Research on Optimal Transport Algorithms

Su Xun

笠井研究室 修士 2 年

2022 年 4 月 13 日

Outline

- Background of OT
- Recent Progress for Speeding up OT
- Experiments
- Ideas and Future Plan

Background of OT

Optimal Transport

$$W(\alpha,\beta) := \min_{\mathbf{P}\in\mathbb{R}_+^{n\times n}} \langle \mathbf{C}, \mathbf{P}\rangle$$

$$\mathbf{P}\mathbb{1} = \alpha, \mathbf{P}^T\mathbb{1} = \beta, \mathbf{P}_{ij} > 0$$

- Applications on GAN, Retrieving information, Domain adaptation and so on.

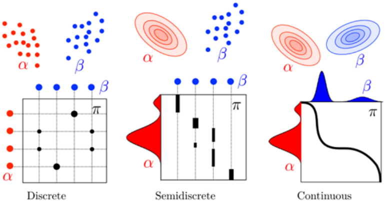

Figure: Different forms of Optimal Transport

Regularized OT

$$W^{\eta}(\alpha,\beta) := \min_{\mathbf{P}\in\mathbb{R}_{+}^{n\times n}} \langle \mathbf{C},\mathbf{P}\rangle - \eta\mathbf{H}(\mathbf{P})$$

$$\mathbf{P}\mathbb{1} = \alpha, \mathbf{P}^{T}\mathbb{1} = \beta, \mathbf{P}_{ij} > 0$$

- For Entropic OT: $\mathbf{H}(\mathbf{P}) := -\sum_{i,j}^{N} \mathbf{P}_{ij} \ln \mathbf{P}_{ij}$
- We hope to get a solution satisfied the constrain condition and $\langle \mathbf{C},\hat{\mathbf{P}}\rangle < \langle \mathbf{C},\mathbf{P}^{*}\rangle + \varepsilon$. Original OT could be solved by Linear Programming (LP), which has a complexity $O(n^3)$
- Adding a regularizer could speed up the computation process to $O(n^2)$

Dual Problem of OT

Dual Form (Entropic OT)

$$
\begin{aligned}
\mathcal{L}(\mathbf{P}, u, v) :=& \langle \mathbf{C}, \mathbf{P} \rangle - \eta H(\mathbf{P}) + \langle u, \alpha \rangle + \langle v, \beta \rangle \\
& - \langle u, \mathbf{P}\mathbf{1} \rangle - \langle v, \mathbf{P}^\top \mathbf{1} \rangle \\
\min_{\mathbf{P}} \mathcal{L}(\mathbf{P}, u, v) =& \min \{ \langle u, \alpha \rangle + \langle v, \beta \rangle \\
& + \max_{\mathbf{P}} (-\langle \mathbf{C}, \mathbf{P} \rangle + \eta H(\mathbf{P}) - \langle u, \mathbf{P}\mathbf{1} \rangle - \langle v, \mathbf{P}^\top \mathbf{1})\} \\
=& - \min \varphi(u, v) \\
=& - \min_{u,v} (\eta \sum_{i,j=1}^{n} e^{-\frac{\mathbf{C}_{ij} - u_i - v_j}{\eta} - 1}) - \langle u, \alpha \rangle - \langle v, \beta \rangle
\end{aligned}
$$

Various Optimization Method for OT

Dual Form

$$\min_{u,v}(\eta \sum_{i,j=1}^{n} e^{-\frac{\mathbf{C}_{ij}-u_i-v_j}{\eta}-1}) - \langle u, \alpha \rangle - \langle v, \beta \rangle$$

- Block Coordinate Descent (BCD)
  There is no direct closed form for the dual form, but exists the closed form for partial gradient. From $\partial\varphi_u = 0$ and $\partial\varphi_v = 0$ we have $u = \frac{\alpha}{Kv}, v = \frac{\beta}{K^T u}, K = -e^{\frac{\mathbf{C}}{\eta}}$.
  We could minimize the problem by alternatively descent along u or v, which is so called Sinkhorn Algorithm.
  The best convergence speed analysis is $O(\frac{n^2}{\varepsilon^2})$, but it is believed that the best convergence rate is $O(\frac{n^2}{\varepsilon})$

Various Optimization Method for OT

Semi-Dual Form (Entropic OT)

$$\max_{v \in \mathbb{R}^J} \varphi_\varepsilon(v) = \sum_{i \in I} \bar{h}_\varepsilon\left(x_i, v\right) \boldsymbol{\alpha}_i, \quad \text{where}$$

$$\bar{h}_\varepsilon(x, v) = \sum_{j \in J} v_j \boldsymbol{\beta}_j + \varepsilon \log\left(\sum_{j \in J} \exp\left(\frac{v_j - c\left(x, y_j\right)}{\varepsilon}\right) \boldsymbol{\beta}_j\right) - \varepsilon$$

- Optimization on Semi-Dual
  The semi-dual problem is smooth and has a sum form.

  It could be regarded as many sum problem. Stochastic methods is suitable for the optimization problem (SGD SAG).[?] These stochastic methods don't have a theoretical complexity analysis.

Various Optimization Method for OT

Dual Form

$$\min_{u,v}(\eta \sum_{i,j=1}^{n} e^{-\frac{\mathbf{C}_{ij}-u_i-v_j}{\eta}-1}) - \langle u, \alpha \rangle - \langle v, \beta \rangle$$

- Gradient Descent (GD)
  We can still use traditional gradient method, they use a full gradient to optimize the dual problem and run really slow compared with Sinkhorn.
  Gradient method would could be easily applied to all kinds of regularizers.

Different Regularizers

- Entorpic Regularizer
  For Entropic OT: $\mathbf{H}(\mathbf{P}) := -\sum_{i,j}^{N} P_{ij}(\ln P_{ij} - 1)$

- L2 Regularizer
  For l2 OT: $\mathbf{H}(\mathbf{P}) := -\sum_{i,j}^{N} \|P_{ij}\|^2$

- Group-Lasso Regularizer (used in Domain adaptation with label)
  For Group-Lasso OT: $\mathbf{H}(\mathbf{P}) := -\sum \|\boldsymbol{Pij}\|^2 + \eta \sum_{G \in \mathcal{G}} \|\boldsymbol{Pij}_G\|$

- Different Regularizers has its own benefits, L2 regularizer could get a sparse solution, Group-Lasso Regularizer is used on Domain Adaption field but don't have a closed form iteration step for Sinkhorn.

Revised Speed Up Method

Algorithm 1 OT by CD

procedure OT's Dual Form F(u, v)
  $u^{k+1} = \text{argmin}\,?$
  $v^{k+1} = \text{argmin}\,?$ ▷ Multiple choice
  Stop
  Output u, v
end procedure

- Sinkhorn(Block CD) $u^{k+1} = \text{argmin}_u(f(u^k, v^k))$
- Greekhorn(Greedy CD) $u_i^{k+1} = \text{argmin}_{u_i}(f(u^k, v^k))$ and i is the steepest direction.
- Stochastic Sinkhorn(Random CD) $u_i^{k+1} = \text{argmin}_{u_i}(f(u^k, v^k))$

Revised Speed Up Method

Gradient Descent

$$x_{k+1} = x_k - \epsilon \nabla f(x_k)$$

Accelerated GD

$$x_{k+1} = y_k - \epsilon \nabla f(y_k)$$
$$y_k = x_k + \frac{k-1}{k+2}(x_k - x_{k-1})$$

Figure: Nesterov Accelerated

- It could speed up the convergence rate from $O(\frac{1}{\epsilon})$ to $O(\frac{1}{\sqrt{\epsilon}})$

Revised Speed Up Method

Primal Average

$$\begin{aligned} \min_{\mathbf{x}\in\mathbb{R}^t} \quad & f(\mathbf{x}), \\ \text{s.t.} \quad & \boldsymbol{B}\mathbf{x} + \mathbf{b} = \mathbf{0}, \\ & g_i(\mathbf{x}) \leq 0, \quad i = 1, \cdots, m. \end{aligned}$$

$$\hat{\mathbf{X}}^K = \frac{\sum_{k=K_0}^{K} \frac{\mathrm{x}^*\left(\mathrm{v}^k\right)}{\theta_k}}{\sum_{k=K_0}^{K} \frac{1}{\theta_k}}$$

- Regularized OT is a primal-dual problem.
  We optimize on dual form and judge the error on primal form, which would harm the convergence speed quadratically. the primal average skill could counteract the problem but require a storage of solution for every iteration,[**?**] used the skill and combined it with Nestrov Accelerate Method to achieve $O(\frac{n^{2.5}}{\epsilon})$

Experiments on CD method

Experiments on Gauss Distributions

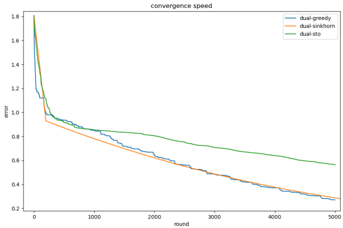

Figure: Dimension m = n = 100

- 1. Random generated Gauss distributions, per round is an update with $O(n)$ operations. error $:= \|\hat{P}\mathbb{1} - \mu\|_1 + \|\hat{P}^T\mathbb{1} - \nu\|_1$
- 2. There is no huge difference on Greekhorn and Sinkhorn, Stochastic method performs not well.

Experiments on CD method

Experiments on Real MNIST pics

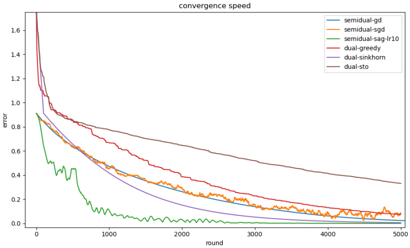

Figure: dimension m = 192, n =162

- 1. In real MNIST, SAG for semi-dual performs well.
- 2. Greekhorn can still not outperform Sinkhorn.
- 3. Stochastic Sinkhorn performs not well.

Experiments on CD method and other regulazor

Experiments on synthetic MNIST pics

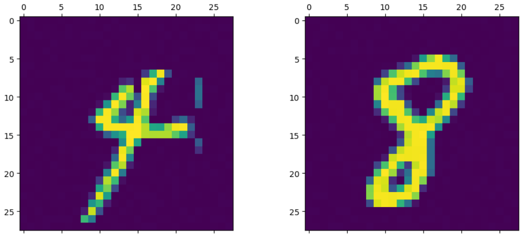

Figure: Synthetic MNIST pics

- 1. Synthetic MNIST pics has a background noise, and the pic is dense.

Experiments on CD method

Experiments on synthetic MNIST pics

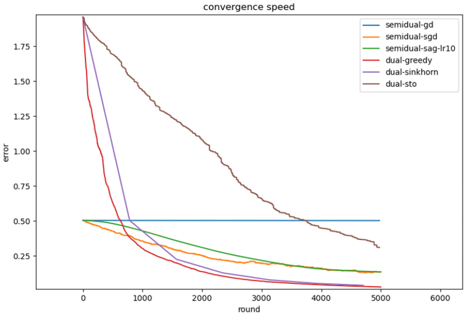

Figure: dimension $m = n = 784$

- 1. Stochastic Semi-dual method and gradient method is slow.
- 2. Greekhorn outperform the Sinkhorn.

Ideas and Questions

- My tools can not get a good convergence analysis for stochastic method on semi-dual form, and its advantage has been covered by Greenkhorn method.
- Traditional ideas on how to speed up OT has been well researched and didn't shake the position of Sinkhorn too much. I think the gradient method should focus on other regularizer which don't have a closed form solution for Sinkhorn method.
- It has been proved that Sinkhorn could achieve $O(\frac{n^2}{\varepsilon})$ complexity[**?**], I wonder that wether gradient method for UOT could achieve the same complexity.
- Using gradient method directly can not alleviate the main drawback of Sinkhorn on Entropic OT with a small ϵ

Future Plan

- Do research on the other regularizers of OT problems or composite problems and try to design useful optimization method for specific application.
- Test the gradient based method for UOT problems and analysis its theoretical convergence.

References I

ご清聴ありがとうございました.

Thank you for listening.