Global Convergence of Direct Policy Search for State-Feedback $\mathcal{H}_\infty$ Robust Control: A Revisit of Nonsmooth Synthesis with Goldstein Subdifferential

Anonymous Author(s)
Affiliation
Address
email

Abstract

Direct policy search has been widely applied in modern reinforcement learning and continuous control. However, the performance of direct policy search on nonsmooth robust control synthesis has not been well understood. The optimal $\mathcal{H}_\infty$ control framework aims at designing a policy to minimize the closed-loop $\mathcal{H}_\infty$ norm, and is arguably the most important robust control paradigm. In this work, we show that direct policy search is guaranteed to find the global solution of the robust $\mathcal{H}_\infty$ state-feedback control design problem. Notice that policy search for optimal $\mathcal{H}_\infty$ control leads to a constrained nonconvex nonsmooth optimization problem where the nonconvex feasible set consists of all the policies stabilizing the closed-loop dynamics. We show that for this nonsmooth optimization problem, all Clarke stationary points are global minimum. Next, we identify the coercivity of the closed-loop $\mathcal{H}_\infty$ objective function, and prove that the sublevel sets of the resultant policy search problem are compact. Based on these properties, we show that the Goldstein subdifferential method and its various implementable variants can be guaranteed to stay in the non-convex feasible set and eventually find the global optimal solution for the $\mathcal{H}_\infty$ state-feedback synthesis problem. Our work builds a new connection between non-convex nonsmooth optimization theory and robust control, leading to the first global convergence result for direct policy search on optimal $\mathcal{H}_\infty$ synthesis.

1 Introduction

Reinforcement learning (RL) has achieved impressive performance on many continuous control tasks [43, 34], and policy optimization is one of the main workhorses for such applications [15, 46, 42, 44]. Recently, there have been extensive research efforts studying the global convergence properties of policy optimization methods on benchmark control problems including linear quadratic regulator (LQR) [18, 7, 35, 48, 37, 19, 33, 25], stabilization [40], linear robust/risk-sensitive control [51, 50, 23, 52, 53], Markov jump linear quadratic control [28, 27, 41], output feedback control [17, 54], and dynamic filtering [47]. For all these benchmark problems, the objective function in the policy optimization formulation is always differentiable over the entire feasible set, and the existing convergence theory heavily relies on this fact. Consequently, an important open question remains whether direct policy search can enjoy similar global convergence properties when applied to the famous $\mathcal{H}_\infty$ control problem whose objective function is not always differentiable in the policy space [1, 2, 3, 24, 9, 12]. Different from LQR which considers stochastic disturbance, $\mathcal{H}_\infty$ control directly addresses the worst-case disturbance, and provides arguably the most fundamental robust control paradigm [55, 16, 45, 4, 14, 20]. Regarding the connection with RL, it has also been shown that $\mathcal{H}_\infty$ control can be applied to stabilize the training of adversarial RL schemes in the linear

Submitted to 36th Conference on Neural Information Processing Systems (NeurIPS 2022). Do not distribute.

quadratic setup [50, Section 5]. Given the fundamental importance of $\mathcal{H}_\infty$ control, we view it as an important benchmark for understanding the theoretical properties of direct policy search in the context of robust control and adversarial RL. The main contribution of our paper is the first global convergence guarantee of direct policy search for the $\mathcal{H}_\infty$ state-feedback control synthesis problem.

The objective of the $\mathcal{H}_\infty$ state-feedback synthesis is to design a linear state-feedback policy that stabilizes the closed-loop system and minimizes the $\mathcal{H}_\infty$ norm from the disturbance to a performance signal at the same time. The design goal is also equivalent to synthesizing a state-feedback policy that minimizes a quadratic cost subject to the worst-case disturbance. We will present the problem formulation for the $\mathcal{H}_\infty$ state-feedback synthesis and discuss such connections in Section 2. Essentially, $\mathcal{H}_\infty$ state-feedback synthesis can be formulated as a constrained policy optimization problem $\min_{K\in\mathcal{K}} J(K)$, where the decision variable K is a matrix parameterizing the linear state-feedback policy, the objective function $J(K)$ is the closed-loop $\mathcal{H}_\infty$-norm for given K, and the feasible set $\mathcal{K}$ consists of all the linear state-feedback policies stabilizing the closed-loop dynamics. Notice that the feasible set for the $\mathcal{H}_\infty$ state-feedback control problem is the same as the nonconvex feasible set for the LQR policy search problem [18, 7]. However, the objective function $J(K)$ for the $\mathcal{H}_\infty$ control can be non-differential over certain feasible points, introducing new difficulty to direct policy search. There has been a large family of nonsmooth $\mathcal{H}_\infty$ policy search algorithms developed based on the concept of Clarke subdifferential [1, 2, 3, 24, 9, 12]. However, no global convergence guarantee has been reported in the literature. Our paper bridges this gap by making the following two contributions.

1. We show that there exists a unique global minimizing set for $\mathcal{H}_\infty$ state-feedback control problem, and any Clarke stationary points are also global minimum.

2. We identify the coercivity property of the $\mathcal{H}_\infty$ cost function and use this property to show that the Goldstein subdifferential method [22] and its implementable variants [49, 13, 9, 10, 31, 32] can be guaranteed to stay in the nonconvex feasible set of stabilizing policies during the optimization process and eventually find the global optimal solution of the $\mathcal{H}_\infty$ state-feedback control problem.

To the best of our knowledge, our work offers the first global convergence guarantee of direct policy search on $\mathcal{H}_\infty$ control problems. We believe that our work also serves as a meaningful initial step towards a general convergence theory of direct policy search on nonsmooth robust control synthesis.

Finally, it is worth clarifying the differences between $\mathcal{H}_\infty$ control and mixed $\mathcal{H}_2/\mathcal{H}_\infty$ design. For mixed $\mathcal{H}_2/\mathcal{H}_\infty$ control, the objective is to design a stabilizing policy that minimizes an $\mathcal{H}_2$ performance bound and satisfies an $\mathcal{H}_\infty$ constraint at the same time [21, 30, 29, 38]. In other words, mixed $\mathcal{H}_2/\mathcal{H}_\infty$ control aims at improving the average $\mathcal{H}_2$ performance while “maintaining" a certain level of robustness by keeping the closed-loop $\mathcal{H}_\infty$ norm to be smaller than a pre-specified number. In contrast, $\mathcal{H}_\infty$ control aims at “improving" the system robustness and the worst-case performance via achieving the smallest closed-loop $\mathcal{H}_\infty$ norm. In [51], it has been shown that the natural policy gradient method initialized from a policy satisfying the $\mathcal{H}_\infty$ constraint can be guaranteed to maintain the $\mathcal{H}_\infty$ requirement during the optimization process and eventually converge to the optimal solution of the mixed design problem. However, notice that the objective function for the mixed $\mathcal{H}_2/\mathcal{H}_\infty$ control problem is still differentiable over all the feasible points, and hence the analysis technique in [51] cannot be applied to our $\mathcal{H}_\infty$ control setting. More discussion on the connections and differences between these two problems will be presented in the supplementary material.

.

2 Problem Formulation and Preliminaries

2.1 Notation

The set of p-dimensional real vectors is denoted as $\mathbb{R}^p$. For a matrix A, we use the notation A^T, $\|A\|$, $\operatorname{tr} A$, $\sigma_{\min}(A)$, $\|A\|_2$, and $\rho(A)$ to denote its transpose, largest singular value, trace, smallest singular value, Frobenius norm, and spectral radius, respectively. When a matrix P is negative semidefinite (definite), we will use the notation $P \preceq (\prec) 0$. When P is positive semidefinite (definite), we use the notation $P \succeq (\succ) 0$.

Consider a (real) sequence $\mathbf{u} := \{u_0, u_1, \cdots\}$ where $u_t \in \mathbb{R}^{n_u}$ for all t. This sequence is said to be in $\ell_2^{n_u}$ if $\sum_{t=0}^{\infty} \|u_t\|^2 < \infty$ where $\|u_t\|$ denotes the standard (vector) 2-norm of u_t. In addition, the 2-norm for $\mathbf{u} \in \ell_2^{n_u}$ is defined as $\|\mathbf{u}\|^2 := \sum_{t=0}^{\infty} \|u_t\|^2$.

2.2 Problem statement: $\mathcal{H}_\infty$ state-feedback synthesis and a policy optimization formulation

We consider the following linear time-invariant (LTI) system

$$x_{t+1} = Ax_t + Bu_t + w_t,\ x_0 = 0 \tag{1}$$

where $x_t \in \mathbb{R}^{n_x}$ is the state, $u_t \in \mathbb{R}^{n_u}$ is the control action, and $w_t \in \mathbb{R}^{n_w}$ is the disturbance. We have $A \in \mathbb{R}^{n_x \times n_x}$ $B \in \mathbb{R}^{n_x \times n_u}$, and $n_w = n_x$. We denote $\mathbf{x} := \{x_0, x_1, \cdots\}$, $\mathbf{u} := \{u_0, u_1, \cdots\}$, and $\mathbf{w} := \{w_0, w_1, \cdots\}$. The initial condition is fixed as $x_0 = 0$. The objective of $\mathcal{H}_\infty$ control is to choose $\{u_t\}$ to minimize the quadratic cost $\sum_{t=0}^{\infty}(x_t^\mathsf{T} Q x_t + u_t^\mathsf{T} R u_t)$ in the presence of the worst-case ℓ_2 disturbance satisfying $\|\mathbf{w}\| \leq 1$. In this paper, the following assumption is made on the matrices (A, B, Q, R).

Assumption 1. *The matrices Q and R are positive definite. The matrix pair (A, B) is stabilizable.*

In $\mathcal{H}_\infty$ control, $\{w_t\}$ is considered to be the worst-case disturbance satisfying the ℓ_2 norm bound $\|\mathbf{w}\| \leq 1$, and can be chosen in an adversarial manner. This is different from LQR which makes stochastic assumptions on $\{w_t\}$. Without loss of generality, we have chosen the ℓ_2 upper bound on $\mathbf{w}$ to be 1. In principle, we can formulate the $\mathcal{H}_\infty$ control problem with any arbitrary ℓ_2 upper bound on $\mathbf{w}$, and there is no technical difference. We will provide more explanations on this fact in the supplementary material. Therefore, $\mathcal{H}_\infty$ control can be formulated as the following minimax problem

$$\min_{\mathbf{u}} \max_{\mathbf{w}:\|\mathbf{w}\|\leq 1} \sum_{t=0}^{\infty}(x_t^\mathsf{T} Q x_t + u_t^\mathsf{T} R u_t) \tag{2}$$

Under Assumption 1, it is well known that the optimal solution for (2) can be achieved using a linear state-feedback policy $u_t = -Kx_t$ (see [4]). Given any K, the LTI system (1) can be rewritten as

$$x_{t+1} = (A - BK)x_t + w_t,\ x_0 = 0. \tag{3}$$

Now we define $z_t = (Q + K^\mathsf{T} R K)^{\frac{1}{2}} x_t$. We have $\|z_t\|^2 = x_t^\mathsf{T}(Q + K^\mathsf{T} R K)x_t = x_t^\mathsf{T} Q x_t + u_t^\mathsf{T} R u_t$. Hence we can denote $\mathbf{z} = \{z_0, z_1, \cdots\}$. If $\mathbf{x} \in \ell_2^{n_x}$, then we have $\|\mathbf{z}\|^2 = \sum_{t=0}^{\infty}(x_t^\mathsf{T} Q x_t + u_t^\mathsf{T} R u_t)$. Therefore, the closed-loop LTI system (3) can be viewed as a linear operator mapping any disturbance sequence $\{w_t\}$ to another sequence $\{z_t\}$. We denote this operator as G_K, where the subscript highlights the dependence of this operator on K. If K is stabilizing, i.e. $\rho(A - BK) < 1$, then G_K is bounded in the sense that it maps any ℓ_2 sequence $\mathbf{w}$ to another sequence $\mathbf{z}$ in $\ell_2^{n_x}$. For any stabilizing K, the $\ell_2 \to \ell_2$ induced norm of G_K can be defined as:

$$\|G_K\|_{2\to 2} := \sup_{0 \neq \|\mathbf{w}\| \leq 1} \frac{\|\mathbf{z}\|}{\|\mathbf{w}\|} \tag{4}$$

Since G_K is a linear operator, it is straightforward to show

$$\|G_K\|_{2\to 2}^2 := \max_{\mathbf{w}:\|\mathbf{w}\|\leq 1} \sum_{t=0}^{\infty} x_t^\mathsf{T}(Q + K^\mathsf{T} R K)x_t = \max_{\mathbf{w}:\|\mathbf{w}\|\leq 1} \sum_{t=0}^{\infty}(x_t^\mathsf{T} Q x_t + u_t^\mathsf{T} R u_t).$$

Therefore, the minimax optimization problem (2) can be rewritten as the policy optimization problem: $\min_{K\in\mathcal{K}} \|G_K\|_{2\to 2}^2$, where $\mathcal{K}$ is the set of all linear state-feedback stabilizing policies, i.e. $\mathcal{K} = \{K \in \mathbb{R}^{n_x \times n_u} :\ \rho(A - BK) < 1\}$. In the robust control literature [1, 2, 3, 24, 9, 12], it is standard to drop the square in the cost function and just reformulate (2) as $\min_{K\in\mathcal{K}} \|G_K\|_{2\to 2}$. This is exactly the policy optimization formulation for $\mathcal{H}_\infty$ state-feedback control. The main reason why this problem is termed as $\mathcal{H}_\infty$ state-feedback control is that in the frequency domain, G_K can be viewed as a transfer function which lives in the Hardy $\mathcal{H}_\infty$ space and has an $\mathcal{H}_\infty$ norm being exactly equal to $\|G_K\|_{2\to 2}$. We denote the $\mathcal{H}_\infty$ norm of G_K as $\|G_K\|_\infty$. We have

$$\|G_K\|_\infty = \sup_{\omega\in[0,2\pi]} \lambda_{\max}^{1/2}\big((e^{-j\omega}I - A + BK)^{-\mathsf{T}}(Q + K^\mathsf{T} R K)(e^{j\omega}I - A + BK)^{-1}\big), \tag{5}$$

where I is the identity matrix, and $\lambda_{\max}$ denotes the largest eigenvalue of a given symmetric matrix. Therefore, eventually the $\mathcal{H}_\infty$ state-feedback control problem can be formulated as

$$\min_{K \in \mathcal{K}} J(K), \tag{6}$$

where $J(K)$ is equal to the $\mathcal{H}_\infty$ norm $\|G_K\|_\infty$ defined by (5). Classical $\mathcal{H}_\infty$ control theory typically solves (6) via introducing extra Lyapunov variables and reparameterizing the problem into a higher-dimensional convex domain over which convex optimization algorithms can be applied [55, 16, 6]. In this paper, we revisit (6) as a benchmark for direct policy search, and discuss how to search the optimal solution of (6) in the policy space directly. Applying direct policy search for (6) leads to a nonconvex nonsmooth optimization problem. A main technical challenge is that the objective function (5) can be non-differentiable over some important feasible points [1, 2, 3, 24, 9, 12].

2.3 Direct policy search: A nonsmooth optimization perspective

Now we briefly review several key facts known for the $\mathcal{H}_\infty$ policy optimization problem (6).

Proposition 1. *The set $\mathcal{K} = \{K : \rho(A - BK) < 1\}$ is open. In general, it can be unbounded and non-convex. The cost function* (5) *is continuous and nonconvex in K.*

See [18, 8] for some related proofs. We have also included more explanations in the supplementary material. An immediate consequence is that (6) becomes a nonconvex optimization problem. Another important fact is that the objective function (5) is also nonsmooth. As a matter of fact, (5) is subject to two sources of nonsmoothness. Namely, the largest eigenvalue for a fixed frequency ω is nonsmooth, and the optimization step over $\omega \in [0, 2\pi]$ is also nonsmooth. As a matter of fact, the $\mathcal{H}_\infty$ objective function (5) can be non-differentiable on important feasible points, e.g. optimal points. Fortunately, it is well known[1] that the $\mathcal{H}_\infty$ objective function (5) has the following desired property so it is Clarke subdifferentiable.

Proposition 2. *The $\mathcal{H}_\infty$ objective function* (5) *is locally Lipschitz and subdifferentially regular over the stabilizing feasible set $\mathcal{K}$*

Recall that $J : \mathcal{K} \to \mathbb{R}$ is locally Lipschitz if for any bounded $S \subset \mathcal{K}$, there exists a constant $L > 0$ such that $|J(K) - J(K')| \leq L\|K - K'\|_2$ for all $K, K' \in S$. Based on Rademacher's theorem, a locally Lipschitz function is differentiable almost everywhere, and the Clarke subdifferential is well defined for all feasible points. Formally, the Clarke subdifferential is defined as

$$\partial_C J(K) = \mathrm{conv}\{\lim_{i \to \infty} \nabla J(K_i) : K_i \to K,\, K_i \in \mathcal{K}\} \tag{7}$$

where conv denotes the convex hull. Then we know the Clarke subdifferential for the $\mathcal{H}_\infty$ objective function (5) exist for all $K \in \mathcal{K}$. We say that K is a Clarke stationary point if $0 \in \partial_C J(K)$. The following fact is also well known.

Proposition 3. *If K is a local min of J, then $0 \in \partial_C J(K)$ and K is a Clarke stationary point.*

Under Assumption 1, it is well known that there exists finite $K^* \in \mathcal{K}$ achieving the minimum of (6). Since $\mathcal{K}$ is an open set, K^* has to be an interior point of $\mathcal{K}$ and hence K^* has to be a Clarke stationary point. In Section 3, we will prove that any Clarke stationary points for (6) are actually global minimum.

Finally, we briefly elaborate on the subdifferentially regular property stated in Proposition 2. For any given direction d (which has the same dimension as K), the Clarke directional derivative of J is defined as

$$J^\circ(K, d) = \lim_{K' \to K} \sup_{t \searrow 0} \frac{J(K' + td) - J(K')}{t}. \tag{8}$$

[1] We cannot find a formal statement of Proposition 2 in the literature. However, based on our discussion with other researchers who have worked on nonsmooth $\mathcal{H}_\infty$ synthesis for long time, this fact is well known and hence we do not claim any credits in deriving this result. As a matter of fact, although not explicitly stated, the proof of Proposition 2 is hinted in the last paragraph of [2, Section III] given the facts that the $\mathcal{H}_\infty$ norm is a convex function over the Hardy $\mathcal{H}_\infty$ space (which is a Banach space) and the mapping from $K \in \mathcal{K}$ to the (infinite-dimensional) Hardy $\mathcal{H}_\infty$ space is smooth. For completeness, a simple proof of Proposition 2 based on Clarke's chain rule is included in the supplementary material.

In contrast, the directional derivative is defined as follows (when existing)

$$J'(K,d) = \lim_{t \searrow 0} \frac{J(K+td) - J(K)}{t}. \tag{9}$$

In general, the Clarke directional derivative can be different from the directional derivative, and the directional derivative does not even exist for some locally Lipschitz functions. The objective function $J(K)$ is subdifferentially regular if for every $K \in \mathcal{K}$, the ordinary directional derivative always exists and coincides with the generalized one for every direction, i.e. $J'(K,d) = J^{\circ}(K,d)$. The most important consequence of the subdifferentially regular property is the following fact.

Corollary 1. *Suppose $K^{\dagger}$ is a Clarke stationary point for J. If J is subdifferentially regular, then the directional derivatives $J'(K^{\dagger}, d)$ are non-negative for all d.*

See [11] for related proofs. Notice that having non-negative directional derivatives does not mean that the point $K^{\dagger}$ is a local minimum. Nevertheless, the above fact will be used in our main theoretical developments. Now we briefly summarize two key difficulties in establishing a global convergence theory for direct policy search on the $\mathcal{H}_\infty$ state-feedback control problem (6). First, it is unclear whether the direct policy search will get stuck at local minimum. Second, it is challenging to guarantee the direct policy search to stay in the nonconvex feasible set $\mathcal{K}$. Since $\mathcal{K}$ is nonconvex, we cannot use the projection to maintain feasibility. Our main results will address these two issues.

2.4 Goldstein subdifferential

Generating a descent direction for nonsmooth optimization is not trivial. Many nonsmooth optimization algorithms are based on the concept of Goldstein subdifferential [22]. Before proceeding to our main result, we briefly review this concept here.

Definition 1 (Goldstein subdifferential)**.** *Suppose J is locally Lipschitz, Given a point $K \in \mathcal{K}$ and a parameter $\delta > 0$, the Goldstein subdifferential of J at K is the set*

$$\partial_\delta J(K) := \operatorname{conv}\left\{\cup_{K' \in \mathbb{B}_\delta(K)} \partial_C J(K')\right\} \tag{10}$$

where $\mathbb{B}_\delta(K)$ denotes the δ-ball around K.

A point K is (δ, ϵ)-stationary if $\operatorname{dist}(0, \partial_\delta J(K)) \leq \epsilon$. It is well-known that the minimal norm element of the Goldstein subdifferential generates a descent direction. This fact is stated as follows.

Proposition 4 ([22])**.** *Let F be the minimal norm element in $\partial_\delta J(K)$. Suppose $K - \alpha F \in \mathcal{K}$ for any $0 \leq \alpha \leq \delta$. Then we have*

$$J(K - \delta F/\|F\|_2) \leq J(K) - \delta \|F\|_2. \tag{11}$$

The idea of Goldstein subdifferential has been used in designing algorithms for nonsmooth $\mathcal{H}_\infty$ control [3, 24, 9, 12]. We will show that the Goldstein subdifferential method can be guaranteed to find the global minimum of (6).

3 Optimization Landscape for $\mathcal{H}_\infty$ State-Feedback Control

In this section, we investigate the optimization landscape of $\mathcal{H}_\infty$ state-feedback control, and show that any Clarke stationary points of (6) are also global minimum. We start by showing the coercivity of the $\mathcal{H}_\infty$ objective function (5).

Lemma 1. *The $\mathcal{H}_\infty$ objective function $J(K)$ defined by* (5) *is coercive over the set $\mathcal{K}$ in the sense that for any sequence $\{K^l\}_{l=1}^{\infty} \subset \mathcal{K}$ we have*

$$J(K^l) \to +\infty$$

if either $\|K^l\|_2 \to +\infty$, or K^l converges to an element in the boundary $\partial\mathcal{K}$.

Proof. Due to the space limit, we will only provide a proof sketch here. A detailed proof is presented in the supplementary material. Suppose we have a sequence $\{K^l\}$ satisfying $\|K^l\|_2 \to +\infty$. We can choose $\mathbf{w} = \{w_0, 0, 0, \cdots\}$ with $\|w_0\| = 1$ and show

$$J(K^l) \geq w_0^{\mathsf{T}}(Q + (K^l)^{\mathsf{T}} R K^l) w_0 \geq \lambda_{\min}(R) \|K^l w_0\|^2. \tag{12}$$

Clearly, we have used the positive definiteness of R in the above derivation. Then by carefully choosing w_0, we can ensure $J(K^l) \to +\infty$ as $\|K^l\|_2 \to +\infty$. Next, we assume $K^l \to K \in \partial\mathcal{K}$. Since $K^l \in \mathcal{K}$, we can use the bounded real lemma [55] to show the existence of a positive definite solution of some Riccati equation. Based on the assumption that Q is positive definite, we can take the limit of some subsequence of these Riccati solutions to construct a Lyapunov function showing K has to be stabilizing, i.e. $K \notin \partial\mathcal{K}$. This leads to a contradiction. More details can be found in the supplementary material. □

We want to emphasize that the positive definiteness of (Q, R) are crucial for proving the above result. Built upon Lemma 1, we can obtain the following nice properties of the sublevel sets of (6).

Lemma 2. *Consider the $\mathcal{H}_\infty$ state-feedback control problem* (6) *with the objective function $J(K)$ defined in* (5)*. Under Assumption 1, the sublevel set defined as $\mathcal{K}_\gamma = \{K \in \mathcal{K} : J(K) \leq \gamma\}$ is compact and path-connected for every $\gamma \geq J(K^*)$ where K^* is the global minimum of* (6)*.*

Proof. The compactness of $\mathcal{K}_\gamma$ directly follows from the continuity and coercivity of $J(K)$, and is actually a consequence of [5, Proposition 11.12]. The path-connectedness of the strict sublevel sets for the continuous-time $\mathcal{H}_\infty$ control problem has been proved in [26]. We can slightly modify the proof in [26] to show that the strict sublevel set $\{K \in \mathcal{K} : J(K) < \gamma\}$ is path-connected. Based on the fact that every non-strict sublevel sets are compact, now we can apply [36, Theorem 5.2] to show $\mathcal{K}_\gamma$ is also path-connected. An independent proof based on the non-strict version of the bounded real lemma is provided in the supplementary material. □

The path-connectedness of $\mathcal{K}_\gamma$ for every γ actually implies the uniqueness of the minimizing set in a certain sense [36, Sections 2&3]. Due to the space limit, we will defer the discussion on the uniqueness of the minimizing set to the supplementary material. Here, we present a stronger result which is one of the main contributions of our paper.

Theorem 1. *Consider the $\mathcal{H}_\infty$ state-feedback control problem* (6)*. Under Assumption 1, any Clarke stationary point of $J(K)$ is a global minimum.*

A detailed proof is presented in the supplementary material. Here we provide a proof sketch. Since Q and R are positive definite, the non-strict version of the bounded real lemma [6, Section 2.7.3] states that $J(K) \leq \gamma$ if and only if there exists a positive definite matrix P such that the following matrix inequality holds

$$\begin{bmatrix} (A-BK)^\mathsf{T} P (A-BK) - P & P(A-BK) \\ (A-BK)^\mathsf{T} P & P \end{bmatrix} + \begin{bmatrix} Q + K^\mathsf{T} R K & 0 \\ 0 & -\gamma^2 I \end{bmatrix} \preceq 0. \tag{13}$$

The above matrix inequality is linear in P but not linear in K. A standard trick from the control theory can be combined with the Schur complement lemma to convert the above matrix inequality condition to another condition which is linear in all the decision variable [6]. Specifically, there exists a matrix function $\mathrm{LMI}(Y, L, \gamma)$ which is linear in (Y, L, γ) and satisfies $\mathrm{LMI}(Y, L, \gamma) \preceq 0$ if and only if (13) is feasible with $K = LY^{-1}$ and $P = Y^{-1}$. The matrix function $\mathrm{LMI}(Y, L, \gamma)$ involves a larger matrix. Hence we present the analytical formula of $LMI(Y, L, \gamma)$ in the supplementary material and skip it here. Since $\mathrm{LMI}(Y, L, \gamma)$ is linear in (Y, L, γ), and hence $\mathrm{LMI}(Y, L, \gamma) \preceq 0$ is just a convex semidefinite programming condition. Based on this convex necessary and sufficient condition for $J(K) \leq gamma$, we can prove the following important lemma.

Lemma 3. *For any $K \in \mathcal{K}$ satisfying $J(K) > J^*$, there exists $d \neq 0$ such that $J'(K, d) = J^* - J(K) < 0$, where $J^* = J(K^*)$ and K^* is the global minimum of* (6)*.*

Proof. Suppose we have $K = LY^{-1}$ where $(Y, L, J(K))$ is a feasible point for the convex regime $\mathrm{LMI}(Y, L, J(K)) \preceq 0$. In addition, we have $K^* = L^*(Y^*)^{-1}$ where $(Y^*, L^*, J(K^*))$ is a point satisfying $\mathrm{LMI}(Y^*, L^*, J(K^*)) \preceq 0$. Since the LMI condition is convex, we know that the line segment between $(Y, L, J(K))$ and (Y^*, Q^*, J^*) is also in this convex set. For any $t > 0$, we know $(Y + t\Delta Y, L + t\Delta L, J(K) + t(J^* - J(K)))$ also satisfies $\mathrm{LMI}(Y + t\Delta Y, L + t\Delta L, J(K) + t(J^* - J(K))) \preceq 0$, where $\Delta L = L - L^*$, and $\Delta Y = Y - Y^*$. Let's choose $d = \Delta L Y + L Y^{-1} \Delta Y Y^{-1}$. Then we have

$$\begin{aligned} J'(K, d) &= \lim_{t \to 0} \left(\frac{J((L + t\Delta L)(Y + t\Delta Y)^{-1}) - J(K)}{t} + o(t) \right) \\ &= J^* - J(K) < 0. \end{aligned}$$

A detailed verification of the above inequality is provided in the supplementary material. Notice $d \neq 0$. Otherwise we have $\Delta L = 0$, $\Delta Q = 0$, and $K = K^*$. This contradicts the fact $J(K) > J^*$. It is also impossible to have $\Delta L + LQ^{-1}\Delta Q Q^{-2} = 0$ because we know $J(K) > J^*$. This completes the proof for this lemma. □

Now we are ready to sketch the rest of the proof for Theorem 1. Suppose K^* is the global minimum, and $K^\dagger$ is a Clarke stationary point. Since J is subdifferentially regular, we know all the directional derivatives at $K^\dagger$ are non-negative. There are two possibilities here. If K has positive directional derivatives at all the directions, then clearly $K^\dagger$ is a strictly local minimum. Based on the path-connectedness of all the non-strict sublevel sets of (6), $K^\dagger$ has to be global minimum. For the second possibility, if $K^\dagger$ has one directional derivative which is 0, then we can apply Lemma 3 to show that $K^\dagger$ has to be a global minimum. Otherwise, there exists one direction such that the directional derivative becomes strictly negative and this contradicts the fact that J is subdifferentially regular. Therefore, in both cases, $K^\dagger$ has to be the global minimum of (6). We will provide a more detailed proof in the supplementary material.

The above proof relies on Lemma 3 and the fact that J is subdifferentially regular. Without using the subdifferential regular property, Lemma 3 itself is not sufficient for proving Theorem 1. It is also worth mentioning that Lemma 3 can be viewed as an extension of the lifting technique in [47, Theorem 4] to non-differentiable points.

4 Global Convergence of Direct Policy Search on $\mathcal{H}_\infty$ State-Feedback Control

In this section, we first show that the Goldstein subdifferential method can be guaranteed to stay in the nonconvex feasible regime $\mathcal{K}$ during the optimization process and eventually converge to the global minimum of (6). Then we further discuss the convergence guarantees for various more implementable forms of the Goldstein subdifferential method. Due to the space limit, a detailed discussion on the results for these more implementable variants of the Goldstein subdifferential method is deferred to the supplementary material.

4.1 Global convergence of the Goldstein subdifferential method

We will investigate the global convergence of the following Goldstein subdifferential method for policy search of the $\mathcal{H}_\infty$ optimal state-feedback policy:

$$K^{n+1} = K^n - \delta^n F^n / \|F^n\|_2 \tag{14}$$

where F_n is the minimum norm element of the Goldstein subdifferential $\partial_{\delta^n} J(K^n)$. We assume that an initial stabilizing policy is available, i.e. $K^0 \in \mathcal{K}$. The same initial policy assumption has also been made in the global convergence theory for direct policy search on LQR [18]. More recently, there has been some provable guarantees for obtaining such stabilizing policy via direct policy search [40]. Hence such an assumption on the initial policy K^0 is reasonable. Our global convergence result relies on the fact that there is a strict separation between any sublevel set of (6) and the boundary of $\mathcal{K}$. This fact is formalized as follows.

Lemma 4. *Consider the $\mathcal{H}_\infty$ state-feedback control problem* (6) *with the objective function $J(K)$ defined in* (5)*. Suppose $\gamma \geq J^*$. Under Assumption 1, there is a strict separation between the sublevel set $\mathcal{K}_\gamma$ and $\mathcal{K}^c$ which is the complement of the feasible set $\mathcal{K}$. In other words, we have $dist(\mathcal{K}_\gamma, \mathcal{K}^c) > 0$.*

Proof. Obviously the set $\mathcal{K}_\gamma \cap \mathcal{K}^c$ is empty (since we know $\mathcal{K}_\gamma \subset \mathcal{K}$). Based on Lemma 2, we know $\mathcal{K}_\gamma$ is compact. Since $\mathcal{K}$ is open, we know $\mathcal{K}^c$ is closed. Therefore, there is a strict separation between $\mathcal{K}_\gamma$ and $\mathcal{K}^c$, and we have $dist(\mathcal{K}_\gamma, \mathcal{K}^c) > 0$. □

Now we are ready to present our main convergence result.

Theorem 2. *Consider the $\mathcal{H}_\infty$ state-feedback control problem* (6) *with the objective function $J(K)$ defined in* (5)*. Suppose an initial stabilizing policy is given, i.e. $K^0 \in \mathcal{K}$. Denote $\Delta_0 := \mathrm{dist}(\mathcal{K}_{J(K^0)}, \mathcal{K}^c) > 0$. Choose $\delta^n = \frac{\Delta_0}{n}$ for all n. Then the Goldstein subdifferential algorithm* (14) *is guaranteed to stay in $\mathcal{K}$ for all n. In addition, we have $J(K^n) \to J^*$ as $n \to \infty$.*

Proof. We have $\delta^n \leq \Delta_0$ for all n. Now we use an induction proof to show $K^n \in \mathcal{K}_{J(K^0)}$ for all n. For $n = 1$, we know $K^0 - \Delta_0 F^0/\|F_0\|_2$ has to be within the Δ_0 ball around K^0 since we know the norm of $F^0/\|F_0\|_2$ is exactly equal to 1. Since $\Delta_0 := \mathrm{dist}(\mathcal{K}_{J(K^0)}, \mathcal{K}^c) > 0$, we know $K^0 - \Delta_0 F^0/\|F_0\|_2 \in \mathcal{K}$. As a matter of fact, we know $\mathbb{B}_{\delta^0}(K^0)$ has to be a subset of $\mathcal{K}$. Hence we can apply (11) to show that K^1 exists and is also in $\mathcal{K}_{J(K^0)}$. Similarly, we can repeat this argument to show $K^n \in \mathcal{K}_{J(K^)}$ for all n. Next, we can apply (11) to every step and then sum the inequalities over all n. . Then the following inequality holds for all N:

$$\sum_{n=0}^{N} \delta_n \|F^n\|_2 \leq J(K^0) - J^* \tag{15}$$

Since we have $\sum_{n=0}^{\infty} \delta^n = \infty$, we know $\liminf_{n\to\infty} \|F^n\|_2 \to 0$. There exists one subsequence of $\{i_n\}$ such that $\|G_{i_n}\|_2 \to 0$. For this subsequence, the resultant policy sequence K is also bounded (think about that the policy parameter sequence stays in the compact set $\mathcal{K}_{J(K^0)}$ for all n) and has a convergent subsequence. The limit of this subsequence has to be a Clarke stationary point. Otherwise, it will contract the fact that $\|G_{i_n}\|_2 \to 0$. Therefore, the function value $\{J(K^{i_n})\}$ associated with this subsequence converges to J^*. Notice that $J(K^n)$ is monotonically decreasing for the entire sequence $\{n\}$. Hence we have $J(K_n) \to J^*$. This completes the proof. □

We have tried to be brief in giving the above proof. We present a more detailed proof in the supplementary material. We believe that this is the first global convergence result for direct policy search on $\mathcal{H}_\infty$ state-feedback control problem.

4.2 Implementable variants and related convergence results

Notice that for practical problems it can be difficult to evaluate the minimum norm element of the Goldstein subdifferential. There are several more implementable variants available. We briefly discuss several implementable variants of the Goldstein subdifferential algorithm and the associated convergence guarantees here. We will present several new global convergence results for these implementable variants in the supplementary material. A detailed discussion on these implementable variants and related global conve

1. The sampling algorithm in [49]: A special sampling method has been proposed in [49] to approximate the minimum norm element in the Goldstein subdifferential for directionally diffeerentiable functions. The $\mathcal{H}_\infty$ objective function (6) is subdifferentially regular and hence the algorithm in [49] can be directly applied. Again, we can choose δ^n based on the distance parameter $\Delta_0 := \mathrm{dist}(\mathcal{K}_{J(K^0)}, \mathcal{K}^c)$ to ensure that the policies generated by this algorithm stays in $\mathcal{K}$ with high probability. Then we can modify the high probability iteration complexity for finding (ϵ, δ) stationary point via this algorithm to obtain a high probability global convergence result.
2. The sampling algorithm in [13]: A new variant of the Goldstein subdifferential method has been proposed in [13] to simplify the oracle and improve the iteration complexity for low dimensional problems (the case where the dimension of K is relatively low). This algorithm also has high probability iteration complexity, and can be modified similarly to yield a global convergence result on the $\mathcal{H}_\infty$ state-feedback control problem.
3. Gradient sampling techniques: There was a large class of old algorithms developed based on the concepts of gradient sampling [9, 10, 31]. See [9] for a comprehensive review. These methods are also based on the concept of Goldstein subdifferential. It is very straightforward to use the information of Δ_0 to force the iterations of such algorithms to stay in $\mathcal{K}$. There are theoretical results showing that the limit points of these methods have to be Clarke stationary points [9]. But the proof requires an extra assumption that $J(K)$ is continuously differentiable on a set of full measure. Under this extra assumption, we can modify the existing theory on asymptotic convergence to stationary points to show that that the gradient sampling methods can be guaranteed to find the global solutions of (6). It is also worth mentioning that the gradient sampling method is the main workhorse for the robust control package HIFOO [3, 24].
4. Derivative-free version of gradient sampling: The gradient sampling method has also been implemented in the derivative-free manner when only zeroth-order oracle is used [32]. The theory in [32] can also be adopted to the $\mathcal{H}_\infty$ control problem (6).

References

[1] P. Apkarian and D. Noll. Controller design via nonsmooth multidirectional search. *SIAM Journal on Control and Optimization*, 44(6):1923–1949, 2006.

[2] P. Apkarian and D. Noll. Nonsmooth $\mathcal{H}_\infty$ synthesis. *IEEE Transactions on Automatic Control*, 51(1):71–86, 2006.

[3] D. Arzelier, G. Deaconu, S. Gumussoy, and D. Henrion. H2 for HIFOO. In *International Conference on Control and Optimization With Industrial Applications (COIA 2011)*, 2011.

[4] T. Başar and P. Bernhard. *H^∞-Optimal Control and Related Minimax Design Problems*. Birkhäuser, 2nd edition, 1995.

[5] H. H. Bauschke, P. L. Combettes, et al. *Convex analysis and monotone operator theory in Hilbert spaces*, volume 408. Springer, 2011.

[6] S. Boyd, L. El Ghaoui, E. Feron, and V. Balakrishnan. *Linear Matrix Inequalities in System and Control Theory*, volume 15 of *Studies in Applied Mathematics*. SIAM, 1994.

[7] J. Bu, A. Mesbahi, M. Fazel, and M. Mesbahi. LQR through the lens of first order methods: Discrete-time case. *arXiv preprint arXiv:1907.08921*, 2019.

[8] J. Bu, A. Mesbahi, and M. Mesbahi. On topological and metrical properties of stabilizing feedback gains: the MIMO case. *arXiv preprint arXiv:1904.02737*, 2019.

[9] J. V. Burke, F. E. Curtis, A. S. Lewis, M. L. Overton, and L. E. Simões. Gradient sampling methods for nonsmooth optimization. *Numerical Nonsmooth Optimization*, pages 201–225, 2020.

[10] J. V. Burke, A. S. Lewis, and M. L. Overton. A robust gradient sampling algorithm for nonsmooth, nonconvex optimization. *SIAM Journal on Optimization*, 15(3):751–779, 2005.

[11] F. H. Clarke. *Optimization and nonsmooth analysis*. SIAM, 1990.

[12] F. E. Curtis, T. Mitchell, and M. L. Overton. A BFGS-SQP method for nonsmooth, nonconvex, constrained optimization and its evaluation using relative minimization profiles. *Optimization Methods and Software*, 32(1):148–181, 2017.

[13] D. Davis and D. Drusvyatskiy. A gradient sampling method with complexity guarantees for general lipschitz functions. *arXiv preprint arXiv:2112.06969*, 2021.

[14] J. Doyle, K. Glover, P. Khargonekar, and B. Francis. State-space solutions to standard $\mathcal{H}_2$ and $\mathcal{H}_\infty$ control problems. *IEEE Transactions on Automatic Control*, 34:831–847, 1989.

[15] Y. Duan, X. Chen, R. Houthooft, J. Schulman, and P. Abbeel. Benchmarking deep reinforcement learning for continuous control. In *International Conference on Machine Learning*, pages 1329–1338, 2016.

[16] G. Dullerud and F. Paganini. *A Course in Robust Control Theory: A Convex Approach*. Springer, 1999.

[17] I. Fatkhullin and B. Polyak. Optimizing static linear feedback: Gradient method. *SIAM Journal on Control and Optimization*, 59(5):3887–3911, 2021.

[18] M. Fazel, R. Ge, S. Kakade, and M. Mesbahi. Global convergence of policy gradient methods for the linear quadratic regulator. In *Proceedings of the 35th International Conference on Machine Learning*, volume 80, pages 1467–1476, 2018.

[19] L. Furieri, Y. Zheng, and M. Kamgarpour. Learning the globally optimal distributed LQ regulator. In *Learning for Dynamics and Control*, pages 287–297, 2020.

[20] P. Gahinet and P. Apkarian. A linear matrix inequality approach to H_∞ control. *International Journal of Robust and Nonlinear Control*, 4:421–448, 1994.

[21] K. Glover and J. C. Doyle. State-space formulae for all stabilizing controllers that satisfy an $\mathcal{H}_\infty$-norm bound and relations to relations to risk sensitivity. *Systems & Control Letters*, 11(3):167–172, 1988.

[22] A. Goldstein. Optimization of lipschitz continuous functions. *Mathematical Programming*, 13(1):14–22, 1977.

[23] B. Gravell, P. M. Esfahani, and T. Summers. Learning optimal controllers for linear systems with multiplicative noise via policy gradient. *IEEE Transactions on Automatic Control*, 66(11):5283–5298, 2020.

[24] S. Gumussoy, D. Henrion, M. Millstone, and M. L. Overton. Multiobjective robust control with HIFOO 2.0. *IFAC Proceedings Volumes*, 42(6):144–149, 2009.

[25] B. Hambly, R. Xu, and H. Yang. Policy gradient methods for the noisy linear quadratic regulator over a finite horizon. *SIAM Journal on Control and Optimization*, 59(5):3359–3391, 2021.

[26] B. Hu and Y. Zheng. Connectivity of the feasible and sublevel sets of dynamic output feedback control with robustness constraints. *arXiv preprint arXiv:2203.11177*, 2022.

[27] J. P. Jansch-Porto, B. Hu, and G. Dullerud. Policy learning of MDPs with mixed continuous/discrete variables: A case study on model-free control of Markovian jump systems. In *Learning for Dynamics and Control*, pages 947–957, 2020.

[28] J. P. Jansch-Porto, B. Hu, and G. E. Dullerud. Convergence guarantees of policy optimization methods for Markovian jump linear systems. In *American Control Conference*, pages 2882–2887, 2020.

[29] I. Kaminer, P. P. Khargonekar, and M. A. Rotea. Mixed $\mathcal{H}_2/\mathcal{H}_\infty$ control for discrete-time systems via convex optimization. *Automatica*, 29(1):57–70, 1993.

[30] P. P. Khargonekar and M. A. Rotea. Mixed $\mathcal{H}_2/\mathcal{H}_\infty$ control: A convex optimization approach. *IEEE Transactions on Automatic Control*, 36(7):824–837, 1991.

[31] K. C. Kiwiel. Convergence of the gradient sampling algorithm for nonsmooth nonconvex optimization. *SIAM Journal on Optimization*, 18(2):379–388, 2007.

[32] K. C. Kiwiel. A nonderivative version of the gradient sampling algorithm for nonsmooth nonconvex optimization. *SIAM Journal on Optimization*, 20(4):1983–1994, 2010.

[33] Y. Li, Y. Tang, R. Zhang, and N. Li. Distributed reinforcement learning for decentralized linear quadratic control: A derivative-free policy optimization approach. *IEEE Transactions on Automatic Control*, 2021.

[34] T. P. Lillicrap, J. J. Hunt, A. Pritzel, N. Heess, T. Erez, Y. Tassa, D. Silver, and D. Wierstra. Continuous control with deep reinforcement learning. *arXiv preprint arXiv:1509.02971*, 2015.

[35] D. Malik, A. Pananjady, K. Bhatia, K. Khamaru, P. Bartlett, and M. Wainwright. Derivative-free methods for policy optimization: Guarantees for linear quadratic systems. *arXiv:1812.08305*, 2018.

[36] D. H. Martin. Connected level sets, minimizing sets, and uniqueness in optimization. *Journal of Optimization Theory and Applications*, 36(1):71–91, 1982.

[37] H. Mohammadi, A. Zare, M. Soltanolkotabi, and M. R. Jovanovic. Convergence and sample complexity of gradient methods for the model-free linear quadratic regulator problem. *IEEE Transactions on Automatic Control*, 2021.

[38] D. Mustafa and D. S. Bernstein. LQG cost bounds in discrete-time $\mathcal{H}_2/\mathcal{H}_\infty$ control. *Transactions of the Institute of Measurement and Control*, 13(5):269–275, 1991.

[39] D. Noll and P. Apkarian. Spectral bundle methods for non-convex maximum eigenvalue functions: second-order methods. *Mathematical Programming*, 104(2):729–747, 2005.

[40] J. Perdomo, J. Umenberger, and M. Simchowitz. Stabilizing dynamical systems via policy gradient methods. *Advances in Neural Information Processing Systems*, 34, 2021.

[41] S. Rathod, M. Bhadu, and A. De. Global convergence using policy gradient methods for model-free markovian jump linear quadratic control. *arXiv preprint arXiv:2111.15228*, 2021.

[42] J. Schulman, S. Levine, P. Abbeel, M. Jordan, and P. Moritz. Trust region policy optimization. In *International Conference on Machine Learning*, pages 1889–1897, 2015.

[43] J. Schulman, P. Moritz, S. Levine, M. Jordan, and P. Abbeel. High-dimensional continuous control using generalized advantage estimation. *arXiv preprint arXiv:1506.02438*, 2015.

[44] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov. Proximal policy optimization algorithms. *arXiv preprint arXiv:1707.06347*, 2017.

[45] S. Skogestad and I. Postlethwaite. *Multivariable Feedback Control: Analysis and Design*, volume 2. Wiley New York, 2007.

[46] R. S. Sutton, D. A. McAllester, S. P. Singh, and Y. Mansour. Policy gradient methods for reinforcement learning with function approximation. In *Advances in Neural Information Processing Systems*, pages 1057–1063, 2000.

[47] J. Umenberger, M. Simchowitz, J. C. Perdomo, K. Zhang, and R. Tedrake. Globally convergent policy search over dynamic filters for output estimation. *arXiv preprint arXiv:2202.11659*, 2022.

[48] Z. Yang, Y. Chen, M. Hong, and Z. Wang. On the global convergence of actor-critic: A case for linear quadratic regulator with ergodic cost. *arXiv preprint arXiv:1907.06246*, 2019.

[49] J. Zhang, H. Lin, S. Jegelka, S. Sra, and A. Jadbabaie. Complexity of finding stationary points of nonconvex nonsmooth functions. In *Proceedings of the 37th International Conference on Machine Learning*, volume 119, pages 11173–11182. PMLR, 2020.

[50] K. Zhang, B. Hu, and T. Başar. On the stability and convergence of robust adversarial reinforcement learning: A case study on linear quadratic systems. *Advances in Neural Information Processing Systems*, 33, 2020.

[51] K. Zhang, B. Hu, and T. Basar. Policy optimization for $\mathcal{H}_2$ linear control with $\mathcal{H}_\infty$ robustness guarantee: Implicit regularization and global convergence. *SIAM Journal on Control and Optimization*, 59(6):4081–4109, 2021.

[52] K. Zhang, X. Zhang, B. Hu, and T. Başar. Derivative-free policy optimization for linear risk-sensitive and robust control design: Implicit regularization and sample complexity. In *Thirty-Fifth Conference on Neural Information Processing Systems*, 2021.

[53] F. Zhao and K. You. Primal-dual learning for the model-free risk-constrained linear quadratic regulator. In *Learning for Dynamics and Control*, pages 702–714, 2021.

[54] Y. Zheng, Y. Tang, and N. Li. Analysis of the optimization landscape of linear quadratic gaussian (LQG) control. *arXiv preprint arXiv:2102.04393*, 2021.

[55] K. Zhou, J. Doyle, and K. Glover. *Robust and Optimal Control*. Prentice-Hall, 1996.

Checklist

1. For all authors...
   (a) Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope? [Yes]
   (b) Did you describe the limitations of your work? [Yes] Our works only addresses the $\mathcal{H}_\infty$ state-feedback control problem. More study is needed for the partial observation case.
   (c) Did you discuss any potential negative societal impacts of your work? [Yes] We do not believe that our current work will create any negative societal impacts.
   (d) Have you read the ethics review guidelines and ensured that your paper conforms to them? [Yes]
2. If you are including theoretical results...
   (a) Did you state the full set of assumptions of all theoretical results? [Yes]
   (b) Did you include complete proofs of all theoretical results? [Yes] See Appendix in the supplementary material.
3. If you ran experiments...
   (a) Did you include the code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL)? [N/A] The nature of our work is theoretical.
   (b) Did you specify all the training details (e.g., data splits, hyperparameters, how they were chosen)? [N/A]
   (c) Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)? [N/A]
   (d) Did you include the total amount of compute and the type of resources used (e.g., type of GPUs, internal cluster, or cloud provider)? [N/A]
4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets...
   (a) If your work uses existing assets, did you cite the creators? [N/A]
   (b) Did you mention the license of the assets? [N/A]
   (c) Did you include any new assets either in the supplemental material or as a URL? [N/A]
   (d) Did you discuss whether and how consent was obtained from people whose data you're using/curating? [N/A]
   (e) Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content? [N/A]
5. If you used crowdsourcing or conducted research with human subjects...
   (a) Did you include the full text of instructions given to participants and screenshots, if applicable? [N/A]
   (b) Did you describe any potential participant risks, with links to Institutional Review Board (IRB) approvals, if applicable? [N/A]
   (c) Did you include the estimated hourly wage paid to participants and the total amount spent on participant compensation? [N/A]