Supplementary material

A Proofs

A.1 Proof of Proposition 1

Proof. Let $(Q, R, g) \in \mathcal{C}(a, b, r, \alpha)$, $P := Q\,\mathrm{diag}(1/g)R^T$. Remarks that for all i, j,

$$\begin{aligned}\sqrt{\sum_{i',j'} |A_{i,i'} - B_{j,j'}|^2 P_{i',j'}} &\geq \left| \sqrt{\sum_{i',j'} |A_{i,i'}|^2 P_{i',j'}} - \sqrt{\sum_{i',j'} |B_{j,j'}|^2 P_{i',j'}} \right| \\ &\geq |\sqrt{\tilde{x}_i} - \sqrt{\tilde{y}_j}|\end{aligned}$$

Therefore we have

$$\begin{aligned}\sqrt{\sum_{i,i',j,j'} |A_{i,i'} - B_{j,j'}|^2 P_{i',j'} P_{i,j}} &= \sqrt{\sum_{i,j} \sum_{i',j'} |A_{i,i'} - B_{j,j'}|^2 P_{i',j'} P_{i,j}} \\ &\geq \sqrt{\sum_{i,j} |\sqrt{\tilde{x}_i} - \sqrt{\tilde{y}_j}|^2 P_{i,j}}\end{aligned}$$

Finally we obtain that

$$\sum_{i,i',j,j'} |A_{i,i'} - B_{j,j'}|^2 P_{i',j'} P_{i,j} - \varepsilon H(Q, R, g) \geq \sum_{i,j} |\sqrt{\tilde{x}_i} - \sqrt{\tilde{y}_j}|^2 P_{i,j} - \varepsilon H(Q, R, g)$$

and by taking the infimum over all $(Q, R, g) \in \mathcal{C}(a, b, r, \alpha)$, the results follows. □

A.2 Proof of Proposition 2

To show the result, we first need to recall some notions linked to the relative smoothness. Let $\mathcal{X}$ a closed convex subset in a Euclidean space $\mathbb{R}^q$. Given a convex function $H : \mathcal{X} \to \mathbb{R}$ continuously differentiable, one can define the *prox-function* associated to H as

$$D_H(x, z) := H(x) - H(z) - \langle \nabla H(z), x - z \rangle.$$

Let us now introduce the definition of the relative smoothness with respect the H.

Definition 2 (Relative smoothness.)**.** *Let $L > 0$ and f continuously differentiable on $\mathcal{X}$. f is said to be L-smooth relatively to H if*

$$f(y) \leq f(x) + \langle \nabla f(x), y - x \rangle + L D_H(y, x)$$

In [33], the authors show the following general result on the non-asymptotic stationary convergence of the mirror-descent scheme defined by the following recursion:

$$x_{k+1} = \operatorname*{argmin}_{x \in \mathcal{X}} \langle \nabla f(x_k), x \rangle + \frac{1}{\gamma_k} D_h(x, x_k)$$

where (γ_k) a sequence of positive step-size.

Proposition 3 ([33])**.** *Let $N \geq 1$, f continuously differentiable on $\mathcal{X}$ which is L-smooth relatively to H. By considering for all $k = 1, \ldots, N$, $\gamma_k = 1/2L$, and by denoting $D_0 = f(x_0) - \min_{x \in \mathcal{X}} f(x)$, we have*

$$\min_{0 \leq k \leq N-1} \Delta_k \leq \frac{4LD_0}{N}.$$

where for all $k = 1, \ldots, N$

$$\Delta_k := \frac{1}{\gamma_k^2}(D_H(x_k, x_{k+1}) + D_H(x_{k+1}, x_k)).$$

Let us now show that our objective function is relatively smooth with respect the the KL divergence [26, 43]. The result of Propostion 2 will then follow from Proposition 3. Here $\mathcal{X} = \mathcal{C}(a, b, r, \alpha)$, H is the negative entropy defined as

$$H(Q, R, g) := \sum_{i,j} Q_{i,j}(\log(Q_{i,j}) - 1) + \sum_{i,j} R_{i,j}(\log(R_{i,j}) - 1) + \sum_j g_j(\log(g_j) - 1),$$

and let us define for all $(Q, R, g) \in \mathcal{C}(a, b, r, \alpha)$

$$F_\varepsilon(Q, R, g) := -2\langle AQ\,\mathrm{diag}(1/g)R^T B, Q\,\mathrm{diag}(1/g)R^T\rangle + \varepsilon H(Q, R, g)\ .$$

Let us now show the following proposition.

Proposition 4. *Let $\varepsilon \geq 0$, $\frac{1}{r} \geq \alpha > 0$ and let us denote $L_{\varepsilon,\alpha} := 27(\|A\|_2\|B\|_2/\alpha^4 + \varepsilon)$. Then for all $(Q_1, R_1, g_1), (Q_2, R_2, g_2) \in \mathcal{C}(a, b, r, \alpha)$, we have*

$$\|\nabla F_\varepsilon(Q_1, R_1, g_1) - \nabla F_\varepsilon(Q_2, R_2, g_2)\|_2 \leq L_{\varepsilon,\alpha}\|H(Q_1, R_1, g_1) - H(Q_2, R_2, g_2)\|_2$$

Proof. Let $(Q, R, g) \in \mathcal{C}(a, b, r, \alpha)$ and let us denote $P = Q\,\mathrm{diag}(1/g)R^T$. We first have that

$$\nabla F_\varepsilon(Q, R, g) = (\nabla_Q F_\varepsilon(Q, R, g), \nabla_R F_\varepsilon(Q, R, g), \nabla_g F_\varepsilon(Q, R, g))$$

where

$$\begin{aligned}
\nabla_Q F_\varepsilon(Q, R, g) &:= -4APBR\,\mathrm{diag}(1/g) + \varepsilon \log Q \\
\nabla_R F_\varepsilon(Q, R, g) &:= -4BP^T AQ\,\mathrm{diag}(1/g) + \varepsilon \log R \\
\nabla_g F_\varepsilon(Q, R, g) &:= -4\mathcal{D}(Q^T APBR)/g^2 + \varepsilon \log g
\end{aligned}$$

First remarks that

$$\begin{aligned}
\|\nabla_Q F_\varepsilon(Q_1, R_1, g_1) - \nabla_Q F_\varepsilon(Q_2, R_2, g_2)\|_2 \leq\ & 4\|AP_1BR_1\,\mathrm{diag}(1/g_1) - AP_2BR_2\,\mathrm{diag}(1/g_2)\|_2 \\
& + \varepsilon\|\log Q_1 - \log Q_2\|_2\ .
\end{aligned}$$

Moreover we have

$$AP_1BR_1\,\mathrm{diag}(1/g_1) - AP_2BR_2\,\mathrm{diag}(1/g_2) = A((P_1 - P_2)BR_1\,\mathrm{diag}(1/g_1) + P_2B(R_1\,\mathrm{diag}(1/g_1) - R_2\,\mathrm{diag}(1/g_2))$$

where

$$P_1 - P_2 = (Q_1 - Q_2)\,\mathrm{diag}(1/g_1)R_1^T + Q_2(\mathrm{diag}(1/g_1)R_1^T - \mathrm{diag}(1/g_2)R_2^T)$$

and

$$R_1\,\mathrm{diag}(1/g_1) - R_2\,\mathrm{diag}(1/g_2) = (R_1 - R_2)\,\mathrm{diag}(1/g_1) + R_2(\mathrm{diag}(1/g_1) - \mathrm{diag}(1/g_2))$$

Finally we obtain that

$$\begin{aligned}
\|AP_1BR_1\,\mathrm{diag}(1/g_1) - AP_2BR_2\,\mathrm{diag}(1/g_2)\| \leq & \frac{\|A\|\|B\|}{\alpha}\left(\frac{\|Q_1 - Q_2\|}{\alpha} + \frac{\|R_1 - R_2\|}{\alpha} + \|1/g_1 - 1/g_2\|\right) \\
& + \|A\|\|B\|\left(\frac{\|R_1 - R_2\|}{\alpha} + \|1/g_1 - 1/g_2\|\right)\ .
\end{aligned}$$

As $Q \to H(Q)$ is 1-strongly convex w.r.t to the ℓ_2-norm on $\Delta_{n\times r}$, we have

$$\begin{aligned}
\|Q_1 - Q_2\|_2^2 &\leq \langle \log Q_1 - \log Q_2, Q_1 - Q_2\rangle \\
&\leq \|\log Q_1 - \log Q_2\|_2\|Q_1 - Q_2\|_2
\end{aligned}$$

from which follows that

$$\|Q_1 - Q_2\|_2 \leq \|\log Q_1 - \log Q_2\|_2.$$

Moreover we have

$$\|1/g_1 - 1/g_2\|_2 \leq \frac{\|g_1 - g_2\|_2}{\alpha^2} \leq \frac{\|\log g_1 - \log g_2\|_2}{\alpha^2}$$

Then we obtain that

$$\begin{aligned}\|\nabla_Q F_\varepsilon(Q_1,R_1,g_1)-\nabla_Q F_\varepsilon(Q_2,R_2,g_2)\|_2 \le &\left(\frac{4\|A\|\|B\|}{\alpha^2}+\varepsilon\right)\|\log Q_1-\log Q_2\|_2\\ &+(1+1/\alpha)\frac{4\|A\|\|B\|}{\alpha}\|\log R_1-\log R_2\|_2\\ &(1+1/\alpha)\frac{4\|A\|\|B\|}{\alpha^2}\|\log g_1-\log g_2\|_2\end{aligned}$$

Similarly we obtain that Then we obtain that

$$\begin{aligned}\|\nabla_R F_\varepsilon(Q_1,R_1,g_1)-\nabla_R F_\varepsilon(Q_2,R_2,g_2)\|_2 \le &\left(\frac{4\|A\|\|B\|}{\alpha^2}+\varepsilon\right)\|\log R_1-\log R_2\|_2\\ &+(1+1/\alpha)\frac{4\|A\|\|B\|}{\alpha}\|\log Q_1-\log Q_2\|_2\\ &(1+1/\alpha)\frac{4\|A\|\|B\|}{\alpha^2}\|\log g_1-\log g_2\|_2\end{aligned}$$

Moreover we have

$$\begin{aligned}\|\nabla_g F_\varepsilon(Q_1,R_1,g_1)-\nabla_g F_\varepsilon(Q_2,R_2,g_2)\|_2 \le &4\|\mathcal{D}(Q_1^TAP_1BR_1)/g_1^2-\mathcal{D}(Q_2^TAP_2BR_2)/g_2^2\|\\ &+\varepsilon\|\log g_1-\log g_2\|\end{aligned}$$

and

$$\begin{aligned}\mathcal{D}(Q_1^TAP_1BR_1)/g_1^2-\mathcal{D}(Q_2^TAP_2BR_2)/g_2^2 =&(1/g_1^2-1/g_2^2)\mathcal{D}(Q_1^TAP_1BR_1)\\ &+\frac{1}{g_2^2}(\mathcal{D}(Q_1^TAP_1BR_1)-\mathcal{D}(Q_2^TAP_2BR_2));.\end{aligned}$$

Note also that

$$\|(1/g_1^2-1/g_2^2)\mathcal{D}(Q_1^TAP_1BR_1)\| \le \frac{2\|A\|\|B\|}{\alpha^3}\|\log g_1-\log g_2\|$$

and

$$\begin{aligned}Q_1^TAP_1BR_1-Q_2^TAP_2BR_2 &= (Q_1^T-Q_2^T)AP_1BR_1+Q_2^TA(P_1BR_1-P_2BR_2)\\ &= (Q_1^T-Q_2^T)AP_1BR_1+Q_2^TA((P_1-P_2)BR_1+P_2B(R_1-R_2))\end{aligned}$$

from which follows that

$$\|\frac{1}{g_2^2}(\mathcal{D}(Q_1^TAP_1BR_1)-\mathcal{D}(Q_2^TAP_2BR_2))\| \le \frac{\|A\|\|B\|}{\alpha^2}\left(\|\log Q_1-\log Q_2\|+\|\log R_1-\log R_2\|+\|P_1-P_2\|\right)$$

and we obtain that

$$\begin{aligned}\|\nabla_g F_\varepsilon(Q_1,R_1,g_1)-\nabla_g F_\varepsilon(Q_2,R_2,g_2)\|_2 \le &\left(\frac{4\|A\|\|B\|}{\alpha^2}+\frac{1}{\alpha}\right)\|\log Q_1-\log Q_2\|\\ &+\left(\frac{4\|A\|\|B\|}{\alpha^2}+\frac{1}{\alpha}\right)\|\log R_1-\log R_2\|\\ &+\left(\frac{4\|A\|\|B\|}{\alpha^4}+\frac{8\|A\|\|B\|}{\alpha^3}+\varepsilon\right)\|\log g_1-\log g_2\|\end{aligned}$$

Finally we have

$$\begin{aligned}\|\nabla F_\varepsilon(Q_1,R_1,g_1)-\nabla F_\varepsilon(Q_2,R_2,g_2)\|_2^2 \le 3&\left[\left(\frac{4\|A\|\|B\|}{\alpha^2}+\varepsilon\right)^2+(1+1/\alpha)^2\frac{16\|A\|^2\|B\|^2}{\alpha^2}+\left(\frac{4\|A\|\|B\|}{\alpha^2}+\frac{1}{\alpha}\right)^2\right]\\ &\left(\|\log Q_1-\log Q_2\|^2+\|\log R_1-\log R_2\|^2\right)\\ &+3\left[2(1+1/\alpha)^2\frac{16\|A\|\|^2B\|^2}{\alpha^4}+\left(\frac{4\|A\|\|B\|}{\alpha^4}+\frac{8\|A\|\|B\|}{\alpha^3}+\varepsilon\right)^2\right]\\ &\|\log g_1-\log g_2\|^2\end{aligned}$$

from which we obtain that

$$\|\nabla F_\varepsilon(Q_1,R_1,g_1)-\nabla F_\varepsilon(Q_2,R_2,g_2)\|_2^2 \le L_{\varepsilon,\alpha}^2\left(\|\log Q_1-\log Q_2\|^2+\|\log R_1-\log R_2\|^2+\|\log g_1-\log g_2\|^2\right)$$

and the result follows. □

B Low-rank Approximation of Distance Matrices

Here we recall the algorithm used to perform a low-rank approximation of a distance matrix [5, 21]. We use the implementation of [33].

Algorithm 4 LR-Distance(X, Y, r, γ) [5, 21]

Inputs: X, Y, r, γ
Choose $i^* \in \{1, \dots, n\}$, and $j^* \{1, \dots, m\}$ uniformly at random.
For $i = 1, \dots, n$, $p_i \leftarrow d(x_i, y_j^*)^2 + d(x_i^*, y_j^*)^2 + \frac{1}{m}\sum_{j=1}^m d(x_i^*, y_j)^2$.
Independently choose $i^{(1)}, \dots, i^{(t)}$ according $(p_1, \dots, p_n)$.
$X^{(t)} \leftarrow [x_{i^{(1)}}, \dots, x_{i^{(t)}}]$, $P^{(t)} \leftarrow [\sqrt{t p_{i^{(1)}}}, \dots, \sqrt{t p_{i^{(t)}}}]$, $S \leftarrow d(X^{(t)}, Y)/P^{(t)}$
Denote $S = [S^{(1)}, \dots, S^{(m)}]$,
For $j = 1, \dots, m$, $q_j \leftarrow \|S^{(j)}\|_2^2 / \|S\|_F^2$
Independently choose $j^{(1)}, \dots, j^{(t)}$ according $(q_1, \dots, q_m)$.
$S^{(t)} \leftarrow [S^{j^{(1)}}, \dots, S^{j^{(t)}}]$, $Q^{(t)} \leftarrow [\sqrt{t q_{j^{(1)}}}, \dots, \sqrt{t q_{j^{(t)}}}]$, $W \leftarrow S^{(t)}/Q^{(t)}$
$U_1, D_1, V_1 \leftarrow \text{SVD}(W)$ (decreasing order of singular values).
$N \leftarrow [U_1(1), \dots, U_1^{(r)}]$, $N \leftarrow S^T N / \|W^T N\|_F$
Choose $j^{(1)}, \dots, j^{(t)}$ uniformly at random in $\{1, \dots, m\}$.
$Y^{(t)} \leftarrow [y_{j^{(1)}}, \dots, y_{j^{(t)}}]$, $D^{(t)} \leftarrow d(X, Y^{(t)})/\sqrt{t}$.
$U_2, D_2, V_2 = \text{SVD}(N^T N)$, $U_2 \leftarrow U_2 / D_2$, $N^{(t)} \leftarrow [(N^T)^{(j^{(1)})}, \dots, (N^T)^{(j^{(t)})}]$, $B \leftarrow U_2^T N^{(t)}/\sqrt{t}$, $A \leftarrow (BB^T)^{-1}$.
$Z \leftarrow AB(D^{(t)})^T$, $M \leftarrow Z^T U_2^T$
Result: M, N

C Nonnegative Low-rank Factorization of the Couplings

In this section, we recall the algorithm presented in [33] to solve problem (8) where we denote $p_1 := a$ and $p_2 := b$.

Algorithm 5 LR-Dykstra$((K^{(i)})_{1 \le i \le 3}, p_1, p_2, \alpha, \delta)$ [33]

Inputs: $K^{(1)}, K^{(2)}, \tilde{g} := K^{(3)}, p_1, p_2, \alpha, \delta, q_1^{(3)} = q_2^{(3)} = \mathbf{1}_r, \forall i \in \{1, 2\}$, $\tilde{v}^{(i)} = \mathbf{1}_r, q^{(i)} = \mathbf{1}_r$
repeat
  $u^{(i)} \leftarrow p_i / K^{(i)} \tilde{v}^{(i)}$ $\forall i \in \{1, 2\}$,
  $g \leftarrow \max(\alpha, \tilde{g} \odot q_1^{(3)})$, $q_1^{(3)} \leftarrow (\tilde{g} \odot q_1^{(3)})/g$, $\tilde{g} \leftarrow g$,
  $g \leftarrow (\tilde{g} \odot q_2^{(3)})^{1/3} \prod_{i=1}^2 (v^{(i)} \odot q^{(i)} \odot (K^{(i)})^T u^{(i)})^{1/3}$,
  $v^{(i)} \leftarrow g / (K^{(i)})^T u^{(i)}$ $\forall i \in \{1, 2\}$,
  $q^{(i)} \leftarrow (\tilde{v}^{(i)} \odot q^{(i)})/v^{(i)}$ $\forall i \in \{1, 2\}$, $q_2^{(3)} \leftarrow (\tilde{g} \odot q_2^{(3)})/g$,
  $\tilde{v}^{(i)} \leftarrow v^{(i)}$ $\forall i \in \{1, 2\}$, $\tilde{g} \leftarrow g$
until $\sum_{i=1}^2 \|u^{(i)} \odot K^{(i)} v^{(i)} - p_i\|_1 < \delta$;
$Q \leftarrow \text{diag}(u^{(1)}) K^{(1)} \text{diag}(v^{(1)})$
$R \leftarrow \text{diag}(u^{(2)}) K^{(2)} \text{diag}(v^{(2)})$
Result: Q, R, g

D Additional Experiements

D.1 Illustration

In Fig. 7, we show the time-accuracy tradeoffs of the two methods presented in Figure 1 on the same example. We see that our method, **Lin GW-LR**, manages to obtain similar accuracy as the one obtained by **Quad Entropic-GW** even when the rank $r = n/1000$ while being much faster with order of magnitude.

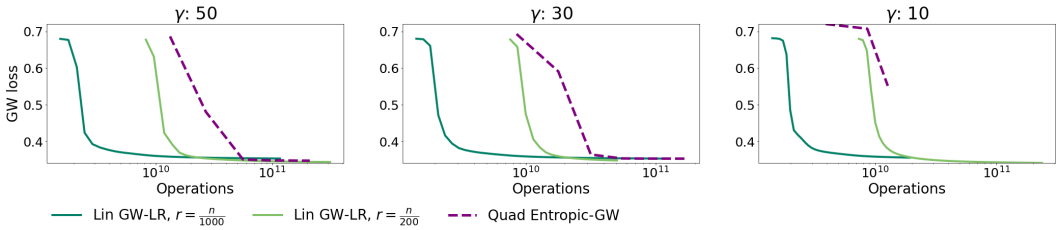

Figure 7: Here $n = m = 10000$, and the ground cost considered is the squared Euclidean distance. Note that for in that case we have an exact low-rank factorization of the cost. Therefore we compare only **Quad Entropic-GW** and **Lin GW-LR**. We plot the time-accuracy tradeoff when varying γ for multiple ranks r. $\varepsilon = 1/\gamma$ for **Quad Entropic-GW** and $\varepsilon = 0$ for **Lin GW-LR**.

D.2 Effect of ε and γ

In Fig. 2, we consider two Gaussian mixture densities samples with $n = m = 5000$ points in respectively 5D and 10D where

$$\mu_{\mathcal{X}}^{(1)} = [0, \dots, 0] \in \mathbb{R}^5,\ \mu_{\mathcal{X}}^{(2)} = [0, 1, 0, \dots, 0] \in \mathbb{R}^5,\ \mu_{\mathcal{X}}^{(3)} = [1, 1, 0, \dots, 0] \in \mathbb{R}^5,$$
$$\nu_{\mathcal{Y}}^{(1)} = [0.5, 0.5, 0, \dots, 0] \in \mathbb{R}^{10},\ \nu_{\mathcal{Y}}^{(2)} = [-0.5, 0.5, 0, \dots, 0] \in \mathbb{R}^{10},$$
$$\Sigma_{\mathcal{X}} = 0.05 \times \mathrm{Id}_5 \quad \text{and} \quad \Sigma_{\mathcal{Y}} = 0.05 \times \mathrm{Id}_{10}.$$

In Figure 8, we compare the time-accuracy tradeoff when varying ε and γ. We show that when $\varepsilon = 0$, the proposed method manage to consistently obtain the smallest GW loss whatever γ is. By allowing $\varepsilon > 0$, the algorithm is able to obtain a better time-accuracy tradeoff. However the choice of ε for the best time-accuracy tradeoff depends highly on γ.

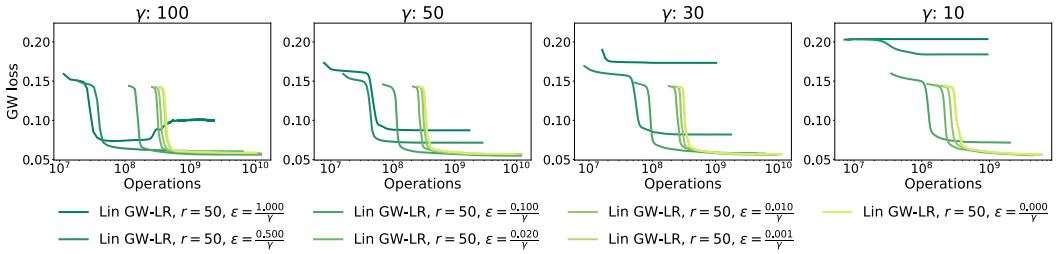

Figure 8: In this experiment, we consider the exact same setting as presented in Figure 2. The ground cost is the squared Euclidean distance. We consider only the linear version of our algorithm. The rank is fixed to be $r = n/100$. We plot the time-accuracy tradeoff when varying ϵ for multiple choices of γ.

D.3 Low-rank Problem

In Fig. 3 and 4, we consider two distributions in respectively 10-D and 15-D where the support is a concatenation of clusters of points. In Fig. 9, we show an illustration of the distributions considered in smaller dimensions.

In Fig. 10, 11, 10, we compare the time-accuracy tradeoffs of our method with the entropic one as proposed in [28] when the underlying cost is the Euclidean distance. The setting considered is the same as the one presented in Fig. 3 but we consider different number of clusters. In Fig. 13, 14, 13, we also compare the time-accuracy tradeoffs of our method with the entropic one when the underlying cost is the squared Euclidean distance. The setting is the same as the one presented in Fig. 4 but we consider different number of clusters. We show that our method consistenly manage to obtain similar accuracy as the entropic method while being orders of magnitude faster.

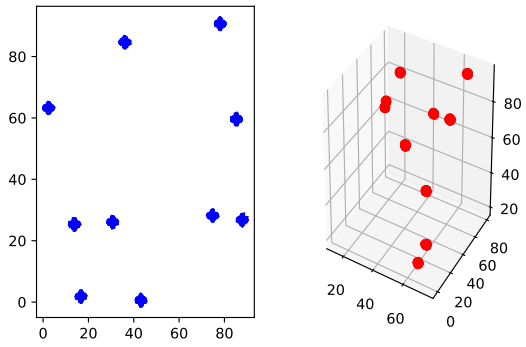

Figure 9: The source distribution and the target distribution live respectively in $\mathbb{R}^2$ and $\mathbb{R}^3$. Both distributions have the same number of samples $n = m = 10000$, the same number of clusters which is set to be 10 here, the same number of points in each cluster, and we force the distance between the centroids of the cluster to be larger than $\beta = 10$ in each distribution.

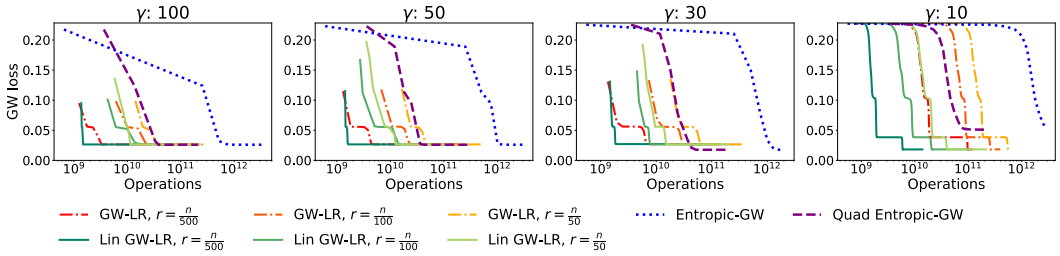

Figure 10: The number of clusters is set to be 5 and the underlying cost is the Euclidean distance.

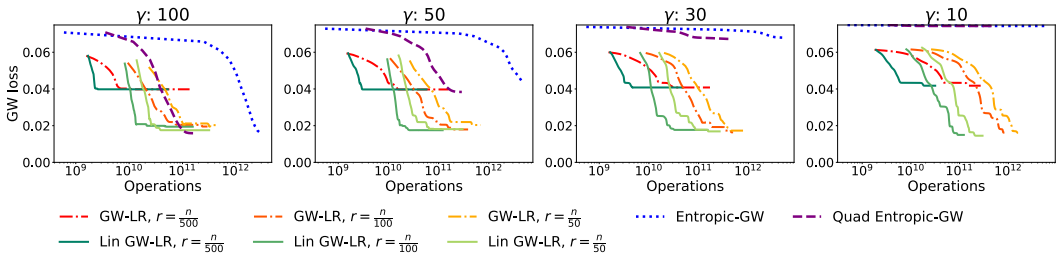

Figure 11: The number of clusters is set to be 20 and the underlying cost is the Euclidean distance.

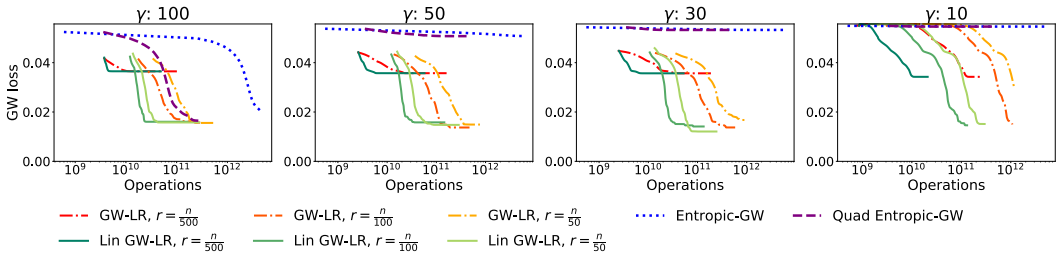

Figure 12: The number of clusters is set to be 30 and the underlying cost is the Euclidean distance.

D.4 Ground Truth Experiment

In this experiment we aim at comparing the different methods when the optimal coupling solving the GW problem has a full rank. For that purpose we consider a certain shape in 2-D which corresponds to the support of the source distribution and we apply two isometric transformations to it, which are a rotation and a translation to obtain the support the target distribution. See Figure 16 (*left*) for an illustration of the dataset. Here we set a and b to be uniform distributions and the underlying cost is

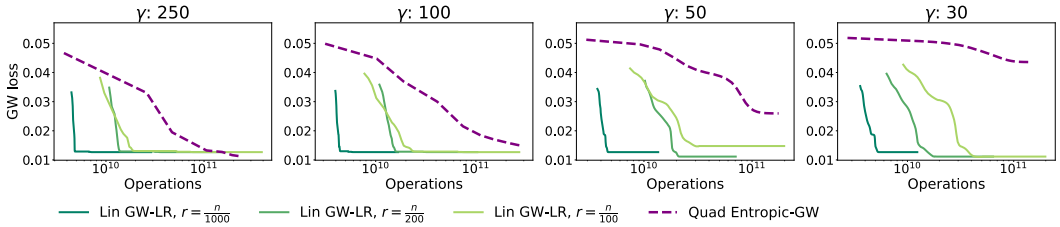

Figure 13: The number of clusters is set to be 10 and the underlying cost is the squared Euclidean distance.

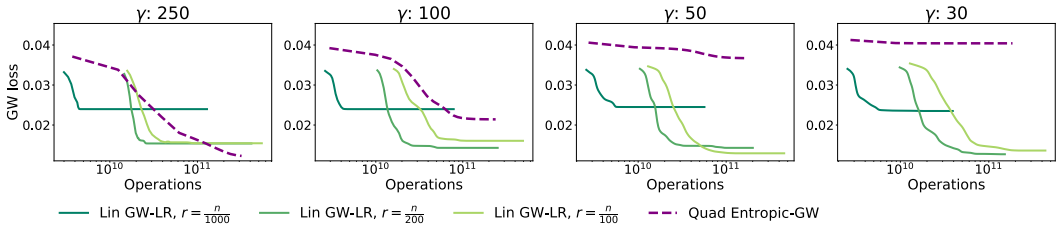

Figure 14: The number of clusters is set to be 20 and the underlying cost is the squared Euclidean distance.

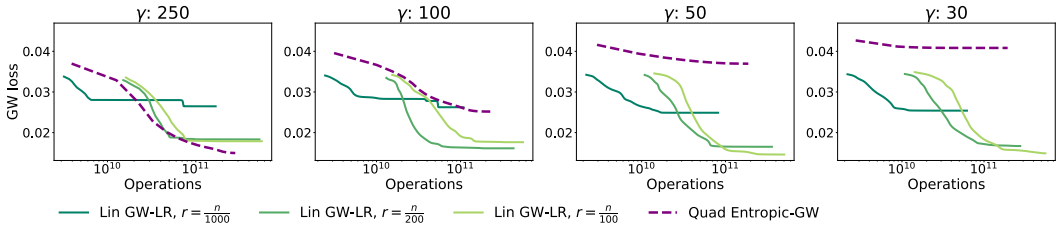

Figure 15: The number of clusters is set to be 30 and the underlying cost is the squared Euclidean distance.

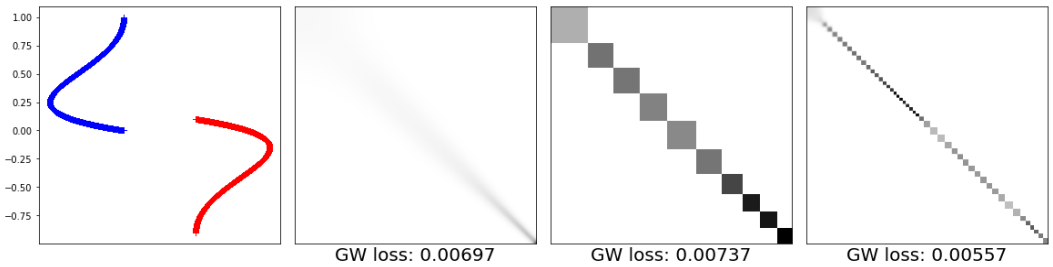

Figure 16: We compare the couplings obtained when the ground truth is the identity matrix in the same setting as in Figure 7. Here the comparison is done when $\gamma = 250$. *Left:* illustration of the dataset considered. *Middle left:* we show the coupling as well as the GW loss obtained by **Quad Entropic-GW**. *Middle right, right:* we show the couplings and the GW losses obtained by **Lin GW-LR** when the rank is respectively $r = 10$ and $r = 100$.

the squared Euclidean distance. Therefore the optimal coupling solution of the GW problem is the identity matrix and the GW loss must be 0. In Figure 17, we compare the time-accuracy tradeoffs, and we show that even in that case, our methods obtain a better time-accuracy tradeoffs for all γ. See also Figure 16 for a comparison of the couplings obtained by the different methods.

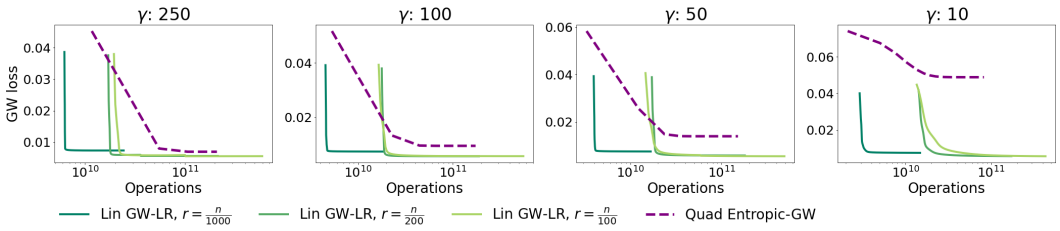

Figure 17: The ground truth here is the identity matrix and the true GW loss to achieve is 0. We set the number of samples to be $n = m = 10000$. As we consider the squared Euclidean distance, only **Quad Entropic-GW** and **Lin GW-LR** are compared. We plot the time-accuracy tradeoff when varying γ for multiple choices of rank r. $\varepsilon = 1/\gamma$ for **Quad Entropic-GW** and $\varepsilon = 0$ for **Lin GW-LR**.