Derivative of Gromov–Wasserstein

Danielle Ezuz,[1] Justin Solomon,[2] Vladimir G. Kim,[3] and Mirela Ben-Chen[1]

[1]Technion - Israel Institute of Technology
[2]MIT
[3]Adobe Research

Computation of Gromov–Wasserstein (GW) distances alternates between two steps, as specified in Algorithm 1. The algorithm is adjusted to our purpose by fixing the number of GW iterations and constraining v, w. The first step is a closed-form exponential formula applied independently to every element of the matrix variable. The other is projection onto the cone of doubly-stochastic matrices. We differentiate the result of GW distance computation by providing derivatives of the two steps independently and composing the formulas during alternation. The exponential is differentiable using formulaic techniques; we work out the derivative of doubly-stochastic projection below.

function GROMOV-WASSERSTEIN($\mu_0, D_0, \mu, D, \alpha, \eta, \Gamma^0$)
  // Computes a local minimizer Γ *of GW distance*
  for $i = 1, 2, \ldots, I$
    $K^i \leftarrow \exp(D_0 [\![\mu_0]\!] \Gamma^{i-1} [\![\mu]\!] D^\top \eta/\alpha) \otimes \left(\Gamma^i\right)^{\wedge(1-\eta)}$
    $\Gamma^i, v^i, w^i \leftarrow$ SINKHORN-PROJECTION($K^i; \mu_0, \mu$)
  return Γ^I,

function SINKHORN-PROJECTION($K; \mu_0, \mu$)
  // Finds Γ *minimizing* $KL(\Gamma|K)$ *subject to* $\Gamma \in \mathscr{M}(\mu_0, \mu)$
  $v, w \leftarrow \mathbf{1}$
  for $j = 1, 2, 3, \ldots$
    $v \leftarrow \mathbf{1} \oslash K(w \otimes \mu)$
    $w \leftarrow \mathbf{1} \oslash K^\top (v \otimes \mu_0)$
  $v \leftarrow v / \sqrt{\frac{\mathbf{1}^\top v}{\mathbf{1}^\top w}}$
  $w \leftarrow w \sqrt{\frac{\mathbf{1}^\top v}{\mathbf{1}^\top w}}$
  return $[\![v]\!] K [\![w]\!], v, w$

Algorithm 1: Iteration for finding regularized Gromov-Wasserstein distances. $\otimes, \oslash$ denote elementwise multiplication and division.

1 Derivative with respect to K_{ij}

Suppose we wish to rescale a kernel matrix K to be doubly stochastic (Sinkhorn projection step). We can think of this as solving the following quadratic system of equations for Γ and dual vectors v, w (we don't use superscript here for simplicity, we assume all the variables were computed at the same iteration):

$$\Gamma = [\![v]\!]\, K\, [\![w]\!] \tag{1}$$
$$\Gamma\mu = [\![v]\!]\, K(w \otimes \mu) = \mathbf{1} \tag{2}$$
$$\Gamma^\top \mu_0 = [\![w]\!]\, K^\top (v \otimes \mu_0) = \mathbf{1} \tag{3}$$
$$\mathbf{1}^\top v = \mathbf{1}^\top w \tag{4}$$

We differentiate these expressions with respect to an element K_{ij}. Then,

$$\frac{d\Gamma}{dK_{ij}} = \left[\!\!\left[\frac{dv}{dK_{ij}}\right]\!\!\right] K\, [\![w]\!] + v_i w_j (e_i e_j^\top) + [\![v]\!]\, K \left[\!\!\left[\frac{dw}{dK_{ij}}\right]\!\!\right] \tag{5}$$
$$0 = \frac{dv}{dK_{ij}} \otimes [K(w \otimes \mu)] + (v_i w_j \mu_j) e_i + [\![v]\!]\, K\, [\![\mu]\!]\, \frac{dw}{dK_{ij}} \tag{6}$$
$$0 = \frac{dw}{dK_{ij}} \otimes [K^\top (v \otimes \mu_0)] + (v_i w_j \mu_{0i}) e_j + [\![w]\!]\, K^\top [\![\mu_0]\!]\, \frac{dv}{dK_{ij}} \tag{7}$$
$$\mathbf{1}^\top \frac{dv}{dK_{ij}} = \mathbf{1}^\top \frac{dw}{dK_{ij}} \tag{8}$$

Let's organize the second two relationships as a matrix equation:

$$\begin{pmatrix} [\![K(w \otimes \mu)]\!] & [\![v]\!]\, K\, [\![\mu]\!] \\ [\![w]\!]\, K^\top [\![\mu_0]\!] & [\![K^\top (v \otimes \mu_0)]\!] \\ \mathbf{1}^\top & -\mathbf{1}^\top \end{pmatrix} \begin{pmatrix} dv/dK_{ij} \\ dw/dK_{ij} \end{pmatrix} = \begin{pmatrix} -v_i w_j \mu_j e_i \\ -v_i w_j \mu_{0i} e_j \\ 0 \end{pmatrix} \tag{9}$$

We can simplify this by leveraging the fact that the diagonal elements of this block 2×2 matrix appear in the Sinkhorn conditions. In particular, $K(w \otimes \mu) = \mathbf{1} \oslash v$ and $K^\top (v \otimes \mu_0) = \mathbf{1} \oslash w$. Hence, we can write the expression as:

$$\begin{pmatrix} [\![v]\!]^{-1} & [\![v]\!]\, K\, [\![\mu]\!] \\ [\![w]\!]\, K^\top [\![\mu_0]\!] & [\![w]\!]^{-1} \\ \mathbf{1}^\top & -\mathbf{1}^\top \end{pmatrix} \begin{pmatrix} dv/dK_{ij} \\ dw/dK_{ij} \end{pmatrix} = \begin{pmatrix} -v_i w_j \mu_j e_i \\ -v_i w_j \mu_{0i} e_j \\ 0 \end{pmatrix} \tag{10}$$

We factor to make it look symmetric:

$$\begin{pmatrix} [\![v]\!] & 0 & 0 \\ 0 & [\![w]\!] & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} [\![v \otimes v \otimes \mu_0]\!]^{-1} & K \\ K^\top & [\![w \otimes w \otimes \mu]\!]^{-1} \\ (\mathbf{1} \oslash \mu 0)^\top & -(\mathbf{1} \oslash \mu)^\top \end{pmatrix} \begin{pmatrix} [\![\mu_0]\!] & 0 \\ 0 & [\![\mu]\!] \end{pmatrix} \begin{pmatrix} dv/dK_{ij} \\ dw/dK_{ij} \end{pmatrix} = \begin{pmatrix} -v_i w_j \mu_j e_i \\ -v_i w_j \mu_{0i} e_j \\ 0 \end{pmatrix} \tag{11}$$

The two vectors on the right-hand side are sparse except element i in the first half and element j in the second half. So, we can invert the first matrix and put the diagonal matrix into the unknowns:

$$\begin{pmatrix} [\![v \otimes v \otimes \mu_0]\!]^{-1} & K \\ K^\top & [\![w \otimes w \otimes \mu]\!]^{-1} \\ (\mathbf{1} \oslash \mu 0)^\top & -(\mathbf{1} \oslash \mu)^\top \end{pmatrix} \begin{pmatrix} [\![\mu_0]\!]\, {}^{dv}\!/_{dK_{ij}} \\ [\![\mu]\!]\, {}^{dw}\!/_{dK_{ij}} \end{pmatrix} = \begin{pmatrix} -w_j \mu_j e_i \\ -v_i \mu_{0i} e_j \end{pmatrix} \tag{12}$$

Denote the left inverse of the left matrix as

$$\begin{pmatrix} [\![v \otimes v \otimes \mu_0]\!]^{-1} & K \\ K^\top & [\![w \otimes w \otimes \mu]\!]^{-1} \\ (\mathbf{1} \oslash \mu 0)^\top & -(\mathbf{1} \oslash \mu)^\top \end{pmatrix}^+ := \begin{pmatrix} A & B \\ C & E \end{pmatrix} \tag{13}$$

Then,

$$\begin{pmatrix} {}^{dv}\!/_{dK_{ij}} \\ {}^{dw}\!/_{dK_{ij}} \end{pmatrix} = \begin{pmatrix} [\![\mathbf{1} \oslash \mu_0]\!] & 0 \\ 0 & [\![\mathbf{1} \oslash \mu]\!] \end{pmatrix} \begin{pmatrix} A & B \\ C & E \end{pmatrix} \begin{pmatrix} -w_j \mu_j e_i \\ -v_i \mu_{0i} e_j \end{pmatrix} \tag{14}$$

$$= \begin{pmatrix} [\![\mathbf{1} \oslash \mu_0]\!] & 0 \\ 0 & [\![\mathbf{1} \oslash \mu]\!] \end{pmatrix} \begin{pmatrix} -w_j \mu_j A_{\text{column } i} - v_i \mu_{0i} B_{\text{column } j} \\ -w_j \mu_j C_{\text{column } i} - v_i \mu_{0i} E_{\text{column } j} \end{pmatrix} \tag{15}$$

2 Gradient with respect to K

Returning to our original derivative, we extract a single element

$$\frac{d\Gamma_{k\ell}}{dK_{ij}} = \frac{dv_k}{dK_{ij}} K_{k\ell} w_\ell + v_i w_j \delta_{ik} \delta_{j\ell} + v_k K_{k\ell} \frac{dw_\ell}{dK_{ij}} \tag{16}$$

In the end, we know $\frac{dL}{d\Gamma_{k\ell}}$ for some function L and want $\frac{dL}{dK_{ij}}$. Write the gradient of L with respect to Γ as $\nabla_\Gamma L$. We start computing

$$\begin{aligned} \frac{dL}{dK_{ij}} &= \sum_{k\ell} (\nabla_\Gamma L)_{k\ell} \frac{d\Gamma_{k\ell}}{dK_{ij}} \\ &= \sum_{k\ell} (\nabla_\Gamma L)_{k\ell} \left[\frac{dv_k}{dK_{ij}} K_{k\ell} w_\ell + v_i w_j \delta_{ik} \delta_{j\ell} + v_k K_{k\ell} \frac{dw_\ell}{dK_{ij}} \right] \end{aligned}$$

Breaking this down term by term,

$$\begin{aligned} \sum_{k\ell} (\nabla_\Gamma L)_{k\ell} \frac{dv_k}{dK_{ij}} K_{k\ell} w_\ell &= \left(\frac{dv}{dK_{ij}} \right)^\top (K \otimes \nabla_\Gamma L) w \\ &= \left(-w_j \mu_j A_{\text{column } i} - v_i \mu_{0i} B_{\text{column } j} \right)^\top [\![\mathbf{1} \oslash \mu_0]\!] (K \otimes \nabla_\Gamma L) w \\ \sum_{k\ell} (\nabla_\Gamma L)_{k\ell} v_i w_j \delta_{ik} \delta_{j\ell} &= (\nabla_\Gamma L)_{ij} v_i w_j \end{aligned}$$

$$\sum_{k\ell}(\nabla_\Gamma L)_{k\ell} v_k K_{k\ell} \frac{dw_\ell}{dK_{ij}} = v^\top (K\otimes\nabla_\Gamma L)\frac{dw}{K_{ij}}$$
$$= v^\top (K\otimes\nabla_\Gamma L) [\![\mathbf{1}\oslash\mu]\!] \left(-w_j\mu_j C_{\text{column } i} - v_i\mu_{0i}E_{\text{column } j}\right)$$

Getting rid of the ij index (and applying symmetry of A and C) shows

$$\boxed{\begin{aligned}\nabla_K L = \quad & -A [\![\mathbf{1}\oslash\mu_0]\!] (K\otimes\nabla_\Gamma L) w (w\otimes\mu)^\top - (v\otimes\mu_0)[(C [\![\mathbf{1}\oslash\mu_0]\!] (K\otimes\nabla_\Gamma L)w]^\top \\ & + [\![v]\!] \nabla_\Gamma L [\![w]\!] \\ & -B [\![\mathbf{1}\oslash\mu]\!] (K\otimes\nabla_\Gamma L)^\top v (w\otimes\mu)^\top - (v\otimes\mu_0)[E [\![\mathbf{1}\oslash\mu]\!] (K\otimes\nabla_\Gamma L)^\top v]^\top\end{aligned}} \tag{17}$$

3 Exponential formula

Given $\nabla_{K^i}L$ we would like to compute $\nabla_{\Gamma^{i-1}}L$, the derivatives with respect to Γ from the previous iteration. In the rest of this section K will be used for K^i and Γ for Γ^{i-1}.

$$K = exp\left(D_0 [\![\mu_0]\!] \Gamma [\![\mu]\!] D^\top \cdot \eta/\alpha\right) \otimes \Gamma^{\wedge 1-\eta} \tag{18}$$

$$\frac{dL}{d\Gamma_{ij}} = \sum_{kl}\frac{dL}{dK_{kl}}\frac{dK_{kl}}{d\Gamma_{ij}} \tag{19}$$

$$\frac{dK_{kl}}{d\Gamma_{ij}} = \delta_{i==k}\delta_{j==l}\,(1-\eta)\,(\Gamma)_{kl}^{-\eta}\exp\left(D_0 [\![\mu_0]\!] \Gamma [\![\mu]\!] D^\top \cdot \eta/\alpha\right)_{kl} + \tag{20}$$
$$\exp\left(D_0 [\![\mu_0]\!] \Gamma [\![\mu]\!] D^\top \cdot \eta/\alpha\right)_{kl}\frac{\eta}{\alpha}[D_0 [\![\mu_0]\!]]_{ki}\left[[\![\mu]\!] D^\top\right]_{jl}\Gamma^{\wedge 1-\eta}$$

$$\nabla_\Gamma L = (1-\eta)\,\nabla_K L\otimes(\Gamma)^{\wedge-\eta}\otimes\exp\left(D_0 [\![\mu_0]\!] \Gamma [\![\mu]\!] D^\top \cdot \eta/\alpha\right) + \tag{21}$$
$$\frac{\eta}{\alpha}\nabla_K L\otimes\left([\![\mu_0]\!] D_0^\top\left(\Gamma^{\wedge 1-\eta}\otimes\exp\left(D_0 [\![\mu_0]\!] \Gamma [\![\mu]\!] D^\top \cdot \eta/\alpha\right)\right) D [\![\mu]\!]\right)$$

4 Gradient with respect to D

Given $\nabla_{K^i}L$, we can compute $\nabla_D L$ by cumulating the following gradients throughout the iterations:

$$\frac{dK_{kl}}{dD_{ij}} = \Gamma_{kl}^{1-\eta} exp\left(D_0 [\![\mu_0]\!] \Gamma [\![\mu]\!] D^\top \cdot \eta/\alpha\right)_{kl}\delta_{i==l}\frac{\eta}{\alpha}[D_0 [\![\mu_0]\!] \Gamma [\![\mu]\!]]_{kj} \tag{22}$$

$$\frac{dL}{dD_{ij}} = \sum_k \frac{dL}{dK_{ki}}\frac{\eta}{\alpha}\Gamma_{ki}^{1-\eta} exp\left(D_0 [\![\mu_0]\!] \Gamma [\![\mu]\!] D^\top \cdot \eta/\alpha\right)_{ki}[D_0 [\![\mu_0]\!] \Gamma [\![\mu]\!]]_{kj} \tag{23}$$

$$\nabla_D L = \frac{\eta}{\alpha}\left(\nabla_K L\otimes\Gamma^{\wedge 1-\eta}\otimes exp\left(D_0 [\![\mu_0]\!] \Gamma [\![\mu]\!] D^\top \cdot \eta/\alpha\right)\right)^\top D_0 [\![\mu_0]\!] \Gamma [\![\mu]\!] \tag{24}$$

We also add the transposed matrix of derivatives to enforce symmetry.