基于输运的方法

黄政宇
北京大学北京国际数学研究中心
北京大学国际机器学习研究中心

本堂课大纲

- 课程内容简介
  - 重要性采样 (Importance sampling)
  - 卡尔曼方法 (Kalman methodology)
  - 标准化流 (Normalizing flow)

贝叶斯反问题

- 贝叶斯反问题

$$y = \mathcal{G}(\theta) + \eta \qquad \eta \sim \rho_\eta \qquad \theta \sim \rho_{\text{prior}}$$

- 假设

高斯先验分布：$\rho_{\text{prior}}(\theta) = \mathcal{N}(\theta; r_0, \Sigma_0)$

高斯噪音：$\rho_\eta = \mathcal{N}(x; 0, \Sigma_\eta)$

- 后验分布

$$\rho_{\text{post}}(\theta; y) \propto \rho(y|\theta)\rho_{\text{prior}}(\theta) \propto e^{-\Phi_R(\theta, y)}$$

$$\rho_{\text{post}}(\theta; y) = \frac{1}{Z} e^{-\Phi_R(\theta, y)}$$

$$\Phi_R(\theta, y) = \frac{1}{2} \| \Sigma_\eta^{-\frac{1}{2}} (y - \mathcal{G}(\theta)) \|^2 + \frac{1}{2} \| \Sigma_0^{-\frac{1}{2}} (\theta - r_0) \|^2$$

贝叶斯采样、推理

- 有未知归一化常数的目标分布

$$\rho^*(\theta) = \frac{1}{Z} e^{-\Phi_R(\theta)}$$

已知

未知

$$\Phi_R(\theta, y) = \frac{1}{2} \| \Sigma_\eta^{-\frac{1}{2}} (y - \mathcal{G}(\theta)) \|^2 + \frac{1}{2} \| \Sigma_0^{-\frac{1}{2}} (\theta - r_0) \|^2$$

- 计算目标分布的期望、协方差等
- 计算目标函数的期望 $\mathbb{E}[f] = \int f(\theta)\rho^*(\theta)d\theta$
- 生成服从目标分布的样本 $\{\theta_j\} \sim \rho^*(\theta)$

贝叶斯采样、推理

- 基于输运的方法（直接近似的方法）

暴力网格搜索：假设$N_\theta = 2,$

$$\theta^{i,j} = [-L + \frac{2(i-1)}{N-1}L, -L + \frac{2(j-1)}{N-1}L]$$

$$Z = \sum \rho^*\left(\theta^{i,j}\right)\Delta x \Delta y$$

输运：$\{\theta^j\} \sim \rho_{\text{prior}}$　　$\{\mathcal{T}\theta^j\} \sim \rho^*$

- 重要性采样
- 卡尔曼方法
- 标准化流方法

……

重要性采样

- 蒙特卡洛方法

计算：$\mathbb{E}_{\rho^*}[f] = \int f(\theta)\rho^*(\theta)d\theta$

采样：$\{\theta^j\} \sim \rho^*(\theta)$

$$\mathbb{E}_{\rho^*}[f] \approx {\rho^*}^J_{\mathrm{MC}}(f) = \frac{1}{J}\sum_{j=1}^{J} f(\theta^j)$$

重要性采样

蒙特卡洛方法收敛性

对于　$f: R^{N\theta} \to R$，$\mathrm{Var}_\rho[f] = \mathbb{E}_\rho\left[\left(f - \mathbb{E}_\rho f\right)^2\right] < +\infty$

我们有

$$\mathbb{E}_\rho\left[\frac{1}{J}\sum_{j=1}^{J} f(\theta^j) - \mathbb{E}_\rho[f]\right] = 0$$

$$\mathbb{E}_\rho\left[\left(\frac{1}{J}\sum_{j=1}^{J} f(\theta^j) - \mathbb{E}_\rho[f]\right)^2\right] = \frac{\mathrm{Var}_\rho[f]}{J}$$

重要性采样

- 重要性采样

$$\rho^*\ (\theta) = \frac{1}{Z} e^{-\Phi(\theta)} \rho(\theta)$$

采样：$\{\theta^j\} \sim \rho(\theta)$

计算：$\Phi\left(\theta^j\right)$

计算权重：$w^j = \frac{e^{-\Phi\left(\theta^j\right)}}{\sum_{j=1}^J e^{-\Phi\left(\theta^j\right)}} = \frac{1}{J} \frac{e^{-\Phi\left(\theta^j\right)}}{\frac{1}{J}\sum_{j=1}^J e^{-\Phi\left(\theta^j\right)}}$

输运：$\left\{\theta^j\right\} \rightarrow \{w^j \theta^j\}$

$$\rho^*(\theta) \approx \sum_{j=1}^J w^j \delta\left(\theta - \theta^j\right)$$

$$\mathbb{E}_{\rho^*}[f] \approx \rho^{*J}_{\mathrm{IS}}(f) = \sum_{j=1}^J w^j f(\theta^j)$$

重要性采样

重要性采样方法收敛性

对于　$f: R^{N_\theta} \to R$，$\chi^2[\rho^* \| \rho] = \int \frac{{\rho^*}^2}{\rho} d\theta - 1$，

我们有

$$\sup_{|f|_\infty \le 1} \mathbb{E}_\rho \left[{\rho^*}^J_{\mathrm{IS}}(f) - \mathbb{E}_{\rho^*}[f] \right] \le 2 \frac{1 + \chi^2[\rho^* \| \rho]}{J}$$

$$\sup_{|f|_\infty \le 1} \mathbb{E}_\rho \left[\left({\rho^*}^J_{\mathrm{IS}}(f) - \mathbb{E}_{\rho^*}[f] \right)^2 \right] \le 4 \frac{1 + \chi^2[\rho^* \| \rho]}{J}$$

重要性采样

- 练习（Rosenbrock 函数）

$$y = \mathcal{G}(\theta) + \eta$$

$$\mathcal{G}(\theta) = \begin{bmatrix} \theta_2 - c_1\theta_1^2 \\ \theta_1 \end{bmatrix} \qquad y = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

高斯先验分布： $\rho_{\text{prior}}(\theta) = \mathcal{N}(\theta; 0, \begin{bmatrix} 10^2 & \\ & 10^2 \end{bmatrix})$

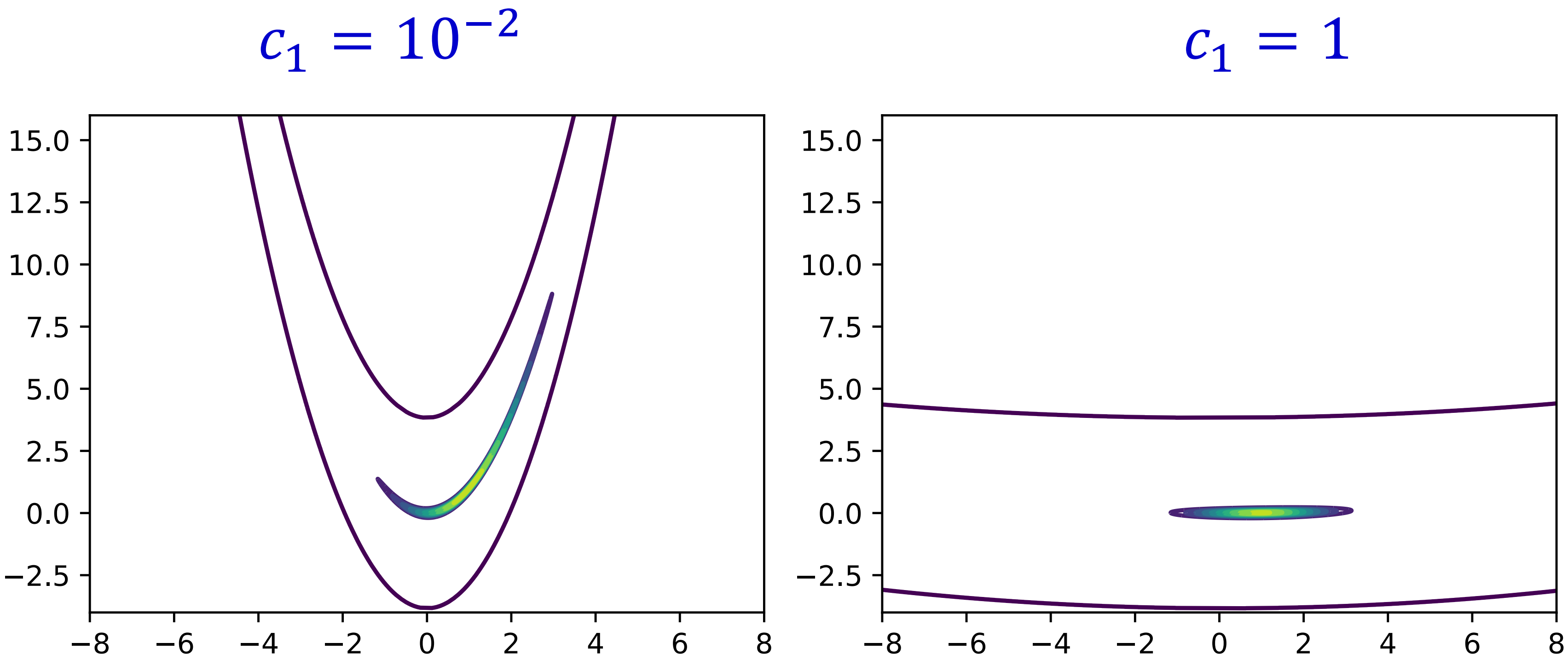
高斯噪音： $\rho_\eta = \mathcal{N}\left(x; 0, \begin{bmatrix} \frac{1}{10^2} & \\ & 1 \end{bmatrix}\right)$

后验分布： $\rho^*(\theta) = \frac{1}{Z} e^{-\Phi(\theta)} \rho_{\text{prior}}(\theta)$

考虑：$c_1 = 10^{-2}$，1，计算θ的期望。

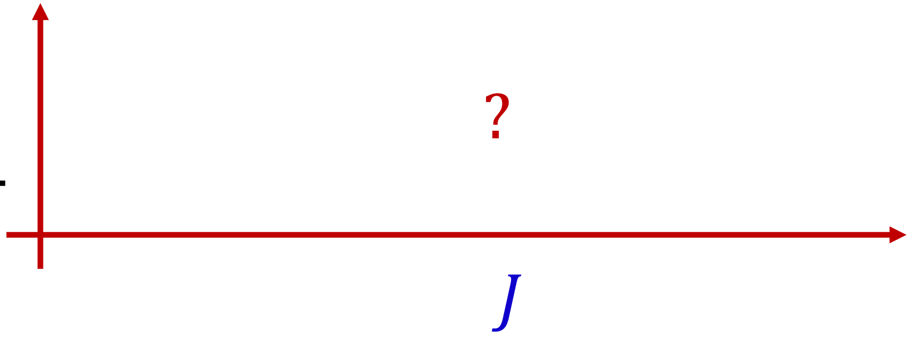
重要性采样

- 练习（Rosenbrock 函数）


$c_1 = 10^{-2}$

$c_1 = 1$

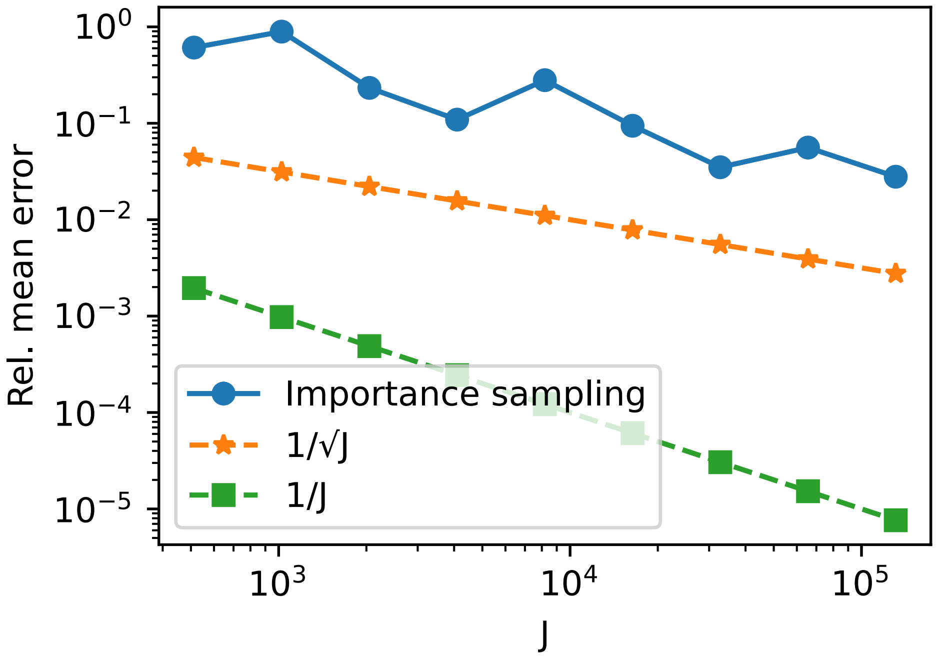
$$\frac{\mathrm{norm}\left({\rho^*}_{\mathrm{IS}}^{J}(\theta) - \mathbb{E}_{\rho^*}[\theta]\right)}{\mathrm{norm}\left(\mathbb{E}_{\rho^*}[\theta]\right)}$$



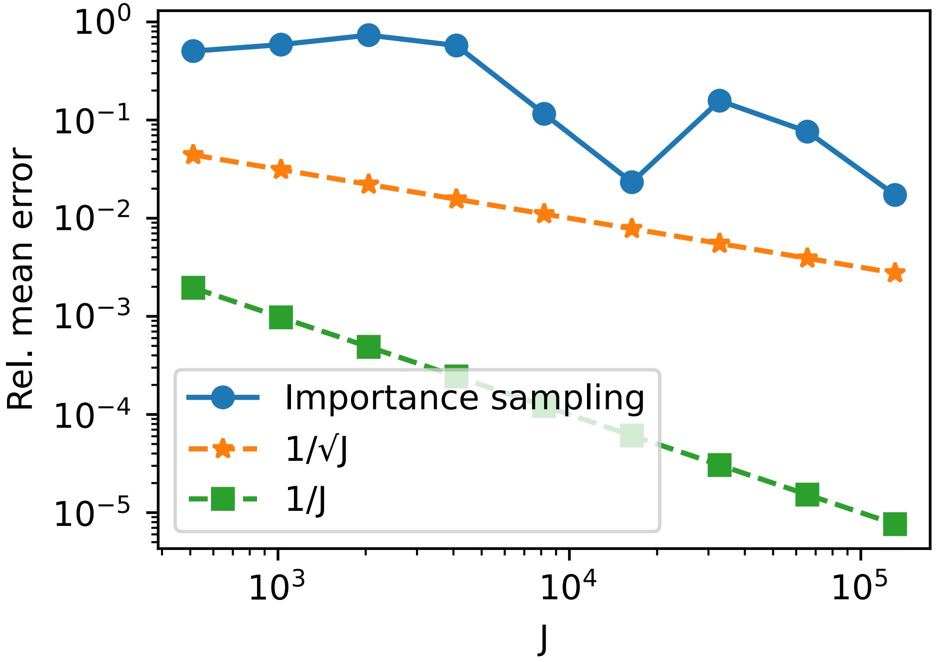
重要性采样

- ➢ 练习（Rosenbrock 函数）

$c_1 = 10^{-2}$



$c_1 = 1$



重要性采样

- 优点

    - 有收敛性保证

- 缺点

    - 收敛很慢，尤其是 $\chi^2[\rho^* \| \rho]$ 较大的时候

卡尔曼(Kalman)方法

- 贝叶斯反问题

$$y = \mathcal{G}(\theta) + \eta \qquad \eta \sim \rho_\eta \qquad \theta \sim \rho_{\text{prior}}$$

- 假设

高斯先验分布: $\rho_{\text{prior}}(\theta) = \mathcal{N}(\theta; r_0, \Sigma_0)$

高斯噪音: $\rho_\eta = \mathcal{N}(x; 0, \Sigma_\eta)$

- 贝叶斯法则

$$\rho(\theta|y) = \frac{\rho(\theta, y)}{\rho(y)} = \frac{\rho(y|\theta)\rho(\theta)}{\rho(y)}$$

$$\rho_{\text{prior}}(\theta) \rightarrow \rho(\theta, y) \rightarrow \rho_{\text{post}}(\theta)$$

卡尔曼方法

- 先验分布

$$\rho_{\text{prior}}(\theta) = \mathcal{N}(r_0, \Sigma_0)$$

- θ 和 $\mathcal{G}(\theta) + \eta$ 的联合分布

$$\rho(\theta, \mathcal{G}(\theta) + \eta) \approx \mathcal{N}\left(\begin{bmatrix} r_0 \\ \hat{y} \end{bmatrix}, \begin{bmatrix} \Sigma_0 & \hat{C}^{\theta y} \\ \hat{C}^{\theta y^T} & \hat{C}^{yy} \end{bmatrix}\right)$$

$$\hat{y} = \mathbb{E}[\mathcal{G}(\theta) + \eta] \quad \hat{C}^{\theta y} = \mathrm{Cov}[\theta, \mathcal{G}(\theta) + \eta] \quad \hat{C}^{yy} = \mathrm{Cov}[\mathcal{G}(\theta) + \eta]$$

- 后验分布（条件分布）

$$\rho(\theta | \mathcal{G}(\theta) + \eta = y) = \mathcal{N}(m, C)$$

$$m = r_0 + \hat{C}^{\theta y}(\hat{C}^{yy})^{-1}(y - \hat{y})$$

$$C = \Sigma_0 - \hat{C}^{\theta y}(\hat{C}^{yy})^{-1}\hat{C}^{\theta y^T}$$

卡尔曼方法

- 贝叶斯反问题

$$y = \mathcal{G}(\theta) + \eta \qquad \eta \sim \rho_\eta \qquad \theta \sim \rho_{\mathrm{prior}}$$

- 假设

高斯先验分布： $\rho_{\mathrm{prior}}(\theta) = \mathcal{N}(\theta; r_0, \Sigma_0)$

高斯噪音： $\rho_\eta = \mathcal{N}(x; 0, \Sigma_\eta)$

- 如何计算

$$\hat{y} = \mathbb{E}[\mathcal{G}(\theta) + \eta] \quad \hat{C}^{\theta y} = \mathrm{Cov}[\theta, \mathcal{G}(\theta) + \eta] \quad \hat{C}^{yy} = \mathrm{Cov}[\mathcal{G}(\theta) + \eta]$$

扩展(Extended)卡尔曼方法

- 泰勒展开线性化

$$\mathcal{G}(\theta) \approx \mathcal{G}(r_0) + \nabla\mathcal{G}(r_0)(\theta - r_0)$$

$$\rho_{\text{prior}}(\theta) = \mathcal{N}(r_0, \Sigma_0) \qquad\qquad \eta \sim \mathcal{N}(0, \Sigma_\eta)$$

$$\hat{y} = \mathbb{E}[\mathcal{G}(\theta) + \eta] \approx \mathcal{G}(r_0)$$

$$\hat{C}^{\theta y} = \mathrm{Cov}[\theta, \mathcal{G}(\theta) + \eta] \approx \Sigma_0 \nabla\mathcal{G}(r_0)^T$$

$$\hat{C}^{yy} = \mathrm{Cov}[\mathcal{G}(\theta) + \eta] \approx \nabla\mathcal{G}(r_0)^T \Sigma_0 \nabla\mathcal{G}(r_0) + \Sigma_\eta$$

扩展(Extended)卡尔曼方法

- 线性贝叶斯反问题

$$
\begin{aligned}
C_{\text{post}} &= \Sigma_0 - \Sigma_0 G^T \left(G \Sigma_0 G^T + \Sigma_\eta\right)^{-1} G \Sigma_0 \\
m_{\text{post}} &= r_0 - \Sigma_0 G^T \left(G \Sigma_0 G^T + \Sigma_\eta\right)^{-1} (G r_0 - y)
\end{aligned}
$$

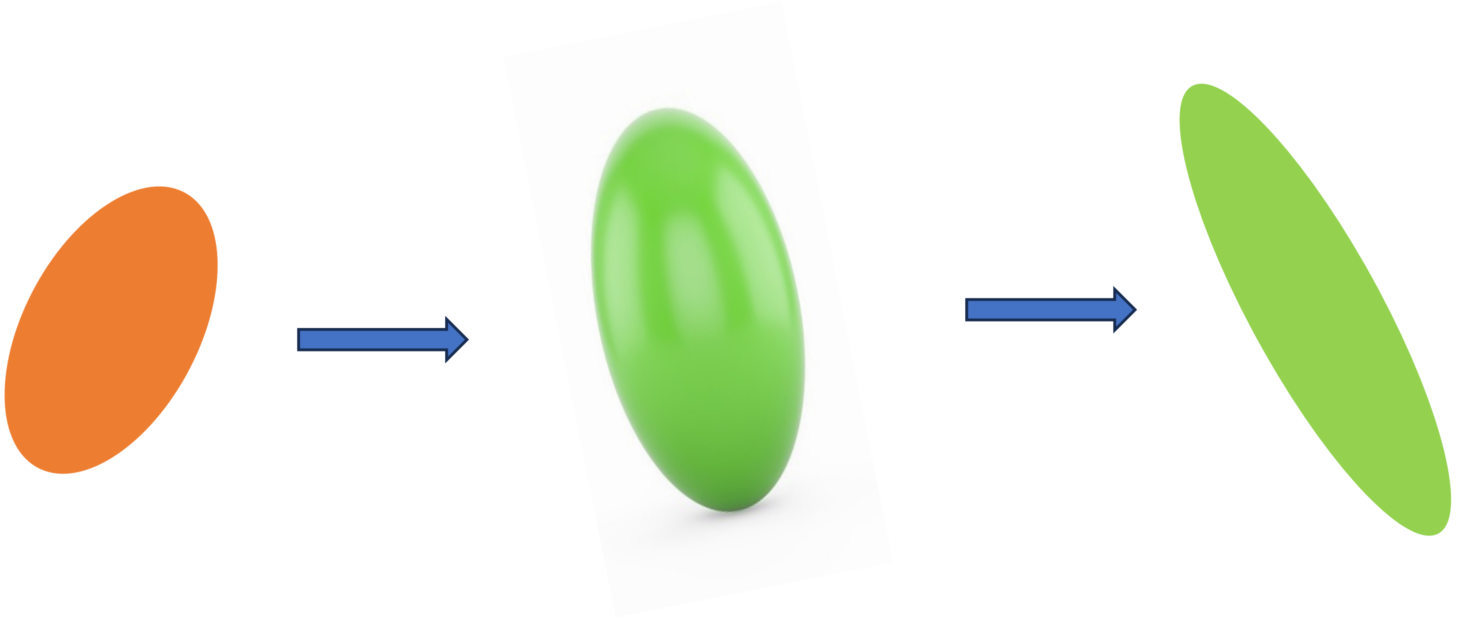
- 扩展卡尔曼方法

$$
\begin{aligned}
C_{\text{post}} &= \Sigma_0 - \Sigma_0 \nabla\mathcal{G}^T \left(\nabla\mathcal{G} \Sigma_0 \nabla\mathcal{G}^T + \Sigma_\eta\right)^{-1} \nabla\mathcal{G} \Sigma_0 \\
m_{\text{post}} &= r_0 - \Sigma_0 \nabla\mathcal{G}^T \left(\nabla\mathcal{G} \Sigma_0 \nabla\mathcal{G}^T + \Sigma_\eta\right)^{-1} (G(r_0) - y)
\end{aligned}
$$

扩展(Extended)卡尔曼方法

- 输运

$$\mathcal{T}: \mathcal{N}(r_0, \Sigma_0) \rightarrow \mathcal{N}(m, C)$$



$\rho_{\text{prior}}(\theta) \quad \rightarrow \quad \rho(\theta, y) \quad \rightarrow \quad \rho_{\text{post}}(\theta)$

无迹(Unscented)卡尔曼方法

无迹变换

对于高斯分布 $\theta \sim \mathcal{N}(m, C) \in R^{N_\theta}$, 我们选取$2N_\theta + 1$ 个σ点， $\theta^0 = m$，对 $j = 1,2\cdots, N_\theta$

$$\theta^j = m + c_j[\sqrt{C}]_j \qquad \theta^{j+N_\theta} = m - c_j[\sqrt{C}]_j$$

其中 $[\sqrt{C}]_j$ 是C的Cholesky分解的第 j个列向量，那么

$$\mathbb{E}[\mathcal{G}(\theta)] \approx \hat{\mathbb{E}}\,[\mathcal{G}(\theta)] := \sum_{i=0}^{2N_\theta} W_i^m \mathcal{G}(\theta^i)$$

$$\mathrm{Cov}[\mathcal{G}_1(\theta), \mathcal{G}_2(\theta)] \approx$$

$$\sum_{i=0}^{2N_\theta} W_i^c (\mathcal{G}_1(\theta^i) - \hat{\mathbb{E}}[\mathcal{G}_1(\theta)])(\mathcal{G}_2(\theta^i) - \hat{\mathbb{E}}[\mathcal{G}_2(\theta)])^T$$

参数： c_i， W_i^m, W_i^c

无迹(Unscented)卡尔曼方法

- 无迹变换

对于高斯分布 $\theta \sim \mathcal{N}(m, C) \in R^{N_\theta}$ ，我们有

$$\mathcal{G}(\theta) = \mathcal{G}(m) + \nabla\mathcal{G}\delta\theta + \frac{1}{2}\nabla^2\mathcal{G}\delta\theta\otimes\delta\theta + \frac{1}{6}\nabla^3\mathcal{G}\delta\theta\otimes\delta\theta\otimes\delta\theta$$
$$+\mathcal{O}(\delta\theta^4)$$

$$\mathbb{E}[\mathcal{G}(\theta)] = \mathcal{G}(m) + \frac{1}{2}\nabla^2\mathcal{G}C + \mathcal{O}(\|C\|^2)$$

$$\mathrm{Cov}[\mathcal{G}_1(\theta), \mathcal{G}_2(\theta)] = \nabla\mathcal{G}_1 C\nabla\mathcal{G}_2^{\mathrm{T}} + \mathcal{O}(\|C\|^2)$$

无迹(Unscented)卡尔曼方法

无迹变换

当 $1 \le j \le N_\theta$

$$W_j^m = W_{j+N_\theta}^m \qquad \sum_{i=0}^{2N_\theta} W_i^m = 1 \qquad W_j^c = W_{j+N_\theta}^c = \frac{1}{2c_j^2}$$

我们有：

$$\sum_{i=0}^{2N_\theta} W_i^m \mathcal{G}(\theta^i) = \mathcal{G}(m) +$$

$$+ \sum_{j=1}^{N_\theta} c_j^2 W_j^m \nabla^2 \mathcal{G}[\sqrt{C}]_j \otimes [\sqrt{C}]_j + \mathcal{O}(\|C\|^2)$$

$$\sum_{i=0}^{2N_\theta} W_i^c (\mathcal{G}_1(\theta^i) - \widehat{\mathbb{E}}[\mathcal{G}_1(\theta)])(\mathcal{G}_2(\theta^i) - \widehat{\mathbb{E}}[\mathcal{G}_2(\theta)])^T$$

$$= \mathrm{Cov}[\mathcal{G}_1(\theta), \mathcal{G}_2(\theta)] + \mathcal{O}(\|C\|^2)$$

无迹(Unscented)卡尔曼方法

无迹变换

当 $1 \le j \le N_\theta$

$$W_j^m = W_{j+N_\theta}^m \qquad \sum_{i=0}^{2N_\theta} W_i^m = 1 \qquad W_j^c = W_{j+N_\theta}^c = \frac{1}{2c_j^2}$$

期望近似误差：

$$\sum_{j=1}^{N_\theta} c_j^2 W_j^m \nabla^2 \mathcal{G}[\sqrt{C}]_j \otimes [\sqrt{C}]_j - \frac{1}{2} \nabla^2 \mathcal{G} C + \mathcal{O}(\|C\|^2)$$

方差近似误差：$\mathcal{O}(\|C\|^2)$

我们选取： $c_j = a$，$W_j^c = \frac{1}{2a^2}$

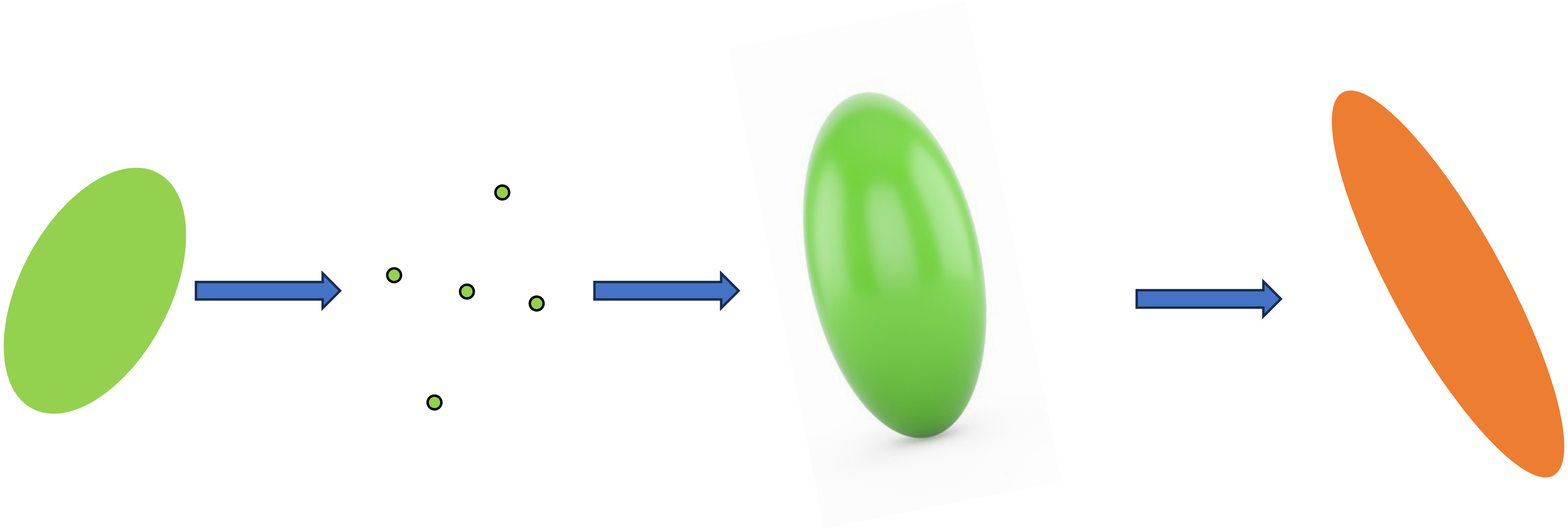
二阶精度：$W_j^m = \frac{1}{2c_j^2}$, $W_0^m = 1 - \sum_j \frac{1}{c_j^2} = 1 - \frac{N_\theta}{a^2}$

一阶精度：$W_0^m = 1$， $W_j^m = 0$

无迹(Unscented)卡尔曼方法

- 输运

$$\mathcal{T}:\mathcal{N}(r_0,\Sigma_0)\rightarrow\mathcal{N}(m,C)$$



$$\rho_{\mathrm{prior}}(\theta)\quad\rightarrow\quad\sigma-\text{点}\qquad\rightarrow\qquad\rho(\theta,y)\qquad\rightarrow\qquad\rho_{\mathrm{post}}(\theta)$$

集合卡尔曼方法

- 蒙特卡洛方法

$$\{\theta_j\} \sim \rho_{\text{prior}}(\theta) = \mathcal{N}(r_0, \Sigma_0) \qquad \eta \sim \mathcal{N}(0, \Sigma_\eta)$$

$$\hat{y} = \frac{1}{J}\sum_{j=1}^{J} y^j \qquad y^j = \mathcal{G}(\theta^j)$$

$$\hat{y} = \mathbb{E}[\mathcal{G}(\theta) + \eta] \approx \hat{y}$$

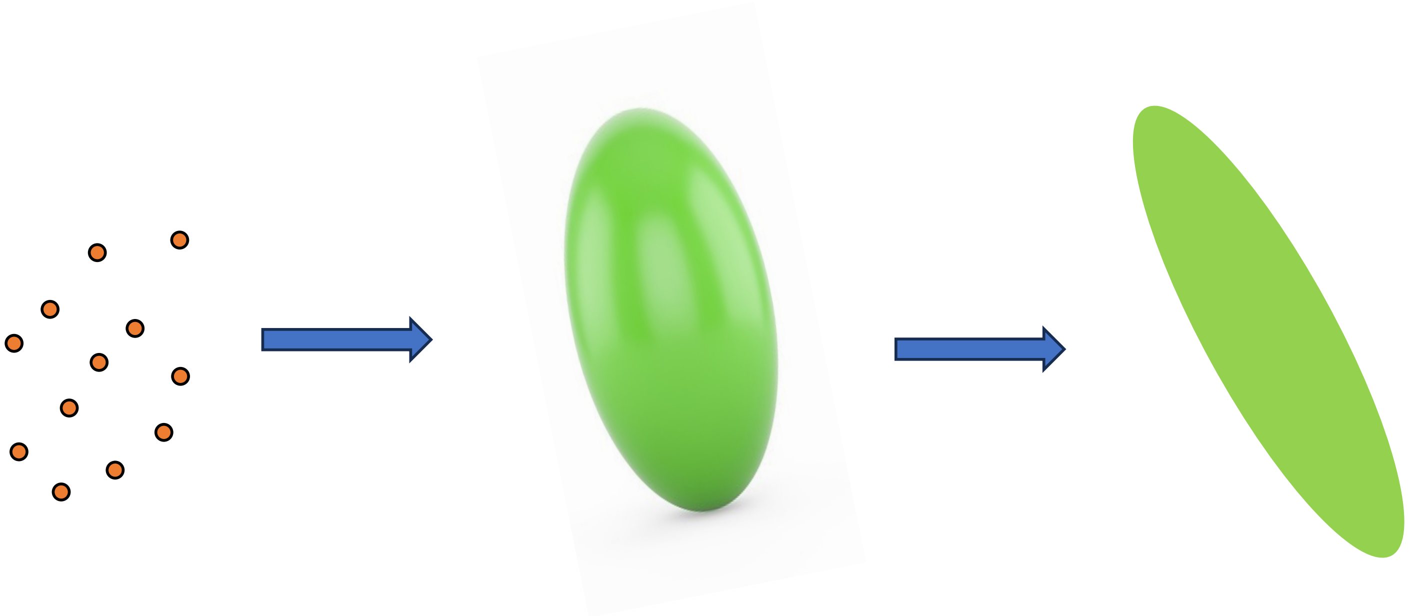
$$\hat{C}^{\theta y} = \mathrm{Cov}[\theta, \mathcal{G}(\theta) + \eta] = \frac{1}{J-1}\sum_{j=1}^{J}(\theta^j - r_0)(y^j - \hat{y})^T$$

$$\hat{C}^{yy} = \mathrm{Cov}[\mathcal{G}(\theta) + \eta] = \frac{1}{J-1}\sum_{j=1}^{J}(y^j - \hat{y})(y^j - \hat{y})^T + \Sigma_\eta$$

集合(Ensemble)卡尔曼方法

- 输运

$$\mathcal{T}:\{\theta^j\} \rightarrow \mathcal{N}(m, C)$$

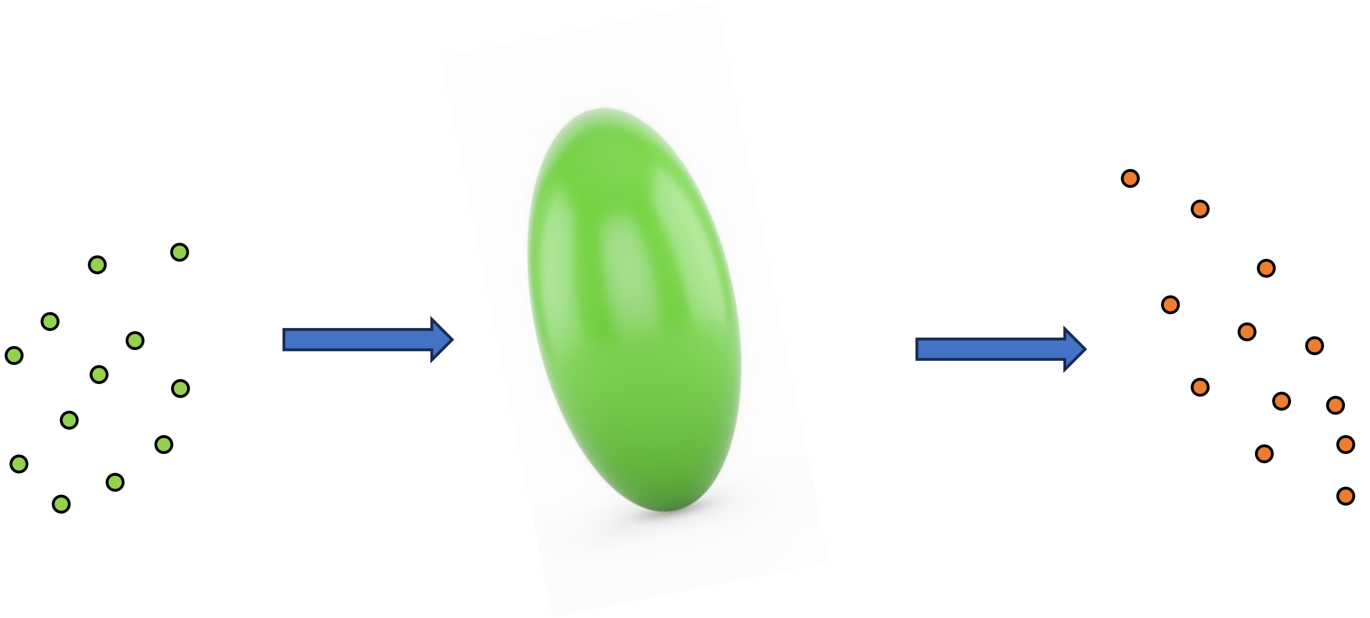


$$\rho_{\text{prior}}(\theta) \quad \rightarrow \quad \rho(\theta, y) \quad \rightarrow \quad \rho_{\text{post}}(\theta)$$

集合(Ensemble)卡尔曼方法

- 输运

$$\mathcal{T}:\{\theta^j\} \to \{\mathcal{T}\theta^j\}$$



$$\rho_{\text{prior}}(\theta) \quad \to \quad \rho(\theta, y) \quad \to \quad \rho_{\text{post}}(\theta)$$

卡尔曼方法

- 练习：计算$\mathcal{G}(\theta)$ 分布

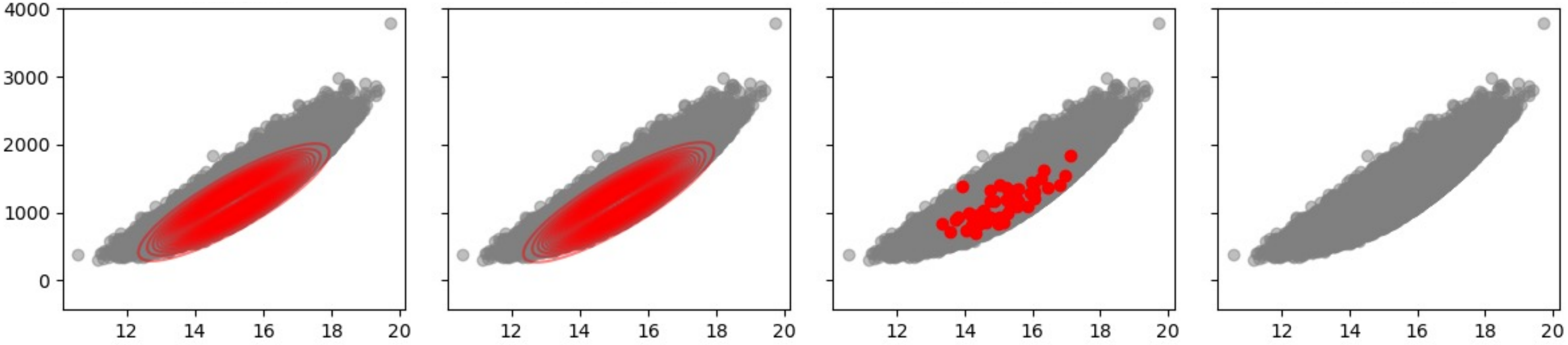
$$\theta \sim \mathcal{N}\left(\begin{bmatrix}10\\10\end{bmatrix}, \begin{bmatrix}1 & 0\\0 & 1\end{bmatrix}\right) \qquad \mathcal{G}(\theta) = \begin{bmatrix} 1+\sqrt{\theta_{(1)}^2+\theta_{(2)}^2} \\ \exp\frac{\theta_{(1)}}{2}+\theta_{(2)}^3 \end{bmatrix}$$

扩展卡尔曼方法
$J = 1$

无迹卡尔曼方法
$J = 5$

集合卡尔曼方法
$J = 50$

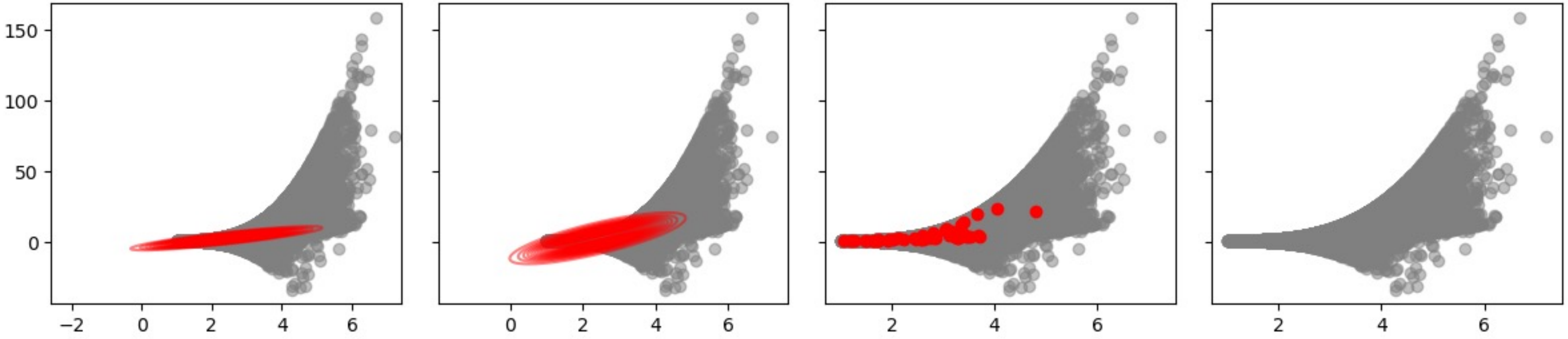
标准
$J = 10^5$



卡尔曼方法

- 练习：计算$\mathcal{G}(\theta)$ 分布

$$\theta \sim \mathcal{N}\left(\begin{bmatrix}1\\1\end{bmatrix}, \begin{bmatrix}1 & 0\\0 & 1\end{bmatrix}\right) \qquad \mathcal{G}(\theta) = \begin{bmatrix} 1+\sqrt{\theta_{(1)}^2+\theta_{(2)}^2} \\ \exp\frac{\theta_{(1)}}{2}+\theta_{(2)}^3 \end{bmatrix}$$



扩展卡尔曼方法 $J=1$

无迹卡尔曼方法 $J=5$

集合卡尔曼方法 $J=50$

标准 $J=10^5$

卡尔曼方法

- 优点

  - 高效、实用

- 缺点

  - 没有收敛性
  - 只适用于后验分布接近高斯分布的情形

更通用的近似

标准化流(Normalizing flow)

- 有未知归一化常数的目标分布

$$\rho^*(\theta) = \frac{1}{Z} e^{-\Phi_R(\theta)}$$

- 标准化流

基于神经网络的映射 $\mathcal{T}_{NN}\colon R^{N_\theta} \to R^{N_\theta}$

$$\{\theta^j\} \sim \rho_{\text{prior}} \qquad \to \qquad \{\mathcal{T}_{NN}\left(\theta^j\right)\} \sim \rho^*$$

神经网络$\mathcal{T}_{NN}$ ：参数化的（非线性）映射，能自动计算关于参数或输入的导数

标准化流

- 诱导测度(Pushforward)

$$\mathcal{T}: \theta \rightarrow \tilde{\theta} = \mathcal{T}(\theta)$$

$$\mathcal{T}: \rho \rightarrow \tilde{\rho} = \mathcal{T}\#\rho \qquad \rho(\theta) = \tilde{\rho}\big(\mathcal{T}(\theta)\big)|\nabla_\theta \mathcal{T}(\theta)|$$

$$\int_{\theta\in A} \rho(\theta)d\theta = \int_{\tilde{\theta}\in\mathcal{T}(A)} \tilde{\rho}(\tilde{\theta})d\tilde{\theta}$$

$$= \int_{\theta\in A} \tilde{\rho}\big(\mathcal{T}(\theta)\big)|\nabla_\theta \mathcal{T}(\theta)|d\theta$$

$$\mathcal{T}^{-1}: \tilde{\theta} \rightarrow \theta = \mathcal{T}^{-1}(\theta)$$

$$\mathcal{T}^{-1}: \tilde{\rho} \rightarrow \rho = \mathcal{T}^{-1}\#\tilde{\rho} \qquad \tilde{\rho}(\tilde{\theta}) = \rho\left(\mathcal{T}^{-1}(\tilde{\theta})\right)|\nabla_{\tilde{\theta}} \mathcal{T}^{-1}(\tilde{\theta})|$$

标准化流

- KL-散度

$$\mathrm{KL}[\rho \parallel \rho^*] = \int \rho \log\left(\frac{\rho}{\rho^*}\right) d\theta$$

- $\mathrm{KL}[\rho^* \parallel \rho^*] = 0$
- $\mathrm{KL}[\rho \parallel \rho^*] \geq 0$
- $\mathrm{KL}(\rho \parallel Z\rho^*) = \mathrm{KL}(\rho \parallel \rho^*) - \log(Z)$

标准化流

- 标准化流

训练神经网络 $\mathcal{T}_{NN}$： $\{\theta^j\} \sim \rho_{\text{prior}} \quad \rightarrow \quad \{\mathcal{T}_{NN}(\theta^j)\} \sim \rho^*$

$$\min_{NN} \text{KL}\left[\mathcal{T}_{NN}\#\rho_{\text{prior}} \parallel \rho^*\right]$$

$$= \min_{NN} \int \mathcal{T}_{NN}\#\rho_{\text{prior}} \left(\log\left(\mathcal{T}_{NN}\#\rho_{\text{prior}}\right) + \Phi_R(\theta)\right) d\theta$$

采样 ：$\tilde{\theta}^j = \mathcal{T}_{NN}(\theta^j) \sim \mathcal{T}_{NN}\#\rho_{\text{prior}}$

计算目标函数，更新神经网络

$$\int \mathcal{T}_{NN}\#\rho_{\text{prior}} \left(\log\left(\mathcal{T}_{NN}\#\rho_{\text{prior}}\right) + \Phi_R(\theta)\right) d\theta \approx$$

$$\frac{1}{J}\sum_{j=1}^{J} \log\left(\rho_{\text{prior}}(\theta^j)\left|\nabla_{\tilde{\theta}}\mathcal{T}_{NN}^{-1}(\tilde{\theta}^j)\right|\right) + \Phi_R(\tilde{\theta}^j)$$

标准化流

- 标准化流

  神经网络设计，计算 $\nabla_\theta \mathcal{T}_{NN}^{-1}(\tilde{\theta}^j)$

  映射可逆、容易计算关于输入的导数

- 实值非体积保持模型

  $\mathcal{T}_{NN} = f_K \circ f_{K-1} \circ \cdots \circ f_1$

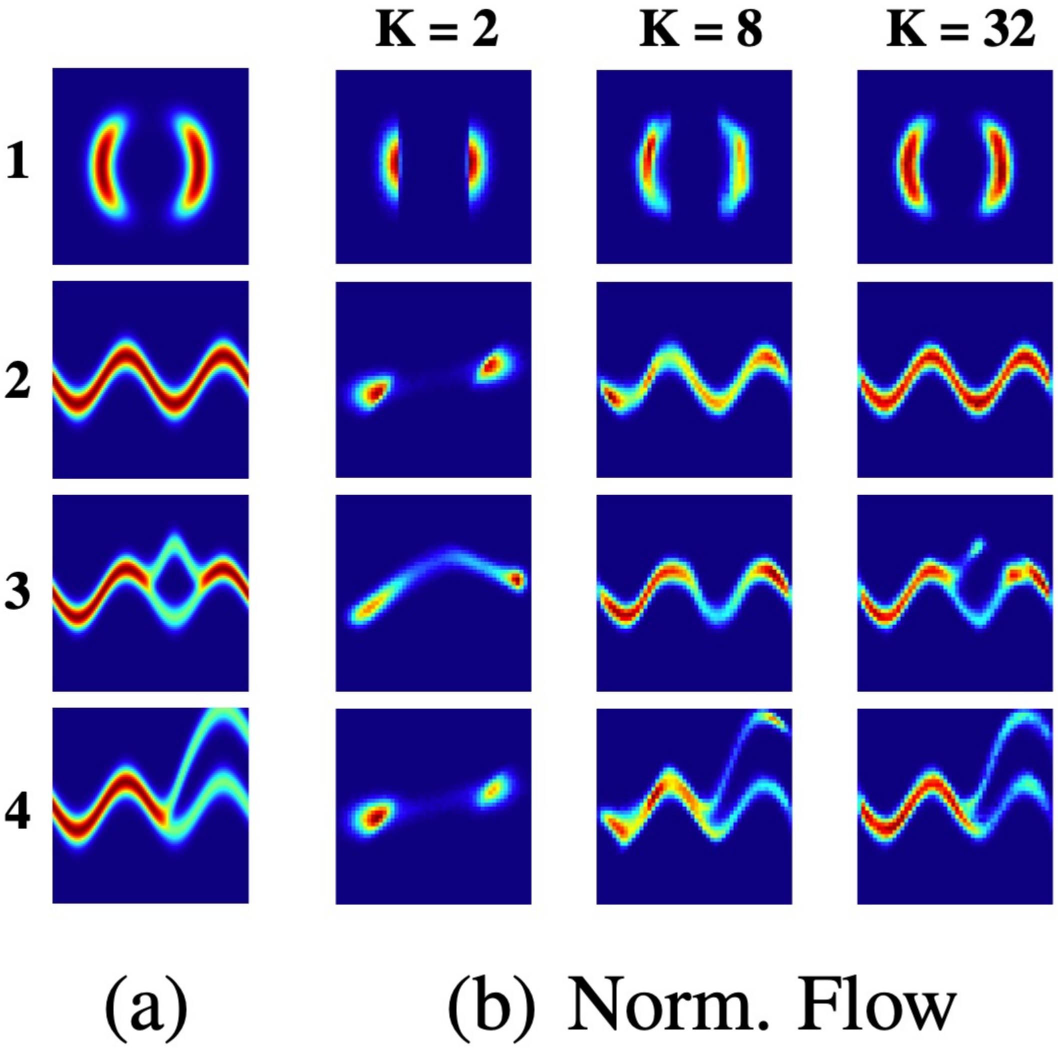
  f_i包含仿射耦合层：

$$y_{1:d} = x_{1:d}$$
$$y_{d+1:N_\theta} = x_{d+1:N_\theta} \odot \exp(s(x_{1:d})) + t(x_{1:d})$$

  随机打乱维度：$y_{1:N_\theta} \to y_{i_1:i_{N_\theta}}$

标准化流

- 标准化流



生成模型 (Generative model)

- 有未知归一化常数的目标分布

$$\rho^*(\theta) = \frac{1}{Z} e^{-\Phi_R(\theta)}$$

- 生成模型

已知：$\{\theta_j^{\text{data}}\} \sim \rho^*(\theta)$

生成：$\{\theta_j\} \sim \rho^*(\theta)$

标准化流

- 标准化流

训练神经网络 $\mathcal{T}_{NN}$: $\{\theta^j\} \sim \rho_{\text{prior}} \quad \to \quad \{\mathcal{T}_{NN}(\theta^j)\} \sim \rho^*$

$$\min_{NN} \text{KL}\left[\rho^* \parallel \mathcal{T}_{NN}\#\rho_{\text{prior}}\right]$$

$$= \max_{NN} \frac{1}{n}\sum_{j=1}^{n} \log\left(\mathcal{T}_{NN}\#\rho_{\text{prior}}(\theta_j^{\text{data}})\right)$$

$$= \max_{NN} \frac{1}{n}\sum_{j=1}^{n} \log\left(\rho_{\text{prior}}\left(\mathcal{T}_{NN}^{-1}\theta_j^{\text{data}}\right)\left|\nabla_{\tilde{\theta}}\mathcal{T}_{NN}^{-1}\left(\theta_j^{\text{data}}\right)\right|\right)$$

标准化流

- 标准化流

训练神经网络 $\mathcal{T}_{NN}$：$\{\theta^j\} \sim \rho_{\text{prior}} \quad \rightarrow \quad \{\mathcal{T}_{NN}(\theta^j)\} \sim \rho^*$

$$\min_{NN} \mathrm{KL}\left[\rho^* \parallel \mathcal{T}_{NN}\#\rho_{\text{prior}}\right]$$

$$= \max_{NN} \frac{1}{n}\sum_{j=1}^{n} \log\left(\mathcal{T}_{NN}\#\rho_{\text{prior}}(\theta_j^{\text{data}})\right)$$

$$= \max_{NN} \frac{1}{n}\sum_{j=1}^{n} \log\left(\rho_{\text{prior}}\left(\mathcal{T}_{NN}^{-1}\theta_j^{\text{data}}\right)\left|\nabla_{\tilde{\theta}}\mathcal{T}_{NN}^{-1}\left(\theta_j^{\text{data}}\right)\right|\right)$$

扩展阅读

重要性采样

迭代的思路: Beskos, Alexandros, et al. "Sequential Monte Carlo methods for Bayesian elliptic inverse problems." *Statistics and Computing* 25 (2015): 727–737.

卡尔曼方法

第四种卡尔曼方法: Arasaratnam, Ienkaran, and Simon Haykin. "Cubature kalman filters." *IEEE Transactions on automatic control* 54.6 (2009): 1254-1269.

迭代的思路: Huang, Daniel Zhengyu, et al. "Efficient derivative-free Bayesian inference for large-scale inverse problems." *Inverse Problems* 38.12 (2022): 125006.

基于非线性映射的输运方法

标准化流: Rezende, Danilo, and Shakir Mohamed. "Variational inference with normalizing flows." *International conference on machine learning*. PMLR, 2015.
下三角映射: Marzouk, Youssef, et al. ”An introduction to Sampling via measure transport: " *Handbook of uncertainty quantification* 1 (2016): 2.