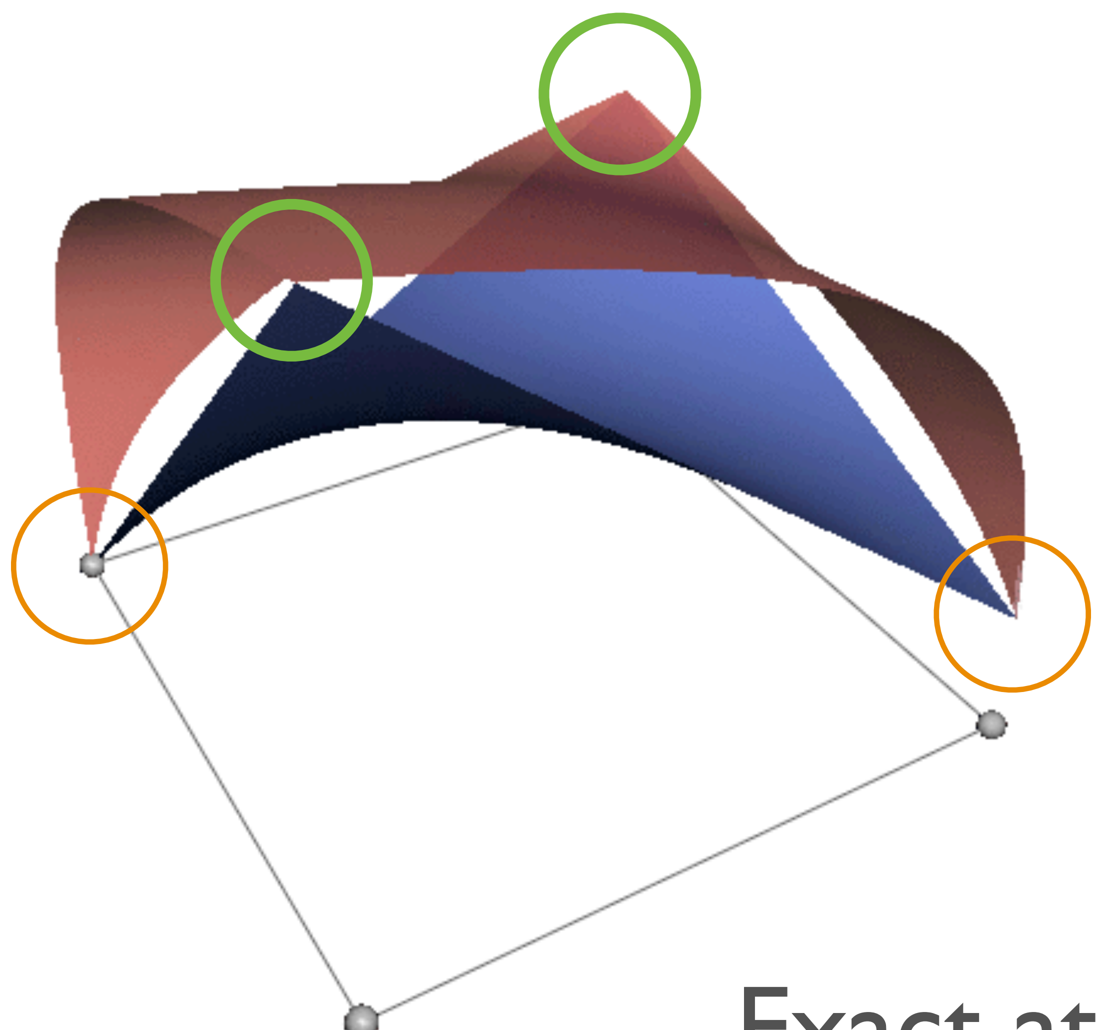


Exact at integral points



Softmax extension

Multilinear extension



Exact at integral points

SOFTMAX EXTENSION

$$\tilde{F}(\mathbf{x}) = \log \sum_Y \prod_{i \in Y} x_i \prod_{i \notin Y} (1 - x_i) f(Y)$$

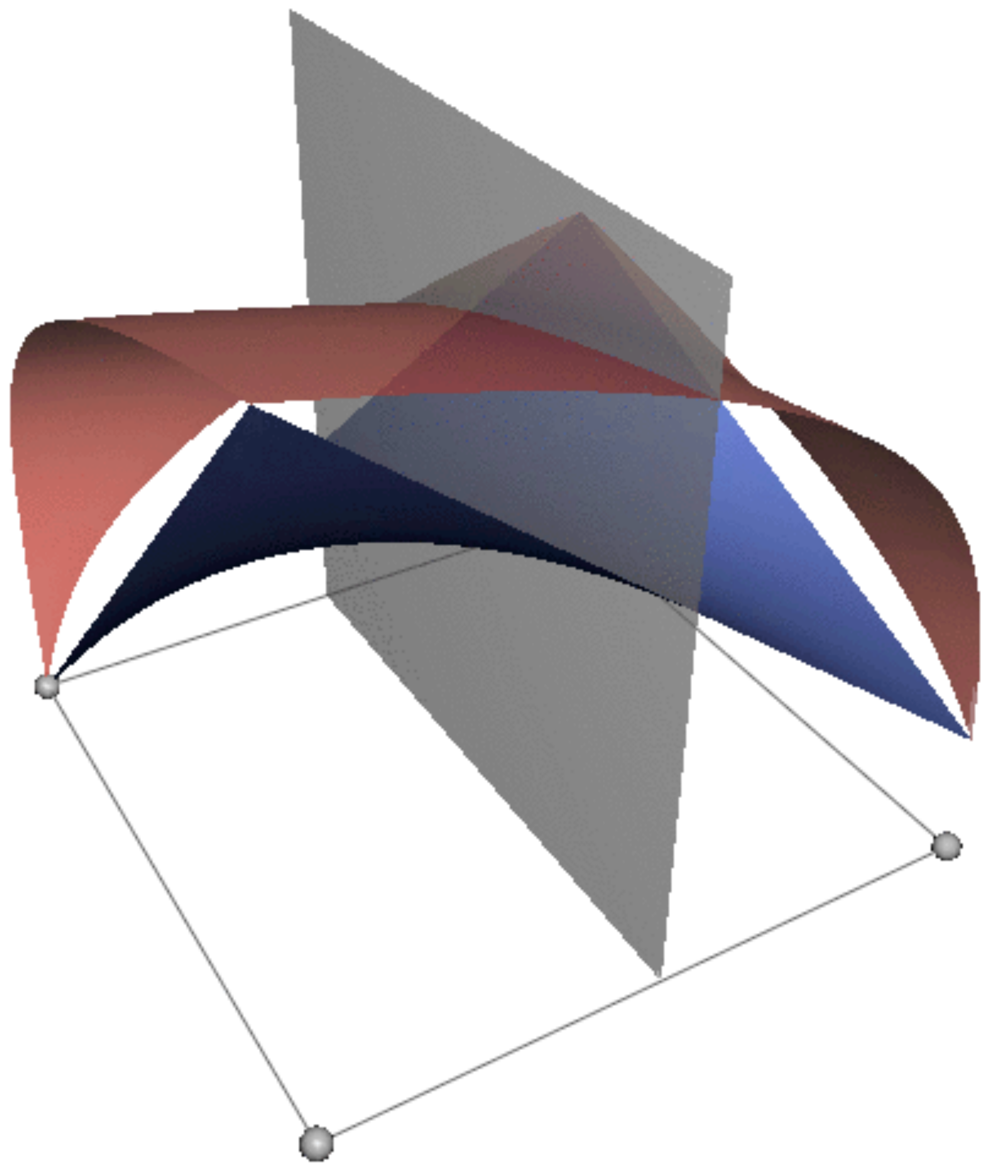
Theorem:

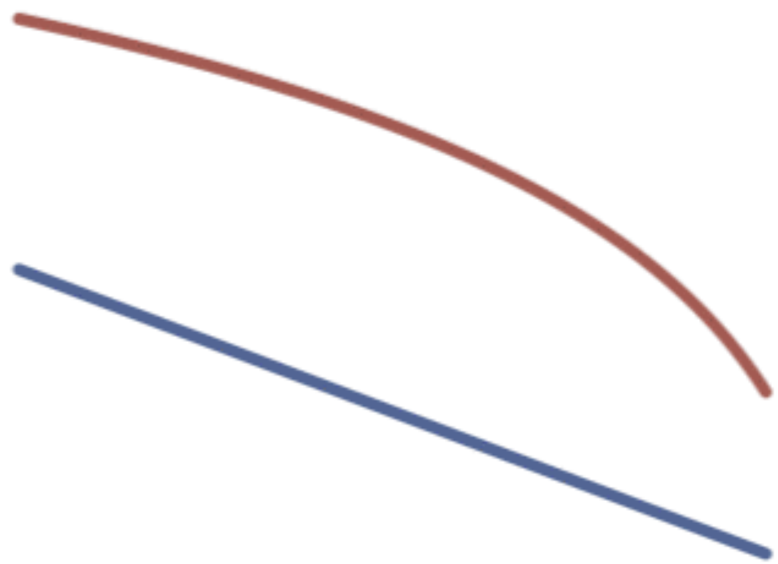
Efficiently computable for $f(Y) = \det(L_Y)$

$$O(N^3)$$

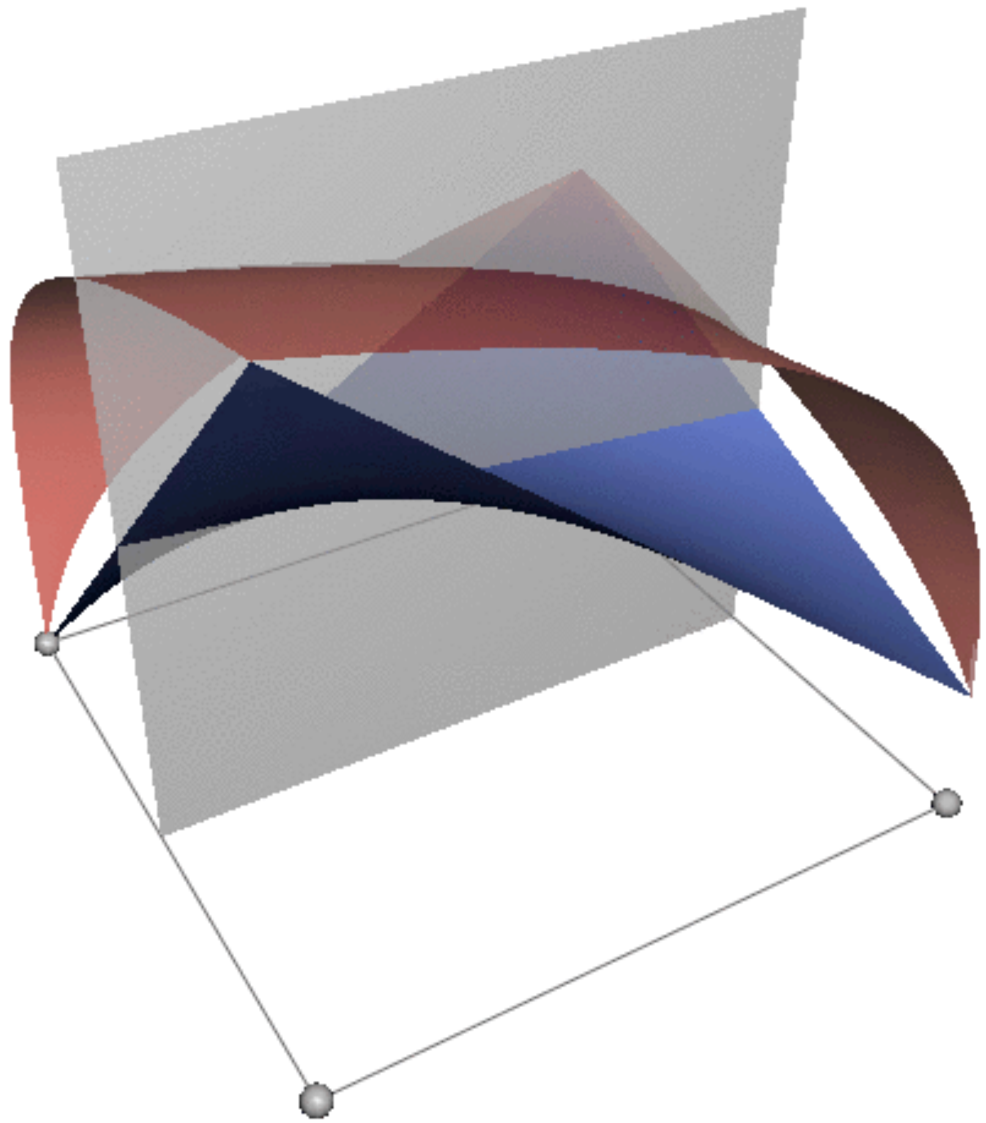
$$\tilde{F}(\mathbf{x}) = \det(\text{diag}(\mathbf{x})(L - I) + I)$$

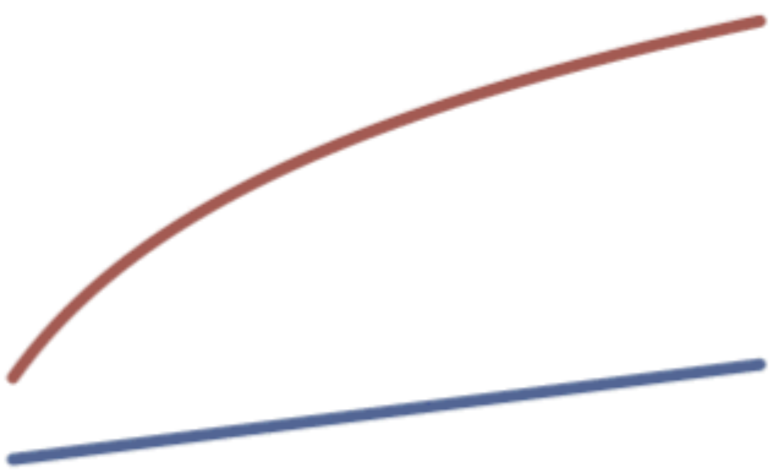
Concave in all-positive/all-negative directions





Not necessarily concave in other directions





APPROXIMATION GUARANTEE

Theorem: Concavity in all-positive directions

+ Submodularity $\Longrightarrow$

$$\text{LOCAL OPT of } \tilde{F} \geq \frac{1}{4} \max_{\mathbf{x}} \tilde{F}(\mathbf{x}) \geq \frac{1}{4} \max_{Y} \log\det(L_Y)$$

Theorem: In the unconstrained case, LOCAL OPT will be integer (no rounding necessary).

Constrained: No guarantees, but in practice pipage rounding and thresholding work well.

$\max_{Y \in S}$

BASELINE

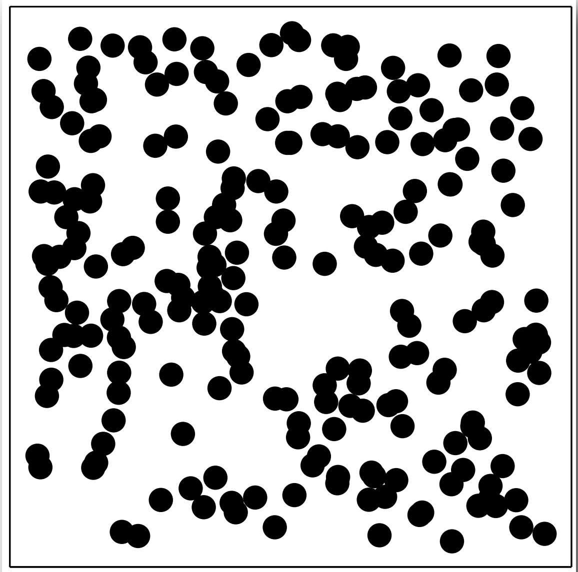
Monotone:
"greedy" (1 - 1/e)-approx
Nemhauser and Wolsey (1978)

Basic idea: Start from Y = {} and find the single item to add that most increases the score. Iterate until no remaining item increases the score.

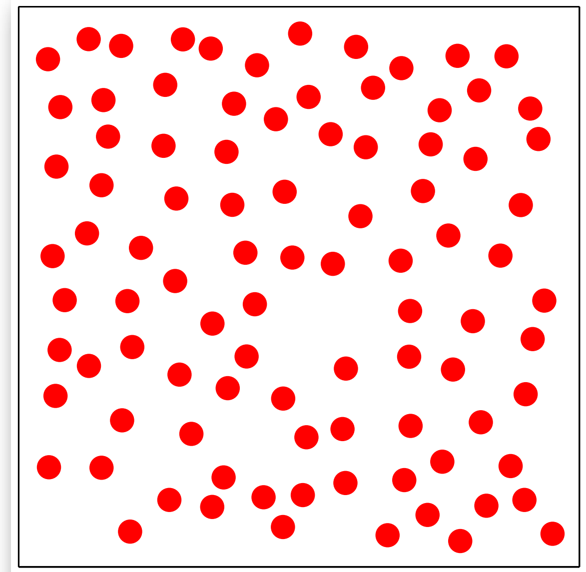
SYNTHETIC EXPERIMENTS

Unconstrained

$\max_Y$



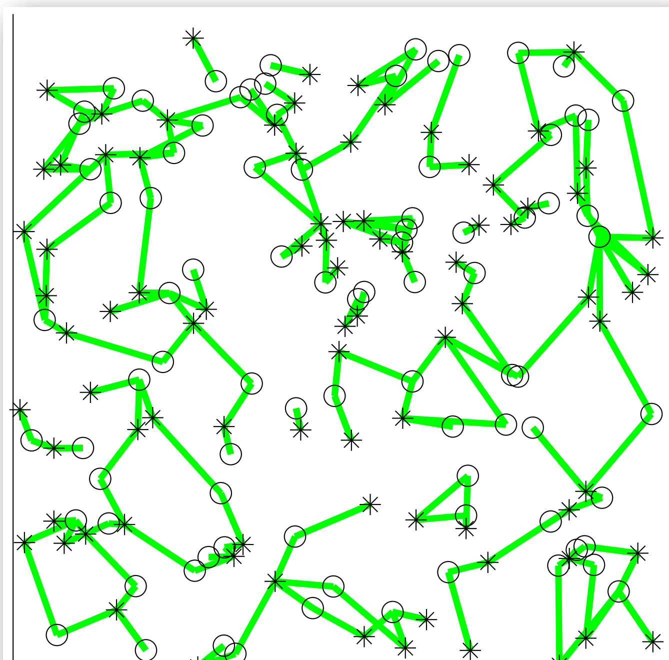
All points



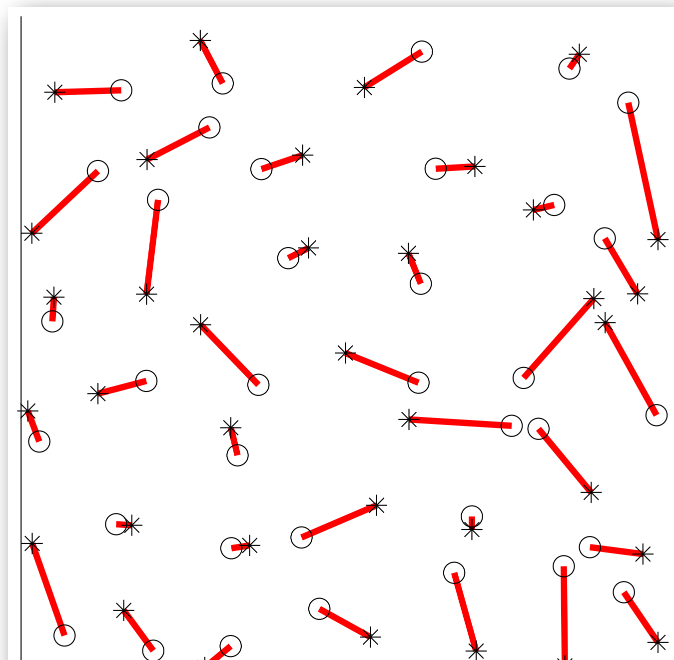
DPP (approx) MAP

Constrained

$\max_{Y \in S}$

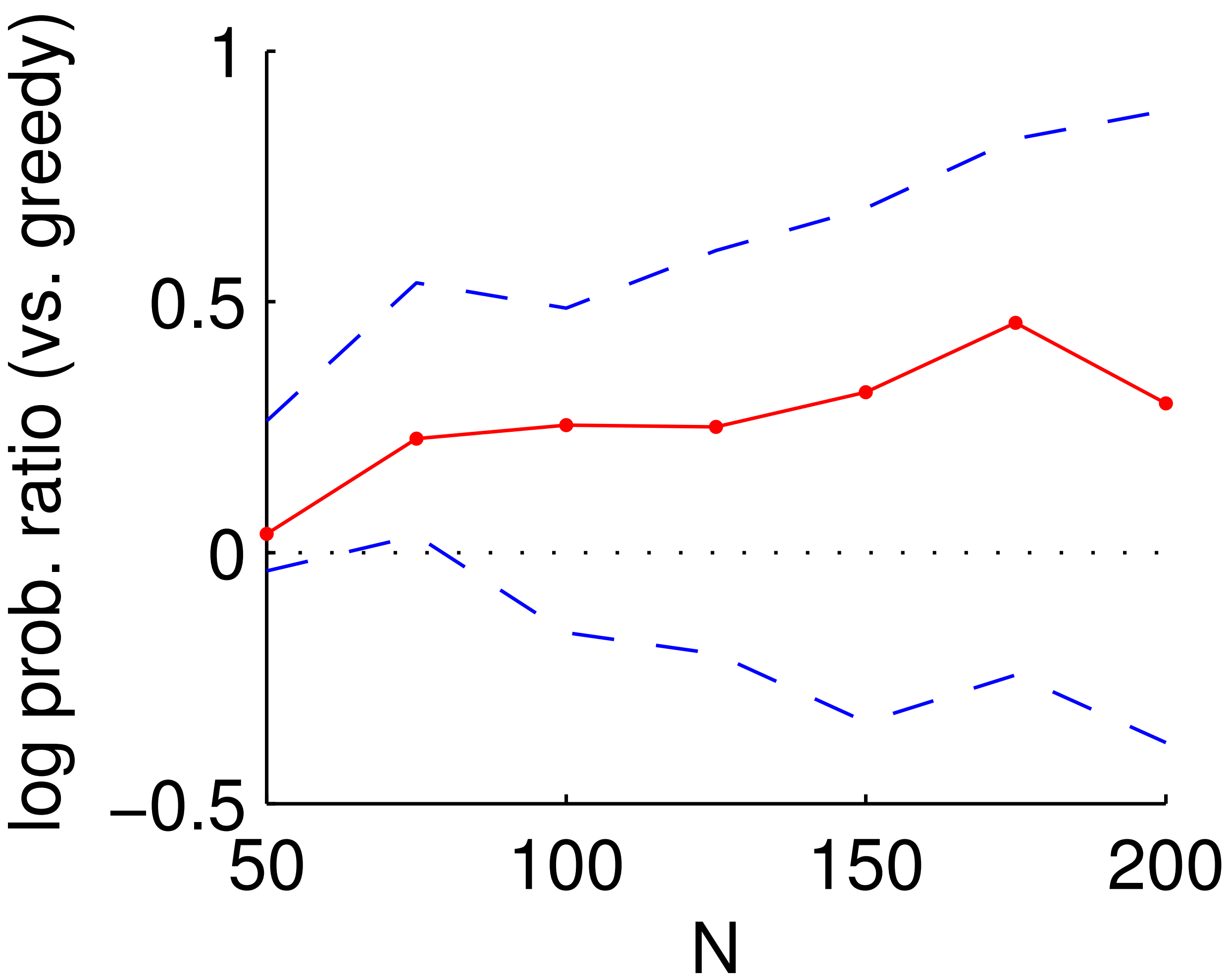


Many matches



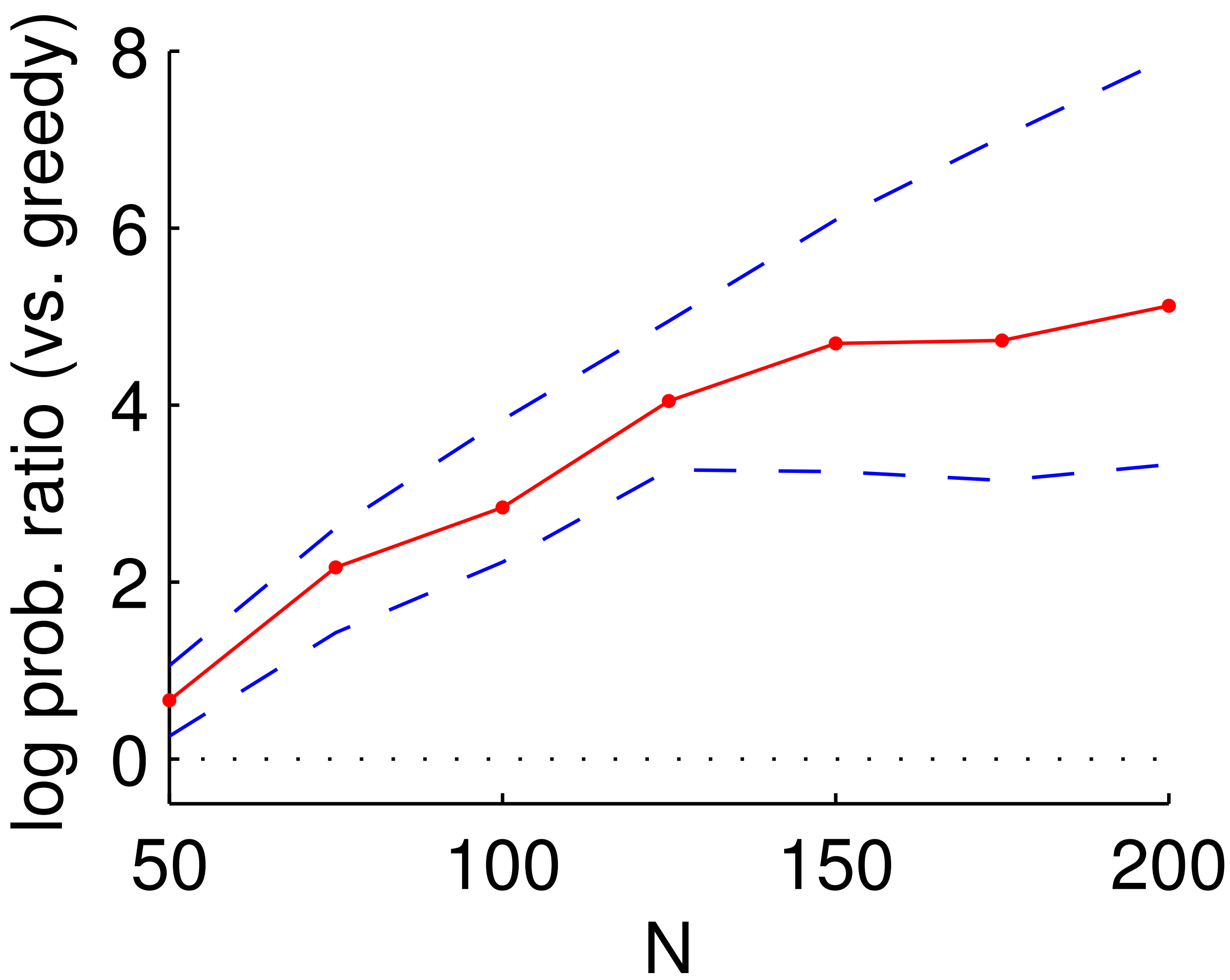
DPP MAP 1:1 matching

EFFECTIVENESS EVAL



Unconstrained

$\max_{Y}$



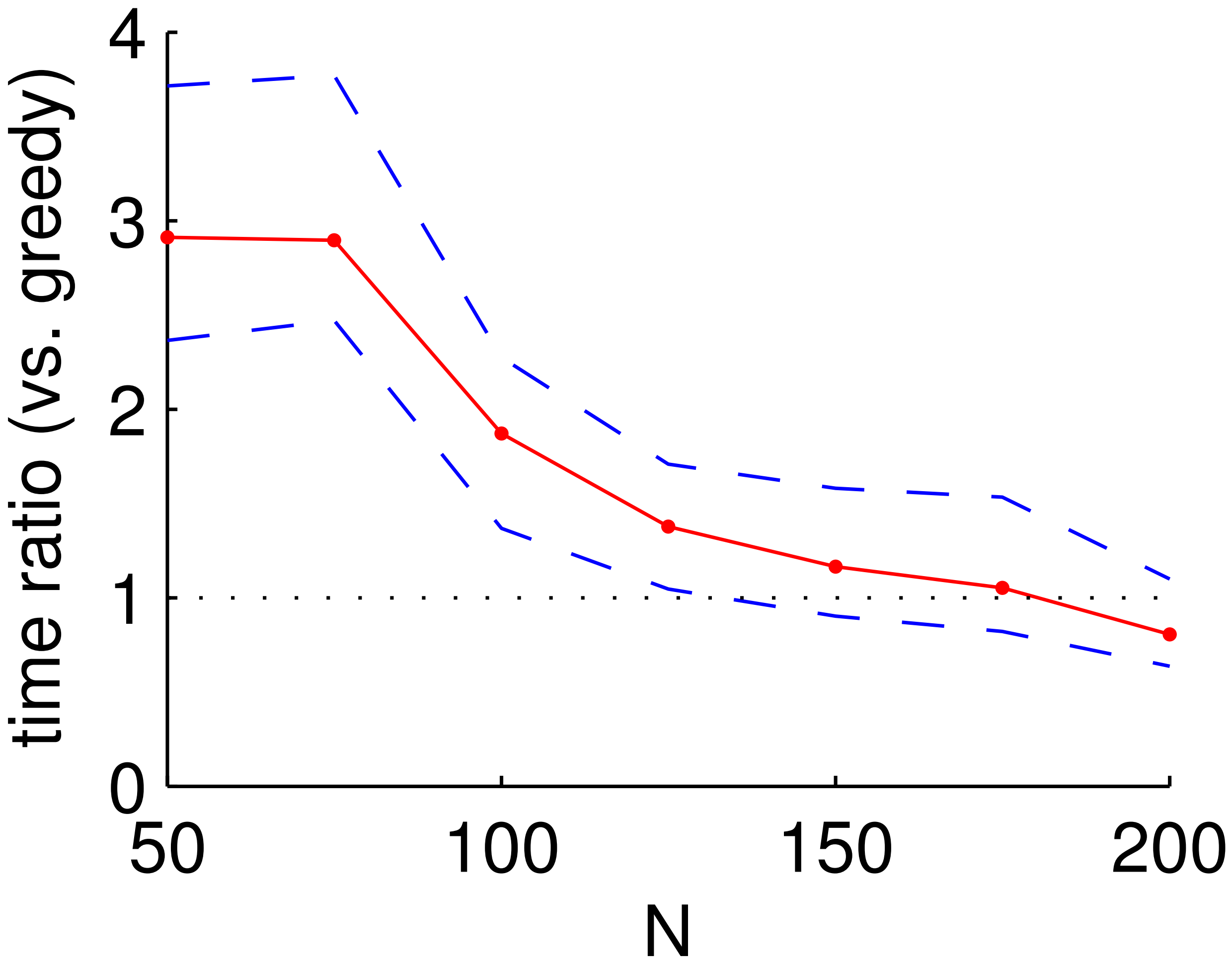
Constrained

$\max_{Y \in S}$

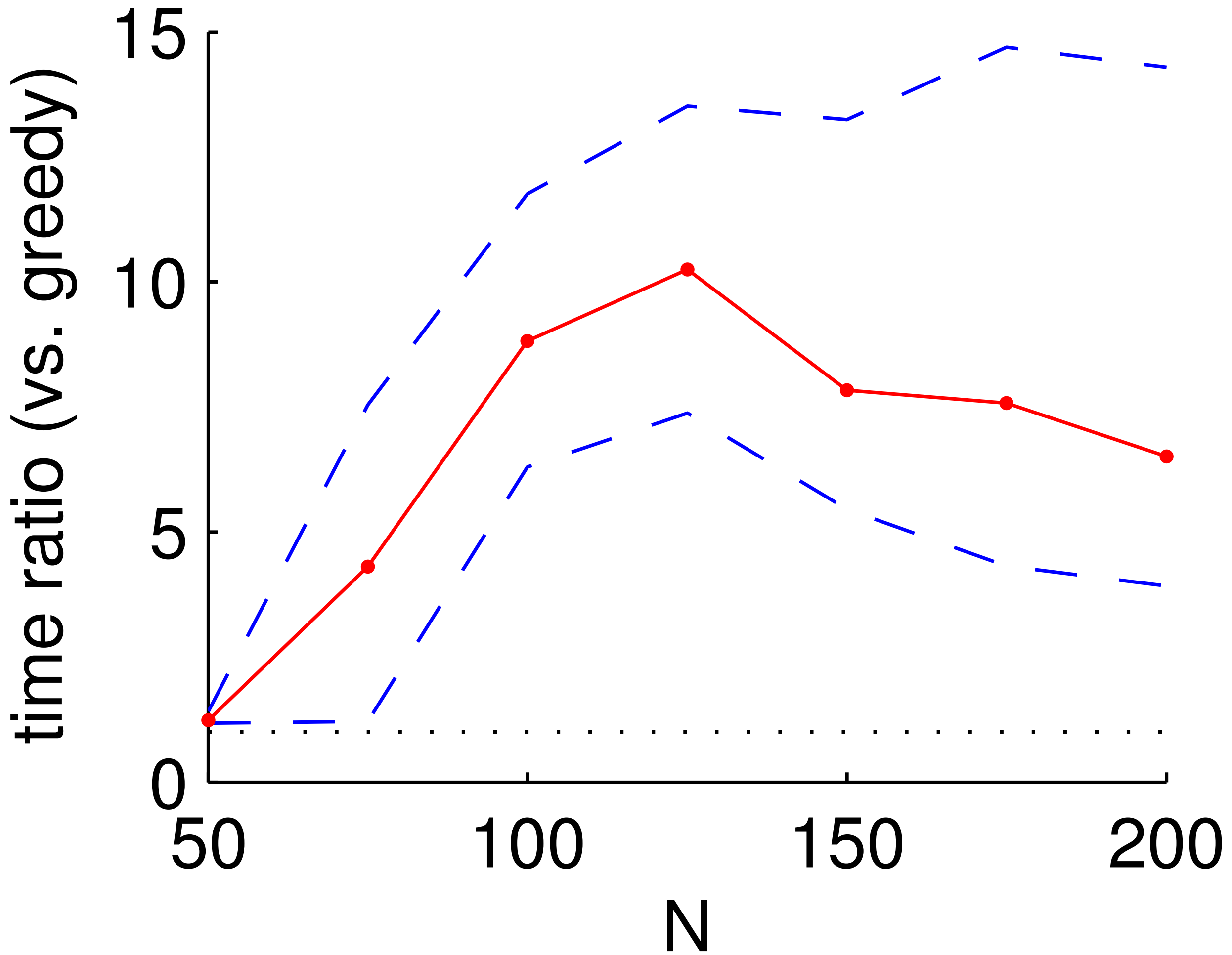
EFFICIENCY EVAL

Naive greedy algorithm: $O(N^5)$
Optimized version: $O(N^4)$

EFFICIENCY EVAL



Unconstrained
$\max_Y$



Constrained
$\max_{Y \in S}$

MATCHED SUMMARIZATION

20 Republican primary debates



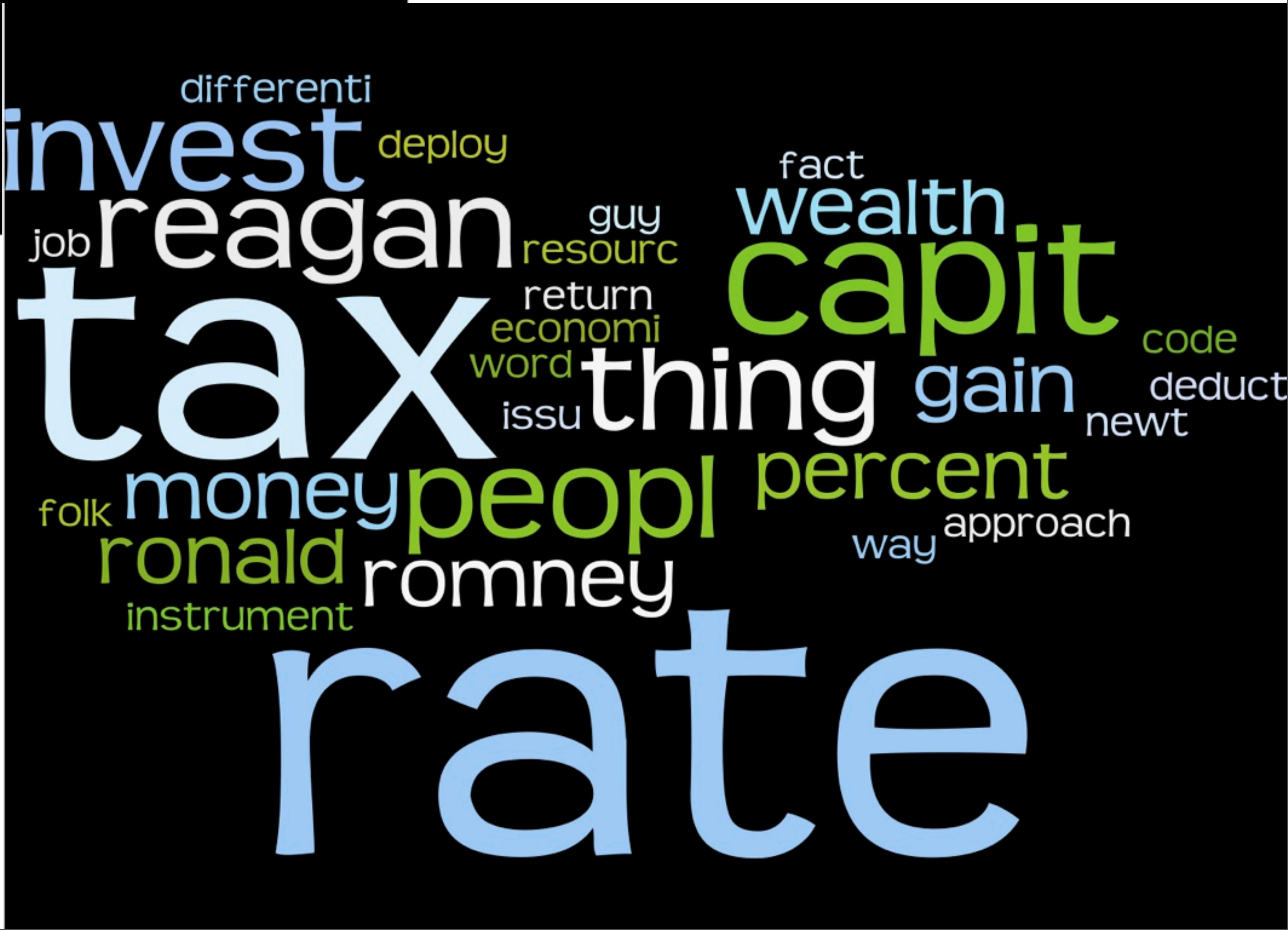




Average of 179 quotes per candidate

MATCHED SUMMARIZATION

bowl
base
friend
nation
middle-class
obama
reduct
dividend
peopl
except
gain
time
interest
way
inch
term
place
amount
save
american
incom
bowles-simpson
govern
scale
code
economi
number
jon
reign
relief
employ
rate
capit
exempt
simpson
incent
democrat
percent
tax



differenti
invest
deploy
fact
wealth
job
reagan
guy
resourc
capit
return
tax
economi
code
word
thing
gain
deduct
issu
newt
percent
folk
money
peopl
approach
way
ronald
romney
instrument
rate

- **R1 (taxes): No tax on interest, dividends, or capital gains.**
- **R2 (law): We're not going to have Sharia law applied in U.S. courts.**
- **R3 (healthcare): I will ... grant a waiver from Obamacare to all 50 states.**
- **R4 (aid): We're spending more on foreign aid than we ought to.**
- **R5(healthcare): If you think what we did in Massachusetts and what President Obama did are the same, boy, take a closer look.**

- **S1 (taxes): I don't believe in a zero capital gains tax rate.**
- **S2 (taxes): Manufacture in America, you aren't going to pay any taxes.**
- **S3 (aid): Zeroing out foreign aid ... that's absolutely the wrong course.**
- **S4 (ethanol): I voted against ethanol subsidies my entire time in Congress.**
- **S5 (healthcare): Obamacare ... is going to blow a hole in the budget.**

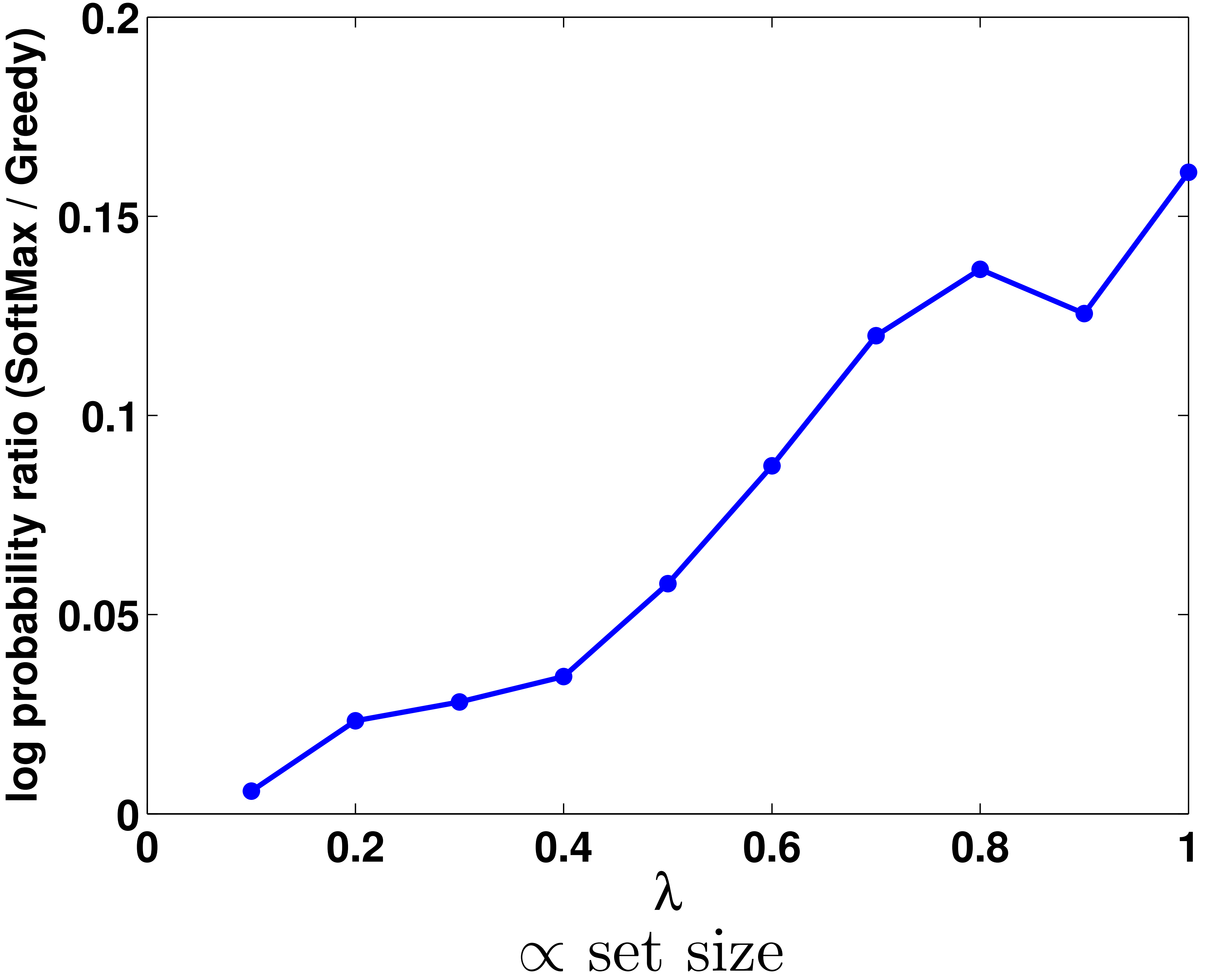
Matched summary

- **R1 (taxes): No tax on interest, dividends, or capital gains.**
- **S1 (taxes): I don’t believe in a zero capital gains tax rate.**

- **R3 (healthcare): I will … grant a waiver from Obamacare to all 50 states.**
- **S5 (healthcare): Obamacare … is going to blow a hole in the budget.**

- **R4 (aid): We’re spending more on foreign aid than we ought to.**
- **S3 (aid): Zeroing out foreign aid … that’s absolutely the wrong course.**

PERFORMANCE



SUMMARY

Efficient, effective approximate DPP MAP algorithm for subset selection problems

Code + data:

http://www.seas.upenn.edu/~jengi/dpp-map.html

SUMMARY

Code + data:

http://www.seas.upenn.edu/~jengi/dpp-map.html

Future work:

- Other applications:
  - sensor selection
    [A. Krause, A. Singh, and C. Guestrin. Near-Optimal Sensor Placements in Gaussian Processes, 2008.]
  - text summarization
    [H. Lin and J. Bilmes. Multi-Document Summarization via Budgeted Maximization of Submodular Functions, 2010.]
- Other submodular functions for which the softmax extension is efficiently computable?

Poster W35