Dynamic Screening Method on the Unbalanced Optimal Transport Problem

Anonymous Author
Anonymous Institution

Abstract

This paper build the the latest dynamic screening framework on the Unbalanced Optimal Transport (UOT) problem. Recently, researchers connected the UOT problem with Lasso problem, which encourage us to combine the widely used technique in Lasso problem, Dynamic Screening, onto the UOT problem. We demonstrate the effectiveness of the screening method for UOT problem and propose improvements based on the unique structure of the UOT problem. We constructed several experiments and prove the effectiveness of our method.

1 Introduction

Optimal Transport (OT) has a long history in mathematics and prevailed recently due to its important role in measuring the distance between histograms in the Machine Learning community. It outperforms the traditional method in many fields like domain adaptation [Courty, 2017], generative model [Arjovsky et al., 2017], graph machine learning [Petric Maretic et al., 2019] and natural language processing. [Chen et al., 2019] Its popularity is attributed to the introduction of the Sinkhorn algorithm to the entropic optimal transport problem, [Cuturi, 2013] which improved the computational speed of OT problem from $\Theta(n^3)$ of Simplex method to $\Theta(n^2)$. However, Optimal transport problem can only deal with balanced samples, which limits its application in various data structures. Unbalanced Optimal Transport (UOT) problem has been promoted to deal with the drawback on unbalanced samples. Traditional Sinkhorn method can deal with an entropic UOT problem as well, but suffered from the slow convergence rate of the large penalty part and a non sparse solution, It can also be solved with other methods like Majorization-Minimization and FISTA method according to the choice of the penalty function in the primal problem and the Lagrange method for dual problem[Chapel et al., 2021]. UOT problem has a similar structure with many other famous problems like Non-negative Matrix Factorization and Lasso problem, which encourage the researchers to use the abundant results in these field to improve it.

Screening is a famous method in Lasso problem field, the L_1 penalize function causes a sparse solution for problem, which constrains many elements of solution equal to zero. The large scale optimization problem suffers from the computational process for manipulating on these zeros elements. [Ghaoui et al., 2010] invented the safe screening, which could theoretically judge whether the elements in solution equal to zero. It freezed the identified elements with linear complexity computation and save optimization time. Many new methods have been promoted to revise the method, [Ndiaye et al., 2017] invented the dynamic screening to dynamically screening out zeros elements, and there are many paper tries to improve it.

Fortunately, the OT function in UOT problem has the same effectiveness as L_1 in lasso and cause a sparse solution. We believe that this method could be applied on UOT problem and have better performance than ordinal Lasso problem for its special structure.

Contribution:

- We systematically combine the framework of Screening method on UOT problem, and we give a correct projection method for UOT screening, which is better than the Lasso one.

- We also imporve the constraints construction method for the specific sparse structure of UOT problems and benefits from it.

Preliminary work. Under review by AISTATS 2022. Do not distribute.

2 Background

2.1 Optimal Transport and Unbalanced Optimal Transport

Given two histograms $\alpha \in \mathbb{R}^m, \beta \in \mathbb{R}^n$, For a cost matrix $C \in \mathbb{R}^{m\times n}$, Optimal transport problem is trying to get a corresponding transport matrix $T \in \mathbb{R}^{m\times n}$ that minimize the whole transport cost, which could be formulated as:

$$W(\alpha, \beta) := \min_{\mathbf{T}\in\mathbb{R}_+^{n\times n}, \mathbf{T}\mathbb{1}=\alpha, \mathbf{T}^T\mathbb{1}=\beta} \langle \mathbf{C}, \mathbf{T}\rangle$$

We can write it into a vector type, set $c, t \in \mathbb{R}^{mn}$:

$$W(\alpha, \beta) := \min_{t\in\mathbb{R}_+^{n^2}, \mathbf{N}t=\alpha, \mathbf{M}t=\beta} c^\top t$$

$\mathbf{N} \in \mathbb{R}^{m\times mn}, \mathbf{N} \in \mathbb{R}^{n\times mn}$ are two matrix consisted with 0 and 1, listed in Appendix.A. We define $y = [\alpha, \beta]^\top$, the UOT problem add a penalty function for the historgrams:

$$W(\alpha, \beta) := \min_{t\in\mathbb{R}_+^{mn}} c^\top t + D_h(\mathbf{X}t, y) \tag{1}$$

D_h is the Bregman divergence and h is the norm, $\mathbf{X} = [\mathbf{M}^\top \mathbf{N}^\top]^\top$.

2.2 Relationship with Lasso

Lasso-like problem has a general formula as:

$$f(t) = g(t) + D_h(\mathbf{X}t, b), t \in \mathbb{R}^{mn}$$

When $g(t) = \lambda\|t\|$ and $D_h(\mathbf{X}t, b) = \|\mathbf{X}t - b\|_2^2$, this is the Euclid regression Lasso problem

2.3 Dynamic Screening Framework

We follow Yamada and Yamada [2021]'s framework to introduce about the whole dynamic screening technique for Lasso-like problem:

$$f(t) = g(t) + d(\mathbf{X}t) \tag{2}$$

By Frenchel-Rockafellar Duality, we get the dual problem

Theorem 1. *(Frenchel-Rockafellar Duality) If d and g are proper convex functions on $\mathbb{R}^{m+n}$ and $\mathbb{R}^{mn}$. Then we have the following:*

$$\min_t g(t) + d(Xt) = \max_\theta -d^*(-\theta) - g^*(X^\top\theta)$$

Because the primal function d is always convex, the dual function d^* is concave. Assuming d^* is an L-strongly concave problem. we design an area for any $\tilde{\theta}$ by the strongly concave property:

Theorem 2. *(L-strongly concave) Considering problem 2, if d and g are both convex, for $\forall\theta \in R^{m+n}$, we have the following:*

$$\theta \in \{\frac{L}{2}\|\theta - \tilde{\theta}\|_2^2 + d^*(-\tilde{\theta}) \leq d^*(-\theta)\}$$

We know that the optimal solution for the dual problem $\hat{\theta}$ satisfied the inequality, so the set is not empty. We can get the dual form of Lasso-like problem for some specific functions:

Lemma 3. *For $d(\mathbf{X}t) = \frac{1}{2}\|\mathbf{X}t - y\|_2^2$, the dual Lasso problem has the following form:*

$$d^*(-\theta) = \frac{1}{2}\|\theta\|_2^2 - y^\top\theta$$

$$g^*(X^\top\theta) = \begin{cases} 0 & (\forall t \quad \theta^\top Xt - g(t) \leq 0) \\ \infty & (\exists t \quad \theta^\top Xt - g(t) \leq 0) \end{cases}$$

3 Dynamic Screening and UOT problem

3.1 Screening for UOT

For UOT problem 1, we could get its dual form.

Lemma 4. *(Dual form of UOT problem)*

$$\begin{aligned} &-d^*(-\theta) - g^*(X^\top\theta) = -\frac{1}{2}\|\theta\|_2^2 - y^\top\theta \\ &\textbf{s.t.} \quad \forall i \quad x_i^\top\theta - \lambda c_i \leq 0 \end{aligned} \tag{3}$$

The equation indicate a dual feasible area constructed by many dual constraints, the optimal solution is inside the constraints.

From the KKT condition, we can make sure that, for the optimal primal solution $\hat{t}$:

Theorem 5. *(Screening) For the dual optimal solution $\hat{\theta}$, we have the following relationship:*

$$x_i^\top\hat{\theta} - \lambda c_i \begin{cases} < 0 & \Rightarrow \hat{t}_i = 0 \\ = 0 & \Rightarrow \hat{t}_i \geq 0 \end{cases} \tag{4}$$

As we do not know the information of $\hat{t}$ directly, we can construct an area $\mathcal{R}^S$ containing the $\hat{t}$, if

$$\max_{t\in\mathcal{R}^S} x_i^\top\theta - \lambda c_i < 0 \tag{5}$$

then we have:

$$x_i^\top\hat{\theta} - \lambda c_i < 0 \tag{6}$$

which means the corresponding $\hat{t}_i = 0$, and can be screening out.

Now we start to construct the area containing $\hat{\theta}$, from 2 we know that, if we can find a $\tilde{\theta}$ inside the dual feasible area, we can construct a circle where $\hat{\theta}$ is. For convenience, we set all $\theta = [u^\intercal, v^\intercal]^\intercal$.

Theorem 6. *(UOT projection) For any any $\theta^k = [{u^k}^\intercal, {v^k}^\intercal]^\intercal$, we can compute the projection $\tilde{\theta}^k = [\tilde{u^k}^\intercal, \tilde{v^k}^\intercal]^\intercal$ onto the dual feasible area.*

$$\begin{aligned} \tilde{u^k}_i &= {u^k}_i - \max_j \frac{{u^k}_i + {v^k}_j - \mathbf{C}_{ij}}{2} \\ \tilde{v^k}_j &= {v^k}_j - \max_i \frac{{u^k}_i + {v^k}_j - \mathbf{C}_{ij}}{2} \end{aligned} \tag{7}$$

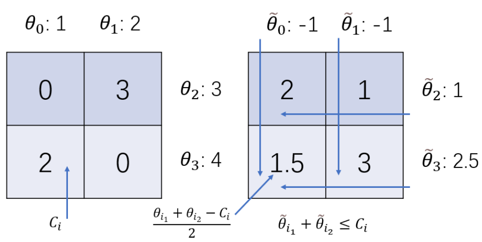

Figure 1: Shifting on a 2×2 matrix

The intersection of the circle and the dual feasible area contain the $\hat{\theta}$, however, the multilinear constraints make it hard to compute the maximum for the problem, We design a relaxation method. which divide the constrains into two parts, then we are maximizing on the intersection of two hyperplane and a hyper-ball.

Theorem 7. *(Screening Area for UOT) With the help of $\tilde{\theta}$, we can construct specific area for every single primal variable as following area $\mathcal{R}^S_{IJ}$, and the optimal dual solution $\hat{\theta}$ must be inside the area.*

$$\begin{aligned} &\theta^\intercal X^{A_{IJ}}\beta - \lambda g^{A_{IJ}}\beta \leq 0 \\ \mathcal{R}^S_{IJ} = \{\theta \| &\theta^\intercal X^{B_{IJ}}\beta - \lambda g^{B_{IJ}}\beta \leq 0 \\ &(\theta - \tilde{\theta})^\intercal(\theta - y) \leq 0\} \end{aligned} \tag{8}$$

We devide the constraints into two group A and B for every single IJ, we have $X^A + X^B = X$ and $g^A + g^B = g$ the computational process is in Appendix.A

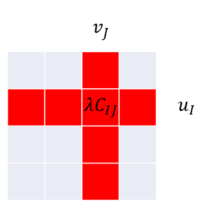

Figure 2: Selection of group A_{IJ}(red) and B_{IJ}(grey)

3.2 Screening Algorithms

Algorithm 1 UOT Dynamic Screening Algorithm

Input: $t_0, S \in R^{n\times m}, S_{ij} = 1$
Output: S

Choose a solver for the problem.
for $t = 0$ to K **do**
  Projection $\tilde{\theta} = \text{Proj}(t^k)$
  if $(i \neq 0)$ **then**
    break
  end if
  $\mathcal{R} \Leftarrow \mathcal{R}^S(\tilde{\theta}, t^k)$
  $S \Leftarrow S_{ij} = 0$ if $\max_{\theta\in\mathcal{R}^S} {x_{k(i,j)}}^\intercal\theta < \lambda c_{k(i,j)}$
  for $a \in A_{ij} \| A_{ij} = 0$ **do**
    $t^k(i,j) \Leftarrow 0$
  end for
  $t^{k+1} = \text{update}(t^k)$
end for
return t^{K+1}, S

Screening method is irrelevent to the optimization solver you choose. We give the specific algorithm for L_2 UOT problem to show the whole optimization process.
Benefiting from the special structure of UOT problem, The computational time of the screening method is $O(dnm)$, d is a constant.

4 Experiments

In this section, we show the efficacy of the proposed methods using a toy Gaussian model and the MNIST dataset.

4.1 Projection Method

In order to prove the rightness of our projection method compared with the traditional projection method in lasso problem, we compared the screening ratio with random generated Gaussian measures by two projection method. We set the $\lambda = \frac{\|\mathbf{X}^\intercal y\|}{100}$ and test for 10 different pairs.We choose FISTA for solving L_2 penalized UOT problem.

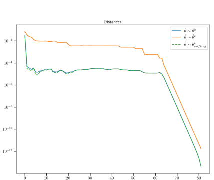

Figure 3: Distance of different projection method

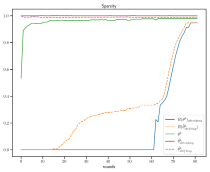

Figure 4: Screening ratio of different projection method

4.2 Divide Method

We compared the screening ratio with three different method, including our Divide method, Dynamic Sasvi method and Gap method.

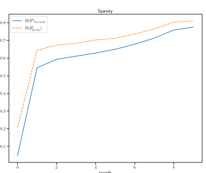

Figure 5: Screening ratio of dividing method

4.3 Best divide Method

We compared the screening ratio with three different method, including our Divide method, Dynamic Sasvi method and a random divide method.

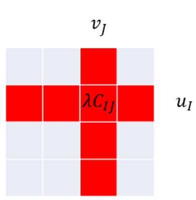

Figure 6: Comparing of our seperation method with random seperation method

4.4 Speed up ratio

We choose FISTA method, Newton method and Language method to test about the screening ratio.

5 Conclusion

Our algorithm is great, we are going to apply the method onto Sinkhorn

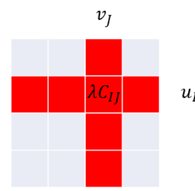

Figure 7: speed up ratio for different solver

References

Martin Arjovsky, Soumith Chintala, and Léon Bottou. Wasserstein generative adversarial networks. In *International conference on machine learning*, pages 214–223. PMLR, 2017.

Laetitia Chapel, Rémi Flamary, Haoran Wu, Cédric Févotte, and Gilles Gasso. Unbalanced optimal transport through non-negative penalized linear regression. In M. Ranzato, A. Beygelzimer, Y. Dauphin, P.S. Liang, and J. Wortman Vaughan, editors, *Advances in Neural Information Processing Systems*, volume 34, pages 23270–23282. Curran Associates, Inc., 2021. URL `https://proceedings.neurips.cc/paper/2021/file/c3c617a9`

Liqun Chen, Yizhe Zhang, Ruiyi Zhang, Chenyang Tao, Zhe Gan, Haichao Zhang, Bai Li, Dinghan Shen, Changyou Chen, and Lawrence Carin. Improving sequence-to-sequence learning via optimal transport. In *7th International Conference on Learning Representations, ICLR 2019*. International Conference on Learning Representations, ICLR, January 2019. Generated from Scopus record by KAUST IRTS on 2021-02-09.

Nicolas Courty. Optimal transport for domain adaptation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(9):1853–1865, 2017. doi: 10.1109/TPAMI.2016.2615921.

Marco Cuturi. Sinkhorn distances: Lightspeed computation of optimal transport. In C.J. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K.Q. Weinberger, editors, *Advances in Neural Information Processing Systems*, volume 26. Curran Associates, Inc., 2013. URL `https://proceedings.neurips.cc/paper/2013/file/af21d0c9`

Laurent El Ghaoui, Vivian Viallon, and Tarek Rabbani. Safe feature elimination for the lasso and sparse supervised learning problems. *arXiv preprint arXiv:1009.4219*, 2010.

Eugene Ndiaye, Olivier Fercoq, Alex, re Gramfort, and Joseph Salmon. Gap safe screening rules for sparsity enforcing penalties. *Journal of Machine Learning Research*, 18(128):1–33, 2017. URL `http://jmlr.org/papers/v18/16-577.html`.

Hermina Petric Maretic, Mireille El Gheche, Giovanni Chierchia, and Pascal Frossard. Got: An optimal transport framework for graph comparison. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019. URL https://proceedings.neurips.cc/paper/2019/file/fdd5b16fc8134339089ef25b3cf0e588-Paper.pdf.

Hiroaki Yamada and Makoto Yamada. Dynamic sasvi: Strong safe screening for norm-regularized least squares. In M. Ranzato, A. Beygelzimer, Y. Dauphin, P.S. Liang, and J. Wortman Vaughan, editors, *Advances in Neural Information Processing Systems*, volume 34, pages 14645–14655. Curran Associates, Inc., 2021. URL https://proceedings.neurips.cc/paper/2021/file/7b5b23f4aadf9513306bcd59afb6e4c9-Paper.pdf.

Supplementary Material: Dynamic Screening Method on the Unbalanced Optimal Transport Problem

A Notation

$$M = \begin{pmatrix} 1 & & & & & \cdots\cdots & 1 & & & & \\ & 1 & & & & \cdots\cdots & & 1 & & & \\ & & \ddots & & & \ddots\,\ddots & & & \ddots & & \\ & & & 1 & & & & & & 1 & \\ & & & & 1 & \cdots\cdots & & & & & 1 \end{pmatrix} \tag{9}$$

$$N = \begin{pmatrix} 1 & 1 & \dots & 1 & \cdots\cdots & & & & \\ & & & & \ddots\,\ddots & & & & \\ & & & & \cdots\cdots & 1 & 1 & \dots & 1 \end{pmatrix} \tag{10}$$

B Maximizing on two plane and a circle

We Generalize the problem as

$$\max_{\theta \in \mathcal{R}_I^S} \theta_{I_1} + \theta_{I_2} \tag{11}$$

Considering the center of the circle as θ^o, we define $\theta = \theta^o + p$, as $\theta^o_{I_1} + \theta^o_{I_2}$ is a constant, the Lagrangrian function of $\max_{\theta \in \mathcal{R}_I^S} p_{I_1} + p_{I_2}$ is.

$$L(p, \eta, \mu, \nu) = p_{I_1} + p_{I_2} + \eta(r - p^\top p) + \mu(a^\top p - e_a) + \nu(b^\top p - e_b) \tag{12}$$

$$\frac{\partial L}{\partial p_i} = \begin{cases} 1 - 2\eta p_i + \mu a_i + \nu b_i & i = I_1, I_2 \\ \quad -2\eta p_i + \mu a_i + \nu b_i & i \neq I_1, I_2 \end{cases} \tag{13}$$

$$p_i{}^* = \begin{cases} \dfrac{1 + \mu a_i + \nu b_i}{2\eta} & i = I_1, I_2 \\ \dfrac{\mu a_i + \nu b_i}{2\eta} & i \neq I_1, I_2 \end{cases} \tag{14}$$

We can get the dual problem and directly solve it:

$$L(\eta, \mu, \nu) = \frac{1 + \mu a_{I_1} + \nu I_1}{2\eta} + \frac{1 + \mu a_{I_2} + \nu I_2}{2\eta} + \eta(r - p^\top p) + \mu(a^\top p - e_a) + \nu(b^\top p - e_b) \tag{15}$$

C FORMATTING INSTRUCTIONS FOR THE SUPPLEMENTARY MATERIAL

Your supplementary material should go here. It may be in one-column or two-column format. To display the supplementary material in two-column format, comment out the line

```
\onecolumn \makesupplementtitle
```

and uncomment the following line:

```
\twocolumn[ \makesupplementtitle ]
```

Please submit your paper (including the supplementary material) as a single PDF file. Besides the PDF file, you may submit a single file of additional non-textual supplementary material, which should be a ZIP file containing, e.g., code.

If you require to upload any video as part of the supplementary material of your camera-ready submission, do not submit it in the ZIP file. Instead, please send us via email the URL containing the video location.

Note that reviewers are under no obligation to examine your supplementary material.

D MISSING PROOFS

The supplementary materials may contain detailed proofs of the results that are missing in the main paper.

D.1 Proof of Lemma 3

In this section, we present the detailed proof of Lemma 3 and then [...]

E ADDITIONAL EXPERIMENTS

If you have additional experimental results, you may include them in the supplementary materials.

E.1 The Effect of Regularization Parameter

Our algorithm depends on the regularization parameter λ. Here we illustrate the effect of this parameter on the performance of our algorithm [...]