大偏差原理和梯度流试讲

黄忠淦
zhonggan@math.utah.edu

University of Utah
Department of Mathematics

目录

1. 大偏差原理
   - 作为大数定律的延伸
   - 一般的原理表述
   - Sanov 定理
2. 最优传输与梯度流
   - 梯度流简介
   - 最优传输简介
   - JKO 格式：热扩散作为熵的梯度流
   - Benamou-Brenier 公式与 Otto 的形式黎曼结构
   - 梯度流理论作为一个统一的框架
3. 浅谈非平衡态物理系统
4. 参考资料

大偏差原理

作为大数定律的延伸

我们考虑一个抛硬币的系统：记 $X_i \overset{iid}{\sim} Ber(1/2)$ 为 n 次抛硬币的结果，那么根据强大数定律，我们知道

$$\frac{S_n}{n} := \frac{\sum_{i=1}^n X_i}{n} \xrightarrow{a.s.} \frac{1}{2}.$$

根据如上定律我们得知当 n 足够大时，我们可以用 $n/2$ 来近似 S_n。

作为大数定律的延伸

我们考虑一个抛硬币的系统：记 $X_i \overset{iid}{\sim} Ber(1/2)$ 为 n 次抛硬币的结果，那么根据强大数定律，我们知道

$$\frac{S_n}{n} := \frac{\sum_{i=1}^n X_i}{n} \xrightarrow{a.s.} \frac{1}{2}.$$

根据如上定律我们得知当 n 足够大时，我们可以用 $n/2$ 来近似 S_n。

可若是我们希望量化其它更为稀少的事件发生的概率呢？比如对于某个 $a > 1/2$，我们计算

$$\begin{aligned} P(S_n = \lfloor an \rfloor) &= \frac{n!}{\lfloor an \rfloor!(n - \lfloor an \rfloor)!} \frac{1}{2^n} \\ &\approx e^{-(?)n}. \end{aligned}$$

作为大数定律的延伸

可若是我们希望量化其它更为稀少的事件发生的概率呢？比如对于某个 $a > 1/2$，我们计算

$$P(S_n = \lfloor an \rfloor) = \frac{n!}{\lfloor an \rfloor!(n - \lfloor an \rfloor)!} \frac{1}{2^n} \\ \approx e^{-(?)n}.$$

那么此时我们的问题就变成了

$$-\frac{1}{n} \log P(S_n = \lfloor an \rfloor) \approx (?).$$

作为大数定律的延伸

根据 Stirling 公式，我们知道

$$\lim_{n\to\infty} -\frac{1}{n}\log P(S_n = \lfloor an \rfloor) = I_{1/2}(a), \smiley$$

其中

$$I_p(a) = a\log\left(\frac{s}{p}\right) + (1-a)\log\left(\frac{1-a}{1-p}\right).$$

式子 ☺ 就是这个系统对应的大偏差原理。而函数 $I_{1/2}$ 被称为这个系统的 rate function，亦可称为熵函数。

一般的原理表述

现在我们来讨论一般的原理表述。假设 $\mathcal{X}$ 是一个 Hausdorff 空间，且记 $\mathcal{B}_{\mathcal{X}}$ 为由其上拓扑生成的 Borel σ-代数。记 $\mathcal{P}(\mathcal{X})$ 为 $\mathcal{X}$ 上所有的 Borel 概率测度。

定义

一个函数 $f : \mathcal{X} \to [-\infty, \infty]$ 被称为**下半连续**若对所有 $c \in \mathbb{R}$ 都有

$$\{f \le c\} = \{x \in \mathcal{X} : f(x) \le c\}$$

是闭子集。

一般的原理表述

定义

一个函数 $f:\mathcal{X}\to[-\infty,\infty]$ 被称为**下半连续**若对所有 $c\in\mathbb{R}$ 都有

$$\{f\le c\}=\{x\in\mathcal{X}:\ f(x)\le c\}$$

是闭子集。

定义

令 $I:\mathcal{X}\to[0,\infty]$ 为一个下半连续的函数，且 $r_n\nearrow\infty$ 是一串正实数。一列概率测度 $(\mu_n)_n\subset\mathcal{P}(\mathcal{X})$ 满足关于熵函数 I 与规范系数 r_n 的**大偏差原理**若如下不等式满足

$$\limsup_{n\to\infty}\frac{1}{r_n}\log\mu_n(F)\le-\inf_{x\in F}I(x),\ \forall\text{闭子集}F\subset\mathcal{X},$$

$$\liminf_{n\to\infty}\frac{1}{r_n}\log\mu_n(G)\ge-\inf_{x\in G}I(x),\ \forall\text{开子集}G\subset\mathcal{X}.$$

Sanov 定理

接下来我们讨论一个结构更复杂的空间的大偏差原理：Sanov 定理。假设 $X_i \overset{iid}{\sim} \mu$ 是实轴上 n 个随机变量，我们想要了解其经验分布 $\nu_n = \dfrac{1}{n}\sum_{i=1}^{n}\delta_{X_i}$ 的渐进性质，其中 δ_{X_i} 是质量在 X_i 的 Dirac 测度。

Sanov 定理

接下来我们讨论一个结构更复杂的空间的大偏差原理：Sanov 定理。假设 $X_i \overset{iid}{\sim} \mu$ 是实轴上 n 个随机变量，我们想要了解其经验分布 $\nu_n = \frac{1}{n}\sum_{i=1}^{n}\delta_{X_i}$ 的渐进性质，其中 δ_{X_i} 是质量在 X_i 的 Dirac 测度。

定理 (Sanov)

经验分布 ν_n 的分布 $P(\nu_n \in \cdot) \in \mathcal{P}(\mathcal{P}(\mathbb{R}))$ 满足如下大偏差原理：

$$\limsup_{n\to\infty} \frac{1}{n}\log P(\nu_n \in F) \le -\inf_{x\in F} H(\cdot|\mu), \forall \text{闭子集} F \subset \mathcal{P}(\mathbb{R}),$$

$$\liminf_{n\to\infty} \frac{1}{n}\log P(\nu_n \in G) \ge -\inf_{x\in G} H(\cdot|\mu), \forall \text{开子集} G \subset \mathcal{P}(\mathbb{R}).$$

Sanov 定理 [6]

定理 (Sanov)

经验分布 ν_n 的分布 $P\left(\nu_n \in \cdot\right) \in \mathcal{P}\left(\mathcal{P}(\mathbb{R})\right)$ 满足如下大偏差原理:

$$\limsup_{n\to\infty} \frac{1}{n} \log P\left(\nu_n \in F\right) \leq -\inf_{\lambda\in F} H(\lambda|\mu),\ \forall 闭子集 F \subset \mathcal{P}(\mathbb{R}),$$

$$\liminf_{n\to\infty} \frac{1}{n} \log P\left(\nu_n \in G\right) \geq -\inf_{\lambda\in G} H(\lambda|\mu),\ \forall 开子集 G \subset \mathcal{P}(\mathbb{R}).$$

其中对于 $\mu, \lambda \in \mathcal{P}(\mathcal{X})$,

$$H(\lambda|\mu) = \begin{cases} \int_{\mathbb{R}} \phi \log \phi d\mu, & 若 \lambda << \mu,\ 且 \phi = \frac{d\lambda}{d\mu}, \\ +\infty, & 其他情况。\end{cases}$$

这里的 $H(\cdot|\cdot)$ 被称为相对熵。注意到 $H(\lambda|\mu) \geq 0$ 且等号成立当且仅当 $\mu = \lambda$。

最优传输与梯度流

梯度流简介

给定一个黎曼流形 (M, g)，我们记上面的度量为

$$d(x,y)^2 := \inf\left\{\int_0^1 g_{x_t}(\dot{x}_t, \dot{x}_t)dt \,\middle|\, x_0 = x,\, x_1 = y\right\}.$$

由于 g_x 是内积，根据 Riesz 表示定理，我们得到

$$G_x : T_x M \to T_x M^*,$$

使得 $G_x(v)w = g_x(v, w)$ 对于所有的 $v, w \in T_x M$ 成立。记

$$K_x = G_x^{-1} : T_x M^* \to T_x M.$$

注：我们可以将度量重新写为

$$d(x,y)^2 := \inf\left\{\int_0^1 \xi_t\left(K_{x_t}(\xi_t)\right)dt \,\middle|\, x_0 = x,\, x_1 = y,\, \xi_t \in T_{x_t}M^*, \text{且} \dot{x}_t = K_{x_t}(\xi_t)\right\}.$$

梯度流简介

由于 g_x 是内积，根据 Riesz 表示定理，我们得到

$$G_x : T_x M \to T_x M^*,$$

使得 $G_x(v)w = g_x(v, w)$ 对于所有的 $v, w \in T_x M$ 成立。记 $K_x = G_x^{-1} : T_x M^* \to T_x M$。

设 $f : M \to \mathbb{R}$ 是一个可微函数，$z_0 \in M$ 是某一个初始位置，那么任给一个 M 中从 z_0 出发的曲线 $x_t : [0, 1] \to M$，我们有

$$\begin{aligned}
\left.\frac{d}{dt}\right|_{t=0} f(x_t) &= \nabla f(z_0) \cdot \dot{x}_0 \\
&= G_{z_0}^{-1} \nabla f(z_0) \cdot G_{z_0} \dot{x}_0 \\
&= g_{z_0}\left(G_{z_0}^{-1} \nabla f(z_0), \dot{x}_0\right) \\
&= g_{z_0}\left(K_{z_0} \nabla f(z_0), \dot{x}_0\right).
\end{aligned}$$

梯度流简介

设 $f : M \to \mathbb{R}$ 是一个可微函数，$z_0 \in M$ 是某一个初始位置，那么任给一个 M 中从 z_0 出发的曲线 $x_t : [0,1] \to M$，我们有

$$
\begin{aligned}
\left.\frac{d}{dt}\right|_{t=0} f(x_t) &= \nabla f(z_0) \cdot \dot{x}_0 \\
&= G_{z_0}^{-1} \nabla f(z_0) \cdot G_{z_0} \dot{x}_0 \\
&= g_{z_0}\left(G_{z_0}^{-1} \nabla f(z_0), \dot{x}_0\right) \\
&= g_{z_0}\left(K_{z_0} \nabla f(z_0), \dot{x}_0\right).
\end{aligned}
$$

这时我们注意到若是想要 x_t 最速下降，那么它应当满足如下常微分方程

$$
\begin{cases}
\dot{x}_t = -K_{x_t} \nabla f(x_t),\ t > 0 \\
x_0 = z_0.
\end{cases}
$$

这时方程的解 x_t 便被称为函数 f 在流形 (M, g) 上的梯度流。

最优传输简介：Monge 问题

给定 $Z \subset \mathbb{R}^n$，我们记 $\mathcal{P}(Z)$ 为 Z 上所有有的概率测度。任给 $X, Y \subset \mathbb{R}^n$ 及 $\mu \in \mathcal{P}(X)$, $\nu \in \mathcal{P}(Y)$，我们考虑传输映射

$$\mathcal{T}(\mu,\nu) := \{T : X \to Y \ ; \ \mu(T^{-1}(U)) = \nu(U) \, \forall U \subset Y\}.$$

一个传输映射的花费记为

$$C(T) := \int_X c(x, T(x)) d\mu(x).$$

而 **Monge 问题** (1781) 则问

$$\min_{T \in \mathcal{T}(\mu,\nu)} C(T).$$

最优传输简介：Kantorovitch 问题

我们现在讨论以上问题的简化版本。我们考虑传输规划 ($\pi_x(x,y) = x$ and $\pi_y(x,y) = y$) 使得

$$\Pi(\mu,\nu) := \{\gamma \in \mathcal{P}(X \times Y)\ ;\ \gamma(\pi_x^{-1}(\cdot)) = \mu,\ \gamma(\pi_y^{-1}(\cdot)) = \nu\}.$$

传输规划的花费则记为

$$C(\gamma) := \int_{X\times Y} c(x,y) d\gamma(x,y).$$

Kantorovitch 问题 (1942) 则问

$$\min_{\gamma\in\Pi(\mu,\nu)} C(\gamma).$$

这个问题确实简化了 Monge 问题，因为任给传输映射 $T \in \mathcal{T}(\mu,\nu)$，那么 $\gamma := \mu((Id_x \times T)^{-1}(\cdot)) \in \Pi(\mu,\nu)$ 且 $C(\gamma) = C(T)$。

最优传输简介：Wasserstein 度量

以上两个问题经过了多年的讨论，已经非常成熟。在这里我们不讨论它们的特性，而是用这个问题来为概率空间赋予合适的度量。

定义

任给 $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$，我们定义

$$W_2(\mu,\nu) := \left(\inf_{\gamma \in \Pi(\mu,\nu)} \int_{\mathbb{R}^d \times \mathbb{R}^d} |\mathrm{x}-\mathrm{y}|^2 d\gamma(\mathrm{x},\mathrm{y}) \right)^{1/2}.$$

注：W_2 是与底层空间中的度量高度契合的：事实上给定 $\mathrm{x},\mathrm{y} \in \mathbb{R}^d$，我们有

$$W_2(\delta_{\mathrm{x}}, \delta_{\mathrm{y}}) = |\mathrm{x}-\mathrm{y}|.$$

这是这个度量与 L^2 度量的一个本质区别！

非平衡态系统的运动

给定一个物理系统，我们知道当这个系统处于平衡态时，我们很容易观测到系统的状态便是熵函数对应的最小值点：也即是随着观察次数增加，出现率减少最慢的那个状态。

非平衡态系统的运动

给定一个物理系统，我们知道当这个系统处于平衡态时，我们很容易观测到系统的状态便是熵函数对应的最小值点：也即是随着观察次数增加，出现率减少最慢的那个状态。

若是有一个处在非平衡态的物理系统，我们希望知道它是如何演化到平衡态的时候，我们怎么做呢？那便是考虑熵函数的梯度流。

非平衡态系统的运动

给定一个物理系统，我们知道当这个系统处于平衡态时，我们很容易观测到系统的状态便是熵函数对应的最小值点：也即是随着观察次数增加，出现率减少最慢的那个状态。

若是有一个处在非平衡态的物理系统，我们希望知道它是如何演化到平衡态的时候，我们怎么做呢？那便是考虑熵函数的梯度流。

简单探讨一下其实不难发现 L^2 并不能作为我们需要的度量，因为它在概率空间上的行为很不自然。这时我们就自然认识到应当讨论熵函数在 Wasserstein 度量下的梯度流！

Jordan-Kinderlehrer-Otto 格式 [3]

由于 Wassertein 度量复杂的形式，我们希望有一个不那么依赖导数的方式来得到熵函数的梯度流。我们首先回到黎曼流形上的梯度流：

$$\dot{x}_t = -K_{x_t}\nabla f(x_t).$$

将其与任意 $w \in T_{x_t}M$ 作内积，我们得到

$$g_{x_t}(\dot{x}_t, w) + \nabla f(x_t)\cdot w = 0,$$

经过计算上式左边等于

$$\nabla\frac{d^2(z_0, x_t)}{2t}\cdot w + \nabla f(x_t)\cdot w = 0.$$

当 $t > 0$ 足够小且 f 足够光滑时，我们知道 $y \mapsto f(y) + \dfrac{d^2(z_0, y)}{2t}$ 就是一个凸函数，那么这时 x_t 便是这个新函数唯一的极小值点。JKO 格式便是由这一思想而来。

Jordan-Kinderlehrer-Otto 格式

任给 $n>0$ 以及初始概率测度 $\rho_0\in\mathcal{P}(\mathbb{R}^d)$，我们考虑格式

$$\rho^n_{m+1}=\mathrm{argmin}_{\rho\in\mathcal{P}(\mathbb{R}^d)}\left\{H(\rho|L)+\frac{W_2^2(\rho^n_m,\rho)}{2/n}\right\},$$

其中 L 为 Lebesgue 测度。记 $\rho^n_t=\rho^n_{\lfloor tn\rfloor}$, $t>0$。

定理 (Jordan-Kinderlehrer-Otto)

给定 $t>0$，我们有在弱的 L^1 意义下，

$$\rho^n_t \overset{n\to\infty}{\rightharpoonup} \rho_t,$$

其中 $\rho_t\in C^\infty((0,+\infty)\times\mathbb{R}^d)$ 满足热方程

$$\begin{cases}\partial_t\rho_t=\Delta\rho_t\\ \rho_t\xrightarrow{L^1}\rho_0,\ \text{随着}t\to 0^+.\end{cases}$$

Benamou-Brenier 公式 [2]

我们应当如何理解 JKO 格式呢？有没有更简洁的解释？

Benamou-Brenier 公式

我们应当如何理解 JKO 格式呢？有没有更简洁的解释？

事实上我们知道 $(\mathcal{P}(\mathbb{R}^d), W_2)$ 是一个测地空间：也即是说任给两个点都存在测地线将它们相连。

定理 (Benamou-Brenier)

假设 $\nabla\psi : \mathbb{R}^d \times [0,1] \to \mathbb{R}^d$ 代表时空中的向量场，那么我们可以将 *Wasserstein* 度量写为

$$W_2(f^0 dx, f^1 dx)^2 = \min\left\{\int_{[0,1]^2}\int_0^1 f_t(x)|\nabla\psi(x,t)|^2 dt dx\right\},$$

限制条件为如下偏微分方程

$$\partial_t f + \nabla\cdot(f\nabla\psi) = 0,\ f_0 = f^0,\ f_1 = f^1.$$

Benamou-Brenier 公式

定理 (Benamou-Brenier)

假设 $V:\mathbb{R}^d\times[0,1]\to\mathbb{R}^d$ 代表时空中的向量场，那么我们可以将 *Wasserstein* 度量写为

$$W_2(f^0dx, f^1dx)^2 = \min\left\{\int_{[0,1]^2}\int_0^1 f_t(x)|\nabla\psi(x,t)|^2 dtdx\right\}, \frownie$$

限制条件为如下偏微分方程

$$\partial_t f + \nabla\cdot(f\nabla\psi) = 0,\ f_0 = f^0,\ f_1 = f^1.$$

进一步地，假设 T_{opt} 是 f^0dx 到 f^1dx 的最优传输映射，那么 $\frownie$ 中最优向量场正好为 $\nabla\psi(\mathrm{x},t) = T_{opt}(\mathrm{x}) - \mathrm{x}$。

Otto 的形式黎曼流形 [4]

利用 Benamou-Brenier 公式，Otto 观察到了 $(\mathcal{P}(\mathbb{R}^d), W_2)$（形式上的）黎曼流形结构。

定义

对于测度 $\mu \in \mathcal{P}(\mathbb{R}^d)$，我们定义这个点上的黎曼结构为

$$K_{Otto}(\mu) =: K_\mu : \psi \mapsto -\nabla \cdot (\mu \nabla \psi) .$$

这个定义与之前的 Benamou-Brenier 公式是契合的。

Otto 的形式黎曼流形

如果 $\Phi(\mu = u(\mathrm{x})d\mathrm{x}) = \int F(u)d\mathrm{x}$，那么我们可以算出它的梯度为

$$\nabla\Phi(\mu) := K_\mu F'(u) = -\nabla_{\mathrm{x}} \cdot (u\nabla_{\mathrm{x}}F'(u)),$$

而其对应的梯度流为

$$\dot{u} = -\nabla\Phi = -\nabla_{\mathrm{x}} \cdot (u\nabla_{\mathrm{x}}F'(u)).$$

若我们将熵函数 $F(u) = u\log u$ 代入，则有

$$\begin{aligned}
\dot{u} &= -\nabla\Phi \\
&= \nabla_{\mathrm{x}} \cdot (u\nabla_{\mathrm{x}}F'(u)) \\
&= \nabla_{\mathrm{x}} \cdot (u\nabla_{\mathrm{x}}(1+\log u)) \\
&= \nabla_{\mathrm{x}} \cdot \left(u\frac{\nabla_{\mathrm{x}}u}{u}\right) \\
&= \Delta u.
\end{aligned}$$

梯度流理论作为一个统一的框架

不难发现 Otto 的形式黎曼流形结构极大地增加了整个理论的可能性，事实上，梯度流理论可以切实地统一非常多的扩散系统，这里我列一部分的方程：

Many diffusive systems are Wasserstein gradient flows

A surprisingly large number of well-known diffusive partial differential equations have the structure of a gradient flow with respect to the Wasserstein metric (or a related metric). These are just a few examples:

convection and nonlinear diffusion [JKO98]	$\partial_t \rho = \operatorname{div} \rho \nabla \left[U'(\rho) + V \right]$
nonlocal pbs ([AGS05, CMV03, CMV06] *et al.*)	$\partial_t \rho = \operatorname{div} \rho \nabla \left[U'(\rho) + V + W * \rho \right]$
thin-film equation [Ott98, GO01]	$\partial_t \rho = -\partial_x \left[\rho \partial_x^3 \rho \right]$
DLSS and related [DLSS91, JM09, MMS09]	$\partial_t \rho = -\operatorname{div} \rho \nabla \left(\rho^{\alpha-1} \Delta \rho^\alpha \right), \quad 1/2 \le \alpha \le 1$
a reactive moving-boundary problem [PP10]	$\partial_t \rho = \Delta \rho$ in $\Omega(t)$, with $\partial_n \rho = -\rho v_n$ and $v_n = f(\rho)$ on $\partial\Omega(t)$
the Cahn-Hilliard equation	$\partial_t \rho = -\operatorname{div} D(\rho) \nabla \left[\Delta \rho + f(\rho) \right]$
two-phase flow [Ott99]	$\partial_t \rho = -\operatorname{div} \rho u, \ \operatorname{div} u = 0, \ u = f(\rho) \left[\nabla p + \rho e_z \right]$

图: 可被囊括的方程 [5]

浅谈非平衡态物理系统

非平衡态物理系统的两个方面

我们可以从两个方面来研究非平衡态物理系统：一个是从多粒子系统

$$(\text{动态})\text{rate function} \overset{\text{大偏差原理}}{\longleftrightarrow} \text{随机多粒子系统}.$$

另一个是连续性的演化方程

$$\text{梯度流结构} \longleftrightarrow \text{连续性的演化方程}.$$

非平衡态物理系统的两个方面

事实上两个方面又明显有如下联系

$$
\begin{array}{ccc}
\text{(动态)rate function} & \xleftrightarrow{\text{大偏差原理}} & \text{随机多粒子系统} \\
\updownarrow ? & & \downarrow \text{连续性极限} \\
\text{梯度流结构} & \longleftrightarrow & \text{连续性的演化方程}
\end{array}
$$

非平衡态物理系统的两个方面

事实上两个方面又有如下联系

$$\begin{tikzcd} \text{(动态)rate function} \arrow[r, leftrightarrow, "\text{大偏差原理}"] \arrow[d, leftrightarrow, "?"] & \text{随机多粒子系统} \arrow[d, "\text{连续性极限}"] \\ \text{梯度流结构} \arrow[r, leftrightarrow] & \text{连续性的演化方程} \end{tikzcd}$$

‘?’ 处的联系由 [1] 给出。

多粒子布朗系统

我们从一个多粒子布朗运动系统来理解这个联系。

考虑 n 个互相独立的布朗运动粒子 $X_{n,i}(t) \in \mathbb{R}^d$，它们的初始位置给定 $X_{n,i}(0) = x_{n,i}$，每一个都满足均值 $x_{n,i}$ 及方差 $2t$（注意到这里我们使用的生成子为 Δ 而非常用的 $(1/2)\Delta$）。

与平衡态系统不同的是，我们希望了解经过一段很短的时间 $h > 0$，经验测度

$$\rho_n(h) := \frac{1}{n} \sum_{i=1}^{n} \delta_{X_{n,i}(h)}$$

的分布。

多粒子布朗系统

与平衡态系统不同的是，我们希望了解经过一段很短的时间 $h > 0$，经验测度

$$\rho_n(h) := \frac{1}{n}\sum_{i=1}^{n}\delta_{X_{n,i}(h)}$$

的分布。

记初始测度为 $\rho^0 = \lim_{n\to\infty}\rho_n(0)$, 经过计算，事实上我们有如下大偏差原理

$$Prob(\rho_n(h) \approx \rho^1) \approx \exp[-nI_h(\rho^1)],$$

其中

$$I_h(\rho^1) \approx \frac{1}{4h}W_2(\rho^0,\rho^1)^2 + \frac{1}{2}H(\rho^1|L) - \frac{1}{2}H(\rho^0|L), \text{ 当} h \to 0.$$

JKO 格式的熵解释

注意到 I_h 约等于 JKO 格式的一半，那么事实上我们就为 JKO 格式给出了熵解释：JKO 格式中经过时间 $h > 0$ 后新泛函 $H + (1/2h)W^2$ 的极小值点正好是对应粒子系统中的熵最小值点！
这样便将非平衡态系统的随机多粒子系统解释和演化方程的解释从熵和梯度流的角度联系起来。

参考资料

[1] Stefan Adams, Nicolas Dirr, Mark Peletier, and Johannes Zimmer. Large deviations and gradient flows. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 371(2005):20120341, dec 2013.

[2] Jean-David Benamou and Yann Brenier. A computational fluid mechanics solution to the monge-kantorovich mass transfer problem. *Numerische Mathematik*, 84:375–393, 2000.

[3] Richard Jordan, David Kinderlehrer, and Felix Otto. The variational formulation of the fokker–planck equation. *SIAM Journal on Mathematical Analysis*, 29(1):1–17, 1998.

[4] Felix Otto. The geometry of dissipative evolution equations: The porous medium equation. *Communications in Partial Differential Equations*, 26(1-2):101–174, 2001.

[5] Mark A. Peletier. Energies, gradient flows, and large deviations: a modelling point of view.

[6] I. N. Sanov. On the probability of large deviations of random variables. 1958.

感谢!