CS 263: Counting and Sampling

Nima Anari

Stanford University

slides for

Continuous Sampling

Review

Stochastic localization

For μ on $\mathbb{R}^n$, and adapted matrix process C_t, we define $\forall x$

$$d\mu_t(x) = \langle x - \mathsf{mean}(\mu_t), C_t\, dB_t \rangle \mu_t(x)$$

Review

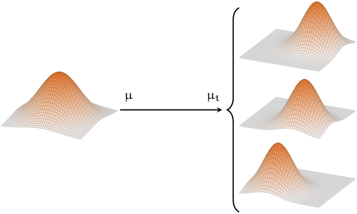
Stochastic localization

For μ on $\mathbb{R}^n$, and adapted matrix process C_t, we define $\forall x$

$$d\mu_t(x) = \langle x - \text{mean}(\mu_t), C_t dB_t \rangle \mu_t(x)$$

Distribution at time t is

$$\mu_t \propto \mu \cdot \exp\left(-\frac{x^\intercal A_t x}{2} + \langle h_t, x \rangle\right)$$



Review

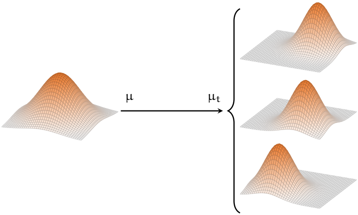
Stochastic localization

For μ on $\mathbb{R}^n$, and adapted matrix process C_t, we define $\forall x$

$$d\mu_t(x) = \langle x - \text{mean}(\mu_t), C_t \, dB_t \rangle \mu_t(x)$$

Distribution at time t is

$$\mu_t \propto \mu \cdot \exp\left(-\frac{x^\intercal A_t x}{2} + \langle h_t, x \rangle\right)$$



- Localization scheme: a measure-valued martingale.

Review

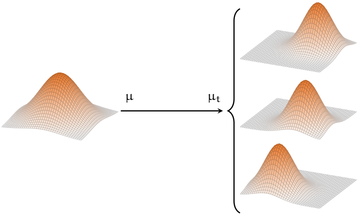
Stochastic localization

For μ on $\mathbb{R}^n$, and adapted matrix process C_t, we define $\forall x$

$$d\mu_t(x) = \langle x - \mathrm{mean}(\mu_t), C_t dB_t \rangle \mu_t(x)$$

Distribution at time t is

$$\mu_t \propto \mu \cdot \exp\left(-\frac{x^\intercal A_t x}{2} + \langle h_t, x \rangle\right)$$



- Localization scheme: a measure-valued martingale.
- Conservation of ϕ-entropies:

$$\mathbb{E}\left[\mathrm{Ent}^{\phi}_{\mu_t}[f]\right] \geqslant \gamma \cdot \mathrm{Ent}^{\phi}_{\mu}[f]$$

Review

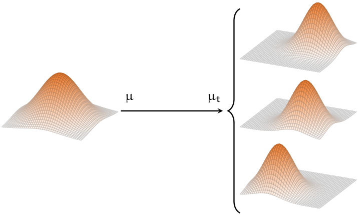
Stochastic localization

For μ on $\mathbb{R}^n$, and adapted matrix process C_t, we define $\forall x$

$$d\mu_t(x) = \langle x - \mathsf{mean}(\mu_t), C_t dB_t \rangle \mu_t(x)$$

Distribution at time t is

$$\mu_t \propto \mu \cdot \exp\left(-\frac{x^\intercal A_t x}{2} + \langle h_t, x \rangle\right)$$



- Localization scheme: a measure-valued martingale.
- Conservation of ϕ-entropies:

$$\mathbb{E}\left[\mathsf{Ent}^{\phi}_{\mu_t}[f]\right] \geqslant \gamma \cdot \mathsf{Ent}^{\phi}_{\mu}[f]$$

- For Markov chains constructed as NN°, we can transfer functional ineqs for μ_t to μ with a loss of γ.

Review

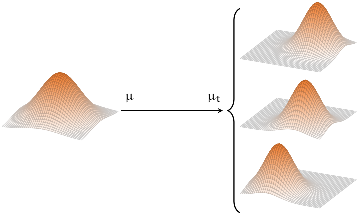
Stochastic localization

For μ on $\mathbb{R}^n$, and adapted matrix process C_t, we define $\forall x$

$$d\mu_t(x) = \langle x - \text{mean}(\mu_t), C_t dB_t \rangle \mu_t(x)$$

Distribution at time t is

$$\mu_t \propto \mu \cdot \exp\left(-\frac{x^\intercal A_t x}{2} + \langle h_t, x \rangle\right)$$



- Localization scheme: a measure-valued martingale.
- Conservation of ϕ-entropies:

$$\mathbb{E}\left[\text{Ent}^{\phi}_{\mu_t}[f]\right] \geqslant \gamma \cdot \text{Ent}^{\phi}_{\mu}[f]$$

- For Markov chains constructed as NN°, we can transfer functional ineqs for μ_t to μ with a loss of γ.

[Eldan-Koehler-Zeitouni]

Glauber for Ising models μ on $\{\pm 1\}^n$

$$\mu(x) \propto \exp\left(\frac{x^\intercal J x}{2} + \langle h, x \rangle\right)$$

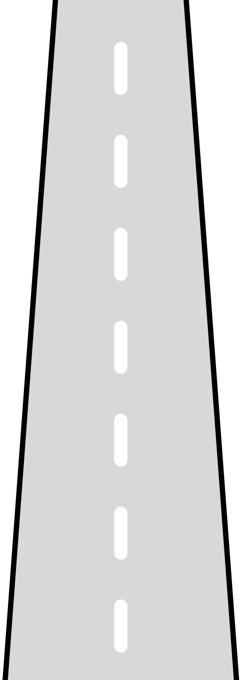
fast when $\lambda_{\max}(J) - \lambda_{\min}(J) < 1$.

Continuous Sampling

- Log-concave distributions
- Restricted Gaussian dynamics

Highlights

- Deterministic methods
- Markov chains

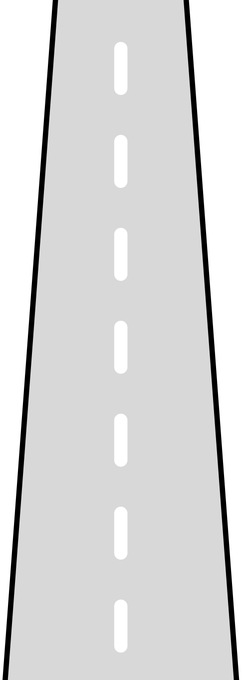


Continuous Sampling

- Log-concave distributions
- Restricted Gaussian dynamics

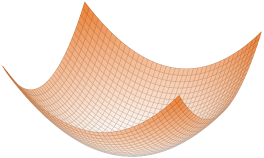
Highlights

- Deterministic methods
- Markov chains



Continuous space

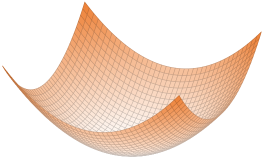
Optimization



$$U : \mathbb{R}^n \to \mathbb{R}$$

Continuous space

Optimization

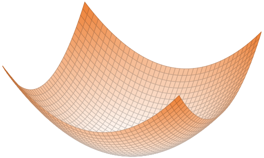


$$\mathfrak{U} : \mathbb{R}^n \to \mathbb{R}$$

- Tractable: $\mathfrak{U}$ convex

Continuous space

Optimization



$$U : \mathbb{R}^n \to \mathbb{R}$$

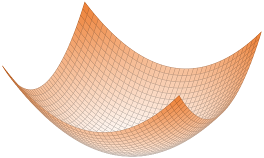
- Tractable: U convex
- Even better (well-conditioned):

$$\alpha I \preceq \nabla^2 U \preceq \beta I.$$

and condition number $\kappa = \beta/\alpha$.
Gradient descent: $\mathrm{poly}(\kappa, \log(1/\epsilon))$.

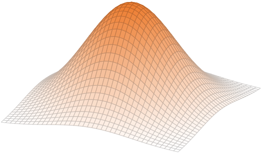
Continuous space

Optimization



$U : \mathbb{R}^n \to \mathbb{R}$

Sampling



μ on $\mathbb{R}^n$, $d\mu \propto e^{-U(x)}dx$

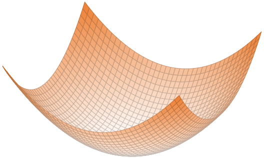
- Tractable: U convex
- Even better (well-conditioned):

$$\alpha I \preceq \nabla^2 U \preceq \beta I.$$

and condition number $\kappa = \beta/\alpha$.
Gradient descent: $\mathrm{poly}(\kappa, \log(1/\epsilon))$.

Continuous space

Optimization



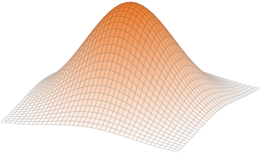
$$U : \mathbb{R}^n \to \mathbb{R}$$

- Tractable: U convex
- Even better (well-conditioned):

$$\alpha I \preceq \nabla^2 U \preceq \beta I.$$

and condition number $\kappa = \beta/\alpha$.
Gradient descent: $\mathrm{poly}(\kappa, \log(1/\epsilon))$.

Sampling

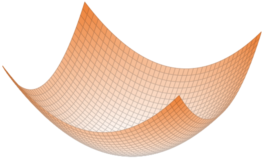


μ on $\mathbb{R}^n$, $d\mu \propto e^{-U(x)} dx$

- Tractable: μ is log-concave, i.e., U convex

Continuous space

Optimization



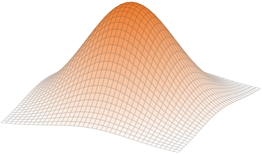
$$U : \mathbb{R}^n \to \mathbb{R}$$

- Tractable: U convex
- Even better (well-conditioned):

$$\alpha I \preceq \nabla^2 U \preceq \beta I.$$

and condition number $\kappa = \beta/\alpha$.
Gradient descent: $\mathsf{poly}(\kappa, \log(1/\epsilon))$.

Sampling

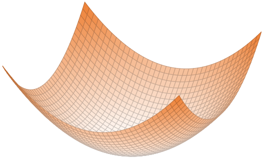


μ on $\mathbb{R}^n$, $d\mu \propto e^{-U(x)} dx$

- Tractable: μ is log-concave, i.e., U convex
- Open: can we sample (say within $d_{TV} \leqslant 0.1$) in $\mathsf{poly}(\kappa)$ steps?

Continuous space

Optimization



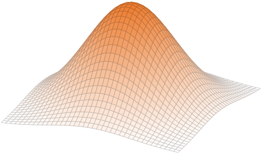
$$U : \mathbb{R}^n \to \mathbb{R}$$

- Tractable: U convex
- Even better (well-conditioned):

$$\alpha I \preceq \nabla^2 U \preceq \beta I.$$

and condition number $\kappa = \beta/\alpha$.
Gradient descent: $\mathsf{poly}(\kappa, \log(1/\epsilon))$.

Sampling



μ on $\mathbb{R}^n$, $d\mu \propto e^{-U(x)} dx$

- Tractable: μ is log-concave, i.e., U convex
- Open: can we sample (say within $d_{\mathsf{TV}} \leqslant 0.1$) in $\mathsf{poly}(\kappa)$ steps?
- Best known: $\widetilde{O}(\sqrt{n}) \cdot \mathsf{poly}(\kappa)$ [Altschuler-Chewi]

We will see how to sample in $\mathsf{poly}(n, \kappa)$.

Rejection sampling

- As a first attempt, let us try rejection sampling.

Rejection sampling

- As a first attempt, let us try rejection sampling.
- Well-conditionedness is measuring local deviation from Gaussianity.

Rejection sampling

- As a first attempt, let us try rejection sampling.
- Well-conditionedness is measuring local deviation from Gaussianity.
- Propose Gaussians and accept/reject into μ?

Rejection sampling

- As a first attempt, let us try rejection sampling.
- Well-conditionedness is measuring local deviation from Gaussianity.
- Propose Gaussians and accept/reject into μ?
- Wlog assume μ has mode at $\underset{\uparrow}{0}$ and $\underset{\uparrow}{U(0)} = 0$.

mode with GD

subtract constant

Rejection sampling

- As a first attempt, let us try rejection sampling.
- Well-conditionedness is measuring local deviation from Gaussianity.
- Propose Gaussians and accept/reject into μ?
- Wlog assume μ has mode at $\underset{\uparrow}{0}$ and $\underset{\uparrow}{U(0)} = 0$.

  mode with GD

  subtract constant

- Proposal dist $\nu = \mathcal{N}(0, I/\alpha)$:

$$d\nu \propto \exp\left(-\frac{\alpha\cdot\|x\|^2}{2}\right) dx$$

Rejection sampling

- As a first attempt, let us try rejection sampling.
- Well-conditionedness is measuring local deviation from Gaussianity.
- Propose Gaussians and accept/reject into μ?
- Wlog assume μ has mode at $\underset{\uparrow}{0}$ and $\underset{\uparrow}{U(0)} = 0$.

mode with GD

subtract constant

- Proposal dist $\nu = \mathcal{N}(0, I/\alpha)$:

$$d\nu \propto \exp\left(-\frac{\alpha \cdot \|x\|^2}{2}\right) dx$$

- Observation: everywhere we have

$$\frac{\alpha\|x\|^2}{2} \leqslant U(x) \leqslant \frac{\beta\|x\|^2}{2}$$

Because $U(x) - \alpha\|x\|^2/2$ is convex and $U(x) - \beta\|x\|^2/2$ is concave.

Rejection sampling

- As a first attempt, let us try rejection sampling.
- Well-conditionedness is measuring local deviation from Gaussianity.
- Propose Gaussians and accept/reject into μ?
- Wlog assume μ has mode at 0 and $U(0) = 0$.

mode with GD

subtract constant

- Proposal dist $\nu = \mathcal{N}(0, I/\alpha)$:

$$d\nu \propto \exp\left(-\frac{\alpha \cdot \|x\|^2}{2}\right) dx$$

- Observation: everywhere we have

$$\frac{\alpha\|x\|^2}{2} \leqslant U(x) \leqslant \frac{\beta\|x\|^2}{2}$$

Because $U(x) - \alpha\|x\|^2/2$ is convex and $U(x) - \beta\|x\|^2/2$ is concave.

- Acceptance prob for sample x:

$$\exp\left(\frac{\alpha\|x\|^2}{2} - U(x)\right) \leqslant 1$$

Rejection sampling

- As a first attempt, let us try rejection sampling.
- Well-conditionedness is measuring local deviation from Gaussianity.
- Propose Gaussians and accept/reject into μ?
- Wlog assume μ has mode at $\underset{\uparrow}{0}$ and $\underset{\uparrow}{U(0)} = 0$.

mode with GD

subtract constant

- Proposal dist $\nu = \mathcal{N}(0, I/\alpha)$:

$$d\nu \propto \exp\left(-\frac{\alpha \cdot \|x\|^2}{2}\right) dx$$

- Observation: everywhere we have

$$\frac{\alpha\|x\|^2}{2} \leqslant U(x) \leqslant \frac{\beta\|x\|^2}{2}$$

Because $U(x) - \alpha\|x\|^2/2$ is convex and $U(x) - \beta\|x\|^2/2$ is concave.

- Acceptance prob for sample x:

$$\exp\left(\frac{\alpha\|x\|^2}{2} - U(x)\right) \leqslant 1$$

- Chance of acceptance:

$$\frac{\int \exp(-U(x))\,dx}{\int \exp\left(-\frac{\alpha\|x\|^2}{2}\right) dx} \geqslant \frac{\int \exp\left(-\frac{\beta\|x\|^2}{2}\right) dx}{\int \exp\left(-\frac{\alpha\|x\|^2}{2}\right) dx}$$

Rejection sampling

- As a first attempt, let us try rejection sampling.
- Well-conditionedness is measuring local deviation from Gaussianity.
- Propose Gaussians and accept/reject into μ?
- Wlog assume μ has mode at 0 and $U(0) = 0$.

  mode with GD

  subtract constant
- Proposal dist $\nu = \mathcal{N}(0, I/\alpha)$:
  $$d\nu \propto \exp\left(-\frac{\alpha \cdot \|x\|^2}{2}\right) dx$$
- Observation: everywhere we have
  $$\frac{\alpha\|x\|^2}{2} \leqslant U(x) \leqslant \frac{\beta\|x\|^2}{2}$$
  Because $U(x) - \alpha\|x\|^2/2$ is convex and $U(x) - \beta\|x\|^2/2$ is concave.
- Acceptance prob for sample x:
  $$\exp\left(\frac{\alpha\|x\|^2}{2} - U(x)\right) \leqslant 1$$
- Chance of acceptance:
  $$\frac{\int \exp(-U(x))\, dx}{\int \exp\left(-\frac{\alpha\|x\|^2}{2}\right) dx} \geqslant \frac{\int \exp\left(-\frac{\beta\|x\|^2}{2}\right) dx}{\int \exp\left(-\frac{\alpha\|x\|^2}{2}\right) dx}$$
- This is $(\sqrt{2\pi\alpha}/\sqrt{2\pi\beta})^n = 1/\kappa^{n/2}$.

Rejection sampling works for $\kappa \leqslant 1 + \widetilde{O}(1/n)$

Restricted Gaussian dynamics [Lee-Shen-Tian]

Markov chain

- μ: dist on $\mathbb{R}^n$
- N: $x \mapsto y = x + g$ for $g \sim \mathcal{N}(0, I/\eta)$
- $P = NN^{\circ}$: then sample z w.p. $\propto$

$$\mu(z) \cdot e^{-\eta \cdot \|z-y\|^2/2}$$

$\uparrow$

restricted Gaussian

Restricted Gaussian dynamics [Lee-Shen-Tian]

Markov chain

- μ: dist on $\mathbb{R}^n$
- N: $x \mapsto y = x + g$ for $g \sim \mathcal{N}(0, I/\eta)$
- $P = NN^\circ$: then sample z w.p. $\propto$

$$\mu(z) \cdot e^{-\eta \cdot \|z-y\|^2/2}$$

↑

restricted Gaussian

- If μ has cond $\kappa = \beta/\alpha$, then cond number for restricted Gaussian is:

$$\kappa' = \frac{\beta+\eta}{\alpha+\eta}$$

Restricted Gaussian dynamics [Lee-Shen-Tian]

Markov chain

- μ: dist on $\mathbb{R}^n$
- N: $x \mapsto y = x + g$ for $g \sim \mathcal{N}(0, I/\eta)$
- $P = NN^\circ$: then sample z w.p. $\propto$

$$\mu(z) \cdot \underset{\substack{\uparrow \\ \text{restricted Gaussian}}}{e^{-\eta \cdot \|z-y\|^2/2}}$$

- If μ has cond $\kappa = \beta/\alpha$, then cond number for restricted Gaussian is:

$$\kappa' = \frac{\beta+\eta}{\alpha+\eta}$$

- As long as $\eta \geqslant n\beta$, we can use rejection sampling: $\kappa' \leqslant 1 + 1/n$.

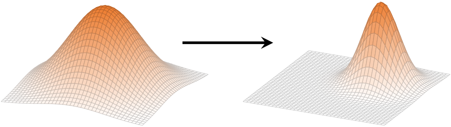
Restricted Gaussian dynamics [Lee-Shen-Tian]

Markov chain

- μ: dist on $\mathbb{R}^n$
- N: $x \mapsto y = x + g$ for $g \sim \mathcal{N}(0, I/\eta)$
- $P = NN^\circ$: then sample z w.p. $\propto$

$$\mu(z) \cdot e^{-\eta \cdot \|z-y\|^2/2}$$

restricted Gaussian



- If μ has cond $\kappa = \beta/\alpha$, then cond number for restricted Gaussian is:

$$\kappa' = \frac{\beta+\eta}{\alpha+\eta}$$

- As long as $\eta \geqslant n\beta$, we can use rejection sampling: $\kappa' \leqslant 1 + 1/n$.

Restricted Gaussian dynamics [Lee-Shen-Tian]

Markov chain

- μ: dist on $\mathbb{R}^n$
- N: $x \mapsto y = x + g$ for $g \sim \mathcal{N}(0, I/\eta)$
- $P = NN^\circ$: then sample z w.p. $\propto$

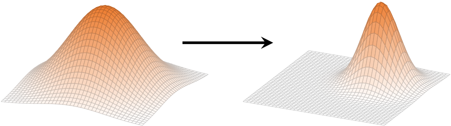
$$\mu(z) \cdot e^{-\eta \cdot \|z-y\|^2/2}$$

restricted Gaussian

- If μ has cond $\kappa = \beta/\alpha$, then cond number for restricted Gaussian is:

$$\kappa' = \frac{\beta+\eta}{\alpha+\eta}$$

- As long as $\eta \geqslant n\beta$, we can use rejection sampling: $\kappa' \leqslant 1 + 1/n$.



- All that remains is to bound mixing time of Markov chain.

Restricted Gaussian dynamics [Lee-Shen-Tian]

Markov chain

- μ: dist on $\mathbb{R}^n$
- N: $x \mapsto y = x + g$ for $g \sim \mathcal{N}(0, I/\eta)$
- $P = NN^\circ$: then sample z w.p. $\propto$

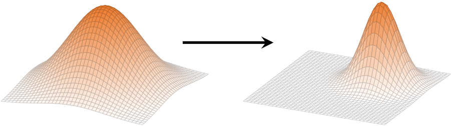
$$\mu(z) \cdot \underset{\uparrow}{e^{-\eta \cdot \|z-y\|^2/2}}$$

restricted Gaussian

- If μ has cond $\kappa = \beta/\alpha$, then cond number for restricted Gaussian is:

$$\kappa' = \frac{\beta+\eta}{\alpha+\eta}$$

- As long as $\eta \geq n\beta$, we can use rejection sampling: $\kappa' \leq 1 + 1/n$.



- All that remains is to bound mixing time of Markov chain.
- Will show

$$t_{\text{mix}} \leq \text{poly}(\eta/\alpha, n)$$

Restricted Gaussian dynamics [Lee-Shen-Tian]

Markov chain

- μ: dist on $\mathbb{R}^n$
- N: $x \mapsto y = x + g$ for $g \sim \mathcal{N}(0, I/\eta)$
- $P = NN^\circ$: then sample z w.p. $\propto$

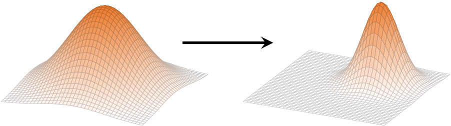
$$\mu(z) \cdot \underset{\uparrow}{e^{-\eta\cdot\|z-y\|^2/2}}$$

restricted Gaussian

- If μ has cond $\kappa = \beta/\alpha$, then cond number for restricted Gaussian is:

$$\kappa' = \frac{\beta+\eta}{\alpha+\eta}$$

- As long as $\eta \geqslant n\beta$, we can use rejection sampling: $\kappa' \leqslant 1 + 1/n$.



- All that remains is to bound mixing time of Markov chain.
- Will show

$$t_{\text{mix}} \leqslant \text{poly}(\eta/\alpha, n)$$

- In fact, we will show χ^2 contraction:

$$t_{\text{rel}} \leqslant \frac{\alpha+\eta}{\alpha}$$

Restricted Gaussian dynamics [Lee-Shen-Tian]

Markov chain

- μ: dist on $\mathbb{R}^n$
- N: $x \mapsto y = x + g$ for $g \sim \mathcal{N}(0, I/\eta)$
- $P = NN^\circ$: then sample z w.p. $\propto$

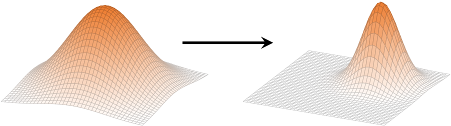
$$\mu(z) \cdot \underset{\uparrow}{e^{-\eta \cdot \|z-y\|^2/2}}$$

restricted Gaussian

- If μ has cond $\kappa = \beta/\alpha$, then cond number for restricted Gaussian is:

$$\kappa' = \frac{\beta+\eta}{\alpha+\eta}$$

- As long as $\eta \geqslant n\beta$, we can use rejection sampling: $\kappa' \leqslant 1 + 1/n$.



- All that remains is to bound mixing time of Markov chain.
- Will show

$$t_{\text{mix}} \leqslant \text{poly}(\eta/\alpha, n)$$

- In fact, we will show χ^2 contraction:

$$t_{\text{rel}} \leqslant \frac{\alpha+\eta}{\alpha}$$

- Warm start: Gaussian!

Restricted Gaussian dynamics [Lee-Shen-Tian]

Markov chain

- μ: dist on $\mathbb{R}^n$
- N: $x \mapsto y = x + g$ for $g \sim \mathcal{N}(0, I/\eta)$
- $P = NN^\circ$: then sample z w.p. $\propto$

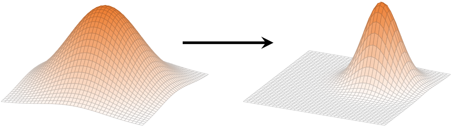
$$\mu(z) \cdot e^{-\eta \cdot \|z-y\|^2/2}$$

restricted Gaussian

- If μ has cond $\kappa = \beta/\alpha$, then cond number for restricted Gaussian is:

$$\kappa' = \frac{\beta+\eta}{\alpha+\eta}$$

- As long as $\eta \geqslant n\beta$, we can use rejection sampling: $\kappa' \leqslant 1 + 1/n$.



- All that remains is to bound mixing time of Markov chain.
- Will show

$$t_{\text{mix}} \leqslant \text{poly}(\eta/\alpha, n)$$

- In fact, we will show χ^2 contraction:

$$t_{\text{rel}} \leqslant \frac{\alpha+\eta}{\alpha}$$

- Warm start: Gaussian!
- [Chen-Eldan]: same for entropy.

Stochastic localization

Stochastic localization

For μ on $\mathbb{R}^n$, and adapted matrix process C_t, we define $\forall x$

$$d\mu_t(x) = \langle x - \mathsf{mean}(\mu_t), C_t dB_t \rangle \mu_t(x)$$

Stochastic localization

Stochastic localization

For μ on $\mathbb{R}^n$, and adapted matrix process C_t, we define $\forall x$

$$d\mu_t(x) = \langle x - \mathsf{mean}(\mu_t), C_t dB_t \rangle \mu_t(x)$$

- We will use $C_t = I$ today!

Stochastic localization

Stochastic localization

For μ on $\mathbb{R}^n$, and adapted matrix process C_t, we define $\forall x$

$$d\mu_t(x) = \langle x - \mathsf{mean}(\mu_t), C_t dB_t \rangle \mu_t(x)$$

- We will use $C_t = I$ today!
- We will always deal with restricted Gaussians:

$$\mu_t \propto \mu \cdot \mathsf{exp}\left(-\frac{t\|x\|^2}{2} + \langle h_t, x \rangle\right)$$

Stochastic localization

Stochastic localization

For μ on $\mathbb{R}^n$, and adapted matrix process C_t, we define $\forall x$

$$d\mu_t(x) = \langle x - \text{mean}(\mu_t), C_t dB_t \rangle \mu_t(x)$$

- We will use $C_t = I$ today!
- We will always deal with restricted Gaussians:
$$\mu_t \propto \mu \cdot \exp\left(-\frac{t\|x\|^2}{2} + \langle h_t, x \rangle\right)$$
- μ_t is a posterior: it is $N^\circ(h_t/t, \cdot)$, where N is the Gaussian noise operator adding $\mathcal{N}(0, I/t)$.

Stochastic localization

Stochastic localization

For μ on $\mathbb{R}^n$, and adapted matrix process C_t, we define $\forall x$

$$d\mu_t(x) = \langle x - \text{mean}(\mu_t), C_t dB_t \rangle \mu_t(x)$$

- We will use $C_t = I$ today!
- We will always deal with restricted Gaussians:
$$\mu_t \propto \mu \cdot \exp\left(-\frac{t\|x\|^2}{2} + \langle h_t, x \rangle \right)$$
- μ_t is a posterior: it is $N^\circ(h_t/t, \cdot)$, where N is the Gaussian noise operator adding $\mathcal{N}(0, I/t)$.
- In fact, h_t/t, the center of the Gaussian, follows nice dist:
$$\frac{h_t}{t} \sim \mu * \mathcal{N}(0, I/t) = \mu N.$$

Stochastic localization

Stochastic localization

For μ on $\mathbb{R}^n$, and adapted matrix process C_t, we define $\forall x$

$$d\mu_t(x) = \langle x - \text{mean}(\mu_t), C_t dB_t \rangle \mu_t(x)$$

- We will use $C_t = I$ today!
- We will always deal with restricted Gaussians:
$$\mu_t \propto \mu \cdot \exp\left(-\frac{t\|x\|^2}{2} + \langle h_t, x\rangle\right)$$
- μ_t is a posterior: it is $N^\circ(h_t/t, \cdot)$, where N is the Gaussian noise operator adding $\mathcal{N}(0, I/t)$.
- In fact, h_t/t, the center of the Gaussian, follows nice dist:
$$\frac{h_t}{t} \sim \mu * \mathcal{N}(0, I/t) = \mu N.$$
- Exercise: h_t follows
$$dh_t = \text{mean}(\mu_t)dt + dB_t$$

Stochastic localization

Stochastic localization

For μ on $\mathbb{R}^n$, and adapted matrix process C_t, we define $\forall x$

$$d\mu_t(x) = \langle x - \text{mean}(\mu_t), C_t dB_t \rangle \mu_t(x)$$

- We will use $C_t = I$ today!
- We will always deal with restricted Gaussians:
$$\mu_t \propto \mu \cdot \exp\left(-\frac{t\|x\|^2}{2} + \langle h_t, x \rangle\right)$$
- μ_t is a posterior: it is $N^\circ(h_t/t, \cdot)$, where N is the Gaussian noise operator adding $\mathcal{N}(0, I/t)$.
- In fact, h_t/t, the center of the Gaussian, follows nice dist:
$$\frac{h_t}{t} \sim \mu * \mathcal{N}(0, I/t) = \mu N.$$
- Exercise: h_t follows
$$dh_t = \text{mean}(\mu_t)dt + dB_t$$
- This means that
$$\text{Ent}^{\phi}_{\mu}[f] - \text{Ent}^{\phi}_{\mu N}[f] = \mathbb{E}\left[\text{Ent}^{\phi}_{\mu_t}[f]\right]$$

Stochastic localization

Stochastic localization

For μ on $\mathbb{R}^n$, and adapted matrix process C_t, we define $\forall x$

$$d\mu_t(x) = \langle x - \text{mean}(\mu_t), C_t dB_t \rangle \mu_t(x)$$

- We will use $C_t = I$ today!
- We will always deal with restricted Gaussians:
$$\mu_t \propto \mu \cdot \exp\left(-\frac{t\|x\|^2}{2} + \langle h_t, x \rangle\right)$$
- μ_t is a posterior: it is $N^\circ(h_t/t, \cdot)$, where N is the Gaussian noise operator adding $\mathcal{N}(0, I/t)$.
- In fact, h_t/t, the center of the Gaussian, follows nice dist:
$$\frac{h_t}{t} \sim \mu * \mathcal{N}(0, I/t) = \mu N.$$
- Exercise: h_t follows
$$dh_t = \text{mean}(\mu_t)dt + dB_t$$
- This means that
$$\text{Ent}^{\phi}_{\mu}[f] - \text{Ent}^{\phi}_{\mu N}[f] = \mathbb{E}\left[\text{Ent}^{\phi}_{\mu_t}[f]\right]$$
- So, as long as we prove γ-approximate conservation of Ent^{ϕ}, we have proved N contracts Ent^{ϕ} by $1 - \gamma$. Do this at $t = \eta$. ☺

Conservation

- Let us specialize to $\mathsf{Ent}^{\Phi} = \mathsf{Var}$.

Conservation

- Let us specialize to $\mathsf{Ent}^{\Phi} = \mathsf{Var}$.
- We saw last time that $d\,\mathsf{Var}_{\mu_t}[f] =$

  $$(\text{martingale term}) - v_t^{\intercal} \Sigma_t v_t \, dt.$$

  Here $\Sigma_t = I$ and

  $$v_t = \mathbb{E}_{x \sim \mu_t}[f(x)(x - \mathsf{mean}(\mu_t))].$$

Conservation

- Let us specialize to $\mathsf{Ent}^{\phi} = \mathsf{Var}$.
- We saw last time that $d\,\mathsf{Var}_{\mu_t}[f] =$

  $$(\text{martingale term}) - v_t^\intercal \Sigma_t v_t \, dt.$$

  Here $\Sigma_t = I$ and

  $$v_t = \mathbb{E}_{x \sim \mu_t}[f(x)(x - \mathsf{mean}(\mu_t))].$$
- Claim: by Cauchy-Schwartz

  $$\|v_t\|^2 \leqslant \mathsf{Var}_{\mu_t}[f] \cdot \lambda_{\mathsf{max}}(\mathsf{cov}(\mu_t)).$$

Conservation

- Let us specialize to $\mathsf{Ent}^{\Phi} = \mathsf{Var}$.
- We saw last time that $d\,\mathsf{Var}_{\mu_t}[f] =$

  $\quad$ (martingale term) $- v_t^{\intercal} \Sigma_t v_t dt$.

  Here $\Sigma_t = I$ and

  $$v_t = \mathbb{E}_{x\sim\mu_t}[f(x)(x - \mathsf{mean}(\mu_t))].$$
- Claim: by Cauchy-Schwartz

  $$\|v_t\|^2 \leqslant \mathsf{Var}_{\mu_t}[f] \cdot \lambda_{\mathsf{max}}(\mathsf{cov}(\mu_t)).$$
- This means that

  $$\mathbb{E}[d\,\mathsf{Var}_{\mu_t}[f] \mid \mathcal{F}_t] \geqslant \\ -\lambda_{\mathsf{max}}(\mathsf{cov}(\mu_t))\,\mathsf{Var}_{\mu_t}[f]dt$$

Conservation

- Let us specialize to $\mathsf{Ent}^{\phi} = \mathsf{Var}$.
- We saw last time that $d\,\mathsf{Var}_{\mu_t}[f] =$

  $(\text{martingale term}) - v_t^\intercal \Sigma_t v_t dt$.

  Here $\Sigma_t = I$ and

  $v_t = \mathbb{E}_{x \sim \mu_t}[f(x)(x - \mathsf{mean}(\mu_t))]$.
- Claim: by Cauchy-Schwartz

  $\|v_t\|^2 \leqslant \mathsf{Var}_{\mu_t}[f] \cdot \lambda_{\mathsf{max}}(\mathsf{cov}(\mu_t))$.
- This means that

$$\begin{gathered}\mathbb{E}[d\,\mathsf{Var}_{\mu_t}[f] \mid \mathcal{F}_t] \geqslant \\ -\lambda_{\mathsf{max}}(\mathsf{cov}(\mu_t))\,\mathsf{Var}_{\mu_t}[f]dt\end{gathered}$$

- If $\lambda_{\mathsf{max}}(\mathsf{cov}(\mu_t)) \leqslant g(t)$, then we get:

$$\gamma \geqslant \exp\left(-\int_0^\eta g(t)dt\right)$$

Conservation

- Let us specialize to $\mathrm{Ent}^{\phi} = \mathrm{Var}$.
- We saw last time that $d\,\mathrm{Var}_{\mu_t}[f] =$
$$\text{(martingale term)} - v_t^\intercal \Sigma_t v_t \, dt.$$
Here $\Sigma_t = I$ and
$$v_t = \mathbb{E}_{x\sim\mu_t}[f(x)(x - \mathrm{mean}(\mu_t))].$$
- Claim: by Cauchy-Schwartz
$$\|v_t\|^2 \leqslant \mathrm{Var}_{\mu_t}[f] \cdot \lambda_{\mathrm{max}}(\mathrm{cov}(\mu_t)).$$
- This means that
$$\mathbb{E}[d\,\mathrm{Var}_{\mu_t}[f] \mid \mathcal{F}_t] \geqslant -\lambda_{\mathrm{max}}(\mathrm{cov}(\mu_t))\,\mathrm{Var}_{\mu_t}[f]dt$$
- If $\lambda_{\mathrm{max}}(\mathrm{cov}(\mu_t)) \leqslant g(t)$, then we get:
$$\gamma \geqslant \exp\left(-\int_0^\eta g(t)dt\right)$$
- But note that μ_t is $(\alpha + t)$-strongly log-concave:
$$-\nabla^2 \log \mu_t \succeq (\alpha + t)I$$

Conservation

- Let us specialize to $\mathsf{Ent}^{\phi} = \mathsf{Var}$.
- We saw last time that $d\,\mathsf{Var}_{\mu_t}[f] =$
  $$(\text{martingale term}) - v_t^T \Sigma_t v_t \, dt.$$
  Here $\Sigma_t = I$ and
  $$v_t = \mathbb{E}_{x \sim \mu_t}[f(x)(x - \mathsf{mean}(\mu_t))].$$
- Claim: by Cauchy-Schwartz
  $$\|v_t\|^2 \leqslant \mathsf{Var}_{\mu_t}[f] \cdot \lambda_{\mathsf{max}}(\mathsf{cov}(\mu_t)).$$
- This means that
  $$\mathbb{E}[d\,\mathsf{Var}_{\mu_t}[f] \mid \mathcal{F}_t] \geqslant -\lambda_{\mathsf{max}}(\mathsf{cov}(\mu_t))\,\mathsf{Var}_{\mu_t}[f] dt$$
- If $\lambda_{\mathsf{max}}(\mathsf{cov}(\mu_t)) \leqslant g(t)$, then we get:
  $$\gamma \geqslant \exp\left(-\int_0^\eta g(t) dt\right)$$
- But note that μ_t is $(\alpha + t)$-strongly log-concave:
  $$-\nabla^2 \log \mu_t \succeq (\alpha + t) I$$
- We can compare $\mathsf{cov}(\mu_t)$ with that of $\mathcal{N}(0, I/(\alpha + t))$:
  $$\mathsf{cov}(\mu_t) \preceq \frac{I}{\alpha+t}$$

Conservation

- Let us specialize to $\mathsf{Ent}^{\phi} = \mathsf{Var}$.
- We saw last time that $d\,\mathsf{Var}_{\mu_t}[f] =$
$$(\text{martingale term}) - v_t^T \Sigma_t v_t dt.$$
Here $\Sigma_t = I$ and
$$v_t = \mathbb{E}_{x\sim\mu_t}[f(x)(x - \mathsf{mean}(\mu_t))].$$
- Claim: by Cauchy-Schwartz
$$\|v_t\|^2 \leqslant \mathsf{Var}_{\mu_t}[f] \cdot \lambda_{\mathsf{max}}(\mathsf{cov}(\mu_t)).$$
- This means that
$$\mathbb{E}[d\,\mathsf{Var}_{\mu_t}[f] \mid \mathcal{F}_t] \geqslant -\lambda_{\mathsf{max}}(\mathsf{cov}(\mu_t))\,\mathsf{Var}_{\mu_t}[f]dt$$
- If $\lambda_{\mathsf{max}}(\mathsf{cov}(\mu_t)) \leqslant g(t)$, then we get:
$$\gamma \geqslant \mathsf{exp}\left(-\int_0^{\eta} g(t)dt\right)$$
- But note that μ_t is $(\alpha + t)$-strongly log-concave:
$$-\nabla^2 \log \mu_t \succeq (\alpha + t)I$$
- We can compare $\mathsf{cov}(\mu_t)$ with that of $\mathcal{N}(0, I/(\alpha + t))$:
$$\mathsf{cov}(\mu_t) \preceq \frac{I}{\alpha+t}$$
- Proof sketch: reduce to 1D using the fact that marginals of log-concave are log-concave.

Conservation

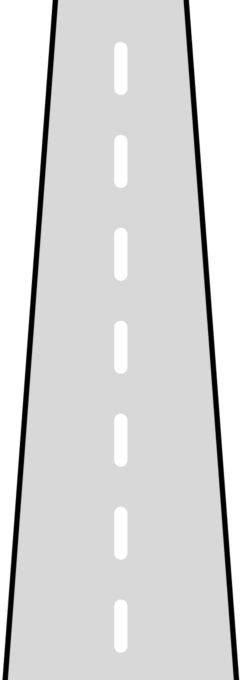
- Let us specialize to $\mathrm{Ent}^{\phi} = \mathrm{Var}$.
- We saw last time that $d\,\mathrm{Var}_{\mu_t}[f] =$
  $$(\text{martingale term}) - v_t^{\intercal} \Sigma_t v_t \, dt.$$
  Here $\Sigma_t = I$ and
  $$v_t = \mathbb{E}_{x \sim \mu_t}[f(x)(x - \mathrm{mean}(\mu_t))].$$
- Claim: by Cauchy-Schwartz
  $$\|v_t\|^2 \leqslant \mathrm{Var}_{\mu_t}[f] \cdot \lambda_{\max}(\mathrm{cov}(\mu_t)).$$
- This means that
  $$\mathbb{E}[d\,\mathrm{Var}_{\mu_t}[f] \mid \mathcal{F}_t] \geqslant -\lambda_{\max}(\mathrm{cov}(\mu_t))\,\mathrm{Var}_{\mu_t}[f] dt$$
- If $\lambda_{\max}(\mathrm{cov}(\mu_t)) \leqslant g(t)$, then we get:
  $$\gamma \geqslant \exp\left(-\int_0^{\eta} g(t) dt\right)$$
- But note that μ_t is $(\alpha + t)$-strongly log-concave:
  $$-\nabla^2 \log \mu_t \succeq (\alpha + t) I$$
- We can compare $\mathrm{cov}(\mu_t)$ with that of $\mathcal{N}(0, I/(\alpha + t))$:
  $$\mathrm{cov}(\mu_t) \preceq \frac{I}{\alpha + t}$$
- Proof sketch: reduce to 1D using the fact that marginals of log-concave are log-concave.
- This gives us $g(t) = 1/(\alpha + t)$, so
  $$\int_0^{\eta} g(t) dt = \log\left(\frac{\alpha + \eta}{\alpha}\right)$$
  which means $\gamma \geqslant \alpha/(\alpha + \eta)$. 🙂

Continuous Sampling

- Log-concave distributions
- Restricted Gaussian dynamics

Highlights

- Deterministic methods
- Markov chains

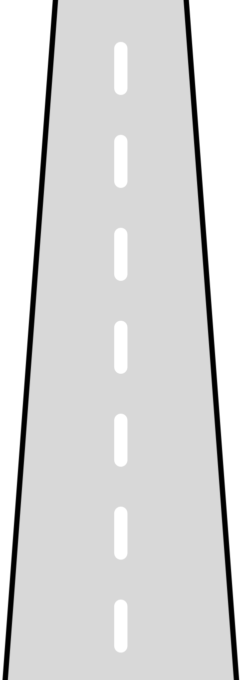


Continuous Sampling

- Log-concave distributions
- Restricted Gaussian dynamics

Highlights

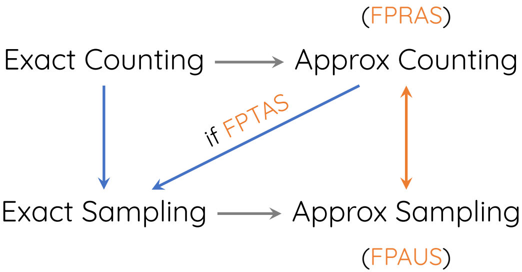
- Deterministic methods
- Markov chains



Theorem [Jerrum-Valiant-Vazirani]

For “self-reducible” problems:

approx counting ≡ approx sampling



(FPRAS)

Exact Counting

Approx Counting

if FPTAS

Exact Sampling

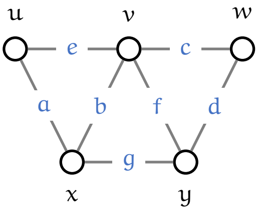
Approx Sampling

(FPAUS)

arrows are poly-time reductions

Deterministic counting

Determinant-based counting

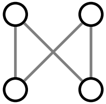


$$\begin{array}{c|ccccccc} & a & b & c & d & e & f & g \\ \hline u & +1 & 0 & 0 & 0 & +1 & 0 & 0 \\ v & 0 & -1 & +1 & 0 & -1 & -1 & 0 \\ w & 0 & 0 & -1 & +1 & 0 & 0 & 0 \\ x & -1 & +1 & 0 & 0 & 0 & 0 & -1 \\ y & 0 & 0 & 0 & -1 & 0 & +1 & +1 \end{array}$$

$$\#\text{spanning trees} = \det(\text{matrix})$$

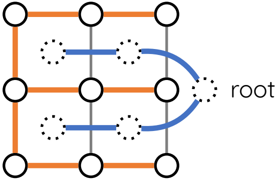
Laplacian, drop one row+col

- [Pólya]'s scheme:

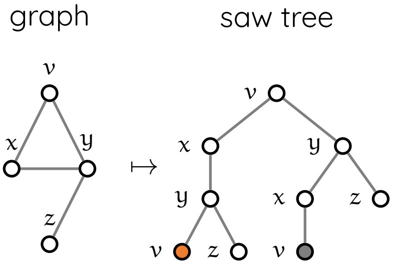


$$\det\left(\begin{bmatrix} +1 & -1 \\ +1 & +1 \end{bmatrix}\right) = \operatorname{per}\left(\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}\right)$$

- Good signing exists for planar graphs [Fisher-Kasteleyn-Temperley].



Correlation decay



- Root marginals are the same.
- Weak spatial mixing:

$$d_{TV}(\text{root} \mid \sigma, \text{root} \mid \sigma') \to 0$$

as the following goes to ∞:

$$\min\{d(\text{root}, u) \mid u \in S\}.$$

- Strong spatial mixing:

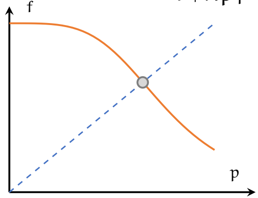
$$d_{TV}(\text{root} \mid \sigma, \text{root} \mid \sigma') \to 0$$

as the following goes to ∞:

$$\min\{d(\text{root}, u) \mid \sigma(u) \neq \sigma'(u)\}.$$

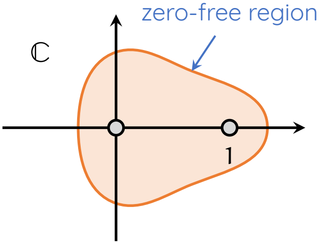
- [Weitz]'s alg forms truncated saw tree and uses recursion:

$$f(p_1, \ldots, p_d) = \frac{1}{1+\lambda p_1 \cdots p_d}$$

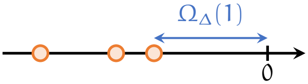


- Attractive exactly when $\lambda < \lambda_c(\Delta)$

[Barvinok]'s method



- [Barvinok]: approx $p(1)$ via $p^{(i)}(0)$ for $i = 0, \ldots, O(\log \deg(p))$
- Idea 1: Riemann map from disk
- Idea 2: trunc Taylor series of $\log p$
- Matchings via [Heilmann-Lieb]:



Markov chains

Transport

- Influence: X, X' differing in coord j:

$$d_{TV}\big(\mathrm{dist}(X_i \mid X_{-i}), \mathrm{dist}(X'_i \mid X'_{-i})\big)$$

- Call maximum value $\mathcal{J}[j \to i]$.

Dobrushin's condition

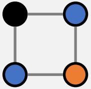
If columns of $\mathcal{J}$ sum to $\leqslant 1 - \delta$, then

$$\mathcal{W}(\nu P, \nu' P) \leqslant (1 - \delta/n)\, \mathcal{W}(\nu, \nu')$$

$$t_{\mathrm{mix}}(\epsilon) = O\big(\tfrac{1}{\delta} n \log(n/\epsilon)\big)$$

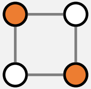
Example: coloring

- $\Omega = [q]^n$
- $\mathcal{J} \leqslant \frac{1}{q-\Delta} \cdot \mathrm{adj}$



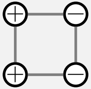
Example: hardcore

- $\Omega = \{0, 1\}^n$
- $\mathcal{J} \leqslant \frac{\lambda}{1+\lambda} \cdot \mathrm{adj}$



Example: Ising

- $\Omega = \{\pm 1\}^n$
- $\mathcal{J}[j \to i] \leqslant |\beta_{ij}|$



- Dobrushin++: if $c\,\mathcal{J} < (1 - \delta)c$

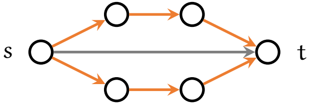
$$t_{\mathrm{mix}}(\epsilon) = O\left(\frac{n}{\delta} \log\left(\frac{n \cdot c_{\mathrm{max}}}{\epsilon \cdot c_{\mathrm{min}}}\right)\right)$$

- Existence: $\lambda_{\mathrm{max}}(\mathcal{J}) < 1$

Comparison

- P, P' reversible with same stationary distribution
- Comparison: route Q' through Q with low congestion and length.

$$\pi(\text{path} \mid X_0 = s, X_\ell = t)$$



Congestion

Suppose π is dist over paths and Q is ergodic flow. Congestion is

$$\max\left\{ \frac{\mathbb{P}_{\text{path}\sim\pi}[(x \to y) \in \text{path}]}{Q(x,y)} \;\middle|\; x \neq y \right\}$$

Lemma: comparison

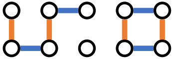
Suppose ρ, ρ' are χ^2 contraction rates:

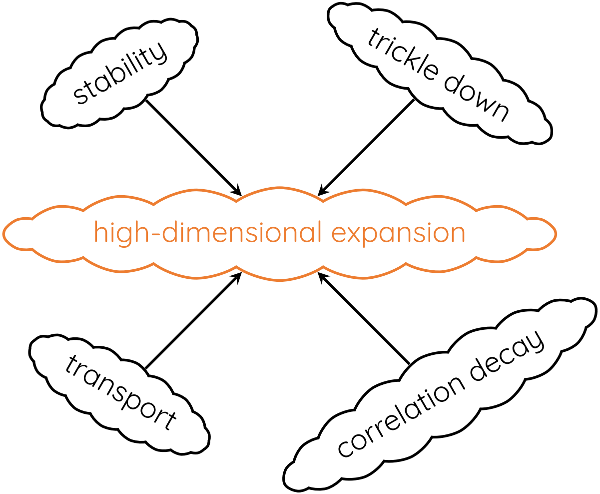
$$\rho \geqslant \frac{\rho'}{(\text{congestion}) \cdot (\text{max length})}$$

- If len $\leqslant 1$, can use any $\mathcal{D}_\phi$.
- Canonical paths: a few-to-one mapping enc from (s,t)-pairs whose path passes $x \to y$ to Ω:

$$\mu(s)\mu(t) \leqslant C \cdot \mu(\mathsf{enc}(s,t))Q(x,y)$$

- If M-to-one, then cong $\leqslant CM$.
- Matching walks mix in $\mathsf{poly}(n)$.





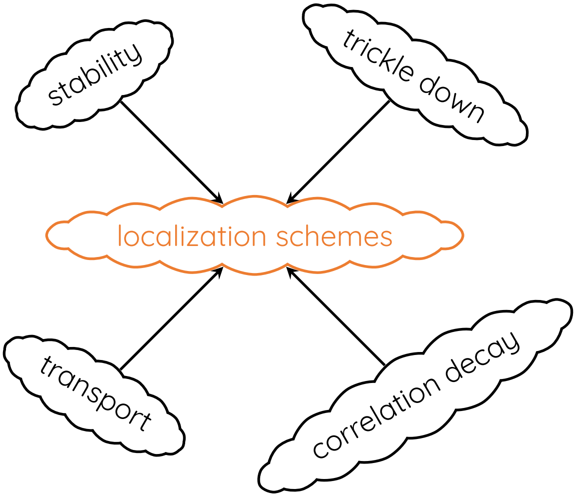
stability

trickle down

high-dimensional expansion

transport

correlation decay



stability

trickle down

localization schemes

transport

correlation decay