Information Geometry of Reversible Markov Chains

International Conference on Information Geometry for Data Science

doi:10.1007/s41884-021-00061-7 *Geoffrey Wolfer*†, *Shun Watanabe*‡

September 19, 2022

† RIKEN AIP – Special Postdoctoral Researcher Program
‡ Tokyo University of Agriculture and Technology

Introduction & Preliminaries

Information geometry

Finite space $\mathcal{X} = [m]$. Simplex over $\mathcal{X}$: $\mathcal{P}(\mathcal{X})$.

Information geometry

Finite space $\mathcal{X} = [m]$. Simplex over $\mathcal{X}$: $\mathcal{P}(\mathcal{X})$.

Regard family $\mathcal{M} = \{\mu_\theta : \theta \in \Theta\} \subseteq \mathcal{P}(\mathcal{X})$ as a smooth manifold.

Fisher information metric g and dual affine connections $\nabla^{(e)}$, $\nabla^{(m)}$

$$\mathrm{g}_{ij}(\mu_\theta) := \sum_{x \in \mathcal{X}} \mu_\theta(x)\, \partial_i \log \mu_\theta(x)\, \partial_j \log \mu_\theta(x),$$

$$\Gamma^{(e)}_{ij,k}(\mu_\theta) := \sum_{x \in \mathcal{X}} \partial_i \partial_j \log \mu_\theta(x)\, \partial_k \mu_\theta(x),$$

$$\Gamma^{(m)}_{ij,k}(\mu_\theta) := \sum_{x \in \mathcal{X}} \partial_i \partial_j \mu_\theta(x)\, \partial_k \log \mu_\theta(x).$$

Finite space $\mathcal{X} = [m]$. Simplex over $\mathcal{X}$: $\mathcal{P}(\mathcal{X})$.

Regard family $\mathcal{M} = \{p_\theta : \theta \in \Theta\} \subset \mathcal{P}(\mathcal{X})$ as a smooth manifold.

Fisher information metric g and dual affine connections $\nabla^{(e)}$, $\nabla^{(m)}$

$$g_{ij}(p_\theta) \triangleq \sum_{x \in \mathcal{X}} p_\theta(x) \partial_i \log p_\theta(x) \partial_j \log p_\theta(x),$$

$$\Gamma^{(e)}_{ij,k}(p_\theta) \triangleq \sum_{x \in \mathcal{X}} \partial_i \partial_j \log p_\theta(x) \partial_k p_\theta(x),$$

$$\Gamma^{(m)}_{ij,k}(p_\theta) \triangleq \sum_{x \in \mathcal{X}} \partial_i \partial_j p_\theta(x) \partial_k \log p_\theta(x).$$

Wide range of applications

1. Higher-order efficiency analysis of estimator
2. Information decomposition / projection
3. Natural gradient algorithms

Information geometry

Finite space $\mathcal{X} = [m]$. Simplex over $\mathcal{X}$: $\mathcal{P}(\mathcal{X})$.

Regard family $\mathcal{M} = \{\mu_\theta : \theta \in \Theta\} \subseteq \mathcal{P}(\mathcal{X})$ as a smooth manifold.

Fisher information metric $\mathfrak{g}$ and dual affine connections $\nabla^{(e)}$, $\nabla^{(m)}$

$$\mathfrak{g}_{ij}(\mu_\theta) \triangleq \sum_{x \in \mathcal{X}} \mu_\theta(x) \partial_i \log \mu_\theta(x) \partial_j \log \mu_\theta(x),$$

$$\Gamma^{(e)}_{ij,k}(\mu_\theta) \triangleq \sum_{x \in \mathcal{X}} \partial_i \partial_j \log \mu_\theta(x) \partial_k \mu_\theta(x),$$

$$\Gamma^{(m)}_{ij,k}(\mu_\theta) \triangleq \sum_{x \in \mathcal{X}} \partial_i \partial_j \mu_\theta(x) \partial_k \log \mu_\theta(x).$$

Wide range of applications

1. Higher-order efficiency analysis of estimator
2. Information decomposition / projection
3. Natural gradient algorithms

What about Markov models?

Irreducible Markov chains

Notation

$E \quad X^2$ such that (X, E) strongly connected.

Functions and positive functions over E: $F(X, E)$, $F_+(X, E)$.

Irreducible row-stochastic matrices over (X, E): $W(X, E)$.

Often we will take $E = X^2$.

Irreducible Markov chains

Notation

$E \subseteq X^2$ such that (X, E) strongly connected.

Functions and positive functions over E: $F(X, E)$, $F_+(X, E)$.

Irreducible row-stochastic matrices over (X, E): $W(X, E)$.

Often we will take $E = X^2$.

Discrete-time, time-homogeneous Markov chain

$$\mathbb{P}(X_1 = x_1, \ldots, X_k = x_k) = \mu(x_1) \prod_{t=1}^{k-1} P(x_t, x_{t+1}),$$

$$(\mu, P) \in (\mathcal{P}(X), W(X, E)).$$

Irreducible Markov chains

Notation

$E \subseteq X^2$ such that (X, E) strongly connected.

Functions and positive functions over E: $\mathcal{F}(X, E)$, $\mathcal{F}_+(X, E)$.

Irreducible row-stochastic matrices over (X, E): $\mathcal{W}(X, E)$.

Often we will take $E = X^2$.

Discrete-time, time-homogeneous Markov chain

$$\mathbb{P}(X_1 = x_1, \dots, X_k = x_k) = \mu(x_1) \prod_{t=1}^{k-1} P(x_t, x_{t+1}),$$

$$(\mu, P) \in (\mathcal{P}(X), \mathcal{W}(X, E)).$$

Stationary distribution: $\pi P = \pi$.

Edge-measure: $Q(x, x') = \pi(x) P(x, x') = \mathbb{P}_\pi(X_t = x, X_{t+1} = x')$.

Reversible Markov chains

Notation

Time-reversal: $P^{?}(x, x') = p(x')P(x', x)/p(x)$.

Reversible: $p(x)P(x, x') = p(x')P(x', x)$ (detailed balance).

$W_{\text{rev}}(X, E)$ the set of all reversible kernels.

Reversible Markov chains

Notation

Time-reversal: $P^{?}(x, x^{0}) = p(x^{0})P(x^{0}, x)/p(x)$.

Reversible: $p(x)P(x, x^{0}) = p(x^{0})P(x^{0}, x)$ (detailed balance).

$W_{\text{rev}}(X, E)$ the set of all reversible kernels.

Numerous interesting properties (e.g.)

Spectrum is real and stable.

Mixing time is governed by spectral gap.

Stationary distribution is easy to compute.

Reversible Markov chains

Notation

Time-reversal: $P^{?}(x, x^{0}) = p(x^{0})P(x^{0}, x)/p(x)$.

Reversible: $p(x)P(x, x^{0}) = p(x^{0})P(x^{0}, x)$ (detailed balance).

$W_{\text{rev}}(X, E)$ the set of all reversible kernels.

Numerous interesting properties (e.g.)

Spectrum is real and stable.

Mixing time is governed by spectral gap.

Stationary distribution is easy to compute.

Historical remarks

Inspiration & origin Schrödinger (1931).

First treatment for MCs Kolmogorov (1936, 1937).

Tribute and historical summary Dobrushin et al. (1988).

Exponential tilting (ET)

ET of distribution
$X \sim m \in \mathcal{P}([m]), f : X \to \mathbb{R}$. Construct exponential family:

$$m_q(x) = m(x)e^{qf(x) - k(q)}, \qquad k(q) = \log \mathrm{E}\, e^{qf(X)} \qquad \text{(CGF)}.$$

Exponential tilting (ET)

ET of distribution
$X \sim \mu \in \mathcal{P}([m])$, $f : \mathcal{X} \to \mathbb{R}$. Construct exponential family:

$$\mu_\theta(x) = \mu(x) e^{\theta f(x) - \kappa(\theta)}, \qquad \kappa(\theta) = \log \mathbb{E} e^{\theta f(X)} \quad \text{(CGF)}$$

$$\lambda > \mathbb{E}[f], \lim_{k \to \infty} \frac{1}{k} \log \mathbb{P}\left(\frac{1}{k} \sum_{t=1}^{k} f(X_t) > \lambda \right) = -\kappa^*(\lambda), \quad \sup_{\theta \in \mathbb{R}} \{\theta\lambda - \kappa(\theta)\}.$$

ET of distribution

$X \sim \mu \in \mathcal{P}([m])$, $f : \mathcal{X} \to \mathbb{R}$. Construct exponential family:

$$\mu_\theta(x) = \mu(x) e^{\theta f(x) - \kappa(\theta)}, \qquad \kappa(\theta) = \log \mathbb{E} e^{\theta f(X)} \qquad \text{(CGF)}.$$

$$\lambda > \mathbb{E}[f], \lim_{k \to \infty} \frac{1}{k} \log \mathbb{P}\left(\frac{1}{k} \sum_{t=1}^{k} f(X_t) > \lambda \right) = -\kappa^{\star}(\lambda), \quad \sup_{\theta \in \mathbb{R}} \{\theta\lambda - \kappa(\theta)\}.$$

ET of stochastic matrix

$P \in \mathcal{W}$, $f : \mathcal{X} \to \mathbb{R}$, $\tilde{P}_\theta(x, x') = P(x, x') \exp(\theta f(x'))$,

Exponential tilting (ET)

ET of distribution
$X \sim \mu \in \mathcal{P}([m])$, $f : \mathcal{X} \to \mathbb{R}$. Construct exponential family:

$$\mu_\theta(x) = \mu(x) e^{\theta f(x) - \kappa(\theta)}, \qquad \kappa(\theta) = \log \mathbb{E} e^{\theta f(X)} \qquad \text{(CGF)},$$

$$l > \mathbb{E}[f], \lim_{k \to \infty} -\frac{1}{k} \log \mathbb{P}\left(\frac{1}{k} \sum_{t=1}^{k} f(X_t) > l \right) = \kappa^\star(l) \triangleq \sup_{\theta \in \mathbb{R}} \{ \theta l - \kappa(\theta) \}.$$

ET of stochastic matrix
$P \in \mathcal{W}$, $f : \mathcal{X} \to \mathbb{R}$, $\tilde{P}_\theta(x, x') = P(x, x') \exp(\theta f(x'))$, $P_\theta = s(\tilde{P}_\theta)$, with Perron-Frobenius (PF) rescaling (Miller, 1961),

$$s : \mathcal{F}_+(\mathcal{X}, \mathcal{E}) \to \mathcal{W}(\mathcal{X}, \mathcal{E}), \ \tilde{P}(x, x') \mapsto P(x, x') = \frac{\tilde{P}(x, x') v(x')}{\rho v(x)},$$

where ρ, v are the PF root and right PF eigenvector of $\tilde{P}$.

Exponential tilting (ET)

ET of distribution

$X \sim \mu \in \mathcal{P}([m])$, $f : \mathcal{X} \to \mathbb{R}$. Construct exponential family:

$$\mu_\theta(x) = \mu(x) e^{\theta f(x) - \kappa(\theta)}, \qquad \kappa(\theta) = \log \mathbb{E} e^{\theta f(X)} \qquad \text{(CGF)},$$

$$\lambda > \mathbb{E}[f], \lim_{k \to \infty} -\frac{1}{k} \log \mathbb{P}\left(\frac{1}{k} \sum_{t=1}^{k} f(X_t) > \lambda \right) = \kappa^\star(\lambda) \triangleq \sup_{\theta \in \mathbb{R}} \{\theta \lambda - \kappa(\theta)\}.$$

ET of stochastic matrix

$P \in \mathcal{W}$, $f : \mathcal{X} \to \mathbb{R}$, $\tilde{P}_\theta(x, x') = P(x, x') \exp(\theta f(x'))$, $P_\theta = s(\tilde{P}_\theta)$, with Perron-Frobenius (PF) rescaling (Miller, 1961),

$$s : \mathcal{F}_+(\mathcal{X}, \mathcal{E}) \to \mathcal{W}(\mathcal{X}, \mathcal{E}), \ \tilde{P}(x, x') \mapsto P(x, x') = \frac{\tilde{P}(x, x') v(x')}{\rho v(x)},$$

where ρ, v are the PF root and right PF eigenvector of $\tilde{P}$.

$$\lambda > \mathbb{E}[f], \lim_{k \to \infty} -\frac{1}{k} \log \mathbb{P}\left(\frac{1}{k} \sum_{t=1}^{k} f(X_t) > \lambda \right) = \sup_{\theta \in \mathbb{R}} \{\theta \lambda - \log \rho(\theta)\}.$$

Information Geometry of Markov Chains

Information Geometry of Markov Chains

1. Large deviations: Miller (1961); Donsker and Varadhan (1975); Gärtner (1977)
2. Information projection: Csiszár et al. (1987)
3. Asymptotic e-families: Ito and Amari (1988); Takeuchi and Barron (1998); Takeuchi and Kawabata (2007); Takeuchi and Nagaoka (2017)
4. One-parameters exponential families: Nakagawa and Kanaya (1993)
5. Dually flat structure: Nagaoka (2005)

Information Geometry of Markov Chains (Nagaoka, 2005)

Distributions

$P(X)$

D

KL divergence

$$D(m_\theta \| m_{\theta'}) = E_{m_\theta} \log \frac{m_\theta(x)}{m_\theta(x')}$$

Information Geometry of Markov Chains (Nagaoka, 2005)

Distributions $\mathcal{P}(\mathcal{X})$

Markov chains $\mathcal{W}(\mathcal{X}, \mathcal{E})$

$$X_1, X_2, \dots, X_t \quad P$$

$$D \longrightarrow D$$

KL divergence

KL divergence rate

$$\lim_{t \to \infty} \frac{1}{t} D\left(X_1, \dots, X_t \quad P_\theta \| X'_1, \dots, X'_t \quad P_{\theta'}\right) = \mathbb{E}_{(X,X') \quad Q_\theta} \log \frac{P_\theta(X,X')}{P_{\theta'}(X,X')}$$

$$\triangleq D\left(P_\theta \| P_{\theta'}\right)$$

Information Geometry of Markov Chains (Nagaoka, 2005)

Distributions	Markov chains
$P(X)$	$W(X, E)$

limit ¥

$r^{(e)}, r^{(m)}$ ⟶ $r^{(e)}, r^{(m)}$

g ⟶ g

D ⟶ D

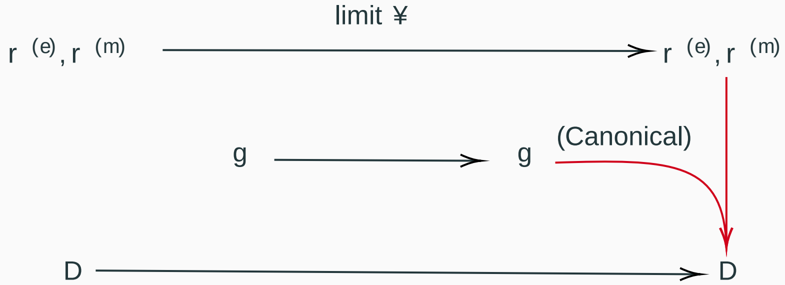
Information Geometry of Markov Chains (Nagaoka, 2005)

Distributions

$P(X)$

Markov chains

$W(X, E)$



View $\mathcal{W}(\mathcal{X}, \mathcal{E})$ as a smooth manifold.

Fisher information metric $\mathfrak{g}$

$$g_{ij}(P_\theta) \triangleq \sum_{(x,x')\in\mathcal{E}} Q_\theta(x,x')\,\partial_i \log P_\theta(x,x')\,\partial_j \log P_\theta(x,x').$$

Dual affine connections $\nabla^{(e)}, \nabla^{(m)}$

$$\Gamma^{(e)}_{ij,k}(P_\theta) \triangleq \sum_{(x,x')\in\mathcal{E}} \partial_i\partial_j \log P_\theta(x,x')\,\partial_k Q_\theta(x,x'),$$

$$\Gamma^{(m)}_{ij,k}(P_\theta) \triangleq \sum_{(x,x')\in\mathcal{E}} \partial_i\partial_j Q_\theta(x,x')\,\partial_k \log P_\theta(x,x').$$

Exponential families of transition kernels (Nagaoka, 2005)

Let $\Theta \subseteq \mathbb{R}^d$, open connected parameter space.

$$\mathcal{V}_e = \left\{ P_\theta : \theta = (\theta^1, \ldots, \theta^d) \in \Theta \right\} \subset \mathcal{W}$$

is an e-family with natural parameter θ, whenever there exist functions $K, R_\theta, \psi_\theta, g_1, \ldots, g_d$ such that

$$\log P_\theta(x, x') = K(x, x') + \sum_{i=1}^{d} \theta^i g_i(x, x') + \underbrace{R_\theta(x') - R_\theta(x) - \psi_\theta}_{\text{rescaling terms}} .$$

Let $\Theta \subset \mathbb{R}^d$, open connected parameter space.

$$\mathcal{V}_e = \left\{ P_\theta : \theta = (\theta^1, \dots, \theta^d) \in \Theta \right\} \subset \mathcal{W}$$

is an e-family with natural parameter θ, whenever there exist functions $K, R_\theta, \psi_\theta, g_1, \dots, g_d$ such that

$$\log P_\theta(x, x') = K(x, x') + \sum_{i=1}^{d} \theta^i g_i(x, x') + \underbrace{R_\theta(x') - R_\theta(x) - \psi_\theta}_{\text{rescaling terms}}.$$

Example 1 (Nagaoka, 2005)

$\mathcal{W}(\mathcal{X}, \mathcal{E})$ forms an e-family of dimension $|\mathcal{E}| - |\mathcal{X}|$. For $\mathcal{E} = \mathcal{X}^2$,

$$\log P(x, x') = \sum_{i=1}^{|\mathcal{X}|} \sum_{\substack{j=1 \\ j \neq x_\star}}^{|\mathcal{X}|} \overbrace{\log \frac{P(i,j)P(j,x_\star)}{P(i,x_\star)P(x_\star,x_\star)}}^{\theta^{ij}} \overbrace{\delta_i(x)\delta_j(x')}^{g_{ij}(x,x')}$$

$$+ \log P(x, x_\star) - \log P(x', x_\star) + \log P(x_\star, x_\star).$$

Mixture families (Nagaoka, 2005)

We say that $\mathcal{V}_m$ is a mixture family when there exists functions $C, F_1, \ldots, F_d$, such that $C, C + F_1, \ldots, C + F_d$ are affinely independent,

$$\sum_{x,x' \in \mathcal{X}} C(x, x') = 1, \qquad \sum_{x,x' \in \mathcal{X}} F_i(x, x') = 0, \forall i \in [d],$$

and

$$\mathcal{V}_m = \left\{ P_\xi \in \Omega : Q_\xi = C + \sum_{i=1}^{d} \xi^i F_i, \xi \in \Xi \right\}$$

where $\Xi = \left\{ \xi \in \mathbb{R}^d : Q_\xi(x, x') > 0, \forall (x, x') \in \mathcal{X}^2 \right\}$, and Q_ξ is the edge measure that pertains to P_ξ.

Mixture families (Nagaoka, 2005)

We say that V_m is a mixture family when there exists functions $C, F_1, \ldots, F_d$, such that $C, C + F_1, \ldots, C + F_d$ are affinely independent,

$$\sum_{x,x' \in X} C(x, x') = 1, \qquad \sum_{x,x' \in X} F_i(x, x') = 0, \forall i \in [d],$$

and

$$V_m = \left\{ P_x \in \Omega : Q_x = C + \sum_{i=1}^{d} x^i F_i, x \in X \right\}$$

where $X = \left\{ x \in \mathbb{R}^d : Q_x(x, x') > 0, \forall (x, x') \in X^2 \right\}$, and Q_x is the edge measure that pertains to P_x.

Example 2
W_{bis} (doubly stochastic) forms an m-family (Hayashi and Watanabe, 2016).

Geometric approach has recently lead to finite sample analysis for:

1. Parameter estimation problem in Markov chains in HMMs Hayashi and Watanabe (2016); Hayashi (2022).

Applications { Inference in Markov chains

Geometric approach has recently lead to finite sample analysis for:

1. Parameter estimation problem in Markov chains in HMMs
   Hayashi and Watanabe (2016); Hayashi (2022).
2. Hypothesis testing problem
   Watanabe and Hayashi (2017).

Geometric approach has recently lead to finite sample analysis for:

1. Parameter estimation problem in Markov chains in HMMs
   Hayashi and Watanabe (2016); Hayashi (2022).
2. Hypothesis testing problem
   Watanabe and Hayashi (2017).
3. Local equivalence problem in hidden Markov models
   Hayashi (2019).

Applications { Inference in Markov chains

Geometric approach has recently lead to finite sample analysis for:

1. Parameter estimation problem in Markov chains in HMMs
   Hayashi and Watanabe (2016); Hayashi (2022).
2. Hypothesis testing problem
   Watanabe and Hayashi (2017).
3. Local equivalence problem in hidden Markov models
   Hayashi (2019).
4. Cherno€ and Hoe€ding bounds with improved pre-factor
   Moulos and Anantharam (2019)

Applications { Inference in Markov chains

Geometric approach has recently lead to finite sample analysis for:

1. Parameter estimation problem in Markov chains in HMMs
   Hayashi and Watanabe (2016); Hayashi (2022).
2. Hypothesis testing problem
   Watanabe and Hayashi (2017).
3. Local equivalence problem in hidden Markov models
   Hayashi (2019).
4. Cherno€ and Hoe€ding bounds with improved pre-factor
   Moulos and Anantharam (2019)

Extension of known conceptsˆ but also new structures emerge!

Time-reversal families

Time-reversal families

Definition
Let V W .

$$V^? , \quad f P^? : P 2 V g.$$

Definition
Let V W .

$$V^{?} , \quad f \, P^{?} : P \; 2 \; V g .$$

Proposition 1

$$V_m \text{ is an m-family } =) \; V_m^{?} \text{ is an m-family.}$$

Definition

Let V W .

$$V^{?} , \quad f P^{?} : P \; 2 \; V g .$$

Proposition 1

V_m is an m-family =) $V_m^{?}$ is an m-family.

V_e is an e-family =) $V_e^{?}$ is an e-family.

Definition
Let V W .

$$V^{?} , \quad \{ P^{?} : P \in V \} .$$

Proposition 1

$$V_m \text{ is an m-family} \implies V_m^{?} \text{ is an m-family.}$$

$$V_e \text{ is an e-family} \implies V_e^{?} \text{ is an e-family.}$$

What about families of fixed points?

Reversible e-families

Reversible e-families

Definition

V is reversible when P is reversible $\forall P \in V$.

Reversible e-families

Definition
$\mathcal{V}$ is reversible when P is reversible $\forall P \in \mathcal{V}$.

Example 3
Unbiased lazy random walks on the d-cycle form a reversible e-family.

Reversible e-families

Definition
V is reversible when P is reversible $\forall P \in V$.

Example 3
Unbiased lazy random walks on the d-cycle form a reversible e-family.

Problem: How to determine whether a given e-family is reversible ?

Definition

$\mathcal{V}$ is reversible when P is reversible $\forall P \in \mathcal{V}$.

Example 3

Unbiased lazy random walks on the d-cycle form a reversible e-family.

Problem: How to determine whether a given e-family is reversible ?

Theorem (Kolmogorov's criterion (Kolmogorov, 1936))

Let P irreducible, and Γ the set of finite directed closed-paths, i.e. $\gamma \in \Gamma$, when $\gamma : [n] \to \mathcal{E}$, $\gamma(t) = (x_t, x_{t+1})$ with $x_{n+1} = x_1$. Write $\gamma^\star$ the time-reversed path. P is reversible if and only if for all $\gamma \in \Gamma$,

$$\prod_{t=1}^{|\gamma|} P(\gamma(t)) = \prod_{t=1}^{|\gamma|} P(\gamma^\star(t)).$$

Definition
$\mathcal{V}$ is reversible when P is reversible $\forall P \in \mathcal{V}$.

Example 3
Unbiased lazy random walks on the d-cycle form a reversible e-family.

Problem: How to determine whether a given e-family is reversible ?

Theorem (Kolmogorov's criterion (Kolmogorov, 1936))
Let P irreducible, and Γ the set of finite directed closed-paths, i.e. $\gamma \in \Gamma$, when $\gamma : [n] \to E$, $\gamma(t) = (x_t, x_{t+1})$ with $x_{n+1} = x_1$. Write $\gamma^\star$ the time-reversed path. P is reversible if and only if for all $\gamma \in \Gamma$,

$$\prod_{t=1}^{|\gamma|} P(\gamma(t)) = \prod_{t=1}^{|\gamma|} P(\gamma^\star(t)).$$

Log-reversible function
A function $h \in \mathcal{F}$ is log-reversible whenever it satisfies that for all $\gamma \in \Gamma$,

$$\sum_{t=1}^{|\gamma|} h(\gamma(t)) = \sum_{t=1}^{|\gamma|} h(\gamma^\star(t)).$$

Characterization of reversible e-families

Theorem 2

Let V be e-family of kernels generated by K and $(g_i)_{i \in [d]}$. The two statements are equivalent:

Characterization of reversible e-families

Theorem 2

Let $\mathcal{V}$ be e-family of kernels generated by K and $(g_i)_{i \in [d]}$. The two statements are equivalent:

(i) $\mathcal{V}$ is reversible.

Characterization of reversible e-families

Theorem 2

Let $\mathcal{V}$ be e-family of kernels generated by K and $(g_i)_{i \in [d]}$. The two statements are equivalent:

(i) $\mathcal{V}$ is reversible.

(ii) $\mathcal{V}$ is such that K and $g_i, \forall i \in [d]$ are all log-reversible functions (and $\mathcal{E} = \mathcal{E}^{\star}$ for more general irreducible kernels).

Characterization of reversible e-families

Theorem 2

Let $\mathcal{V}$ be e-family of kernels generated by K and $(g_i)_{i \in [d]}$. The two statements are equivalent:

(i) $\mathcal{V}$ is reversible.

(ii) $\mathcal{V}$ is such that K and $g_i, \forall i \in [d]$ are all log-reversible functions (and $\mathcal{E} = \mathcal{E}^{\star}$ for more general irreducible kernels).

Naively, the number of distinct equations that must be checked would be large

$$(d+1) \sum_{k=3}^{|\mathcal{X}|} \binom{|\mathcal{X}|}{k} \frac{(k-1)!}{2}.$$

Theorem 2

Let V be e-family of kernels generated by K and $(g_i)_{i \in [d]}$. The two statements are equivalent:

(i) V is reversible.

(ii) V is such that K and $g_i, \forall i \in [d]$ are all log-reversible functions (and $E = E^{\star}$ for more general irreducible kernels).

Naively, the number of distinct equations that must be checked would be large

$$(d+1) \sum_{k=3}^{|X|} \binom{|X|}{k} \frac{(k-1)!}{2}.$$

Lemma 2 (Polynomial time procedure for reversibility checking)

Let $h \in F_+$ (more generally irreducible). $\log[h]$ is log-reversible if and only if $P_h \circ h^\top$ is a symmetric matrix, with $P_h = v_h^\top u_h$ the PF projection of h, where u_h and v_h are respectively the left and right PF eigenvectors of h, normalized such that $u_h v_h^\top = 1$.

Lemma 4 (Vector space of log-reversible functions)

$$\mathcal{F}_{\text{rev}} \triangleq \{ h \in \mathcal{F} : h \text{ is log-reversible} \},$$
$$\mathcal{N} \triangleq \{ h \in \mathcal{F} : \exists (c, f),\, h(x, x') = f(x') - f(x) + c \}.$$

Affine structures

Lemma 4 (Vector space of log-reversible functions)

$$\mathcal{F}_{\text{rev}} \triangleq \{h \in \mathcal{F} : h \text{ is log-reversible}\},$$
$$\mathcal{N} \triangleq \{h \in \mathcal{F} : \exists (c, f),\ h(x, x') = f(x') - f(x) + c\}.$$

$$\mathcal{N} \subset \mathcal{F}_{\text{rev}} \subset \mathcal{F} \quad \text{(vector space inclusions).}$$

Lemma 4 (Vector space of log-reversible functions)

$$\mathcal{F}_{rev} := \{h \in \mathcal{F} : h \text{ is log-reversible}\},$$
$$\mathcal{N} := \{h \in \mathcal{F} : \exists (c, f),\ h(x, x') = f(x') - f(x) + c\}.$$

$$\mathcal{N} \subset \mathcal{F}_{rev} \subset \mathcal{F} \quad \text{(vector space inclusions)}.$$

Quotient space: $\mathcal{G}_{rev} := \mathcal{F}_{rev} / \mathcal{N}$

$\mathcal{D}(\mathcal{G}_{rev}) = \mathcal{W}_{rev}$, with $\mathcal{D}(f) := s(\exp[f])$ diffeomorphism.

Lemma 4 (Vector space of log-reversible functions)

$$\mathcal{F}_{rev} := \{h \in \mathcal{F} : h \text{ is log-reversible}\},$$
$$\mathcal{N} := \{h \in \mathcal{F} : \exists (c, f), h(x, x') = f(x') - f(x) + c\}.$$

$$\mathcal{N} \subset \mathcal{F}_{rev} \subset \mathcal{F} \quad \text{(vector space inclusions).}$$

Quotient space: $\mathcal{G}_{rev} := \mathcal{F}_{rev} / \mathcal{N}$

$D(\mathcal{G}_{rev}) = \mathcal{W}_{rev}$, with $D(f) := s(\exp[f])$ diffeomorphism.

Theorem 3 & 4

Let $T = \{x \to x'\} \setminus \{(m, 1)\}$, $|T| = |X|(|X| + 1)/2 - 1$.

Lemma 4 (Vector space of log-reversible functions)

$$\mathcal{F}_{rev} \coloneqq \{h \in \mathcal{F} : h \text{ is log-reversible}\},$$
$$\mathcal{N} \coloneqq \{h \in \mathcal{F} : \exists (c, f),\ h(x, x') = f(x') - f(x) + c\}.$$

$$\mathcal{N} \subset \mathcal{F}_{rev} \subset \mathcal{F} \quad \text{(vector space inclusions)}.$$

Quotient space: $\mathcal{G}_{rev} \coloneqq \mathcal{F}_{rev} / \mathcal{N}$

$$D(\mathcal{G}_{rev}) = \mathcal{W}_{rev}, \text{ with } D(f) \coloneqq s(\exp[f]) \text{ diffeomorphism}.$$

Theorem 3 & 4
Let $\mathcal{T} = \{x \leq x'\} \setminus \{(m, 1)\}$, $|\mathcal{T}| = |\mathcal{X}|(|\mathcal{X}| + 1)/2 - 1$.

(i) $\mathcal{G}_{rev}$ is a $|\mathcal{T}|$-dimensional vector space.

Affine structures

Lemma 4 (Vector space of log-reversible functions)

$$\mathcal{F}_{\text{rev}} \triangleq \{h \in \mathcal{F} : h \text{ is log-reversible}\},$$
$$\mathcal{N} \triangleq \{h \in \mathcal{F} : \exists (c, f), h(x, x') = f(x') - f(x) + c\}.$$

$$\mathcal{N} \subset \mathcal{F}_{\text{rev}} \subset \mathcal{F} \quad \text{(vector space inclusions)}.$$

Quotient space: $\mathcal{G}_{\text{rev}} \triangleq \mathcal{F}_{\text{rev}} / \mathcal{N}$

$$\Delta(\mathcal{G}_{\text{rev}}) = \mathcal{W}_{\text{rev}}, \text{ with } \Delta(f) \triangleq \mathfrak{s}(\exp[f]) \text{ diffeomorphism}.$$

Theorem 3 & 4

Let $\mathcal{T} = \{x \to x'\} \setminus \{(m, 1)\}$, $|\mathcal{T}| = |\mathcal{X}|(|\mathcal{X}| + 1)/2 - 1$.

(i) $\mathcal{G}_{\text{rev}}$ is a $|\mathcal{T}|$-dimensional vector space.

(ii) $\mathcal{W}_{\text{rev}}$ is an e-family, and is an m-family.

Affine structures

Lemma 4 (Vector space of log-reversible functions)

$$\mathcal{F}_{\mathrm{rev}} \triangleq \{h \in \mathcal{F} : h \text{ is log-reversible}\},$$
$$\mathcal{N} \triangleq \{h \in \mathcal{F} : \exists (c, f), h(x, x') = f(x') - f(x) + c\}.$$

$$\mathcal{N} \subset \mathcal{F}_{\mathrm{rev}} \subset \mathcal{F} \quad \text{(vector space inclusions).}$$

Quotient space: $\mathcal{G}_{\mathrm{rev}} \triangleq \mathcal{F}_{\mathrm{rev}} / \mathcal{N}$

$$\mathfrak{D}(\mathcal{G}_{\mathrm{rev}}) = \mathcal{W}_{\mathrm{rev}}, \text{ with } \mathfrak{D}(f) \triangleq \mathfrak{s}(\exp[f]) \text{ diffeomorphism.}$$

Theorem 3 & 4

Let $\mathcal{T} = \{x \to x'\} \setminus \{(m, 1)\}$, $|\mathcal{T}| = |\mathcal{X}|(|\mathcal{X}| + 1)/2 - 1$.

(i) $\mathcal{G}_{\mathrm{rev}}$ is a $|\mathcal{T}|$-dimensional vector space.

(ii) $\mathcal{W}_{\mathrm{rev}}$ is an e-family, and is an m-family.

(iii) $\dim \mathcal{W}_{\mathrm{rev}} = |\mathcal{T}|$.

Affine structures

Lemma 4 (Vector space of log-reversible functions)

$$F_{rev} \triangleq \{ h \in F : h \text{ is log-reversible} \},$$
$$N \triangleq \{ h \in F : \exists (c, f), h(x, x') = f(x') - f(x) + c \}.$$

$$N \subset F_{rev} \subset F \quad \text{(vector space inclusions)}.$$

Quotient space: $G_{rev} \triangleq F_{rev} / N$

$D(G_{rev}) = W_{rev}$, with $D(f) \triangleq s(\exp[f])$ diffeomorphism.

Theorem 3 & 4

Let $T = \{x \leq x'\} \setminus \{(m, 1)\}$, $|T| = |X|(|X| + 1)/2 - 1$.

(i) G_{rev} is a $|T|$-dimensional vector space.

(ii) W_{rev} is an e-family, and is an m-family.

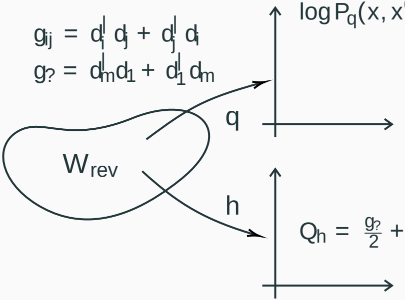
(iii) $\dim W_{rev} = |T|$.

(iv) Basis: $\left\{ g_{ij} \triangleq d_i^\top d_j + d_j^\top d_i : (i, j) \in T \right\}$.

Parametrization of $\mathcal{W}_{rev}$ (Theorem 5)

$$g_{ij} = \delta_i^{\intercal}\delta_j + \delta_j^{\intercal}\delta_i$$

$$g_{?} = \delta_m^{\intercal}\delta_1 + \delta_1^{\intercal}\delta_m$$



$$\log P_\theta(x,x') = \sum_{(i,j)\in T} \overbrace{\frac{1}{2(1+\delta_i(j))}\log\frac{P(i,j)P(j,i)}{P(m,1)P(1,m)}}^{=\theta^{ij}} g_{ij}(x,x') + \frac{1}{2}\log p(x') \quad \frac{1}{2}\log p(x) + \frac{1}{2}\log P(m,1)P(1,m)$$

$$Q_\eta = \frac{g_{?}}{2} + \sum_{(i,j)\in T}(g_{ij} \quad g_{?})\frac{Q(i,j)}{1+\delta_i(j)}$$

$$\theta^i(\eta) = \sum_{x,x'\in E} \frac{Q_\eta(x,x')}{\partial\eta_i} \log P_\eta(x,x')$$

$$\eta_i(\theta) = \sum_{x,x'\in E} Q_\theta(x,x') g_i(x,x')$$

Reversible e/m-projections

Closest reversible chain

Reversiblization
Given an irreducible (non-reversible) P, we can construct

1. $(P + P^{?})/2$, the additive reversiblization of P.
2. $PP^{?}$, the multiplicative reversiblization of P (Fill, 1991).
3. $S_p(P)$, the reversible dilation of P (Wolfer and Kontorovich, 2022).

Closest reversible chain

Reversiblization
Given an irreducible (non-reversible) P, we can construct

1. $(P + P^?)/2$, the additive reversiblization of P.
2. $PP^?$, the multiplicative reversiblization of P (Fill, 1991).
3. $S_p(P)$, the reversible dilation of P (Wolfer and Kontorovich, 2022).

Problem
Given an irreducible P, what is the reversible kernel that is the closest to P?

Closest reversible chain

Reversiblization
Given an irreducible (non-reversible) P, we can construct

1. $(P + P^{?})/2$, the additive reversiblization of P.
2. $PP^{?}$, the multiplicative reversiblization of P (Fill, 1991).
3. $S_p(P)$, the reversible dilation of P (Wolfer and Kontorovich, 2022).

Problem
Given an irreducible P, what is the reversible kernel that is the closest to P?

Motivation
Properties of reversible chains are powerful but bri,le

(e.g. real spectrum, Weyl's inequality).

Closest reversible chain

Reversiblization
Given an irreducible (non-reversible) P, we can construct

1. $(P + P^{?})/2$, the additive reversiblization of P.
2. $PP^{?}$, the multiplicative reversiblization of P (Fill, 1991).
3. $S_p(P)$, the reversible dilation of P (Wolfer and Kontorovich, 2022).

Problem
Given an irreducible P, what is the reversible kernel that is the closest to P?

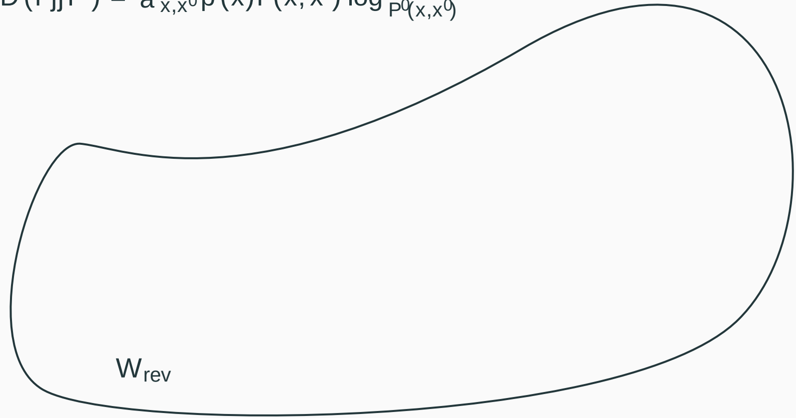
Motivation
Properties of reversible chains are powerful but bri,le

(e.g. real spectrum, Weyl's inequality).

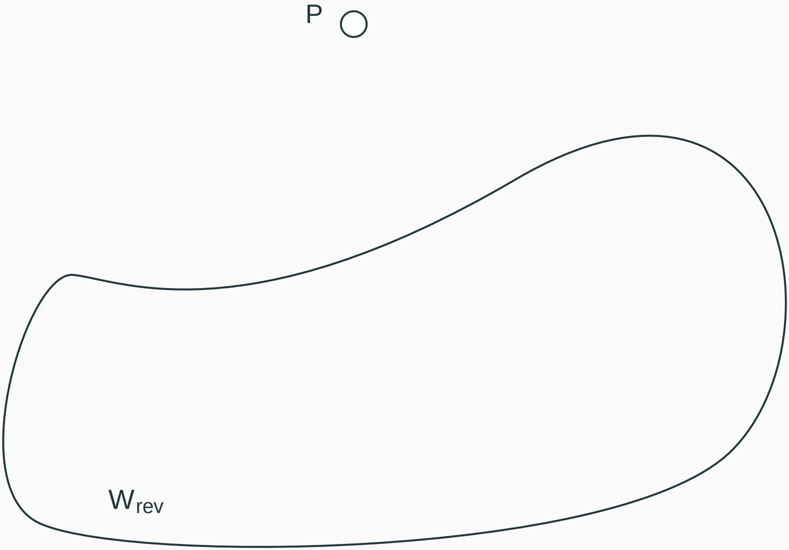
Related work
Analyzed by Nielsen and Weber (2015) for matrix norms induced by scalar products (unique, obtained by solving a convex minimization problem).

Reversible e-projection / m-projection (Theorem 7)

$$D(P \| P') = \sum_{x,x'} \pi(x) P(x,x') \log \frac{P(x,x')}{P'(x,x')}$$



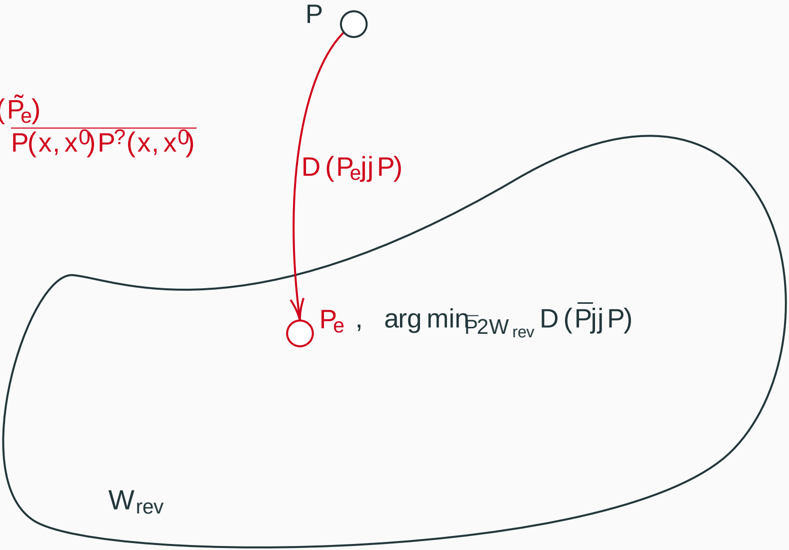
Reversible e-projection / m-projection (Theorem 7)



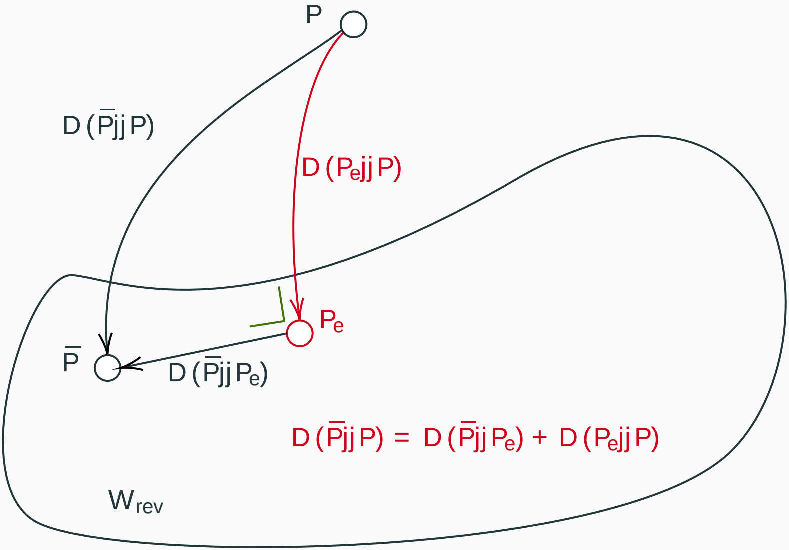
Reversible e-projection / m-projection (Theorem 7)

$$P_e = s(\tilde{P}_e)$$

$$\tilde{P}_e = \sqrt{P(x,x')P^{\star}(x,x')}$$

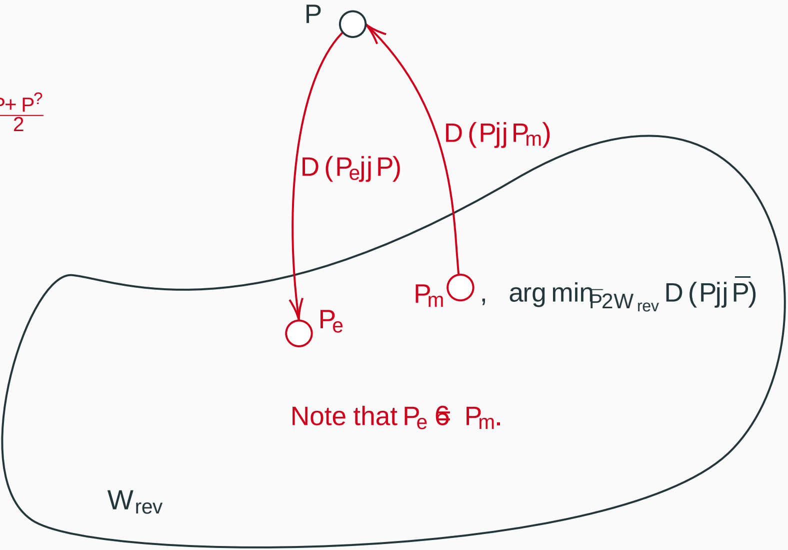


Reversible e-projection / m-projection (Theorem 7)

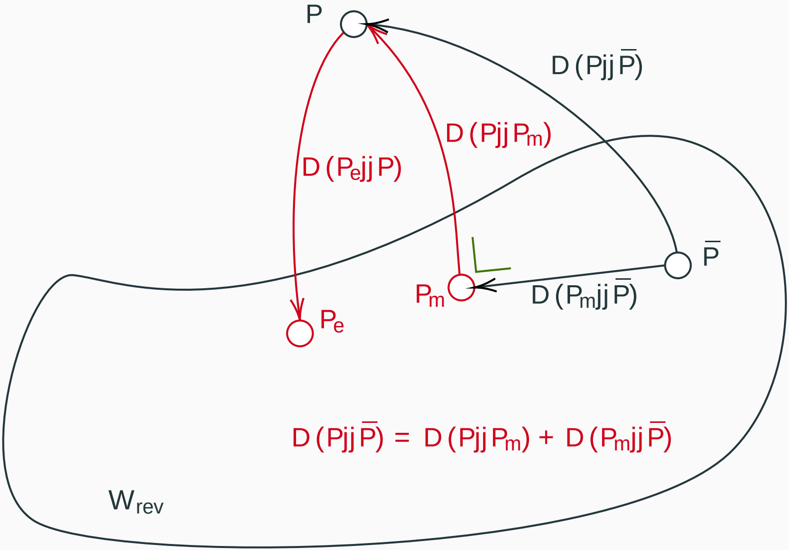


Reversible e-projection / m-projection (Theorem 7)

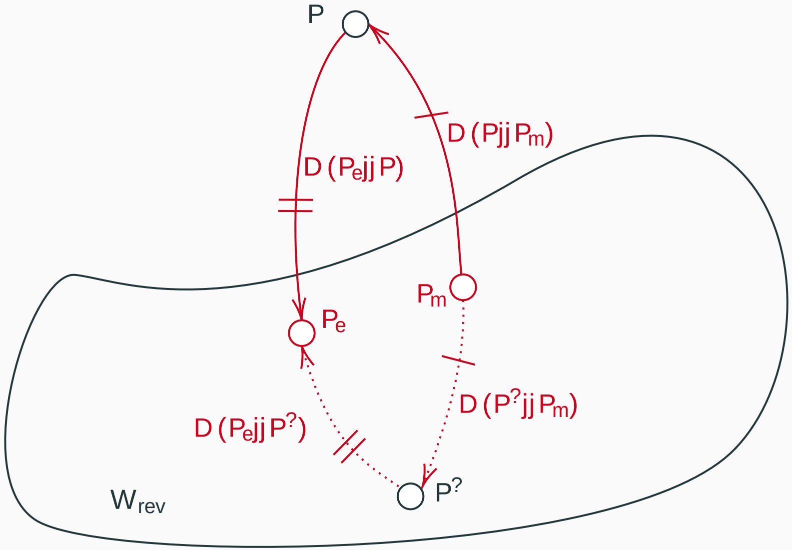
$$P_m = \frac{P + P^{\star}}{2}$$



Reversible e-projection / m-projection (Theorem 7)



Reversible e-projection / m-projection (Theorem 7)



Usual families of kernels

Usual families of kernels

$\mathcal{W}_{bis}$: bi-stochastic kernels ($P^\top \in \mathcal{W}$)
$\mathcal{W}_{sym}$: symmetric kernels ($P = P^\top$)
$\mathcal{W}_{iid}$: memoryless kernels ($P = \mathbf{1}^\top p$)

Usual families of kernels

- $\mathcal{W}_{bis}$: bi-stochastic kernels ($P^\top \in \mathcal{W}$)
- $\mathcal{W}_{sym}$: symmetric kernels ($P = P^\top$)
- $\mathcal{W}_{iid}$: memoryless kernels ($P = 1^\top p$)

Mfd.	m-family	e-family	Dimension
$\mathcal{W}$	✓	✓	$\lvert\mathcal{X}\rvert(\lvert\mathcal{X}\rvert - 1)$
$\mathcal{W}_{rev}$	✓	✓	$\lvert\mathcal{X}\rvert(\lvert\mathcal{X}\rvert + 1)/2 - 1$
$\mathcal{W}_{bis}$	✓	✗	$(\lvert\mathcal{X}\rvert - 1)^2$
$\mathcal{W}_{sym}$	✓	✗	$\lvert\mathcal{X}\rvert(\lvert\mathcal{X}\rvert - 1)/2$
$\mathcal{W}_{iid}$	✗	✓	$\lvert\mathcal{X}\rvert - 1$

Usual families of kernels

$\mathcal{W}_{bis}$: bi-stochastic kernels ($P^\intercal \in \mathcal{W}$)
$\mathcal{W}_{sym}$: symmetric kernels ($P = P^\intercal$)
$\mathcal{W}_{iid}$: memoryless kernels ($P = 1^\intercal p$)

Mfd.	m-family	e-family	Dimension
$\mathcal{W}$	✓	✓	$\lvert\mathcal{X}\rvert(\lvert\mathcal{X}\rvert - 1)$
$\mathcal{W}_{rev}$	✓	✓	$\lvert\mathcal{X}\rvert(\lvert\mathcal{X}\rvert + 1)/2 - 1$
$\mathcal{W}_{bis}$	✓	✗	$(\lvert\mathcal{X}\rvert - 1)^2$
$\mathcal{W}_{sym}$	✓	✗	$\lvert\mathcal{X}\rvert(\lvert\mathcal{X}\rvert - 1)/2$
$\mathcal{W}_{iid}$	✗	✓	$\lvert\mathcal{X}\rvert - 1$

Hierarchies

$$\mathcal{W}_{iid} \overset{\text{e-family}}{\subsetneq} \mathcal{W}_{rev} \overset{\text{e-family}}{\subsetneq} \mathcal{W},$$

$$\mathcal{W}_{sym} \overset{\text{m-family}}{\subsetneq} \mathcal{W}_{rev} \overset{\text{m-family}}{\subsetneq} \mathcal{W},$$

$$\mathcal{W}_{sym} \overset{\text{m-family}}{\subsetneq} \mathcal{W}_{bis} \overset{\text{m-family}}{\subsetneq} \mathcal{W}.$$

Generation of reversible kernels

Generation of the reversible family from W_{iid}

Definition 4 (Mixture hull)
Let $V \quad W$.

$$\text{m-hull}(V) = \left(P : Q \in Q, Q = \sum_{i=1}^{k} a_i Q_i, k \in N, a_1, \ldots, a_k \in R, P_1, \ldots, P_k \in V \right),$$

where Q (resp. Q_i) pertains to P (resp. P_i).

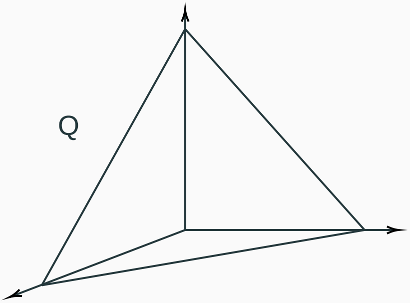
Generation of the reversible family from W_{iid}

Definition 4 (Mixture hull)

Let $V \subseteq W$.

$$\text{m-hull}(V) = \left\{ P : Q \in \mathcal{Q},\ Q = \sum_{i=1}^{k} a_i Q_i,\ k \in \mathbb{N},\ a_1, \ldots, a_k \in \mathbb{R},\ P_1, \ldots, P_k \in V \right\},$$

where Q (resp. Q_i) pertains to P (resp. P_i).

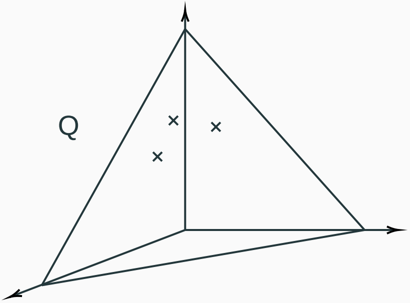


Definition 4 (Mixture hull)

Let $V \quad W$.

$$\text{m-hull}(V) = \left\{ P : Q \in Q, Q = \sum_{i=1}^{k} a_i Q_i, k \in N, a_1, \dots, a_k \in R, P_1, \dots, P_k \in V \right\},$$

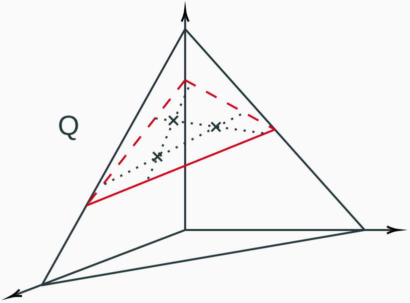
where Q (resp. Q_i) pertains to P (resp. P_i).



Definition 4 (Mixture hull)
Let V W .

$$\text{m-hull}(V) = \left(P : Q \in Q , Q = \sum_{i=1}^{k} a_i Q_i , k \in N , a_1, \ldots, a_k \in R , P_1, \ldots, P_k \in V \right),$$

where Q (resp. Q_i) pertains to P (resp. P_i).

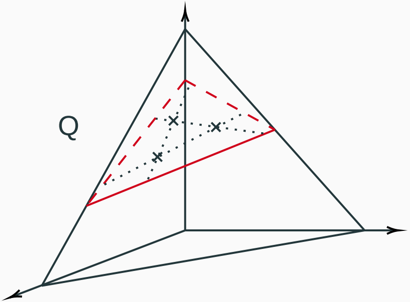


Definition 4 (Mixture hull)

Let V W .

$$\text{m-hull}(V) = \left\{ P : Q \in \mathcal{Q}, Q = \sum_{i=1}^{k} a_i Q_i, k \in \mathbb{N}, a_1, \dots, a_k \in \mathbb{R}, P_1, \dots, P_k \in V \right\},$$

where Q (resp. Q_i) pertains to P (resp. P_i).



Theorem 10

$$\text{m-hull}(W_{iid}) = W_{rev}.$$

Definition 5 (Exponential hull)

Let $V \subseteq \Omega$.

$$\text{e-hull}(V) = \left\{ \sigma(\tilde{P}) : \log[\tilde{P}] = \sum_{i=1}^{k} \alpha_i \log[P_i], \right.$$

$$\left. k \in \mathbb{N}, \alpha_1, \ldots, \alpha_k \in \mathbb{R}, \sum_{i=1}^{k} \alpha_i = 1, P_1, \ldots P_k \in V \right\}.$$

Generation of the reversible family from W_{sym}

Definition 5 (Exponential hull)
Let $V \subseteq W$.

$$\text{e-hull}(V) = \Big\{ s(P) : \log[P] = \sum_{i=1}^{k} a_i \log[P_i], \; k \in \mathbb{N}, a_1, \ldots, a_k \in \mathbb{R}, \sum_{i=1}^{k} a_i = 1, P_1, \ldots P_k \in V \Big\}.$$

Definition 5 (Exponential hull)
Let $V \subset W$.

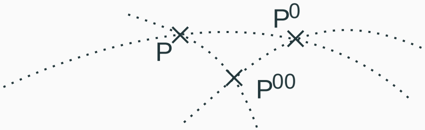
$$\text{e-hull}(V) = \left\{ \sigma(P) : \log[P] = \sum_{i=1}^{k} a_i \log[P_i], \right.$$
$$\left. k \in \mathbb{N}, a_1, \ldots, a_k \in \mathbb{R}, \sum_{i=1}^{k} a_i = 1, P_1, \ldots P_k \in V \right\}.$$

e-geodesic $g^{(e)}_{P,P'}(t), t \in (-\infty, +\infty)$

Definition 5 (Exponential hull)

Let $V \subseteq W$.

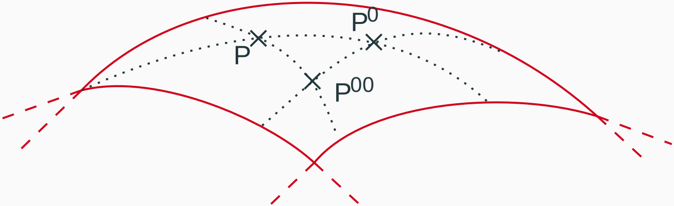
$$\text{e-hull}(V) = \left\{ s(P) : \log[P] = \sum_{i=1}^{k} \alpha_i \log[P_i], \right.$$
$$\left. k \in \mathbb{N}, \alpha_1, \ldots, \alpha_k \in \mathbb{R}, \sum_{i=1}^{k} \alpha_i = 1, P_1, \ldots P_k \in V \right\}.$$



Definition 5 (Exponential hull)

Let $V \subset W$.

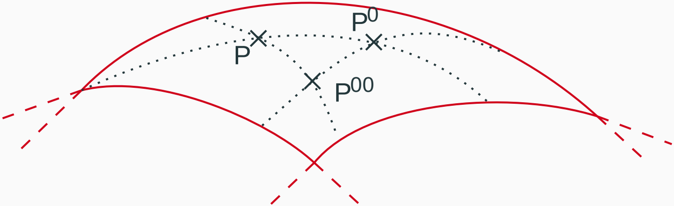
$$\text{e-hull}(V) = \left\{ \sigma(P) : \log[P] = \sum_{i=1}^{k} a_i \log[P_i], \right.$$
$$\left. k \in \mathbb{N}, a_1, \ldots, a_k \in \mathbb{R}, \sum_{i=1}^{k} a_i = 1, P_1, \ldots P_k \in V \right\}.$$



Generation of the reversible family from $\mathcal{W}_{sym}$

Definition 5 (Exponential hull)
Let $\mathcal{V} \subseteq \mathcal{W}$.

$$\text{e-hull}(\mathcal{V}) = \left\{ s(P) : \log[P] = \sum_{i=1}^{k} \alpha_i \log[P_i], \right.$$
$$\left. k \in \mathbb{N}, \alpha_1, \dots, \alpha_k \in \mathbb{R}, \sum_{i=1}^{k} \alpha_i = 1, P_1, \dots P_k \in \mathcal{V} \right\}.$$



Theorem 9

$$|\mathcal{X}| \geq 3 \implies \text{e-hull}(\mathcal{W}_{sym}) = \mathcal{W}_{rev}.$$

Reversible edge measures

Reversible family of edge measures

Edge-measures

P reversible $\implies Q \iff \operatorname{diag}(\pi) P$ is symmetric.

$$Q = \left(Q : \sum_{x' \in X} Q(x, x') = \sum_{x' \in X} Q(x', x) \right),$$

$$Q_{rev} = \{ Q \in Q : Q^{\star} = Q \}.$$

Reversible family of edge measures

Edge-measures

P reversible $\implies Q \iff \mathrm{diag}(\pi) P$ is symmetric.

$$\mathcal{Q} = \left(Q : \sum_{x' \in \mathcal{X}} Q(x, x') = \sum_{x' \in \mathcal{X}} Q(x', x) \right),$$

$$\mathcal{Q}_{\mathrm{rev}} = \{ Q \in \mathcal{Q} : Q^{\star} = Q \}.$$

Lemma 6

$$\mathcal{Q}_{\mathrm{rev}} \cong \mathcal{P} \left(\frac{|\mathcal{X}| (|\mathcal{X}| + 1)}{2} \right),$$

where $\cong$ denotes Markov equivalence.

Reversible family of edge measures

Edge-measures

P reversible $\implies \mathcal{Q} \iff \mathrm{diag}(\pi)P$ is symmetric.

$$\mathcal{Q} = \left\{ Q : \sum_{x' \in \mathcal{X}} Q(x, x') = \sum_{x' \in \mathcal{X}} Q(x', x) \right\},$$

$$\mathcal{Q}_{rev} = \{ Q \in \mathcal{Q} : Q^{\star} = Q \}.$$

Lemma 6

$$\mathcal{Q}_{rev} = \mathcal{P} \quad \frac{|\mathcal{X}|(|\mathcal{X}|+1)}{2},$$

where $=$ denotes Markov equivalence.

Theorem 8

(i) $\mathcal{Q}_{rev}$ is e-family and m-family of $\mathcal{P}(\mathcal{X}^2)$.

Edge-measures

P reversible $\Rightarrow$ $Q \Leftrightarrow \mathrm{diag}(\pi)P$ is symmetric.

$$Q = \left(Q : \sum_{x' \in X} Q(x, x') = \sum_{x' \in X} Q(x', x) \right),$$

$$Q_{rev} = \{ Q \in Q : Q^{?} = Q \}.$$

Lemma 6

$$Q_{rev} = P \frac{|X|(|X|+1)}{2},$$

where $=$ denotes Markov equivalence.

Theorem 8

(i) Q_{rev} is e-family and m-family of $P(X^2)$.

(ii) $\dim Q_{rev} = |X|(|X|+1)/2 \quad 1$.

Edge-measures

P reversible $\Rightarrow$ $Q \Leftrightarrow \mathrm{diag}(\pi) P$ is symmetric.

$$Q = \left(Q : \sum_{x' \in X} Q(x, x') = \sum_{x' \in X} Q(x', x) \right),$$

$$Q_{rev} = \{ Q \in Q : Q^{\star} = Q \}.$$

Lemma 6

$$Q_{rev} = P \quad \frac{|X|(|X|+1)}{2} ,$$

where $=$ denotes Markov equivalence.

Theorem 8

(i) Q_{rev} is e-family and m-family of $P(X^2)$.

(ii) $\dim Q_{rev} = |X|(|X|+1)/2 \quad 1$.

(iii) Q is not e-family in $P(X^2)$ (except when $|X| = 2$).

Summary

(*i*) Property of being an e/m-family is invariant under time-reversal operation.

(*ii*) We introduce reversible e-families, and provide a characterization (efficiently verifiable).

(*iii*) Perhaps surprisingly, W_{rev} forms an e/m-family. We further construct a basis, and derive an explicit parametrization.

(*iv*) We show that e/m-projections, possess closed-form expressions that verify Pythagorean identities, and are always equidistant from an irreducible Markov kernel and its time-reversal (bisection property).

(*v*) We characterize the reversible family as both the minimal exponential family that comprises symmetric kernels, and the smallest mixture family that contains memoryless Markov kernels.

(*vi*) We show that, unlike the set of all irreducible edge measures, the restriction to reversible edge measures forms an e-family (in distributions over pairs).

References

Imre Csiszár, Th Cover, and Byoung-Seon Choi. Conditional limit theorems under Markov conditioning. *IEEE Transactions on Information Theory*, 33(6):788–801, 1987.

Roland L Dobrushin, Yu M Sukhov, and József Fritz. A.N. Kolmogorov - the founder of the theory of reversible Markov processes. *Russian Mathematical Surveys*, 43(6):157–182, 1988.

Monroe D Donsker and SR Srinivasa Varadhan. Asymptotic evaluation of certain Markov process expectations for large time, i. *Communications on Pure and Applied Mathematics*, 28(1):1–47, 1975.

James Allen Fill. Eigenvalue bounds on convergence to stationarity for nonreversible Markov chains, with an application to the exclusion process. *The annals of applied probability*, pages 62–87, 1991.

Jürgen Gärtner. On large deviations from the invariant measure. *Theory of Probability & Its Applications*, 22(1):24–39, 1977.

Masahito Hayashi. Local equivalence problem in hidden Markov model. *Information Geometry*, 2(1):1–42, 2019.

Masahito Hayashi. Information geometry approach to parameter estimation in hidden Markov model. *Bernoulli*, 28(1):307–342, 2022.

Masahito Hayashi and Shun Watanabe. Information geometry approach to parameter estimation in Markov chains. *The Annals of Statistics*, 44(4): 1495 – 1535, 2016. doi: 10.1214/15-AOS1420.

Hisashi Ito and Shun-ichi Amari. Geometry of information sources. In *Proceedings of the 11th Symposium on Information Theory and Its Applications (SITA '88)*, pages 57–60, 1988.

A Kolmogorov. Zur umkehrbarkeit der statistischen naturgesetze. *Mathematische Annalen*, 113(1):766–772, 1937.

Andrei Kolmogorov. Zur theorie der Markoffschen ketten. *Mathematische Annalen*, 112(1):155–160, 1936.

HD Miller. A convexity property in the theory of random variables defined on a finite Markov chain. *The Annals of mathematical statistics*, pages 1260–1270, 1961.

Vrettos Moulos and Venkat Anantharam. Optimal Chernoff and Hoeffding bounds for finite state Markov chains. *arXiv preprint arXiv:1907.04467*, 2019.

Hiroshi Nagaoka. The exponential family of Markov chains and its information geometry. In *The proceedings of the Symposium on Information Theory and Its Applications*, volume 28, pages 601–604, 2005. (available on arXiv:1701.06119).

Kenji Nakagawa and Fumio Kanaya. On the converse theorem in statistical hypothesis testing for Markov chains. *IEEE transactions on information theory*, 39(2):629–633, 1993.

AJN Nielsen and Marcus Weber. Computing the nearest reversible Markov chain. *Numerical Linear Algebra with Applications*, 22(3):483–499, 2015.

Erwin Schrödinger. Über die umkehrung der naturgesetze. *Sitzungsberichte der preussischen Akademie der Wissenschaften, physikalische mathematische Klasse*, 8(N9):144–153, 1931.

Jun-ichi Takeuchi and Andrew R Barron. Asymptotically minimax regret by Bayes mixtures. In *Proceedings. 1998 IEEE International Symposium on Information Theory (Cat. No. 98CH36252)*, page 318. IEEE, 1998.

Jun'ichi Takeuchi and Tsutomu Kawabata. Exponential curvature of Markov models. In *2007 IEEE International Symposium on Information Theory*, pages 2891–2895. IEEE, 2007.

Jun'ichi Takeuchi and Hiroshi Nagaoka. On asymptotic exponential family of Markov sources and exponential family of Markov kernels, 2017.

Shun Watanabe and Masahito Hayashi. Finite-length analysis on tail probability for Markov chain and application to simple hypothesis testing. *Ann. Appl. Probab.*, 27(2):811–845, 04 2017. doi: 10.1214/16-AAP1216.

Geoffrey Wolfer and Aryeh Kontorovich. Improved estimation of relaxation time in non-reversible Markov chains, 2022.

Thank you for listening!

doi:10.1007/s41884-021-00061-7

geoffrey.wolfer@riken.jp