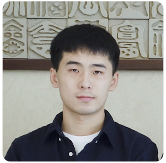
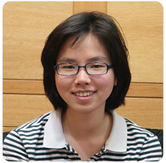
Parallel Discrete Sampling via Continuous Walks

Nima Anari

Stanford University

joint work with





Yizhi Huang

Tianyu Liu

June Vuong

Brian Xu

Katherine Yu

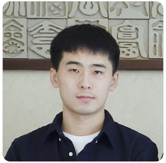
DALL·E for Spanning Trees

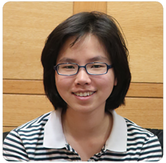
Nima Anari

Stanford University

joint work with


Yizhi Huang


Tianyu Liu

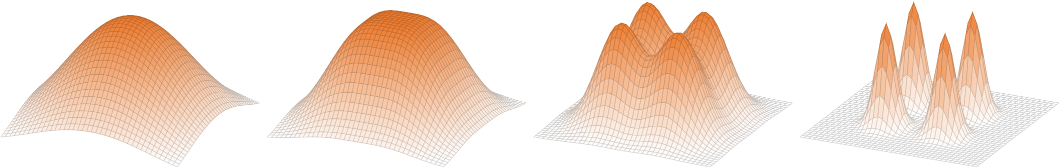

June Vuong


Brian Xu


Katherine Yu

sampling in diffusion models [image by Andy Shih]

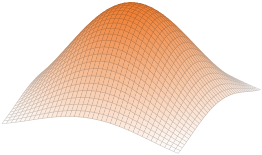
sampling in diffusion models [image by Andy Shih]



stochastic localization [Eldan'13]

Sampling

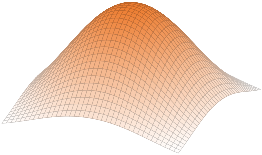
Continuous



$$\mu : \mathbb{R}^n \to \mathbb{R}_{\geqslant 0}$$

Sampling

Continuous

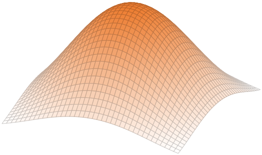


$$\mu : \mathbb{R}^n \to \mathbb{R}_{\geqslant 0}$$

- Tractable: $\log \mu$ concave

Sampling

Continuous



$$\mu : \mathbb{R}^n \to \mathbb{R}_{\geqslant 0}$$

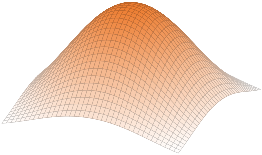
- Tractable: $\log \mu$ concave
- Even better (well-conditioned):

$$-\beta I \preceq \nabla^2 \log \mu \preceq -\alpha I.$$

and β/α is small.

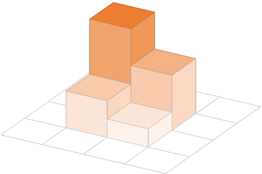
Sampling

Continuous



$\mu : \mathbb{R}^n \to \mathbb{R}_{\geqslant 0}$

Discrete



$\mu : \{\pm 1\}^n \to \mathbb{R}_{\geqslant 0}$

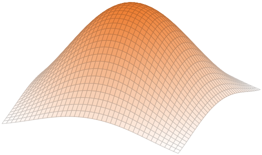
- Tractable: $\log \mu$ concave
- Even better (well-conditioned):

$$-\beta I \preceq \nabla^2 \log \mu \preceq -\alpha I.$$

and β/α is small.

Sampling

Continuous



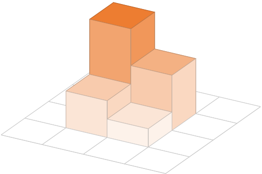
$$\mu : \mathbb{R}^n \to \mathbb{R}_{\geqslant 0}$$

- Tractable: $\log \mu$ concave
- Even better (well-conditioned):

$$-\beta I \preceq \nabla^2 \log \mu \preceq -\alpha I.$$

and β/α is small.

Discrete



$$\mu : \{\pm 1\}^n \to \mathbb{R}_{\geqslant 0}$$

- Tractable: ? (patchwork)

Sampling via counting

Counting

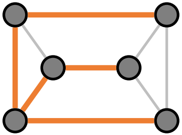
Sub-cube $C \subseteq \{\pm 1\}^n \mapsto \sum_{x \in C} \mu(x)$.

Sampling via counting

Counting

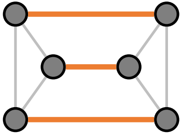
Sub-cube $C \subseteq \{\pm 1\}^n \mapsto \sum_{x \in C} \mu(x)$.

Spanning trees



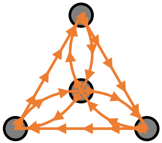
Matrix-tree thm.

Planar PMs



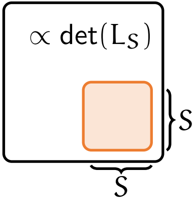
[FKT] thm.

Eulerian tours*



[BEST] thm.

Det. Point Process

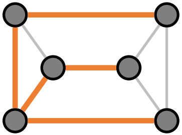
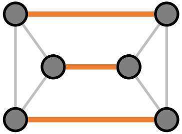
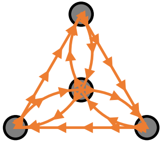
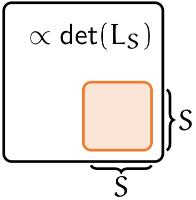


$\det(L + I)$

Sampling via counting

Counting

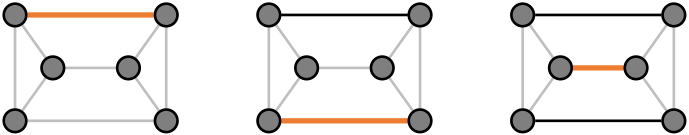
Sub-cube $C \subseteq \{\pm 1\}^n \mapsto \sum_{x \in C} \mu(x)$.

Spanning trees	Planar PMs	Eulerian tours*	Det. Point Process
			
Matrix-tree thm.	[FKT] thm.	[BEST] thm.	$\det(L + I)$

[Jerrum-Valiant-Vazirani'89]

Polynomial-time counting $\Longrightarrow$ polynomial-time sampling.

- The standard reduction [Jerrum-Valiant-Vazirani'89] is sequential. ☹



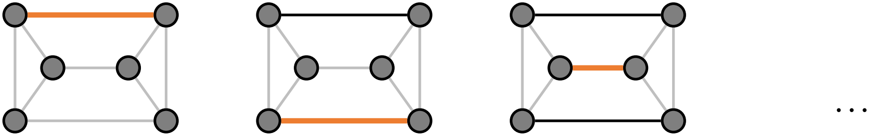
$\mathbb{P}[e_1]$?

$\mathbb{P}[e_2 \mid e_1]$?

$\mathbb{P}[e_3 \mid e_1, e_2]$?

...

- The standard reduction [Jerrum-Valiant-Vazirani'89] is sequential. ☹

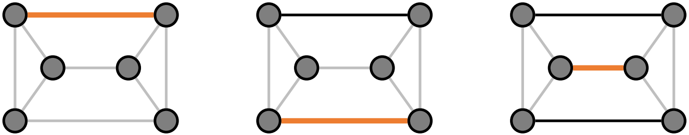


$\mathbb{P}[e_1]$? $\mathbb{P}[e_2 \mid e_1]$? $\mathbb{P}[e_3 \mid e_1, e_2]$?

- Counting doable in parallel: $\log(n)^{O(1)}$ time with $n^{O(1)}$ processors (NC).

[Csanky'75]
Linear algebra is parallelizable.

- The standard reduction [Jerrum-Valiant-Vazirani'89] is sequential. ☹



$\mathbb{P}[e_1]$? $\mathbb{P}[e_2 \mid e_1]$? $\mathbb{P}[e_3 \mid e_1, e_2]$? $\cdots$

- Counting doable in parallel: $\log(n)^{O(1)}$ time with $n^{O(1)}$ processors (NC).

[Csanky'75]

Linear algebra is parallelizable.

- Question: Can we sample in parallel (RNC)?

Main result (informal)

We can sample spanning trees, DPPs, Eulerian tours, and more in parallel by moving to continuous space.

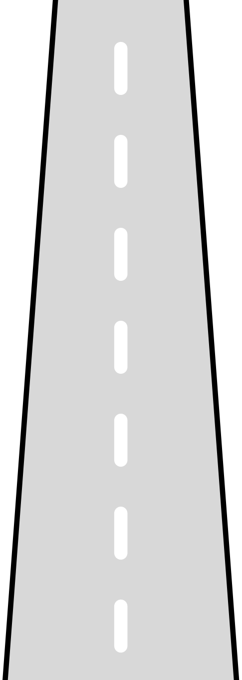
Note: list excludes planar perfect matchings.

Discrete to Continuous

- Exponential Tilts
- Interlude: Eulerian Tours
- Transport Stability

Sampling Algorithm

- Stochastic Localization
- Parallel Continuous Sampling

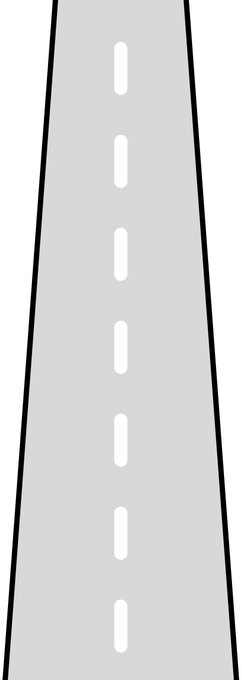


Discrete to Continuous

- Exponential Tilts
- Interlude: Eulerian Tours
- Transport Stability

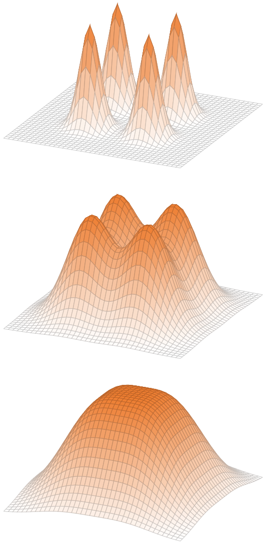
Sampling Algorithm

- Stochastic Localization
- Parallel Continuous Sampling



From discrete to continuous

Take convolution of μ with normal $\mathcal{N}(0, cI)$.

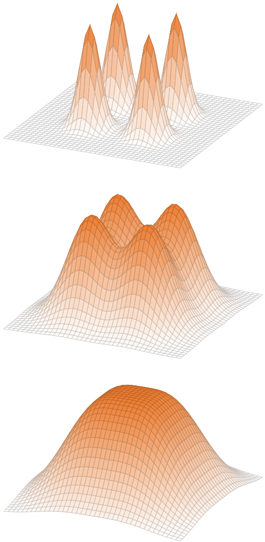


From discrete to continuous

Take convolution of μ with normal $\mathcal{N}(0, cI)$.

Main lemma

$\nu := \mu * \mathcal{N}(0, cI)$ log-concave for $c \geqslant c_0 = O(1)$.

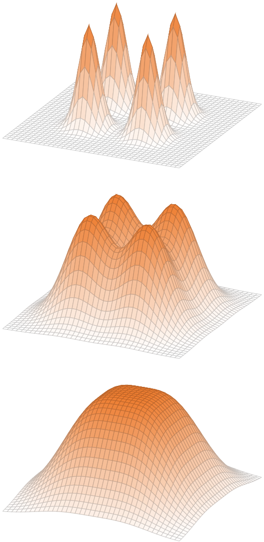


From discrete to continuous

Take convolution of μ with normal $\mathcal{N}(0, cI)$.

Main lemma

$\nu := \mu * \mathcal{N}(0, cI)$ log-concave for $c \geqslant c_0 = O(1)$.



From discrete to continuous

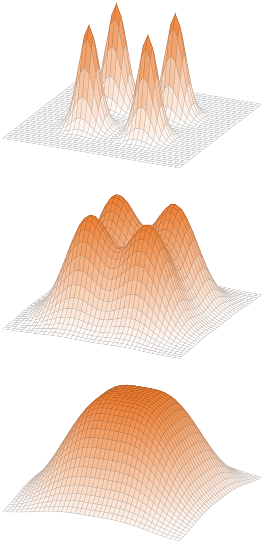
Take convolution of μ with normal $\mathcal{N}(0, cI)$.

Main lemma

$\nu := \mu * \mathcal{N}(0, cI)$ log-concave for $c \geqslant c_0 = O(1)$.

- The p.d.f. of ν at w is $\propto \sum_x e^{-\|w-x\|^2/2c} \mu(x)$

$$\propto e^{-\|w\|^2/2c} \cdot \underbrace{\sum_x e^{\langle w/c, x\rangle} \mu(x)}_{\text{count of weighted } \mu}.$$



From discrete to continuous

Take convolution of μ with normal $\mathcal{N}(0, cI)$.

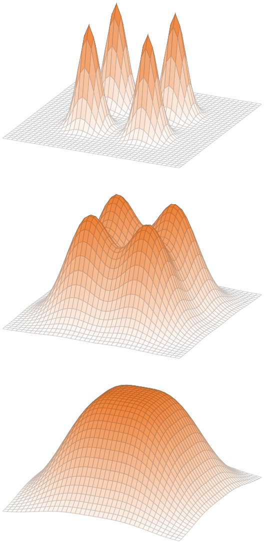
Main lemma

$\nu := \mu * \mathcal{N}(0, cI)$ log-concave for $c \geqslant c_0 = O(1)$.

- The p.d.f. of ν at w is $\propto \sum_x e^{-\|w-x\|^2/2c} \mu(x)$

$$\propto e^{-\|w\|^2/2c} \cdot \underbrace{\sum_x e^{\langle w/c, x\rangle} \mu(x)}_{\text{count of weighted } \mu} .$$

- $$\nabla^2 \log \nu \big|_{w=0} = -I/c + \mathrm{cov}(\mu)/c^2$$



From discrete to continuous

Take convolution of μ with normal $\mathcal{N}(0, cI)$.

Main lemma

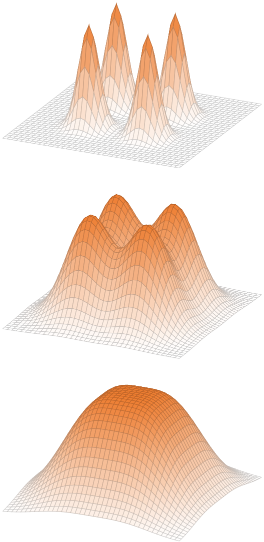
$\nu := \mu * \mathcal{N}(0, cI)$ log-concave for $c \geqslant c_0 = O(1)$.

- The p.d.f. of ν at w is $\propto \sum_x e^{-\|w-x\|^2/2c}\mu(x)$

$$\propto e^{-\|w\|^2/2c} \cdot \underbrace{\sum_x e^{\langle w/c, x\rangle}\mu(x)}_{\text{count of weighted } \mu}.$$

- $$\nabla^2 \log \nu\big|_{w=0} = -I/c + \mathrm{cov}(\mu)/c^2$$

- For larger variance, e.g., $\mu * \mathcal{N}(0, 2c_0 I)$, we have well-conditioned log-concavity (easy to sample).



Exponential tilts

For μ on $\{\pm 1\}^n$, an exponential tilt is $\tau_w \mu$ for $w \in \mathbb{R}^n$ defined as

$$\tau_w \mu(x) \propto e^{\langle w, x \rangle} \mu(x).$$

Exponential tilts

For μ on $\{\pm 1\}^n$, an exponential tilt is $\tau_w \mu$ for $w \in \mathbb{R}^n$ defined as

$$\tau_w \mu(x) \propto e^{\langle w, x \rangle} \mu(x).$$

- Spanning trees $\to$ weighted spanning trees.

Exponential tilts

For μ on $\{\pm 1\}^n$, an exponential tilt is $\tau_w \mu$ for $w \in \mathbb{R}^n$ defined as

$$\tau_w \mu(x) \propto e^{\langle w, x \rangle} \mu(x).$$

- Spanning trees $\to$ weighted spanning trees.
- DPP $\to$ DPP with rows and columns scaled.

Exponential tilts

For μ on $\{\pm 1\}^n$, an exponential tilt is $\tau_w \mu$ for $w \in \mathbb{R}^n$ defined as

$$\tau_w \mu(x) \propto e^{\langle w, x \rangle} \mu(x).$$

- Spanning trees $\rightarrow$ weighted spanning trees.
- DPP $\rightarrow$ DPP with rows and columns scaled.
- Eulerian tours: switching networks $\ldots$

Exponential tilts

For μ on $\{\pm 1\}^n$, an exponential tilt is $\tau_w \mu$ for $w \in \mathbb{R}^n$ defined as

$$\tau_w \mu(x) \propto e^{\langle w, x \rangle} \mu(x).$$

- Spanning trees $\rightarrow$ weighted spanning trees.
- DPP $\rightarrow$ DPP with rows and columns scaled.
- Eulerian tours: switching networks $\ldots$

Determinant-based counting is closed under external fields.

Exponential tilts

For μ on $\{\pm 1\}^n$, an exponential tilt is $\tau_w \mu$ for $w \in \mathbb{R}^n$ defined as

$$\tau_w \mu(x) \propto e^{\langle w, x \rangle} \mu(x).$$

- Spanning trees $\to$ weighted spanning trees.
- DPP $\to$ DPP with rows and columns scaled.
- Eulerian tours: switching networks ...

Determinant-based counting is closed under external fields.

Covariance bound

We just need all of these $\tau_w \mu$ to have bounded covariance (semi-log-concavity [Eldan-Shamir'20]):

$$\text{cov}(\tau_w \mu) \preceq O(1) \cdot I.$$

Exponential tilts

For μ on $\{\pm 1\}^n$, an exponential tilt is $\tau_w \mu$ for $w \in \mathbb{R}^n$ defined as

$$\tau_w \mu(x) \propto e^{\langle w, x\rangle} \mu(x).$$

- Spanning trees $\rightarrow$ weighted spanning trees.
- DPP $\rightarrow$ DPP with rows and columns scaled.
- Eulerian tours: switching networks ...

Determinant-based counting is closed under external fields.

Covariance bound

We just need all of these $\tau_w \mu$ to have bounded covariance (semi-log-concavity [Eldan-Shamir'20]):

$$\mathrm{cov}(\tau_w \mu) \preceq O(1) \cdot I.$$

Spectral independence [A-Liu-OveisGharan'20] is even stronger:

$$\mathrm{cov}(\tau_w \mu) \preceq O(1) \cdot \mathrm{diag}(\mathrm{cov}(\tau_w \mu)).$$

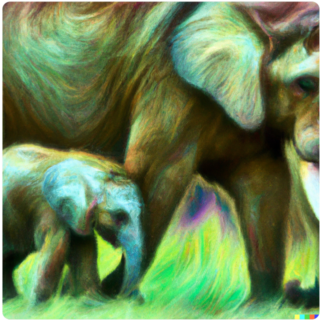
All except Planar PMs.

[Alimohammadi-A-Shiragur-Vuong'21]

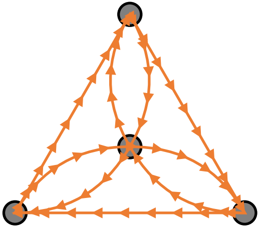
What are switching networks?

What are switching networks?

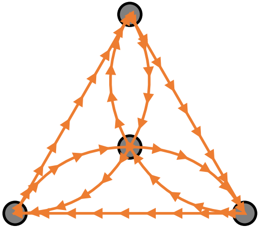
What are switching networks?



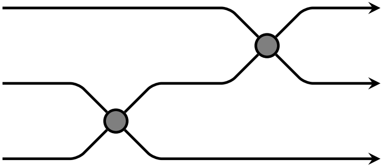
- Represent Eulerian tours as members of $\{\pm 1\}^n$?

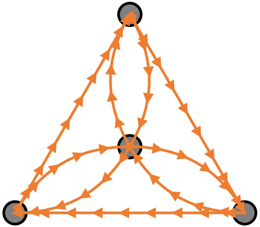


- Represent Eulerian tours as members of $\{\pm 1\}^n$?
- General case reducible to $\mathbf{deg}_{\mathsf{in}} = \mathbf{deg}_{\mathsf{out}} = 2$.

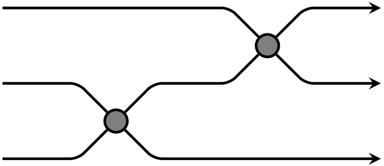


- Represent Eulerian tours as members of $\{\pm 1\}^n$?
- General case reducible to $\mathbf{deg}_{\text{in}} = \mathbf{deg}_{\text{out}} = 2$.
- Replace each vertex by "switching network" gadget:

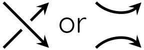




- Represent Eulerian tours as members of $\{\pm 1\}^n$?
- General case reducible to $\mathbf{deg}_{\text{in}} = \mathbf{deg}_{\text{out}} = 2$.
- Replace each vertex by “switching network” gadget:

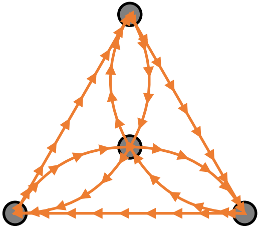


- Binary choice per vertex:

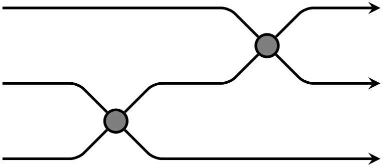


[Bouchet]: $\exists\ n \times n$ skew-symmetric L, such that

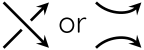
$$\det(L_{S,S}) = \mathbb{1}[S \text{ indicates Eulerian tour}].$$



- Represent Eulerian tours as members of $\{\pm 1\}^n$?
- General case reducible to $\mathbf{deg}_{\text{in}} = \mathbf{deg}_{\text{out}} = 2$.
- Replace each vertex by "switching network" gadget:



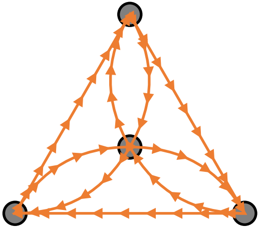
- Binary choice per vertex:



[Bouchet]: $\exists\ n \times n$ skew-symmetric L, such that

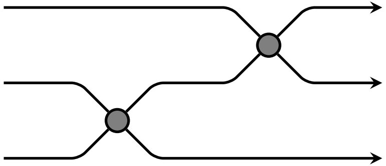
$$\mathbf{det}(L_{S,S}) = \mathbb{1}[S \text{ indicates Eulerian tour}].$$

- Exponential tilt becomes biased switching.



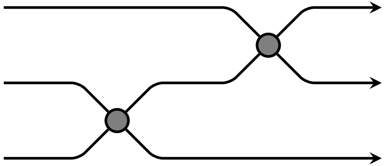
Switching networks

- Want: random switching $\equiv$ uniformly random permutation.



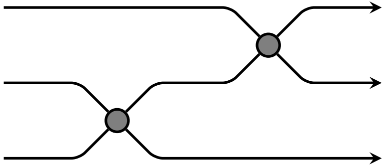
Switching networks

- Want: random switching $\equiv$ uniformly random permutation.
- Randomly constructed unbiased $\widetilde{O}(\mathbf{deg})$-sized network $\simeq$ uniform permutation [Czumaj'15].



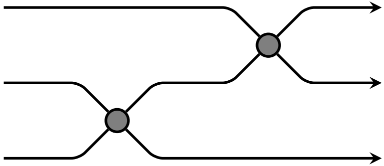
Switching networks

- Want: random switching $\equiv$ uniformly random permutation.
- Randomly constructed unbiased $\widetilde{\mathrm{O}}(\mathbf{deg})$-sized network $\simeq$ uniform permutation [Czumaj'15].
- With biases, $\widetilde{\mathrm{O}}(\mathbf{deg}^2)$ enough to get exactly uniform.



Switching networks

- Want: random switching $\equiv$ uniformly random permutation.
- Randomly constructed unbiased $\widetilde{O}(\mathbf{deg})$-sized network $\simeq$ uniform permutation [Czumaj'15].
- With biases, $\widetilde{O}(\mathbf{deg}^2)$ enough to get exactly uniform.
- Open: What is the minimum size for exactly uniform permutations?



Standard Counting

Sub-cube $C \subseteq \{\pm 1\}^n \mapsto \sum_{x \in C} \mu(x)$.

Weighted Counting

$w \mapsto \sum_x e^{\langle w, x \rangle} \mu(x)$.

Standard Counting

Sub-cube $C \subseteq \{\pm 1\}^n \mapsto \sum_{x \in C} \mu(x)$.

Weighted Counting

$w \mapsto \sum_x e^{\langle w, x \rangle} \mu(x)$.

Main result 1

Approx. sampling (ϵ in d_{TV}) via weighted counting in $\textbf{polylog}(n/\epsilon)$ time and $\textbf{quasipoly}(n/\epsilon)$ processors, for μ spectrally independent under exponential tilts.

Standard Counting

Sub-cube $C \subseteq \{\pm 1\}^n \mapsto \sum_{x \in C} \mu(x)$.

Weighted Counting

$w \mapsto \sum_x e^{\langle w, x \rangle} \mu(x)$.

Main result 1

Approx. sampling (ϵ in d_{TV}) via weighted counting in **polylog**(n/ϵ) time and **quasipoly**(n/ϵ) processors, for μ spectrally independent under exponential tilts.

- Spectral independence [A-Liu-OveisGharan'20] under exponential tilts is also known as "fractional log-concavity" [Alimohammadi-A-Shiragur-Vuong'21].

Standard Counting

Sub-cube $C \subseteq \{\pm 1\}^n \mapsto \sum_{x \in C} \mu(x)$.

Weighted Counting

$w \mapsto \sum_x e^{\langle w, x \rangle} \mu(x)$.

Main result 1

Approx. sampling (ϵ in d_{TV}) via weighted counting in $\mathbf{polylog}(n/\epsilon)$ time and $\mathbf{quasipoly}(n/\epsilon)$ processors, for μ spectrally independent under exponential tilts.

- Spectral independence [A-Liu-OveisGharan'20] under exponential tilts is also known as "fractional log-concavity" [Alimohammadi-A-Shiragur-Vuong'21].
- Weaker condition "semi-log-concavity" [Eldan-Shamir'20] is also enough.

Applications

- Quasi-RNC sampling of Eulerian tours in digraphs.

Applications

- Quasi-RNC sampling of Eulerian tours in digraphs.
- Quasi-RNC sampling of DPPs on skew-symmetric matrix.

Applications

- Quasi-RNC sampling of Eulerian tours in digraphs.
- Quasi-RNC sampling of DPPs on skew-symmetric matrix.

Main result 2

The number of processors can be improved from **quasipoly**(n/ϵ) to **poly**(n/ϵ) if μ is "transport-stable".

Applications

- Quasi-RNC sampling of Eulerian tours in digraphs.
- Quasi-RNC sampling of DPPs on skew-symmetric matrix.

Main result 2

The number of processors can be improved from **quasipoly**(n/ϵ) to **poly**(n/ϵ) if μ is "transport-stable".

- RNC sampling of DPPs on symmetric PSD matrix.

Applications

- Quasi-RNC sampling of Eulerian tours in digraphs.
- Quasi-RNC sampling of DPPs on skew-symmetric matrix.

Main result 2

The number of processors can be improved from $\mathrm{quasipoly}(n/\epsilon)$ to $\mathrm{poly}(n/\epsilon)$ if μ is "transport-stable".

- RNC sampling of DPPs on symmetric PSD matrix.
- RNC sampling of spanning trees (already known via parallelization of Aldous-Broder alg. [Teng'95, A-Hu-Saberi-Schild'21]).

Applications

- Quasi-RNC sampling of Eulerian tours in digraphs.
- Quasi-RNC sampling of DPPs on skew-symmetric matrix.

Main result 2

The number of processors can be improved from **quasipoly**(n/ϵ) to **poly**(n/ϵ) if μ is "transport-stable".

- RNC sampling of DPPs on symmetric PSD matrix.
- RNC sampling of spanning trees (already known via parallelization of Aldous-Broder alg. [Teng'95, A-Hu-Saberi-Schild'21]).
- Conjecture: Eulerian tours and non-symmetric DPPs also "transport-stable".

Applications

- Quasi-RNC sampling of Eulerian tours in digraphs.
- Quasi-RNC sampling of DPPs on skew-symmetric matrix.

Main result 2

The number of processors can be improved from $\mathsf{quasipoly}(n/\epsilon)$ to $\mathsf{poly}(n/\epsilon)$ if μ is "transport-stable".

- RNC sampling of DPPs on symmetric PSD matrix.
- RNC sampling of spanning trees (already known via parallelization of Aldous-Broder alg. [Teng'95, A-Hu-Saberi-Schild'21]).
- Conjecture: Eulerian tours and non-symmetric DPPs also "transport-stable".
- Corollary of ongoing work [A-Chewi-Vuong]: "Quasi" can be dropped.

Transport stability

We call μ transport-stable if

$$\underbrace{\mathcal{W}_1(\tau_w\mu, \tau_{w'}\mu)}_{\text{Wasserstein distance w.r.t. Hamming metric}} \leqslant C \cdot \|w - w'\|_1.$$

Transport stability

We call μ transport-stable if

$$\underbrace{\mathcal{W}_1(\tau_w \mu, \tau_{w'} \mu)}_{\text{Wasserstein distance w.r.t. Hamming metric}} \leqslant C \cdot \|w - w'\|_1 .$$

Lemma

$C = O(1)$ for spanning trees, etc.

Fact

$C = O(n)$ for any distribution.

Transport stability

We call μ transport-stable if

$$\underbrace{\mathcal{W}_1(\tau_w \mu, \tau_{w'} \mu)}_{\text{Wasserstein distance w.r.t. Hamming metric}} \leqslant C \cdot \|w - w'\|_1.$$

Lemma

$C = O(1)$ for spanning trees, etc.

Fact

$C = O(n)$ for any distribution.

- In contrast, semi-log-concavity is:

$$\|\text{mean}(\tau_w \mu) - \text{mean}(\tau_{w'} \mu)\|_2 \leqslant C \cdot \|w - w'\|_2.$$

Transport stability

We call μ transport-stable if

$$\underbrace{\mathcal{W}_1(\tau_w \mu, \tau_{w'} \mu)}_{\text{Wasserstein distance w.r.t. Hamming metric}} \leqslant C \cdot \|w - w'\|_1.$$

Lemma

$C = O(1)$ for spanning trees, etc.

Fact

$C = O(n)$ for any distribution.

- In contrast, semi-log-concavity is:

$$\|\text{mean}(\tau_w \mu) - \text{mean}(\tau_{w'} \mu)\|_2 \leqslant C \cdot \|w - w'\|_2.$$

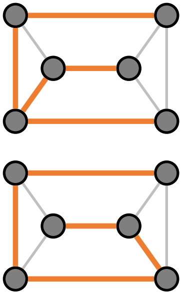
- Aside: $\|\cdot\|_2$ can be replaced by $\|\cdot\|_1$ in our dists.

Transport stability

[Feder-Mihail'92]

For edge e, $\exists$ random spanning trees T, T', such that

- T is uniformly random conditioned on $e \in T$.
- T' is uniformly random conditioned on $e \notin T'$.
- Almost surely $|T \Delta T'| = 2$.

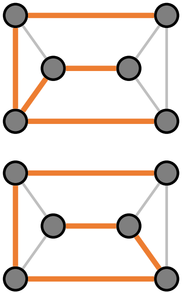


Transport stability

[Feder-Mihail'92]

For edge e, $\exists$ random spanning trees T, T', such that

- T is uniformly random conditioned on $e \in T$.
- T' is uniformly random conditioned on $e \notin T'$.
- Almost surely $|T \Delta T'| = 2$.



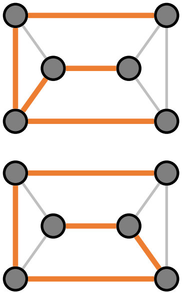
- By gluing these couplings, we get transport-stability.

Transport stability

[Feder-Mihail'92]

For edge e, $\exists$ random spanning trees T, T', such that

- T is uniformly random conditioned on $e \in T$.
- T' is uniformly random conditioned on $e \notin T'$.
- Almost surely $|T \Delta T'| = 2$.



- By gluing these couplings, we get transport-stability.
- Transport stability $\implies$ semi-log-concavity.

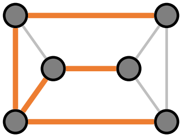
$$\|\text{each row of } \mathrm{cov}(\mu)\|_1 \leqslant O(1).$$

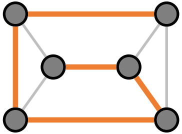
Transport stability

[Feder-Mihail'92]

For edge e, $\exists$ random spanning trees T, T', such that

- T is uniformly random conditioned on $e \in T$.
- T' is uniformly random conditioned on $e \notin T'$.
- Almost surely $|T \Delta T'| = 2$.





- By gluing these couplings, we get transport-stability.
- Transport stability $\implies$ semi-log-concavity.

$$\|\text{each row of } \mathsf{cov}(\mu)\|_1 \leqslant O(1).$$

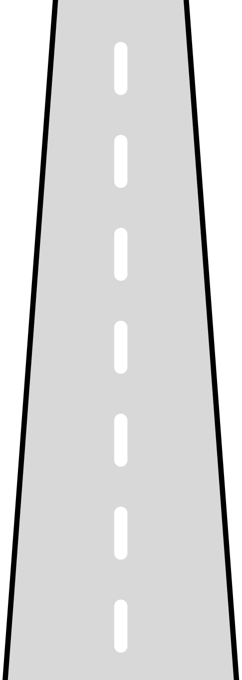
- Conjecture: the same holds for Eulerian tours, etc.

Discrete to Continuous

- Exponential Tilts
- Interlude: Eulerian Tours
- Transport Stability

Sampling Algorithm

- Stochastic Localization
- Parallel Continuous Sampling

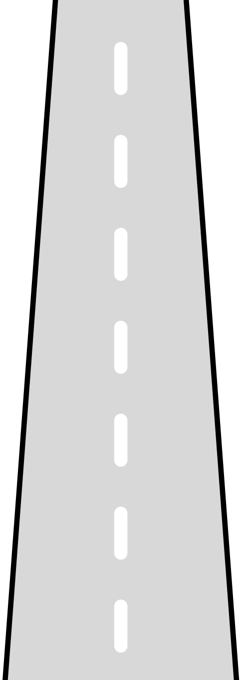


Discrete to Continuous

- Exponential Tilts
- Interlude: Eulerian Tours
- Transport Stability

Sampling Algorithm

- Stochastic Localization
- Parallel Continuous Sampling



How do we turn continuous samples into discrete ones?

Stochastic localization (i.e., DALL·E-for-theorists)

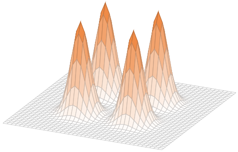
Stochastic localization [Eldan'13] in discrete time steps. Different discretization used by [ElAlaoui-Montanari-Sellke'22].

$w_0 \leftarrow 0$
for $i = 0, \dots, T-1$ **do**
  $x \leftarrow$ sample from $\tau_{w_i}\mu * \mathcal{N}(0, cI)$
  $w_{i+1} \leftarrow w_i + x/c$

return $\mathrm{sign}(w_T)$

Stochastic localization (i.e., DALL·E-for-theorists)

Stochastic localization [Eldan'13] in discrete time steps. Different discretization used by [ElAlaoui-Montanari-Sellke'22].



$w_0 \leftarrow 0$
for $i = 0, \dots, T-1$ **do**
  $x \leftarrow$ sample from $\tau_{w_i}\mu * \mathcal{N}(0, cI)$
  $w_{i+1} \leftarrow w_i + x/c$
return sign(w_T)

Stochastic localization (i.e., DALL·E-for-theorists)

Stochastic localization [Eldan'13] in discrete time steps. Different discretization used by [ElAlaoui-Montanari-Sellke'22].

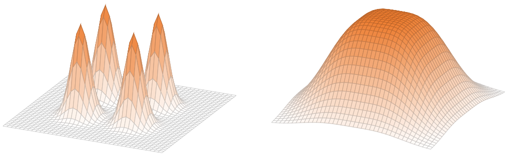
$w_0 \leftarrow 0$

for $i = 0, \ldots, T-1$ **do**

  $x \leftarrow$ sample from $\tau_{w_i}\mu * \mathcal{N}(0, cI)$

  $w_{i+1} \leftarrow w_i + x/c$

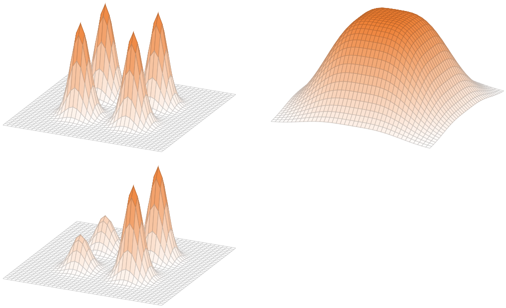
return $\mathrm{sign}(w_T)$



Stochastic localization (i.e., DALL·E-for-theorists)

Stochastic localization [Eldan'13] in discrete time steps. Different discretization used by [ElAlaoui-Montanari-Sellke'22].

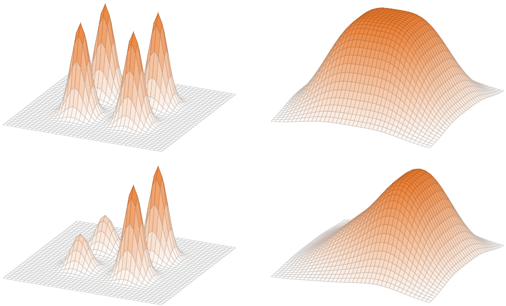
$w_0 \leftarrow 0$
for $i = 0, \dots, T-1$ **do**
  $x \leftarrow$ sample from $\tau_{w_i}\mu * \mathcal{N}(0, cI)$
  $w_{i+1} \leftarrow w_i + x/c$
return $\mathrm{sign}(w_T)$



Stochastic localization (i.e., DALL·E-for-theorists)

Stochastic localization [Eldan'13] in discrete time steps. Different discretization used by [ElAlaoui-Montanari-Sellke'22].

$w_0 \leftarrow 0$
for $i = 0, \ldots, T-1$ **do**
  $x \leftarrow$ sample from $\tau_{w_i}\mu * \mathcal{N}(0, cI)$
  $w_{i+1} \leftarrow w_i + x/c$
return $\mathrm{sign}(w_T)$



Stochastic localization (i.e., DALL·E-for-theorists)

Stochastic localization [Eldan'13] in discrete time steps. Different discretization used by [ElAlaoui-Montanari-Sellke'22].

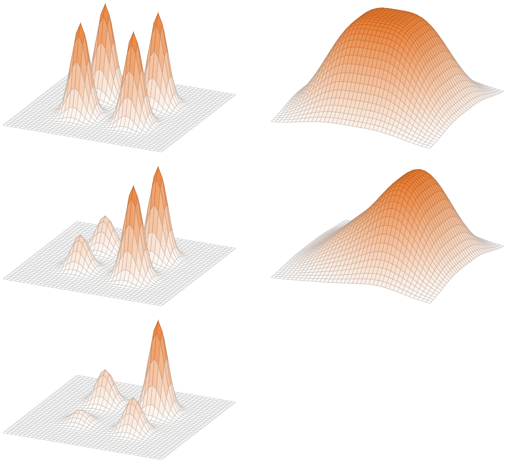
$w_0 \leftarrow 0$

for $i = 0, \ldots, T-1$ **do**

  $x \leftarrow$ sample from $\tau_{w_i}\mu * \mathcal{N}(0, cI)$

  $w_{i+1} \leftarrow w_i + x/c$

return $\mathrm{sign}(w_T)$



Stochastic localization (i.e., DALL·E-for-theorists)

Stochastic localization [Eldan'13] in discrete time steps. Different discretization used by [ElAlaoui-Montanari-Sellke'22].

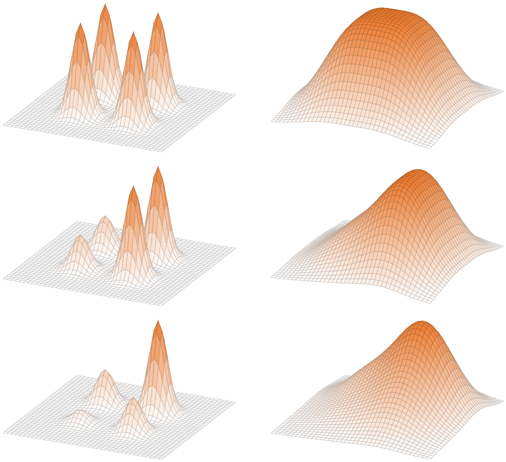
$w_0 \leftarrow 0$

for $i = 0, \dots, T-1$ **do**

  $x \leftarrow$ sample from $\tau_{w_i}\mu * \mathcal{N}(0, cI)$

  $w_{i+1} \leftarrow w_i + x/c$

return $\mathrm{sign}(w_T)$



Stochastic localization (i.e., DALL·E-for-theorists)

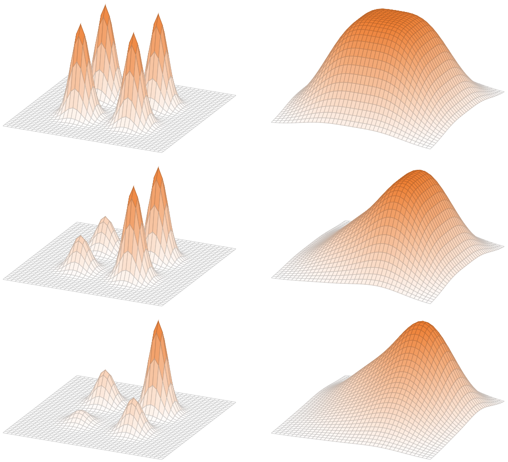
Stochastic localization [Eldan'13] in discrete time steps. Different discretization used by [ElAlaoui-Montanari-Sellke'22].

$w_0 \leftarrow 0$
for $i = 0, \ldots, T-1$ **do**
  $x \leftarrow$ sample from $\tau_{w_i}\mu * \mathcal{N}(0, cI)$
  $w_{i+1} \leftarrow w_i + x/c$
return $\mathrm{sign}(w_T)$

Lemma [cf. ElAlaoui-Montanari'21]

$$cw_T/T \sim \mu * \mathcal{N}(0, cI/T).$$

▷ Enough to stop at $T \simeq c\log(n)$.



How do we sample from $\mu * \mathcal{N}(0, cI)$ in parallel?

Parallel continuous sampling

- Open: For a well-conditioned log-concave ν on $\mathbb{R}^n$, what is the minimum number of $\nabla \log \nu$ we need to query to sample? We do not know if $\mathrm{polylog}(n)$ is possible. ☹
- Fortunately parallel time $\mathrm{polylog}(n)$ is possible. ☺ We use randomized midpoint of [Shen-Lee'19], but others such as Lagenvin can be parallelized too [A-Chewi-Vuong]. Picard iterations change the sequential version:

$$x_{t+dt} \leftarrow x_t + dt \nabla \log \nu(x_t) + \mathcal{N}(0, 2dt \cdot I)$$

to iterations for $i = 1, \ldots, O(\mathrm{poly} \log n)$ of

$$x_{t+dt}^i \leftarrow x_t^i + dt \nabla \log \nu(x_t^{i-1}) + \mathcal{N}(0, 2dt \cdot I).$$

Error propagation

Recall that μ transport-stable if

$$\underbrace{\mathcal{W}_1(\tau_w\mu, \tau_{w'}\mu)}_{\text{Wasserstein distance w.r.t. Hamming metric}} \qquad \leqslant C \cdot \|w - w'\|_1.$$

Error propagation

Recall that μ transport-stable if

$$\underbrace{\mathcal{W}_1(\tau_w\mu, \tau_{w'}\mu)}_{\text{Wasserstein distance w.r.t. Hamming metric}} \leqslant C \cdot \|w - w'\|_1.$$

- The sampling error in one step gets multiplied by C in every future step.

Error propagation

Recall that μ transport-stable if

$$\underbrace{\mathcal{W}_1(\tau_w \mu, \tau_{w'} \mu)}_{\text{Wasserstein distance w.r.t. Hamming metric}} \leqslant C \cdot \|w - w'\|_1 .$$

▷ The sampling error in one step gets multiplied by C in every future step.

Lemma

$C = O(1)$ for spanning trees, etc.

Fact

$C = O(n)$ for any distribution.

Error propagation

Recall that μ transport-stable if

$$\underbrace{W_1(\tau_w \mu, \tau_{w'} \mu)}_{\text{Wasserstein distance w.r.t. Hamming metric}} \leqslant C \cdot \|w - w'\|_1 .$$

- The sampling error in one step gets multiplied by C in every future step.

Lemma

$C = O(1)$ for spanning trees, etc.

Fact

$C = O(n)$ for any distribution.

- Wasserstein accuracy $\mathbf{quasipoly}(n)^{-1}$ enough in continuous sampler.

Error propagation

Recall that μ transport-stable if

$$\underbrace{\mathcal{W}_1(\tau_w\mu, \tau_{w'}\mu)}_{\text{Wasserstein distance w.r.t. Hamming metric}} \leqslant C \cdot \|w - w'\|_1.$$

- The sampling error in one step gets multiplied by C in every future step.

Lemma

$C = O(1)$ for spanning trees, etc.

Fact

$C = O(n)$ for any distribution.

- Wasserstein accuracy $\mathsf{quasipoly}(n)^{-1}$ enough in continuous sampler.
- For $C = O(1)$, it is enough to have Wasserstein accuracy $\mathsf{poly}(n)^{-1}$.

Error propagation

Recall that μ transport-stable if

$$\underbrace{\mathcal{W}_1(\tau_w \mu, \tau_{w'} \mu)}_{\text{Wasserstein distance w.r.t. Hamming metric}} \leqslant C \cdot \|w - w'\|_1.$$

- The sampling error in one step gets multiplied by C in every future step.

Lemma

$C = O(1)$ for spanning trees, etc.

Fact

$C = O(n)$ for any distribution.

- Wasserstein accuracy $\mathbf{quasipoly}(n)^{-1}$ enough in continuous sampler.
- For $C = O(1)$, it is enough to have Wasserstein accuracy $\mathbf{poly}(n)^{-1}$.
- [A-Chewi-Vuong]: we can get TV-accurate samples in parallel.

Conclusion

- Parallel reduction of sampling to counting for a class of distributions.

Conclusion

- Parallel reduction of sampling to counting for a class of distributions.
- Open: Planar perfect matchings.

Conclusion

- Parallel reduction of sampling to counting for a class of distributions.
- Open: Planar perfect matchings.
- Open: With no assumption on μ, what is the parallel round complexity of sampling given $\mathsf{poly}(n)$ queries of $\sum_x e^{\langle w,x\rangle}\mu(x)$?

Conclusion

- Parallel reduction of sampling to counting for a class of distributions.
- Open: Planar perfect matchings.
- Open: With no assumption on μ, what is the parallel round complexity of sampling given $\textbf{poly}(n)$ queries of $\sum_x e^{\langle w,x \rangle} \mu(x)$?
- Open: With no assumption on μ, what is the parallel round complexity of sampling given $\textbf{poly}(n)$ queries of $\sum_{x \in \text{sub-cube}} \mu(x)$?

Conclusion

- Parallel reduction of sampling to counting for a class of distributions.
- Open: Planar perfect matchings.
- Open: With no assumption on μ, what is the parallel round complexity of sampling given $\mathbf{poly}(n)$ queries of $\sum_x e^{\langle w,x\rangle}\mu(x)$?
- Open: With no assumption on μ, what is the parallel round complexity of sampling given $\mathbf{poly}(n)$ queries of $\sum_{x\in\text{sub-cube}}\mu(x)$?

Thank you!