On Gradient Flows and Variational Structures for Markov Chains

Marcus Kaiser

University of Bath

Gradient Flows: Challenges and New Directions,
ICMS, Edinburgh
11.09.2018

Joint work with Rob Jack (Cambridge) and Johannes Zimmer (Bath).

Classical Gradient Flow

A gradient flow describes the time evolution of some quantity ρ_t (usually a density) in terms of an **energy functional** $\mathcal{F}(\rho)$ and a **metric** $M(\rho)$:

$$\dot{\rho}_t = -M(\rho_t)\frac{\delta \mathcal{F}}{\delta \rho_t}.$$

The classical example is the diffusion equation

$$\dot{\rho}_t = \Delta \rho_t + \nabla \cdot (\rho_t \nabla V),$$

for a (smooth and confining) external potential V. In this case one choice for the metric and energy are given by

$$M(\rho) = -\nabla \cdot \rho \nabla \quad \text{and} \quad \mathcal{F}(\rho) = \int \rho \log\left(\frac{\rho}{\mathrm{e}^{-V}}\right) \mathrm{d}x.$$

Physics perspective

From a physics perspective, the above diffusion equation

$$\dot{\rho}_t = \Delta \rho_t + \nabla \cdot (\rho_t \nabla V)$$

can be interpreted as a continuity equation $\dot{\rho}_t = -\nabla \cdot J(\rho_t)$, where the current (or flux)

$$J(\rho) = -\nabla \rho - \rho \nabla V$$

describes the flow of particles (on a macroscopic scale).

We can define a 'force' $F(\rho) = -\nabla \log \rho - \nabla V = -\nabla \frac{\delta \mathcal{F}}{\delta \rho}$, such that

$$J(\rho) = \rho F(\rho).$$

(This is a version of the (linear) Onsager relation between fluxes and forces.)

Comparison of two different representations

Gradient Flow:

$$\dot{\rho}_t = -M(\rho_t)\frac{\delta \mathcal{F}}{\delta \rho_t}$$

$$M(\rho_t) = -\nabla \cdot \rho_t \nabla$$

describes duality between velocities and 'potentials'

$$\dot{\rho}_t \Longleftrightarrow \frac{\delta \mathcal{F}}{\delta \rho_t}$$

Flux-Force:

$$\dot{\rho}_t = -\nabla \cdot J(\rho_t)$$

$$J(\rho_t) = \rho_t F(\rho_t)$$

$$F(\rho_t) = -\nabla \frac{\delta \mathcal{F}}{\delta \rho_t}$$

describes duality between fluxes and forces

$$J(\rho_t) \Longleftrightarrow F(\rho_t)$$

Non-conservative case (without time-reversal symmetry)

For $E \neq -\nabla V$ consider (the non-equilibrium system)

$$\dot{\rho}_t = \Delta\rho_t - \nabla \cdot (\rho_t E).$$

The right hand side above is still valid for $F(\rho) = -\nabla \log \rho + E$, but there exists no gradient flow representation...

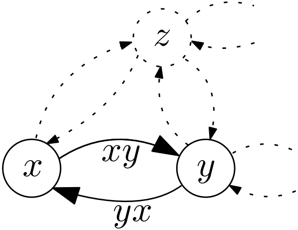
$\Rightarrow$ **This suggests that the representation on the right hand side is more general (and that one in general might want to consider the duality between fluxes and forces).**

(This statement is supported by e.g. results from *Large Deviation Theory*.)

Markov chains

Consider a finite state Markov chain (in continuous time) with transition rates r_{xy} such that

- $r_{xy} > 0$ if and only if $r_{yx} > 0$
- there exists a unique steady state π.



Markov chains

The associated generator $\mathcal{L}$ is a matrix with off-diagonal entries given by the rates r_{xy}.

The time-evolution of the probability density μ_t is in this case given by

$$\dot{\mu}_t = \mathcal{L}^\dagger \mu_t$$

(where $\mathcal{L}^\dagger$ is the adjoint of $\mathcal{L}$).

Equivalently, we can write the dynamics as

$$\dot{\mu}_t(x) = -\nabla \cdot J(\mu_t)(x) := -\sum_y J_{xy}(\mu_t)$$

for the "probability current"

$$J_{xy}(\mu) = \mu(x) r_{xy} - \mu(y) r_{yx}.$$

Maas-Mielke Gradient flow

Assume the above Markov chain satisfies detailed balance, i.e.

$$\pi(x)r_{xy} = \pi(y)r_{yx},$$

or equivalently $J(\pi) = 0$. In this case we can write the time-evolution as the gradient flow

$$\dot{\mu}_t = -\mathbb{M}(\mu_t)\frac{\delta\mathcal{F}}{\delta\mu_t},$$

for the free energy (relative entropy)

$$\mathcal{F}(\mu) = \sum_x \mu(x)\log\Big(\frac{\mu(x)}{\pi(x)}\Big)$$

and a metric $\mathbb{M}(\mu)$ that acts on some f as

$$\big[\mathbb{M}(\mu)f\big](x) = \sum_y \theta\big(\mu(x)r_{xy}, \mu(y)r_{yx}\big)\big(f(y) - f(x)\big) \approx \nabla\cdot\theta(\mu)\nabla f$$

(Here $\theta(a, b) = (a - b)/(\log(a) - \log(b))$ is the logarithmic mean.)

Representation in terms of fluxes and forces

For a general Markov chain (without detailed balance), we can define a force $F(\mu)$ and a mobility $a(\mu)$ as

$$F_{xy}(\mu) = \log\Big(\frac{\mu(x)r_{xy}}{\mu(y)r_{yx}}\Big) \quad \text{and} \quad a_{xy}(\mu) = 2\sqrt{\mu(x)r_{xy}\mu(y)r_{yx}}.$$

With these quantities, we can rewrite the probability current $J(\mu)$ as

$$J(\mu) = a(\mu)\sinh\big(\tfrac{1}{2}F(\mu)\big).$$

Note that this yields a **non-linear flux-force relation**, as opposed to the prior cases, where flux and force satisfy a linear relation.

Nonetheless, this is a structure, which can again be derived from Large Deviations (which we will discuss below).

Comparison for Markov chains (reversible case)

Gradient Flow:

$$\dot{\mu}_t = -\mathbb{M}(\mu_t)\frac{\delta \mathcal{F}}{\delta \mu_t}$$

$$\mathbb{M}(\mu_t) \approx -\nabla \cdot \theta(\mu_t)\nabla$$

describes duality between velocities and 'potentials'

$$\dot{\mu}_t \Longleftrightarrow \frac{\delta \mathcal{F}}{\delta \mu_t}$$

linear relation

Flux-Force:

$$\dot{\mu}_t = -\nabla \cdot J(\mu_t)$$

$$J(\mu_t) = a(\mu_t)\sinh\left(\tfrac{1}{2}F(\mu_t)\right)$$

$$F(\mu_t) = -\nabla\frac{\delta \mathcal{F}}{\delta \mu_t}$$

describes duality between fluxes and forces

$$J(\mu_t) \Longleftrightarrow F(\mu_t)$$

non-linear relation

Overview for general case

Gradient Flow (Diffusion):

$$\dot{\rho}_t = -M(\rho_t)\frac{\delta \mathcal{F}}{\delta \rho_t}$$

$$M(\rho_t) = -\nabla \cdot \rho_t \nabla$$

Flux-Force (Diffusion):

$$\dot{\rho}_t = -\nabla \cdot J(\rho_t)$$

$$J(\rho_t) = \rho_t F(\rho_t)$$

Gradient Flow (MC):

$$\dot{\mu}_t = -\mathbb{M}(\mu_t)\frac{\delta \mathcal{F}}{\delta \mu_t}$$

$$\mathbb{M}(\mu_t) \approx -\nabla \cdot \theta(\mu_t)\nabla$$

Flux-Force (MC):

$$\dot{\mu}_t = -\nabla \cdot J(\mu_t)$$

$$J(\mu_t) = a(\mu_t)\sinh\left(\tfrac{1}{2}F(\mu_t)\right)$$

Large Deviation Principle

For both diffusions and Markov Chains, a pathwise LDP **for density and current** can be derived. The **rate function** is of the form

$$I_{[0,\tau]}\big((\rho_t, j_t)_{t\in[0,\tau]}\big) = \frac{1}{2}\int_0^\tau \Phi\big(\rho_t, j_t, F(\rho_t)\big)\,\mathrm{d}t,$$

where

$$\Phi(\rho, j, F) = \Psi(\rho, j) - \langle j, F\rangle + \ \Psi^\star(\rho, F) \geq 0$$

is defined in terms of a Ψ-$\Psi^\star$ structure (via **Legendre duality**).

For diffusions, we have

$$\Psi^\star(\rho, F) := \frac{1}{2}\int \rho |F|^2\,\mathrm{d}x$$

and for Markov chains

$$\Psi^\star(\mu, F) = \sum_{x,y} a_{xy}(\mu)\big(\cosh\big(\tfrac{1}{2}F_{xy}\big) - 1\big).$$

Minimisers for Rate Function

The minimisers of $\Phi(\rho, j, F) = \Psi(\rho, j) - \langle j, F\rangle + \Psi^\star(\rho, F)$ satisfy

$$j = \frac{\delta\Psi^\star(\rho,\cdot)}{\delta F(\rho)}$$

(up to a multiplicative factor that depends on the way the inner product $\langle j, F\rangle$ is defined...)

For diffusions this finally leads to

$$J(\rho) = \frac{\delta\Psi^\star(\rho,\cdot)}{\delta F(\rho)} = \rho F(\rho)$$

and for Markov chains we have

$$J(\mu) = 2\frac{\delta\Psi^\star(\mu,\cdot)}{\delta F(\mu)} = a(\mu)\sinh\big(\tfrac{1}{2}F(\mu)\big).$$

(Further examples covered by Ψ-$\Psi^\star$ are the classical Onsager Theory and Macroscopic Fluctuation Theory).

Thank you!

Some references:

- Gradient flow structure for Markov chains:

  Maas, 2011, Journal of Functional Analysis 8.261, 2250-2292.

  Mielke, 2011, Nonlinearity 24.4, 1329

- Ψ-$\Psi^\star$ formulas:

  Maes & Netočný, 2008, EPL (Europhysics Letters) 82.3, 30003.

  Mielke, Peletier & Renger, 2014, Potential Analysis 41.4, 1293-1327.

  K., Jack & Zimmer, 2018, Journal of Statistical Physics 170.6, 1019-1050.

- Convergence of Ψ-$\Psi^\star$ to hydrodynamic limits:

  K., Jack & Zimmer, 2018, arXiv preprint arXiv:1805.01411.