Composable Coresets for Constrained Determinant Maximization and Beyond

Sepideh Mahabadi

(Microsoft Research)

Thuy Duong Vuong

(University of California San Diego)

Goal

Design **Composable Coresets**

for **Determinant Maximization**

under **Partition Constraint**

Coresets [Agarwal, Har-Peled, Varadarajan'05]

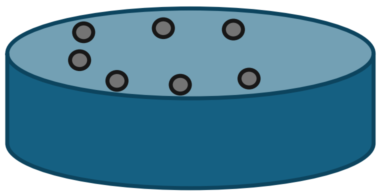
Coreset: a small subset $\boldsymbol{U}$ of the data $\boldsymbol{V}$ that represents it well.

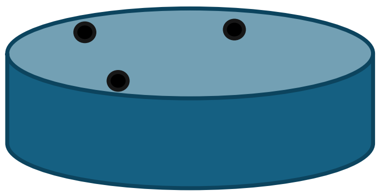
Coresets [Agarwal, Har-Peled, Varadarajan'05]

Coreset: a small subset U of the data V that represents it well.

Solving the problem over dataset V (approximately)

Solving the problem over coreset U





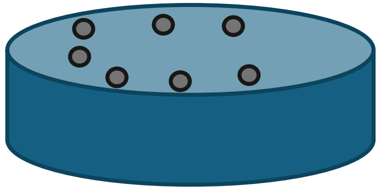
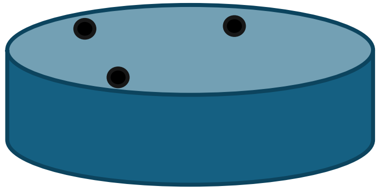
Coresets [Agarwal, Har-Peled, Varadarajan'05]

α –**Coreset:** a small subset U of the data V that represents it well.

- Task specific

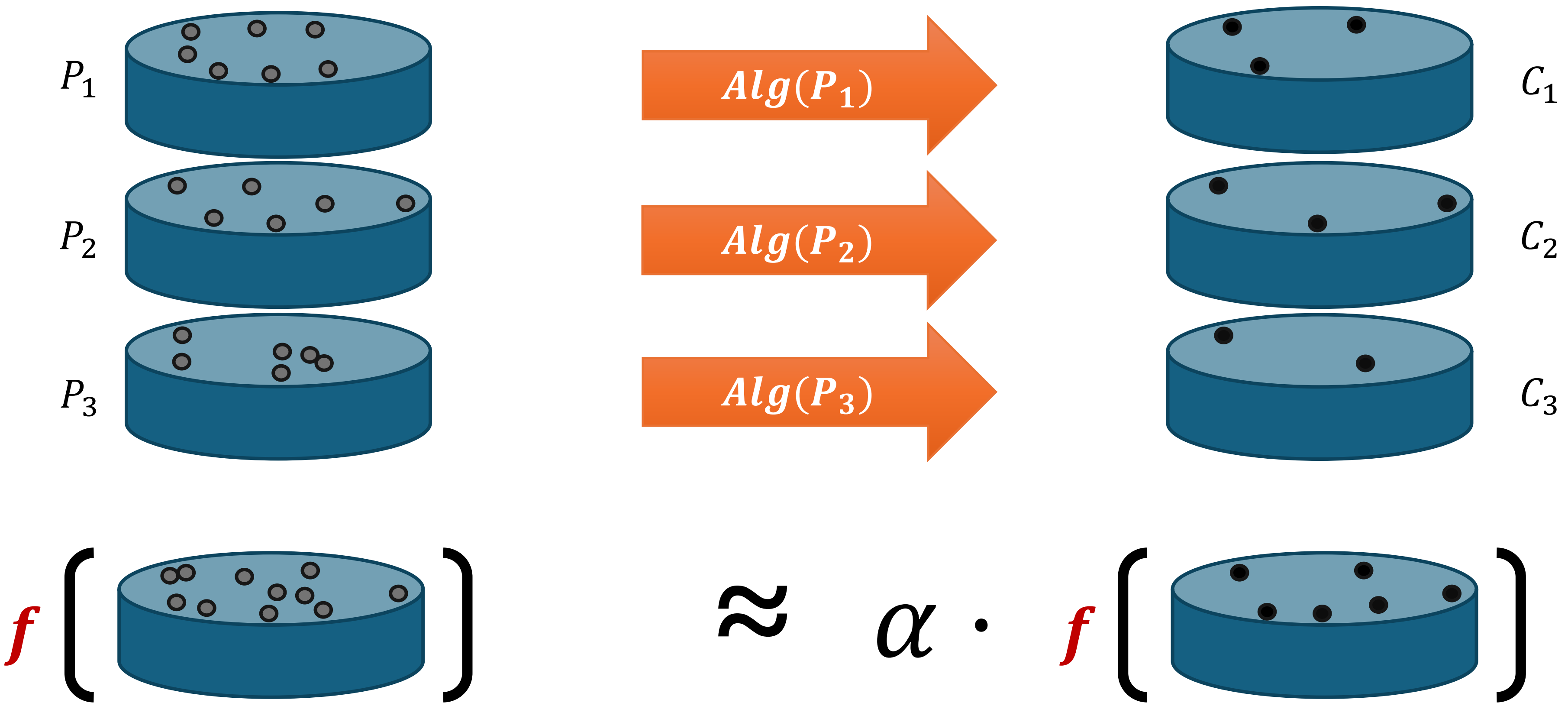
Solving the problem over dataset V (approximately)

Solving the problem over coreset U

$f($  $)$ $\approx$ $\alpha \cdot$ $f($  $)$

Composable Coresets [Indyk, M., Mahdian, Mirrokni'14]

- **Coresets with composability property:** "The union of coresets is a coreset for the union"
- Summarization Algorithm



Goal

Design **Composable Coresets**

for **Determinant Maximization**

under **Partition Constraint**

The Determinant Maximization problem

Input: a set of n points P in $\mathbb{R}^d$ and a parameter k,

Goal: pick a subset $S \subseteq P$ of k points while maximizing their determinant

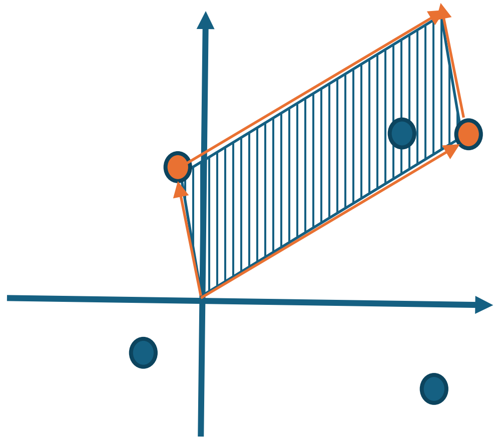
The Determinant Maximization problem

Input: a set of n points P in $\mathbb{R}^d$ and a parameter k,

Goal: pick a subset $S \subseteq P$ of k points while maximizing their determinant

- For $\boldsymbol{k} \leq \boldsymbol{d}$, this is their volume

$\boldsymbol{k} = \mathbf{2}$



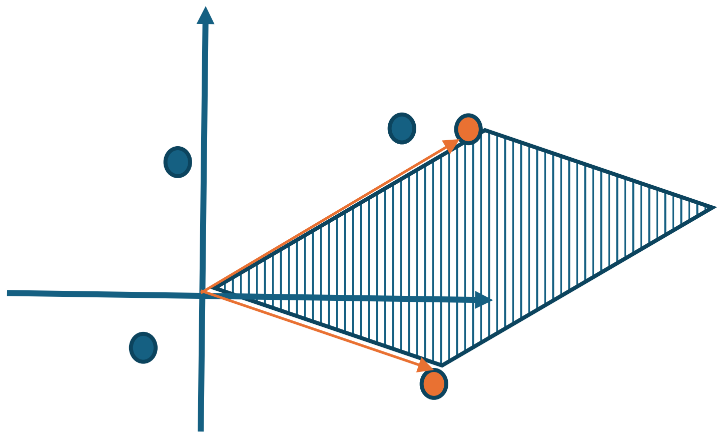
The Determinant Maximization problem

Input: a set of n points P in $\mathbb{R}^d$ and a parameter k,

Goal: pick a subset $S \subseteq P$ of k points while maximizing their determinant

- For $\boldsymbol{k \leq d}$, this is their volume

$\boxed{\boldsymbol{k = 2}}$



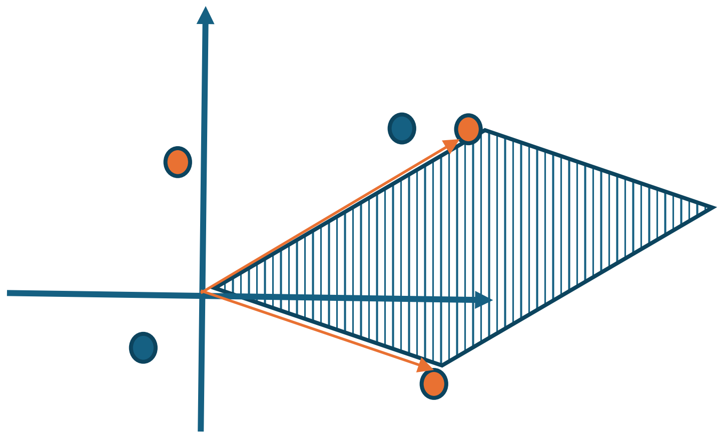
The Determinant Maximization problem

Input: a set of n points P in $\mathbb{R}^d$ and a parameter k,

Goal: pick a subset $S \subseteq P$ of k points while maximizing their determinant

- For $\boldsymbol{k \leq d}$, this is their volume

$\boxed{\boldsymbol{k = 3, d = 2}}$



- For $\boldsymbol{k > d}$, this is the sum of the volumes of all $\binom{\boldsymbol{k}}{\boldsymbol{d}}$ subsets.

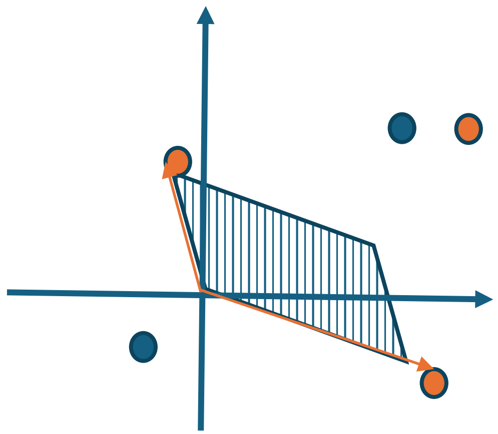
The Determinant Maximization problem

Input: a set of n points P in $\mathbb{R}^d$ and a parameter k,

Goal: pick a subset $S \subseteq P$ of k points while maximizing their determinant

- For $\boldsymbol{k \leq d}$, this is their volume

$\boldsymbol{k = 3, d = 2}$



- For $\boldsymbol{k > d}$, this is the sum of the volumes of all $\binom{\boldsymbol{k}}{\boldsymbol{d}}$ subsets.

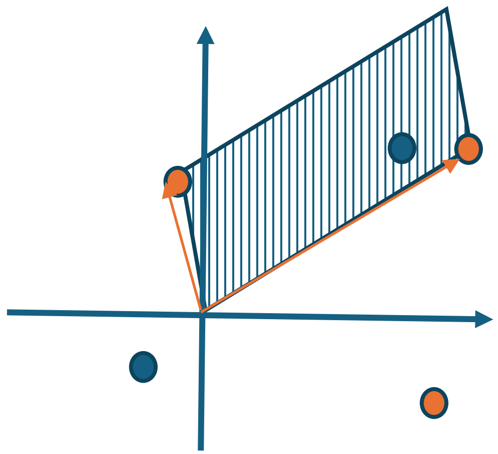
The Determinant Maximization problem

Input: a set of n points P in $\mathbb{R}^d$ and a parameter k,

Goal: pick a subset $S \subseteq P$ of k points while maximizing their determinant

- For $\boldsymbol{k \leq d}$, this is their volume

$\boldsymbol{k = 3, d = 2}$



- For $\boldsymbol{k > d}$, this is the sum of the volumes of all $\binom{\boldsymbol{k}}{\boldsymbol{d}}$ subsets.

Applications

- Determinant is used as a measure for **Diversity** of a set of objects.

Goal

Design **Composable Coresets**

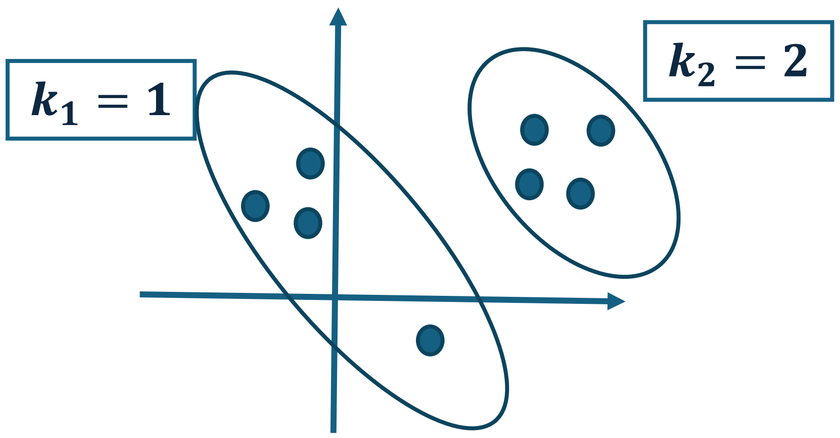
for **Determinant Maximization**

under **Partition Constraint**

Constrained Determinant Maximization

Input:

- sets of vectors $P_1, \cdots, P_m$, $P = \bigcup_i P_i$
- and $k_1, \cdots, k_m \leq d$, $k = \sum_i k_i$

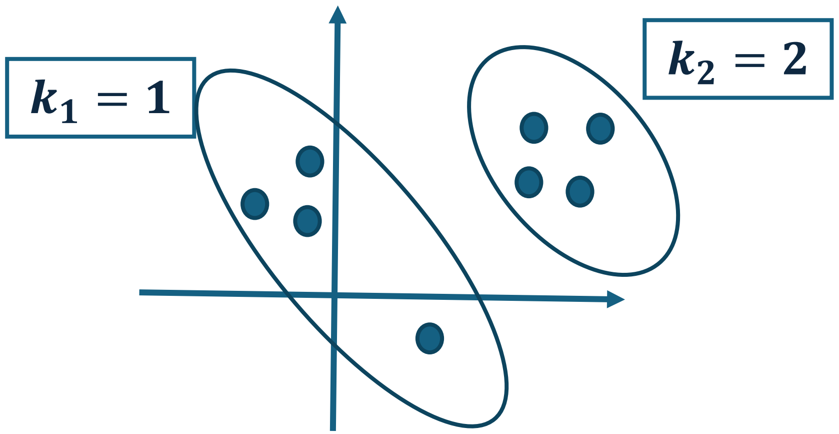


Constrained Determinant Maximization

Input:

- sets of vectors $P_1, \cdots, P_m$, $P = \bigcup_i P_i$
- and $k_1, \cdots, k_m \leq d$, $k = \sum_i k_i$

Goal: pick k_i points $S_i \subset P_i$ s.t. the determinant of the picked points $S = \bigcup_i S_i$ is maximized.

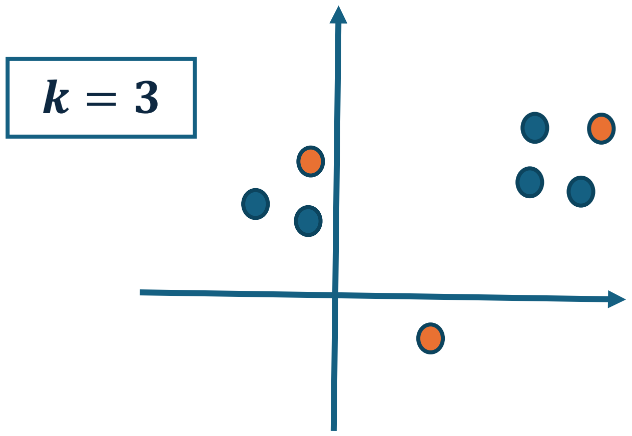


Constrained Determinant Maximization

Input:

- sets of vectors $P_1, \cdots, P_m$, $P = \bigcup_i P_i$
- and $k_1, \cdots, k_m \leq d$, $k = \sum_i k_i$

Goal: pick k_i points $S_i \subset P_i$ s.t. the determinant of the picked points $S = \bigcup_i S_i$ is maximized.

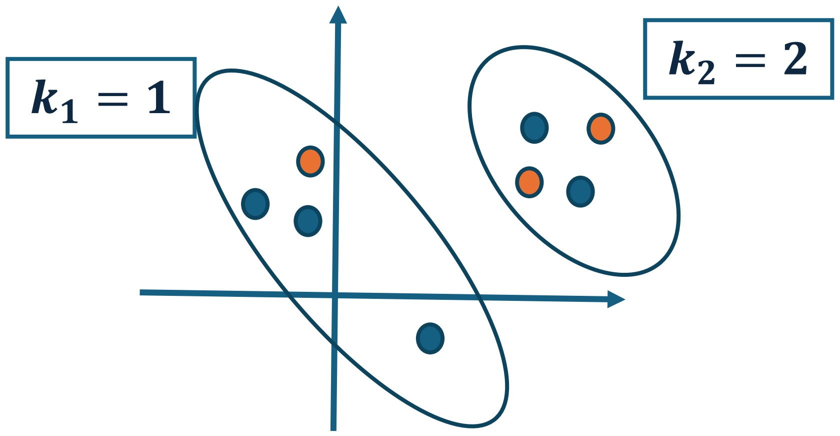


Constrained Determinant Maximization

Input:

- sets of vectors $P_1, \cdots, P_m$, $P = \bigcup_i P_i$
- and $k_1, \cdots, k_m \leq d,\ k = \sum_i k_i$

Goal: pick k_i points $S_i \subset P_i$ s.t. the determinant of the picked points $S = \bigcup_i S_i$ is maximized.



Applications

- Determinant is used as a measure for Diversity of a set of objects.

- Partition Constraint used for assigning **budgets** to different classes of the input. For example,

Applications

- Determinant is used as a measure for Diversity of a set of objects.
- Partition Constraint used for assigning **budgets** to different classes of the input. For example,
    1. Different genres of movies in a movie recommendation system, e.g. comedy, action, drama.

Applications

- Determinant is used as a measure for Diversity of a set of objects.
- Partition Constraint used for assigning **budgets** to different classes of the input. For example,
    1. Different genres of movies in a movie recommendation system, e.g. comedy, action, drama.
    2. Modeling Recency in feed summarization [M., Trajanovski'23]:
        - Each message has a posted time, and the goal is to show more recent messages and less old ones. We still need diversity.
        - Divide the messages in a month into four groups based on the week they have been posted
        - Set k_i to be higher for more recent weeks
    - ...

Results

1. When $\boldsymbol{k \leq d}$, get a coreset of **size $\boldsymbol{mk}$** and **approximation factor $\boldsymbol{O(k^{2k})}$,**
   - Optimal size
   - Optimal approximation factor

Results

1. When $\boldsymbol{k \leq d}$, get a coreset of **size $\boldsymbol{mk}$** and **approximation factor $\boldsymbol{O(k^{2k})}$,**
    - Optimal size
    - Optimal approximation factor
2. When $\boldsymbol{k \geq d}$, get a coreset of **size $\boldsymbol{kd}$** with **approximation factor $\boldsymbol{O(d^{2d})}$,**
    - Peeling algorithm
    - Optimal approximation factor
    - Size lower bound of $\boldsymbol{k + d(d-1)}$ (optimal for $\boldsymbol{k = O(d)}$)

Results

1. When $\boldsymbol{k \le d}$, get a coreset of **size** $\boldsymbol{mk}$ and **approximation factor** $\boldsymbol{O(k^{2k})}$,
   - Optimal size
   - Optimal approximation factor
2. When $\boldsymbol{k \ge d}$, get a coreset of **size** $\boldsymbol{kd}$ with **approximation factor** $\boldsymbol{O(d^{2d})}$,
   - Peeling algorithm
   - Optimal approximation factor
   - Size lower bound of $\boldsymbol{k + d(d-1)}$ (optimal for $\boldsymbol{k = O(d)}$)

- Get new results for the **unconstrained version** of the problem.
- Generalize to other **experimental design** functions
- Generalize to get new results for **Laminar Matroids**
- Generalize to all **Strongly Rayleigh distributions**, obtained via an improved exchange inequality.

Results

1. When $k \leq d$, get a coreset of **size mk** and **approximation factor $O(k^{2k})$,**
   - Optimal size
   - Optimal approximation factor
2. When $k \geq d$, get a coreset of **size kd** with **approximation factor $O(d^{2d})$,**
   - Peeling algorithm
   - Optimal approximation factor
   - Size lower bound of $k + d(d-1)$ (optimal for $k = O(d)$)

➢ Get new results for the **unconstrained version** of the problem.

➢ Generalize to other **experimental design** functions

➢ Generalize to get new results for **Laminar Matroids**

➢ Generalize to all **Strongly Rayleigh distributions**, obtained via an improved exchange inequality.

❑ Use to improve the runtime of local-search based algorithm of [Anari-Vuong–COLT'22] for determinant maximization under partition constraint from $O(n^{2^m} k^{2^m})$ to $O(nk^{2^m})$

Thank you!

	Size	Approx.	Constraint
$k \leq d$	k	$O(k)^{2k}$	Cardinality [Indyk et al. 2020, Mahabadi et al. 2019]
$k \geq d$	$\tilde{O}(d)$	$\tilde{O}(d)^{2d}$	Cardinality (with repetition) [Indyk et al. 2020]
$k \geq d$	kd	d^{2d}	Cardinality (without repetition) **This work**
$k \leq d$	mk	k^{2k}	Partition, **This work**
$k \geq d$	kd	d^{2d}	Partition, **This work**
$k \leq d$	k^{2k}	k^{2k}	Laminar, **This work**
$k \geq d$	$(kd)^k$	d^{2d}	Laminar, **This work**