Formation Shape Control using the Gromov-Wasserstein Metric

Haruto Nakashima   NAKASHIMA.HARUTO.72V@ST.KYOTO-U.AC.JP
Siddhartha Ganguly   GANGULY.SIDDHARTHA.7P@KYOTO-U.AC.JP
Kohei Morimoto   KOHEI.MORIMOTO.73R@ST.KYOTO-U.AC.JP
Kenji Kashima   KK@I.KYOTO-U.AC.JP
Graduate School of Informatics, Kyoto University, Kyoto, Japan

Editors: N. Ozay, L. Balzano, D. Panagou, A. Abate

Abstract

This article introduces a formation shape control algorithm, in the optimal control framework, for steering an initial population of agents to a desired configuration via employing the Gromov-Wasserstein distance. The underlying dynamical system is assumed to be a constrained linear system and the objective function is a sum of quadratic control-dependent stage cost and a Gromov-Wasserstein terminal cost. The inclusion of the Gromov-Wasserstein cost transforms the resulting optimal control problem into a well-known NP-hard problem, making it both numerically demanding and difficult to solve with high accuracy. Towards that end, we employ a recent semi-definite relaxation-driven technique to tackle the Gromov-Wasserstein distance. A numerical example is provided to illustrate our results.

Keywords: Formation shape control, Optimal control, Optimal transport, Gromov-Wasserstein distance

1. Introduction

The Gromov-Wasserstein (GW) distance (Mémoli (2011)) is an important and popular optimal transport (OT) distance which is often employed to compare distributions whose underlying spaces are *incomparable*, as opposed to the classical OT distances, for example, the Wasserstein distance. The GW distance recently gained ample amount of attention in various scientific communities ranging from machine learning Xu et al. (2019a) to mathematical finance Hendrikson (2016) and system biology Demetci et al. (2020). The GW distance is invariant under isometric transformations, such as translation, rotation, and reflection, and thus it is especially useful in applications where the underlying metric geometry, for example, shape or formation, needs to be preserved; see Alvarez-Melis and Jaakkola (2018), Solomon et al. (2016), Mémoli (2009), Xu et al. (2019b).

This article establishes a formation/shape control algorithm in the minimum energy optimal control framework by employing the GW distance and leveraging its invariance and metric properties. We outline our primary motivation:

1. Traditional control approaches, such as leader-follower or distance-based methods Dimarogonas and Johansson (2008); Deghat et al. (2016), rely on specifying and pre-assigning precise target positions or inter-agent distances within a fixed reference frame. While these methods work well for simple, static formations, they are often rigid and struggle to adapt to dynamic environments or objectives that do not require strict positional accuracy. This rigidity makes them less effective in scenarios where the overall shape or configuration of the agents is essential, while the distances between each pair of agents are of secondary importance. In contrast, the GW distance allows for comparing distributions based on their relative geometries, making it invariant to translations, reflections, and rotations.

2. The invariance properties of the GW distance can indeed be useful in the context of controlling the shape of given formation, where, for example, for an ensemble of homogeneous agents (for example, a multi-robotic swarm), the goal may be to create a specific shape, such as a circle, line, or more complex configuration, without concern for the precise orientation or placement of the formation. For example, consider a scenario where a group of drones is required to spread out into a circular formation for optimal coverage in a search-and-rescue mission. Using conventional control techniques, each drone would need to follow precise, fixed coordinates or inter-agent distances, which can be cumbersome and inflexible if the environment changes or if the entire formation needs to rotate or translate. In the existing literature on formation control of ensemble systems such as Mosteo et al. (2017); Chen et al. (2017), only configurations with fixed orientations have been considered, while the optimization over distance-preserving transformations, such as rotation and translation, has not been addressed. On the other hand, in our formulation, incorporating GW distance as the objective function in formation control adds flexibility because the distance itself accounts only for the shape similarity between two distributions rather than for absolute positions and orientations. This means that agents can achieve the desired formation shape with minimal dependency on external coordinate frames or absolute positioning. This can be viewed as a simultaneous optimization of the dynamical steering and the configuration of agents.

With this motivation, we introduce the GW distance in the context of the multi-agent formation control problem of linear discrete-time dynamical systems from a constrained optimal control perspective. Our approach treats the formation control problem as a *controlled optimal transport* problem which offers us a systematic way to compare and align multi-agent configurations by utilizing the inherent geometric structure of the agent formations. The GW-optimization problem (see (2) ahead) is nonconvex and presents significant computational challenges. To address this, we employ a semidefinite programming (SDP) relaxation of the GW-distance, developed in Chen et al. (2024), which enables an accurate and computationally tractable approach for solving the optimal transport problem. Furthermore, under certain conditions on the solutions of the SDP, we can achieve *globally optimal solutions* to the original GW problem (2) unlike conditional gradient based methods Titouan et al. (2019) that searches for locally optimal solutions or entropy regularization-based techniques Peyré et al. (2016) which does not solve the original problem and instead solve an approximate problem, minimizing a different regularized objective.

The remainder of the paper is organized as follows: §2 introduces the concept of GW distance, §3 formulates the problem as a minimum-energy optimal control problem. In Section §4, we present our primary findings, including the SDP relaxation-based algorithm. We also provide several results that ensure the existence of optimizers for specific optimal control problems. We demonstrate our results through a numerical example in §5. Our findings and results suggest that there is ample scope for improvement, both theoretically and numerically, with several potential directions for future research. These possibilities are discussed in detail in §6.

2. Background on the Gromov-Wasserstein distance

We fix some notations first. For us $\mathbb{N}$ is the set of natural numbers and $\mathbb{N}^* := \mathbb{N} \setminus \{0\} = \{1, 2, \ldots\}$ denote the set of positive integers. For any $n \in \mathbb{N}^*$, we define the finite set of positive integers up to n by $[\![n]\!] := \{1, \ldots, n\}$. Given $d \in \mathbb{N}^*$, let $\mathbb{R}^d_{>0}$ and $\mathbb{R}^d_{\geq 0}$ denote the sets of positive and

nonnegative vectors in the Euclidean space $\mathbb{R}^d$, respectively. For any $x \in \mathbb{R}^d$ and for any positive definite matrix $M \in \mathbb{R}^{d \times d}$, we denote the standard quadratic form $\langle x, Mx \rangle$ by $\|x\|_M^2$; when M is not necessarily positive (semi) definite we employ $\langle x, Mx \rangle := x^\top M x$. Let $d_1, d_2 \in \mathbb{N}^*$ and consider the vector space of all $d_1 \times d_2$ matrices with real entries; we denote a inner product of such matrices by $(A_1, A_2) \mapsto \langle A_1, A_2 \rangle := \mathrm{tr}(A_2^\top A_1)$. We denote the standard probability simplex by $\Sigma_d := \{\zeta \in \mathbb{R}_{\succeq 0}^d \mid \sum_{i=1}^d \zeta_i = 1\}$, and $\mathbf{1}_N \in \mathbb{R}^N$ represents the vector with all entries equal to 1. The standard Euclidean basis vector whose i-th entry is 1 is denoted by e_i.

Let $\mu_{\mathcal{X}}(\cdot)$ and $\mu_{\mathcal{Y}}(\cdot)$ be two given probability density functions specified on metric spaces $\mathcal{X}$ and $\mathcal{Y}$ with respective metrics $d_{\mathcal{X}} : \mathcal{X} \times \mathcal{X} \longrightarrow \mathbb{R}$ and $d_{\mathcal{Y}} : \mathcal{Y} \times \mathcal{Y} \longrightarrow \mathbb{R}$. Define the set of probability measures with marginals $\mu_{\mathcal{X}}(\cdot)$ and $\mu_{\mathcal{Y}}(\cdot)$ by $\Pi(\mu_{\mathcal{X}}, \mu_{\mathcal{Y}}) := \{\pi(x, y) \mid \int_{\mathcal{Y}} \pi(x, y) \, \mathrm{d}y = \mu_{\mathcal{X}}(x), \int_{\mathcal{X}} \pi(x, y) \, \mathrm{d}x = \mu_{\mathcal{Y}}(y)\}$. Then for $p \geqslant 1$, the GW distance, GW_p is defined by:

$$\mathsf{GW}_p\big(\mu_{\mathcal{X}}, \mu_{\mathcal{Y}}\big)^p := \inf_{\pi \in \Pi(\mu_{\mathcal{X}}, \mu_{\mathcal{Y}})} \int_{\mathcal{X}^2 \times \mathcal{Y}^2} \big|d_{\mathcal{X}}(x, x') - d_{\mathcal{Y}}(y, y')\big|^p \, \mathrm{d}\pi(x, y)\mathrm{d}\pi(x', y'). \quad (1)$$

We refer the readers to Mémoli (2011) for an in-depth analysis of the geometric and metric properties of $\mathsf{GW}(\cdot, \cdot)$. For computational purposes, we focus on the setting of *discrete* metric-measure spaces. To this end, fix natural numbers $n, m \geqslant 1$, and let $\mathsf{a} := (\mathsf{a}_1, \dots, \mathsf{a}_n)^\top \in \Sigma_n$ and $\mathsf{b} := (\mathsf{b}_1, \dots, \mathsf{b}_m)^\top \in \Sigma_m$. On $(\mathcal{X}, d_{\mathcal{X}})$ and $(\mathcal{Y}, d_{\mathcal{Y}})$ define the discrete probability measures $\mu_{\mathcal{X}} := \sum_{i=1}^n \mathsf{a}_i \delta_{x_i}$ and $\mu_{\mathcal{Y}} := \sum_{j=1}^m \mathsf{b}_j \delta_{y_j}$. Given two non-negative cost matrices $C^{\mathcal{X}} \in \mathbb{R}^{n \times n}$ and $C^{\mathcal{Y}} \in \mathbb{R}^{m \times m}$ associated with the discrete metric-measure spaces $(C^{\mathcal{X}}, \mathsf{a})$ and $(C^{\mathcal{Y}}, \mathsf{b})$, define the set of couplings by $\Pi(\mathsf{a}, \mathsf{b}) := \{P = [P_{ij}] \in \mathbb{R}^{n \times m} \mid P\mathbf{1}_m = \mathsf{a}, P^\top \mathbf{1}_n = \mathsf{b}\}$.

Consider a smooth *loss* function $\mathbb{R} \times \mathbb{R} \ni (\xi, \mu) \mapsto g(\xi, \mu)$, where $g\big(C_{i,j}^{\mathcal{X}}, C_{k,l}^{\mathcal{Y}}\big)$ measures the difference between the metrics $d_{\mathcal{X}}(x^i, x^{i'})$ and $d_{\mathcal{Y}}(y^j, y^{j'})$ for $1 \leqslant i, i' \leqslant n$ and $1 \leqslant j, j' \leqslant m$. The values of $g(\cdot, \cdot)$ can be put in a tensor $G \in \mathbb{R}^{mn \times mn}$ where $G := [G_{iji'j'}] = g\big(C_{i,i'}^{\mathcal{X}}, C_{j,j'}^{\mathcal{Y}}\big)$. With these ingredients, the discrete analogue of (1) is given by

$$\mathsf{GW}\big(C^{\mathcal{X}}, C^{\mathcal{Y}}, \mathsf{a}, \mathsf{b}\big) := \inf_{P \in \Pi(\mathsf{a}, \mathsf{b})} g\big(C_{i,i'}^{\mathcal{X}}, C_{j,j'}^{\mathcal{Y}}\big) P_{i,j} P_{i',j'} = \inf_{P \in \Pi(\mathsf{a}, \mathsf{b})} \sum_{i,j,i',j'} G_{iji'j'} P_{i,j} P_{i',j'}. \quad (2)$$

Defining the vectorization of P by $\mathrm{vec}(P) := [P_{11}, P_{21}, \dots, P_{m1}, \dots, P_{mn}]^\top \in \mathbb{R}^{mn}$ we rewrite (2) as

$$\mathsf{GW}\big(C^{\mathcal{X}}, C^{\mathcal{Y}}, \mathsf{a}, \mathsf{b}\big) = \inf_{P \in \Pi(\mathsf{a}, \mathsf{b})} \langle \mathrm{vec}(P), G\mathrm{vec}(P) \rangle, \quad (3)$$

which is a nonconvex NP-hard optimization problem and computationally intractable in general. We note that if the distributions are assumed to be Gaussian, the expression in (1) admits a closed-form solution; see Delon et al. (2022) and Morimoto and Kashima (2024) for an application. While approximation-driven methods Ryner et al. (2023) and entropic regularization techniques Peyré et al. (2016) offer computationally feasible ways to solve (3), they either depend on loss-specific algorithms or lack a certificate of global optimality. In the following section, we adopt an SDP-relaxation approach introduced in Chen et al. (2024) (which provides a certificate of global optimality, contingent on satisfying a specific condition; see §3, Remark 3, Equation (12)) to numerically solve (3).

3. Problem formulation

Fix $N \in \mathbb{N}^*$ and consider an N-agent discrete-time control system given by the recursion

$$x_{t+1}^i = Ax_t^i + Bu_t^i, \quad x_0^i = x_{\text{init}}^i \text{ given}, \ t \in \mathbb{N}, \text{ and } i \in [\![N]\!], \tag{4}$$

along with the following data:

((4)-a) for each $i \in [\![N]\!]$, the state and control input vectors at time t are denoted by x_t^i and u_t^i, respectively. Moreover, for each $i \in [\![N]\!]$ and $t \in \{0, \ldots, T-1\}$, we have $x_t^i \in \mathbb{X} \subset \mathbb{R}^d$ and $u_t^i \in \mathbb{U} \subset \mathbb{R}^m$, where $\mathbb{X}$ and $\mathbb{U}$ are compact and convex sets with $0 \in \mathbb{R}^d$ and $0 \in \mathbb{R}^m$ in their respective interiors.

((4)-b) At the final time T, we impose $x_T \in \mathbb{X}_F \subset \mathbb{R}^d$, where $\mathbb{X}_F$ is a compact and convex set containing $0 \in \mathbb{R}^d$ in its interior.

((4)-c) $A \in \mathbb{R}^{d \times d}$ and $B \in \mathbb{R}^{d \times m}$ are respectively the system and the control matrices which are, for simplicity, assumed to be the same for all the agents, and the (A, B) pair is stabilizable.

For a fixed number of agents $N \in \mathbb{N}^*$ and a fixed horizon $T \in \mathbb{N}^*$ we consider steering the initial point clouds $x_{\text{init}}^i \in \mathbb{R}^d$ to the same shape as the target point group $x_\star^i \in \mathbb{R}^d$, for each $i \in [\![N]\!]$, in an energy-efficient manner. Define the set of initial and final states for the agents by $x_{\text{init}} := (x_{\text{init}}^1, \ldots, x_{\text{init}}^N)$ and $x_d := (x_T^1, \ldots, x_T^N)$, and let $x_\star := (x_\star^1, \ldots, x_\star^N)$. Consider the N-agent, T-horizon, minimum energy optimal control problem (OCP):

$$\inf_{x_d} \inf_{(u_t^i)_{t=0}^{T-1}} \quad \sum_{i=0}^{N} \sum_{t=0}^{T-1} \|u_t^i\|_R^2 + \mathsf{GW}(x_d, x_\star)$$

$$\text{s.t.} \quad \begin{cases} x_{t+1}^i = Ax_t^i + Bu_t^i & \text{for each } t = 0, \ldots, T-1, \text{ and } i \in [\![N]\!], \\ x_0^i = x_{\text{init}}^i, \text{ and } x_T^i = x_d^i, \ x_T^i \in \mathbb{X}_F, \\ x_t^i \in \mathbb{X}, \ u_t^i \in \mathbb{U} & \text{for each } t = 0, \ldots, T-1, \text{ and } i \in [\![N]\!], \end{cases} \tag{5}$$

where $R \in \mathbb{R}_{\succ 0}^m$ is the control weighting matrix, and for compact notation in our context, we denote the GW distance in (3) by $\mathsf{GW}(x_d, x_\star)$. The OCP (5) can be compactly written as:

$$\inf_{x_d} c(x_{\text{init}}, x_d) + \mathsf{GW}(x_d, x_\star), \tag{6}$$

where quadratic control cost is given by

$$c(x_{\text{init}}, x_d) := \inf_{(u_t^i)_{t=0}^{T-1}} \quad \sum_{i=0}^{N} \sum_{t=0}^{T-1} \|u_t^i\|_R^2$$

$$\text{s.t.} \quad \begin{cases} x_{t+1}^i = Ax_t^i + Bu_t^i & \text{for each } t = 0, \ldots, T-1, \text{ and } i \in [\![N]\!], \\ x_0^i = x_{\text{init}}^i, \text{ and } x_T^i = x_d^i, \ x_T^i \in \mathbb{X}_F, \\ x_t^i \in \mathbb{X}, \ u_t^i \in \mathbb{U} & \text{for each } t = 0, \ldots, T-1, \text{ and } i \in [\![N]\!], \end{cases} \tag{7}$$

and for a fixed x_d, the GW-cost is given by (recall the notation established in §2)

$$\mathsf{GW}(x_d, x_\star) := \inf_{P \in \Pi\left(\frac{\mathbf{1}_N}{N}, \frac{\mathbf{1}_N}{N}\right)} g\big(C_{i,i'}^{\mathcal{X}}, C_{j,j'}^{\mathcal{Y}}\big) P_{i,j} P_{i',j'}$$

$$= \inf_{P \in \Pi\left(\frac{\mathbf{1}_N}{N}, \frac{\mathbf{1}_N}{N}\right)} \langle \text{vec}(P), \, G(x_d, x_\star) \text{vec}(P) \rangle, \tag{8}$$

4

where $g(\cdot, \cdot)$ is a loss function that accounts for possible discrepancies between $C^{\mathcal{X}}_{i,i'} := \big(d_{\mathcal{X}}(x^i_d, x^{i'}_d)\big)_{i,i'}$, and $C^{\mathcal{Y}}_{j,j'} := \big(d_{\mathcal{Y}}(x^j_\star, x^{j'}_\star)\big)_{j,j'}$ and $G(x_d, x_\star) = g\big(C^{\mathcal{X}}_{i,i'}, C^{\mathcal{Y}}_{j,j'}\big)$ captures the $N^2 \times N^2$ distance matrix as per the notation established in §2. Note that, for us $\mathsf{a} = \mathsf{b} = \frac{\mathbf{1}_N}{N}$ and thus P belongs to the set of all couplings $\Pi\big(\frac{\mathbf{1}_N}{N}, \frac{\mathbf{1}_N}{N}\big)$. In the sequel, we will denote the optimizer of (8) by P^*. The optimal control problem (OCP) in (6) can be understood as determining the destination for each agent such that the states are optimally guided to resemble the target shape x_{target} with minimal control cost.

Remark 1 *The OCP* (7) *is a standard constrained quadratic program which, for a fixed x_{init} and x_d, under the problem data* ((4)-a)–((4)-c), *admits a unique solution. When both $\mathcal{X}$ and $\mathcal{Y}$ are d-dimensional Euclidean spaces, $(\xi, \mu) \mapsto g(\xi, \mu) := \frac{1}{2}|\xi - \mu|^2$, and $d_{\mathcal{X}}(\cdot, \cdot)$ and $d_{\mathcal{Y}}(\cdot, \cdot)$ are quadratic Euclidean functions, G in* (8) *reduces to*

$$G(x_d, x_\star)_{ii'jj'} := \frac{1}{2}\left|d_{\mathcal{X}}(x^i_d, x^{i'}_d) - d_{\mathcal{Y}}(x^j_\star, x^{j'}_\star)\right|^2, \tag{9}$$

and the corresponding optimization problem is known as a quadratic assignment *problem Ryner et al. (2023), which is still an NP-hard problem. Our formulation, based on the SDP-relaxation technique established in Chen et al. (2024) does not depend on a specific choice of $g(\cdot, \cdot)$.*

4. Main results

The steering problem (8) is inherently NP-hard and challenging to solve numerically. To address this, we build on the approach in Chen et al. (2024) to develop a tight SDP relaxation for (8), enabling globally optimal solutions in particular cases.

4.1. SDP relaxation of (8)

Recall that for a given x_d, the GW-problem is given by

$$\mathsf{GW}(x_d, x_\star) := \inf_{P \in \Pi\big(\frac{\mathbf{1}_N}{N}, \frac{\mathbf{1}_N}{N}\big)} \langle \text{vec}(P), G(x_d, x_\star)\text{vec}(P)\rangle$$

where $G(x_d, x_\star) := g\big(d_{\mathcal{X}}(x^i_d, x^{i'}_d), d_{\mathcal{Y}}(x^j_\star, x^{j'}_\star)\big)$ for each $i, i', j, j' \in [\![N]\!]$. By representing the quadratic term with a rank-one positive semi-definite matrix and reformulating the linear marginal equalities in vectorized form, we derive a relaxed SDP of the following structure:

$$\inf_{(P, \widehat{P}) \in \mathbb{R}^{N \times N} \times \mathbb{R}^{N^2 \times N^2}} \quad \left\langle G(x_d, x_\star), \widehat{P}\right\rangle$$

$$\text{s. t.} \quad \begin{cases} \begin{pmatrix} \widehat{P} & \text{vec}(P) \\ \text{vec}(P)^\top & 1 \end{pmatrix} \succeq 0, \ \widehat{P} \succeq 0 \\ P \in \Pi\big(\frac{1}{N}\mathbf{1}_N, \frac{1}{N}\mathbf{1}_N\big), \\ \widehat{P}\,\text{vec}(e_i\mathbf{1}_N^\top) = \frac{1}{N}\text{vec}(P), \\ \widehat{P}\,\text{vec}(\mathbf{1}_N e_j^\top) = \frac{1}{N}\text{vec}(P) \end{cases} \tag{10}$$

One can interpret the relaxation (10) as a Lagrangian dual formulation of the GW problem (8), where additional constraints are introduced to connect the linear and quadratic components of transportation plans. We refer the readers to (Chen et al., 2024, §3) for more details on the relaxation procedure and (Chen et al., 2024, Appendix B) for duality results.

Proposition 2 *Consider the GW problem (8) and its SDP relaxation (10). Then (10) admits a global optimal point over its feasible set.*

Proof For a fixed x_d, denote the set of all feasible pairs $(P, \widehat{P}) \in \mathbb{R}^{N \times N} \times \mathbb{R}^{N^2 \times N^2}$ for the SDP by

$$\mathcal{F}_{x_d} := \left\{ (P, \widehat{P}) \left| \begin{array}{l} \begin{pmatrix} \widehat{P} & \mathrm{vec}(P) \\ \mathrm{vec}(P)^\top & 1 \end{pmatrix} \succeq 0, \ P \in \Pi\left(\frac{1}{N}\mathbf{1}_N, \frac{1}{N}\mathbf{1}_N\right), \ P \succeq 0 \\ \widehat{P}\,\mathrm{vec}(e_i \mathbf{1}_N^\top) = \frac{1}{N}\,\mathrm{vec}(P), \ \widehat{P}\,\mathrm{vec}(\mathbf{1}_N e_j^\top) = \frac{1}{N}\,\mathrm{vec}(P) \end{array} \right. \right\}.$$

Then, the SDP can be compactly written as

$$\inf_{(P,\widehat{P}) \in \mathbb{R}^{N \times N} \times \mathbb{R}^{N^2 \times N^2}} \quad \left\langle G(x_d, x_\star), \widehat{P} \right\rangle \tag{11}$$

$$\text{s.\,t.} \qquad (P, \widehat{P}) \in \mathcal{F}_{x_d}$$

Observe that:

- The set $\Pi\left(\frac{1}{N}\mathbf{1}_N, \frac{1}{N}\mathbf{1}_N\right)$ forms a convex and compact polytope.

- The set of matrices $\widehat{P} \in \mathbb{R}^{N^2 \times N^2}$ that satisfy the linear constraints $\widehat{P}\,\mathrm{vec}(e_i\mathbf{1}_N^\top) = \frac{1}{N}\,\mathrm{vec}(P)$ and $\widehat{P}\,\mathrm{vec}(\mathbf{1}_N e_j^\top) = \frac{1}{N}\,\mathrm{vec}(P)$ have a trace of at most one. Thus the feasible set $\mathcal{F}_{x_d}$ consists of positive semi-definite matrices with trace at most one which is closed and bounded and thus compact. Moreover $\mathcal{F}_{x_d}$ is convex due to the linearity of trace.

- Finally, the map $(P, \widehat{P}) \mapsto \left\langle G(x_d, x_\star), \widehat{P} \right\rangle$ is continuous.

Thus, solutions to (10) exist by Weierstrass's Theorem (Güler, 2010, Theorem 2.2). ■

Remark 3 (Approximation ratio and strong duality) *The SDP relaxation (10) of the original GW problem (8) provides a framework for deriving a certificate of optimality, which can confirm whether a solution to the relaxed problem is also optimal for the original problem. To this end, let $P^* \in \Pi\left(\frac{\mathbf{1}_N}{N}, \frac{\mathbf{1}_N}{N}\right)$ denote an optimizer of the original GW problem (8), and let $(P_r^*, \widehat{P}_r^*) \in \mathcal{F}_{x_d}$ represent the optimizer of the relaxed problem (10). By construction, $P_r^* \in \Pi\left(\frac{\mathbf{1}_N}{N}, \frac{\mathbf{1}_N}{N}\right)$, indicating that $(P_r^*, \widehat{P}_r^*)$ is feasible for (8) as well. Consequently, we obtain the inequality $\left\langle \widehat{P}_r^*, G(x_d, x_\star) \right\rangle \leqslant \left\langle \mathrm{vec}(P_r^*), G(x_d, x_\star)\mathrm{vec}(P_r^*) \right\rangle$. This implies that if the ratio*

$$\frac{\left\langle \mathrm{vec}(P_r^*), G(x_d, x_\star)\mathrm{vec}(P_r^*) \right\rangle}{\left\langle \widehat{P}_r^*, G \right\rangle} = 1, \tag{12}$$

then P_r^* is a globally optimal *solution to (8) Chen et al. (2024). In §5, we will observe that in our numerical experiments, this ratio remains very close to one. Moreover, one can obtain* exact solution *if the optimizers $(P_r^*, \widehat{P}_r^*)$ satisfies the rank condition:*

$$\mathrm{rank}\begin{pmatrix} \widehat{P}_r^* & \mathrm{vec}(P_r^*) \\ \mathrm{vec}(P_r^*)^\top & 1 \end{pmatrix} = 1. \tag{13}$$

The condition (13) guarantees that the strong duality holds, i.e., duality gap between the SDP (10) and its dual is zero; see (Chen et al., 2024, Proposition B.3). We will also demonstrate in our numerical example that the condition (13) holds.

Using the SDP relaxation (10), we aim to solve the relaxed optimal control problem (OCP):

$$\inf_{x_d \in \mathbb{X}_F} \quad J(x_d) \coloneqq c(x_{\text{init}}, x_d) + \mathsf{GWSDP}\left(x_d, x_\star\right), \tag{14}$$

where $\mathsf{GWSDP}\left(x_d, x_\star\right)$ corresponds to the relaxation (10).

Theorem 4 *Consider the OCP (14) with the problem data ((4)-a)–((4)-c) and assume that it is strictly feasible. Then, (14) admits a globally optimal solution over its feasible set.*

Proof Note that, x_d is constrained to the compact set $\mathbb{X}_F$ and observe that:

- For each fixed x_d, in the OCP (7), the map $\mu \mapsto \langle \mu, \, R\mu \rangle$ is convex and the constraint sets $\mathbb{X}, \mathbb{U}$, and $\mathbb{X}_F$ are all convex sets, and the Slater condition hold.

- For each fixed x_d, in (10), the set $\mathcal{F}_{x_d}$ is convex; see Proposition 2, and the function $\widehat{P} \mapsto \left\langle G(x_d, x_\star), \, \widehat{P} \right\rangle$ is convex. Moreover, (10) is strictly feasible.

Thus, the maps $x_d \mapsto c(x_{\text{init}}, x_d,)$ and $x_d \mapsto \mathsf{GWSDP}(x_d, x_\star)$ are lower semicontinuous (Still, 2018, Lemma 5.7). By a version of Weierstrass's Theorem, the existence of a solution follows (Santambrogio, 2023, Box 1.3, pp.9–10). ∎

Remark 5 *The choice of an algorithm to solve (14) depends on the smoothness and convexity properties of the objective function. In the absence of additional assumptions on the problem data, $J(\cdot)$ is generally nonsmooth and nonconvex with respect to the decision variable. Several algorithms are available for tackling such problems, including interior point methods Karmitsa et al. (2008); Schmidt (2015), proximal-like methods Jia and Grimmer (2024), and projective penalty methods Norkin (2024). In fact, with $\mathbb{X}_F$ convex, the OCP (14) admits the same form given in (Norkin, 2024, §3), and the penalty method established therein is exact (Norkin, 2024, Theorem 3), which can be utilized to solve (14). These approaches ensure that the sequence of iterates converges to a stationary point, which may correspond to a KKT point or a Fritz-John point depending on the constraint qualification conditions. A comprehensive analysis of the structural and regularity properties of $J(\cdot)$, along with the design of tailored algorithms to solve (14), will be the subject of future work.*

5. Numerical experiments

We consider a constrained formation shape control problem with 10 agents which follows the linear dynamics

$$x_{t+1}^i = \begin{pmatrix} 0.5 & 0.2 \\ 0.1 & 0.4 \end{pmatrix} x_t^i + \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} u_t^i \quad \text{for } i \in [\![10]\!] \text{ and } t \in \mathbb{N}^*. \tag{15}$$

The constraint sets are given by $\mathbb{X} \coloneqq \{\xi \mid \|\xi\|_\infty \leqslant 20\}$, $\mathbb{U} \coloneqq \{\mu \mid \|\mu\|_\infty \leqslant 20\}$ and $\mathbb{X}_F \coloneqq \{\eta \mid \|\eta\|_\infty \leqslant 20\}$. Fix $T = 10$ and consider the OCP:

$$\inf_{x_d} \inf_{(u_t)_{t=0}^{T-1}} \quad \sum_{i=0}^{N} \sum_{t=0}^{T-1} \varepsilon \left\|u_t^i\right\|_{I_2}^2 + \mathsf{GW}\left(x_d, x_\star\right)$$

$$\text{s. t.} \quad \begin{cases} \text{dynamics (15)}, \ x_t^i \in \mathbb{X}, \ u_t^i \in \mathbb{U} \text{ for each } i \in [\![10]\!], \ t \in \{0, \ldots, T-1\}, \\ \text{and } x_d \in \mathbb{X}_F \end{cases} \tag{16}$$

where $I_2 \in \mathbb{R}^{2 \times 2}$ is the identity matrix, and $\varepsilon > 0$ is a parameter that we employ to strike a balance between the control cost and GW distance. In the GW cost, we employed $g(x, y) \coloneqq \frac{1}{2} |x - y|^2$.

As the agents' initial states, we randomly selected ten points from the set $\{(x, y) \in [-15, -12] \times [-2, 2]\}$ and fixed a target formation given in Figure 2 (see the left-hand subfigure in the first row) which our agents need to mimic at the final time. We employed Julia v1.11.1 Bezanson et al. (2017) and its `JuMP` library to perform all our numerical computations and optimization tasks. To convert the GW distance in (8) we performed the relaxation given in (10) and the ensuing SDP was solved using MOSEK MOSEK ApS (2019). To solve the intermediate OCP (7), which is a convex quadratic program, we employed MOSEK as well. Finally, we employed the interior-point solver IPOPT Wächter and Biegler (2006) to solve the problem (14). The tolerance of MOSEK for solving both the problems (7) and (10) was set at 10^{-5}; the tolerance for the interior-point algorithm was set at 10^{-6}.

Figure 1: Approximation ratio

We calculated the approximation ratio for 10 different runs using the optimizers $(P_r^*, \widehat{P}_r^*)$ of the SDP to check the accuracy of the solution and we found that, in average $\frac{\langle \text{vec}(P_r^*), G(x_d, x_\star) \text{vec}(P_r^*) \rangle}{\langle \widehat{P}_r^*, G(x_d, x_\star) \rangle} \approx 1$, which indicates that $\widehat{P}_r^*$ is a globally optimal solution to (8) and our method offers a high-quality approximation of the optimal transport solution for the GW distance; see Figure 1. Consequently, the final configuration of the agents nearly mirrors the target formation; see Figure 2 (the right-hand subfigure in the first row, and the subfigure in the second row). Moreover, we observed that the optimizers $(P_r^*, \widehat{P}_r^*)$ satisfies the rank condition (13) which indicates that strong duality holds for this example. As noted in Chen et al. (2024), while the SDP is solvable in polynomial time, computational efficiency could be further improved, for example, using sparsification of the SDP solutions. We will explore this in our future investigations. We also record the control cost and the cost associated with solving the SDP problem in Table 1 for some ε in (16). Table 1 highlights a trade-off between the overall control cost and the GW distance between the target and final formation.

The previous numerical illustration presented a case where both $d_{\mathcal{X}}(\cdot, \cdot)$ and $d_{\mathcal{Y}}(\cdot, \cdot)$ were standard Euclidean distances. However, the flexibility of the GW distance allows us to adopt different underlying metrics. For example, let us consider a scenario in which 6 agents are to be separated into 2 groups in the final state, each consisting of 3 agents (the numbers can be adjusted as long as they are meaningful). To this end, we fixed fully connected graph structure as given in Figure 3 (left-hand subfigure) where we assigned weights to all pairs of edges in the graph in the following manner:

$$d_{\text{graph}}(i, i^{'}) \coloneqq \begin{cases} 2 & \text{if } \{i, i'\} \subset \{1, 2, 3\} \text{ or } \{i, i'\} \subset \{4, 5, 6\}, \\ 4 & \text{otherwise.} \end{cases} \tag{17}$$

To achieve the desired grouping, we assign smaller weights to edges connecting nodes within the same group, and larger weights to edges that span across groups. Note that d_{graph} is a metric. With this premise, we considered the following GW distance:

$$\text{GW}_C(x_d) \coloneqq \inf_{P \in \Pi\left(\frac{\mathbf{1}_N}{N}, \frac{\mathbf{1}_N}{N}\right)} \langle \text{vec}(P), G_C(x_d) \text{vec}(P) \rangle, \tag{18}$$
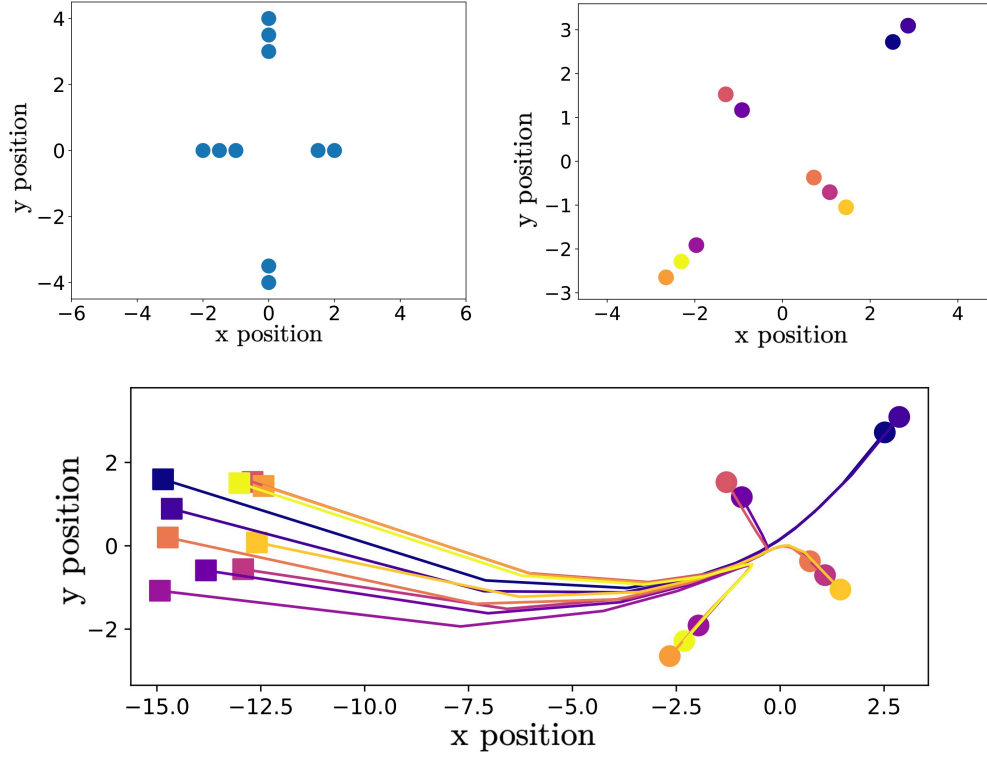
8
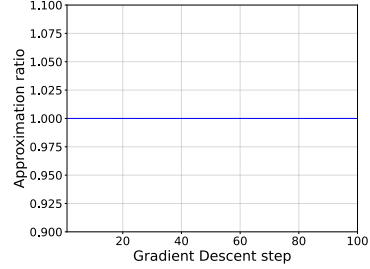
Figure 2: In the top row, the left-hand and right-hand subfigures depict the target shape and the final shape, respectively. The bottom row depicts the randomly chosen initial points (squares) and their trajectories up to the final formation (circles), with $\varepsilon = 0.5$.

where $G_C(x_d) := g\big(d_{\text{graph}}(i, i'), d_{\mathcal{Y}}(x_d^j, x_d^{j'})\big)$ for each $i, i', j, j' \in [\![N]\!]$; and $g(\cdot, \cdot)$ is same as before. Finally, to achieve separation of agents, we consider the following problem.

$$\inf_{x_d \in \mathbb{X}_F} c(x_{\text{init}}, x_d) + \mathsf{GW}_C(x_d), \tag{19}$$

where $c(x_{\text{init}}, x_d)$ takes care of the control cost while $\mathsf{GW}_C(x_d)$ accounts for the separation of agents. Consequently, we employed the same SDP relaxation strategy given in §4 for GW_C and solved (19) with exactly the same problem data and solver specifications.

ε	Control cost	GW distance
0.1	59.82	0.02
0.5	51.63	0.50
1.0	48.12	2.51
2.5	40.60	15.21

Table 1: Trade-off between control cost and GW distance between agents' state and target formation, with 10 agents. Observe that increasing ε leads to a lower control cost; however, this comes at the expense of a higher GW cost, resulting in a misaligned formation.

6. Discussion and future work

We developed an SDP-relaxation approach for multi-agent shape control, guided by the Gromov-Wasserstein (GW) distance, to achieve a target formation within a constrained, minimum-energy optimal control framework. Drawing on insights from Chen et al. (2024), our algorithm provides an optimality certificate, ensuring the exactness and global optimality of the SDP solutions. While
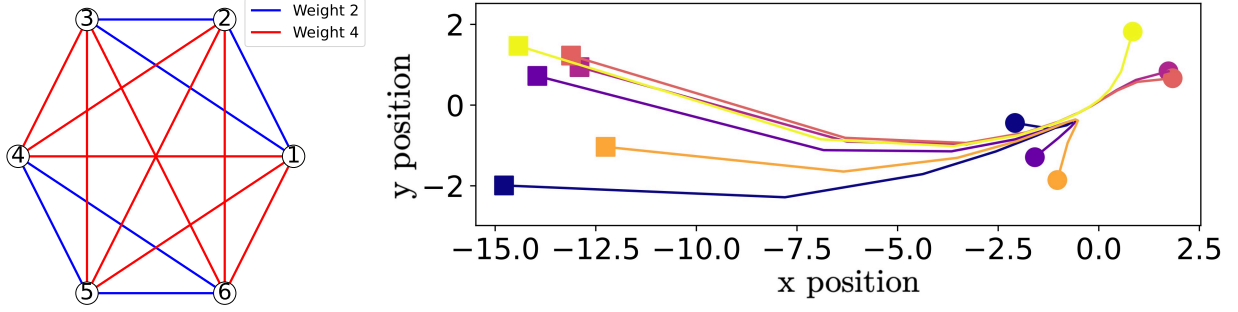
Figure 3: The left-hand subfigure depicts the target graph structure considered in (17), the right-hand subfigures shows the trajectory achieved, transitioning from squares to circles, with $\varepsilon = 0.5$. Note that in the final shape, the agents are successfully separated into 2 groups, respecting the graph structure. This is ensured by the term $\text{GW}_c(x_d)$.

the NP-hardness associated with the original GW problem is mitigated by this relaxation, several challenges remain, opening up various avenues for theoretical and numerical enhancements:

- Enhancing the computational efficiency of our algorithm by introducing sparsification techniques to the SDP solutions. By selectively reducing the complexity of the SDP structure while retaining essential features, it may be possible to lower the computational burden without compromising accuracy.

- Incorporating entropy regularization into our framework, combined with sparsification-driven Sinkhorn-like algorithms, to address an approximate version of the original Gromov-Wasserstein (GW) problem. This approach could leverage the computational advantages of entropy-regularized optimal transport while maintaining a balance between precision and efficiency, and consequently could enable faster and more scalable solutions.

- Investigating scenarios where external disturbances influence the system and designing a robust model predictive control (MPC) algorithm to stabilize the system while maintaining the desired formation. In such cases, the use of an unbalanced GW cost Sejourne et al. (2021) would be more appropriate, as it allows for greater flexibility in handling mismatched distributions and system perturbations. This extension could enable the algorithm to adapt to practical challenges and ensure reliable performance in real-world applications.

Acknowledgments

References

D. Alvarez-Melis and T. S. Jaakkola. Gromov-Wasserstein alignment of word embedding spaces. In Ellen Riloff, David Chiang, Julia Hockenmaier, and Jun'ichi Tsujii, editors, *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 1881–1890, Brussels, Belgium, October-November 2018. Association for Computational Linguistics. doi: 10.18653/v1/D18-1214. URL https://aclanthology.org/D18-1214.

J. Bezanson, A. Edelman, S. Karpinski, and V. B. Shah. Julia: A fresh approach to numerical computing. *SIAM review*, 59(1):65–98, 2017. doi: https://doi.org/10.1137/141000671. URL https://julialang.org/.

J. Chen, B. T. Nguyen, S. S. Koh, and Y. S. Soh. Semidefinite relaxations of the Gromov-Wasserstein distance. *Advances in Neural Information Processing Systems*, 2024. URL https://doi.org/10.48550/arXiv.2312.14572.

Y. Chen, T. T. Georgiou, and M. Pavon. Optimal transport over a linear dynamical system. *IEEE Transactions on Automatic Control*, 62(5):2137–2152, 2017. URL https://doi.org/10.1109/TAC.2016.2602103.

M. Deghat, B. D. O. Anderson, and Z. Lin. Combined flocking and distance-based shape control of multi-agent formations. *IEEE Transactions on Automatic Control*, 61(7):1824–1837, 2016. URL https://doi.org/10.1109/TAC.2015.2480217.

Julie Delon, Agnes Desolneux, and Antoine Salmona. Gromov–wasserstein distances between gaussian distributions. *Journal of Applied Probability*, 59(4):1178–1198, 2022. URL https://doi.org/10.1017/jpr.2022.16.

P. Demetci, R. Santorella, B. Sandstede, W. S. Noble, and R. Singh. Gromov-Wasserstein optimal transport to align single-cell multi-omics data. *BioRxiv*, pages 2020–04, 2020. URL https://doi.org/10.1101/2020.04.28.066787.

D. V. Dimarogonas and K. H. Johansson. On the stability of distance-based formation control. In *2008 47th IEEE Conference on Decision and Control*, pages 1200–1205, 2008. URL https://doi.org/10.1109/CDC.2008.4739215.

O. Güler. *Foundations of Optimization*, volume 258 of *Graduate Texts in Mathematics*. Springer, New York, 2010. ISBN 978-0-387-34431-7. doi: 10.1007/978-0-387-68407-9. URL https://doi.org/10.1007/978-0-387-68407-9.

R. Hendrikson. Using Gromov-Wasserstein distance to explore sets of networks. *University of Tartu, Master Thesis*, 2016. URL https://core.ac.uk/download/pdf/79113246.pdf.

Z. Jia and B. Grimmer. First-order methods for nonsmooth nonconvex functional constrained optimization with or without slater points, 2024. URL https://arxiv.org/abs/2212.00927.

N. M. S. Karmitsa, M. M. Mäkelä, and M. M. Ali. Limited memory interior point bundle method for large inequality constrained nonsmooth minimization. *Applied Mathematics and Computation*, 198(1):382–400, 2008. URL https://doi.org/10.1016/j.amc.2007.08.044.

F. Mémoli. Spectral Gromov-Wasserstein distances for shape matching. In *2009 IEEE 12th International Conference on Computer Vision Workshops, ICCV Workshops*, pages 256–263. IEEE, 2009. URL https://doi.org/10.1109/ICCVW.2009.5457690.

F. Mémoli. Gromov-Wasserstein distances and the metric approach to object matching. *Foundations of Computational Mathematics*, 11:417–487, 2011. URL https://doi.org/10.1007/s10208-011-9093-5.

K. Morimoto and K. Kashima. Minimum energy density steering of linear systems with Gromov–Wasserstein terminal cost. *IEEE Control Systems Letters*, 8:586–591, 2024. URL https://doi.org/10.1109/LCSYS.2024.3397228.

MOSEK ApS. The MOSEK optimization toolbox for MATLAB manual. version 9.0., 2019. url: http://docs.mosek.com/9.0/toolbox/index.html.

A. R. Mosteo, E. Montijano, and D. Tardioli. Optimal role and position assignment in multi-robot freely reachable formations. *Automatica*, 81:305–313, 2017. ISSN 0005-1098. doi: https://doi.org/10.1016/j.automatica.2017.03.040. URL https://www.sciencedirect.com/science/article/pii/S0005109817301711.

V. Norkin. The exact projective penalty method for constrained optimization. *Journal of Global Optimization*, 89(2):259–276, 2024. URL https://doi.org/10.1007/s10898-023-01350-4.

G. Peyré, M. Cuturi, and J. Solomon. Gromov-Wasserstein averaging of kernel and distance matrices. In M. F. Balcan and K. Q. Weinberger, editors, *Proceedings of The 33rd International Conference on Machine Learning*, volume 48 of *Proceedings of Machine Learning Research*, pages 2664–2672, New York, New York, USA, 20–22 Jun 2016. PMLR. URL https://proceedings.mlr.press/v48/peyre16.html.

M. Ryner, J. Kronqvist, and J. Karlsson. Globally solving the Gromov-Wasserstein problem for point clouds in low dimensional Euclidean spaces. In A. Oh, T. Naumann, A. Globerson, K. Saenko, M. Hardt, and S. Levine, editors, *Advances in Neural Information Processing Systems*, volume 36, pages 7930–7946. Curran Associates, Inc., 2023. URL https://proceedings.neurips.cc/paper_files/paper/2023/file/188409d2ad91db4fb13644d024d99074-Paper-Conference.pdf.

F. Santambrogio. *A Course in the Calculus of Variations: Optimization, Regularity, and Modeling*. Universitext, Springer Nature, 2023. URL https://doi.org/10.1007/978-3-031-45036-5.

M. Schmidt. An interior-point method for nonlinear optimization problems with locatable and separable nonsmoothness. *EURO Journal on Computational Optimization*, 3:309–348, 2015. URL https://doi.org/10.1007/s13675-015-0039-6.

T. Sejourne, F.-X. Vialard, and G. Peyré. The unbalanced gromov wasserstein distance: Conic formulation and relaxation. In M. Ranzato, A. Beygelzimer, Y. Dauphin, P.S. Liang, and J. Wortman Vaughan, editors, *Advances in Neural Information Processing Systems*, volume 34, pages 8766–8779. Curran Associates, Inc., 2021. URL https://proceedings.neurips.cc/paper_files/paper/2021/file/4990974d150d0de5e6e15a1454fe6b0f-Paper.pdf.

J. Solomon, G. Peyré, V. G. Kim, and S. Sra. Entropic metric alignment for correspondence problems. *ACM Transactions on Graphics (ToG)*, 35(4):1–13, 2016. URL https://doi.org/10.1145/2897824.2925903.

G. Still. Lectures on Parametric Optimization: An Introduction. *Optimization Online*, 2018. URL https://optimization-online.org/?p=15154.

V. Titouan, N. Courty, R. Tavenard, and R. Flamary. Optimal transport for structured data with application on graphs. In K. Chaudhuri and R. Salakhutdinov, editors, *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pages 6275–6284. PMLR, 09–15 Jun 2019. URL https://proceedings.mlr.press/v97/titouan19a.html.

A. Wächter and L. T. Biegler. On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming. *Mathematical programming*, 106:25–57, 2006. URL https://doi.org/10.1007/s10107-004-0559-y.

H. Xu, D. Luo, and L. Carin. Scalable Gromov-Wasserstein learning for graph partitioning and matching. 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., 2019a. URL https://proceedings.neurips.cc/paper_files/paper/2019/file/6e62a992c676f611616097dbea8ea030-Paper.pdf.

H. Xu, D. Luo, H. Zha, and L. Carin. Gromov-Wasserstein learning for graph matching and node embedding. In K. Chaudhuri and R. Salakhutdinov, editors, *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pages 6932–6941. PMLR, 09–15 Jun 2019b. URL https://proceedings.mlr.press/v97/xu19b.html.