

非线性贝叶斯反问题

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

- 贝叶斯反问题
- 贝叶斯采样、推理
- 贝叶斯后验分布



贝叶斯反问题

➤ 贝叶斯反问题

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

➤ 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

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

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

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



贝叶斯反问题

➤ 优化方法

最大似然估计：

$$\operatorname{argmin} \frac{1}{2} \|\Sigma_{\eta}^{-\frac{1}{2}} (y - \mathcal{G}(\theta))\|^2$$

最大后验估计：

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

非线性最小二乘问题：高斯-牛顿方法，Levenberg Marquardt 方法等

➤ 贝叶斯方法

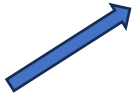

$$\text{采样：} \quad \rho_{\text{post}}(\theta; y) = \frac{1}{Z} e^{-\Phi_R(\theta, y)}$$



贝叶斯采样、推理

➤ 有未知归一化常数的目标分布

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

未知   已知

- 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

➤ 参数化(parametric)方法

高斯近似：

$$\rho^*(\theta) \approx \mathcal{N}(\theta; m, C)$$

混合高斯近似：

$$\rho^*(\theta) \approx \sum_{j=1}^J w_j \mathcal{N}(\theta; m_j, C_j) \quad \sum_{j=1}^J w_j = 1$$

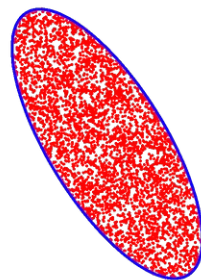
.....

➤ 非参数化(nonparametric)方法

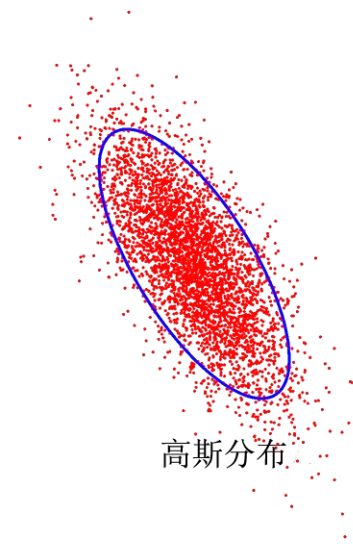
粒子近似 $J \gg 1$ ：

$$\rho^*(\theta) \approx \{\theta_j\}_{j=1}^J$$

$$\rho^*(\theta) \approx \frac{1}{J} \sum_{j=1}^J \delta(\theta - \theta_j)$$



均匀分布



高斯分布



贝叶斯采样、推理

► 输运的方法（直接近似的方法）

暴力网格搜索：假设 $N_\theta = 2$,

$$\theta_{i,j} = \left[-L + \frac{2(i-1)}{N-1}L, -L + \frac{2(j-1)}{N-1}L\right] \quad Z = \sum \rho^*(\theta_{i,j})$$

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

- 重要性采样
- 卡尔曼方法
- 标准化流方法
-



贝叶斯采样、推理

➤ 马氏链蒙特卡洛(Markov chain Monte Carlo)方法

暴力网格搜索：假设 $N_\theta = 2$,

$$\theta_{i,j} = \left[-L + \frac{2(i-1)}{N-1}L, -L + \frac{2(j-1)}{N-1}L\right] \quad Z = \sum \rho(\theta_{i,j})$$

随机游走： $\theta_n \rightarrow \theta_{n+1}$ 概率分布 $\psi_I(\theta_n, \theta_{n+1})$

不变分布是目标分布

$$\rho_{\text{post}}(\theta) = \int \psi_I(\theta', \theta) \rho_{\text{post}}(\theta') d\theta'$$

遍历性质

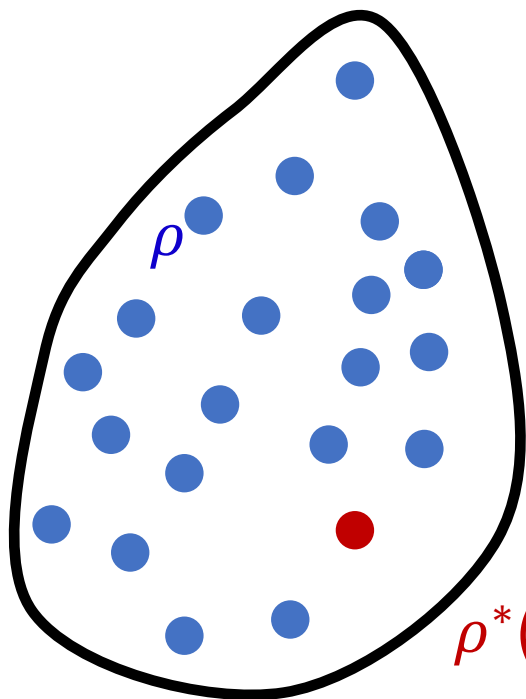
$$\{\theta_n\}_{n \geq N} \sim \rho_{\text{post}}(\theta)$$



贝叶斯采样、推理

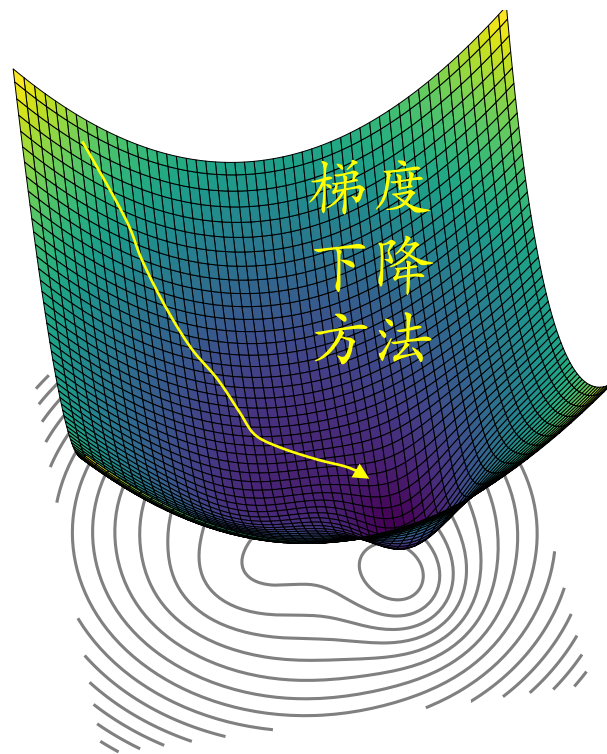
变分推理 (Variational Inference)

$$\text{minimize}_{\rho} \mathcal{E}(\rho; \rho^*)$$



概率密度空间 \mathcal{P}

度量 $M(\rho)$
距离 $\mathcal{D}(\rho_A, \rho_B)$





贝叶斯后验分布

➤ Rosenbrock 函数 ($N_\theta = 2$)

$$y = G(\theta) + \eta$$

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

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: $\rho_\eta = \mathcal{N}(x; 0, c_2 \begin{bmatrix} \frac{1}{10^2} & \\ & 1 \end{bmatrix})$

考虑:

$$c_1 = 1, 10^{-1}, 10^{-2}, 10^{-3} \quad c_2 = 1$$

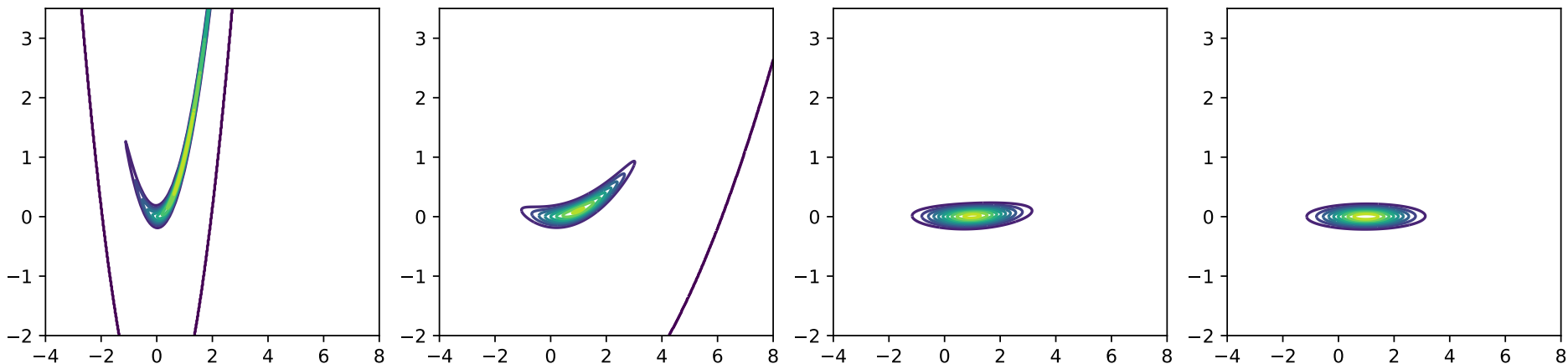
$$c_2 = 1, 10^{-1}, 10^{-2}, 10^{-3} \quad c_1 = 1$$



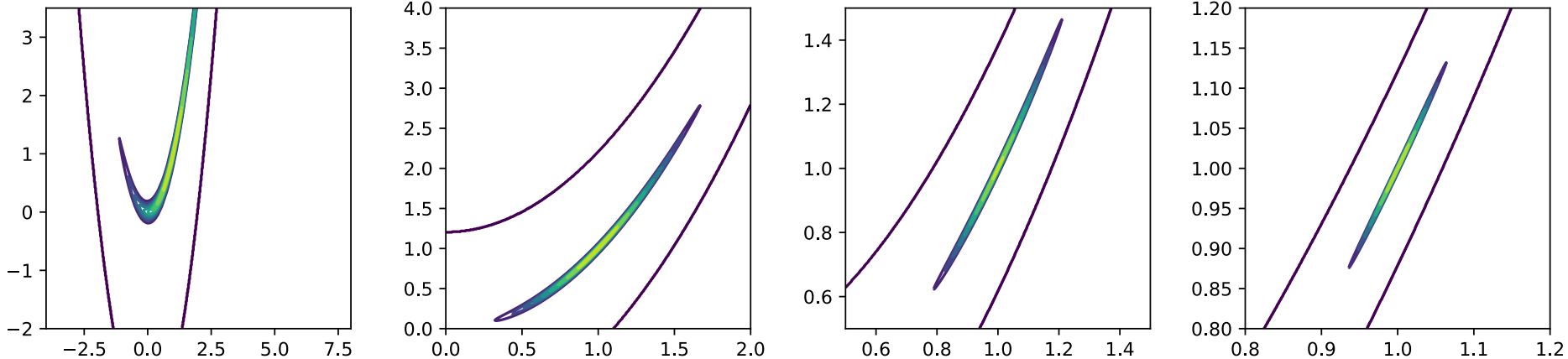
贝叶斯后验分布

➤ Rosenbrock 函数 ($N_\theta = 2$)

$c_1 = 1 \quad c_2 = 1$ $c_1 = 10^{-1} \quad c_2 = 1$ $c_1 = 10^{-2} \quad c_2 = 1$ $c_1 = 10^{-3} \quad c_2 = 1$



$c_1 = 1 \quad c_2 = 1$ $c_1 = 1 \quad c_2 = 10^{-1}$ $c_1 = 1 \quad c_2 = 10^{-2}$ $c_1 = 1 \quad c_2 = 10^{-3}$





贝叶斯后验分布

► Bernstein Von Mises 理论

对于适定反问题, $N_y = N_\theta$

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

我们假设噪音满足 $\eta \sim \mathcal{N}(x; 0, \gamma^2 \Sigma_\eta)$, 先验分布满足 $\theta \sim \mathcal{N}(\theta; r_0, \Sigma_0)$ 。

在一定条件下, 当 $\gamma \rightarrow 0$

$$\rho_{\text{post}}(\theta) \rightarrow \mathcal{N}(\theta; \theta^\dagger, \gamma^2 \left(\nabla_\theta \mathcal{G}(\theta^\dagger)^T \Sigma_\eta^{-1} \nabla_\theta \mathcal{G}(\theta^\dagger) \right)^{-1})$$

其中 θ^\dagger 是 $(y - \mathcal{G}(\theta)) \Sigma_\eta^{-1} (y - \mathcal{G}(\theta))$ 的唯一极小值。



贝叶斯后验分布

➤ 多峰函数 ($N_\theta = 2$)

$$y = G(\theta) + \eta$$

$$G(\theta) = (\theta_1 - \theta_2)^2 \quad y = 4$$

高斯先验分布: $\rho_{\text{prior}}(\theta) = \mathcal{N}(\theta; \begin{bmatrix} c \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \\ & 1 \end{bmatrix})$

高斯噪音: $\rho_\eta = \mathcal{N}(x; 0, 1)$

考虑:

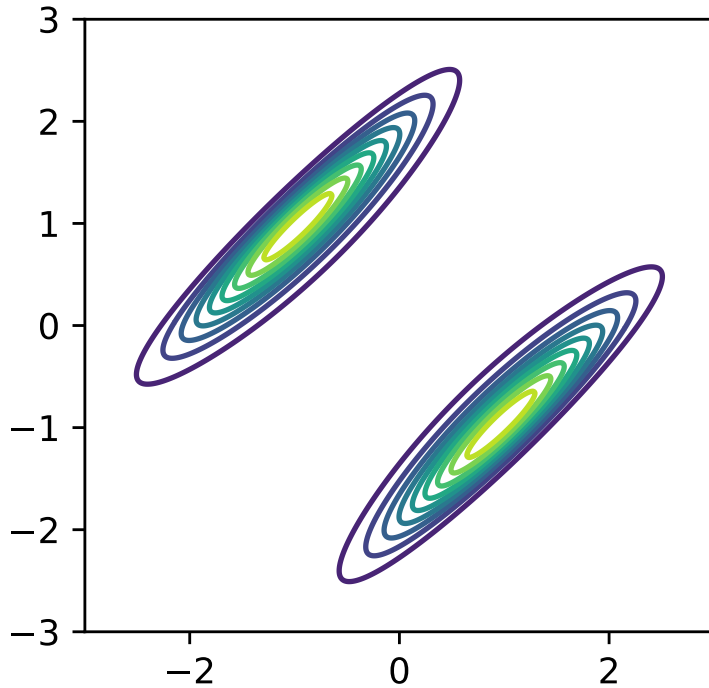
$$c = 0, 0.5$$



贝叶斯后验分布

➤ 多峰函数 ($N_\theta = 2$)

$c = 0$



$c = 0.5$

