Dynamic Screening Method on the Unbalanced Optimal Transport Problem

Anonymous Author
Anonymous Institution

Abstract

This paper promote a dynamic screening framework for Unbalanced Optimal Transport (UOT) problem. Recently, researchers connected the UOT problem with Lasso problem, which encourage us to apply the common speeding up method screening on the UOT problem. We demonstrate the effectiveness of the screening method and propose a better improvement to it based on the unique structure of the UOT problem. We constructed several experiments on the problem and they indicate that...

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)Arjovsky, Chintala, and Bottou], graph machine learning and natural language processing. Its popularity is attributed to the introduction of the Sinkhorn algorithm to the entropic optimal transport problem. which improve the computational speed of OT problem from $\Theta(n^3)$ to $\Theta(n^2)$. However, Optimal transport problem can only deal with balanced samples, which limits its application in the various data structures. Unbalanced Optimal Transport (UOT) problem has been promoted to deal with the drawback. Except for the traditional Sinkhorn method, UOT can be solved with other method like MM for primal problem and Lagrange method for dual problem. 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, theL_1 penalize function cause a sparse solution for lasso 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. [] invented a technique called safe screening, which could theoretically judeg whether the elements in solution equal to zero. It freeze the identified elements before computation and save computational time. Many new methods have been promoted to revise the method, [] invented a 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

Contribution: We systematically provide the framework for Screening method on UOT problem. We give the correct projection method for UOT screening, which is better than the Lasso one. We also edit the constraints construtioin 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

Optimal transport problem has a formula:

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

$$\mathbf{T}\mathbb{1} = \alpha, \mathbf{T}^T\mathbb{1} = \beta$$

We can write it into a vector type:

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

The UOT problem can be wrote as:

$$W(\alpha,\beta) := \min_{t\in\mathbb{R}_+^{n^2}} c^\top t + D_h(Xt,y) \tag{1}$$

D_h is the Bregman divergence and h is the norm.

2.2 Relationship with Lasso

Lasso like problem can be write as:

$$f(t) = g(t) + D_h(Xt,b), t\in\mathbb{R}^{n^2}$$

When $g(t)=\lambda\|t\|$ and $D_h(Xt,b)=\|Xt-b\|_2^2$, this is the L_2 penalized Lasso problem

2.3 Dynamic Screening Framework

We follow the [Yamada and Yamada(2021)]'s method to introduce about the whole dynamic framework, The Lasso like problem has a common formula like:

$$f(t) = g(t) + d(Xt)$$

By Frenchel-Rockafellar Duality, we get the dual problem

Theorem 1.

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

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 $\tilde{\theta}$:

Theorem 2.

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

we know that the optimal solution for the dual problem $\hat{\theta}$ satisfied the inequality. So it is not empty. For $d(Xt)=\frac{1}{2}\|Xt-y\|_2^2$, the Lasso-like problem have:

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

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

3 Dynamic Screening and UOT problem

3.1 Screening for UOT

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

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

The equation indicate that the dual problem has many dual constraints, the optimal solution is inside the constaints.
From the KKT condition, we can make sure that, for the optimal primal solution $\hat{t}$:

$$x_i^\top\theta - \lambda c_i \begin{cases} <0 & \Rightarrow \hat{t}_i = 0 \\ =0 & \Rightarrow \hat{t}_i \ge 0 \end{cases} \tag{3}$$

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

$$\max_{t\in\mathcal{R}^S} x_i^\top\theta - \lambda c_i < 0 \tag{4}$$

then we have:

$$x_i^\top\hat{\theta} - \lambda c_i < 0 \tag{5}$$

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 the $\hat{\theta}$ is inside the intersection of the area of 2 and the dual feasible area. 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 3.

$$\mathcal{R}^S = \{nihao\}$$

the computational process is in Appendix.A

3.2 Algorithms

screening method is irrelevent to the optimization solver you choose. We gave the specific algorithm for L_2 UOT problem to show the whole optimization process.
An important part is that we use the primal solution to compute a dual solution, which might not inside the dual feasible area, a projection method is necessary. We promote a new projection method for the UOT problem for its sparse matrix structure.

Theorem 4.

$$\tilde{\theta} = 1$$

Algorithm 1 Algorithm for ...

Input: in
Output: out

Initialisation :
first statement
LOOP Process
for $i = l - 2$ to 0 **do**
  statements..
  if $(i \neq 0)$ **then**
    statement..
  end if
end for
return P

4 Experiments

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

4.1 Screening Ratio

5 Conclusion

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

References

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

[Courty(2017)] 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.

[Yamada and Yamada(2021)] 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 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.

B MISSING PROOFS

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

B.1 Proof of Lemma 3

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

C ADDITIONAL EXPERIMENTS

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

C.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 [...]