Near-Optimal MAP Inference for Determinantal Point Processes

Jennifer Gillenwater Alex Kulesza Ben Taskar

{jengi,kulesza,taskar} @ cis.upenn.edu

TASK: SUBSET SELECTION

two goals

quality diversity

EX: IMAGE SEARCH -- "JAGUAR"

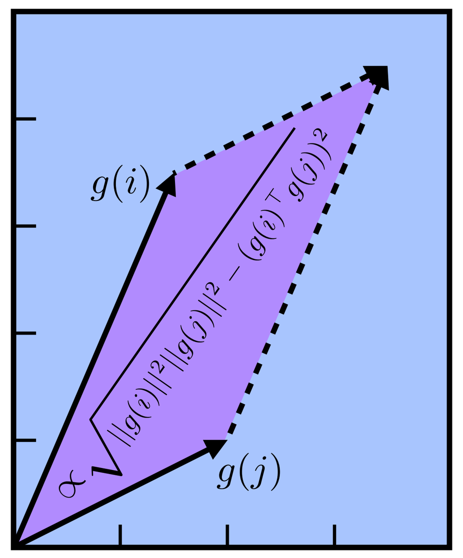
Relevance only:

...

Relevance + diversity:

...

FORMALIZING



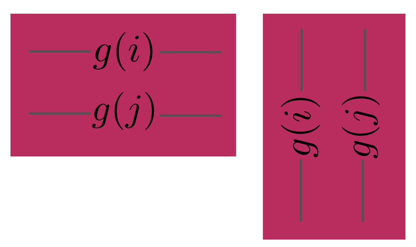
feature space

quality $= ||g(i)||^2 = g(i)^\top g(i)$

similarity $= g(i)^\top g(j)$

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

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

$= \det\Big($  $\Big)$

"goodness" of set Y = quality & diversity of Y

$\propto$ volume$_{|Y|}(Y)^2$

volume$_1$ = length

volume$_2$ = area

etc.

$\propto \det((GG^\top)_Y)$

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

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

DPP INFERENCE

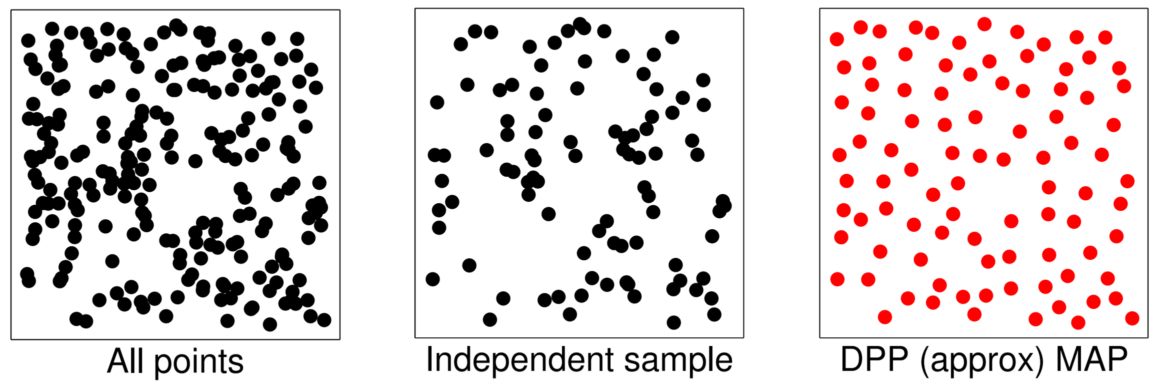
Exact and efficient $O(N^3)$

- normalization: $\sum_Y \det(L_Y) = \det(L + I)$
- marginalization: $\mathcal{P}(A \subseteq Y)$
- conditioning: $\mathcal{P}(A \mid B \subseteq Y)$
- sampling: $Y \sim \det(L_Y)$

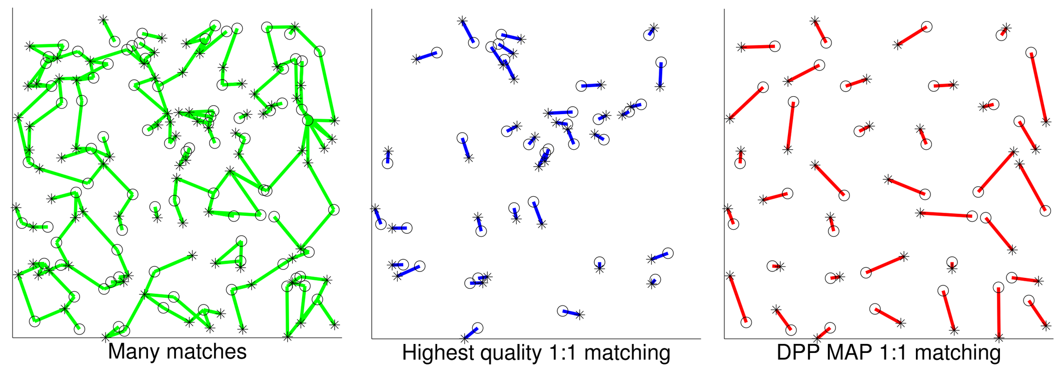
But: DPP MAP is NP-hard

DPP MAP: NP-HARD

$\arg\max_Y \det(L_Y)$



EVEN HARDER: CONSTRAINED MAP



SUBMODULARITY TO THE RESCUE

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

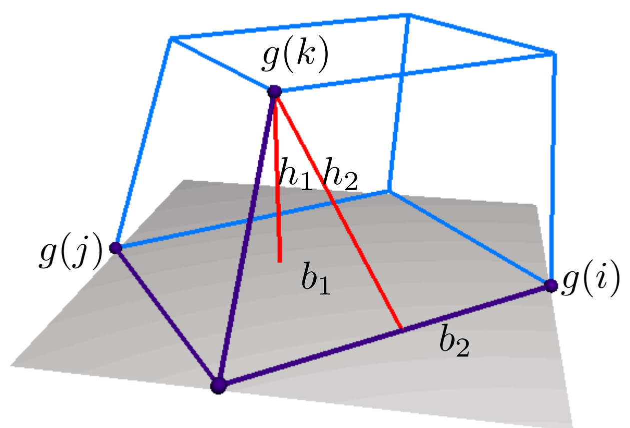
Diminishing returns:

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

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

$$\frac{\mathrm{vol}(g(i), g(j), g(k))}{\mathrm{vol}(g(i), g(j))} = \frac{b_1 h_1}{b_1} = h_1$$

$$\frac{\mathrm{vol}(g(i), g(k))}{\mathrm{vol}(g(i))} = \frac{b_2 h_2}{b_2} = h_2$$



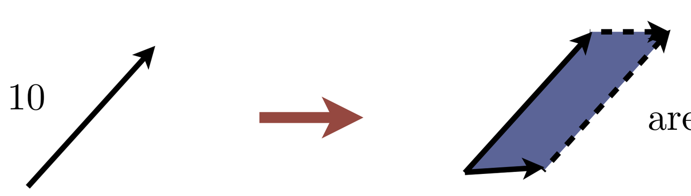
$h_1 \leq h_2$

MONOTONICITY

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

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

length = 10



area = 5

PRIOR WORK

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

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

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

CHEKURI ET AL. 2011

Step 1: Relax to [0, 1]

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

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

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

Step 2: Extend objective

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

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

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 $x_i \in \{0, 1\}$

SOFTMAX EXTENSION

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

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

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

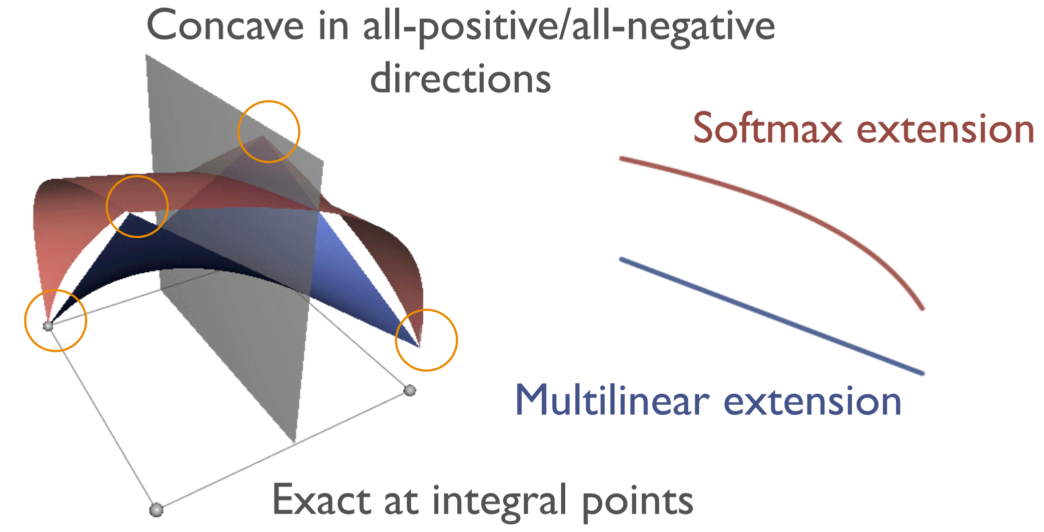
$$\frac{\partial}{\partial x_i} \tilde{F}(\mathbf{x}) = \mathrm{tr}((\mathrm{diag}(\mathbf{x})(L - I) + I)^{-1}(L - I)_i)$$

APPROXIMATION GUARANTEE

Lemma: When $\mathbf{u}, \mathbf{v} \geq \mathbf{0}$, we have $\frac{\partial^2}{\partial s \partial t} \tilde{F}(\mathbf{x} + s\mathbf{u} + t\mathbf{v}) \leq 0$ whenever $\mathbf{0} \leq \mathbf{x} + s\mathbf{u} + t\mathbf{v} \leq \mathbf{1}$.

Corollary: $\tilde{F}(\mathbf{x} + t\mathbf{v})$ is concave along any direction $\mathbf{v} \geq \mathbf{0}$.

Concave in all-positive/all-negative directions



Exact at integral points

Lemma: If $\mathbf{x}$ is a local optimum of $\tilde{F}(\cdot)$, then for any $\mathbf{y} \in [0, 1]^N$,

$$2\tilde{F}(\mathbf{x}) \geq \tilde{F}(\mathbf{x} \vee \mathbf{y}) + \tilde{F}(\mathbf{x} \wedge \mathbf{y}),$$

where $(\mathbf{x} \vee \mathbf{y})_i = \max(x_i, y_i)$ and $\mathbf{x} \wedge \mathbf{y})_i = \min(x_i, y_i)$.

Theorem: Let $\tilde{F}(\mathbf{x})$ be the softmax extension of a nonnegative submodular function $f(Y) = \log\det(L_Y)$, let S be the polytope $[0, 1]^N$, let $OPT = \max_{\mathbf{x}' \in S}$, and let $\mathbf{x}$ and $\mathbf{y}$ be local optima of $\tilde{F}$ in S and $S \cap \{\mathbf{y}' \mid \mathbf{y}' \leq (\mathbf{1} - \mathbf{x})\}$, respectively. Then:

$$\max(\tilde{F}(\mathbf{x}), \tilde{F}(\mathbf{y})) \geq \tfrac{1}{4} OPT \geq \tfrac{1}{4} \max_{Y \in S} \log\det(L_Y).$$

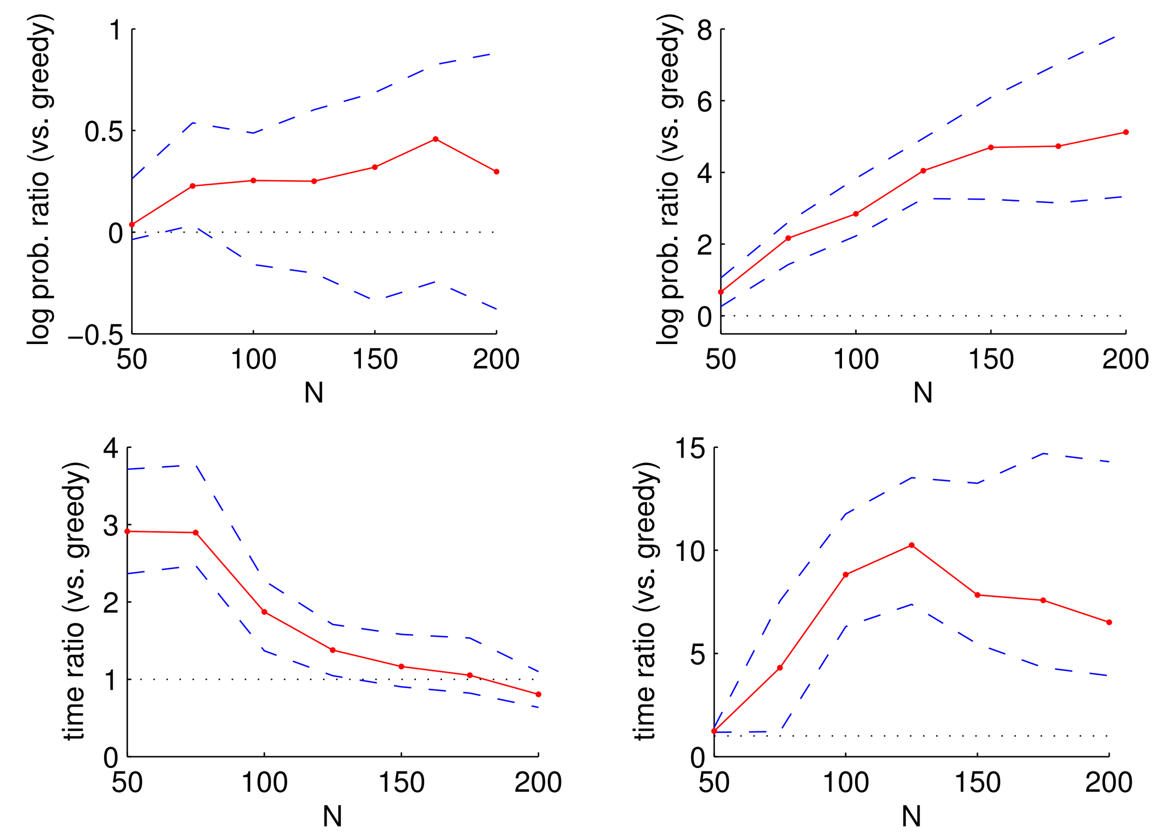
Theorem: If $S = [0, 1]^N$, then for any local optimum $\mathbf{x}$ of $\tilde{F}$, either $\mathbf{x}$ is integral or at least one fractional coordinate x_i can be set to 0 or 1 without lowering the objective.

Summary: concavity in positive directions + submodularity = 1/4-approx sans constraints

No guarantees for constrained setting for softmax, but in practice pipage rounding and thresholding work well.

SYNTHETIC EXPERIMENTS

Unconstrained Constrained



MATCHED SUMMARIZATION

20 Republican primary debates

Average of 179 quotes per candidate

Task: Given statements made by candidate A and statements made by candidate B, select a set of ***pairs*** such that the two elements ***within*** a pair are ***similar***, but the ***set*** of pairs is ***diverse***.

- **Romney 1: No tax on interest, dividends, or capital gains.**
- **Romney 2: We're not going to have Sharia law applied in U.S. courts.**
- **Romney 3: I will ... grant a waiver from Obamacare to all 50 states.**
- **Romney 4: We're spending more on foreign aid than we ought to.**
- **Romney 5: If you think what we did in Massachusetts and what President Obama did are the same, boy, take a closer look.**

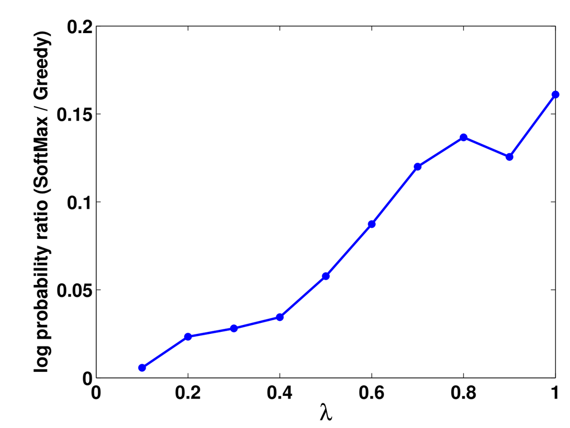
- **Santorum 1: I don't believe in a zero capital gains tax rate.**
- **Santorum 2: Manufacture in America, you aren't going to pay any taxes.**
- **Santorum 3: Zeroing out foreign aid ... that's absolutely the wrong course.**
- **Santorum 4: I voted against ethanol subsidies my entire time in Congress.**
- **Santorum 5: Obamacare ... is going to blow a hole in the budget.**

 +  =

Matched summary

R1 & S1, R3 & S5, R4 & S3

Code + Data:
www.seas.upenn.edu/~jengi/dpp-map.html



$\propto$ set size