Dynamic Screening for Unbalanced Optimal Transport Problem

Anonymous Author
Anonymous Institution

Abstract

Safe Screening is a famous technique on Lasso problem by frezzing the zero elements in the solution. Recently, researchers have linked the UOT problem to the Lasso problem. In this paper, we applies the the newest dynamic screening framework on the L_2 penalized Unbalanced Optimal Transport (UOT) problem. We proposed new projection and feasible area construction method for UOT problem and demonstrate its extraordinary effectiveness. Benefiting from the unique structure of the UOT problem, our proposed improved algorithm substantially improves the screening efficiency compared to the ordinary Lasso algorithm without significantly increasing the computational complexity. We demonstrate the advantages of the algorithm through some constructed experiments on Gaussian model and MNIST dataset.

1 Introduction

Optimal Transfer (OT) has a long history in mathematics and has recently become prevalent due to its important role in the machine learning community for measuring distances between histograms. It has outperformed traditional methods in many different areas such as domain adaptation (Courty, 2017), generative models (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 Sinkhorn's algorithm for the entropy optimal transmission problem, (Cuturi, 2013) which improves the computational speed of the OT problem from $\Theta(n^3)$ of Simplex's method to $\Theta(n^2)$. In order to extend the optimal transmission problem, which can only handle balanced samples, to a wider range of unbalanced samples. The unbalanced optimal transport (UOT) is proposed by modifying the restriction term to a penalty function term. UOT has been used in several applications like computational biology(Schiebinger et al., 2019) , machine learning (Janati et al., 2019) and deep learning (Yang and Uhler, 2019).

The UOT problem is a regularized version of Kantorovich formulation which replaced the equality constraints with penalty functions on the marginal distributions with a divergence. Many different divergence has been taken into consider for UOT problem like KL divergence, l_1 norm, and L_2 norm. When it comes to the solving method, KL penalty with the entropy form can also be solved by the Sinkhorn algorithm. It provide the UOT computation with scalability and differentiability, but suffers from a larger error of KL divergence and lack of sparsity compared with other regularizers (Blondel et al., 2018). L_2 norm could bring a sparse solution, which attracted the attention of researchers and many new algorithms were designed for it, such as FISTA, Majorization-Minimization method and Lagrange pairwise method. (Chapel et al., 2021) The link between UOT problem with many other well-known problems such as non-negative matrix decomposition and Lasso problem has been discovered, which also encourages researchers to improve it by using the rich results in these fields.

Screening is a well-known technique proposed by (Ghaoui et al., 2010) in the field of lasso problems, where the L_1 regularizer leads to a sparse solution for the problem. It can preselect solutions that must be zero from theory and freeze them before computation. The solutions of many large-scale optimization problems are sparse, and a large amount of computation is wasted on updating the zero elements. With the safe screening method, we can identify and freeze the elements that are zero with linear complexity computation before enabling the algorithm, thus saving optimization time. the Screening method get attention in recent years and have been improved a lot, such as Dynamic Screening (Bonnefoy et al., 2015), Gap

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

screening method (Ndiaye et al., 2017) and Dynamic Sasvi (Yamada and Yamada, 2021)

The OT and UOT problems produce extremely sparse solutions due to the effectiveness of their similar operator to Lasso problem. We believe that indicate the potential effectiveness of applying screening technical in the Lasso problem onto the UOT problem. Furthermore, Different from the Lasso problem which has a dense constraints matrix, UOT problem's constraint matrix is extremely sparse and has a unique transport matrix structure, which would benefit the design of screening and the outcome.

Contribution:

- We systematically provide the newest framework for Screening method on UOT problem. Considering the sparse and specific structure of UOT problem, we design a new projection method for UOT screening, which hugely improve the screening performance then the general Lasso method.
- We promoted a two plane screening method for UOT problems, which benefits from its sparse constraints and outperforms the ordinary methods adding only a negligible amount of computation

2 Background

2.1 Optimal Transport and Unbalanced Optimal Transport

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

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

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

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

$\mathbf{N} \in \mathbb{R}^{m\times mn}, \mathbf{N} \in \mathbb{R}^{n\times mn}$ are two matrix consisted with 0 and 1, listed in Appendix.A. When the $\|\mathbf{a}\|_2 = \|\mathbf{b}\|_2$, it is the OT problem. When $\|\mathbf{a}\|_2 \neq \|\mathbf{b}\|_2$, the solution $\hat{\mathbf{t}}$ is not exist. We define $\mathbf{y} = [\mathbf{a}, \mathbf{b}]^\mathsf{T}$, the UOT problem uses a penalty function for the historgrams:

$$W(\mathbf{a},\mathbf{b}) := \min_{t\in\mathbb{R}_+^{mn}} \mathbf{c}^\mathsf{T} t + D_h(\mathbf{X}\mathbf{t}, \mathbf{y}) \tag{1}$$

D_h is the Bregman divergence derived from the norm h, $\mathbf{X} = [\mathbf{M}^\mathsf{T}\mathbf{N}^\mathsf{T}]^\mathsf{T}$.

2.2 Relationship with Lasso

Lasso-like problem has a general formula as:

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

When $g(\mathbf{t}) = \lambda\|\mathbf{t}\|_1$ and $D_h(\mathbf{X}\mathbf{t}, \mathbf{y}) = \|\mathbf{X}\mathbf{t} - \mathbf{y}\|_2^2$, this is the L_2 regression Lasso problem. It is important to note that $\mathbf{X}$ in UOT is a bit different from the $\mathbf{X}$ in Lasso problem, the former $\mathbf{X}$ has a specific structure and has only two non-zero elements and equal to 1, which is quite different to the irregular and dense $\mathbf{X}$ in 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:

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

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:*

$$\min_{\mathbf{t}} g(\mathbf{t}) + d(\mathbf{X}\mathbf{t}) = \max_{\theta} -d^*(-\theta) - g^*(\mathbf{X}^\mathsf{T}\theta)$$

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

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

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

We know that the optimal solution for the dual problem $\hat{\theta}$ satisfied the inequality, so the set is not empty.

3 Dynamic Screening and UOT problem

3.1 Screening for UOT

We can get the dual form of UOT problem:

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

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

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

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

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

$$\begin{aligned} &-d^*(-\theta) - g^*(\mathbf{X}^\mathsf{T}\theta) = -\frac{1}{2}\|\theta\|_2^2 - \mathbf{y}^\mathsf{T}\theta \\ &\mathbf{s.t.} \quad \forall p \quad \mathbf{x}_p^\mathsf{T}\theta - \lambda\mathbf{c}_p \le 0 \end{aligned} \tag{3}$$

$\mathbf{x}_p$ is the p-th column of $\mathbf{X}$, these inequations 9 make up a dual feasible area written as $\mathcal{R}^D$, and the optimal solution definitely satisfied them.
From the KKT condition, we konw that, for the optimal primal solution $\hat{\mathbf{t}}$:

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

$$\mathbf{x}_p^\mathsf{T}\hat{\theta} - \lambda\mathbf{c}_p \begin{cases} < 0 & \Rightarrow \hat{\mathbf{t}}_p = 0 \\ = 0 & \Rightarrow \hat{\mathbf{t}}_p \ge 0 \end{cases} \tag{4}$$

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

$$\max_{\mathbf{t} \in \mathcal{R}^S} \mathbf{x}_p^\mathsf{T}\theta - \lambda\mathbf{c}_p < 0 \tag{5}$$

then we have:

$$\mathbf{x}_p^\mathsf{T}\hat{\theta} - \lambda\mathbf{c}_p < 0 \tag{6}$$

which means the corresponding $\hat{t}_i = 0$, and can be screening out. As for the UOT problem, $x_p = [..., 0, 1, 0, ..., 0, 1, 0, ...]^\mathsf{T}$, which has only two elements p_1, p_2 equal to 1, we can set $\theta = [\mathbf{u}^\mathsf{T}, \mathbf{v}^\mathsf{T}]^\mathsf{T}$ and $\mathbf{u} \in \mathbb{R}^m, \mathbf{v} \in \mathbb{R}^n$, assuming $p = (I, J), I = p \mid m, J = p \bmod m$. then we could rewrite 6 as

$$\mathbf{u}_I + \mathbf{v}_J - \lambda\mathbf{c}_p < 0 \tag{7}$$

Before we start to construct the area containing $\hat{\theta}$, from 2 we know that, we have to find a $\tilde{\theta}$ in the dual feasible area before we construct any area, there is a relationship between the primal variable and dual variable $\theta = \mathbf{y} - \mathbf{X}\mathbf{t}$, however, the outcome θ might not inside the dual feasible area, which encourage us to project. In lasso problem, as the constraints limit the $\|\mathbf{x}_p\theta\|_1$, and every elements of θ is multiplied by the dense x_i, researchers has to use a shrinking method to obtain a $\tilde{\theta} \in \mathcal{R}^D$ for further constructing the dual screening area:

$$\tilde{\theta} = \frac{(\mathbf{y} - \mathbf{X}\mathbf{t})}{\max(\lambda\mathbf{c}, \|\mathbf{X}^\mathsf{T}(\mathbf{y} - \mathbf{X}\mathbf{t})\|_\infty)} \tag{8}$$

As for UOT problem, it only allow $\mathbf{t}_p \ge 0$, and its x_p only consists of two non-zero elements, which allows us to adapt a better projection method:

Theorem 6. *(UOT projection) For any any $\theta = [\mathbf{u}^\mathsf{T}, \mathbf{v}^\mathsf{T}]^\mathsf{T}$, we can compute the projection $\tilde{\theta} = [\tilde{\mathbf{u}}^\mathsf{T}, \tilde{\mathbf{v}}^\mathsf{T}]^\mathsf{T} \in \mathcal{R}^D$.*

$$\begin{aligned} \tilde{\mathbf{u}}_I &= \mathbf{u}_I - \max_{0 \ge j \ge n} \frac{\mathbf{u}_I + \mathbf{v}_j - \lambda\mathbf{c}_p}{2} \\ &= \frac{\mathbf{u}_I + \lambda\mathbf{c}_p}{2} - \frac{1}{2}\max_{0 \ge j \ge n} \mathbf{v}_j \\ \tilde{\mathbf{v}}_J &= \mathbf{v}_J - \max_{0 \ge i \ge m} \frac{\mathbf{u}_i + \mathbf{v}_J - \lambda\mathbf{c}_p}{2} \\ &= \frac{\mathbf{v}_J + \lambda\mathbf{c}_p}{2} - \frac{1}{2}\max_{0 \ge i \ge m} \mathbf{u}_j \end{aligned} \tag{9}$$

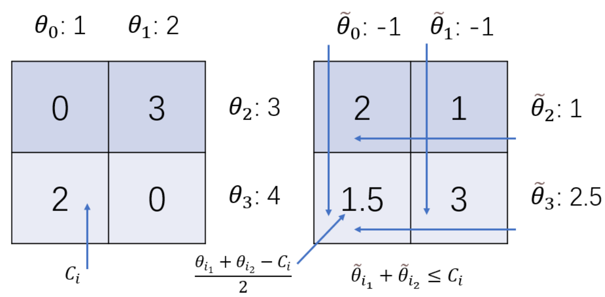

Figure 1: Shifting on a 2×2 matrix

As we have got the $\tilde{\theta}$ in the R^D and we also have another constraint area $\mathcal{R}^C$, we are sure that the $\hat{\mathbf{t}} \in \mathcal{R}^C \cap \mathcal{R}^D$. However, The intersection of a sphere and a polytope can not be compute in $O(knm)$, where k is a constant. We design a relaxation method. which divide the constrains into two parts, then we are maximizing on the intersection of two hyperplanes 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 $\mathcal{R}_{IJ}^S$, and the optimal dual solution $\hat{\theta}$ must be inside the area.*

$$\begin{aligned} &\theta^\mathsf{T}\mathbf{X}^{A_{IJ}}\mathbf{t} - \lambda g^{A_{IJ}}\mathbf{t} \le 0 \\ \mathcal{R}_{IJ}^S = \{\theta \| &\theta^\mathsf{T}\mathbf{X}^{B_{IJ}}\mathbf{t} - \lambda g^{B_{IJ}}\mathbf{t} \le 0 \\ &(\theta - \tilde{\theta})^\mathsf{T}(\theta - \mathbf{y}) \le 0\} \end{aligned} \tag{10}$$

We devide the constraints into two group A and B for every single IJ, we have $X^A + X^B = X$ and $g^A + g^B =$

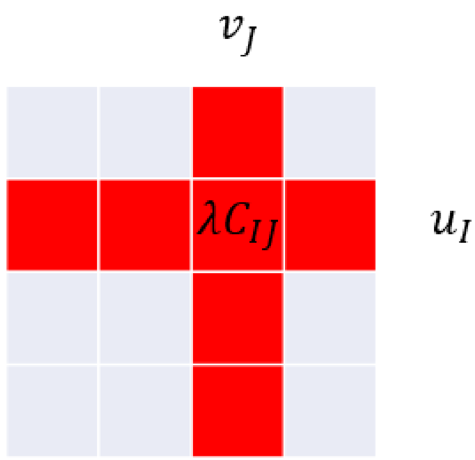

Figure 2: Selection of group A_{IJ}(red) and B_{IJ}(grey)

g This problem can be solved easily by Lagrangian method in constant time, the computational process is in Appendix.A

3.2 Screening Algorithms

Algorithm 1 UOT Dynamic Screening Algorithm

Input: $\mathbf{t}_0, S \in R^{n\times m}, S_{ij} = 1, (i,j) = mi + j$
Output: S
  Choose a solver for the problem.
  for $k = 0$ to K **do**
    Projection $\tilde{\theta} = \text{Proj}(t^k)$
    for $i = 0$ to m **do**
      for $j = 0$ to n **do**
        $\mathcal{R}^S \Leftarrow \mathcal{R_{IJ}}^S(\tilde{\theta}, t^k)$
        $S \Leftarrow S_{ij} = 0$ if $\max_{\theta \in \mathcal{R}^S} x_{(i,j)}{}^\top \theta < \lambda c_{(i,j)}$
      end for
    end for
    for $(i,j) \in \{(i,j) \| S_{ij} = 0\}$ **do**
      $\mathbf{t}^k_{(i,j)} \Leftarrow 0$
    end for
    $\mathbf{t}^{k+1} = \text{update}(\mathbf{t}^k)$
  end for
  return $\mathbf{t}^{K+1}, S$

The 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.

4 Experiments

In this section, we show the efficacy of the proposed methods using toy Gaussian models 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 projection distance and screening ratio with random generated Gaussian measures by two projection method. We set the $\lambda = \frac{\|\mathbf{X}^\top y\|}{100}$ and test for 10 different pairs.We choose the FISTA for solving the L_2 penalized UOT problems. Our projection method have only moved the dual point by a very small order of magnitude. It ensures that the points are kept at a smaller distance from the optimal solution and cause a better screening effects.

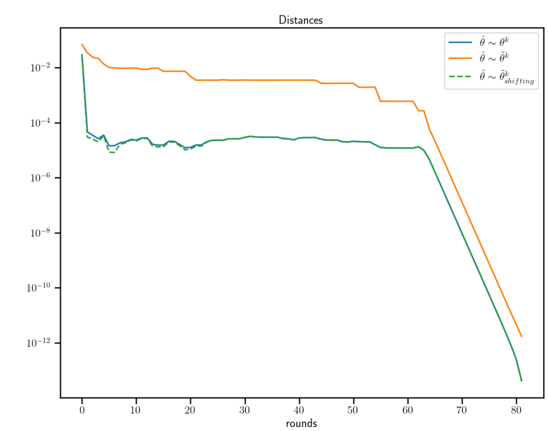

Figure 3: Distance of different projection method

4.2 Divide Method

We compared the screening ratio with three different method, including our Divide method, Dynamic Sasvi method and Gap method. Every method would use our projection method to get a better outcome, which also make sure the difference in performance is only in the construction of the feasible domain.

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.

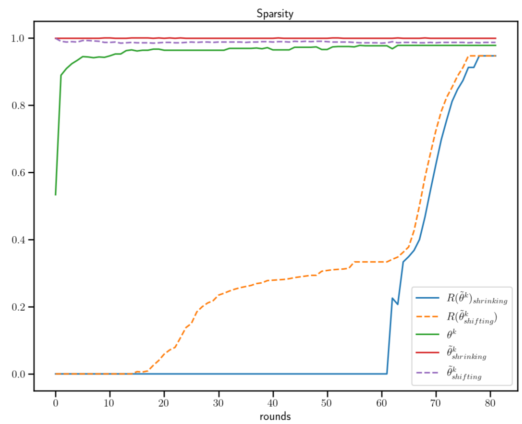

Figure 4: Screening ratio of different projection method

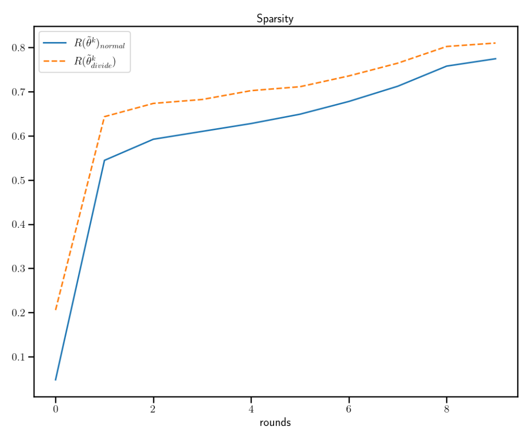

Figure 5: Screening ratio of dividing method

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

References

Arjovsky, M., Chintala, S., and Bottou, L. (2017). Wasserstein generative adversarial networks. In *International conference on machine learning*, pages 214–223. PMLR.

Blondel, M., Seguy, V., and Rolet, A. (2018). Smooth

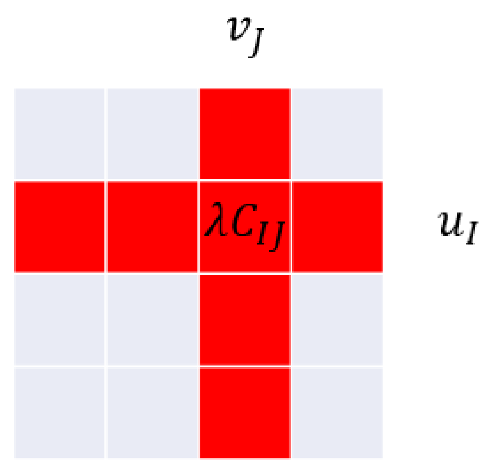

Figure 6: Comparing of our seperation method with random seperation method

and sparse optimal transport. In Storkey, A. J. and Pérez-Cruz, F., editors, *International Conference on Artificial Intelligence and Statistics, AISTATS 2018, 9-11 April 2018, Playa Blanca, Lanzarote, Canary Islands, Spain*, volume 84 of *Proceedings of Machine Learning Research*, pages 880–889. PMLR.

Bonnefoy, A., Emiya, V., Ralaivola, L., and Gribonval, R. (2015). Dynamic screening: Accelerating first-order algorithms for the lasso and group-lasso. *IEEE Transactions on Signal Processing*, 63(19):5121–5132.

Chapel, L., Flamary, R., Wu, H., Févotte, C., and Gasso, G. (2021). Unbalanced optimal transport through non-negative penalized linear regression. In Ranzato, M., Beygelzimer, A., Dauphin, Y., Liang, P., and Vaughan, J. W., editors, *Advances in Neural Information Processing Systems*, volume 34, pages 23270–23282. Curran Associates, Inc.

Chen, L., Zhang, Y., Zhang, R., Tao, C., Gan, Z., Zhang, H., Li, B., Shen, D., Chen, C., and Carin, L. (2019). Improving sequence-to-sequence learning via optimal transport. In *7th International Conference on Learning Representations, ICLR 2019*. International Conference on Learning Representations, ICLR. Generated from Scopus record by KAUST IRTS on 2021-02-09.

Courty, N. (2017). Optimal transport for domain adaptation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(9):1853–1865.

Cuturi, M. (2013). Sinkhorn distances: Lightspeed computation of optimal transport. In Burges, C.,

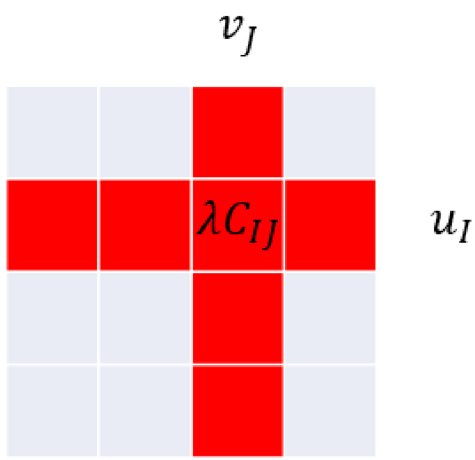

Figure 7: speed up ratio for different solver

Bottou, L., Welling, M., Ghahramani, Z., and Weinberger, K., editors, *Advances in Neural Information Processing Systems*, volume 26. Curran Associates, Inc.

Ghaoui, L. E., Viallon, V., and Rabbani, T. (2010). Safe feature elimination for the lasso and sparse supervised learning problems. *arXiv preprint arXiv:1009.4219*.

Janati, H., Cuturi, M., and Gramfort, A. (2019). Wasserstein regularization for sparse multi-task regression. In Chaudhuri, K. and Sugiyama, M., editors, *The 22nd International Conference on Artificial Intelligence and Statistics, AISTATS 2019, 16-18 April 2019, Naha, Okinawa, Japan*, volume 89 of *Proceedings of Machine Learning Research*, pages 1407–1416. PMLR.

Ndiaye, E., Fercoq, O., Alex, re Gramfort, and Salmon, J. (2017). Gap safe screening rules for sparsity enforcing penalties. *Journal of Machine Learning Research*, 18(128):1–33.

Petric Maretic, H., El Gheche, M., Chierchia, G., and Frossard, P. (2019). Got: An optimal transport framework for graph comparison. In Wallach, H., Larochelle, H., Beygelzimer, A., d'Alché-Buc, F., Fox, E., and Garnett, R., editors, *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc.

Schiebinger, G., Shu, J., Tabaka, M., Cleary, B., Subramanian, V., Solomon, A., Gould, J., Liu, S., Lin, S., Berube, P., Lee, L., Chen, J., Brumbaugh, J., Rigollet, P., Hochedlinger, K., Jaenisch, R., Regev, A., and Lander, E. S. (2019). Optimal-transport analysis of single-cell gene expression identifies developmental trajectories in reprogramming. *Cell*, 176(4):928–943.e22.

Yamada, H. and Yamada, M. (2021). Dynamic sasvi: Strong safe screening for norm-regularized least squares. In Ranzato, M., Beygelzimer, A., Dauphin, Y., Liang, P., and Vaughan, J. W., editors, *Advances in Neural Information Processing Systems*, volume 34, pages 14645–14655. Curran Associates, Inc.

Yang, K. D. and Uhler, C. (2019). Scalable unbalanced optimal transport using generative adversarial networks. In *7th International Conference on Learning Representations, ICLR 2019, New Orleans, LA, USA, May 6-9, 2019*. OpenReview.net.

Supplementary Material:
Dynamic Screening for Unbalanced Optimal Transport Problem

A Notation

$$M = \begin{pmatrix} 1 & & & & \cdots\cdots & 1 & & & \\ & 1 & & & \cdots\cdots & & 1 & & \\ & & \ddots & & \ddots\ \ddots & & & \ddots & \\ & & & 1 & & & & 1 & \\ & & & & 1 \ \cdots\cdots & & & & 1 \end{pmatrix} \tag{11}$$

$$N = \begin{pmatrix} 1 & 1 & \dots & 1 & 1 & \cdots\cdots & & & & & \\ & & & & & \ddots\ \ddots & & & & & \\ & & & & & \cdots\cdots & 1 & 1 & \dots & 1 & 1 \end{pmatrix} \tag{12}$$

B Proof of Theorem 6

For any $p \in 0, 1, ..., nm-1$ we assume that $p = (I, J)$, then we can compute that:

$$\begin{aligned} \mathbf{x}_p^\mathsf{T}\tilde{\theta} &= \tilde{\mathbf{u}}_I + \tilde{\mathbf{v}}_J \\ &= \mathbf{u}_I + \mathbf{v}_J - \max_{0 \geq j \geq n} \frac{\mathbf{u}_I + \mathbf{v}_j - \lambda \mathbf{c}_p}{2} - \max_{0 \geq i \geq m} \frac{\mathbf{u}_i + \mathbf{v}_J - \lambda \mathbf{c}_p}{2} \\ &= \frac{\mathbf{u}_I + \mathbf{v}_J}{2} - \max_{0 \geq j \geq n} \frac{\mathbf{v}_j}{2} - \max_{0 \geq i \geq m} \frac{\mathbf{u}_i}{2} + \lambda \mathbf{c}_p \\ &= \frac{1}{2}\mathbf{x}_p^\mathsf{T}\theta - \max_{0 \geq j \geq n} \frac{\mathbf{v}_j}{2} - \max_{0 \geq i \geq m} \frac{\mathbf{u}_i}{2} + \lambda \mathbf{c}_p \\ &\leq \lambda \mathbf{c}_p \end{aligned} \tag{13}$$

As it holds for $\forall p$, $\tilde{\theta} \in \mathcal{R}^D$

C Proof of Theorem 7

We Generalize the problem as

$$\max_{\theta \in \mathcal{R}_I^S} \theta_{I_1} + \theta_{I_2} \tag{14}$$

Considering the center of the circle as θ^o, we define $\theta = \theta^o + q$, as $\theta^o_{I_1} + \theta^o_{I_2}$ is a constant, the problem is equal to $\max_{\theta \in \mathcal{R}_I^S} q_{I_1} + q_{I_2}$, we compute the Lagrangrian function of later:

$$\min_t \max_{\eta,\mu,\nu} L(q, \eta, \mu, \nu) = \min_t \max_{\eta,\mu,\nu} q_{I_1} + q_{I_2} + \eta(r - q^\mathsf{T} q) + \mu(a^\mathsf{T} q - e_a) + \nu(b^\mathsf{T} q - e_b) \tag{15}$$

$$\frac{\partial L}{\partial q_i} = \begin{cases} 1 - 2\eta q_i + \mu a_i + \nu b_i & i = I_1, I_2 \\ -2\eta q_i + \mu a_i + \nu b_i & i \neq I_1, I_2 \end{cases} \tag{16}$$

$${q_i}^* = \begin{cases} \dfrac{1 + \mu a_i + \nu b_i}{2\eta} & i = I_1, I_2 \\ \dfrac{\mu a_i + \nu b_i}{2\eta} & i \neq I_1, I_2 \end{cases} \tag{17}$$

We can get the Lagrangrian dual problem:

$$\max_{\eta,\mu,\nu} L(\eta,\mu,\nu) = \max_{\eta,\mu,\nu} \frac{1+\mu a_{I_1}+\nu b_{I_1}}{2\eta} + \frac{1+\mu a_{I_2}+\nu b_{I_2}}{2\eta} + \eta(r - {q^*}^\mathsf{T} q^*) + \mu(a^\mathsf{T} q^* - e_a) + \nu(b^\mathsf{T} q^* - e_b) \tag{18}$$

Compute the gradient, we can get three equality:

$$\begin{aligned}
\frac{\partial L}{\partial \eta} &= -\frac{1+\mu a_{I_1}+\nu I_1}{4\eta^2} - \frac{1+\mu a_{I_2}+\nu I_2}{4\eta^2} - 2\frac{\partial q^*}{\partial \eta} q^*(r - {q^*}^\mathsf{T} q^*) + \mu a \frac{\partial q^*}{\partial \eta} + + \nu b \frac{\partial q^*}{\partial \eta} \\
\frac{\partial L}{\partial \mu} &= \frac{a_{I_1}+a_{I_2}}{2\eta} - 2q^* \frac{\partial q^*}{\partial \mu} + (a^\mathsf{T} q^* - e_a)\frac{\partial q^*}{\partial \mu} + \nu b \frac{\partial q^*}{\partial \mu} \\
\frac{\partial L}{\partial \nu} &= \frac{b_{I_1}+b_{I_2}}{2\eta} - 2q^* \frac{\partial q^*}{\partial \nu} + (b^\mathsf{T} q^* - e_b)\frac{\partial q^*}{\partial \nu} + \mu a \frac{\partial q^*}{\partial \nu}
\end{aligned} \tag{19}$$

As