Lecture 11. Wed. 4/20/2022

Today: Stability of Sinkhorn's algorithm. Log-domain method

Given $a \in \mathcal{P}_m$, $b \in \mathcal{P}_n$, $C \in \mathbb{R}^{m\times n}$, $C \geq 0$, $0 < \varepsilon < 1$.

Entropy regularized OT: $\min_{P \in \mathcal{A}(a,b)} E_\varepsilon[P]$.

$$E_\varepsilon[P] = \langle P, C\rangle + \varepsilon \langle P(\log P - 1), \mathbb{1}_{m\times n}\rangle.$$

If $a > 0$, $b > 0$, then using the method of Lagrange multipliers, we have that the unique minimizer is $P_\varepsilon = [P_{ij}] \in \mathcal{A}(a,b)$: $P_{ij} = e^{-(C_{ij} - f_i - g_j)/\varepsilon} = e^{f_i/\varepsilon} e^{-C_{ij}/\varepsilon} e^{g_j/\varepsilon}$.

Sinkhorn's algorithm: Let $K = e^{-C_{ij}/\varepsilon}$. Find u, v s.t $(u_i K_{ij} v_j) \in \mathcal{A}(a,b)$: $\quad u^{(k)} = \dfrac{a}{Kv^{(k-1)}}$, $v^{(k)} = \dfrac{b}{K^T u^{(k)}}$ $(k = 1, 2, \cdots)$

Potential issue: $\varepsilon > 0$ small $\Rightarrow$ instability in computing C_{ij}/ε, and hence u, v.

Work on the dual problem.

The dual problem of a constrained convex optimization problem.

Primal: $\quad \min_x f(x)$

subject to: $h(x) = 0$

The Lagrangian function

$$L(x, \lambda) = f(x) - \lambda \cdot h(x)$$

It's dual function is

$$g(\lambda) := \inf_x L(x, \lambda).$$

Dual: $\quad \max_\lambda g(\lambda)$.

A special case: $h(x) = Ax - c$.

$$g(\lambda) = \inf_x \left[f(x) - \lambda \cdot (Ax - c) \right]$$

$$= - \sup_x \left[-f(x) + \lambda \cdot (Ax - c) \right]$$

$$= - \sup_x \left[A^T\lambda \cdot x - f(x) \right] + c^T x$$

$$= -f^*(A^T\lambda) + c^T x.$$

$f^*(\xi) = \sup_x (\xi \cdot x - f(x))$ — Legendre transform of f.

Back to $F_\varepsilon[P]$.

$$\mathcal{L}(P, \lambda) = \mathcal{L}(P, f, g) = \sum_{i,j} C_{ij} P_{ij} + \varepsilon \sum_{ij} P_{ij} (\log P_{ij} - 1)$$

$$+ \sum_{i,j} f_i (P_{ij} - a_i) + \sum_{ij} g_j \cdot (P_{ij} - b_j)$$

$\frac{\partial \mathcal{L}}{\partial P_{kl}} = 0 \quad C_{kl} - f_k - g_l + \varepsilon \log P_{kl} = 0.$

$$P_{kl} = e^{-(C_{kl} - f_k - g_l)/\varepsilon} =: e^{q_{kl}}$$

$$q_{kl} = -(C_{kl} - f_k - g_l)/\varepsilon.$$

$$L(f, g) = L_\varepsilon(f, g) = \sum_{ij} C_{ij} e^{q_{ij}} + \varepsilon \sum_{ij} e^{q_{ij}} (q_{ij} - 1)$$

$$+ \sum_{ij} f_i (e^{q_{ij}} - a_i) + \sum_{ij} g_j \cdot (e^{q_{ij}} - b_j)$$

$$= \sum_i f_i a_i + \sum_j g_j \cdot b_j - \varepsilon \sum_{ij} e^{q_{ij}}$$

$$= \langle f, a \rangle + \langle g, b \rangle - \varepsilon \langle e^{f/\varepsilon}, e^{-C/\varepsilon} e^{g/\varepsilon} \rangle$$

Or: $f(x) = x(\log x - 1) \Longrightarrow f^*(\xi) = e^\xi$.

The dual problem (dual to the entropy-regularized OT problem):

$$\max_{(f,g) \in \mathbb{R}^m \times \mathbb{R}^n} L_\varepsilon(f, g).$$

<u>Proposition</u> $\exists!\,(f_\varepsilon, g_\varepsilon) \in \mathbb{R}^m \times \mathbb{R}^n$ s.t.

$$(f_\varepsilon, g_\varepsilon) = \underset{(f,g)\in\mathbb{R}^m\times\mathbb{R}^n}{\arg\max}\, L_\varepsilon(f,g).$$

Moreover, for $u_\varepsilon = e^{f_\varepsilon/\varepsilon}$, $v_\varepsilon = e^{g_\varepsilon/\varepsilon}$, and $K_\varepsilon = e^{-C/\varepsilon}$,

$$P_\varepsilon := \mathrm{diag}(u_\varepsilon) K_\varepsilon \,\mathrm{diag}(v_\varepsilon) = \underset{P\in\mathcal{A}(a,b)}{\arg\min}\, E_\varepsilon[P].$$

<u>Proof</u> L_ε is smooth, convex, and bounded above. As $(f,g)\to\infty$, $L_\varepsilon(f,\infty)\to-\infty$. Thus, L_ε has a unique maximizer $(f_\varepsilon, g_\varepsilon)$. It is also the unique critical point of L_ε. Hence, $\partial L/\partial f_{\varepsilon i} = 0$, $\partial L/\partial g_{\varepsilon j} = 0$ $\forall i,j$.

$$\frac{\partial L_\varepsilon}{\partial f_{\varepsilon i}} = 0 \Rightarrow a_i - \sum_j e^{f_i/\varepsilon} e^{-C_{ij}/\varepsilon} e^{g_j/\varepsilon} = 0 \Rightarrow a_i = \sum_j P_{\varepsilon ij}\ \forall i$$

$$\frac{\partial L_\varepsilon}{\partial g_{\varepsilon j}} = 0 \Rightarrow b_j = \sum_i P_{\varepsilon ij}\ \forall j.$$ So, $P_\varepsilon \in \mathcal{A}(a,b)$. But, the unique minimizer of $E_\varepsilon[\cdot]$ over $\mathcal{A}(a,b)$ is exactly $\mathrm{diag}(u) K_\varepsilon \,\mathrm{diag}(v) \in \mathcal{A}(a,b)$ for some $u \in \mathbb{R}^m_+$ and $v \in \mathbb{R}^n_+$. <u>QED</u>.

<u>A dual Sinkhorn algorithm</u>. The maximizer $(f_\varepsilon, g_\varepsilon)$ of $L_\varepsilon(f,g)$ satisfies

$$\nabla_f L_\varepsilon(f,g) = a - e^{f/\varepsilon} \odot K_\varepsilon e^{g/\varepsilon} = 0$$

$$\nabla_g L_\varepsilon(f,g) = b - e^{g/\varepsilon} \odot K_\varepsilon^T e^{f/\varepsilon} = 0$$

Given $g^{(0)}$:
$$f^{(k)} = \varepsilon \log a - \varepsilon \log\left(K_\varepsilon e^{g^{(k-1)}/\varepsilon}\right),$$
$$g^{(k)} = \varepsilon \log b - \varepsilon \log\left(K_\varepsilon^T e^{f^{(k)}/\varepsilon}\right),\quad k = 1, 2, \cdots$$

Primal vs. Dual: $(f^{(k)}, g^{(k)}) = \varepsilon(\log u^{(k)}, \log v^{(k)})$.

But the dual prob. still has $1/\varepsilon$.

observe

$$-\varepsilon \log\left(K_\varepsilon e^{f_i^{(k)}/\varepsilon}\right) = -\varepsilon \log\left(\sum_j K_{\varepsilon ij}\, e^{f_i^{(k)}/\varepsilon}\right)$$

$$= -\varepsilon \log\left(\sum_j e^{-(C_{ij}-f_i^{(k)})/\varepsilon}\right) = -\varepsilon \log\left(\sum_j e^{-((C^T e_i)_j - f^{(k)})/\varepsilon}\right)$$

$$= -\varepsilon \log \sum_j \left(e^{-z_j/\varepsilon}\right) = \min\nolimits_\varepsilon z \quad (\text{soft-min})$$

This is familiar, e.g. in statistical mechanics.

Let $z = (z_1, \cdots, z_N)^T \in \mathbb{R}^N$. Define

$$\underset{\uparrow\ \text{free energy}}{F_\varepsilon(z)} = -\varepsilon \log \underbrace{\sum_{i=1}^N e^{-z_i/\varepsilon}}_{Z_\varepsilon:\ \text{partition function}}, \qquad \varepsilon = T: \text{temperature}.$$

<u>Proposition</u> Given $z = (z_1, \cdots, z_N) \in \mathbb{R}^N$, we have

$$\min z := \min_{1 \le i \le N} z_i = \lim_{\varepsilon \to 0^+} F_\varepsilon(z) = \lim_{\varepsilon \to 0^+}\left[-\varepsilon \log\left(\sum_i e^{-z_i/\varepsilon}\right)\right].$$

<u>Proof</u> Exercise. <u>QED</u>

Given $g^{(0)}$:
$$f^{(k)} = \varepsilon \log a - \min\nolimits_\varepsilon\left(C e_j - g^{(k-1)}\right),$$
$$g^{(k)} = \varepsilon \log b - \min\nolimits_\varepsilon\left(C^T e_i - f^{(k)}\right), \quad k = 1, 2, \cdots.$$

Now, $\min_\varepsilon z = F_\varepsilon(z) = -\varepsilon \log\left(\sum_i e^{-z_i/\varepsilon}\right)$

$$= -\varepsilon \log\left(\sum_{i=1}^N e^{-(z_i - \min z)/\varepsilon} \cdot e^{-\min z/\varepsilon}\right)$$

$$= \min z - \varepsilon \log \left(\sum_{i=1}^{N} e^{-(z_i - \min z)/\varepsilon} \right)$$

Note: ⊙ $z_i - \min z = 0$ or strictly positive.

If $z_i > \min z$, then $e^{-(z_i - \min z)/\varepsilon}$ remains uniformly bounded for all $\varepsilon > 0$. So, stable!

⊙ $\min z$ is unknown, but can be approximated by an iteration process.

The log-domain method

$$f^{(k)} = \min_\varepsilon^{row} \left(C - f^{(k-1)} \oplus g^{(k-1)} \right) - f^{(k-1)} + \varepsilon \log a,$$

$$g^{(k)} = \min_\varepsilon^{col} \left(C - f^{(k)} \oplus g^{(k-1)} \right) - g^{(k-1)} + \varepsilon \log b.$$

stable now! But price to pay: evaluate log $m \cdot n$ times for each k. — Slower!