Sinkhorn Algorithm as a Special Case of Stochastic Mirror Descent

Konstantin Mishchenko,
KAUST

جامعة الملك عبدالله
للعلوم والتقنية
King Abdullah University of
Science and Technology

Problem 1

Let $X \in \mathbb{R}_{++}^{n \times n}$

Find vectors $u, v \in \mathbb{R}_{+}^{n}$ such that

$W = \mathrm{diag}(u) X \mathrm{diag}(v)$ is doubly stochastic

Problem 1

Let $X \in \mathbb{R}_{++}^{n \times n}$

Find vectors $u, v \in \mathbb{R}_{+}^{n}$ such that

$$W = \mathrm{diag}(u) X \mathrm{diag}(v) \text{ is doubly stochastic}$$

$$\sum_{i=1}^{n} W_{ij} = 1 \quad \text{for any } j$$

$$W_{ij} \geq 0$$

$$\sum_{j=1}^{n} W_{ij} = 1 \quad \text{for any } i$$

Problem 2

$$\min_{X \in \mathbb{R}^{n \times n}} \sum_{i,j=1}^{n} C_{ij} X_{ij} \quad \text{s.t.} \quad X\mathbf{1} = \mathbf{1}, X^\top \mathbf{1} = \mathbf{1}, X \geq 0$$

doubly stochastic

Problem 2

$$\min_{X \in \mathbb{R}^{n \times n}} \sum_{i,j=1}^{n} C_{ij} X_{ij} \quad \text{s.t. } X\mathbf{1} = \mathbf{1}, X^\top \mathbf{1} = \mathbf{1}, X \geq 0$$

$$\min_{X \in \mathbb{R}^{n \times n}} \sum_{i,j=1}^{n} (C_{ij} X_{ij} + \gamma X_{ij} \log X_{ij}) \quad \text{s.t. } X\mathbf{1} = \mathbf{1}, X^\top \mathbf{1} = \mathbf{1}$$

Problem 2

$$\min_{X \in \mathbb{R}^{n \times n}} \sum_{i,j=1}^{n} C_{ij} X_{ij} \quad \text{s.t.} \;\; X\mathbf{1} = \mathbf{1}, X^\top \mathbf{1} = \mathbf{1}, X \geq 0$$

$$\min_{X \in \mathbb{R}^{n \times n}} \sum_{i,j=1}^{n} (C_{ij} X_{ij} + \gamma X_{ij} \log X_{ij}) \quad \text{s.t.} \;\; X\mathbf{1} = \mathbf{1}, X^\top \mathbf{1} = \mathbf{1}$$

$$\min_{X \in \mathbb{R}^{n \times n}} \mathcal{KL}(X || X^0) \quad \text{s.t.} \;\; X\mathbf{1} = \mathbf{1}, X^\top \mathbf{1} = \mathbf{1}$$

$$\text{where } X^0 \stackrel{\text{def}}{=} \exp\left(-\frac{C}{\gamma}\right) \text{ coordinate-wise}$$

$$\mathcal{KL}(X || X^0) \stackrel{\text{def}}{=} \sum_{i,j=1}^{n} \left(X_{ij} \log \frac{X_{ij}}{X^0_{ij}} - X_{ij} + X^0_{ij} \right)$$

Problem 2

$$\min_{X \in \mathbb{R}^{n \times n}} \sum_{i,j=1}^{n} C_{ij} X_{ij} \quad \text{s.t.} \;\; X\mathbf{1} = \mathbf{1}, X^\top \mathbf{1} = \mathbf{1}, X \geq 0$$

$$\min_{X \in \mathbb{R}^{n \times n}} \sum_{i,j=1}^{n} \left(C_{ij} X_{ij} + \gamma X_{ij} \log X_{ij} \right) \quad \text{s.t.} \;\; X\mathbf{1} = \mathbf{1}, X^\top \mathbf{1} = \mathbf{1}$$

$$\min_{X \in \mathbb{R}^{n \times n}} \mathcal{KL}(X \| X^0) \quad \text{s.t.} \;\; X\mathbf{1} = \mathbf{1}, X^\top \mathbf{1} = \mathbf{1}$$

$$\text{where } X^0 \stackrel{\text{def}}{=} \exp\left(-\frac{C}{\gamma} \right) \text{ coordinate-wise}$$

$$\mathcal{KL}(X \| X^0) \stackrel{\text{def}}{=} \sum_{i,j=1}^{n} \left(X_{ij} \log \frac{X_{ij}}{X_{ij}^0} - X_{ij} + X_{ij}^0 \right)$$

Problem 3

$$x = \text{vec}(X) \in \mathbb{R}^d, \; d = n \cdot n$$

$$X\mathbf{1} = \mathbf{1}$$

$$a_1 = (1, 1, \ldots, 1, 0, \ldots, 0)$$

$$\bullet \quad \bullet \quad \bullet$$

$$a_i = (0, \ldots, 0, \underbrace{1, \ldots, 1}_{i\text{−th block}}, 0, \ldots, 0)$$

Problem 3

$$x = \text{vec}(X) \in \mathbb{R}^d, \ d = n \cdot n$$

$$\boldsymbol{X1} = \boldsymbol{1}$$

$$a_1 = (1, 1, \ldots, 1, 0, \ldots, 0)$$

• • •

$$a_i = (0, \ldots, 0, \underbrace{1, \ldots, 1}_{i\text{−th block}}, 0, \ldots, 0)$$

same for $X^\top \boldsymbol{1} = \boldsymbol{1}$

$$A = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_m \end{pmatrix}$$

Problem 3

$$x = \mathrm{vec}(X) \in \mathbb{R}^d,\ d = n \cdot n$$

$$X\mathbf{1} = \mathbf{1}$$

$$a_1 = (1, 1, \ldots, 1, 0, \ldots, 0)$$

$$\bullet \quad \bullet \quad \bullet$$

$$a_i = (0, \ldots, 0, \underbrace{1, \ldots, 1}_{i\text{–th block}}, 0, \ldots, 0)$$

same for $X^\top \mathbf{1} = \mathbf{1}$

$$\min_x \left\{ \mathcal{KL}(Ax || b) = \sum_i^{2n} \left(\langle a_i, x \rangle \log \frac{\langle a_i, x \rangle}{b_i} - \langle a_i, x \rangle + b_i \right) \right\}$$

One method to solve them all

$$W = \mathrm{diag}(u) X \mathrm{diag}(v) \text{ is doubly stochastic}$$

$$\text{Loop} \quad \begin{cases} 1. \ \text{Normalize all rows} \\ 2. \ \text{Normalize all columns} \end{cases}$$

One method to solve them all

$$\text{Loop}\quad \begin{cases} \text{1. Normalize all rows} \\ \text{2. Normalize all columns} \end{cases}$$

$$\min_{X \in \mathbb{R}^{n \times n}} \mathcal{KL}(X || X^0) \quad \text{s.t. } X\mathbf{1} = \mathbf{1}, X^\top \mathbf{1} = \mathbf{1}$$

$$\text{Loop} \begin{cases} X^{k+1} = \arg\min_{X\mathbf{1}=\mathbf{1}} \{\mathcal{KL}(X||X^k)\} \\ X^{k+2} = \arg\min_{X^\top\mathbf{1}=\mathbf{1}} \{\mathcal{KL}(X||X^{k+1})\} \end{cases}$$

One method to solve them all

$$\text{Loop} \begin{cases} 1. \ \text{Normalize all rows} \\ 2. \ \text{Normalize all columns} \end{cases}$$

$$\text{Loop} \begin{cases} X^{k+1} = \arg\min_{X\mathbf{1}=\mathbf{1}} \{\mathcal{KL}(X||X^k)\} \\ X^{k+2} = \arg\min_{X^\top\mathbf{1}=\mathbf{1}} \{\mathcal{KL}(X||X^{k+1})\} \end{cases}$$

$$\min_x \left\{ \mathcal{KL}(Ax||b) = \sum_i^{2n} \left(\langle a_i, x\rangle \log \frac{\langle a_i, x\rangle}{b_i} - \langle a_i, x\rangle + b_i \right) \right\}$$

$$f_1(x) = \mathcal{KL}(A_{\text{rows}}x||\mathbf{1}), \ f_2(x) = \mathcal{KL}(A_{\text{cols}}x||\mathbf{1})$$

$$\log(x^{k+1}) = \log(x^k) - \eta\nabla f_i(x^k), \ i \sim U(\{1,2\})$$

Intuition

$$\min_x \frac{1}{2}\|x - x^0\|^2 \text{ s.t. } Ax = b$$

Kaczmarz Algorithm

$$x^{k+1} = \Pi_{\langle a_i, x\rangle = b_i}(x^k),\ \ i \sim U(\{1, \ldots, m\})$$

Intuition

$$\min_x \frac{1}{2}\|x - x^0\|^2 \text{ s.t. } Ax = b$$

Kaczmarz Algorithm

$$x^{k+1} = \Pi_{\langle a_i, x\rangle = b_i}(x^k), \;\; i \sim U(\{1, \ldots, m\})$$

$$x^{k+1} = x^k + \frac{b_i - \langle a_i, x\rangle}{\|a_i\|^2} a_i$$

$$x^k \in x^0 + \text{Range}(A^\top)$$

$$x^{k+1} = x^k - \eta \nabla f_i(x^k), \quad f_i(x) = \frac{1}{2\|a_i\|^2}(\langle a_i, x\rangle - b_i)^2$$

Sinkhorn Algorithm

$$\min_x \left\{ \mathcal{KL}(Ax||b) = \sum_i^{2n} \left(\langle a_i, x \rangle \log \frac{\langle a_i, x \rangle}{b_i} - \langle a_i, x \rangle + b_i \right) \right\}$$

$$f_1(x) = \mathcal{KL}(A_{\text{rows}} x || \mathbf{1}), \;\; f_2(x) = \mathcal{KL}(A_{\text{cols}} x || \mathbf{1})$$

$$\log(x^{k+1}) = \log(x^k) - \eta \nabla f_i(x^k), \;\; i \sim U(\{1, 2\})$$

Sinkhorn Algorithm

$$\min_x \left\{ \mathcal{KL}(Ax||b) = \sum_i^{2n} \left(\langle a_i, x \rangle \log \frac{\langle a_i, x \rangle}{b_i} - \langle a_i, x \rangle + b_i \right) \right\}$$

$$f_1(x) = \mathcal{KL}(A_{\text{rows}} x || \mathbf{1}), \;\; f_2(x) = \mathcal{KL}(A_{\text{cols}} x || \mathbf{1})$$

$$\log(x^{k+1}) = \log(x^k) - \eta \nabla f_i(x^k), \;\; i \sim U(\{1, 2\})$$

$$\log x^k \in \log x^0 + \text{Range}(A^\top)$$

$$A_{\text{rows}}, A_{\text{cols}} \in \{0, 1\}^{n \times n}$$

$$X^{k+1} = \text{diag}(u^k) X^0 \text{diag}(v^k)$$

Reference

R. Sinkhorn, Diagonal equivalence to matrices with prescribed row and column sums, 1967

K. M., Sinkhorn Algorithm as a Special Case of Stochastic Mirror Descent, 2019
arXiv:1909.06918

NeurIPS workshop on Optimal Transport and its Application to Machine Learning

Canada, December 2019

https://sites.google.com/view/otml2019/