Minimum Energy Density Steering of Linear Systems With Gromov-Wasserstein Terminal Cost

Kohei Morimoto and Kenji Kashima, *Senior Member, IEEE*

Abstract—In this letter, we newly formulate and solve the optimal density control problem with Gromov-Wasserstein (GW) terminal cost in discrete-time linear Gaussian systems. Differently from the Wasserstein or Kullback-Leibler distances employed in the existing works, the GW distance quantifies the difference in *shapes* of the distribution, which is invariant under translation and rotation. Consequently, our formulation allows us to find small energy inputs that achieve the desired shape of the terminal distribution, which has practical applications, e.g., robotic swarms. We demonstrate that the problem can be reduced to a Difference of Convex (DC) programming, which is efficiently solvable through the DC algorithm. Through numerical experiments, we confirm that the state distribution reaches the terminal distribution that can be realized with the minimum control energy among those having the specified shape.

Index Terms—Optimal density control, optimal transport, Gromov-Wasserstein distance.

I. INTRODUCTION

OPTIMAL density control is defined as the problem of controlling the probability distribution of state variables to the desired distribution in a dynamic system. Promising applications of optimal density control include systems in which it is important to manage errors in the state, such as quality control and aircraft control, as well as quantum systems in which the distribution of the state itself is the object of control [1].

The problem addressed in this letter is a variant of the (finite-time) covariance steering problem for discrete-time linear Gaussian systems. Among the long history of this line of research [2], [3], the most related recent works are as follows: hard constraint formulations, as seen in [4], [5],

[6], [7], where the terminal state distribution is enforced as a constraint; and soft constraint formulations, as seen in [8], [9], [10], where the Wasserstein distance between the terminal distribution and the target distribution is incorporated as a cost. In particular, Balci and Bakolas [10] presented an optimal density control problem for discrete-time linear Gaussian systems using Wasserstein distance as the terminal cost, formulated as semidefinite programming (SDP), a form of convex programming, to derive globally optimal control policies.

To motivate the present work, let us consider the state distribution as an ensemble of particles or a multi-robotic swarm [11]. In such applications, a particular *shape* of the formation is required to be achieved, but its location and orientation are often irrelevant. For example, they may seek to align in a single row in a two-dimensional region (See Fig. 3 below), or only the configuration may be specified based on inter-agent distance [12]. The aforementioned formulations can address the realization of the configuration with a fixed orientation but cannot address the optimization with respect to the rotation. To tackle this issue, we propose a novel density control problem incorporating the Gromov-Wasserstein distance (GW distance) as the terminal cost [13]. The GW distance is the distance between probability distributions and can measure the closeness of the shape of the probability distributions. By integrating the GW distance between the state and target distributions into the terminal cost, we can formulate the problem of controlling the shape of the state distribution. This problem can be viewed as a simultaneous optimization of the dynamical steering and the rotation of the target shape, which clearly contrasts the existing formulations.

In this letter, we focus on scenarios where the initial and target distributions are Gaussian and seek the optimal control policy among linear feedback control laws. While computing the GW distance between arbitrary distributions is challenging, it has recently been shown that the Gaussian Gromov-Wasserstein (GGW) distance, which is a relaxation of the GW distance for normal distributions, can be easily calculated [14]. We show that the optimal density control problem with GW terminal cost can be formulated as a difference of convex (DC) programming problem. We solve the problem by the DC algorithm (DCA) [15], a technique for solving DC

programming problems through iterative convex relaxation. Remarkably, the convexified problem is transformed into a SDP form, which can be efficiently solved using standard convex programming solvers.

The rest of this letter is organized as follows. In Section II, we introduce the concept of the GW distance and present the optimal density steering problem with the GW distance as the terminal cost. Section III discusses the formulation of the problem as a DC programming problem, highlighting the objective function's nature as a difference of convex function and deriving the convexified sub-problem used in the DC algorithm. The numerical simulations are presented in Section IV. Finally, we conclude our paper in Section V.

Notation: Let $\mathbb{S}_n^+$ and $\mathbb{S}_n^{++}$ denote the sets of n-dimensional positive semidefinite matrices and positive definite matrices, respectively. Let $O(n)$ denote the set of n-dimensional orthogonal matrices. For matrices, $\|\cdot\|_F$ denotes the Frobenius norm. For a convex function f, $\partial f(x)$ denotes the set of subgradients of f at x. Let $\mathcal{P}(\mathcal{X})$ denote the set of all probability distributions over $\mathcal{X}$. Let $\mathcal{N}_n(\xi, \Sigma)$ denote the multivariate normal distribution with mean $\xi \in \mathbb{R}^n$ and covariance $\Sigma \in \mathbb{S}_n^+$. Let $\mathcal{N}_n$ denote the set of all n-dimensional multivariate normal distributions.

II. PROBLEM SETTING

A. Gromov-Wasserstein Distance

The optimal transport distance is generally defined as the minimized transport cost of transporting one probability distribution to another probability distribution. The GW distance is the distance between probability distributions and is a variant of optimal transport distance, similar to the Wasserstein distance. Given two metric spaces $\mathcal{X}, \mathcal{Y}$, the set of transports Π between probability distributions $\mu \in \mathcal{P}(\mathcal{X})$ and $\nu \in \mathcal{P}(\mathcal{Y})$ is defined by

$$\Pi(\mu, \nu) := \{\pi(x, y) \mid \pi(x, y) \, dx = \nu(y),$$
$$\pi(x, y) \, dy = \mu(x)\}. \quad (1)$$

Each element $\pi(x, y)$ in $\Pi(\mu, \nu)$ represents how the weight $\mu(x)$ at x is transported to y, with the condition that $\pi(x, y) \, dx = \nu(y)$ ensuring that the transported destination becomes $\nu(y)$. The GW distance is defined by

$$GW^2(\mu, \nu) := \inf_{\pi \in \Pi(\mu, \nu)} \int \int \left(\|x - x'\|_{\mathcal{X}} - \|y - y'\|_{\mathcal{Y}} \right)^2 \pi(x, y)\pi\left(x', y'\right) dx dy dx' dy', \quad (2)$$

where $\|x - x'\|_{\mathcal{X}}$ and $\|y - y'\|_{\mathcal{Y}}$ represent the norms in the spaces $\mathcal{X}$ and $\mathcal{Y}$, respectively. The GW distance is small when points that are close (resp. farther apart) before transportation are brought closer together (resp. farther apart) after transportation. Conversely, the GW distance increases when points that were initially close are moved farther apart after transportation. Therefore, this definition quantifies the shape difference between two probabilistic distributions. For comparison, recall that the Wasserstein distance is defined as

$$W^2(\mu, \nu) := \inf_{\pi \in \Pi(\mu, \nu)} \int \int d(x, y)^2 \pi(x, y) dx dy, \quad (3)$$

where $\mathcal{X} = \mathcal{Y}$ and $d(\cdot, \cdot)$ is a suitable distance on $\mathcal{X}$. While the Wasserstein distance is sensitive to the absolute positions or orientations of the distributions, the GW distance is invariant under isometric transformations such as translations and rotations.

The GW distance involves a non-convex quadratic program over transport π, making it challenging to compute the GW distance between arbitrary probability distributions. Recently, it has been shown that the Gaussian Gromov-Wasserstein (GGW) distance, where the transport is constrained to be Gaussian distribution only, can be explicitly expressed in terms of the parameters of the normal distributions [14]. Specifically, the GGW between Gaussian distributions $\mu \in \mathcal{N}_m$ and $\nu \in \mathcal{N}_n$ is defined by

$$GGW^2(\mu, \nu) := \inf_{\pi \in \Pi(\mu, \nu) \cap \mathcal{N}_{m+n}} \int \int \left(\|x - x'\| - \|y - y'\| \right)^2 \pi(x, y)\pi\left(x', y'\right) dx dy dx' dy' \quad (4)$$

where $\Pi(\mu, \nu) \cap \mathcal{N}_{m+n}$ represents the restriction of the transport to the $(m + n)$-dimensional Gaussian distribution $\mathcal{N}_{m+n}$. For $\mu = \mathcal{N}_m(\xi_0, \Sigma_0), \nu = \mathcal{N}_n(\xi_1, \Sigma_1)$, it holds that

$$GGW^2(\mu, \nu) = 4(\text{tr}(\Sigma_0) - \text{tr}(\Sigma_1))^2 + 8 \| D_0 - D_1 \|^2, (5)$$

where D_0, D_1 are the diagonal matrices with the eigenvalues of Σ_0, Σ_1 sorted in descending order, and if $m \neq n$, the missing elements are filled with zeros.

B. Optimal Density Steering With Gromov-Wasserstein Terminal Cost

Let n_x be the dimension of the state space and n_u the dimension of the input, and consider the following discrete-time linear Gaussian system.

$$x_{k+1} = Ax_k + Bu_k + w_k \quad (6a)$$
$$x_0 = \mathcal{N}(0, \Sigma_0) \quad (6b)$$
$$w_k \sim \mathcal{N}(0, W_k) \quad (6c)$$

Here, the covariance matrix of the initial Gaussian distribution is $\Sigma_0 \in \mathbb{S}_{n_x}^{++}$ and that of the noise is $W_k \in \mathbb{S}_{n_x}^+$. For control input u_k, We use a stochastic linear control policy as

$$u_k(x) = \mathcal{N}(K_k x, Q_k), \quad (7)$$

where K_k is feedback gain and $Q_k \succeq O$ is covariance of Gaussian distribution. We consider the problem of minimizing the sum of the control costs and the Gromov-Wasserstein distance between the terminal distribution $\rho_N = \mathcal{N}(\xi_N, \Sigma_N)$ $(\Sigma_N \in \mathbb{S}_{n_x}^{++})$ and the target distribution $\rho_r = \mathcal{N}(0, \Sigma_r)$ $(\Sigma_r \in \mathbb{S}_{n_x}^+)$. The objective function is represented as

$$\min_{K_k, Q_k} J(K_k, Q_k) \quad (8a)$$

$$J(K_k, Q_k) := \lambda \mathbb{E}\left[\sum_{k=0}^{N-1} u_k^T R_k u_k \right] + GGW^2(\rho_N, \rho_r), \quad (8b)$$

where $R_k \in \mathbb{S}_{n_x}^{++}$ denotes the weights for control cost. Using the control policy (7) in system (6a), the probability distribution of the state x_N at the terminal time N will also be

the Gaussian distribution. Thus, by substituting equations (5) and (7) into equation (8b), we obtain

$$
\begin{aligned}
J(K_k, Q_k) = \lambda \sum_{k=0}^{N-1} \mathrm{tr}\big(R_k\big(K_k \Sigma_k K_k^T + Q_k\big)\big) \\
+ 4(\mathrm{tr}(\Sigma_N) - \mathrm{tr}(\Sigma_r))^2 \\
+ 8 \parallel \Sigma_N \parallel^2 - 16\,\mathrm{tr}\,(D_N D_r),
\end{aligned} \tag{9}
$$

where Σ_k is the covariance matrix of the state x_k, and D_N, D_r are diagonal matrices with the eigenvalues of Σ_N, Σ_r arranged in descending order. The dynamics of Σ_k is given by

$$
\begin{aligned}
\Sigma_{k+1} = A\Sigma_k A^T + BK_k \Sigma_k A^T + A\Sigma_k K_k^T B^T \\
+ BK_k \Sigma_k K_k^T B^T + BQ_k B^T + W_k.
\end{aligned} \tag{10}
$$

Here, we introduce the variable transformations $M_k := P_k \Sigma_k^{-1} P_k^T + Q_k$ and $P_k := K_k \Sigma_k$ as in the Ref. [10], [16]. To guarantee the invertibility of the variable transformation, we need $M_k \succeq O$ and $M_k - P_k \Sigma_k^{-1} P_k^T \succeq O$, which implies that the following condition must be satisfied:

$$
\begin{bmatrix} M_k & P_k \\ P_k^T & \Sigma_k \end{bmatrix} \succeq O. \tag{11}
$$

This condition is added to the optimization problem to ensure the feasibility of the solution. Finally, from the (9), (10) and (11), we can write the optimization problem to be solved as follows:

$$
\min_{\Sigma_k, M_k, P_k} \quad J(\Sigma_N, M_k) \tag{12a}
$$

$$
\begin{aligned}
J(\Sigma_N, M_k) = \lambda \sum_{k=0}^{N-1} \mathrm{tr}(R_k M_k) \\
+ 4(\mathrm{tr}(\Sigma_N) - \mathrm{tr}(\Sigma_r))^2 \\
+ 8 \parallel \Sigma_N^2 - 16\,\mathrm{tr}\,(D_N D_r)
\end{aligned} \tag{12b}
$$

$$
\mathrm{s.t.} \quad \begin{aligned} \Sigma_{k+1} = A_k \Sigma_k A_k^T + A_k P_k^T B_k^T \\ + B_k P_k A_k^T + B_k M_k B_k^T + W_k \end{aligned} \tag{12c}
$$

$$
\begin{bmatrix} M_k & P_k \\ P_k^T & \Sigma_k \end{bmatrix} \succeq O \tag{12d}
$$

Although we assumed a stochastic strategy as a control law in (7), the optimal solution turns out to be a deterministic strategy.

Theorem 1: Suppose $\{A_k\}_{k=0}^{N-1}$ are invertible. Then, the optimal policy in problem (12) is deterministic, that is, the optimal solution satisfies $Q_k = M_k - K_k \Sigma_k K_k^T = O$.

Proof: The proof is provided using the same arguments in the [10]. We utilize the Karush-Kuhn-Tucker conditions. Let E_k denote the Lagrange multiplier for constraint (12c), and let F_k denote the Lagrange multiplier for constraint (12d), represented as

$$
F_k = \begin{bmatrix} F_k^{00} & F_k^{01} \\ F_k^{10} & F_k^{11} \end{bmatrix}.
$$

From the stationarity condition, we obtain

$$
B_k^T E_k A_k + F_k^{01} = 0 \tag{13}
$$

$$
R_k - B_k^T E_k B_k + F_k^{00} = 0. \tag{14}
$$

The complementary slackness condition yields

$$
\begin{bmatrix} F_k^{00} & F_k^{01} \\ F_k^{10} & F_k^{11} \end{bmatrix} \begin{bmatrix} M_k & P_k \\ P_k^T & \Sigma_k \end{bmatrix} = O, \tag{15}
$$

implying

$$
\begin{bmatrix} F_k^{00} & F_k^{01} \\ F_k^{10} & F_k^{11} \end{bmatrix} \begin{bmatrix} I & P_k \Sigma_k^{-1} \\ O & I \end{bmatrix} \begin{bmatrix} M_k - P_k \Sigma_k^{-1} P_k^T & O \\ O & \Sigma_k \end{bmatrix} = O
$$

from the definiteness of Σ_k. Thus, we obtain

$$
F_{00}\big(M_k - P_k \Sigma_k^{-1} P_k^T\big) = O \tag{16}
$$

$$
F_{01}^T\big(M_k - P_k \Sigma_k^{-1} P_k^T\big) = O. \tag{17}
$$

Subsequently, by combining (13), (17), and the invertibility of A_k, we derive

$$
E_k B_k\big(M_k - P_k \Sigma_k^{-1} P_k^T\big) = O. \tag{18}
$$

Finally, from (14), (16), and (18), we conclude that

$$
\begin{aligned}
\big(R_k - B_k^T E_k B_k\big)\big(M_k - P_k \Sigma_k^{-1} P_k^T\big) \\
= R_k\big(M_k - P_k \Sigma_k^{-1} P_k^T\big) \\
= O.
\end{aligned}
$$

Then, due to the positive definiteness of R_k, we have $M_k - P_k \Sigma_k^{-1} P_k^T = O$. ∎

III. FORMULATION AS DIFFERENCE OF CONVEX PROGRAMMING

In this section, we show that problem (12) is a DC programming problem and solve it using the DCA, an optimization method for DC programming. The DC programming problem is an optimization problem whose objective function is a DC function, which is expressed as the difference between two convex functions. Since the DC programming problem is a non-convex optimization problem, finding a global optimum is generally challenging. However, several optimization methods that efficiently find solutions by exploiting the properties of DC functions have been proposed. These include global optimization techniques using branch and bound methods [17], and methods for finding sub-optimal solutions, such as the DCA [15] and the Concave-Convex Procedure (CCCP) [18].

The DCA iteratively constructs a convex upper bound for the objective function, minimizes this upper bound, and then updates the upper bound using the minimizer of the previous iteration. Assume the objective function h to be optimized is expressed as $h(z) = f(z) - g(z)$ by convex functions $f(z)$ and $g(z)$ defined on a convex set Ω. The DCA iterates the following steps until convergence:

1) Construct the upper bound $\hat{h}(z)$ of $h(z)$ as

$$
\hat{h}(z) = f(z) - s_n^T(z - z_n),
$$

   where $s_n \in \partial g(z_n)$.
2) Set $z_{n+1} = \min_{z \in \Omega} \hat{h}(z)$.

Because $\hat{h}(z)$ is a convex function over the convex set, we are able to minimize this convex subproblem efficiently. It is known [15] that when the optimal value of the problem is finite and the sequences z_n, s_n are bounded, any accumulation
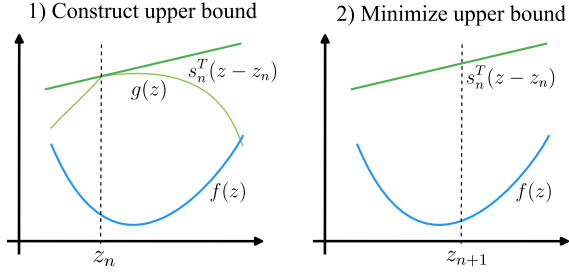
Fig. 1. Visualization for the algorithm of DCA.

points z^∞ of z_n are critical points of $f - g$, which implies $0 \in \partial(f - g)(z^\infty)$. In Figure 1, we show the visualization of the DCA algorithm.

In the next proposition and theorem, we show that problem (12) is a DC programming problem.

Proposition 1 (Anstreicher and Wolkowicz [19]): Let A and B be $n \times n$ symmetric matrices decomposed into their eigenvalues as $A = V\Lambda V^T$ and $B = W\Xi W^T$, respectively. Assume the eigenvalues in W and V are arranged in descending order. Then,

$$\max_{U \in O(n)} \text{tr}\left(UAU^T B\right)$$

has an optimal solution $U^* = WV^T$, and the optimal value is $\text{tr}(\Lambda\Xi)$.

Theorem 2: $J(\Sigma_N, M_k)$ in (12b) is a DC function.

Proof: Since $\lambda \sum_{k=0}^{N-1} (\text{tr}(R_k M_k)) + 4(\text{tr}(\Sigma_N) - \text{tr}(\Sigma_r))^2 + 8 \| \Sigma_N \|^2$ is clearly a convex function, it suffices to show that

$$g(\Sigma_N) := \text{tr}(D_N D_r) \tag{19}$$

is a convex function. From Proposition 1, we have

$$\text{tr}(D_N D_r) = \max_{U \in O(n)} \text{tr}\left(U\Sigma_N U^T \Sigma_r\right).$$

Thus, $\text{tr}(D_N D_r)$ is the maximum of linear functions, making it a convex function with respect to Σ_N. Precisely, let us consider a scalar $\alpha \in [0, 1]$ and two positive definite matrices Σ and Σ'. Then, the convex combination $\alpha\Sigma + (1 - \alpha)\Sigma'$ satisfies the following inequality:

$$\max_{U \in O(n)} \text{tr}\left(U(\alpha\Sigma + (1 - \alpha)\Sigma')U^T \Sigma_r\right)$$
$$\leq \max_{U \in O(n)} \alpha \text{tr}\left(U\Sigma U^T \Sigma_r\right) + (1 - \alpha) \text{tr}\left(U\Sigma' U^T \Sigma_r\right)$$
$$\leq \alpha \max_{U \in O(n)} \text{tr}\left(U\Sigma U^T \Sigma_r\right)$$
$$+ (1 - \alpha) \max_{U' \in O(n)} \text{tr}\left(U'\Sigma' U'^T \Sigma_r\right).$$
∎

Furthermore, in the next theorem, we derive a subgradient of the concave part of $J(M_k, \Sigma_k)$ to construct the upper bound in DCA.

Theorem 3: Assume the eigenvalue decompositions: $\Sigma_N = V_N D_N V_N^T$ and Σ_r is $\Sigma_r = V_r D_r V_r^T$. Then, $V_N D_r V_N^T \in \mathbb{S}_{n_x}^+$ is a subgradient of $g(\Sigma_N)$ in (19).

Proof: From Proposition 1, we have

$$U^* := \arg \max_{U \in O(n)} \text{tr}\left(U\Sigma_N U^T \Sigma_r\right) = V_r V_N^T.$$

Therefore, by Danskin's theorem [20], a subgradient can be obtained by differentiating the function inside the max operation with respect to Σ_N and then substituting U^*. Hence,

$$U^{*T}\Sigma_r U^* = V_N V_r^T \Sigma_r V_r V_N^T = V_N D_r V_N^T \in \partial g(\Sigma_N).$$
∎

Therefore, the convex subproblem in DCA is formulated as

$$\min_{\Sigma_k, M_k, P_k} \quad \lambda \sum_{k=0}^{N-1} \text{tr}(R_k M_k)$$
$$+ 4(\text{tr}(\Sigma_N) - \text{tr}(\Sigma_r))^2$$
$$+ 8 \| \Sigma_N \|^2 - 16 \text{tr}\left(\Sigma_N V_N^{(n)T} D_r V_N^{(n)}\right) \tag{20a}$$
$$\text{s.t.} \quad \Sigma_{k+1} = A_k \Sigma_k A_k^T + A_k P_k^T B_k^T + B_k P_k A_k^T$$
$$+ B_k M_k B_k^T + W_k \tag{20b}$$
$$\begin{bmatrix} M_k & P_k \\ P_k^T & \Sigma_k \end{bmatrix} \succeq 0 \tag{20c}$$

where $V_N^{(n)}$ is the matrix obtained by decomposing the optimal Σ_N in the n-th iteration of DCA. The term $\text{tr}\left(\Sigma_N V_N^{(n)T} D_r V_N^{(n)}\right)$ represents linear lower bound of convex function $l(\Sigma_N)$ using a subgradient obtained in Theorem 3. The subproblem is a semidefinite programming problem (SDP), which can be efficiently solved. We use the solution from each optimization step to update the value of V_N. By iteratively applying the optimization process, the solution progressively approaches a sub-optimal solution for the original problem (12).

In Theorem 1, we showed that the optimal policy for Problem (12) is deterministic. The following theorem shows that the proposed algorithm generates a sequence of deterministic control policies:

Theorem 4: When $\{A_k\}_{k=0}^{N-1}$ are invertible, the optimal policy of the subproblem (20a) is also deterministic.

Proof: The KKT conditions used in the proof of Theorem 1 also hold in the subproblem (20a). ∎

IV. NUMERICAL EXPERIMENTS

In this section, we perform numerical optimization for problem (12) using DCA. We set the parameters in this experiment as

$$A_k = \begin{bmatrix} 1.0 & 0.1 \\ -0.3 & 1.0 \end{bmatrix}, \quad B_k = \begin{bmatrix} 0.7 \\ 0.4 \end{bmatrix}$$
$$\Sigma_0 = \begin{bmatrix} 3 & 0 \\ 0 & 3 \end{bmatrix}, \quad W_k = 0.5I_2,$$
$$R_k = 1.0, \quad N = 10.$$

Figure 2 shows the time evolution of state covariance of the uncontrolled system. For the implementation of the convex subproblem in DCA, we used the MOSEK solver [21] and the CVXPY modeler [22].

1) Line Alignment: First, we consider the case where the desired density is Gaussian $\rho_r = \mathcal{N}(0, 10)$, which is not on $\mathbb{R}^2$, but on $\mathbb{R}$. The problem seeks the optimal policy to align the terminal distribution into one line. Note that $W(\rho_N, \rho_r)$
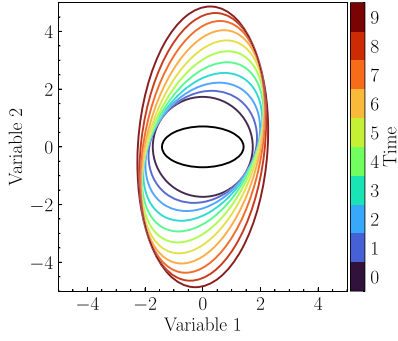
Fig. 2. The snapshots of the covariance matrices of uncontrolled system. The colored lines indicate the 2σ range of Σ_k, while the black line represents Σ_r in (22).
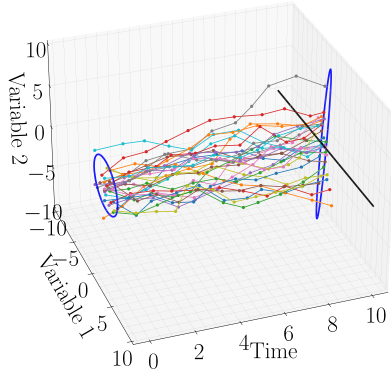

Fig. 3. One hundred sample paths of the systems controlled by the optimal policy for $\lambda = 1$. The blue circle denotes 2σ range of Σ_0 and Σ_N and the black line denotes the degenerated target distribution Σ_r in (21).

in (3) does not make sense[1] because $\mathcal{X} \neq \mathcal{Y}$. In contrast, the GW distance $GW(\rho_N, \rho_r)$ in (2) is well-defined and, thanks to (5), equivalent to take

$$\Sigma_r = \begin{bmatrix} 10 & 0 \\ 0 & 0 \end{bmatrix}. \tag{21}$$

Figure 3 presents the trajectories of one hundred samples from the controlled process when the target distribution is degenerate distribution. It can be seen that the distribution of states actually stretches vertically to achieve a one-line alignment.

2) Comparison With the Wasserstein Formulation: Let us consider

$$\Sigma_r = \begin{bmatrix} 2 & 0 \\ 0 & 0.5 \end{bmatrix} \tag{22}$$

for which

$$GGW^2(\rho_N, \rho_r) = 6711.44 \tag{23}$$

for the uncontrolled system. Figure 4 shows the snapshots of state covariance under the optimal control input obtained by DCA. It can be observed that the shape of the terminal distribution approaches that of the target distribution as λ

[1]One may think we can embed it onto $\mathbb{R}^2$ (e.g., by (21)) and consider the Wasserstein distance. However, there is a rotational degree of freedom, which affects the resulting distance. We can interpret that our formulation optimizes this rotation in the sense of the required control energy; See Fig. 3.
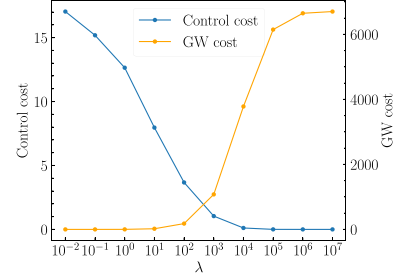

Fig. 5. Relationship between the optimized control cost term and GW cost term in Eq. (8b) for each λ.

decreases. As shown below, the terminal distribution is the one requiring the least energy among the rotated distributions of the target due to the rotational invariance of the GW distance.

Figure 5 shows the relationship between the optimized control cost term and GW cost term in Eq. (8b) for each λ. As the value of λ increases, the control cost rises, while the GW cost decreases. Conversely, as the value of λ decreases, the control cost diminishes, and the GW cost increases. Also, as λ becomes smaller, the GW cost is almost close to zero. It is noteworthy that, in comparison to the uncontrolled system in (23), our algorithm achieves a significant reduction in the GW cost.

Finally, we clarify the advantage of our approach over the Wasserstein terminal cost problem [10]. In Fig. 4(a), the obtained terminal distribution is $\rho_N \approx \mathcal{N}(0, \hat{\Sigma}_r(\theta_{GW}))$ with $\theta_{GW} = 1.20$ [rad] where $\hat{\Sigma}_r(\theta)$ is obtained by rotating Σ_r by an angle θ, i.e.,

$$\hat{\Sigma}_r(\theta) := R(\theta)^T \Sigma_r R(\theta), \quad R(\theta) := \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}.$$

It is shown in [10] that we can solve

$$\min_{K_k, Q_k} \lambda \mathbb{E}\left[\sum_{k=0}^{N-1} u_k^T R_k u_k\right] + W^2\left(\rho_N, \mathcal{N}\left(0, \hat{\Sigma}_r(\theta)\right)\right) \tag{24}$$

by an SDP. Then, we solved this problem for a sufficiently small λ (i.e., large terminal cost). The required control energy for the obtained optimal control input (i.e., the first term without λ in (24)) is denoted by $W_{opt}(\theta)$, which is shown in Fig. 6. It is noteworthy that the function exhibits a non-convexity. We can also observe

$$\theta_{GW} \approx \theta^* := \arg\min_{\theta} W_{opt}(\theta),$$

which implies that the rotation angle obtained by the GW terminal cost problem minimizes the resulting control energy needed to realize the required shape (specified by Σ_r). From a computation cost point of view, while finding θ^* using the Wasserstein terminal cost approach requires performing optimization to compute $W_{opt}(\theta)$ for each θ, our GW terminal cost framework only necessitates solving a single optimization problem. Moreover, our approach remains computationally tractable even in high-dimensional settings, where the Wasserstein terminal cost approach becomes computationally intractable due to the exponential growth of the search space of the rotation matrix.
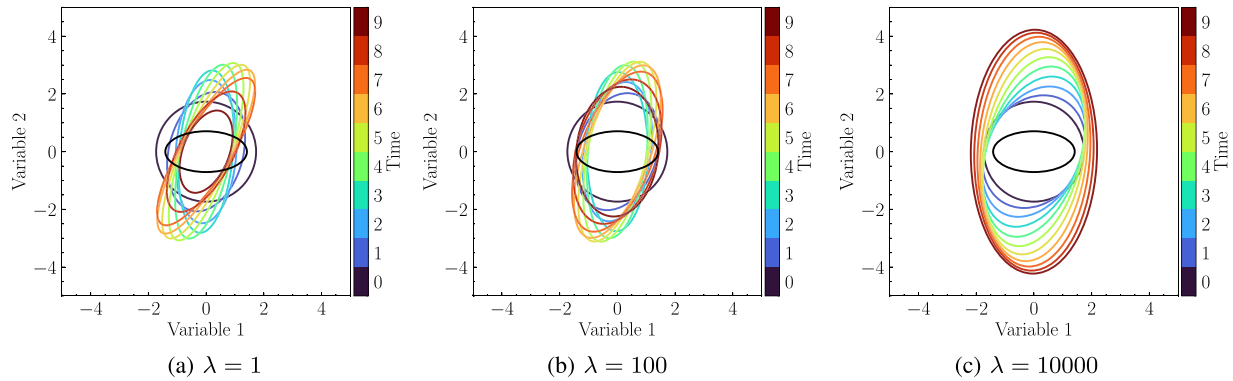
(a) $\lambda = 1$ (b) $\lambda = 100$ (c) $\lambda = 10000$

Fig. 4. The snapshots of the covariance matrices under the optimized policy for different values of λ. The colored lines indicate the 2σ range of Σ_k, while the black line represents Σ_r in (22).
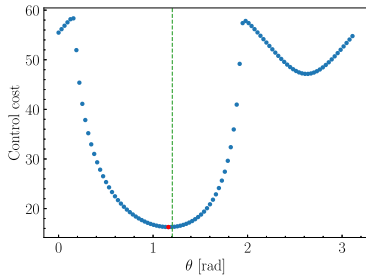

Fig. 6. The required control energy $W_{\text{opt}}(\theta)$ of the optimal input for the Wasserstein terminal cost problem in (24). The green line and the red dot represent θ_{GW} and θ^*, respectively.

V. CONCLUSION

In this letter, we addressed the optimal density control problem with the Gromov-Wasserstein distance as the terminal cost. We showed that the problem is a DC programming problem and proposed an optimization method based on the DC algorithm. Numerical experiments confirmed that the state distribution reaches the terminal distribution that can be realized with the minimum control energy among those having the specified shape.

Future work includes the application of the proposed GW framework to the transport between spaces equipped with different Riemannian metric structures or point clouds. Model predictive formation control based on a fast algorithm for optimal transport [23] is also a promising direction [24]. The convergence and computation complexity of the proposed DC algorithm should also be investigated.

REFERENCES

[1] Y. Chen, T. T. Georgiou, and M. Pavon, "Optimal steering of a linear stochastic system to a final probability distribution, part I," *IEEE Trans. Autom. Control*, vol. 61, no. 5, pp. 1158–1169, May 2016.

[2] B. D. O. Anderson, "The inverse problem of stationary covariance generation," *J. Stat. Phys.*, vol. 1, no. 1, pp. 133–147, 1969.

[3] A. F. Hotz and R. E. Skelton, "A covariance control theory," in *Proc. 24th IEEE Conf. Decision Control*, 1985, pp. 552–557.

[4] F. Liu, G. Rapakoulias, and P. Tsiotras, "Optimal covariance steering for discrete-time linear stochastic systems," 2022, *arXiv:2211.00618*.

[5] G. Rapakoulias and P. Tsiotras, "Discrete-time optimal covariance steering via semidefinite programming," 2023, *arXiv:2302.14296*.

[6] K. Ito and K. Kashima, "Maximum entropy density control of discrete-time linear systems with quadratic cost," 2023, *arXiv:2309.10662*.

[7] K. Ito and K. Kashima, "Maximum entropy optimal density control of discrete-time linear systems and Schrödinger bridges," *IEEE Trans. Autom. Control*, vol. 69, no. 3, pp. 1536–1551, Mar. 2024.

[8] A. Halder and E. D. B. Wendel, "Finite horizon linear quadratic Gaussian density regulator with Wasserstein terminal cost," in *Proc. Amer. Control Conf. (ACC)*, 2016, pp. 7249–7254.

[9] I. M. Balci and E. Bakolas, "Covariance steering of discrete-time stochastic linear systems based on Wasserstein distance terminal cost," *IEEE Control Syst. Lett.*, vol. 5, no. 6, pp. 2000–2005, Dec. 2021.

[10] I. M. Balci and E. Bakolas, "Exact SDP formulation for discrete-time covariance steering with Wasserstein terminal cost," 2022, *arXiv:2205.10740*.

[11] V. Krishnan and S. Martínez, "Distributed optimal transport for the deployment of swarms," in *Proc. IEEE Conf. Decision Control (CDC)*, 2018, pp. 4583–4588.

[12] D. V. Dimarogonas and K. H. Johansson, "On the stability of distance-based formation control," in *Proc. 47th IEEE Conf. Decision Control*, 2008, pp. 1200–1205.

[13] F. Mémoli, "Gromov-Wasserstein distances and the metric approach to object matching," *Found. Comput. Math.*, vol. 11, pp. 417–487, Apr. 2011.

[14] J. Delon, A. Desolneux, and A. Salmona, "Gromov-Wasserstein distances between Gaussian distributions," *J. Appl. Probab.*, vol. 59, no. 4, pp. 1178–1198, 2022.

[15] P. D. Tao and L. H. An, "Convex analysis approach to DC programming: Theory, algorithms and applications," *Acta Math. Vietnamica*, vol. 22, no. 1, pp. 289–355, 1997.

[16] Y. Chen, T. T. Georgiou, and M. Pavon, "Optimal steering of a linear stochastic system to a final probability distribution, part II," *IEEE Trans. Autom. Control*, vol. 61, no. 5, pp. 1170–1180, May 2016.

[17] R. Horst and N. V. Thoai, "DC programming: Overview," *J. Optim. Theory Appl.*, vol. 103, pp. 1–43, Oct. 1999.

[18] A. L. Yuille and A. Rangarajan, "The concave-convex procedure (CCCP)," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 14, 2001, pp. 1–8.

[19] K. Anstreicher and H. Wolkowicz, "On lagrangian relaxation of quadratic matrix constraints," *SIAM J. Matrix Anal. Appl.*, vol. 22, no. 1, pp. 41–55, 2000.

[20] D. P. Bertsekas, *Nonlinear Programming*, 2nd ed. Nashua, NH, USA: Athena Sci., 1999.

[21] *MOSEK Optimizer API for Python. Version 10.1*, MOSEK ApS, Copenhagen, Denmark, 2023.

[22] S. Diamond and S. Boyd, "CVXPY: A python-embedded modeling language for convex optimization," *J. Mach. Learn. Res.*, vol. 17, no. 83, pp. 1–5, 2016.

[23] J. Solomon, G. Peyré, V. G. Kim, and S. Sra, "Entropic metric alignment for correspondence problems," *ACM Trans. Graph.*, vol. 35, no. 4, pp. 1–13, 2016.

[24] K. Ito and K. Kashima, "Entropic model predictive optimal transport over dynamical systems," *Automatica*, vol. 152, Jun. 2023, Art. no. 110980.