

Virtual Attention Points: Bridging Human Movement Characteristics and Dexterous Robot Motion Generation

Abed Soleymani, Yafei Ou, Xingyu Li, and Mahdi Tavakoli

Abstract—In this study, we introduce Virtual Attention Points (VAPs) as a novel technique for characterizing the essence of dexterous human movements through mathematical encoding. This method focuses on pivotal points to capture movement dynamics, resulting in the generation of versatile and human-like motions for robotic systems. The proposed method inspired by the idea of human movement primitives (MPs) generates an interpretable low-dimensional representation for a given complex movement based on a new encoding basis function. Our approach achieves a remarkable 97% improvement in encoding accuracy for dexterous demonstrations with agile maneuvers and sharp turns, surpassing existing MP-based methods, enhancing the precision of fine manipulation and elevating the fidelity of encoded actions to the human movement. The precise replication of crucial poses and underlying behaviors highlights the efficacy of our approach in faithfully capturing the intricacies of expert human demonstrations. Our approach also generates a meaningful and interpretable representation of each demonstration, which encapsulates the skills-related features for performance assessment purposes. We propose a novel trajectory cloning algorithm that minimally warps various movement demonstrations such that starting and end points of motions will be manipulated to desired locations. Our work holds transformative potential, especially in enhancing surgical training and autonomous surgical systems, where precision and human-like dexterity are paramount. As surgical operations necessitate dexterous trajectories to execute specific functional tasks like suturing, we implement the proposed method to assess its performance in surgical skills and autonomous surgery tasks.

Index Terms—Dexterous Robotics, Trajectory Encoding, Generation, and Optimization, Movement Primitives, Surgical Skills Assessment.

I. INTRODUCTION

New application areas for robots demand versatility in handling complex tasks, whether autonomously or through collaboration with humans in real-world environments. For example, in surgical robotics, robots are expected to exhibit human-like behavior while functioning autonomously to enhance human-robot communication, integration, and co-adaptation. Human-like behavior of a collaborative robot can directly enhance the quality of information exchange between the human and the robot (communication), enhance how the human's decision-making process connects with the robot decision-making process leverage this hybrid natural-artificial intelligence fusion to improve the task performance (integration), and facilitate a mutual adaptation process in which the robot adapts to the human's actions and vice versa (co-adaptation). However, despite their fast actuators and real-time signal processing, robots are still far behind humans who suffer from slow muscles and neural delays in terms of dexterous object manipulation tasks

This research was supported by the Canada Foundation for Innovation (CFI), the Natural Sciences and Engineering Research Council (NSERC) of Canada, the Canadian Institutes of Health Research (CIHR), Alberta Innovates, and the Alberta Jobs, Economy, and Trade Ministry's Major Initiatives Fund A-Medico.

All authors are with the Electrical and Computer Engineering Department, University of Alberta, Edmonton, Alberta, Canada. {zsoleymani, yafei.ou, xingyu, mahdi.tavakoli}@ualberta.ca

[1]. In this paper, *dexterity* is referred to as the ability to follow spatial trajectories with fast maneuvers and occasionally sharp turns, such as tying a knot. Purely data-directed imitation learning (i.e., reliance on human demonstrations) in complex scenarios such as (semi-)autonomous robotic surgeries does not necessarily generate a natural and task-appropriate behavior in the real world [2].

In addition to dexterity, the optimality in terms of the minimum effort or minimum wear and tear is another aspect of human-performed tasks [3]. To achieve optimality in robot-performed tasks, the perspective of incorporating optimal control concepts may seem somewhat narrow, as such approaches primarily focus on optimizing the performance of a single task.

By mimicking human-like motions, we aim to generate optimal and dexterous trajectories for robots across a wide array of tasks. These tasks may include encoding human hands' demonstrations, enabling robots to autonomously replicate closely-followed trajectories (the main focus of this paper), or facilitating human-robot interactions that emulate the natural dynamics observed in human-human interactions (the focus of future work). To this end, one elegant approach involves distilling the essence of expert demonstrations into a compact and understandable representation, allowing the robot to gain a tangible grasp of the nuanced dexterity and optimization inherent in skilled trajectories. This paper introduces a novel approach that translates complex human demonstrations into simplified representations in code space. This method not only reduces the complexity and computational cost for downstream tasks, whether involving robot autonomy, task optimization, or human-robot collaboration, but also preserves the dexterity exhibited in the original demonstration and maintains the optimal behaviors and skills of the human demonstrator.

One reason biological systems, such as humans and animals, execute complex movements in a versatile and optimal manner, is that motor trajectories in their nervous system are encoded in terms of a combination of parameterized movement primitives (MPs). This encoding serves as a crucial simplification for learning, performance, and retention of complex skills [1]. MPs have simple time profiles and can be viewed as attractors that contribute to the variations within the trajectory. Recent research on patients recovering after a paralyzing stroke and infant reaching behaviors showed that their earliest movements were composed of simple submovements with streamlined profiles [4]. The time between adjacent submovements decreased and became more overlapped as the stroke patients recovered or the infants became older [4]. Among all possible operators for combining MPs, linear vector superposition of submovements has been shown to provide an accurate description of human-performed trajectories [4].

In mathematical terms, MPs are a set of overlapping basis functions through time that, by a proper linear combination, can produce

a wide range of motion profiles. Such mathematical representation of human movements dramatically reduces the dimensionality of the trajectory by encoding complicated demonstrations onto the parameter space. This allows for efficient modeling of desired movements, compact parameterization, fast trajectory generation, and motor skills generalization. Such minimal expression of human trajectories is ideal in terms of memory efficiency and computational considerations for downstream robot control tasks (e.g., machine/deep learning classifier/regressor or utilizing policy search algorithm for training an agent) compared to storing all timestamps of a given demonstration.

The concept of MPs for representing human motor skills raises an important question: what is the best basis function? The basis should ideally have several properties, including but not limited to the preservation of trajectory’s shape, explainability in human terms, generalizability to new but similar situations, and low computational complexity. The first trivial option is the famous monomial (power) basis (i.e., $[1, t, \dots, t^n]$), which leads to spline fitting. However, spline fitting might not be an intuitive choice, as spline parameters are obtained by solving a set of intangible mathematical equations which usually returns interrelated parameters [5]. Another choice is Gaussian basis which leads to the popular approach of dynamic movement primitives (DMP) [6] and its variants such as probabilistic movement primitives (ProMP) [7] and compliant movement primitives (CMP) [8]. In short, DMP approaches characterize a human demonstration via a second-order dynamical system, formulated with position error, velocity, and a nonlinear forcing term, to obtain the acceleration profile of a smooth movement. The forcing term utilizes a Gaussian basis to encode local variations within the demonstration. DMP-based approaches require first and second derivatives of a trajectory which can be a source of inaccuracy in human-performed trajectories with high frequency content caused by sharp turns or hand tremors. Additionally, DMPs have a relatively large number of hyperparameters and to the best of the authors’ knowledge, there is no tangible interpretation for the weights of the basis functions in the forcing term. For more elaboration on DMP, please see Section IV-B.

Another choice for selecting basis functions is the *Bernstein* basis, with its favorable geometric properties and computationally efficient algorithms, which enables the trajectory generator to efficiently learn demonstrations and enforce constraints along the robots’ trajectories [9]. However, the Bernstein basis often falls short in accurately encoding dexterous movements, leading to inefficiencies and errors in applications demanding high precision, such as robotic surgery. The inherently ultra-smooth nature of these approximations fails to represent the abrupt changes characteristic of human-performed trajectories, thereby necessitating an advanced encoding strategy that can more faithfully replicate the nuanced movements involved.

Inspired by the aforementioned advantages of the MP theory and to address the limitations of existing methods such as DMP, this paper will outline a new basis function so that the resultant encoded trajectory can execute complex movements in a versatile, energy efficient, and adjustable manner and convey valuable insight into the user’s behavior over a finite time interval, yielding numerous useful properties that can be adopted in downstream tasks. The proposed basis offers many advantages including but not limited to: (1) it enjoys a trajectory shape preserving property specifically for dexterous demonstrations, (2) the parameters learned from the observed behavior can be associated with what we define as virtual

attention points (VAPs) of the user that convey tangible information about the executive task and its temporally or spatially similar tasks, (3) the generated VAPs provide an interpretable low-dimensional representation of the original dexterous trajectory which facilitate further optimization and trial-and-error learning of trajectories with low computational demand. Please note that we primarily chose robotic surgery applications to test our method due to its critical necessity for generating trajectories to precisely mimic the nuanced movements of expert human surgeon, ensuring error-free task execution during delicate operations. Our work means robots will be able to perform complex tasks with a level of precision closer to human capabilities, making them safer and more effective for tasks like surgery and other delicate operations. It should be noted that throughout this paper, motion is represented in the Cartesian space or the robot’s joint space, and it is assumed that there exists a low-level controller that converts kinematic variables into motor commands (e.g. force or torque).

The outline of the remainder of the paper is as follows: In Section II, motivations behind VAP method, its theoretical foundation, and implementation details will be discussed. In Section III, the applications of our proposed method in surgical skills assessment and autonomous robotics will be presented. In Section IV, several discussions about the properties and practical details of the proposed method as well as comparative analysis with existing methods will be investigated. Concluding remarks are provided in Section V.

II. METHODOLOGY

A. Motivation

Humans generate fast, dexterous, and optimal trajectories for complex tasks in real world which are beyond the capability of existing MP encoding approaches. It has been shown that incorporating human attention into the imitation learning of the robot can enhance the dexterity and naturalness of the generated trajectories [2]. Such attentive approach leads us to the idea of virtual attention points (VAPs), or namely, primitive intent, which plays an important role in human motion perception and action representation [2]. Beyond the biological motivation, a compact and vectorized representation of human motor skills in terms of specific basis functions can be considered as a feature extraction method. A properly-designed set of basis function that generates VAPs allows for interpretable and explainable knowledge discovery, targeted feedback, and enhances the model’s transparency in downstream tasks [10]. Moreover, the VAP paradigm is also in alignment with the attractor behavior of MPs, which shapes local variations within trajectories, as discussed in Section I.

The central idea of VAP is shown in Fig. 1 in which a given trajectory $P(t)$ starts at source point σ and converges to sink ζ while being attracted by VAPs p_1 and p_2 by a given *attention dynamics*: the mathematical equations that guide how robots focus on different parts of a task. Two sample attention dynamics Ψ and Φ are illustrated in Fig. 1(a) with different temporal width that specify the attraction strength of each VAP as a time-dependent profile. Each ψ_i and ϕ_i reaches its maximum peak at a specific time and diminishes while gaining attraction towards the next VAP. The attractional behavior of VAPs with attention dynamics Ψ is depicted in Fig. 1(b). The effect of attention dynamics on the generated trajectory is shown in Fig. 1(c) with the same set of points $\sigma \rightarrow p_1 \rightarrow p_2 \rightarrow \zeta$ but different attention dynamics Φ . It can be concluded that more spread attention dynamics result in smoother trajectories with less sharp

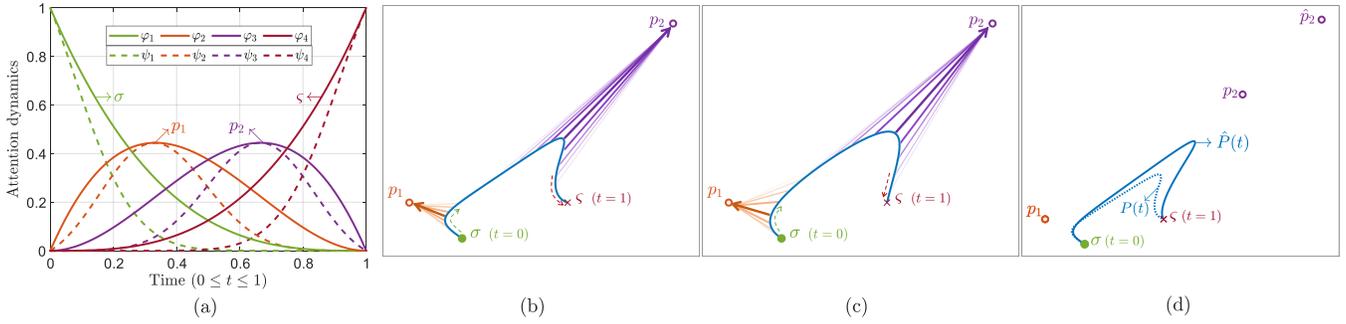


Fig. 1: The concept of VAP in trajectories starting from the source point σ , converging to the sink point ζ , while being attracted towards attention points p_i with specific attention dynamics. (a) Sample illustration of two different attention dynamics $\Psi = [\psi_1, \psi_2, \psi_3, \psi_4]$ and $\Phi = [\varphi_1, \varphi_2, \varphi_3, \varphi_4]$. Attention dynamics represent the influence of each VAP on the trajectory over time, with Ψ being more concentrated and Φ more spread out, indicating different strengths and durations of attraction. (b) The illustration of VAP for four points $\sigma \rightarrow p_1 \rightarrow p_2 \rightarrow \zeta$ based on attention dynamics Ψ . (c) The effect of changing the attention dynamics Ψ to a more spread dynamics Φ on the generated trajectory is shown; spread dynamics yield a smoother, less abrupt path. (d) The effect of changing attention point p_2 on the trajectory $P(t)$ to a distant attention point \hat{p}_2 with the fixed attention dynamics Ψ illustrates how VAPs alter local trajectory shapes depending on their locations and associated dynamics.

turns. Fig. 1(d) investigates the effect of VAPs on the generated trajectory with a fixed set of attention dynamics. If we move p_2 further to point \hat{p}_2 , it generates stronger attractions that rigorously affect the local shape of the trajectory. Each VAP is analogous to the virtual nocking point of a bow and as the bowman pulls them back, the bow string (trajectory) shapes accordingly.

The presented VAP paradigm intuitively makes sense from the brain's computational perspective. The human brain is not advanced enough to simultaneously process all the incoming stimuli, to generate the entire desired path, and execute it ahead of time. Such attentive processing turns out to be a powerful natural heuristic to constrain the search and computational demand, rendering the general object manipulating problem doable for human and tractable for robots [11].

B. Problem Statement

According to the encoding paradigm discussed so far, each continuous trajectory, utilizing a proper basis matrix \mathbf{A} as defined in (3), will be mapped to a finite set of VAPs as coefficients. The basis matrix \mathbf{A} in general represents the fundamental building blocks of a curve-fitting model, such as power or Bernstein basis functions. Each basis function in \mathbf{A} scales with corresponding control gains (VAPs in our case) to reconstruct the desired trajectory with either smooth or sharp transitions as required. In this way, each VAP acts as a natural *shape handle*, permitting intuitive generation or manipulation of the trajectory to satisfy the reconstruction loss. This states that, for a given continuous trajectory $\mathbf{F}(t)$ on a time interval $t \in [t_0, t_1]$ and an error tolerance $\epsilon > 0$, we want to find a parametric polynomial curve $\mathbf{P}_n^{\mathbf{A}}(t)$ of sufficiently high degree n with respect to basis \mathbf{A} , such that

$$\|\mathbf{F}(t) - \mathbf{P}_n^{\mathbf{A}}(t)\| \leq \epsilon, \quad t \in [t_0, t_1]. \quad (1)$$

Replacing the time variable t in (1) with $(t - t_0)/(t_1 - t_0)$ maps $t \in [t_0, t_1]$ to $t \in [0, 1]$ without changing the max norm of function $\mathbf{F}(\cdot)$. As a result, we can restrict our approximation to $t \in [0, 1]$

without loss of generality and define $\mathbf{P}_n^{\mathbf{A}}(t)$ by

$$\mathbf{P}_n^{\mathbf{A}}(t) = \sigma \mathbf{a}_0^n(t) + \sum_{k=1}^{n-1} p_k \mathbf{a}_k^n(t) + \zeta \mathbf{a}_n^n(t), \quad t \in [0, 1] \quad (2)$$

where σ is source, ζ is sink, p_k are gain vectors, and $\mathbf{a}_i^n(t)$ are basis functions for all $0 \leq i \leq n$. (2) can be represented in a matrix multiplication format as

$$\begin{aligned} \mathbf{F}(t) &\approx \mathbf{P}_n^{\mathbf{A}}(t)_{d \times m} = \mathbf{P}_{d \times (n+1)} \mathbf{A}_{(n+1) \times m} \\ &= \begin{bmatrix} | & | & & | & | \\ \sigma & p_1 & \dots & p_{n-1} & \zeta \\ | & | & & | & | \end{bmatrix} \begin{bmatrix} - & \mathbf{a}_0^n(t) & - \\ & \vdots & \\ - & \mathbf{a}_n^n(t) & - \end{bmatrix} \end{aligned} \quad (3)$$

where d is the dimensionality of the trajectory $\mathbf{F}(t)$ encoded by $\mathbf{P}_n^{\mathbf{A}}(t)$ and m is the number of its sample points which is typically much larger than n . Finding points p_i in matrix \mathbf{P} in (3) is equivalent to the problem of least squares fit of polynomial $\mathbf{P}_n^{\mathbf{A}}(t)$ to the trajectory $\mathbf{F}(t)$ with the cost function of $\|\mathbf{F}(t) - \mathbf{P}_n^{\mathbf{A}}(t)\|_2^2$. This problem has a unique solution:

$$\mathbf{P} \approx \mathbf{F}(t) \mathbf{A}^\dagger = \mathbf{F}(t) \mathbf{A}^\top (\mathbf{A} \mathbf{A}^\top)^{-1} \quad (4)$$

where \mathbf{A}^\dagger is the pseudo-inverse of matrix \mathbf{A} .

C. The Choice of Attention Dynamics

The choice of attention dynamics is fundamental to a successful encoding procedure in terms of extracting meaningful and stable attention points with minimal reconstruction loss. \mathbf{A} in (3) should derive benefit from several simple but important attributes. We now briefly describe these key properties and reasons why they are crucial for our attention dynamics.

- 1) Generated points in \mathbf{P} should resemble VAPs in the executed task, so the encoding approach benefits from the mentioned advantages in Section II-A.
- 2) The final approximation presented in (3) should preserve essential properties of the demonstrated trajectory $\mathbf{F}(t)$, such

as monotonicity or convexity. This kind of approximation is referred to as a *shape-preserving* approximation, which, for our specific problem formulated in (1), should have two following properties:

$$\mathbf{P}_n^{\mathbf{A}}(t_i) = \mathbf{F}(t_i), \quad 0 = t_1 < t_2 < \dots < t_m = 1$$

$$\text{sgn}\left(\frac{d\mathbf{P}_n^{\mathbf{A}}(t)}{dt}\bigg|_{t=\hat{t}}\right) = \text{sgn}(\mathbf{F}(t_{l+1}) - \mathbf{F}(t_l)), \quad \hat{t} \in [t_l, t_{l+1}]$$

for all $l = 1, \dots, m-1$ [12]. This suggests that the polynomial basis $\{\mathbf{a}_i^n(t)\}_{i=0}^n$ becomes *normalized totally positive* on $[0, 1]$ which yields $\sum_{i=0}^n \mathbf{a}_i^n(t) = 1$ and $\mathbf{a}_i^n(t) \geq 0$ for $\forall i = 0, 1, \dots, n$ [13]. For such a basis, the total number of sign changes of $\mathbf{P}_n^{\mathbf{A}}(t)$ in (2) cannot exceed the total number of sign variations in coefficients $\{\mathbf{a}_i^n(t)\}_{i=0}^n$. Therefore, the resultant approximation will be specified as shape-preserving. If property 1 holds, connecting consecutive points yields the *attention polygon*, which can be viewed as a directed graph of the user's intent, sketching the caricature of the original trajectory that exaggerates its shape. The assumption that $0 \leq \mathbf{a}_i^n(t) \leq 1$ maps the encoded trajectory inside the attention polygon which is useful for analyzing and manipulating the curve. Furthermore, this feature prevents extreme extrapolations by bounding the trajectory in the convex hull of attention points.

- 3) It is desirable that the encoded trajectory $\mathbf{P}_n^{\mathbf{A}}(t)$ starts from the source at $t = 0$ and converges to the sink at $t = 1$. The combination of this fact and property 2 suggests that $\mathbf{a}_0^n(0) = \mathbf{a}_n^n(1) = 1$. Due to the numerical rounding errors of the pseudo-inverse operation in (4), the first and last columns of the calculated \mathbf{P} matrix (i.e., $\boldsymbol{\sigma}$ and $\boldsymbol{\varsigma}$) exhibit a slight deviation from the first and last points of the trajectory, namely $\mathbf{F}(t_1)$ and $\mathbf{F}(t_m)$, respectively. As a practical perspective, we have to manually set $\boldsymbol{\sigma}$ and $\boldsymbol{\varsigma}$ to $\mathbf{F}(t_1)$ and $\mathbf{F}(t_m)$, respectively.

According to [14], if we use Bernstein basis for attention dynamics \mathbf{A} in (3) defined as follows

$$\mathbf{a}_i^n(t) := \binom{n}{i} (1-t)^{n-i} t^i = \frac{n!}{i!(n-i)!} (1-t)^{n-i} t^i \quad (5)$$

for $\forall i = 0, 1, \dots, n$, the generated points \mathbf{p}_i simulate VAP properties discussed in Section II-A. These properties are illustrated in Fig. 1 for very smooth and simple but not dexterous trajectories. The final estimation $\mathbf{P}_n^{\mathbf{A}}(t)$ based on Bernstein basis is optimally shape-preserving, and the resultant attention polygon gives a better sense of the trajectory shape than the polygon associated with any other polynomial basis. This arises from the fact that property 2 makes \mathbf{A} a column-wise stochastic totally positive matrix—i.e., its elements are non-negative and sum to unity across each column [15] (to see why this fact is correct, please refer to Appendix A). The illustration of Bernstein basis for $n = 4$ is provided as φ_i in Fig. 1(d). Regarding the second required property, as it is clear from (5) and graphical illustration in Fig. 1(d), we will have $\mathbf{a}_0^n(0) = \mathbf{a}_n^n(1) = 1$. The fact that each attention dynamics is exclusive for its attention point or $\arg \max_t \mathbf{a}_i^n(t) = i/n$ is proved in Appendix B for the Bernstein basis. Moreover, the Bernstein polynomial-based approximation $\mathbf{P}_n^{\mathbf{A}}(t)$ is always at least as smooth as the given demonstration $\mathbf{F}(t)$. In other words, if $\mathbf{F}(t)$ has C^r continuity (i.e., $\frac{d^k \mathbf{F}(t)}{dt^k}$ exists and is continuous on $[0, 1]$ for $k = 1, \dots, r$) all

derivatives of $\mathbf{P}_n^{\mathbf{A}}(t)$ up to order r uniformly converge to their corresponding derivatives of $\mathbf{F}(t)$ [16]. In view of the above-mentioned properties, the Bernstein basis is an ideal choice for attention dynamics in the presented VAP trajectory encoding paradigm for very smooth trajectories, however, as we will elaborate later, it may not be optimal for encoding dexterous trajectories with sharp turns.

D. Contributions

Our methodology introduces significant improvements to the problem of MP-based trajectory generation by utilizing the explainable and mathematically intuitive features of Bernstein basis to address challenges in encoding dexterous maneuvers and sharp turns. Please note that trajectory encoding using the Bernstein basis is not a new idea and has been applied in various domains, including simple hand-reaching motion modeling [17], mobile robot path planning [18]–[20], autonomous driving [21], [22], and other fields, often referred to as B'ezier curves in computer graphics. Our investigations suggest that, despite the aforementioned advantages of the Bernstein basis, it may not be ideal for the accurate and meaningful encoding of dexterous human hand trajectories, particularly in the context of surgical trajectories. This is attributed to the extra-smooth nature of B'ezier curves, which is not analogous to the frequent abrupt changes observed in trajectories performed by humans. This contradiction between the nature of B'ezier curves and that of human-performed trajectories causes three main problems. First, to compensate for the ultra-smooth behavior of B'ezier curves, the optimization (4) generates a far-reaching VAP for sharp turns within the trajectory, which is not meaningful in the human sense (we will elaborate on this later). Secondly, to maintain tracking of the trajectory after a sharp turn, the optimization process described by (4) generates an additional distant VAP in the opposite direction, aimed at neutralizing the impact of the preceding far VAP. The iterative back-and-forth pattern in attention points results in a squiggly approximation for the trajectory, leading to the third problem: high encoding error in the vicinity of sharp turns.

One contributing factor that renders the Bernstein basis imperfect for encoding human trajectories is the substantial overlap between bases. This overlap results in an ultra-smooth approximation but comes at the expense of merging the attention effects of nearby VAPs, where 'near' is defined in terms of order and not, for instance, Euclidean distance. In other words, the significant overlap between attention dynamics prevents samples on $\mathbf{P}_n^{\text{Bernstein}}(t)$ from exclusively belonging to one particular VAP, resulting in the generation of zigzag VAPs in the vicinity of sharp turns. To address this limitation, we propose a modification to the Bernstein basis that regulates the overlap between attention dynamics, resulting in meaningful VAPs and accurate trajectory encoding. To achieve this, we multiply each attention dynamics $\mathbf{a}_i^n(t)$ by a Gaussian distribution centered at $t = i/n$ (the time at which the maximum value of $\mathbf{a}_i^n(t)$ occurs, as proven in Appendix B) with a variance of d/δ . The hyperparameter δ will be determined for each given trajectory through an optimization process aimed at minimizing the encoding reconstruction error. As a result, the $\mathcal{N}_{\text{Bernstein}}$ attention dynamics is given by

$$\mathbf{z}_i^{n,\delta}(t) := \frac{\mathcal{N}\left(\frac{i}{n}, \frac{d}{\delta}\right) \mathbf{a}_i^n(t)}{\sum_{j=0}^n \mathcal{N}\left(\frac{j}{n}, \frac{d}{\delta}\right) \mathbf{a}_j^n(t)} = \frac{\binom{n}{i} \mathcal{N}\left(\frac{i}{n}, \frac{d}{\delta}\right) (1-t)^{n-i} t^i}{\sum_{j=0}^n \binom{n}{j} \mathcal{N}\left(\frac{j}{n}, \frac{d}{\delta}\right) (1-t)^{n-j} t^j} \quad (6)$$

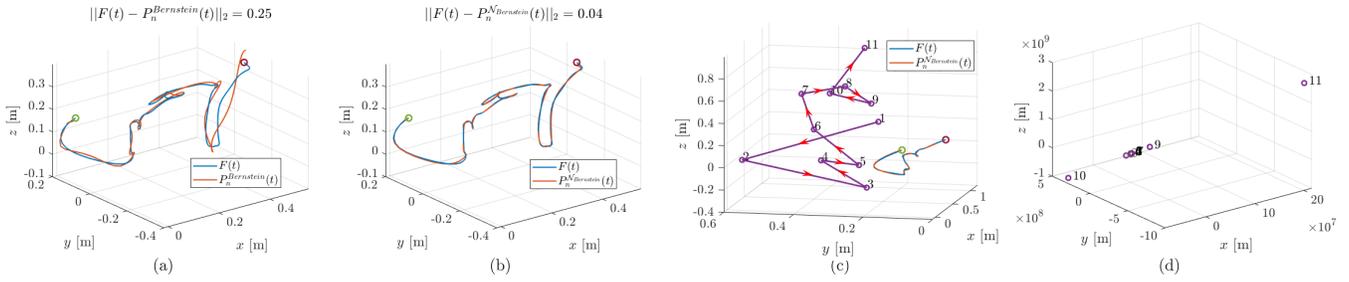


Fig. 2: The advantages of the proposed attention dynamics (6) compared to Bernstein basis in terms of encoding error and generating meaningful attention points. (a) Trajectory encoding quality and reconstruction error based on Bernstein basis (0.25) versus (b) our proposed attention dynamics (0.04), achieving an 84% improvement, especially in sharp turns and the right side of the trajectory. (c) Attention points from (6) accurately replicate human focus on key trajectory segments, offering a more intuitive and interpretable approach ideal for high-risk tasks like surgery, unlike (d) Bernstein basis, which lacks this intuitive alignment.

where

$$\delta = \underset{\delta}{\operatorname{argmin}} \left\| \mathbf{F}(t) - \mathbf{P}_n^{\mathcal{N}_{\text{Bernstein}}}(t) \right\|_2^2 \quad (7)$$

and the denominator in (6) is included to satisfy the normalized totally positive condition required for the final basis. Bases ψ_i in Fig. 1(d) represent $\mathcal{N}_{\text{Bernstein}}$ basis, while ϕ_i represent Bernstein basis. To introduce “spread” dynamics, we modify the Gaussian component in (6) to allow for a broader or narrower focus on specific VAPs. By adjusting the variance parameter δ , we control how long each attention point p_i influences the trajectory over time. A higher δ value results in more spread-out attention dynamics, which generates smoother transitions between VAPs, reducing abrupt changes in trajectory, as shown in Fig. 1(c). For multiple VAPs, the final attention dynamics balance all VAP effects over time. Each VAP p_i has a Gaussian component with a center and spread, where adjusting variance δ controls the influence of each VAP, yielding smoother or sharper transitions. The attention dynamics (6) exhibits all three properties of the Bernstein basis discussed in Section II-C, with significantly reduced encoding loss for dexterous trajectories. Please note that our approach differs from related work, such as [23], which assumes a Gaussian distribution over the weights of Bernstein basis functions (i.e., elements of the matrix \mathbf{P}) and attempts to learn those parameters during the trajectory encoding process.

The VAP approach offers key benefits over direct trajectory mimicry, especially for human-like, dexterous movements (see Fig. 2). Unlike curve fitting, VAPs simplify the trajectory into critical, interpretable points, making it efficient, adjustable, and precise. Fig. 2 illustrates two advantages of the proposed $\mathcal{N}_{\text{Bernstein}}$ basis compared to the original Bernstein basis—specifically, in terms of lumped encoding error and the generation of meaningful VAPs in human terms. Further comparisons of the advantages of the $\mathcal{N}_{\text{Bernstein}}$ basis against other MP-based approaches, as discussed in Section I, will be explored in Section IV-B. Please note that Fig. 2 illustrates a realistic and challenging suturing task from the JIGSAWS dataset [24]. This dataset includes skill-intensive tasks performed by participants across a wide range of expertise, providing a rigorous basis to validate our approach. Deeper and further applications of the VAP method on complex trials from JIGSAWS will be investigated in Section III, with additional theoretical and practical details. According to Fig. 2(a) and Fig. 2(b), we achieved 63% reduction in reconstruction error (i.e., $\|\mathbf{F}(t) - \mathbf{P}_n^{\mathcal{A}}(t)\|_2$) for

the encoding of a minimally invasive suturing task and preserved important local variations within the trajectory. While visual differences in Fig. 2(a) may appear subtle, the discrepancy on the trajectory’s right side can impact high-precision tasks, like robotic surgery, where even minor deviations risk task failure or error accumulation. Moreover, the $\mathcal{N}_{\text{Bernstein}}$ -based encoding avoids the issue of squiggly reconstruction. This not only reduces the reconstruction error but also mitigates the risk of potential danger arising from unnecessary and uncontrolled motions, particularly in safety-critical tasks such as autonomous robot-assisted surgeries. The $\mathcal{N}_{\text{Bernstein}}$ attention dynamics also outperforms the Bernstein basis in terms of generating more meaningful and interpretable VAPs. This issue is illustrated in Fig. 2(c) and Fig. 2(d). As illustrated in Fig. 2(c), the generated VAPs for the trajectory based on the $\mathcal{N}_{\text{Bernstein}}$ basis form a rough sketch of the trajectory and are situated in its vicinity. Furthermore, we observe distant VAPs for sharp turns and nearby ones for smooth and predictable segments, aligning with our intuitive expectations. However, as depicted in Fig. 2(d), the VAPs generated by the original Bernstein basis can reach magnitudes on the order of 10^9 meters for a simple minimally invasive surgery task. In addition to the previously mentioned problems, this behavior is not intuitive in human terms. Please note that this issue is specific to dexterous trajectories. For super-smooth motions, such as reaching tasks, there is no significant difference between $\mathcal{N}_{\text{Bernstein}}$ and the Bernstein basis in terms of reconstruction loss or interpretability of generated VAPs.

E. Properties of the VAPs

Building upon the simplicity and intuitiveness of the VAP approach in encoding dexterous trajectories, there are no predefined selection criteria for attention points; instead, these points are dynamically determined and optimized by the intrinsic nature of the expert’s demonstration. The inherent skills-related behaviors exhibited during the demonstration, especially in how the user performs maneuvers and addresses critical points in delicate scenarios, guide the optimization process to generate explainable and task-related attention points. The VAP approach also exhibits other appealing properties and applications, as follows:

1) *Spatial and Temporal Invariance*: An interesting property of the VAP encoding approach is that generating a scaled version of the demonstrations does not require collecting new data or computations from scratch; we can simply multiply all VAPs to create similar geo-

metric trajectories. In terms of temporal invariance, we can generate trajectories with an arbitrary number of sample points m based on the given VAPs. The execution speed of the trajectory is influenced by the assigned time intervals between consecutive timestamps.

2) *Movement Recognition*: Given the temporal and spatial invariance of the VAP approach, trajectories that are *topologically* similar tend to be represented by a similar set of VAPs. In other words, trajectories that are similar in shape but differ in speed and/or amplitude will result in a similar tensor of VAPs. This feature can be incorporated into applications such as task identification and classification of the user's skill level, which will be elaborated in Section III. Further theoretical and technical advantages of the proposed VAP approach are detailed in the two following sections.

III. APPLICATIONS OF VAP TRAJECTORY ENCODING

A. Basic Application: Highlighting Surgical Skills

In this section, we will discuss how VAPs analyze the precision and consistency of surgical trajectories, offering a nuanced and reliable identification of surgical skill levels compared to prior supervised and unsupervised methods. As emphasized in our previous studies, dependable methods for autonomously assessing surgical skills with informative and instructive feedback are crucial in surgical training programs [10], [25]–[29]. Discovering skills-related features that are meaningful in human terms is a fundamental step in studying human performance trajectories and finding an effective way for automating them [30]–[32]. In this section, we will show that the amount by which a trajectory $\mathbf{F}(t)$ deviates from its ideal smooth encoded polynomial $\mathcal{P}_n^{\mathcal{N}_{\text{Bernstein}}}(t)$ is expressive in terms of revealing skills-related features of executed surgical tasks. Intuitively, this deviation can be interpreted as the error between the human-executed trajectory from its nearest machine-executed counterpart. This suggests that encoding surgical trajectories using $\mathcal{N}_{\text{Bernstein}}$ attention dynamics preserves connections with the underlying skillful behaviors of the user, providing a meaningful and intuitive representation of surgical trajectories. By concatenating the introduced deviations for both translational (i.e., along the \vec{i} , \vec{j} , and \vec{k} directions in Cartesian space) and rotational (i.e., around the roll, pitch, and yaw angles) components of each hand and feeding the resulting feature vector to the t -distributed stochastic neighbor embedding (t -SNE) data visualization technique, we will arrive at the illustration shown in Fig. 3(c). To understand why we employ the t -SNE method, please refer to [25]. As depicted in Fig. 3, the outcomes of this study align with the previous skills evaluation results presented in Fig. 3(a) and Fig. 3(b), which were published in [25] and [10], respectively.

At first glance, trials of the participants with the same level of expertise (i.e., experts Ex_i , intermediates In_i , and novices No_i) tend to cluster near each other. Furthermore, the trials of each user, especially those of the experts, exhibit a high level of coherence, indicating consistency in performing the executive task. In addition to the interpretations mentioned earlier, the results presented in Fig. 3 align with the outcomes of the objective assessment of technical skills method, known as global rating scores (GRS), assigned by a professional gynecologic surgeon during a suturing task. For instance, the trials of In_2 cluster near expert trials, suggesting that this user, despite the assigned skills class label, may have a higher level of expertise comparable to experts. Please note that assigned labels in JIGSAWS are roughly based on working hours with the da Vinci

Surgical Systems. This hypothesis is well supported by the fact that $\text{GRS}_{\text{In}_2} = 3.1 \pm 0.57$ out of 5 is very close to $\text{GRS}_{\text{Ex}_1} = 2.64 \pm 0.47$ and $\text{GRS}_{\text{Ex}_2} = 3.2 \pm 0.3$, and evidently higher than $\text{GRS}_{\text{In}_1} = 2 \pm 0.54$ or other novice participants ($\text{GRS}_{\text{No}_1} = 1.75 \pm 1.07$, $\text{GRS}_{\text{No}_2} = 1.66 \pm 0.3$, and $\text{GRS}_{\text{No}_3} = 2.8 \pm 0.84$).

The merit of employing such an explainable approach, as opposed to its purely data-driven counterparts that prioritize high classification accuracy, lies in its ability to highlight the aforementioned label-free information. Please note that our presented method, which utilizes these meaningful representations, is unaware of the skills labels of participants. In other words, both feature extraction and visualization are entirely unsupervised. It is noteworthy that under situations in which the assigned labels are coarse-grained (e.g., if a participant has 20 to 100 working hours with the da Vinci Surgical System, they are classified as intermediate regardless of their inherent skill level), high-capacity classifiers such as deep learning models tend to be less reliable as they try to extract features (or maybe artifacts) to classify, for instance, In_2 as intermediate when he/she is actually an expert.

B. Advanced Application: Autonomous Robotics

Autonomous robotic surgery remains an open research area due to its clear technical challenges and issues arising from ethics and regulation. In several instances, such as autonomous suturing or knot-tying tasks, we heavily relied on expert demonstrations to generate the final trajectory commands. Unfortunately, demonstrations are imprecise, containing sharp turns, being contaminated with hand tremors, and are not optimal. To overcome such challenges, we require a trajectory encoding approach that returns the nearest smooth trajectory to a given demonstration and encapsulates dexterous maneuvers in terms of a low-dimensional representation for tractable further optimizations.

We considered the 2D navigating task of the da Vinci Surgical System through a squiggly narrow passage in a tissue phantom made from plastisol. This phantom creates a soft environment similar to a real surgical scene (see Fig. 4(a)). This task is a simplified version of trajectory planning in a densely filled environment [33] and simulates the blood suctioning task in a surgical scenario [34]. Moreover, in contrast to other simple surgical benchmark tasks such as peg-in-the-hole or peg transfer, this task involves sharp turns. In the following sections, we address and solve the technical challenges associated with this task step by step.

1) *Trajectory Cloning*: One problem with human demonstrations is that they are not exact. In other words, the source or sink points may not be the same for several trials due to human errors (see Fig. 4(a)). For the source, without loss of generality, we can address this problem by setting all the initial points to zero (please note that in our method, shifting trajectories will result in a corresponding shift in their attention points). For the sink point, we need to relocate it to the desired stop point while maintaining all the relative variations within the trajectory. In simpler terms, we aim to shift the endpoint and adjust all preceding trajectory points so that the relative local variations within the trajectory remain unchanged after the modification (see Fig. 4(b)). Having VAPs simplifies this problem in terms of dimensionality, complexity, and computational cost, as we can modify the VAPs instead of all trajectory timestamps. This means we can set the sink to the desired endpoint and modify VAPs while preserving their relative displacements. Taking minimal

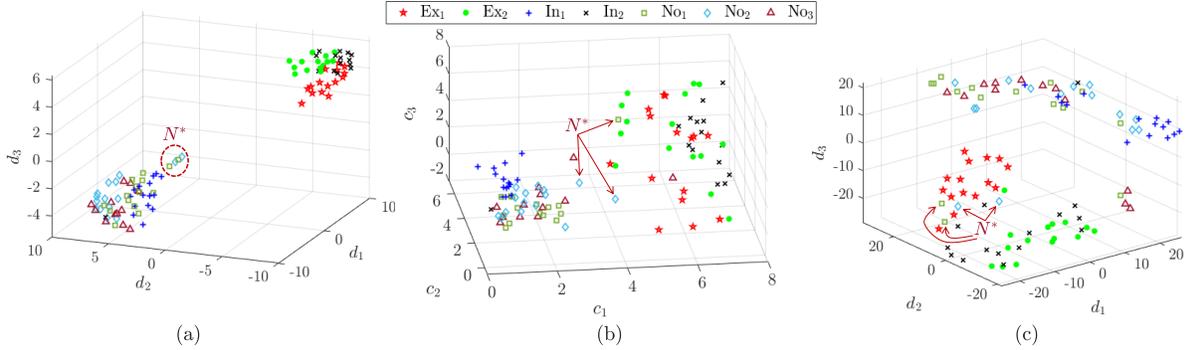


Fig. 3: Comparison of skill evaluation results across different methods using the JIGSAWS suturing task dataset [24]. Each subplot visualizes the clustering of trajectories based on skill level, with t -SNE used to reduce dimensionality for clearer differentiation between novice, intermediate, and expert performances. (a) Skills clustering from a previous domain-adaptive model [25], showing sharp differentiation between skill levels. (b) Clustering from a surgical task-focused method [10], reflecting precision in skill classification. (c) The proposed VAP-based approach, achieving similar sharp separation and consistent clustering.

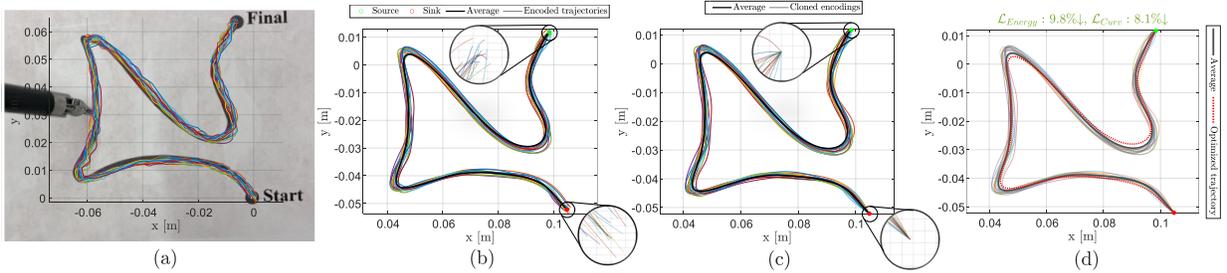


Fig. 4: The process of trajectory cloning to address boundary point mismatches in human demonstrations. (a) Initial human demonstrations in a 2D navigating task performed with the da Vinci Surgical System, highlighting variability in boundary point locations. (b) Adjustment of a sample demonstration's boundary points to align with desired targets, preserving key trajectory characteristics. (c) The outcome of trajectory cloning across all demonstrations, showing consistent alignment to the intended boundary points while maintaining the original movement patterns. (d) Result of applying the optimization method (13) to achieve a smoother, energy-efficient trajectory that retains core trajectory variations from the demonstrations.

inspiration from the pose-graph optimization in the landmark-based simultaneous localization and mapping (SLAM) problem [35], we generate the *incidence matrix* as follows

$$\mathbf{C}^{\mathbf{P}} = \begin{bmatrix} | & | & & | & | \\ \xi_{0p_1} & \xi_{p_1p_2} & \cdots & \xi_{p_{n-2}p_{n-1}} & \xi_{p_{n-1}\varsigma} \\ | & | & & | & | \end{bmatrix}_{d \times n} \quad (8)$$

where $\xi_{xy} = \mathbf{y} - \mathbf{x}$ for arbitrary vectors \mathbf{x} and \mathbf{y} . In other words, each row of $\mathbf{C}^{\mathbf{P}}$ contains all of the relative variations between corresponding axes of source, attention, and sink points. If we change ς to $\bar{\varsigma}$, rows of $\mathbf{C}^{\mathbf{P}}$ no longer are in the same direction of their corresponding rows in $\mathbf{C}^{\mathbf{P}}$. Our goal is to modify VAPs p_i in \mathbf{P} to fabricate \bar{p}_i for all $1 \leq i \leq n-1$ and create $\bar{\mathbf{P}}$ in such a way that corresponding rows in $\mathbf{C}^{\mathbf{P}}$ and $\mathbf{C}^{\bar{\mathbf{P}}}$ are in the same direction (i.e., identical local relative variations between VAPs before and after the modification). As a result, our problem will turn into an optimization problem as follows

$$\bar{p}_1, \dots, \bar{p}_{n-1} = \underset{p_1, \dots, p_{n-1}}{\operatorname{argmin}} \sum_{i=1}^d \operatorname{cosineLoss}(\mathbf{C}_{i,:}^{\mathbf{P}}, \mathbf{C}_{i,:}^{\bar{\mathbf{P}}}) \quad (9)$$

where $\mathbf{C}_{i,:}^{\mathbf{P}}$ and $\mathbf{C}_{i,:}^{\bar{\mathbf{P}}}$ are i^{th} rows of $\mathbf{C}^{\mathbf{P}}$ and $\mathbf{C}^{\bar{\mathbf{P}}}$, respectively. The results of this approach are depicted in the transformation of trajectories from Fig. 4(b) to those in Fig. 4(c). One interesting observation is that the deviation resulting from the endpoint modification diminishes as we backpropagate through the trajectory in time. This procedure will *clone* a similar trajectory with identical local variations compared to its ancestor but with a different endpoint. This observation also makes intuitive sense: the user's deviation from the intended sink point does not occur instantaneously at the end. The error onset happens before completing the task and propagates while approaching the final goal. That is why modifying the sink while keeping all VAPs the same is not a good idea.

2) *Trajectory Curvature Constraints*: Due to task constraints and kinematic limitations, such as nonholonomic constraints in mobile robots [36] or joint limits in manipulator robots [37], a feasible end effector path must exhibit bounded curvature characteristics. The curvature is the amount by which a curve deviates from being a straight line; sharper turns result in higher curvatures. Assuming each demonstration component (see Section III-B) as

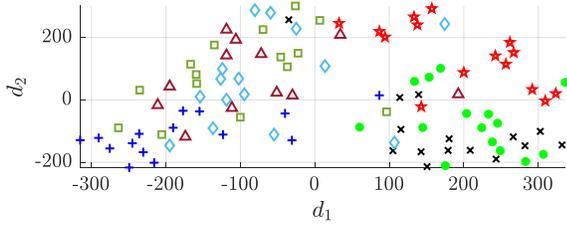


Fig. 5: t -SNE visualization of suturing trials using curvature (κ) analysis for translational and rotational hand data. The plot reveals an accurate skill-based clustering, with curvature features distinguishing expertise. The same legend from Fig. 3 applies here.

$\mathbf{F} = [\mathbf{x}, \mathbf{y}, \mathbf{z}]^\top$, the curvature will be defined as

$$\kappa(\mathbf{F}) := \frac{\|\dot{\mathbf{F}} \times \ddot{\mathbf{F}}\|_2}{\|\dot{\mathbf{F}}\|_3^2} = \frac{\sqrt{\sum_{\text{cyc}} (\dot{\mathbf{x}}\dot{\mathbf{z}} - \dot{\mathbf{z}}\dot{\mathbf{x}})^2}}{(\dot{\mathbf{x}}^2 + \dot{\mathbf{y}}^2 + \dot{\mathbf{z}}^2)^{\frac{3}{2}}} \quad (10)$$

where the dot denotes differentiation with respect to the time parameter t , ‘ \times ’ is the vector cross product, and \sum_{cyc} denotes the cyclic sum that cycles through all the variables \mathbf{x} , \mathbf{y} , and \mathbf{z} . The unit of κ can be interpreted as radians per meter. Additionally, we can interpret the curvature of a curve as the reciprocal radius of a circle with the same curvature tangent to the given point of that curve—commonly known as the osculating circle, which is the circle that best approximates the curve near this point.

In addition to the previously mentioned significance of curvature in autonomous path planning and subjective studies on the relationship between curvature and the user’s skill level [38], our observations suggest that the maximum value, minimum value, and temporal variations of κ convey crucial information regarding the execution quality of surgical tasks. Please note that due to the presence of a few points in κ that are much larger than the bulk of the data (i.e., sparse sharp turns within the trajectory), we investigate κ in logarithmic scales to respond to skewness towards large values. To highlight the temporal variation of κ , we use the fast Fourier transform (FFT) for each trajectory component [27]. If we calculate the minimum and maximum values of each component for two hands, concatenate them with temporal features extracted by FFT, and feed the resultant feature vector to t -SNE data visualization method, we will have an illustration shown in Fig. 5. It is beyond dispute that the pattern of surgical trials shown in Fig. 5 is similar to the patterns shown in Fig. 3. The consistency between the representation based on curvature analysis and other surgical skills evaluation methods presented in this paper and other work of authors reinforces the fact that trajectory curvature analysis plays an important role in synthesising trajectories for autonomous robotic tasks.

3) *Trajectory Economy of Motion*: As authors stated in [25], trajectory economy of motion reflects the total energy demand for accomplishing a task and is akin to the level of expertise of the user, i.e., more energy economic motions reflects higher user’s skills level. This is because surgeon’s high energy consumption results in higher energy injection and higher execution velocity in the patient-side robot which increases the risk of danger and trauma in an operation. Inspired by [25], we define energy loss for a given

discrete trajectory $\tau[t] = \{\tau_i\}_{i=1}^T$ as

$$\mathcal{L}_{\text{Energy}} = \frac{1}{2} \sum_{t=1}^T \dot{\tau}^2[t] \quad (11)$$

where $\dot{\tau}[t]$ and T are time derivative and the total number of time samples of demonstrated trajectory $\tau[t]$, respectively. For further details on this loss and possible complications, please read Section II-C of the authors’ recent work [25]. It is worth mentioning that although the clinically accepted metric of total path length [38] is considered as a plausible skills-related feature, we do not consider it in our optimization procedure. This is because empirically and theoretically we observed that this constraint will force the optimization algorithm to return a straight line which deteriorates important variations within the trajectory (for the theoretical proof see Appendix C). Moreover, based on (A.2) in Appendix C, minimizing $\mathcal{L}_{\text{Energy}}$ as an upper bound for the total path length, eventually minimizes the total path length of the trajectory without destroying its valuable temporal variations.

4) *Trajectory Optimization*: In addition to the mentioned desired improvements, it is crucial for the final optimized trajectory τ^* to remain similar to the original demonstrations. As a result, the lumped reconstruction error between τ^* and all of the demonstrations τ_i

$$\mathcal{L}_{\text{Recons}} = \frac{1}{N} \sum_{i=1}^N \|\tau^* - \tau_i\|_2^2 \quad (12)$$

should be minimized in the trajectory optimization process as well. As a result, the optimized trajectory τ^* will be achieved from the following optimization

$$\mathbf{p}_1^*, \dots, \mathbf{p}_{n-1}^* = \underset{\mathbf{p}_1, \dots, \mathbf{p}_{n-1}}{\text{argmin}} \left(\mathcal{L}_{\text{Recons}} + \alpha \underbrace{\max(\kappa)}_{\mathcal{L}_{\text{Curv}}} + \beta \mathcal{L}_{\text{Energy}} \right) \quad (13)$$

where α and β are weights determining the relative importance of curvature and energy losses compared to the reconstruction loss, respectively. The value of the hyperparameters α and β totally depends on the precision of the task and physical properties of the surgical environment (e.g., presence of sensitive organs or hard objects such as bones).

5) *Experimental Results*: Due to the flexible environment of the surgical scene shown in Fig. 4(a), we set $\alpha = 5e-4$ and $\beta = 50$ to achieve a well-behaved trajectory in terms of energy demand and curvature properties out of original demonstration with no concern about damaging the surgical tool or tissue. We initialized our optimization with the average trajectory $\bar{\tau}$ shown in 4(c) and converged to the optimal trajectory τ^* , as depicted in Fig 4(d), which is 8.1% and 9.8% superior to the average of the total demonstrations τ_i ($\forall i = 1, 2, \dots, 10$) in terms of curvature properties and economy of motion, respectively. Moreover, it exhibits minimal kinematic deviation from its original human demonstrations (video).

IV. DISCUSSIONS

A. Encoding Manipulability

From another perspective, we can think of attention dynamics \mathbf{A} in (3) as a scaling factor of the input VAPs within matrix \mathbf{P} to produce the output encoded trajectory $\mathbf{P}_n^{\mathbf{A}}(t)$. In this case it is beneficial to isolate the element-wise amplitude of \mathbf{P} to quantitatively

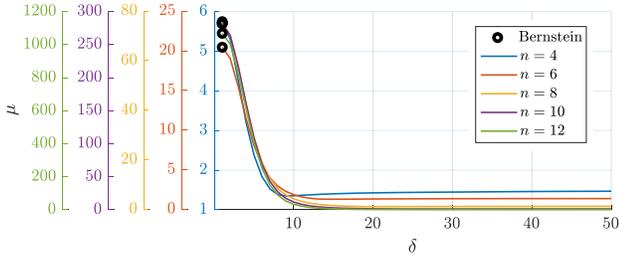


Fig. 6: The advantage of $\mathcal{N}_{\text{Bernstein}}$ relative to the original Bernstein attention dynamics in terms of enhanced encoding manipulability μ . Lower μ values for \mathcal{N} Bernstein indicate improved balance in encoding, allowing it to handle sharp trajectory changes more effectively without generating extreme virtual attention points.

characterize the nature of \mathbf{A} . To do so, we assume that source, sink, and all VAPs are located inside a unit hypersphere or equivalently

$$\|\boldsymbol{\sigma}\|_2, \|\boldsymbol{\varsigma}\|_2, \|\mathbf{p}_i\|_2 \leq 1, \quad i = 1, 2, \dots, n. \quad (14)$$

(14) yields

$$\|\boldsymbol{\sigma}\|_2^2 + \|\boldsymbol{\varsigma}\|_2^2 + \sum_{i=1}^{n-1} \|\mathbf{p}_i\|_2^2 = \text{Tr}(\mathbf{P}^\top \mathbf{P}) \leq n+1 \quad (15)$$

where $\text{Tr}(\cdot)$ is the trace function of a given matrix. According to (4), we have

$$\begin{aligned} \text{Tr}(\mathbf{P}^\top \mathbf{P}) &= \text{Tr}\left(\left[\mathbf{F}(t) \mathbf{A}^\dagger\right]^\top \left[\mathbf{F}(t) \mathbf{A}^\dagger\right]\right) = \text{Tr}\left(\left[\left(\mathbf{A}^\dagger\right)^\top \mathbf{F}^\top(t)\right] \left[\mathbf{F}(t) \mathbf{A}^\dagger\right]\right) \\ &\stackrel{*}{=} \text{Tr}\left(\mathbf{F}(t) \mathbf{A}^\dagger \left(\mathbf{A}^\dagger\right)^\top \mathbf{F}^\top(t)\right) \end{aligned} \quad (16)$$

where equality (*) comes from the fact that the trace function is invariant under cyclic permutations. If we replace \mathbf{A}^\dagger with its singular-value decomposition form, i.e., $\mathbf{A}^\dagger = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}$, (16) yields

$$\begin{aligned} \text{Tr}(\mathbf{P}^\top \mathbf{P}) &= \text{Tr}\left(\mathbf{F}(t) \left[\mathbf{U}\boldsymbol{\Sigma}\underbrace{\mathbf{V}\mathbf{V}^\top}_{=\mathbf{I}}\boldsymbol{\Sigma}^\top \mathbf{U}^\top\right] \mathbf{F}^\top(t)\right) \\ &= \text{Tr}\left(\mathbf{F}(t) \mathbf{U} \left(\boldsymbol{\Sigma}\boldsymbol{\Sigma}^\top\right) \mathbf{U}^\top \mathbf{F}^\top(t)\right) \\ &\stackrel{*}{=} \text{Tr}\left(\boldsymbol{\zeta}\boldsymbol{\Sigma}^2\boldsymbol{\zeta}^\top\right) \stackrel{(15)}{\leq} n+1 \end{aligned} \quad (17)$$

where \mathbf{I} is identity matrix, $\boldsymbol{\zeta} = \mathbf{F}(t)\mathbf{U}$, and (*) holds since $\boldsymbol{\Sigma}$ is a diagonal matrix. Considering $\varphi_0^2, \varphi_1^2, \dots, \varphi_n^2$ as diagonal elements of $\boldsymbol{\Sigma}^2$, (17) yields

$$\sum_{i=0}^n \frac{w_i^2}{n+1} \varphi_i^2 \leq 1 \quad (18)$$

where w_i are constant weights from matrix multiplications and trace function in (17) and φ_i are singular values of matrix \mathbf{A}^\dagger that shape the diagonal elements of matrix $\boldsymbol{\Sigma}$. Since we have

$$\mathbf{A} = \left(\mathbf{A}^\dagger\right)^\dagger = (\mathbf{U}\boldsymbol{\Sigma}\mathbf{V})^\dagger = \mathbf{V}^\dagger \boldsymbol{\Sigma}^\dagger \mathbf{U}^\dagger \stackrel{**}{=} \mathbf{V}^\top \boldsymbol{\Sigma}^{-1} \mathbf{U}^\top, \quad (19)$$

singular values of \mathbf{A} are equal to φ_i^{-1} . Please note that equality (**) holds from the fact that \mathbf{U} and \mathbf{V} are orthogonal matrices and hence, their pseudo-inverses are equal to their transpose. In this way, (18) represents the equation for an axis-aligned hyperellipse in a

new coordinate system obtained by rotating (i.e., via multiplying matrices \mathbf{U} and \mathbf{V}) and scaling (i.e., via multiplying matrix $\boldsymbol{\Sigma}$) the original hypersphere introduced in (14). In other words, source, sink, and VAPs defined under the condition (14) will generate the encoded trajectory $\mathbf{P}_n^{\mathbf{A}}(t)$ inside the hyperellipse (18) with the radii of φ_i^{-1} , $i=0, 1, \dots, n$, which according to (19) are singular values of attention dynamics \mathbf{A} .

One elegant investigation here is examining the shape of hyperellipse (18) to investigate whether its radii are proportionate or not. For a given attention dynamics \mathbf{A} , we define its *encoding manipulability* as

$$\mu_{\mathbf{A}} = \frac{\max(\varphi_i^{-1})}{\min(\varphi_i^{-1})} = \frac{\max(\varphi_i)}{\min(\varphi_i)}. \quad (20)$$

For large values of μ , we end up with a hyperellipse which is simultaneously very thin and very stretched in particular directions. This means that under the confined VAPs assumption, the maneuverability of encoded trajectory $\mathbf{P}_n^{\mathbf{A}}(t)$ in thin directions is very low and basis \mathbf{A} fails to accommodate large and abrupt variations in those directions. As a result, the algorithm needs to generate far VAPs to compensate the attention dynamics shortcomings in those directions which is the source of other encoding problems that was elaborated previously. On the other hand, when $\mu_{\mathbf{A}}$ decreases and approaches to 1, the hyperellipse transfigures to a hypersphere (or isotropic hyperellipse) and we have a balanced space for accommodating desired trajectories with variations in arbitrary directions which results better encoding manipulability.

As shown in Fig. 6, the proposed $\mathcal{N}_{\text{Bernstein}}$ attention dynamics enjoys better encoding manipulability (i.e., low $\mu_{\mathcal{N}_{\text{Bernstein}}}$ value defined in (20)) compared to the original Bernstein basis for all possible polynomial orders. Moreover, as the hyperparameter δ defined in (6) increases (i.e., less overlap between bases), the $\mu_{\mathcal{N}_{\text{Bernstein}}}$ metric decreases as well. One may question that why we do not use familiar power basis for our investigations. According to Fig. 7, $\mathcal{N}_{\text{Bernstein}}$ attention dynamics for $\delta = 2$ enjoys lower μ metric compared to power basis for all polynomial with the order of $n \geq 2$ (since we have at least source and sink points, the order of polynomial should be at least 1). Please note that $\delta = 2$ is a conservative choice since according to (6), each normal distribution has a variance equal to the half of the length of the given trajectory which means a mild affect on bases due to the long tail of $\mathcal{N}\left(\frac{i}{n}, \frac{d}{\delta}\right)$ in (6). In total, the proposed VAP approach benefits from better encoding manipulability compared to the original Bernstein basis, power basis, and, as will be shown in Section IV-B, the DMP method. This enables more precise and skillful maneuvers around critical points of interest, such as sensitive anatomical structures in robotic surgeries.

B. VAP vs. DMP

In this section we will highlight the advantages of our proposed method compared to the popular trajectory encoding approach DMP. The DMP formulation of a given demonstration \mathbf{y} with forcing term \mathbf{f} is as follows [6]

$$\begin{aligned} \ddot{\mathbf{y}} &= \alpha_y(\beta_y(\mathbf{g} - \mathbf{y}) - \dot{\mathbf{y}}) + \mathbf{f} \\ \text{where: } \mathbf{f} &= \frac{\sum_{i=1}^n \boldsymbol{\xi}_i w_i}{\sum_{i=1}^n \boldsymbol{\xi}_i} \mathbf{x} (\mathbf{g} - \mathbf{y}_0) \\ \boldsymbol{\xi}_i &= \exp(-(\mathbf{x} - \boldsymbol{\mu}_i)/\sigma_i^2), \quad \dot{\mathbf{x}} = -\alpha_x \mathbf{x} \end{aligned} \quad (21)$$

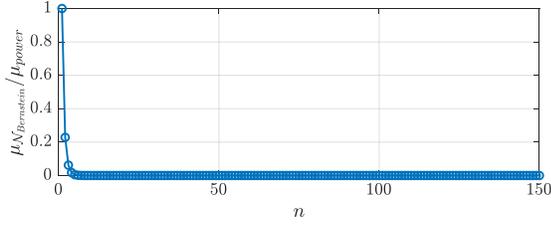


Fig. 7: Relative encoding manipulability defined in (20) between $\mathcal{N}_{\text{Bernstein}}$ and power attention dynamics for $\delta=2$ and orders $n \geq 1$. Lower μ values indicate that $\mathcal{N}_{\text{Bernstein}}$ achieves a more isotropic encoding space, balancing trajectory variations across all directions and reducing the need for exaggerated VAPs to handle abrupt changes.

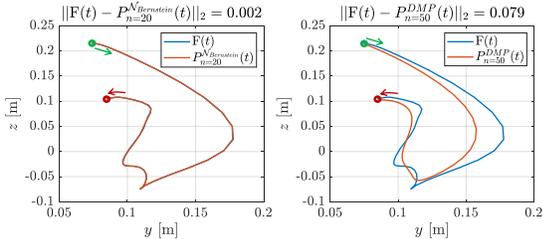


Fig. 8: Trajectory encoding performance of VAP approach vs. DMP. The VAP method achieves significantly lower reconstruction error for complex, sharp movements, thanks to basis functions tailored for capturing rapid trajectory shifts and critical points of dexterity.

where \mathbf{g} is goal point (i.e., sink), $\alpha_y \beta_y$ is the spring coefficient (or the gain for P controller), α_y is the damping term (or the gain for D controller), and α_x is the coefficient for the canonical dynamical system, denoted as \mathbf{x} which decays from 1 to 0 through time. According to (21), the forcing function \mathbf{f} is a normalized weighted summation of Gaussian bases ξ_i that are activated as \mathbf{x} converges to zero. \mathbf{x} in \mathbf{f} is a diminishing term that benefits the stability of the dynamical system and guarantees that the contribution of the forcing term goes to zero when we approach to the sink point \mathbf{g} . The term $\mathbf{g} - \mathbf{y}_0$ in \mathbf{f} is spatial scaling term which is for the sake of scalability of the solution and enables us to move the goal farther away or closer to get a scaled version of the trajectory.

The first and the most important drawback of DMP compared to our method is its need for sensing the acceleration profile of the trajectory which can be troublesome in encoding human dexterous demonstrations with high frequency contents caused by hand tremors and sharp turns. This feature makes DMP to remain a good method for robotic applications with streamlined monotonic movements realized with symmetric bell-shaped velocity profiles such as reaching tasks or wheeled robot motions. Fig. 8 illustrates the fact that our method with lower number of basis compared to DMP can encode dexterous trajectories with 97% lower reconstruction error. This limitation of DMP can be theoretically investigated via the concept of encoding manipulability discussed in Section IV-A. As shown in Fig. 9, for $n \geq 4$ VAP encoding approach with $\mathcal{N}_{\text{Bernstein}}$ basis enjoys lower encoding manipulability than DMP with Gaussian basis. Since dexterous trajectories require high-order polynomial estimations, our method significantly outperforms DMP in generating accurate encoding for such complicated demonstrations.

Apart from such limitations, DMP has extra hyperparameters

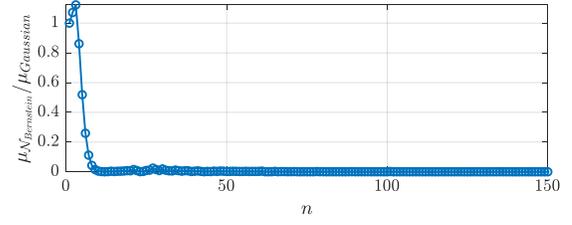


Fig. 9: Relative encoding manipulability defined in (20) between VAP encoding approach with $\delta = 2$ and DMP for orders $n \geq 1$. Lower μ values indicate higher encoding flexibility of the VAP approach compared to DMP to accommodate abrupt trajectory turns.

such as α_x , α_y , and β_y which makes it hard for the user to tune the model for achieving desired outcome. As mentioned in Section I, there is no meaningful interpretation for weights w_i of the basis functions ξ_i as their temporal locations (i.e., μ_i) constantly change with different values for α_x . However, as we showed before, the coefficients in our method can be viewed as VAPs of the user while performing the task and convey rough but important information about the executive trajectory.

Admittedly, DMP provides an online approach that enables real-time adaptation of trajectories in response to changes in goal position, orientation, or other parameters, making it robust in handling variable environmental conditions and unstructured scenarios. This makes DMPs particularly effective for tasks like robotic reaching or grasping, where continuous, smooth adjustments may be required mid-execution. However, in highly structured tasks such as surgical procedures, a pre-planned, consistent trajectory is often preferable. Our VAP approach, as an offline encoding method, captures intricate details of dexterous tasks and enables precise, modifiable execution speed without altering the trajectory in response to new information. This makes it particularly suitable for robotic surgery tasks, where adherence to pre-defined paths, such as suturing or incision placements, aligns with procedural checklists, anatomical constraints, and dexterity requirements inherent in the surgical environment. Additionally, our VAP approach allows precise modifications to execution speed, timing, and trajectory scaling within a pre-planned framework, preserving time-accurate precision and ensuring smoothness through the differentiable basis functions. This guarantees first and second derivative continuity across the trajectory, making it ideally suited for surgical tasks where consistency and smooth, precise motion are critical. In terms of computational efficiency, DMP is roughly twice as fast as VAP due to its simpler basis function, making it ideal for real-time applications. For offline tasks, however, VAP's added processing time is justifiable, as it delivers detailed trajectory encoding suited for pre-planned, intricate movements, which don't require real-time adaptability.

C. Connections Between $\mathcal{L}_{\text{Curv}}$ and $\mathcal{L}_{\text{Energy}}$

In (13), sharp turns or sudden uncontrolled motions within a given trajectory will be penalized explicitly by $\mathcal{L}_{\text{Curv}}$ and implicitly by $\mathcal{L}_{\text{Energy}}$ by minimizing instantaneous changes between two consecutive timestamps. Under these circumstances, one may wonder if one of these metrics is redundant for the trajectory optimization process. However, the ablation study on hyperparameters α and β shown in Table I suggests that not only they are not redundant but also they

TABLE I: Ablation study on the correlation of $\mathcal{L}_{\text{Energy}}$ and $\mathcal{L}_{\text{Curv}}$.

α	β	$\mathcal{L}_{\text{Energy}}$ reduction	$\mathcal{L}_{\text{Curv}}$ reduction	$\mathcal{L}_{\text{Recons}}$
5e-4	50	11.1*	5.3*	1e-5
5e-4	0	1.3	4.1	6e-6
0	50	8.8	0.9	4e-6

have a positive correlation with each other and together, they generate τ^* which is better than just optimizing one parameter at a time.

D. Limitations and Applicability

The VAP method is tailored to replicate human expert movements, generally within a frequency range of approximately 20 Hz [27], which aligns well with the natural dynamics of human actions in dexterous tasks. This design enhances VAP's suitability for applications focused on human-performed tasks. However, this frequency constraint may limit its ability to achieve high-fidelity reconstruction for movements that exceed typical human motion dynamics. For example, applications generating synthetic trajectories or modeling rapid, non-human-like motions might not benefit optimally from VAP's encoding approach. Moreover, VAP may face challenges when applied to datasets generated by novice users, who may perform unpredictable or unsteady movements. These novice-level actions often include random, jagged, or unintended motions, as well as potential hand tremors, which can result in high-frequency fluctuations. Such variability could hinder VAP's ability to generalize effectively and may lead to suboptimal reconstruction performance. Researchers are advised to consider these limitations when applying VAP to datasets that involve non-expert performers or high-frequency actions outside the scope of expert-level human behavior. Notably, these limitations do not contradict the primary focus of this study, which is to accurately encode dexterous trajectories performed by users with a baseline level of expertise, ultimately supporting autonomous replication of such tasks in surgical and robotic settings.

V. CONCLUSIONS

A new basis function for encoding dexterous surgical trajectories to the low-dimensional user's virtual attention points was presented in this paper. It has been demonstrated that the proposed method outperforms other existing MP-based trajectory encoding approaches, such as DMP, showing a 97% improvement in the accuracy of encoding agile movements with possible sharp turns. Such representation allowed us to investigate, preprocess, and optimize a given demonstration in a tractable, meaningful, and intuitive manner. Incorporating the VAP approach complemented by the trajectory cloning algorithm improves the curvature properties and economy of motion of the raw demonstrations by 8.1% and 9.8%, respectively, resulting in the generality and enhanced performance of the proposed framework for real-world applications. According to performed experiments in the fields of surgical skills assessment and autonomous robotic surgery, the performance of the proposed approach was tested in practice. While VAP currently handles non-periodic trajectories effectively, incorporating a periodic variant—similar to periodic DMP models—could extend its utility for repetitive surgical tasks. The introduction of VAPs as a basis for encoding trajectories is not limited to robotic surgery; it also paves

the way for the development of dexterous robotic systems across various application domains, including exoskeletons, collaborative robots, and futuristic technologies such as games or virtual reality.

APPENDIX

A. Stochastic Totally Positivity of Matrix \mathbf{A}

To prove that the matrix \mathbf{A} is a column-wise stochastic totally positive, we have to show that $\sum_{i=0}^n \mathbf{a}_i^n(t) = 1$ and $\mathbf{a}_i^n(t) \geq 0$ for $\forall i = 0, 1, \dots, n$ and all $t \in [0, 1]$. These facts will be proved by the following expression:

$$1 = (t + (1-t))^n = \sum_{i=0}^n \binom{n}{i} (1-t)^{n-i} t^i = \sum_{i=0}^n \mathbf{a}_i^n(t). \quad (\text{A.1})$$

B. Calculating $\text{argmax}_t \mathbf{a}_i^n(t)$

According to (5), for $i = 0$ and $i = n$ it is obvious that $\text{argmax}_t \mathbf{a}_i^n(t)$ is equal to 0 and 1, respectively. Assuming that $0 < i < n$, we will have

$$\begin{aligned} \frac{d\mathbf{a}_i^n(t)}{dt} &= \binom{n}{i} (1-t)^{n-i-1} t^{i-1} [i(1-t) - (n-i)t] = 0 \\ \Rightarrow [i(1-t) - (n-i)t] &= 0 \Rightarrow t = \frac{i}{n}. \end{aligned}$$

C. Trajectory Total Path Length and User's Economy of Motion

According to the definition of the economy of motion for a given trajectory $\tau = \{\tau_i\}_{i=0}^T$ discussed in Section III-B3, we have

$$\begin{aligned} \mathcal{L}_{\text{Energy}} &= \frac{1}{2} \sum_{i=1}^T (\tau_i - \tau_{i-1})^2 \\ &\stackrel{\star\star}{\geq} \frac{1}{2T} \left(\sum_{i=1}^T |\tau_i - \tau_{i-1}| \right)^2 = \frac{1}{2T} \mathcal{L}_{\text{Path}}^2 \end{aligned} \quad (\text{A.2})$$

where $\mathcal{L}_{\text{Path}}$ is the total path length of the executive trajectory and inequality ($\star\star$) is derived from the famous RMS-AM inequality. (A.2) suggests that $\sqrt{\mathcal{L}_{\text{Energy}}}$ is an upper bound for $\mathcal{L}_{\text{Path}}$ and minimizing the total energy demand for executing the task eventually minimizes the total path length. However, jointly minimizing total path length and total energy demand is not a good idea since the algorithm tends to generate straight line connecting the source and sink points which might mitigate important temporal variations. This is because equality for ($\star\star$) holds when all summation elements in (A.2) become the same, i.e., $|\tau_i - \tau_{i-1}| = |\tau_j - \tau_{j-1}|$ for all possible i and j which under the constraint of minimizing the total kinetic energy τ should be a straight line.

REFERENCES

- [1] Neville Hogan and Dagmar Sternad. Dynamic primitives of motor behavior. *Biological cybernetics*, 106:727–739, 2012.
- [2] Tetsunari Inamura, Naoki Kojo, Tomoyuki Sonoda, Kazuyuki Sakamoto, Kei Okada, and Masayuki Inaba. Intent imitation using wearable motion capturing system with on-line teaching of task attention. In *5th IEEE-RAS International Conference on Humanoid Robots, 2005.*, pages 469–474. IEEE, 2005.
- [3] Peter Pastor, Mrinal Kalakrishnan, Franziska Meier, Freek Stulp, Jonas Buchli, Evangelos Theodorou, and Stefan Schaal. From dynamic movement primitives to associative skill memories. *Robotics and Autonomous Systems*, 61(4):351–361, 2013.
- [4] Tamar Flash and Ealan Henis. Arm trajectory modifications during reaching towards visual targets. *Journal of cognitive Neuroscience*, 3(3):220–230, 1991.

- [5] Adrià Colomé and Carme Torras. *Reinforcement Learning of Bimanual Robot Skills*. Springer, 2020.
- [6] Stefan Schaal, Jan Peters, Jun Nakanishi, and Auke Ijspeert. Learning movement primitives. In *Robotics Research. The Eleventh International Symposium: With 303 Figures*, pages 561–572. Springer, 2005.
- [7] Alexandros Paraschos, Christian Daniel, Jan R Peters, and Gerhard Neumann. Probabilistic movement primitives. *Advances in neural information processing systems*, 26, 2013.
- [8] Tadej Petrič, Andrej Gams, Leon Žlajpah, and Aleš Ude. Online learning of task-specific dynamics for periodic tasks. In *2014 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 1790–1795. IEEE, 2014.
- [9] Calvin Jensen. *BeBOT: Bernstein Polynomial-Based Method for Solving Complex Optimal Motion Planning Problems*. PhD thesis, The University of Iowa, 2023.
- [10] Abed Soleymani, Xingyu Li, and Mahdi Tavakoli. Surgical procedure understanding, evaluation, and interpretation: A dictionary factorization approach. *IEEE Transactions on Medical Robotics and Bionics*, 4(2):423–435, 2022.
- [11] John K Tsotsos. Motion understanding: Task-directed attention and representations that link perception with action. *International Journal of Computer Vision*, 45:265–280, 2001.
- [12] Lj M Kocić and GV Milovanović. Shape preserving approximations by polynomials and splines. *Computers & Mathematics with Applications*, 33(11):59–97, 1997.
- [13] TNT Goodman. Shape preserving representations. In *Mathematical methods in computer aided geometric design*, pages 333–351. Elsevier, 1989.
- [14] Jesús M Carnicer and Juan Manuel Pena. Shape preserving representations and optimality of the bernstein basis. *Advances in Computational Mathematics*, 1:173–196, 1993.
- [15] Marie-Laurence Mazure. Blossoms and optimal bases. *Advances in Computational Mathematics*, 20:177–203, 2004.
- [16] PJ Davis. Interpolation & approximation dover publication. *Inc., New York*, 1975.
- [17] Julian J Faraway, Matthew P Reed, and Jing Wang. Modelling three-dimensional trajectories by using bézier curves with application to hand motion. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 56(5):571–585, 2007.
- [18] KG Jolly, R Sreerama Kumar, and R Vijayakumar. A bezier curve based path planning in a multi-agent robot soccer system without violating the acceleration limits. *Robotics and Autonomous Systems*, 57(1):23–33, 2009.
- [19] Fengyu Zhou, Baoye Song, and Guohui Tian. Bézier curve based smooth path planning for mobile robot. *Journal of Information & Computational Science*, 8(12):2441–2450, 2011.
- [20] Lin Xu, Maoyong Cao, and Baoye Song. A new approach to smooth path planning of mobile robot based on quartic bezier transition curve and improved pso algorithm. *Neurocomputing*, 473:98–106, 2022.
- [21] Trent Weiss, Varundev Suresh Babu, and Madhur Behl. Bezier curve based end-to-end trajectory synthesis for agile autonomous driving. In *NeurIPS 2020 Machine Learning for Autonomous Driving Workshop*, 2020.
- [22] Il Bae, Jin Hyo Kim, Jaeyoung Moon, and Shiho Kim. Lane change maneuver based on bezier curve providing comfort experience for autonomous vehicle users. In *2019 IEEE intelligent transportation systems conference (ITSC)*, pages 2272–2277. IEEE, 2019.
- [23] Ronny Hug, Wolfgang Hübner, and Michael Arens. Introducing probabilistic bézier curves for n-step sequence prediction. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 10162–10169, 2020.
- [24] Yixin Gao, S Swaroop Vedula, Carol E Reiley, Narges Ahmidi, Balakrishnan Varadarajan, Henry C Lin, Lingling Tao, Luca Zappella, Benjamin Béjar, David D Yuh, et al. Jhu-isi gesture and skill assessment working set (jigsaws): A surgical activity dataset for human motion modeling. In *MICCAI workshop: M2cai*, volume 3, 2014.
- [25] Abed Soleymani, Xingyu Li, and Mahdi Tavakoli. A domain-adapted machine learning approach for visual evaluation and interpretation of robot-assisted surgery skills. *IEEE Robotics and Automation Letters*, 7(3):8202–8208, 2022.
- [26] Abed Soleymani, Xingyu Li, and Mahdi Tavakoli. Deep neural skill assessment and transfer: Application to robotic surgery training. In *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 8822–8829. IEEE, 2021.
- [27] Abed Soleymani, Ali Akbar Sadat Asl, Mojtaba Yeganejou, Scott Dick, Mahdi Tavakoli, and Xingyu Li. Surgical skill evaluation from robot-assisted surgery recordings. In *2021 International Symposium on Medical Robotics (ISMR)*, pages 1–6. IEEE, 2021.
- [28] Abed Soleymani, Mahdi Tavakoli, Farzad Aghazadeh, Yafei Ou, Hossein Rouhani, Bin Zheng, and Xingyu Li. Hands collaboration evaluation for surgical skills assessment: An information theoretical approach. *IEEE Transactions on Medical Robotics and Bionics*, 2024.
- [29] Abed Soleymani, Xingyu Li, and Mahdi Tavakoli. Artificial intelligence in robot-assisted surgery: Applications to surgical skills assessment and transfer. *Medical and Healthcare Robotics*, pages 183–200, 2023.
- [30] Ran Tao, Renz Ocampo, Jason Fong, Abed Soleymani, and Mahdi Tavakoli. Modeling and emulating a physiotherapist’s role in robot-assisted rehabilitation. *Advanced Intelligent Systems*, 2(7):1900181, 2020.
- [31] Mahdi Chalaki, Abed Soleymani, Xingyu Li, Vivian Mushahwar, and Mahdi Tavakoli. Evaluating gait symmetry with a smart robotic walker: A novel approach to mobility assessment. In *2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 4686–4692. IEEE, 2024.
- [32] Mahdi Chalaki, Amir Zakerimanesh, Abed Soleymani, Vivian Mushahwar, and Mahdi Tavakoli. Vision-based fuzzy control system for smart walkers: Enhancing usability for stroke survivors with unilateral upper limb impairments. *arXiv preprint arXiv:2502.00233*, 2025.
- [33] Bogdan Maris, Debora Botturi, and Paolo Fiorini. Trajectory planning with task constraints in densely filled environments. In *2010 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 2333–2338. IEEE, 2010.
- [34] Yafei Ou, Abed Soleymani, Xingyu Li, and Mahdi Tavakoli. Autonomous blood suction for robot-assisted surgery: A sim-to-real reinforcement learning approach. *IEEE Robotics and Automation Letters*, 2024.
- [35] Nils Rottmann, Ralf Bruder, Achim Schweikard, and Elmar Rueckert. Loop closure detection in closed environments. In *2019 European Conference on Mobile Robots (ECMR)*, pages 1–8. IEEE, 2019.
- [36] Cheng Chen, Yuqing He, Chunguang Bu, Jianda Han, and Xuebo Zhang. Quartic bézier curve based trajectory generation for autonomous vehicles with curvature and velocity constraints. In *2014 IEEE International Conference on Robotics and Automation (ICRA)*, pages 6108–6113. IEEE, 2014.
- [37] Marlene Pinzi, Stefano Galvan, and Ferdinando Rodriguez y Baena. The adaptive hermite fractal tree (ahft): a novel surgical 3d path planning approach with curvature and heading constraints. *International journal of computer assisted radiology and surgery*, 14:659–670, 2019.
- [38] Timothy N Judkins, Dmitry Oleynikov, and Nick Stergiou. Objective evaluation of expert and novice performance during robotic surgical training tasks. *Surgical endoscopy*, 23:590–597, 2009.