Efficient approximate sampling of projection determinantal point processes

Guillaume Gautier[1,2], Rémi Bardenet[1], and Michal Valko[2]

[1]CNRS & CRIStAL, Univ. Lille
[2]Inria Lille

Motivation

Image search/recommendation systems

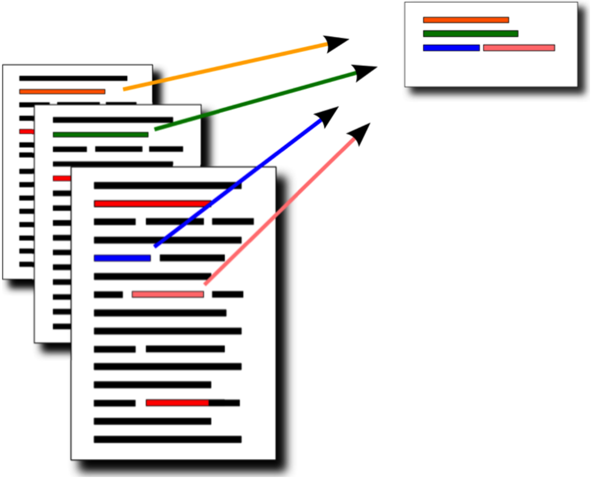
relevance

relevance
+
diversity

'bolt' query

Motivation

Extractive text summarization (Kulesza & Taskar, 2012)



Definition

- $\{1, \ldots, N\}$ indices/labels of items
  - images
  - sentences
  - edges of a graph
- $\mathrm{DPP}(\mathbf{K})$ a measure on subsets of $\{1, \ldots, N\}$
- $\mathbf{K}$ a PSD similarity kernel
- $\mathcal{X} \sim \mathrm{DPP}(\mathbf{K})$ if $\forall S \subseteq \{1, \ldots, N\}$,

$$\mathbb{P}\left[S \subseteq \mathcal{X}\right] = \det \mathbf{K}_S$$

- Existence is guaranteed when $\mathbf{0}_N \preceq \mathbf{K} \preceq \mathbf{I}_N$

Projection DPPs

- $\mathbf{K}$ is an orthogonal projection matrix
  - $\operatorname{Spec} \mathbf{K} \in \{0,1\}^N$

Projection DPPs

- $\mathbf{K}$ is an orthogonal projection matrix
  - $\operatorname{Spec} \mathbf{K} \in \{0,1\}^N$
- $|\mathcal{X}| \overset{a.s.}{=} \operatorname{Tr} \mathbf{K} = \operatorname{rank} \mathbf{K} = r$
  - Summaries made of r sentences
  - Bags of r images

Projection DPPs

- $\mathbf{K}$ is an orthogonal projection matrix
  - $\operatorname{Spec}\mathbf{K} \in \{0,1\}^N$
- $|\mathcal{X}| \overset{a.s.}{=} \operatorname{Tr}\mathbf{K} = \operatorname{rank}\mathbf{K} = r$
  - Summaries made of r sentences
  - Bags of r images
- Gram matrix

$$\mathbf{K} = \sum_{i=1}^{r} u^{(i)} u^{(i)\mathsf{T}} = \mathbf{\Phi}^{\mathsf{T}}\mathbf{\Phi}$$

with $\varphi_n = \left(u_n^{(1)}, \dots, u_n^{(r)}\right)^{\mathsf{T}}$

Projection DPPs

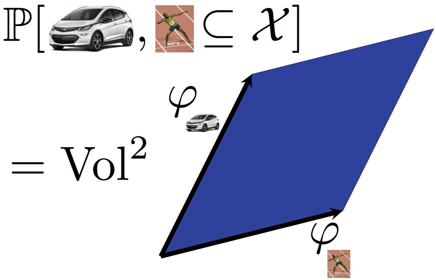
- $\mathbf{K}$ is an orthogonal projection matrix
  - $\operatorname{Spec} \mathbf{K} \in \{0,1\}^N$
- $|\mathcal{X}| \overset{a.s.}{=} \operatorname{Tr} \mathbf{K} = \operatorname{rank} \mathbf{K} = r$
  - Summaries made of r sentences
  - Bags of r images
- Gram matrix

$$\mathbf{K} = \sum_{i=1}^{r} u^{(i)} {u^{(i)}}^{\mathsf{T}} = \mathbf{\Phi}^{\mathsf{T}} \mathbf{\Phi}$$

with $\varphi_n = \left(u_n^{(1)}, \dots, u_n^{(r)}\right)^{\mathsf{T}}$

- Geometrical interpretation

$$\mathbb{P}\left[S \subseteq \mathcal{X}\right] = \det \mathbf{K}_S$$
$$= \operatorname{Vol}^2 \{\varphi_i \,;\, i \in S\}$$



Diversity

- Negative association

$$\begin{aligned}\mathbb{P}\left[\{i,j\} \subseteq \mathcal{X}\right] &= \begin{vmatrix}\mathbb{P}\left[i \in \mathcal{X}\right] & \mathbf{K}_{ij} \\ \mathbf{K}_{ij} & \mathbb{P}\left[j \in \mathcal{X}\right]\end{vmatrix} \\ &= \mathbb{P}\left[i \in \mathcal{X}\right]\mathbb{P}\left[j \in \mathcal{X}\right] - \mathbf{K}_{ij}^2 \\ &\leq \mathbb{P}\left[i \in \mathcal{X}\right]\mathbb{P}\left[j \in \mathcal{X}\right]\end{aligned}$$

Diversity

- Negative association

$$\mathbb{P}\left[\{i,j\} \subseteq \mathcal{X}\right] = \begin{vmatrix} \mathbb{P}\left[i \in \mathcal{X}\right] & \mathbf{K}_{ij} \\ \mathbf{K}_{ij} & \mathbb{P}\left[j \in \mathcal{X}\right] \end{vmatrix}$$
$$= \mathbb{P}\left[i \in \mathcal{X}\right] \mathbb{P}\left[j \in \mathcal{X}\right] - \mathbf{K}_{ij}^2$$
$$\leq \mathbb{P}\left[i \in \mathcal{X}\right] \mathbb{P}\left[j \in \mathcal{X}\right]$$

- The larger $|\mathbf{K}_{ij}|$ the smaller $\mathbb{P}\left[\{i,j\} \subseteq \mathcal{X}\right]$
  - Diversity/repulsion
  - $|\mathbf{K}_{ij}|$ yields departure from independence

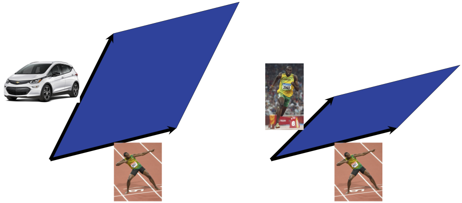
Diversity

- Negative association

$$\mathbb{P}[\{i,j\} \subseteq \mathcal{X}] = \begin{vmatrix} \mathbb{P}[i \in \mathcal{X}] & \mathbf{K}_{ij} \\ \mathbf{K}_{ij} & \mathbb{P}[j \in \mathcal{X}] \end{vmatrix}$$
$$= \mathbb{P}[i \in \mathcal{X}]\,\mathbb{P}[j \in \mathcal{X}] - \mathbf{K}_{ij}^2$$
$$\leq \mathbb{P}[i \in \mathcal{X}]\,\mathbb{P}[j \in \mathcal{X}]$$

- The larger $|\mathbf{K}_{ij}|$ the smaller $\mathbb{P}[\{i,j\} \subseteq \mathcal{X}]$
  - Diversity/repulsion
  - $|\mathbf{K}_{ij}|$ yields departure from independence

$\mathbb{P}[$  ,  $\subseteq \mathcal{X}] \geq \mathbb{P}[$  ,  $\subseteq \mathcal{X}]$



Setup

- Build the $r \times N$ feature matrix

$$\mathbf{A} = \left(\sqrt{q_1}\phi_1 | \dots | \sqrt{q_N}\phi_N\right)$$

  - If $\|\phi_i\|^2 = 1$, 'angles' encode diversity
  - q_i measures relevance of item i

Setup

- Build the $r \times N$ feature matrix

$$\mathbf{A} = \left(\sqrt{q_1}\phi_1 | \dots | \sqrt{q_N}\phi_N\right)$$

  - If $\|\phi_i\|^2 = 1$, 'angles' encode diversity
  - q_i measures relevance of item i

Assumption

$\mathbf{A}$ *is full row rank i.e.* $\operatorname{rank}\mathbf{A} = r$

Setup

- Build the $r \times N$ feature matrix

$$\mathbf{A} = \left(\sqrt{q_1}\phi_1 | \dots | \sqrt{q_N}\phi_N\right)$$

  - If $\|\phi_i\|^2 = 1$, 'angles' encode diversity
  - q_i measures relevance of item i

- Build the $N \times N$ projection kernel $\mathbf{K} = \mathbf{A}^\intercal \left[\mathbf{A}\mathbf{A}^\intercal\right]^{-1} \mathbf{A}$

Assumption

$\mathbf{A}$ *is full row rank i.e.* $\operatorname{rank} \mathbf{A} = r$

Setup

- Build the $r \times N$ feature matrix

$$\mathbf{A} = \left(\sqrt{q_1}\phi_1 | \dots | \sqrt{q_N}\phi_N\right)$$

  - If $\|\phi_i\|^2 = 1$, 'angles' encode diversity
  - q_i measures relevance of item i

- Build the $N \times N$ projection kernel $\mathbf{K} = \mathbf{A}^\top \left[\mathbf{A}\mathbf{A}^\top\right]^{-1} \mathbf{A}$

- If $\mathcal{X} \sim \mathrm{DPP}(\mathbf{K})$,
  - $|\mathcal{X}| \overset{a.s.}{=} r$
  - For $B = \{i_1, \dots, i_r\}$,

$$\mathbb{P}\left[\mathcal{X} = B\right] \propto \left|\det \mathbf{A}_{:B}\right|^2 = \mathrm{Vol}^2 \left\{\sqrt{q_i}\phi_i \, ; \, i \in B\right\}$$

Assumption

$\mathbf{A}$ *is full row rank i.e.* $\operatorname{rank} \mathbf{A} = r$

Setup

- Build the $r \times N$ feature matrix

$$\mathbf{A} = \left(\sqrt{q_1}\phi_1 | \dots | \sqrt{q_N}\phi_N\right)$$

  - If $\|\phi_i\|^2 = 1$, 'angles' encode diversity
  - q_i measures relevance of item i

Assumption

$\mathbf{A}$ *is full row rank i.e.* $\operatorname{rank} \mathbf{A} = r$

- Build the $N \times N$ projection kernel $\mathbf{K} = \mathbf{A}^{\mathsf{T}} \left[\mathbf{A}\mathbf{A}^{\mathsf{T}}\right]^{-1} \mathbf{A}$

- If $\mathcal{X} \sim \operatorname{DPP}(\mathbf{K})$,
  - $|\mathcal{X}| \overset{a.s.}{=} r$
  - For $B = \{i_1, \dots, i_r\}$,

$$\mathbb{P}\left[\mathcal{X} = B\right] \propto \left|\det \mathbf{A}_{:B}\right|^2 = \operatorname{Vol}^2 \left\{\sqrt{q_i}\phi_i \,;\, i \in B\right\}$$

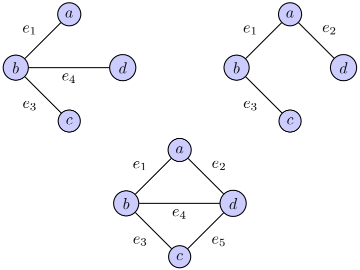
- $\operatorname{DPP}(\mathbf{K})$ has support

$$\mathcal{B} \triangleq \left\{B \,;\, |B| = r, \det \mathbf{A}_{:B} \neq 0\right\}$$

i.e. collection of columns of $\mathbf{A}$ forming a basis

One example

Uniform spanning trees (Lyons, 2003)

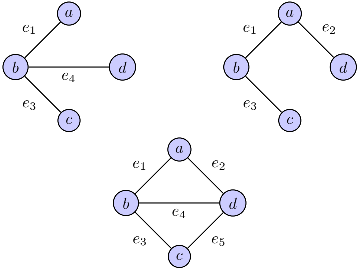


Vertex-edge incidence matrix

$$
\begin{array}{c} \\ a \\ b \\ c \\ d \end{array}
\begin{array}{c} \begin{array}{ccccc} e_1 & e_2 & e_3 & e_4 & e_5 \end{array} \\
\left[\begin{array}{ccccc} -1 & -1 & 0 & 0 & 0 \\ 1 & 0 & -1 & -1 & 0 \\ 0 & 0 & 1 & 0 & -1 \\ 0 & 1 & 0 & 1 & 1 \end{array}\right] \end{array}
$$

One example

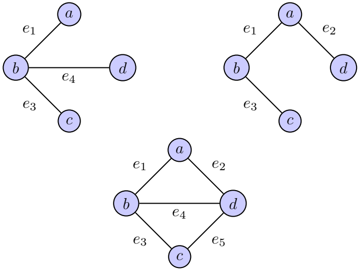
Uniform spanning trees (Lyons, 2003)



$$\mathbf{A} = \begin{array}{c} \\ a \\ b \\ c \\ d \end{array} \begin{bmatrix} e_1 & e_2 & e_3 & e_4 & e_5 \\ -1 & -1 & 0 & 0 & 0 \\ 1 & 0 & -1 & -1 & 0 \\ 0 & 0 & 1 & 0 & -1 \\ 0 & 1 & 0 & 1 & 1 \end{bmatrix}$$

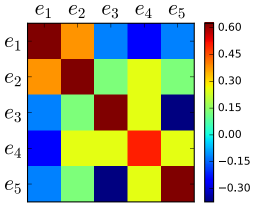
One example

Uniform spanning trees (Lyons, 2003)



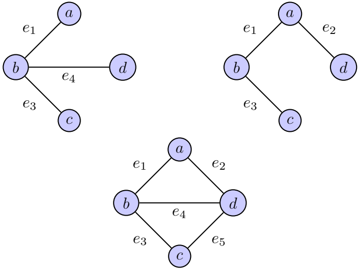
$$\mathbf{A} = \begin{array}{c} \\ a \\ b \\ c \\ d \end{array} \begin{array}{c} \begin{array}{ccccc} e_1 & e_2 & e_3 & e_4 & e_5 \end{array} \\ \begin{bmatrix} -1 & -1 & 0 & 0 & 0 \\ 1 & 0 & -1 & -1 & 0 \\ 0 & 0 & 1 & 0 & -1 \\ 0 & 1 & 0 & 1 & 1 \end{bmatrix} \end{array}$$

$$\mathbf{K} = \mathbf{A}^\top \left[\mathbf{A}\mathbf{A}^\top\right]^{-1} \mathbf{A}$$



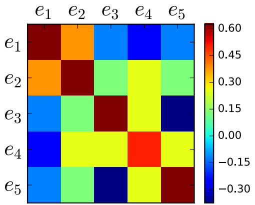
One example

Uniform spanning trees (Lyons, 2003)



$$\mathbf{A} = \begin{array}{c} \\ a \\ b \\ c \\ d \end{array} \begin{array}{c} \begin{array}{ccccc} e_1 & e_2 & e_3 & e_4 & e_5 \end{array} \\ \left[\begin{array}{ccccc} -1 & -1 & 0 & 0 & 0 \\ 1 & 0 & -1 & -1 & 0 \\ 0 & 0 & 1 & 0 & -1 \\ 0 & 1 & 0 & 1 & 1 \end{array}\right] \end{array}$$

$$\mathbf{K} = \mathbf{A}^\top \left[\mathbf{A}\mathbf{A}^\top\right]^{-1} \mathbf{A}$$



- DPP($\mathbf{K}$) is a measure on the edge set of G
  - $\mathcal{B}$ = collection of spanning trees of G
- $\mathbb{P}\left[\mathcal{X} = B\right] = \dfrac{\left|\det \mathbf{A}_{:B}\right|^2}{\det \mathbf{A}\mathbf{A}^\top} = \dfrac{1}{|\mathcal{B}|} \mathbb{1}_{B \in \mathcal{B}}$

Rk: Unimodularity, transfer current matrix, matrix tree theorem, Laplacian solvers

Exact sampling

- From $r \times N$ feature matrix $\mathbf{A} = \left(\phi_1 | \dots | \phi_N\right)$
- To $N \times N$ projection kernel $\mathbf{K} = \mathbf{A}^\top \left[\mathbf{A}\mathbf{A}^\top\right]^{-1} \mathbf{A}$

Exact sampling (Hough et al., 2006; Kulesza & Taskar, 2012)

Sample $\mathcal{X} \sim \mathrm{DPP}(\mathbf{K})$

- Marginals
$$\mathbb{P}\left[\mathcal{X} = B\right] = \det \mathbf{K}_B$$
- Chain rule, $J = \{i_1, \dots, i_k\}$
$$\mathbb{P}\left[i_{k+1} = i | J\right] \propto \mathbf{K}_{ii} - \mathbf{K}_{i,J} {\mathbf{K}_J}^{-1} \mathbf{K}_{J,i}$$
- Costly: eigen-decomposition + Gram-Schmidt = $\mathcal{O}(N^3 + Nr^2)$

Rk: Uniform spanning trees (Aldous, 1990; Propp & Wilson, 1998), generation of mazes

Approximate sampling - 1

- From $r \times N$ feature matrix $\mathbf{A} = \left(\phi_1 | \dots | \phi_N\right)$

Approximate sampling (Anari et al., 2016; Li et al., 2016)

Build a Markov chain, $\mathbf{B} \triangleq \mathbf{A}_{:B}$

- State space $\mathcal{B} \triangleq \{B\,;\, \det \mathbf{B} \neq 0\}$
- Stationary distribution

$$\propto |\det \mathbf{B}|^2 = \mathrm{Vol}^2 \{\phi_i\,;\, i \in B\} \cdot \mathbb{1}_{B \in \mathcal{B}}$$

- Basis-exchange graph
  - $B \leftrightarrow B' = (B \setminus \{i\}) \cup \{j\}$
  - Full analysis: polynomial mixing time
  - Local and correlated moves on $\mathcal{B}$

Rk: Basis-exchange graph (Feder & Mihail, 1992), matroid theory (Oxley, 2003)

Approximate sampling - 1

But a dream within a dream?
Is all that we see or seem
One from the pitiless wave?
O God! can I not save
Them with a tighter clasp?
O God! can I not grasp
While I weep--while I weep!
Through my fingers to the deep,
How few! yet how they creep
Grains of the golden sand-
And I hold within my hand
Of a surf-tormented shore,
I stand amid the roar
Is but a dream within a dream.
All that we see or seem
Is it therefore the less gone?
In a vision, or in none,
In a night, or in a day,
Yet if hope has flown away
That my days have been a dream;
You are not wrong, who deem
Thus much let me avow-
And, in parting from you now,
Take this kiss upon the brow!

Approximate sampling - 1

But a dream within a dream?
Is all that we see or seem
One from the pitiless wave?
O God! can I not save
Them with a tighter clasp?
O God! can I not grasp
While I weep--while I weep!
Through my fingers to the deep,
How few! yet how they creep
Grains of the golden sand-
And I hold within my hand
Of a surf-tormented shore,
I stand amid the roar
Is but a dream within a dream.
All that we see or seem
Is it therefore the less gone?
In a vision, or in none,
In a night, or in a day,
Yet if hope has flown away
That my days have been a dream;
You are not wrong, who deem
Thus much let me avow-
And, in parting from you now,
Take this kiss upon the brow!

Approximate sampling - 1

But a dream within a dream?
Is all that we see or seem
One from the pitiless wave?
O God! can I not save
Them with a tighter clasp?
O God! can I not grasp
While I weep--while I weep!
Through my fingers to the deep,
How few! yet how they creep
Grains of the golden sand-
And I hold within my hand
Of a surf-tormented shore,
I stand amid the roar
Is but a dream within a dream.
All that we see or seem
Is it therefore the less gone?
In a vision, or in none,
In a night, or in a day,
Yet if hope has flown away
That my days have been a dream;
You are not wrong, who deem
Thus much let me avow-
And, in parting from you now,
Take this kiss upon the brow!

Approximate sampling - 1

But a dream within a dream?
Is all that we see or seem
One from the pitiless wave?
O God! can I not save
Them with a tighter clasp?
O God! can I not grasp
While I weep--while I weep!
Through my fingers to the deep,
How few! yet how they creep
Grains of the golden sand-
And I hold within my hand
Of a surf-tormented shore,
I stand amid the roar
Is but a dream within a dream.
All that we see or seem
Is it therefore the less gone?
In a vision, or in none,
In a night, or in a day,
Yet if hope has flown away
That my days have been a dream;
You are not wrong, who deem
Thus much let me avow-
And, in parting from you now,
Take this kiss upon the brow!

Approximate sampling - 2

- From $r \times N$ feature matrix $\mathbf{A} = \left(\phi_1 | \dots | \phi_N\right)$

Approximate sampling (G., Bardenet & Valko, 2017)

Build a Markov chain, $\mathbf{B} \triangleq \mathbf{A}_{:B}$

- State space $\mathcal{B} \triangleq \{B\,;\, \det \mathbf{B} \neq 0\}$
- Stationary distribution

$$\propto \left|\det \mathbf{B}\right|^2 = \mathrm{Vol}^2 \{\phi_i\,;\, i \in B\} \cdot \mathbb{1}_{B \in \mathcal{B}}$$

- Wander in a continuous embedding of $\mathcal{B}$
  - Geometrical representation of $\mathcal{B}$
  - More decorelated moves, empirically faster mixing

Approximate sampling - 2

But a dream within a dream?
Is all that we see or seem
One from the pitiless wave?
O God! can I not save
Them with a tighter clasp?
O God! can I not grasp
While I weep--while I weep!
Through my fingers to the deep,
How few! yet how they creep
Grains of the golden sand-
And I hold within my hand
Of a surf-tormented shore,
I stand amid the roar
Is but a dream within a dream.
All that we see or seem
Is it therefore the less gone?
In a vision, or in none,
In a night, or in a day,
Yet if hope has flown away
That my days have been a dream;
You are not wrong, who deem
Thus much let me avow-
And, in parting from you now,
Take this kiss upon the brow!

Approximate sampling - 2

But a dream within a dream?
Is all that we see or seem
One from the pitiless wave?
O God! can I not save
Them with a tighter clasp?
O God! can I not grasp
While I weep--while I weep!
Through my fingers to the deep,
How few! yet how they creep
Grains of the golden sand-
And I hold within my hand
Of a surf-tormented shore,
I stand amid the roar
Is but a dream within a dream.
All that we see or seem
Is it therefore the less gone?
In a vision, or in none,
In a night, or in a day,
Yet if hope has flown away
That my days have been a dream;
You are not wrong, who deem
Thus much let me avow-
And, in parting from you now,
Take this kiss upon the brow!

Approximate sampling - 2

But a dream within a dream?
Is all that we see or seem
One from the pitiless wave?
O God! can I not save
Them with a tighter clasp?
O God! can I not grasp
While I weep--while I weep!
Through my fingers to the deep,
How few! yet how they creep
Grains of the golden sand-
And I hold within my hand
Of a surf-tormented shore,
I stand amid the roar
Is but a dream within a dream.
All that we see or seem
Is it therefore the less gone?
In a vision, or in none,
In a night, or in a day,
Yet if hope has flown away
That my days have been a dream;
You are not wrong, who deem
Thus much let me avow-
And, in parting from you now,
Take this kiss upon the brow!

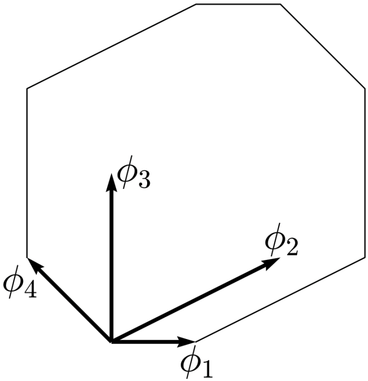
Approximate sampling - 2

But a dream within a dream?
Is all that we see or seem
One from the pitiless wave?
O God! can I not save
Them with a tighter clasp?
O God! can I not grasp
While I weep--while I weep!
Through my fingers to the deep,
How few! yet how they creep
Grains of the golden sand-
And I hold within my hand
Of a surf-tormented shore,
I stand amid the roar
Is but a dream within a dream.
All that we see or seem
Is it therefore the less gone?
In a vision, or in none,
In a night, or in a day,
Yet if hope has flown away
That my days have been a dream;
You are not wrong, who deem
Thus much let me avow-
And, in parting from you now,
Take this kiss upon the brow!

Continuous embedding of the state space $\mathcal{B}$

Volume spanned by feature vectors

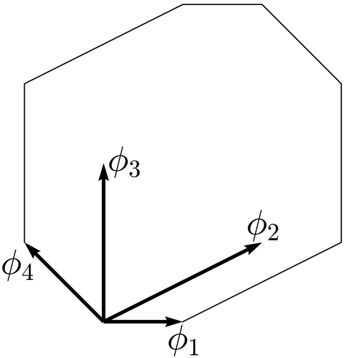
$$\mathcal{Z}(\mathbf{A}) \triangleq \mathbf{A}[0,1]^N$$

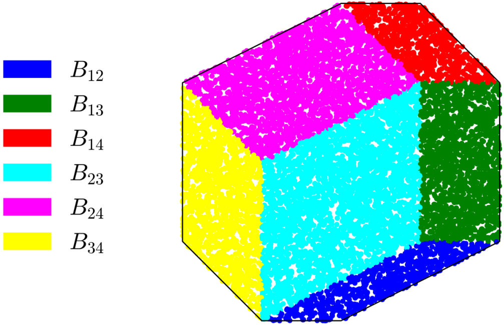


Continuous embedding of the state space $\mathcal{B}$

Volume spanned by feature vectors

$$\mathcal{Z}(\mathbf{A}) \triangleq \mathbf{A}[0,1]^N$$





admits a natural tiling (Dyer & Frieze, 1994), $\mathbf{B} \triangleq \mathbf{A}_{:B}$

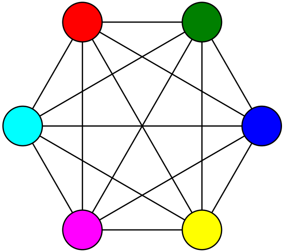
$$\mathrm{Vol}\,\mathcal{Z}(\mathbf{A}) = \sum_{B\in\mathcal{B}} \mathrm{Vol}\,\mathbf{B} = \sum_{B\in\mathcal{B}} |\det \mathbf{B}|$$

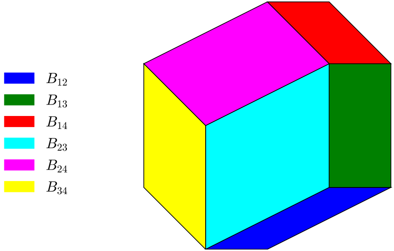
Random walk on $\mathcal{B}$ i.e. on tiles

- From $r \times N$ feature matrix $\mathbf{A} = \left(\phi_1 | \ldots | \phi_N\right)$
- Limiting distribution, $\mathbf{B} \triangleq \mathbf{A}_{:B}$

$$\mathbb{P}\left[\mathcal{X} = B\right] \propto \mathrm{Vol}^2 \, \mathbf{B} \cdot \mathbb{1}_{B \in \mathcal{B}}$$

- State space $\mathcal{B} \triangleq \{B\,;\, \det \mathbf{B} \neq 0\}$
- Continuous embedding of $\mathcal{B}$ via tiling of $\mathcal{Z}(\mathbf{A}) = \mathbf{A}[0,1]^N$

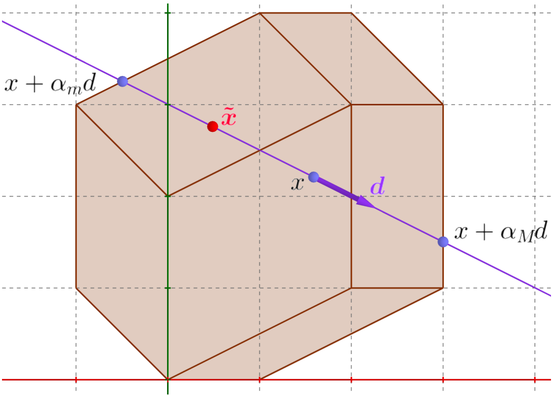




Random walk on $\mathcal{B}$ i.e. on tiles

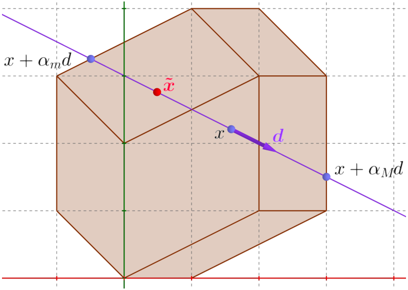
Underlying continuous walk

- $\mathcal{Z}(\mathbf{A})$ is a polytope (convex)
- Hit-and-run is efficient for convex bodies (Lovász & Vempala, 2003)



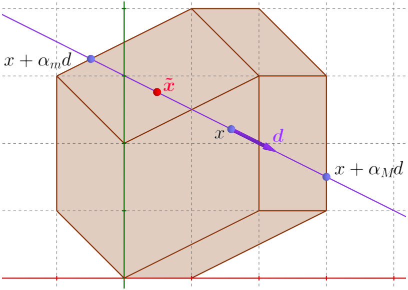
Random walk on $\mathcal{B}$ i.e. on tiles

Continuous random walk on $\mathcal{Z}(\mathbf{A})$



Random walk on $\mathcal{B}$ i.e. on tiles

Continuous random walk on $\mathcal{Z}(\mathbf{A})$



Discrete random walk on $\mathcal{B}$

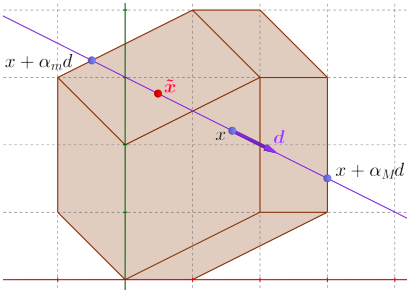
- Identify the tile in which x lies

$$\begin{aligned} \min_{y \in \mathbb{R}^N} \quad & c^\top y \\ \text{s.t.} \quad & \mathbf{A} y = x \\ & 0 \leq y \leq 1 \end{aligned}$$

- $B_x = \{i\,;\, y_i^* \in]0,1[\}$

Random walk on $\mathcal{B}$ i.e. on tiles

Continuous random walk on $\mathcal{Z}(\mathbf{A})$



Continuous target distribution

$$\pi(x)\,\mathrm{d}x = \sum_{B \in \mathcal{B}} C_B \times \mathbb{1}_{\mathbf{B}}(x)\,\mathrm{d}x$$

Discrete random walk on $\mathcal{B}$

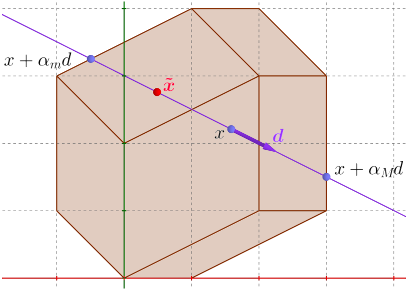
- Identify the tile in which x lies

$$\begin{aligned} \min_{y \in \mathbb{R}^N} \quad & c^\top y \\ \text{s.t.} \quad & \mathbf{A} y = x \\ & 0 \leq y \leq 1 \end{aligned}$$

- $B_x = \{i \,;\, y_i^* \in]0,1[\}$

Random walk on $\mathcal{B}$ i.e. on tiles

Continuous random walk on $\mathcal{Z}(\mathbf{A})$



Discrete random walk on $\mathcal{B}$

- Identify the tile in which x lies

$$\begin{aligned} \min_{y \in \mathbb{R}^N} \quad & c^\top y \\ \text{s.t.} \quad & \mathbf{A} y = x \\ & 0 \leq y \leq 1 \end{aligned}$$

- $B_x = \{ i \,;\, y_i^* \in]0,1[\}$

Continuous target distribution

$$\pi(x)\,\mathrm{d}x = \sum_{B \in \mathcal{B}} C_B \times \mathbb{1}_{\mathbf{B}}(x)\,\mathrm{d}x$$

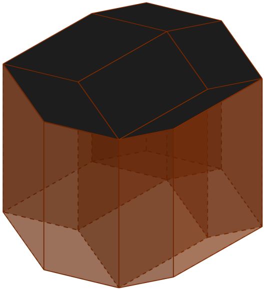
Discrete target distribution

$$\mathbb{P}\left[B_x = B\right] \propto \int_{\mathbf{B}} \pi(x)\,\mathrm{d}x = C_B \times \operatorname{Vol} \mathbf{B}$$

Acceptance = 1

Continuous target distribution

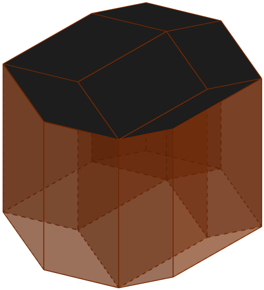
$$\pi(x)\,\mathrm{d}x = \mathbb{1}_{\mathcal{Z}(\mathbf{A})}(x)\,\mathrm{d}x = \sum_{B\in\mathcal{B}} 1 \times \mathbb{1}_{\mathbf{B}}(x)\,\mathrm{d}x$$



Acceptance $= 1$

Continuous target distribution

$$\pi(x)\,\mathrm{d}x = \mathbb{1}_{\mathcal{Z}(\mathbf{A})}(x)\,\mathrm{d}x = \sum_{B \in \mathcal{B}} 1 \times \mathbb{1}_{\mathbf{B}}(x)\,\mathrm{d}x$$

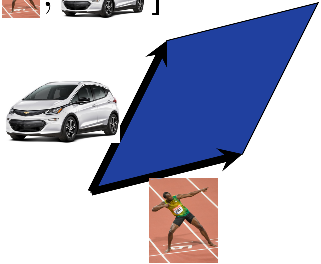


Discrete target distribution

$$\mathbb{P}\left[B_x = B\right] \propto 1 \times \mathrm{Vol}\,\mathbf{B} = \mathrm{Vol}^{1}\,\mathbf{B}$$

$\mathbb{P}[B_x =$ ,  $]$

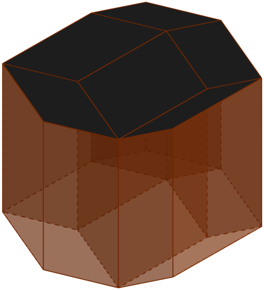
$\propto \mathrm{Vol}^{1}$



Acceptance $= 1$

Continuous target distribution

$$\pi(x)\,\mathrm{d}x = \mathbb{1}_{\mathcal{Z}(\mathbf{A})}(x)\,\mathrm{d}x = \sum_{B \in \mathcal{B}} 1 \times \mathbb{1}_{\mathbf{B}}(x)\,\mathrm{d}x$$



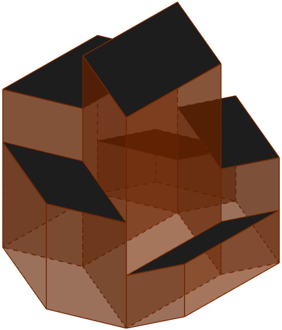
Discrete target distribution

$$\mathbb{P}\left[B_x = B\right] \propto 1 \times \mathrm{Vol}\,\mathbf{B} = \mathrm{Vol}^{1}\,\mathbf{B}$$

Acceptance $= \frac{\text{Vol}\, B(\tilde{x})}{\text{Vol}\, B(x)}$

Continuous target distribution

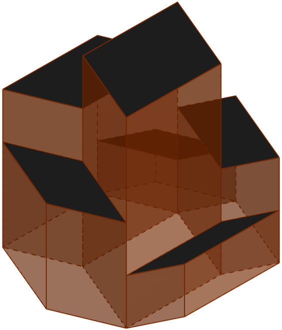
$$\pi(x)\,\mathrm{d}x = \sum_{B \in \mathcal{B}} \text{Vol}\,\mathbf{B} \times \mathbb{1}_{\mathbf{B}}(x)\,\mathrm{d}x$$



Acceptance $= \frac{\text{Vol}\, B(\tilde{x})}{\text{Vol}\, B(x)}$

Continuous target distribution

$$\pi(x)\,\mathrm{d}x = \sum_{B \in \mathcal{B}} \text{Vol}\,\mathbf{B} \times \mathbb{1}_{\mathbf{B}}(x)\,\mathrm{d}x$$

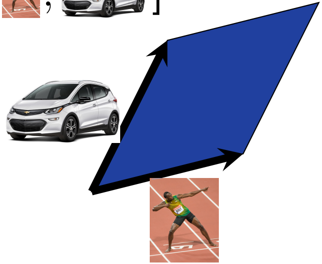


Discrete target distribution

$$\mathbb{P}[B_x = B] \propto \text{Vol}\,\mathbf{B} \times \text{Vol}\,\mathbf{B} = \text{Vol}^2\,\mathbf{B}$$

$\mathbb{P}[B_x =$ ,  $]$

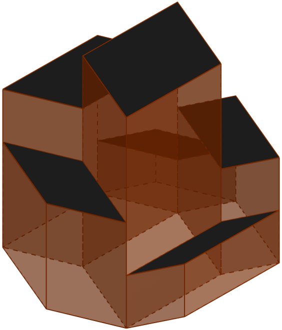
$\propto \text{Vol}^2$



Acceptance $= \frac{\text{Vol } B(\tilde{x})}{\text{Vol } B(x)}$

Continuous target distribution

$$\pi(x)\,\mathrm{d}x = \sum_{B \in \mathcal{B}} \text{Vol } \mathbf{B} \times \mathbb{1}_{\mathbf{B}}(x)\,\mathrm{d}x$$

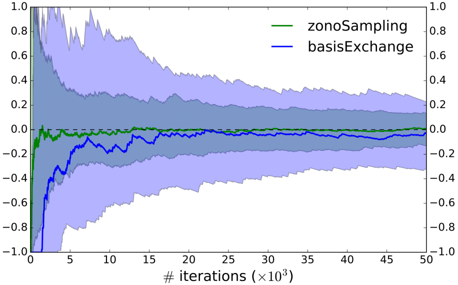


Discrete target distribution

$$\mathbb{P}\left[B_x = B\right] \propto \text{Vol } \mathbf{B} \times \text{Vol } \mathbf{B} = \text{Vol}^2\, \mathbf{B}$$

Behaviour of our chain

Relative error of the estimation of $\mathbb{P}\left[\{i_1, i_2, i_3\} \subseteq \mathcal{X}\right] = \det \mathbf{K}_{\{i_1,i_2,i_3\}}$

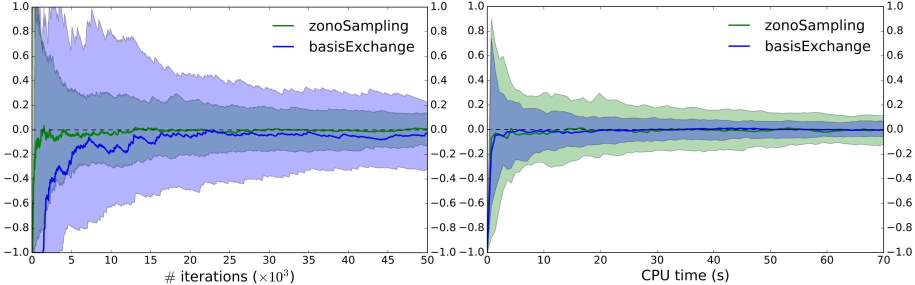


- Better mixing
- More decorelated

Fast sampling of projection DPPs?

Behaviour of our chain

Relative error of the estimation of $\mathbb{P}\left[\{i_1, i_2, i_3\} \subseteq \mathcal{X}\right] = \det \mathbf{K}_{\{i_1,i_2,i_3\}}$



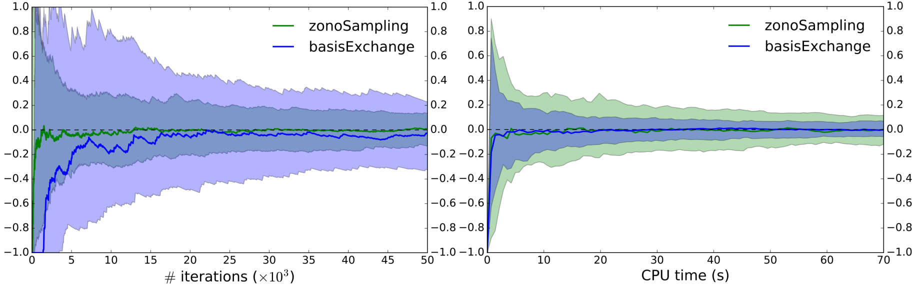
- Better mixing
- More decorelated

- Slower
- Sovle LPs

Fast sampling of projection DPPs?

Behaviour of our chain

Relative error of the estimation of $\mathbb{P}\left[\{i_1, i_2, i_3\} \subseteq \mathcal{X}\right] = \det \mathbf{K}_{\{i_1, i_2, i_3\}}$



- Better mixing
- More decorelated

- Slower
- Sovle LPs

Efficient sampling of projection DPPs!

Summarizing a news article from Slate

Find Y to maximize (Kulesza & Taskar, 2012)

$$\int \textsc{Rouge-1F}(Y, Z)\,\mathrm{DPP}(Z)dZ \approx \frac{1}{N}\sum_{i=1}^{N} \textsc{Rouge-1F}\,(Y, Y_i)$$

where Y_i are samples from our Markov chain

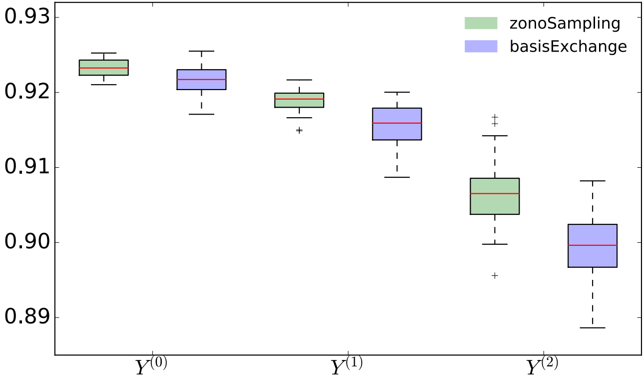


Figure 1: Estimation of the integrated cost

Conclusion

- Provide feature matrix $\mathbf{A}$ (full row rank)
  - Build DPP($\mathbf{A}^{\top}(\mathbf{A}\mathbf{A}^{\top})^{-1}\mathbf{A}$)
  - Continuous embedding of the state space
  - New bridge MCMC $\cap$ Optimization $=$ hit-and-run $+$ LPs
  - Efficient sampling

Conclusion

- Provide feature matrix $\mathbf{A}$ (full row rank)
  - Build DPP($\mathbf{A}^{\top}(\mathbf{A}\mathbf{A}^{\top})^{-1}\mathbf{A}$)
  - Continuous embedding of the state space
  - New bridge MCMC $\cap$ Optimization $=$ hit-and-run $+$ LPs
  - Efficient sampling
- Applications:
  - ML (Kulesza & Taskar, 2012; Kathuria et al., 2016)
  - Graph sampling (Tremblay et al., 2017)
  - Monte Carlo with DPPs (Bardenet & Hardy, 2016)

Conclusion

- Provide feature matrix $\mathbf{A}$ (full row rank)
  - Build DPP($\mathbf{A}^{\top}(\mathbf{A}\mathbf{A}^{\top})^{-1}\mathbf{A}$)
  - Continuous embedding of the state space
  - New bridge MCMC $\cap$ Optimization $=$ hit-and-run $+$ LPs
  - Efficient sampling
- Applications:
  - ML (Kulesza & Taskar, 2012; Kathuria et al., 2016)
  - Graph sampling (Tremblay et al., 2017)
  - Monte Carlo with DPPs (Bardenet & Hardy, 2016)
- Potential speed-up?
  - Starting points
  - Specific implementation of the simplex method

Conclusion

- Provide feature matrix $\mathbf{A}$ (full row rank)
  - Build DPP($\mathbf{A}^\top(\mathbf{A}\mathbf{A}^\top)^{-1}\mathbf{A}$)
  - Continuous embedding of the state space
  - New bridge MCMC $\cap$ Optimization $=$ hit-and-run $+$ LPs
  - Efficient sampling
- Applications:
  - ML (Kulesza & Taskar, 2012; Kathuria et al., 2016)
  - Graph sampling (Tremblay et al., 2017)
  - Monte Carlo with DPPs (Bardenet & Hardy, 2016)
- Potential speed-up?
  - Starting points
  - Specific implementation of the simplex method
- Mixing time?
  - Difficult because of LPs
  - Choice of linear objective c in LP (identification of the tile)

Conclusion

- Provide feature matrix $\mathbf{A}$ (full row rank)
  - Build DPP($\mathbf{A}^\top(\mathbf{A}\mathbf{A}^\top)^{-1}\mathbf{A}$)
  - Continuous embedding of the state space
  - New bridge MCMC $\cap$ Optimization = hit-and-run + LPs
  - Efficient sampling
- Applications:
  - ML (Kulesza & Taskar, 2012; Kathuria et al., 2016)
  - Graph sampling (Tremblay et al., 2017)
  - Monte Carlo with DPPs (Bardenet & Hardy, 2016)
- Potential speed-up?
  - Starting points
  - Specific implementation of the simplex method
- Mixing time?
  - Difficult because of LPs
  - Choice of linear objective c in LP (identification of the tile)
- Generalization
  - k-DPPs (Kulesza & Taskar, 2011)
  - Generic and continuous DPPs (Hough et al., 2006)

Conclusion

- Provide feature matrix $\mathbf{A}$ (full row rank)
  - Build DPP($\mathbf{A}^\top(\mathbf{A}\mathbf{A}^\top)^{-1}\mathbf{A}$)
  - Continuous embedding of the state space
  - New bridge New bridge MCMC $\cap$ Optimization $=$ hit-and-run $+$ LPs
  - Efficient sampling
- Applications:
  - ML (Kulesza & Taskar, 2012; Kathuria et al., 2016)
  - Graph sampling (Tremblay et al., 2017)
  - Monte Carlo with DPPs (Bardenet & Hardy, 2016)
- Potential speed-up?
  - Starting points
  - Specific implementation of the simplex method
- Mixing time?
  - Difficult because of LPs
  - Choice of linear objective c in LP (identification of the tile)
- Generalization
  - k-DPPs (Kulesza & Taskar, 2011)
  - Continuous DPPs (Hough et al., 2006)

POSTER #80

References [1]

Aldous, D. J. The random walk construction of uniform spanning trees and uniform labelled trees. *SIAM Journal on Discrete Mathematics*, 3(4):450–465, 1990.

Anari, N., Gharan, S. O., and Rezaei, A. Monte-Carlo Markov chain algorithms for sampling strongly Rayleigh distributions and determinantal point processes. In *Conference on Learning Theory*, pp. 23–26, 2016.

Bardenet, R. and Hardy, A. Monte-Carlo with determinantal point processes. *arXiv preprint arXiv:1605.00361*, 2016.

Dyer, M. and Frieze, A. Random walks, totally unimodular matrices, and a randomised dual simplex algorithm. *Mathematical Programming*, 64(1-3):1–16, 1994.

Feder, T. and Mihail, M. Balanced matroids. *Proceedings of the twenty-fourth annual ACM*, pp. 26–38, 1992.

Hough, J. B., Krishnapur, M., Peres, Y., and Virág, B. Determinantal processes and independence. *Probability surveys*, 2006.

Kathuria, T., Deshpande, A., and Kohli, P. Batched gaussian process bandit optimization via determinantal point processes. *Neural Information Processing Systems*, pp. pp. 4206–4214, 2016.

Kulesza, A. and Taskar, B. Determinantal point processes for machine learning. *Foundations and Trends in Machine Learning*, 5(2-3):123–286, 2012.

Kulesza, A. and Taskar, B. *k*-dpps: Fixed-size determinantal point processes. *International Conference on Machine Learning*, pp. 1193–1200, 2011.

References [2]

Li, C., Jegelka, S., and Sra, S. Fast mixing markov chains for strongly rayleigh measures, dpps, and constrained sampling. In *Neural Information Processing Systems*, pp. 4188–4196, 2016.

Lovász, L. and Vempala, S. Hit and run is fast and fun. Technical Report MSR-TR-2003-05, 2003.

Lyons, R. Determinantal probability measures. *Publications Mathématiques de l'Institut des Hautes Études Scientifiques*, 2003.

Oxley, J. What is a matroid? *Cubo Matemática Educacional*, 5.3:179–218, 2003.

Propp, J. G. and Wilson, D. B. How to get a perfectly random sample from a generic Markov chain and generate a random spanning tree of a directed graph. *Journal of Algorithms*, 27(2): 170–217, 1998.

Tremblay, N., Amblard, P.-O., and Barthelme, S. Graph sampling with determinantal processes. working paper or preprint, March 2017.