Arm-Aware Guided Dexterous Grasp Generation with Arm-Agnostic Grasp Models

***Abstract*—Dexterous grasp generation that considers arm-related constraints is crucial in real-world scenarios involving arm-environment collision avoidance, workspace boundary grasps, and consecutive grasping. Existing hand-centric grasp models, which primarily focus on the floating hand's pose, are insufficient for such cases. Conventional arm-aware methods either rely on rejection sampling to discard infeasible samples or require retraining on arm-specific data, leading to low sample efficiency under adverse conditions or limited generalization across different robots and environments. To overcome these limitations, this letter presents an arm-aware dexterous grasp generation framework that leverages pretrained arm-agnostic grasp models while integrating arm and environmental information only at inference time. Specifically, we formulate arm-aware constrained grasp generation as a joint optimization of hand pose and arm configuration, and derive closed-form gradients for arm-related constraints. Assuming the hand pose distribution is represented by a diffusion model, we prove that gradient-based optimization is equivalent to guided diffusion sampling, steering near-feasible samples toward the feasible region. Through comprehensive evaluation involving 10k objects across 6 scenarios, we demonstrate that the proposed framework generates feasible grasps in highly constrained settings with significantly higher probability, highlighting its advantages in real-world applications. Supplementary materials are available at `https://arm-aware-dexgrasp.github.io/`.**

***Index Terms*—Dexterous grasp generation, arm-aware manipulation, guided diffusion optimization.**

I. Introduction

DEXTEROUS grasp generation, which generates grasp poses for dexterous hands based on object and environment information, provides a target grasp configuration for grasp execution [1], [2], and serves as a precondition for the subsequent robotic manipulation [3].

Existing grasp generation methods predominantly employ a *hand-centric* scheme, which focuses primarily on learning the distribution of a free-floating hand's grasp poses, overlooking the robotic arm and critical environmental context, such as obstacles [2], [4], [5], [6]. However, focusing solely on the hand is often insufficient in practice. For example, when grasping in a constrained space, the arm must avoid collisions with surrounding objects to ensure safety. Moreover, grasping objects near the boundary of the arm's reachable space constrains the wrist orientation to ensure the existence of inverse kinematics (IK) solutions. Additionally, minimizing arm motion is important for execution efficiency, particularly in tasks involving consecutive grasping and confined pick-and-place. Therefore, it is essential to incorporate arm-related constraints into the grasp generation process.

To address this issue, a straightforward solution is *rejection sampling*, where multiple candidate grasps are generated and those violating arm-related constraints are filtered out before execution. Although this strategy is widely used in practice, it significantly reduces sampling efficiency—especially in highly constrained environments where the feasible region occupies a small volume relative to the full solution space, as many near-feasible samples are discarded without correction. An alternative is to directly train a model on synthetic data that involves both the hand and arm, which already satisfies the constraints [1]. However, this approach suffers from limited generalizability, as the grasp model tends to overfit to a specific arm and environment setup. Adapting to a different robotic arm or environment necessitates additional data synthesis and retraining the model, which is time-consuming.

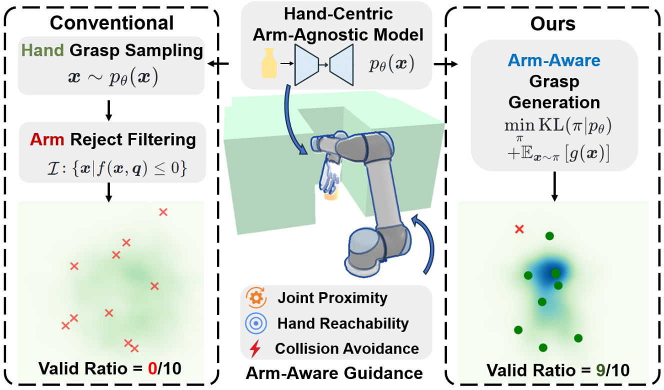

Conventional
Hand Grasp Sampling
$\boldsymbol{x} \sim p_\theta(\boldsymbol{x})$
Arm Reject Filtering
$\mathcal{I}$: $\{\boldsymbol{x} | f(\boldsymbol{x}, \boldsymbol{q}) \leq 0\}$
Valid Ratio = 0/10
Hand-Centric
Arm-Agnostic Model
$p_\theta(\boldsymbol{x})$
Joint Proximity
Hand Reachability
Collision Avoidance
Arm-Aware Guidance
Ours
Arm-Aware
Grasp
Generation
$\min_\pi \mathrm{KL}(\pi | p_\theta) + \mathbb{E}_{\boldsymbol{x}\sim\pi}[g(\boldsymbol{x})]$
Valid Ratio = 9/10

Fig. 1: Motivation of arm-aware dexterous grasp generation. Real-world grasp execution requires considering arm-related constraints, yet conventional methods rely on rejection sampling to discard even near-feasible hand poses, leading to low efficiency. Our approach guides pretrained arm-agnostic models with these constraints at inference, greatly improving sampling efficiency.

In contrast, this letter proposes an *arm-aware* dexterous grasp generation method (shown in Fig. 1) by reformulating the problem as a joint optimization of grasp pose and arm configuration, utilizing only pre-trained arm-agnostic (i.e., hand-centric) grasp generation models. Our approach enhances sampling efficiency by transforming near-feasible samples into feasible ones through inference-phase guidance with arm-related constraints. Additionally, we express all arm-related constraints using closed-form equations and gradients, enabling our approach to be applicable to any robotic arms and environments, based on pre-modeled analytical arm kinematics and environment SDFs, without the need for costly data synthesis and re-training. Assuming the arm-agnostic grasp distribution is expressed as a diffusion model, we demonstrate that solving the proposed optimization problem using a primal–dual (PD) method corresponds to guided grasp-pose denoising within the diffusion framework. While incorporating guidance in the diffusion framework is common, designing guidance that maps between joint-space (arm configuration) constraints and Cartesian-space (hand pose) denoising is challenging.

Contributions of the letter can be summarized as follows:

1) We formulate arm-aware grasp generation as a joint optimization of grasp pose and arm configuration, deriving its relation to guided sampling on the pre-trained arm-agnostic grasp diffusion model with added arm constraints.
2) We derive analytical forms and gradients for three commonly used arm-related constraints (i.e., collision avoidance, hand reachability, and joint proximity) to create the gradient for guidance, addressing the complex mapping between joint-space constraints and Cartesian-space denoising.
3) We design comprehensive benchmark scenarios for simulation and real-world evaluation, featuring high obstacle coverage and grasps near arm limits, which thoroughly verifies that our method generates successful grasps that satisfy constraints with a significantly higher probability than the commonly used rejection sampling strategy. The proposed approach is applicable to various robotic arms (e.g., UR5 and Franka) and environments, utilizing a single hand-centric grasp generation model.

II. Related Works

A. Dexterous Grasp Generation

Dexterous grasp generation aims to predict feasible grasp poses for robotic hands from object meshes or point clouds, typically leveraging large-scale datasets. Early methods relied on supervised learning to directly regress grasp poses and evaluate grasp quality [7], [8], or to construct object-centric contact representations that are later converted into grasp configurations through optimization or regression [9]. To better capture the multimodal nature of feasible grasp distributions, generative models have been adopted, improving grasp diversity and generalization [2]. For instance, Li et al. [10] proposed a conditional variational autoencoder (CVAE) that jointly models hand rotation, translation, and joint articulation. More recently, diffusion models and normalizing flows have shown strong scene-conditioned distribution modeling capabilities and are rapidly becoming prevalent for grasp generation. Some recent approaches directly leverage diffusion and flow-based models to jointly generate grasp poses and their associated hand joint configurations, improving the diversity and expressiveness of grasp distributions [5], [11], [12]. Meanwhile, other methods learn contact geometry and force distributions through diffusion, allowing the resulting grasp representations to generalize across different dexterous hand configurations [13], [14]. However, despite these advances in cross-hand generalization, few methods explicitly account for arm kinematics or environmental obstacles during generation. This hand-centric assumption ultimately restricts execution success, particularly in spatially or kinematically constrained settings. While some recent works attempt to incorporate the robotic arm during data synthesis to produce arm-hand joint datasets [1], they are typically limited to a specific arm and simple tabletop scenes, hindering generalization to different arm embodiments or diverse environments.

B. Guided Diffusion Generation in Robotics

An advantage of diffusion models is their ability to integrate guidance during sampling, allowing the generated trajectories to satisfy task-specific constraints. Classifier guidance [15] and classifier-free guidance (CFG) [16] steer the sampling process toward a desired category using log-probability gradients, while energy-based guidance injects the gradients of differentiable cost functions directly into the denoising steps. In robotics, guided diffusion has been successfully used in motion planning by incorporating collision penalties, kinematic constraints, and joint limits into sampling, yielding feasible trajectories in joint space [17], [18], [19], [20]. Diffusion guidance has also been applied to grasp generation. Weng et al. employed a learning-based evaluator to provide score-based guidance during diffusion sampling [21], whereas Zhong et al. introduced a physics-guided sampler that leverages explicit gradients of differentiable physical grasp metrics to guide the sampling process [22]. Other works leverage language instructions as guidance to identify the target object or part, enabling semantically meaningful and functionally relevant grasping [23], [24]. However, most of these methods primarily focus only on the local optimization of hand configurations to improve grasp quality, while overlooking the kinematic feasibility and environmental constraints imposed by the robotic arm bodies, especially the need for global optimization over the whole-arm configurations in highly constrained spaces.

III. Methods

In this section, we detail the problem formulation and solution for arm-aware dexterous grasp generation, followed by the derivation of analytic forms and gradients for arm-related constraints. See Fig. 2 for an overview of the proposed method.

A. Preliminaries

Classical dexterous grasp generation aims to learn a distribution $p(\mathcal{G}|\mathcal{O})$ over a synthesized grasp dataset, given the partially observed object information $\mathcal{O}$ (i.e., a single-view point-cloud) [22]. A dexterous grasp $\mathcal{G} = \left(\boldsymbol{q}^{\text{hand}}, \boldsymbol{x}\right)$ contains the hand joint angles $\boldsymbol{q}^{\text{hand}} \in \mathbb{R}^{n_q}$ and the wrist pose $\boldsymbol{x} \in \mathrm{SE}(3)$. As observed in [1], the wrist pose distribution $p(\boldsymbol{x}|\mathcal{O})$ typically exhibits a multi-modal pattern, which can be effectively captured by diffusion probabilistic models [25]. Once a wrist pose $\boldsymbol{x}$ is sampled, the joint angles $\boldsymbol{q}^{\text{hand}}$ can be deterministically predicted by a regression network. We therefore adopt a diffusion model to generate wrist poses. During training, a clean pose $\boldsymbol{x}_0$ is randomly drawn from the dataset and perturbed with Gaussian noise $\boldsymbol{\epsilon}_t$ as

$$\boldsymbol{x}_t = \sqrt{\bar{\alpha}_t}\boldsymbol{x}_0 + \sqrt{1-\bar{\alpha}_t}\boldsymbol{\epsilon}_t. \tag{1}$$

Here, $\bar{\alpha}_t = \prod_{s=1}^{t}(1-\beta_s)$ denotes the cumulative noise decay factor, and $\beta_s \in (0, 1)$ determines the variance of the Gaussian noise added in diffusion step s. The optimization objective is derived to minimize the difference between the added noise and predicted noise

$$\mathcal{L}_\theta := \mathrm{MSELoss}(\boldsymbol{\epsilon}_t, \boldsymbol{\epsilon}_\theta(\boldsymbol{x}_t, \mathcal{O}, t)). \tag{2}$$

After training, the learned noise predictor $\boldsymbol{\epsilon}_\theta$ is used during inference to perform iterative denoising. Specifically, we follow

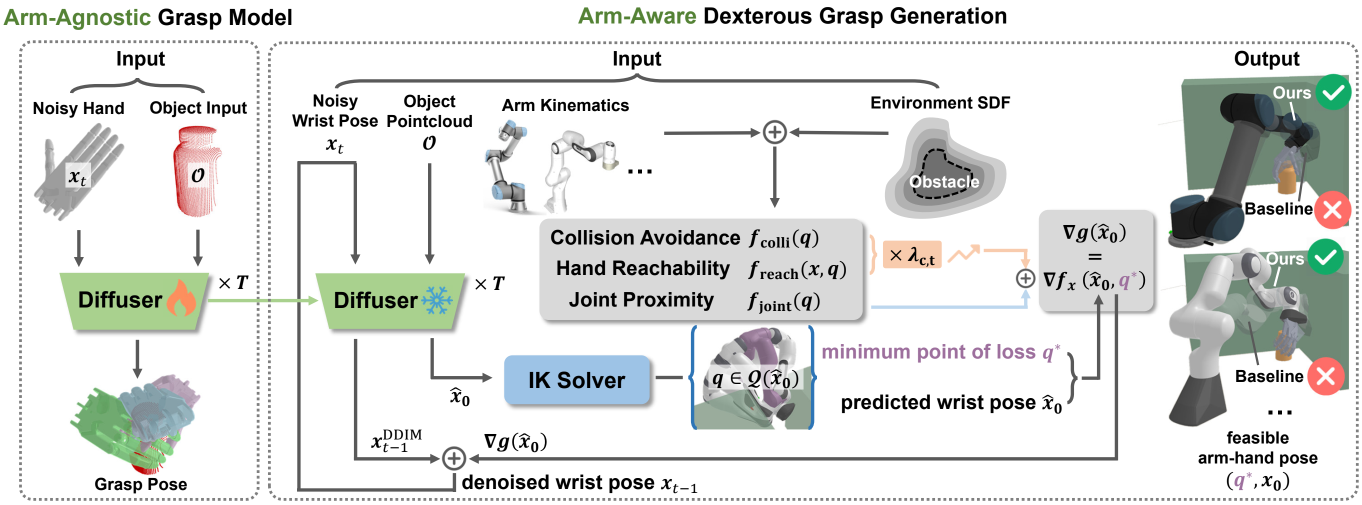

Arm-Agnostic Grasp Model
Arm-Aware Dexterous Grasp Generation
Input
Noisy Hand
Object Input
x_t
$\mathcal{O}$
Diffuser
× T
Grasp Pose
Input
Noisy Wrist Pose x_t
Object Pointcloud $\mathcal{O}$
Arm Kinematics
...
Environment SDF
Obstacle
Diffuser
× T
Collision Avoidance $f_{\text{colli}}(q)$
Hand Reachability $f_{\text{reach}}(x, q)$
Joint Proximity $f_{\text{joint}}(q)$
× $\lambda_{c,t}$
$\nabla g(\hat{x}_0) = \nabla f_x(\hat{x}_0, q^*)$
$\hat{x}_0$
IK Solver
$q \in \mathcal{Q}(\hat{x}_0)$
minimum point of loss q^*
predicted wrist pose $\hat{x}_0$
x_{t-1}^{DDIM}
$\nabla g(\hat{x}_0)$
denoised wrist pose x_{t-1}
Output
Ours
Baseline
Ours
Baseline
...
feasible arm-hand pose (q^*, x_0)

Fig. 2: Overview of the proposed arm-aware dexterous grasp generation method. Initially, we pretrain an arm-agnostic diffusion model to capture the distribution of wrist poses for floating hands. During sampling, arm kinematics and environment SDF are integrated as constraints, with their gradients guiding the denoising process. This approach significantly enhances the proportion of feasible grasps, adaptable to various arm-hand configurations and constrained environments.

the DDIM sampling framework [26] to predict the clean wrist pose $\hat{\boldsymbol{x}}_0$ and update the sample $\boldsymbol{x}_{t-1}^{\text{DDIM}}$ deterministically

$$\hat{\boldsymbol{x}}_0 = \frac{1}{\sqrt{\bar{\alpha}_t}}\left(\boldsymbol{x}_t - \sqrt{1-\bar{\alpha}_t}\,\boldsymbol{\epsilon}_\theta(\boldsymbol{x}_t, \mathcal{O}, t)\right), \tag{3}$$

$$\boldsymbol{x}_{t-1}^{\text{DDIM}} = \sqrt{\bar{\alpha}_{t-1}}\,\hat{\boldsymbol{x}}_0 + \sqrt{1-\bar{\alpha}_{t-1}}\,\boldsymbol{\epsilon}_\theta(\boldsymbol{x}_t, \mathcal{O}, t). \tag{4}$$

The estimated $\hat{\boldsymbol{x}}_0$ is then used to enforce arm-related constraints, which will be detailed in Sec. III-B. From the perspective of Langevin dynamics, the DDIM update can be seen as moving the sample along the log-probability gradient [27]

$$\boldsymbol{x}_{t-1}^{\text{DDIM}} \approx \boldsymbol{x}_t + \frac{\beta_t}{2}\nabla_{\boldsymbol{x}} \log p_\theta(\boldsymbol{x}_t|\mathcal{O}), \tag{5}$$

i.e., toward regions of higher likelihood under the learned wrist pose distribution. After denoising, the corresponding joint angles $\boldsymbol{q}^{\text{hand}}$ are then predicted by an MLP.

B. Problem Formulation of the Arm-Aware Grasp Generation

The learned hand-centric diffusion model $p_\theta(\boldsymbol{x} \mid \mathcal{O})$ for wrist-pose generation does not ensure that a sampled pose $\boldsymbol{x} \sim p_\theta$ satisfies the constraints of the robotic arm. To address this limitation, we extend the diffusion sampling process to incorporate *arm awareness*. Under arm-related constraints, some samples $\boldsymbol{x}$ lie within the feasible set, while others do not. The key idea is that, through gradient-based guidance, *nearly-feasible samples are gradually pulled toward the feasible region*, resulting in a higher proportion of feasible samples and thus forming a *modified distribution*. Hence, our goal is to directly sample wrist poses from this *modified distribution* π, subject to two requirements: 1) The modified distribution π should remain close to the arm-agnostic distribution $p_\theta(\boldsymbol{x} \mid \mathcal{O})$ to maintain grasp quality; 2) Each wrist pose $\boldsymbol{x}$ must be reachable by the robotic arm while satisfying arm-related constraints.

We formulate this as a bi-level optimization problem that separately optimizes the diffusion sampling distribution π and the arm configuration $\boldsymbol{q}$. The outer optimization aims to update the diffusion sampling distribution π by minimizing the Kullback–Leibler (KL) divergence to the pretrained model p_θ, together with the expected arm-aware loss $g(\boldsymbol{x})$ computed at each wrist pose $\boldsymbol{x} \sim \pi$. Formally, the outer problem is defined as a variational optimization over π

$$\min_{\pi} \mathrm{KL}(\pi \| p_\theta) + \lambda_g \mathbb{E}_{x\sim\pi}[g(\boldsymbol{x})], \tag{6}$$

where the value function $g(\boldsymbol{x})$ represents the minimum loss achievable by the robotic arm under feasibility constraints and λ_g is a positive weight coefficient. For each sampled wrist pose $\boldsymbol{x}$, we first consider the following inner constrained optimization over the joint configuration $\boldsymbol{q}$

$$g(\boldsymbol{x}) = \min_{\boldsymbol{q}\in\mathcal{Q}(\boldsymbol{x})} f_s(\boldsymbol{x}, \boldsymbol{q}) \quad \text{s.t.} \quad f_c(\boldsymbol{x}, \boldsymbol{q}) \le 0, \tag{7}$$

where

$$\mathcal{Q}(\boldsymbol{x}) = \{\boldsymbol{q} \mid \mathrm{FK}(\boldsymbol{q}) = \boldsymbol{x}\} \tag{8}$$

denotes the feasible configuration set defined by forward kinematics $\mathrm{FK}(\cdot)$. The terms $f_s(\cdot)$ and $f_c(\cdot)$ represent the soft and hard arm-related cost, respectively. For this inner optimization problem, we introduce the Lagrangian function to handle the inequality constraint

$$\mathcal{L}(\boldsymbol{x}, \boldsymbol{q}; \lambda_c) = f_s(\boldsymbol{x}, \boldsymbol{q}) + \lambda_c f_c(\boldsymbol{x}, \boldsymbol{q}), \tag{9}$$

where $\lambda_c \ge 0$ is the Lagrange multiplier associated with the inequality constraint. The optimal arm configuration is then obtained by minimizing this Lagrangian within the feasible set

$$\begin{gathered} \boldsymbol{q}^* = \arg\min_{\boldsymbol{q}\in\mathcal{Q}(\boldsymbol{x})} f(\boldsymbol{x}, \boldsymbol{q}), \\ \text{where } f(\boldsymbol{x}, \boldsymbol{q}) = f_s(\boldsymbol{x}, \boldsymbol{q}) + \lambda_c f_c(\boldsymbol{x}, \boldsymbol{q}). \end{gathered} \tag{10}$$

Following the primal–dual method [28], the dual variable for the inequality constraint is updated via gradient ascent on the Lagrangian in (9)

$$\lambda_{c,t-1} = \lambda_{c,t} + \eta_\lambda f_c(\boldsymbol{x}, \boldsymbol{q}^*), \tag{11}$$

where $\eta_\lambda > 0$ is the learning rate. This update gradually increases the penalty on infeasible wrist poses as denoising proceeds, thereby enforcing the hard constraint more strictly. This inner optimization thus defines the value function $g(\boldsymbol{x}) = f(\boldsymbol{x}, \boldsymbol{q}^*)$, which then guides the outer update of the diffusion distribution over wrist poses. Following [29], solving the variational optimization in (6) yields a pointwise closed-form solution

$$\pi^*(\boldsymbol{x}) \,\propto\, p_\theta(\boldsymbol{x}|\mathcal{O})\, \exp[\,-\,\lambda_g g(\boldsymbol{x})], \tag{12}$$

which corresponds to an exponentially tilted version of the pre-trained distribution. By descending the negative log-probability of this modified distribution in (12) with the substitution of DDIM update (5), we obtain the following denoising update

$$\boldsymbol{x}_{t-1} \;=\; \boldsymbol{x}_{t-1}^{\text{DDIM}} - \frac{\beta_t \lambda_g}{2}\Big[\nabla g(\hat{\boldsymbol{x}}_0)\Big], \tag{13}$$

where $\boldsymbol{x}_{t-1}^{\text{DDIM}}$ is obtained from (4). This update resembles guided sampling, integrating both the original denoising process and guidance from arm-related feasibility constraints. Compared to [29], we compute gradients using the predicted wrist pose $\hat{\boldsymbol{x}}_0$ instead of the noised sample $\boldsymbol{x}_t$, ensuring smoother and more consistent gradient updates [1].

C. Analytical Formulation of Arm-Related Constraints

Building on the hierarchical formulation in Sec. III-B, we now describe how the arm-related constraints are incorporated into each denoising step (13). The feasible set $\mathcal{Q}(\boldsymbol{x})$ is obtained via IK mapping from wrist poses to arm configurations. To handle multiple IK branches efficiently, we use the geometry-based analytical solver [30]. After obtaining the optimal joint configuration $\boldsymbol{q}^*$ over the feasible set according to (10), the gradient of the value function can then be computed as

$$\nabla g(\boldsymbol{x}) = \nabla_{\boldsymbol{x}} f\left(\boldsymbol{x}, \boldsymbol{q}^*\right). \tag{14}$$

We further evaluate this gradient by applying the chain rule

$$\nabla g(\boldsymbol{x}) = \frac{\partial f}{\partial \boldsymbol{x}}(\boldsymbol{x}, \boldsymbol{q}^*) + \left(\frac{d\boldsymbol{q}^*}{d\boldsymbol{x}}\right)^{\top} \frac{\partial f}{\partial \boldsymbol{q}}(\boldsymbol{x}, \boldsymbol{q}^*). \tag{15}$$

To compute the second term, observe that the IK solution $\boldsymbol{q}^*$ can be viewed as the local minimizer of the following quadratic program, with $\boldsymbol{q}^{\text{ref}}$ denoting a reference configuration

$$\boldsymbol{q}^* = \arg\min_{\boldsymbol{q}} \frac{1}{2}\|\text{FK}(\boldsymbol{q}) - \boldsymbol{x}\|_2^2 + \frac{\lambda_q^2}{2}\|\boldsymbol{q} - \boldsymbol{q}_{\text{ref}}\|_2^2. \tag{16}$$

Here, λ_q serves as a damping coefficient that stabilizes the solution. Through sensitivity analysis of (16), the gradient of $\boldsymbol{q}^*$ can be derived as the damped pseudo-inverse, which is independent of the reference configuration $\boldsymbol{q}^{\text{ref}}$

$$\frac{d\boldsymbol{q}^*}{d\boldsymbol{x}} = \boldsymbol{J}_{\lambda_q}^{\dagger} = (\boldsymbol{J}^\top \boldsymbol{J} + \lambda_q^2 \boldsymbol{I})^{-1} \boldsymbol{J}^\top, \tag{17}$$

where $\boldsymbol{J}$ denotes the geometric Jacobian of the arm. Substituting this result back into (15), the closed-form gradient with respect to the wrist pose is obtained as

$$\nabla g(\boldsymbol{x}) = \frac{\partial f}{\partial \boldsymbol{x}}(\boldsymbol{x}, \boldsymbol{q}^*) + (\boldsymbol{J}_{\lambda_q}^{\dagger})^{\top} \frac{\partial f}{\partial \boldsymbol{q}}(\boldsymbol{x}, \boldsymbol{q}^*). \tag{18}$$

This indicates that the arm-related gradient comprises a direct term with respect to the wrist pose $\boldsymbol{x}$ and an indirect term propagated through the joint configuration $\boldsymbol{q}$ via the damped Jacobian pseudo-inverse.

Based on this gradient formulation, we now present the specific forms of the three arm-related constraints.

Hand Reachability: This constraint ensures that the wrist pose remains within the reachable workspace of the arm by penalizing the forward kinematics error

$$f_{\text{reach}}(\boldsymbol{x}, \boldsymbol{q}) = \frac{1}{2}\|\text{FK}(\boldsymbol{q}) - \boldsymbol{x}\|^2. \tag{19}$$

The partial derivatives of the reachability loss are written as

$$\frac{\partial f_{\text{reach}}}{\partial \boldsymbol{x}} = \boldsymbol{x} - \text{FK}(\boldsymbol{q}),\; \frac{\partial f_{\text{reach}}}{\partial \boldsymbol{q}} = \boldsymbol{J}^\top(\text{FK}(\boldsymbol{q}) - \boldsymbol{x}). \tag{20}$$

Arm–Environment Collision Avoidance: This constraint prevents potential collisions between the robotic arm and the environment. It is determined by the minimum value $d(\boldsymbol{q})$ of the points on the arm's sphere-based collision model within the environment's signed distance field (SDF). We apply a sigmoid shaping function that smoothly increases the cost as the robot approaches obstacles

$$f_{\text{colli}}(\boldsymbol{q}) = \sigma\Big(-d(\boldsymbol{q})\Big) - \frac{1}{2}, \tag{21}$$

where $\sigma(z) = \frac{1}{1+e^{-\kappa z}}$ denotes a sigmoid function with slope parameter κ. The gradient of the collision cost is expressed as

$$\frac{\partial f_{\text{colli}}}{\partial \boldsymbol{q}} = -\sigma'(-d)\boldsymbol{J}_p^\top \boldsymbol{n}, \tag{22}$$

where $\boldsymbol{J}_p \in \mathbb{R}^{3\times n}$ denotes the translational Jacobian of the point on the robotic arm with the minimum SDF, and $\boldsymbol{n}$ is the unit vector which aligns with the SDF gradient at that point.

Joint Proximity: To reduce unnecessary joint-space movement, we introduce a penalty on the deviation of the solved arm configuration $\boldsymbol{q}$ from the current configuration $\boldsymbol{q}_{\text{cur}}$

$$f_{\text{joint}}(\boldsymbol{q}) = \frac{1}{2}\|\boldsymbol{q} - \boldsymbol{q}_{\text{cur}}\|^2, \tag{23}$$

whose gradient is given by

$$\frac{\partial f_{\text{joint}}}{\partial \boldsymbol{q}} = \boldsymbol{q} - \boldsymbol{q}_{\text{cur}}. \tag{24}$$

The **Hand Reachability** and **Collision Avoidance** terms are treated as hard feasibility constraints, whereas the **Joint Proximity** term is considered a soft constraint in (6). Therefore, we have

$$f(\boldsymbol{x}, \boldsymbol{q}) = \lambda_c^{\text{reach}} f_{\text{reach}}(\boldsymbol{x}, \boldsymbol{q}) + \lambda_c^{\text{colli}} f_{\text{colli}}(\boldsymbol{q}) + f_{\text{joint}}(\boldsymbol{q}). \tag{25}$$

According to Sec. III-B, the coefficient $\lambda_c^{\text{reach}}, \lambda_c^{\text{colli}}$ of the hard constraints are dynamically updated following (11). One can adjust the relative weights or disable certain constraint terms in (25) by assigning different schedules for λ. Among these, the **Collision Avoidance** term is the most commonly used, as real-world grasping scenarios often occur in constrained environments. The **Hand Reachability** term addresses extreme situations involving grasping near workspace boundaries, while the **Joint Proximity** term plays a significant role during consecutive grasping. Finally, the gradient $\nabla g(\boldsymbol{x})$ required by (13) can be computed with (18), (20), (22) and (24). The overall grasp generation procedure is summarized in Alg. 1.

[1] Compared to $\nabla g(\boldsymbol{x})$, $\nabla g(\hat{\boldsymbol{x}}_0)$ changes more gently, especially when considering the collision avoidance constraints.

```
Algorithm 1: Arm-Aware Dexterous Grasp Generation
Input: noise predictor $\epsilon_\theta$ of the pre-trained diffusion
       grasp generation model, environment SDF,
       object partial observation $\mathcal{O}$, initial value and
       learning rate of $\lambda_c$, current configuration $\boldsymbol{q}_{\text{cur}}$
for $t$ from $T$ to $1$ do
    Compute $\boldsymbol{x}_{t-1}^{\text{DDIM}}$ using the denoising step (4)
    Compute the IK solution set $\mathcal{Q}$ (8) and the cost
      function $f(\boldsymbol{x}, \boldsymbol{q})$ according to (25)
    Compute $\boldsymbol{q}^*$ according to (10)
    Update $\lambda_c$ with (11)
    Compute $\nabla g(\boldsymbol{x})$ with (18), (20), (22), and (24)
    Update the noised sample $\boldsymbol{x}_{t-1}$ with (13)
end
return $\boldsymbol{x}_0, \boldsymbol{q}^*$
```

IV. Experiments

We design the experiments to answer the following questions: 1) Can the proposed method generate feasible grasps that satisfy arm-related constraints with a higher probability? 2) How does incorporating these constraints affect physical grasp quality? 3) Is the proposed method easily adaptable to different robotic arms and environments without additional data synthesis or training? Additional evaluation results can be found in the Appendix (available on our Project Website).

Experiment Setup. We use the Shadow Hand in simulations and the LEAP Hand for real-world evaluations, considering two common robotic arms, UR5 and Franka. We assume the environment can be represented as a combination of closed geometries, enabling a well-defined SDF. We adopt the sphere robot collision models from cuRobo [31] for collision avoidance. We train the hand-centric diffusion model p_θ on roughly 50k tabletop grasps synthesized with BODex [1] on the processed DGN dataset [12]. The simulation evaluation is performed using DGN's test set, containing 10,892 objects with varying geometries, poses and scales. For each object, 10 random grasp poses and arm configurations are sampled, yielding a total of approximately 100k grasps.

Evaluation Metrics. We are concerned about four types of metrics: 1) *Constraint Satisfaction*. For the hard constraints, we filter out grasps that violate these constraints and report the percentage of arm-feasible grasps among all sampled grasps that satisfy the Collision Avoidance constraint (Collision Feasible Rate, CFR), the Hand Reachability constraint (Reachability Feasible Rate, RFR), and both (Feasible Rate, FR). For the soft constraint, we report the average Joint Proximity loss of all sampled grasps (Average Joint Proximity, AJP). 2) *Grasp Success Rate* (GSR). Among all feasible grasps that satisfy the hard constraints after filtering, we report the percentage of grasps that are successfully executed in MuJoCo. This metric indicates the quality of the arm-feasible grasps. 3) *Success Rate* (SR). This metric, akin to GSR, measures the percentage of feasible and successful grasps from all sampled grasps and is equal to the product of FR and GSR. 4) *Object Success Rate* (OSR). We report the percentage of the 10,892 objects that, among all sampled grasps, have at least one that is both feasible and successfully executed in MuJoCo.

Table I: Performance of grasp generation in constrained environments (UR5)

Scene	Method	C. Satisfaction(%)↑			GSR	SR	OSR
		CFR	RFR	FR	(%)↑	(%)↑	(%)↑
S1	Baseline	14.75	99.52	14.75	**50.10**	7.39	49.92
	Ours	**84.91**	**100.00**	**84.91**	44.69	**37.95**	**89.64**
S2	Baseline	33.69	95.60	33.69	**52.75**	17.77	77.39
	Ours	**98.56**	**100.00**	**98.56**	40.34	**39.76**	**91.25**
S3	Baseline	28.64	72.82	28.64	**52.44**	15.02	73.04
	Ours	**83.18**	**98.82**	**83.18**	41.72	**34.70**	**91.22**

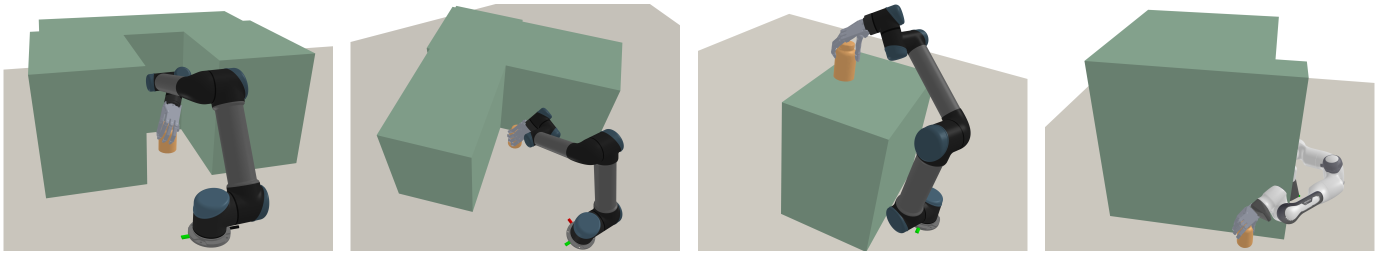
Table II: Performance of grasp generation in constrained environments (Franka)

Scene	Method	C. Satisfaction(%)↑			GSR	SR	OSR
		CFR	RFR	FR	(%)↑	(%)↑	(%)↑
S1	Baseline	11.60	78.34	11.40	**49.33**	5.62	40.98
	Ours	**62.55**	**87.79**	**60.15**	38.97	**23.44**	**81.49**
S2	Baseline	33.03	78.33	31.91	**52.90**	16.88	76.34
	Ours	**96.58**	**94.86**	**93.92**	40.96	**38.46**	**91.32**
S3	Baseline	29.42	58.50	29.41	**53.60**	15.76	74.05
	Ours	**80.50**	**87.88**	**80.45**	48.16	**38.75**	**91.34**
S4	Baseline	23.56	85.31	23.54	**52.59**	12.38	65.78
	Ours	**68.55**	**96.50**	**68.48**	41.91	**28.70**	**85.96**

Baselines. In the experiments, the *baseline* refers to rejection sampling from the pretrained grasp distribution p_θ. We select the arm configuration as the IK solution with the least constraint loss, defined as (10). Additionally, the 'w/o' variants represent ablations that omit the guidance of the associated constraint from our method.

A. Simulation Studies

It is worth noting that Collision Avoidance is essential for constrained grasping, while Hand Reachability and Joint Proximity play critical roles for their respective purposes. We first evaluate Collision Avoidance in Sec. IV-A1, followed by an assessment of the other two constraints and their interaction with Collision Avoidance in Secs. IV-A2 and IV-A3.



(a) Narrow Corridor (S1) (b) Room Corner (S2) (c) Shelf Top (S3) (d) Partition Board (S4)

Fig. 3: Visualization of four evaluation scenes (S1-S4) with one successful grasp involving the robotic arm and dexterous hand, the obstacles painted in green, and the grasped object highlighted in orange. Scene S4 is specifically designed to showcase Franka's collision avoidance using its kinematic redundancy.

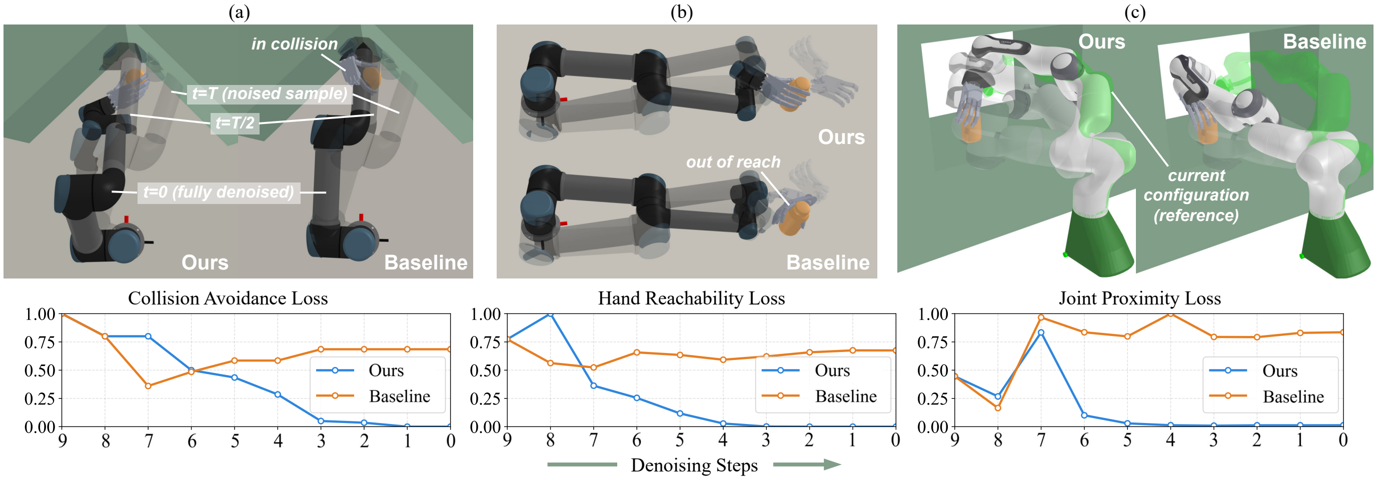


Fig. 4: Illustration of representative cases highlighting the effects of various constraints. The top row illustrates the arm configurations $\boldsymbol{q}^*$ during the denoising process of a certain noised sample both with our method and the baseline, where $\boldsymbol{q}^*$ at $t = T, T/2, 0$ are visualized from transparent to opaque. The denoising process starts from the same initial arm configuration shown with the highest transparency. The fully denoised hand and arm configuration are depicted with the lowest transparency. For clarity of illustration, we only display the evolution of hand configuration in (b). The green arm in (c) represents the reference joint configuration $\boldsymbol{q}_{\text{cur}}$ based on a hypothetical previous grasp attempt. The bottom row presents the evolution of normalized constraint losses over time.

1) Performance of the Collision Avoidance Constraint: To evaluate the performance of guided grasp generation in constrained environments, we design four challenging scenes, as shown in Fig. 3. The feasible regions in these scenarios are relatively small, which significantly impacts the efficiency of rejection sampling. The results are shown in Tab. I and Tab. II. On the one hand, the proposed method significantly enhances constraint satisfaction, particularly with regard to the CFR and FR. On the other hand, the incorporation of guidance slightly degrades quality of the feasible grasps, as indicated by GSR, by altering the direction in the original denoising process. It is worth noting that, while grasp quality may be affected, the proposed method significantly improves SR and the probability of achieving at least one successful grasp for a given object—measured by the Object Success Rate (OSR)—by increasing the proportion of feasible grasps. In the Appendix, we explore a preliminary solution for maintaining grasp quality through null-space projection, which merits further investigation. Additionally, the proposed method demonstrates consistent performance on both UR5 (Tab. I) and Franka (Tab. II), demonstrating its adaptability to robotic arms with different kinematics. For Franka, both the proposed method and the baseline show more noticeable constraint violations on S1 (Tab. II), likely due to Franka's larger collision volume and limited workspace, which reduce the feasible region.

2) Performance of the Hand Reachability Constraint: To evaluate our method's ability to generate grasps with higher reachability, we randomly place objects near the edge of the robotic arm's workspace, as illustrated in the tabletop scene in Fig. 4 (b). Under these placements, the arm can only perform grasps aligned with the object's front-facing direction and cannot reach grasp poses from behind, while still needing to avoid collisions with the tabletop. As shown in Tab. III, applying only the Collision Avoidance constraint (w/o Reachability) or only the Hand Reachability constraint (w/o Collision) results in optimal performance for their respective metrics (CFR and RFR). However, this comes at the expense of the other metric, indicating that a solution addressing both constraints more effectively remains to be found. In contrast, our method incorporates both constraints during inference, guiding sampling toward feasible regions that satisfy both constraints. This leads to a substantial increase in FR and further improves OSR.

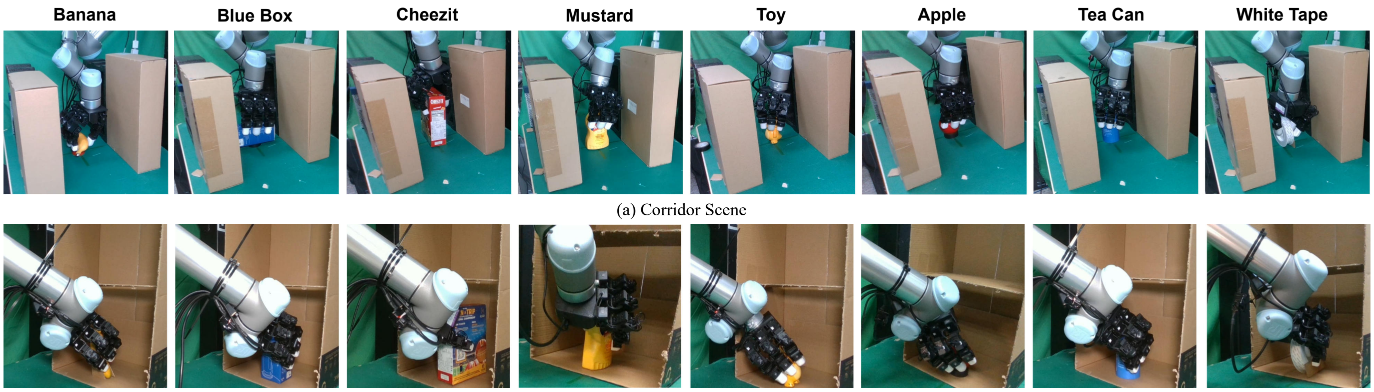
3) Performance of the Joint Proximity Constraint: To evaluate our method's ability to generate grasps with arm configuration proximity, we randomly place objects on a shelf and define a reference joint configuration based on a hypothetical previous grasp attempt, as visualized in Fig. 4 (c). The results in Tab. IV indicate that our method achieves the lowest joint proximity loss (AJP) and delivers the best performance in both CFR and FR. We observe that applying only the Joint Proximity constraint (w/o Collision) also improves collision avoidance performance (CFR), as the reference configuration is collision-free and implicitly guides sampling toward safer

Table III: Performance of reachable grasp generation near workspace boundaries

Arm	Method	C. Satisfaction(%)↑			GSR (%)↑	SR (%)↑	OSR (%)↑
		CFR	RFR	FR			
UR5	Baseline	83.96	27.20	19.47	47.33	9.22	49.94
	Ours	83.87	69.64	**56.20**	43.68	**24.54**	**79.44**
	w/o Reachability	**91.30**	26.36	22.49	**48.16**	10.83	53.73
	w/o Collision	71.13	**72.33**	46.42	41.85	19.43	74.54
Franka	Baseline	78.42	23.60	13.42	**48.03**	6.45	35.81
	Ours	80.66	58.39	**47.62**	44.47	**21.18**	**71.47**
	w/o Reachability	**91.48**	18.81	15.95	47.11	7.51	40.93
	w/o Collision	61.57	**70.45**	40.84	44.52	18.18	60.43

Table IV: Performance of grasp generation with reference joint configuration

Arm	Method	C. Satisfaction			GSR (%)↑	SR (%)↑	OSR (%)↑
		CFR (%)↑	FR (%)↑	AJP (rad^2)↓			
UR5	Baseline	43.95	41.80	0.093	**45.90**	19.19	72.51
	Ours	**74.14**	**73.71**	**0.045**	36.04	26.57	80.43
	w/o Collision	53.67	53.07	0.054	42.40	22.50	77.33
	w/o Joint	65.61	63.81	0.075	42.66	**27.22**	**81.43**
Franka	Baseline	45.04	43.98	0.099	46.64	20.52	58.99
	Ours	**52.67**	**52.01**	**0.081**	44.76	**23.27**	**66.37**
	w/o Collision	49.85	48.84	0.092	**46.67**	22.79	63.78
	w/o Joint	47.26	46.59	0.087	43.40	20.22	60.79



(a) Corridor Scene

(b) Shelf Scene

Fig. 5: Snapshots of the real-world grasp execution with UR5 and LEAP Hand, under two challenging scenarios. (a) Corridor Scene, (b) Shelf Scene.

Table V: Per-object grasp execution success rate with the grasps generated for eight daily objects in two real-world constrained scenarios (Fig. 5)

Corridor Scene	Object	Banana	Blue Box	Cheezit	Mustard	Toy	Apple	Tea Can	White Tape
	Top-10 Success Num	8/10	6/10	8/10	8/10	9/10	10/10	7/10	10/10
Shelf Scene	Object	Banana	Blue Box	Cheezit	Mustard	Toy	Apple	Tea Can	White Tape
	Top-10 Success Num	10/10	9/10	8/10	9/10	9/10	10/10	9/10	10/10

regions. We further demonstrate potential applications of this constraint through real-world experiments.

4) Illustration of Representative Cases: For better understanding, we visualize representative cases that illustrate the effect of incorporating various constraints, as depicted in Fig. 4. The denoising process for both our method and the baseline begins with the same Gaussian noise sample. In Fig. 4 (a), the collision is progressively resolved with guidance, whereas the baseline leads to significant penetration. In Fig. 4 (b), the wrist pose is incrementally projected into the arm's reachable space, whereas the baseline wrist pose remains out of reach. In Fig. 4 (c), the fully denoised arm with guidance is closer to the green reference configuration, indicating a shorter joint-space movement from the last grasp attempt. Among all these cases, the constraint losses converge to zero using our method, while the baseline results in large violations.

B. Real-World Experiments

In the real-world experiments, we use the LEAP Hand mounted on a UR5 robotic arm. An Azure Kinect depth camera captures the object's partial point cloud. We employ cuRobo to plan collision-free hand-arm trajectories from the initial configuration to the pre-grasp configuration, where the fingers are slightly spread apart without contacting the object. In addition, we train a grasp quality evaluator that predicts the probability of successfully executing a given grasp based on the object's partial point cloud and the grasp pose, enabling the selection of higher-quality grasps from a batch of candidates for real-world execution. This module addresses the issue of imperfect grasp generation (i.e., low GSR) in practice.

1) Evaluation of Grasp Quality in the Real-World: We first evaluated the quality of the generated grasps in the real world, which is necessary as simulations cannot accurately capture the contact-rich interactions present in the physical world. The evaluation involved grasping eight everyday objects with varying geometries and masses, positioned in two challenging scenarios: (a) Corridor Scene and (b) Shelf Scene. For each object, we sampled 40 candidate grasps in a batch and executed the top 10 arm-feasible grasps with the highest predicted success probabilities. Fig. 5 shows snapshots of executed grasp poses, while Tab. V highlights the success rates, demonstrating reasonable performance in challenging real-world scenarios. Further discussion of failure cases can be found in the Appendix.

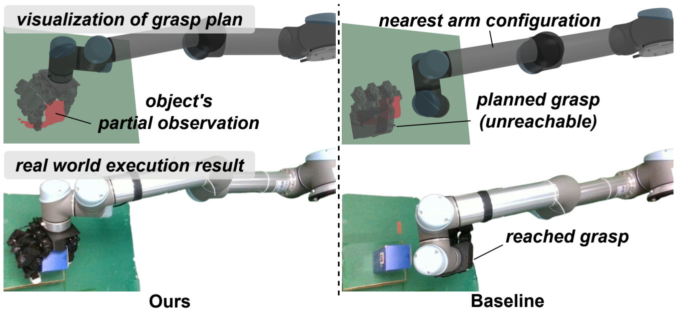

visualization of grasp plan
object's
partial observation
nearest arm configuration
planned grasp
(unreachable)
real world execution result
reached grasp
Ours
Baseline

Fig. 6: Comparison of grasp generation near the arm's workspace boundary. The top right figure shows a planned grasp that the robotic arm cannot reach.

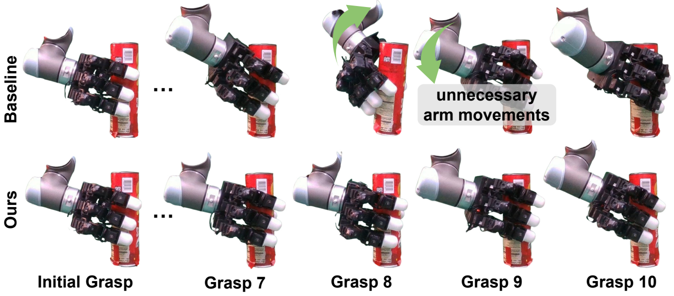


Fig. 7: Illustration of guided grasp generation for continuously regrasping with the Joint Proximity constraint. The reference initial grasp and the four final grasps are displayed from left to right, with additional grasps omitted for brevity. Please refer to the attached video for the complete grasp sequence.

2) Generating Reachable Grasps Near Workspace Boundaries: Additionally, we showcased the effectiveness of the proposed method in generating reachable grasps near the arm's workspace boundary. We executed a generated grasp both with and without guidance, as shown in Fig. 6. Without guidance, the system tends to produce unreachable grasps behind the object, resulting in failures even when the arm is fully extended.

3) Generating Proximal Grasp Configurations for Regrasping: Finally, we demonstrated the proposed method's ability to generate grasps with adjacent arm configurations in joint space during regrasping. Using the initial grasp's arm configuration as a reference, we consecutively sampled grasps 10 times, selecting the one with the lowest joint proximity loss from 40 candidates per sample for execution, as shown in Fig. 7. Without guidance refinement, the baseline method leads to unnecessary arm movements. In contrast, our approach reduces the average joint space distance between grasps from 0.88 rad to 0.56 rad, resulting in a smoother execution trajectory.

V. Conclusions

Conclusion: In this work, we present an arm-aware dexterous grasp generation framework that integrates arm kinematics and environmental constraints into diffusion-based sampling while retaining the generalization ability of pretrained arm-agnostic grasp models. By jointly optimizing wrist poses and arm configurations and injecting closed-form arm-related gradients during denoising, the method guides sampling toward feasible regions. Experiments in simulation and real-world settings featuring constrained scenarios show improved constraint satisfaction and arm configuration proximity, without requiring arm-specific data synthesis and training, while maintaining reasonable success rates across different arms and environments.

Discussion: We note that the incorporation of guidance may slightly perturb the learned distribution and affect grasp quality. To address this, null-space projection and evaluator-based gradient guidance can be introduced during denoising to maintain alignment with high-quality grasp regions. Moreover, as the refinement induced by the guidance is inherently local, the method is most effective at converting near-feasible samples into feasible ones, motivating the use of multiple initial samples for broader feasible coverage. Future work will explore integrating our guidance with improved initial sampling strategies to further enhance global performance.

References

[1] J. Chen, Y. Ke, and H. Wang, "Bodex: Scalable and efficient robotic dexterous grasp synthesis using bilevel optimization," in *IEEE Int. Conf. Robot. Autom.* IEEE, 2025, pp. 01–08.

[2] W. Wei, P. Wang, S. Wang, Y. Luo, W. Li, D. Li, Y. Huang, and H. Duan, "Learning human-like functional grasping for multifinger hands from few demonstrations," *IEEE Trans. Robot.*, vol. 40, pp. 3897–3916, 2024.

[3] Y. Jiang, M. Yu, X. Zhu, M. Tomizuka, and X. Li, "Contact-implicit model predictive control for dexterous in-hand manipulation: A long-horizon and robust approach," in *IEEE/RSJ Int. Conf. Intell. Robots Syst.* IEEE, 2024, pp. 5260–5266.

[4] Z. Wei, Z. Xu, J. Guo, Y. Hou, C. Gao, Z. Cai, J. Luo, and L. Shao, "D (r, o) grasp: A unified representation of robot and object interaction for cross-embodiment dexterous grasping," *arXiv preprint arXiv:2410.01702*, 2024.

[5] J. Zhang, H. Liu, D. Li, X. Yu, H. Geng, Y. Ding, J. Chen, and H. Wang, "Dexgraspnet 2.0: Learning generative dexterous grasping in large-scale synthetic cluttered scenes," in *Conf. Robot. Learn.*, 2024.

[6] Y. Zhong, Q. Jiang, J. Yu, and Y. Ma, "Dexgrasp anything: Towards universal robotic dexterous grasping with physics awareness," in *IEEE Conf. Comput. Vis. Pattern Recognit.*, 2025, pp. 22 584–22 594.

[7] Y. Li, W. Wei, D. Li, P. Wang, W. Li, and J. Zhong, "Hgc-net: Deep anthropomorphic hand grasping in clutter," in *IEEE Int. Conf. Robot. Autom.* IEEE, 2022, pp. 714–720.

[8] Q. Chen, K. Van Wyk, Y.-W. Chao, W. Yang, A. Mousavian, A. Gupta, and D. Fox, "Learning robust real-world dexterous grasping policies via implicit shape augmentation," in *Conference on Robot Learning*. PMLR, 2023, pp. 1222–1232.

[9] H. Li, X. Lin, Y. Zhou, X. Li, Y. Huo, J. Chen, and Q. Ye, "Contact2grasp: 3d grasp synthesis via hand-object contact constraint," in *Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence*, 2023, pp. 1053–1061.

[10] P. Li, T. Liu, Y. Li, Y. Geng, Y. Zhu, Y. Yang, and S. Huang, "Gendexgrasp: Generalizable dexterous grasping," in *IEEE Int. Conf. Robot. Autom.* IEEE, 2023, pp. 8068–8074.

[11] Y. Xu, W. Wan, J. Zhang, H. Liu, Z. Shan, H. Shen, R. Wang, H. Geng, Y. Weng, J. Chen *et al.*, "Unidexgrasp: Universal robotic dexterous grasping via learning diverse proposal generation and goal-conditioned policy," in *IEEE Conf. Comput. Vis. Pattern Recognit.*, 2023, pp. 4737–4746.

[12] J. Chen, Y. Ke, L. Peng, and H. Wang, "Dexonomy: Synthesizing all dexterous grasp types in a grasp taxonomy," *Robotics: Science and Systems*, 2025.

[13] J. Lu, H. Kang, H. Li, B. Liu, Y. Yang, Q. Huang, and G. Hua, "Ugg: Unified generative grasping," in *Eur. Conf. Comput. Vis.* Springer, 2024, pp. 414–433.

[14] Y. Ma, Z. Chen, X. Zhang, Z. Xu, Y. Zhang, H. Wang, and X. Wang, "Contact map transfer with conditional diffusion model for generalizable dexterous grasp generation," *arXiv preprint arXiv:2511.01276*, 2025.

[15] P. Dhariwal and A. Nichol, "Diffusion models beat gans on image synthesis," *Adv. Neural Inf. Process. Syst.*, vol. 34, pp. 8780–8794, 2021.

[16] X. Liu, D. H. Park, S. Azadi, G. Zhang, A. Chopikyan, Y. Hu, H. Shi, A. Rohrbach, and T. Darrell, "More control for free! image synthesis with semantic diffusion guidance," in *Proceedings of the IEEE/CVF winter conference on applications of computer vision*, 2023, pp. 289–299.

[17] Y. Luo, C. Sun, J. B. Tenenbaum, and Y. Du, "Potential based diffusion motion planning," in *Int. Conf. Mach. Learn.* PMLR, 2024, pp. 33 486–33 510.

[18] X. Zhang, X. Mou, R. Wang, T. Wo, N. Gu, T. Wang, C. Xu, and X. Liu, "Robotdiffuse: Motion planning for redundant manipulator based on diffusion model," *arXiv preprint arXiv:2412.19500*, 2024.

[19] K. Saha, V. Mandadi, J. Reddy, A. Srikanth, A. Agarwal, B. Sen, A. Singh, and M. Krishna, "Edmp: Ensemble-of-costs-guided diffusion for motion planning," in *IEEE Int. Conf. Robot. Autom.* IEEE, 2024, pp. 10 351–10 358.

[20] J. Carvalho, A. T. Le, M. Baierl, D. Koert, and J. Peters, "Motion planning diffusion: Learning and planning of robot motions with diffusion models," in *IEEE/RSJ Int. Conf. Intell. Robots Syst.* IEEE, 2023, pp. 1916–1923.

[21] Z. Weng, H. Lu, D. Kragic, and J. Lundell, "Dexdiffuser: Generating dexterous grasps with diffusion models," *IEEE Robot. Autom. Lett.*, 2024.

[22] Y. Zhong, Q. Jiang, J. Yu, and Y. Ma, "Dexgrasp anything: Towards universal robotic dexterous grasping with physics awareness," in *IEEE Conf. Comput. Vis. Pattern Recognit.*, 2025, pp. 22 584–22 594.

[23] G. Singh, S. Kalwar, M. F. Karim, B. Sen, N. Govindan, S. Sridhar, and K. M. Krishna, "Constrained 6-dof grasp generation on complex shapes for improved dual-arm manipulation," in *IEEE/RSJ Int. Conf. Intell. Robots Syst.* IEEE, 2024, pp. 7344–7350.

[24] H. Li, Q. Feng, Z. Zheng, J. Feng, Z. Chen, and A. Knoll, "Language-guided object-centric diffusion policy for generalizable and collision-aware manipulation," in *IEEE Int. Conf. Robot. Autom.* IEEE, 2025, pp. 12 834–12 841.

[25] J. Ho, A. Jain, and P. Abbeel, "Denoising diffusion probabilistic models," *Adv. Neural Inf. Process. Syst.*, vol. 33, pp. 6840–6851, 2020.

[26] J. Song, C. Meng, and S. Ermon, "Denoising diffusion implicit models," in *International Conference on Learning Representations*.

[27] C. Luo, "Understanding diffusion models: A unified perspective," *arXiv preprint arXiv:2208.11970*, 2022.

[28] S. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge university press, 2004.

[29] J. Zhang, L. Zhao, A. Papachristodoulou, and J. Umenberger, "Constrained diffusers for safe planning and control," *arXiv preprint arXiv:2506.12544*, 2025.

[30] A. J. Elias and J. T. Wen, "Ik-geo: Unified robot inverse kinematics using subproblem decomposition," *Mechanism and Machine Theory*, vol. 209, p. 105971, 2025.

[31] B. Sundaralingam, S. K. S. Hari, A. Fishman, C. Garrett, K. Van Wyk, V. Blukis, A. Millane, H. Oleynikova, A. Handa, F. Ramos *et al.*, "Curobo: Parallelized collision-free robot motion generation," in *IEEE Int. Conf. Robot. Autom.* IEEE, 2023, pp. 8112–8119.