

111Equation Chapter 2 Section 1中图分类号:

TP242

**基于肌腱驱动的双足机器人跳跃运
动控制研究**

控制科学与工程

朱玉迪

李清都 教授

二〇二五年十二月

学校代码：10252
学 号：191550059

上海理工大学博士学位论文

基于肌腱驱动的双足机器人跳跃运动控制研究

姓 名	朱玉迪
系 别	光电信息与计算机工程学院
专 业	控制科学与工程
研究方向	模式识别与智能系统
指导教师	李清都 教授

学位论文完成日期 2025 年 12 月

RESEARCH ON JUMPING MOTION CONTROL OF TENDON-
DRIVEN BIPEDAL ROBOTS

by

Yudi Zhu

A Thesis Submitted to University of Shanghai for Science & Technology in

Partial Fulfillment of the Requirements for

the Degree of Doctor of Philosophy

Under the Supervision of

Professor Qingdu Li

University of Shanghai for Science & Technology

December 2025

学位论文版权使用授权书

本学位论文作者完全了解学校有关保留、使用学位论文的规定，同意学位论文保留并向国家有关部门或机构送交论文的复印件和电子版。允许论文被查阅和借阅。本人授权上海理工大学可以将本学位论文的全部内容或部分内容编入有关数据库进行检索，可以采用影印、缩印或扫描等复制手段保存和汇编本学位论文。

本学位论文属于 保 密 _ 年 ☐
 不保密 ☐

学位论文作者签名：

指导教师签名：

年 月 日

年 月 日

声 明

本人郑重声明：所呈交的学位论文，是本人在导师的指导下，独立进行研究工作所取得的成果。除文中已注明引用的内容外，本论文不包含任何其他个人或集体已经公开发表或撰写过的作品成果。对本文的研究做出重要贡献的个人和集体，均已在文中以明确方式标明。

本声明的法律责任由本人承担。

学位论文作者签名：

年 月 日

摘 要

关键词：双足机器人 控制策略

Abstract

Key Words: Bipedal robot, Control Strategy

目 录

中文摘要

ABSTRACT

第一章 绪论	1
1.1 研究背景与意义.....	1
1.2 腿臂协同控制的研究现状	2
1.2.1 基于零空间分解的腿臂协同控制	3
1.2.2 基于分层二次规划的腿臂协同控制	6
1.2.3 基于强化学习的腿臂协同控制.....	8
1.2.4 基于模仿学习的腿臂协同控制.....	11
1.3 本文研究内容及章节安排	14

第一章 绪论

1.1 研究背景与意义

近年来,人工智能技术在感知、认知与决策层面取得了突破性进展。现有系统已经能够在复杂场景中实现高精度的目标识别^[1],合成高分辨率、逼真的照片级图像^[2-3],根据自然语言描述自动生成具有实用价值的程序代码^[4-5],并在多种具有挑战性的游戏环境中超越人类顶尖水平^[6-8]。然而,与其在感知与认知能力上的快速发展相比,人工智能体在物理世界中的运动能力仍显著落后于人类及其他生物所展现出的灵活性与适应性。

人类之所以能够在复杂、动态且高度不确定的环境中完成多样而精细的运动行为,得益于人类长期进化形成的高度协调的全身运动控制机制,尤其体现在上下肢之间的协同配合与任务分工上。相比之下,当前的机器人系统,无论是在仿真环境还是现实部署中,往往只能执行少量预定义或高度专用的动作,其行为形式僵化、鲁棒性有限,难以应对真实环境中的复杂扰动和突发状况。这种差距在双足人形机器人上尤为突出。

具备类人运动能力的双足人形机器人,被普遍认为是未来服务机器人、特种机器人以及人机协作系统的重要发展方向。若能够显著提升其全身运动控制能力,使其在非结构化环境中稳定完成诸如坐下、起立、跌倒与爬起等高风险、高自由度动作,将有助于机器人从实验室和受控工业场景中走向更加真实、复杂的应用环境。同时,具有自然、协调运动模式的虚拟人和机器人智能体,也将在计算机图形学、虚拟现实、生物力学以及康复医学等领域展现重要价值。

围绕上述目标,研究者提出了多种控制器设计方法。传统的基于模型的手工设计控制策略在复现人类敏捷性方面已取得一定成果^[9-17]。然而,尽管人类能够熟练完成各类全身运动技能,其内在控制机理却难以被清晰刻画,更难以将其系统性地编码进控制器之中。

在机器人领域,手工设计控制器同样取得了一些令人信服的成果^[18-22]。但此类方法通常高度依赖专家经验与大量精细调参,所得到的控制器往往针对特定任务或技能进行定制,难以自然扩展到更大规模、更丰富的技能集合。当研究对象转向难度更高、专业性更强的行为(如坐起或爬起)时,由于接触关系更加复杂、系统动力学更难精确建模,控制器的设计与调试过程将显著复杂化。总体而言,尽管手工设计方法取得了诸多成功,其所能达到的运动表现仍与人类所展现出的丰富性、灵活性与优雅性存在明显差距。

为降低控制器设计的人工成本,基于优化的方法,如模型预测控制(MPC)^[23-29]与强化学习(RL)^[30-37],逐渐成为合成复杂运动技能的重要工具。这类方法通过优化目标函数来自动搜索控制策略,将设计者的主要工作从“如何控制”转变为“希望智能体表现出何种行为”。其中,强化学习在合成高自由度、强非线性系统的运动控制策略方面表现出强大潜力,已被成功应用于多种仿人运动技能的学习^[38-43]。

然而,基于优化的方法在实际应用中仍面临显著挑战。目标函数的设计往往需要融合大量领域知识,并对任务高度敏感。例如,仅为了诱导智能体产生自然的人类步态,就需要在能量效率、对称性、冲击力抑制以及稳定性等多个维度之间进行精细权衡^[25,44-48]。当任务扩展至坐下起立、跌倒爬起等更具挑战性的动作时,这种基于目标函数的设计方式不仅调参成本高,而且难以系统性地刻画人类动作中隐含的腿部与上肢的协同控制规律。这不禁引人思考:是否有一种方法能够根据任务来实现高效的腿臂协同控制。

1.2 腿臂协同控制的研究现状

人形机器人通常具有高自由度、多关节串并联混合结构,其运动控制涉及躯干、双臂与双腿等多个子系统的协同配合。在执行站立、行走与操作等任务过程中,机器人不仅需要满足系统动力学约束和接触约束,还需同时兼顾平衡维持、末端轨迹跟踪以及构形与能量优化等多重目标。因此,与传统机械臂或移动机器人控制问题不同,人形机器人的运动控制难以被分解为相互独立的子任务,而必须在统一框架下对全身多个关节和任务进行协调,这类方法通常被统称为全身控制(Whole-Body Control, WBC)。

在双足人形机器人中,腿部主要承担支撑与平衡维持的功能,而手臂除执行操作任务外,还可通过调节角动量、改变质心分布等方式辅助稳定性控制。尤其在外部扰动或复杂接触环境下,腿部与手臂之间往往存在显著的动力学耦合关系:腿部的支撑状态直接影响上肢可用的运动空间,而手臂的摆动和操作动作又会反作用于系统整体的稳定性。因此,如何在保证动态平衡与接触约束的前提下,实现腿臂之间的高效协同控制,是人形机器人全身控制研究中的关键问题之一。

从控制理论角度看,腿臂协同控制问题的本质可归结为多任务、多约束条件下的优先级分配与协调问题。在实际应用中,不同控制目标的重要性存在显著差异,例如保持接触约束、维持系统稳定性和避免关节超限通常具有更高优先级,而姿态优化、能量最小化或构形调整等目标则可在不破坏核心约束的前提下执行。围绕这一思想,研究者提出了以任务优先级为核心的全身控制范式,其关键在于确保高优先级任务得到满足,同时在其可行空间内协调低优先级任务。需要指出的

是，任务优先级并非某一类特定控制方法所独有，而是一种贯穿人形机器人全身控制设计的基础思想，其具体实现形式随控制框架的不同而有所差异。

1.2.1 基于零空间分解的腿臂协同控制

在模型驱动的全身控制方法中，任务优先级通常以显式形式加以刻画。早期研究多基于解析方法，通过雅可比矩阵投影或零空间分解的方式实现多任务协调，使低优先级任务在不干扰高优先级任务的前提下执行。这类方法具有清晰的物理含义和良好的可解释性，在人形机器人站立控制和简单运动生成中得到了广泛应用。然而，当任务数量增加或系统约束变得复杂时，基于解析投影的方法在数值稳定性和可扩展性方面逐渐暴露出局限。

1.2.1.1 方法起源

基于零空间的多任务优先级控制方法最早起源于冗余机械臂的逆运动学求解问题。在具有冗余自由度的机械臂系统中，当关节自由度高于任务空间维度时，系统存在多组满足主任务的解。为在完成主要任务的同时执行次级目标（如避障、关节限位规避或构形优化），研究者提出利用雅可比零空间结构进行分层控制。该思想在早期的冗余机器人控制研究中被系统化提出^[49-50]。

随后，任务优先级控制（Task Priority Control）框架逐渐形成，其核心思想是通过零空间投影保证高优先级任务的严格满足，同时在其可行空间内嵌入低优先级任务^[51]。这一方法后来被扩展至复杂多任务机器人系统，并在“Stack-of-Tasks”结构中得到进一步发展。

在人形机器人领域，由于系统自由度远高于单一任务维度，同时存在质心控制、接触保持、姿态稳定与末端操作等多种任务需求，零空间方法被自然引入全身控制问题，用以构建显式任务优先级结构^[52-53]。

1.2.1.2 数学定义

设机器人关节空间自由度为 $q \in \mathbb{R}^n$ ，任务空间向量为 $x \in \mathbb{R}^m$ ，满足运动学约束：

$$x = f(q), \quad (1.1)$$

对其求导可得关节速度与任务速度的映射关系：

$$\dot{x} = J(q)\dot{q}, \quad (1.2)$$

其中 $J \in \mathbb{R}^{m \times n}$ 为任务雅可比矩阵。当 $n > m$ 时，任务存在冗余自由度，此时雅可比矩阵的右伪逆 J^\dagger 可用于求解最小范数解：

$$\dot{q} = J^\dagger \dot{x}. \quad (1.3)$$

零空间 $\mathcal{N}(J)$ 定义为满足 $\dot{x} = 0$ 的关节速度集合，即：

$$\mathcal{N}(J) = \{\dot{q} \in \mathbb{R}^n \mid J\dot{q} = 0\}. \quad (1.4)$$

对应的零空间投影矩阵为：

$$N = I - J^\dagger J, \quad (1.5)$$

它可以将任意关节速度投影到任务零空间。

对于两个任务 x_1 和 x_2 ，其中 x_1 优先级高于 x_2 ，其关节速度解可迭代表示为：

$$\dot{q}_1 = J_1^\dagger \dot{x}_1 + N_1 \dot{q}_0, \quad (1.6)$$

$$\dot{q}_2 = J_2^\dagger \dot{x}_2 + N_2 \dot{q}_1, \quad (1.7)$$

其中 \dot{q}_0 为任意关节速度向量， N_1 为任务 1 的零空间投影矩阵。通过这种方式，低优先级任务可以在高优先级任务零空间内执行，从而保证高优先级任务不被破坏。对于 k 个任务的情况，可推广为迭代形式：

$$\dot{q} = \dot{q}_k = J_k^\dagger \dot{x}_k + N_k \dot{q}_{k-1}, \quad N_k = I - J_k^\dagger J_k. \quad (1.8)$$

为更直观地说明零空间递归投影在腿臂协同控制中的层级结构关系，图 1-1 给出了基于零空间分解的任务优先级全身控制框架示意图。

如图所示，该框架采用自上而下的层级结构。最上层为约束层（Constraint Layer），通常包括接触约束、质心控制或整体动量调节等稳定性相关任务。这一层对应公式中的最高优先级任务 x_1 ，其解通过雅可比伪逆 J_1^\dagger 求得，并生成零空间投影矩阵 N_1 。

在此基础上，任务层（Task Layer）中的手臂操作或轨迹跟踪任务被投影至约束层的零空间中执行，对应公式中的递归项 $N_1 J_2^\dagger \dot{x}_2$ 。这种结构确保操作任务不会破坏系统稳定性与接触一致性。

进一步地，姿态层（Posture Layer）利用剩余冗余自由度进行构形优化或冗余利用，其数学形式对应 $N_1 N_2 J_3^\dagger \dot{x}_3$ 。最终，各层任务的递归叠加形成关节空间控制指令（如关节速度、加速度或力矩命令）。

在腿臂协同控制场景下，该结构体现为：腿部相关任务（支撑与平衡）优先满

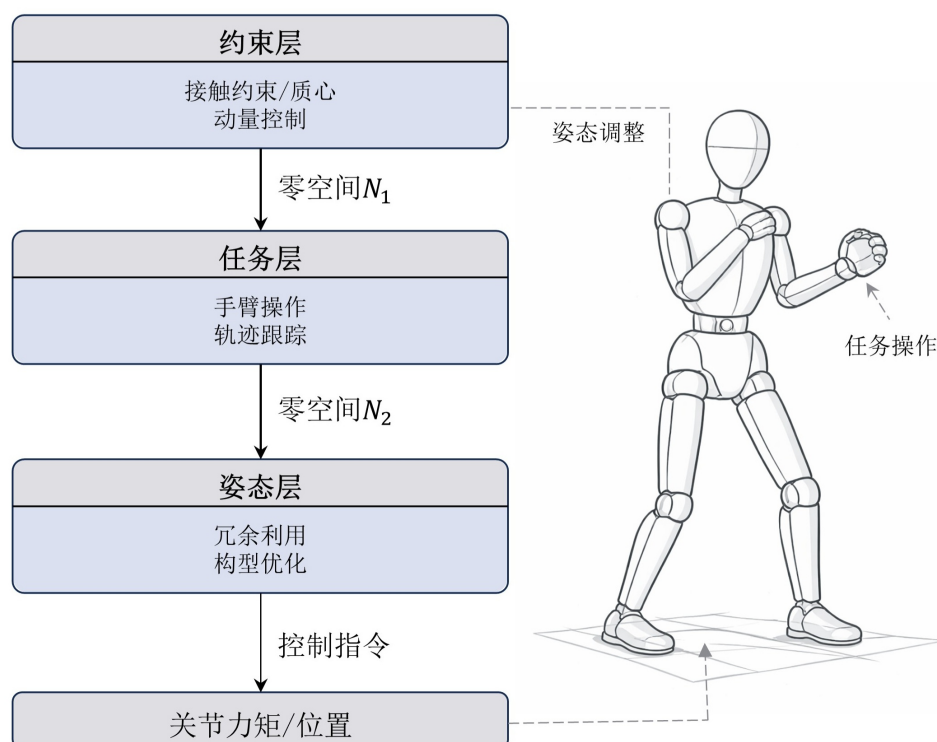


图 1-1 基于零空间分解的任务优先级全身控制框架示意图。高优先级约束任务首先被严格满足，低优先级任务依次投影至高优先级任务零空间中执行。

足；手臂任务嵌入稳定性零空间；姿态调节利用剩余冗余自由度优化整体构型。

因此，图 1-1 不仅直观展示了零空间分解的数学结构，也揭示了其在腿臂协同控制中的物理含义。

1.2.1.3 方法优势与局限性分析

零空间优先级控制方法被广泛应用于早期人形机器人全身控制系统中。Khatib 等人提出的 *Operational Space Control* 框架将任务空间动力学与关节空间控制统一建模，并为后续多任务扩展奠定理论基础^[52,54]。Sentis 与 Khatib 将该思想推广至人形机器人全身动力学控制中，实现了多任务分层控制结构^[53]。在上述工作基础上，基于零空间的任务优先级全身控制方法逐渐形成了一种相对固定的结构模式。在人形机器人腿臂协同控制场景中，该模式通常体现为“稳定性任务优先、操作任务嵌入零空间”的层级结构，并在多种平台上得到验证。为更加清晰地梳理该方法在腿臂协同控制中的典型结构特征、应用场景以及方法优势与局限性，本文将归纳整理如表 1-1 所示。

因此，尽管零空间方法为显式任务优先级建模提供了重要理论基础，但在复杂动态腿臂协同场景下，其扩展性与鲁棒性仍面临挑战。

表 1-1 零空间优先级全身控制在腿臂协同中的结构与特性分析

类别	内容	说明
优先级结构	CoM 与接触约束为最高优先级；躯干姿态为次级；手臂任务嵌入零空间；构形优化最低层级	典型严格优先级 Stack-of-Tasks 结构
典型应用	站立平衡、双臂操作、简单行走	HRP、Atlas 等平台验证 ^[55-56]
优势	严格保证高优先级任务；结构清晰；实时性强	适合早期人形机器人控制框架
局限	难处理不等式约束；数值稳定性问题；优先级固定	复杂动态场景扩展性有限

1.2.2 基于分层二次规划的腿臂协同控制

1.2.2.1 方法起源

随着人形机器人控制任务复杂度的不断提升，单纯基于零空间投影的解析式优先级控制方法逐渐难以满足多接触约束与不等式约束的处理需求。为克服零空间方法在数值稳定性和约束表达方面的局限，研究者开始将多任务控制问题转化为优化问题，通过二次规划（Quadratic Programming, QP）框架统一求解多任务与多约束问题^[57-58]。

分层二次规划（Hierarchical QP, HQP）方法在保持严格优先级结构的同时，能够自然引入不等式约束、摩擦锥约束与关节力矩限制等工程约束条件。该方法逐渐成为现代人形机器人全身控制的主流实现方式^[56,59]。

1.2.2.2 数学定义

在分层 QP 框架中，任务空间控制目标被转化为优化问题。对于单一任务，其运动学控制可表示为：

$$\min_{\dot{q}} \|J\dot{q} - \dot{x}_{\text{des}}\|^2, \quad (1.9)$$

其中 \dot{x}_{des} 为任务期望速度。引入松弛变量 ϵ 可处理不可行约束：

$$\min_{\dot{q}, \epsilon} \|J\dot{q} - \dot{x}_{\text{des}} + \epsilon\|^2, \quad \epsilon \geq 0. \quad (1.10)$$

对于多任务情况，可按照优先级依次求解：

$$\dot{q}_1^* = \arg \min_{\dot{q}, \epsilon_1} \|J_1\dot{q} - \dot{x}_1 + \epsilon_1\|^2, \quad (1.11)$$

$$\dot{q}_2^* = \arg \min_{\dot{q}, \epsilon_2} \|J_2\dot{q} - \dot{x}_2 + \epsilon_2\|^2, \quad \text{s.t. } J_1\dot{q} = \dot{x}_1^*, \quad (1.12)$$

该结构保证高优先级任务的最优解不被低优先级任务破坏。

进一步地，在全身动力学控制中，可将机器人动力学方程纳入约束条件：

$$M(q)\ddot{q} + h(q, \dot{q}) = S^T \tau + J_c^T f_c, \quad (1.13)$$

并将接触一致性与摩擦锥约束写为不等式条件：

$$f_c \in \mathcal{F}_{\text{friction}}. \quad (1.14)$$

此时，优化变量通常包括 \ddot{q} 、接触力 f_c 以及松弛变量，形成完整的分层逆动力学控制框架^[56,58]。

1.2.2.3 方法优势与局限性分析

综合近年来分层 QP 方法在人形机器人中的应用，可将其在腿臂协同控制中的典型结构与方法特性归纳如表 1-2 所示。

表 1-2 分层 QP 全身控制方法在腿臂协同中的结构与特性分析

类别	内容	说明
优先级结构	质心与动量控制、接触一致性为最高优先级；躯干姿态为次级；手臂末端操作嵌入较低层级；关节限位与力矩限制作为不等式约束	典型分层逆动力学（Hierarchical Inverse Dynamics）结构
典型应用	稳定行走、扰动恢复、双臂操作、多接触运动控制	Torque-controlled humanoid 平台验证 ^[56,59]
方法优势	自然处理等式与不等式约束；可纳入接触力与摩擦锥约束；数值稳定性高；易扩展至完整逆动力学与力控制框架	基于模型控制主流实现方式
方法局限	计算复杂度较高；权重与松弛变量依赖人工设计；优先级结构通常固定；接触切换场景可能出现层级冲突	高动态腿臂耦合场景下调参复杂

尽管全身控制方法通常依赖精确的系统模型和显式约束求解，其所蕴含的控制思想对于更一般的运动生成问题具有重要启示意义。特别地，任务分解、优先级建模以及在可行空间内协调多目标行为的理念，为理解和设计复杂全身运动策

略提供了一种清晰的结构化视角。该思想为后续引入数据驱动和学习方法奠定了重要的理论指导，使得学习到的控制策略能够在隐式层面体现类似的协调与优先级行为。

1.2.3 基于强化学习的腿臂协同控制

随着大规模并行仿真技术的发展，强化学习（Reinforcement Learning, RL）逐渐成为类人机器人全身控制的重要研究方向。不同于基于模型的解析优先级方法，强化学习通过与环境交互并最大化长期累积回报，在高维状态-动作空间中直接搜索控制策略，使复杂动力学耦合关系在训练过程中以“涌现”的形式形成。

基于强化学习的类人机器人控制最早主要集中于双足步态生成问题。针对 Cassie 等双足机器人，大量研究表明，仅通过精心设计的奖励函数，即可在无显式参考轨迹的情况下学习到站立、行走、跑步与跳跃等多种步态，并实现稳定的仿真到实机迁移^[60]。这一阶段的研究奠定了“基于奖励设计驱动步态生成”的技术路线，证明了在高度非线性动力学系统中，强化学习能够替代传统模型控制方法学习出稳定、鲁棒的运动行为。

随着类人机器人自由度的提升以及动力学仿真能力的增强，强化学习逐渐从“下肢 locomotion”扩展至“全身控制”。Gu 等提出的 Humanoid-Gym 框架基于 Isaac Gym 构建大规模并行仿真环境，采用 PPO 算法训练高自由度类人机器人行走策略，并在多种机器人平台上实现零样本迁移^[61]。该框架展示了强化学习在高维全身控制问题中的可扩展性与工程可行性。

在此基础上，研究者进一步探索强化学习在更复杂全身运动任务中的应用，逐渐形成了从单一 locomotion 技能向多步态统一控制、高动态全身动作以及复杂多接触交互任务扩展的发展趋势。Siekmann 等通过统一奖励设计实现多步态切换^[60]；Xie 等提出 PBHC 框架提升高动态动作稳定性^[62]；Zhang 等提出 WoCoCo 框架处理多接触序列任务^[63]；Fu 等提出 HumanPlus 系统实现真实机器人稳定部署^[64]。这些工作共同表明，强化学习能够在不显式构建腿臂协调规则的情况下，自主学习复杂全身运动模式。

1.2.3.1 基础原理

在强化学习框架下，类人机器人控制问题通常被建模为马尔可夫决策过程（MDP）：

$$\mathcal{M} = \langle \mathcal{S}, \mathcal{A}, \mathcal{T}, r, \gamma \rangle, \quad (1.15)$$

策略 $\pi_\theta(a_t|s_t)$ 通过最大化期望累积回报

$$J(\theta) = \mathbb{E}_{\pi_\theta} \left[\sum_{t=0}^{\infty} \gamma^t r_t \right] \quad (1.16)$$

进行优化。

图 1-2 所示为基于强化学习的腿臂协同控制框架。该框架可划分为“训练过程”和“执行过程”两个阶段。在训练阶段，机器人根据当前状态 s_t 通过策略 π_θ 产生动作 a_t ，环境反馈新的状态与奖励 r_t 。奖励函数通常由多个子项加权构成：

$$r_t = \sum_i w_i r_t^{(i)}, \quad (1.17)$$

其中不同奖励项分别刻画稳定性、步态性能、角动量调节、摆臂协调以及能耗等因素。策略优化模块通过最大化累计折扣回报 $J(\theta)$ 更新参数。在腿臂协同控制问题中，常见奖励构成包括：

- 质心速度或姿态跟踪奖励；
- 足端接触一致性或周期性奖励；
- 角动量抑制或躯干姿态稳定奖励；
- 手臂摆动对称性或相位协调奖励；
- 能耗与关节速度正则化项。

在执行阶段，训练得到的策略直接根据实时状态输出控制动作，实现闭环控制。

与零空间分解或分层 QP 方法不同，该框架中并不存在显式的“约束层-任务层-姿态层”优先级结构。腿部支撑稳定性、躯干姿态保持与手臂摆动协调并非通过解析投影或层级约束保证，而是通过奖励函数在优化过程中形成权衡关系。换言之，强化学习通过奖励加权结构在目标函数层面实现“隐式优先级”建模，使腿臂协同关系在训练过程中逐渐涌现。

1.2.3.2 方法优势与局限性分析

综合上述研究可以发现，强化学习在腿臂协同控制中的核心特征在于：通过奖励函数的加权组合，在统一优化目标下对稳定性、任务性能与能耗进行整体建模，从而在高维动力学系统中自动学习腿部与手臂之间的协调关系。

该范式在工程实践中展现出若干优势。首先，强化学习能够处理强非线性动力学与频繁接触切换问题，在多接触与高动态场景中具有较强适应能力^[63]。其次，

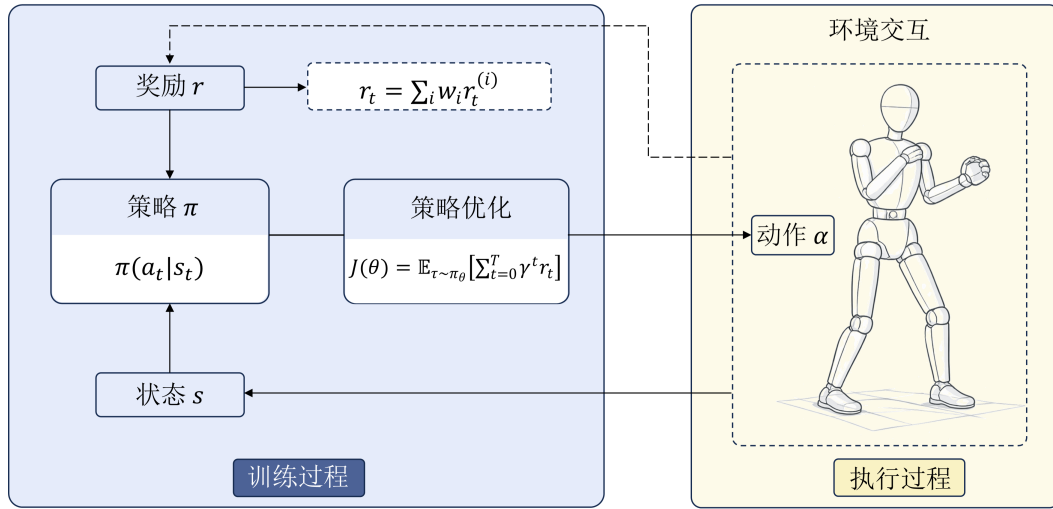


图 1-2 基于强化学习的腿臂协同控制框架示意图

通过大规模并行仿真与域随机化技术，策略可以获得较强的鲁棒性，并在不同机器人平台实现迁移^[61]。此外，通过奖励结构设计，可在单一策略中统一编码多步态与多技能行为，实现全身动作的一体化生成^[60]。

然而，强化学习方法同样存在局限。由于不同控制目标被压缩为奖励加权结构，其相对重要性依赖人工权重设计，缺乏显式优先级所具备的层级保障。当奖励项之间存在冲突时，策略可能出现不可预测行为。同时，强化学习训练通常需要大量仿真交互与计算资源，训练成本较高。此外，硬约束难以通过奖励形式严格保证，在安全性要求较高的实际场景中仍需结合模型方法进行补偿。

为更加系统地总结其技术特征，相关优势与局限性可归纳如表 1-3 所示。

表 1-3 基于强化学习的腿臂协同控制方法特征分析

类别	技术特征	说明与代表工作
方法优势	统一优化框架；可处理强非线性与复杂接触；适应多步态与多技能统一控制；协同关系可在单策略中隐式形成	多步态统一控制 ^[60] ；复杂接触任务 ^[63] ；高自由度平台验证 ^[61]
方法局限	高度依赖奖励设计与权重调节；训练成本高；难以显式保证硬约束；协同关系缺乏结构化解释与理论可证明性	奖励冲突可能导致不稳定；策略可解释性弱；安全约束难以严格满足

1.2.4 基于模仿学习的腿臂协同控制

随着大规模动作捕捉数据与深度生成模型的发展，模仿学习（Imitation Learning, IL）逐渐成为类人机器人全身控制的重要技术路线。与强化学习依赖人工构造奖励函数不同，模仿学习通过直接建模人类运动数据，使机器人策略能够继承人类动作的结构特征，从而在不显式构建协调规则的情况下形成自然的腿臂协同模式。

在人类运动中，腿部动作、躯干稳定与手臂摆动之间存在显著的动力学耦合关系。例如在摔倒后的起身过程中，人类通常会先利用手臂支撑地面以形成稳定支点，通过上肢产生反作用力抬升躯干，同时配合髋、膝关节伸展实现质心上移；在由坐姿或跪姿转为站立的过程中，手臂的前摆或撑地动作能够改变全身角动量分布，从而降低下肢负担并提升起身稳定性。此类动作中，手臂不仅承担支撑或辅助功能，还参与全身动量重分配与平衡调节，其运动与下肢发力之间存在紧密耦合关系。

这种腿臂协同关系天然存在于人类运动数据中，因此通过学习运动分布本身，策略可以在动作空间中隐式编码腿臂之间的协调结构，而无需显式构建协调规则或耦合约束。

1.2.4.1 基础原理

模仿学习的核心目标是使策略 $\pi_\theta(a|s)$ 逼近专家策略 $\pi_E(a|s)$ ，从而在状态空间中复现专家行为。根据建模方式的不同，现有方法大致可分为三类：基于监督回归的行为克隆方法、基于分布匹配的对抗式模仿学习方法，以及基于潜在空间建模的生成式运动先验方法。

1) 行为克隆 (Behavior Cloning, BC) 在监督学习框架下，策略通过最小化专家动作回归误差进行训练：

$$\mathcal{L}_{BC}(\theta) = \mathbb{E}_{(s,a^*) \sim \mathcal{D}} [\|\pi_\theta(s) - a^*\|^2], \quad (1.18)$$

其中 \mathcal{D} 为专家数据集， a^* 为专家动作。该方法实现简单，训练稳定，能够直接复制专家动作模式。然而，由于策略训练仅在专家数据分布上进行，当执行过程中状态分布发生偏移时，误差可能逐步累积，从而导致性能退化。

2) 对抗式模仿学习 (Adversarial Imitation Learning) 为缓解分布偏移问题，对抗式模仿学习通过匹配专家与策略生成轨迹的状态分布进行训练。Peng 等提出的

Adversarial Motion Priors (AMP) 方法^[65] 在此框架下引入运动判别器 D_ϕ ，用于区分专家运动与策略生成运动，并将判别结果转化为奖励信号：

$$r_{AMP}(s) = -\log(1 - D_\phi(s)). \quad (1.19)$$

判别器的优化目标为：

$$\mathcal{L}_D = \mathbb{E}_{s_E} [\log D_\phi(s_E)] + \mathbb{E}_{s_\pi} [\log(1 - D_\phi(s_\pi))], \quad (1.20)$$

其中 s_E 表示专家状态， s_π 表示策略生成状态。通过分布匹配机制，该方法能够在强化学习框架下继承人类运动统计特征，使腿臂协同结构在物理仿真环境中自然涌现。

3) 生成式运动先验 (Generative Motion Prior) 另一类方法通过生成模型学习低维运动潜在空间，从而对复杂动作分布进行结构化建模。例如，基于变分自编码器 (Variational Autoencoder, VAE) 的 Motion Variational Autoencoder (MVAE) 方法^[66] 将运动生成建模为：

$$p_\theta(x_{t+1} | x_t, z), \quad (1.21)$$

其中 z 为潜变量，控制运动演化方向。通过在潜在空间中进行策略优化，控制问题被转化为对结构化动作空间的搜索，从而在保持运动自然性的同时实现目标导向控制。由于人类腿臂协同关系已在潜空间中被隐式编码，控制过程中无需额外构造显式协调约束。

1.2.4.2 典型技术路线与方法特征分析

近年来，基于模仿学习的类人机器人全身控制逐渐形成三类主要技术路线，并在腿臂协同控制问题中展现出不同的技术特征与优势。

第一类为对抗式运动先验方法。AMP 框架^[65] 首次在物理仿真角色中验证了对抗式运动先验的有效性，通过判别器约束策略生成的状态分布，使机器人动作在统计意义上接近人类运动数据。随后，ASE 框架^[67] 将该思想扩展至多技能统一控制场景，使单一策略能够在共享潜在空间中编码多种行为模式。在腿臂协同控制中，该类方法的核心优势在于协同结构来源于真实运动分布，而非人工奖励构造，从而显著提升动作自然性。

第二类为多行为蒸馏与专家融合方法。Zhao 等提出 Adaptive Humanoid Control (AHC) 框架^[68]，采用多行为专家训练、策略蒸馏融合与强化微调相结合的方式，实现起身、行走等多技能统一控制。通过蒸馏机制，不同行为中的腿臂协同模式能够在共享策略中得到整合与保留，从而提高控制器的行为一致性与泛化能力。

第三类为生成式潜空间控制方法。基于变分自编码器 (Variational Autoencoder, VAE) 的 Motion Variational Autoencoder (MVAE) 框架^[66]通过学习低维运动潜空间，对复杂人类动作分布进行结构化建模。在此框架下，策略优化过程在受约束的潜在空间中进行，从而在保证运动风格一致性的同时提升训练稳定性。在复杂腿臂耦合动作中，该方法能够自然继承人类运动中的协调节律与动量分配特征。

综合上述技术路线可以发现，模仿学习方法在腿臂协同控制中的核心优势在于：通过数据驱动方式直接继承人类运动分布中的协同结构，使策略在生成动作时无需显式构造腿臂耦合规则即可获得自然协调模式。然而，该类方法同样面临数据依赖性强、泛化能力受限以及约束可控性不足等问题。

为更加系统地对比分析不同技术路线的特征与影响，本文将其优势与局限总结如表 1-4 所示。

表 1-4 基于模仿学习的腿臂协同控制方法特征分析

类别	技术特征	代表工作与影响
方法优势	继承人类运动分布中的腿臂协同结构，动作自然性高；提供结构化运动先验，提升高维训练稳定性；减少复杂奖励函数设计负担；易与强化学习结合形成混合训练范式	AMP ^[65] ；ASE ^[67] ；MVAE ^[66]
方法局限	高度依赖数据质量与覆盖范围；泛化能力受限于数据分布；难以显式保证动力学与接触约束；协同结构虽被学习但缺乏明确物理可解释性	多行为蒸馏框架 ^[68] 的泛化问题；复杂场景下的约束处理挑战

总体而言，模仿学习为类人机器人腿臂协同控制提供了强有力的数据驱动手段，在动作自然性与人类风格保持方面具有显著优势。相较于纯强化学习方法，其在结构继承与运动稳定性方面更具优势；然而在复杂环境泛化与显式约束建模方面仍存在不足。因此，在实际系统中，将模仿学习与强化学习或任务优先级控制方法进行融合，构建兼具自然性、稳定性与可解释性的混合控制框架，可能是未来的重要发展方向。

1.3 本文研究内容及章节安排

本文围绕人形机器人在复杂全身运动任务中的腿臂协同控制问题展开研究,重点探讨如何在人体与机器人存在显著结构与动力学差异的条件下,通过运动重定向与模仿强化学习方法,实现具有隐式任务优先级特性的全身控制策略。全文的研究内容主要分为三个层次:理论基础与问题建模、基于模仿学习的全身控制方法设计,以及仿真与真实机器人系统上的实验验证。

第2章介绍本文研究所涉及的理论基础与问题建模方法。首先,对强化学习的基本原理进行概述,包括值函数、策略表示、策略梯度方法以及训练过程,并进一步介绍基于目标的强化学习思想,为后续任务导向的控制策略设计奠定基础。随后,介绍与机器人运动控制相关的建模方法,包括机器人动力学模型、状态特征设计以及动作参数化形式。最后,对模仿学习的基本思想进行总结,为后续结合参考动作数据进行全身控制策略学习提供理论支撑。

第3章研究人形机器人全身动作重定向问题,重点解决人体与机器人在结构、自由度及运动约束方面存在差异所导致的动作失真问题。针对仅在关节空间或笛卡尔空间进行重定向各自存在的局限性,本章提出了一种关节空间与笛卡尔空间相融合的动作表示方式,在保证动作自然性与可行性的同时,提高末端执行精度与接触一致性。在此基础上,构建了融合多种约束条件的动作优化模型,并通过定量与定性分析对重定向动作的质量进行了系统评估。本章为后续模仿学习阶段提供了高质量、可控的参考动作。

第4章提出了一种基于模仿强化学习的全身控制框架,是全文的核心内容之一。本章首先介绍了整体学习流程,包括参考动作数据的重定向、状态与动作空间设计以及控制频率设置。随后,提出了一种极简但信息密度较高的奖励设计方法,仅依赖少量关键奖励项即可引导智能体学习稳定且自然的全身运动。针对复杂动作中不同阶段难度分布不均的问题,本章进一步提出了一种难点采样的参考动作调度策略,通过动态调整参考动作中关键阶段的采样概率,有效提升了训练稳定性与收敛速度,并通过对比实验进行了验证。

第5章围绕论文题目,系统分析了基于任务优先级的腿臂协同控制机制。首先对人形机器人在全身运动任务中的腿臂协同任务进行建模,明确腿部支撑、上肢支撑与平衡调节在不同动作阶段中的功能分工。随后,从学习结果的角度分析隐式任务优先级在策略中的体现方式,包括奖励函数中隐含的优先级关系以及动作阶段切换过程中的协调行为。通过对摔倒后爬起、坐下与站起等典型全身任务的分析,对比是否使用上肢参与支撑的控制策略,展示了所提出方法在复杂任务中实现自然腿臂协同控制的能力。

第 6 章对所提出方法在仿真环境中的实验结果进行了系统分析。首先介绍仿真平台、机器人模型及接触参数设置。随后，通过动作重定向实验与模仿学习对比实验，验证所提出方法在动作精度、稳定性与学习效率方面的优势。最后，通过消融实验分析关节一笛卡尔空间融合表示以及难点采样策略对整体性能的影响。

第 7 章进一步在真实双足人形机器人平台上对所提出方法进行实验验证。首先介绍实验硬件系统与控制架构，并说明学习到的策略在真实系统中的部署方式。随后，通过爬起、坐站等典型高风险全身动作实验，从成功率、鲁棒性及动作协调性等方面评估策略在真实环境中的表现，验证所提方法的实际可行性。

第 8 章对全文工作进行总结，并讨论了本文方法的局限性以及未来在复杂交互任务、自主技能扩展和人机协作等方向上的进一步研究展望。

参考文献

- [1] Russakovsky O, Deng J, Su H, et al. Imagenet large scale visual recognition challenge[J]. International journal of computer vision, 2015, 115(3): 211-252.
- [2] Brock A, Donahue J, Simonyan K. Large scale gan training for high fidelity natural image synthesis[J]. arXiv preprint arXiv:1809.11096, 2018.
- [3] Karras T, Laine S, Aila T. A style-based generator architecture for generative adversarial networks[C]. //Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2019: 4401-4410.
- [4] Brown T, Mann B, Ryder N, et al. Language models are few-shot learners[J]. Advances in neural information processing systems, 2020, 33: 1877-1901.
- [5] Chen M. Evaluating large language models trained on code[J]. arXiv preprint arXiv:2107.03374, 2021.
- [6] Berner C, Brockman G, Chan B, et al. Dota 2 with large scale deep reinforcement learning[J]. arXiv preprint arXiv:1912.06680, 2019.
- [7] Arulkumaran K, Cully A, Togelius J. Alphastar: an evolutionary computation perspective[C]. //Proceedings of the genetic and evolutionary computation conference companion. 2019: 314-315.
- [8] Silver D, Huang A, Maddison C J, et al. Mastering the game of go with deep neural networks and tree search[J]. nature, 2016, 529(7587): 484-489.
- [9] Raibert M H. Hopping in legged systems—modeling and simulation for the two-dimensional one-legged case[J]. IEEE Transactions on Systems, Man, and Cybernetics, 2012, (3): 451-463.
- [10] McGeer T. Passive bipedal running[J]. Proceedings of the Royal Society of London. B. Biological Sciences, 1990, 240(1297): 107-134.
- [11] Raibert M H, Hodgins J K. Animation of dynamic legged locomotion[C]. //Proceedings of the 18th annual conference on Computer graphics and interactive techniques. 1991: 349-358.
- [12] Schwind W J. Spring loaded inverted pendulum running: a plant model[B]. University of Michigan, 1998.

-
- [13] Geyer H, Seyfarth A, Blickhan R. Positive force feedback in bouncing gaits?[J]. Proceedings of the Royal Society of London. Series B: Biological Sciences, 2003, 270(1529): 2173-2183.
- [14] Vukobratović M, Borovac B. Zero-moment point—thirty five years of its life[J]. International journal of humanoid robotics, 2004, 1(01): 157-173.
- [15] Yin K, Loken K, Van de Panne M. Simbicon: simple biped locomotion control[J]. ACM Transactions on Graphics (TOG), 2007, 26(3): 105-es.
- [16] Da Silva M, Abe Y, Popović J. Simulation of human motion data using short-horizon model-predictive control[C]. //Computer Graphics Forum: vol. 27: 2. 2008: 371-380.
- [17] Bledt G, Powell M J, Katz B, et al. Mit cheetah 3: design and control of a robust, dynamic quadruped robot[C]. //2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). 2018: 2245-2252.
- [18] Raibert M H, Brown Jr H B, Chepponis M. Experiments in balance with a 3d one-legged hopping machine[J]. The International Journal of Robotics Research, 1984, 3(2): 75-92.
- [19] Miura H, Shimoyama I. Dynamic walk of a biped[J]. The International Journal of Robotics Research, 1984, 3(2): 60-74.
- [20] Hodgins J K, Wooten W L, Brogan D C, et al. Animating human athletics[C]. //Proceedings of the 22nd annual conference on Computer graphics and interactive techniques. 1995: 71-78.
- [21] Coros S, Beaudoin P, Van de Panne M. Generalized biped walking control[J]. ACM Transactions On Graphics (TOG), 2010, 29(4): 1-9.
- [22] Hutter M, Gehring C, Jud D, et al. Anymal-a highly mobile and dynamic quadrupedal robot[C]. //2016 IEEE/RSJ international conference on intelligent robots and systems (IROS). 2016: 38-44.
- [23] De Lasa M, Mordatch I, Hertzmann A. Feature-based locomotion controllers[J]. ACM transactions on graphics (TOG), 2010, 29(4): 1-10.
- [24] Sreenath K, Park H W, Poulakakis I, et al. A compliant hybrid zero dynamics controller for stable, efficient and fast bipedal walking on mabel[J]. The International Journal of Robotics Research, 2011, 30(9): 1170-1193.

- [25] Mordatch I, Todorov E, Popović Z. Discovery of complex behaviors through contact-invariant optimization[J]. *ACM Transactions on Graphics (ToG)*, 2012, 31(4): 1-8.
- [26] Al Borno M, De Lasa M, Hertzmann A. Trajectory optimization for full-body movements with complex contacts[J]. *IEEE transactions on visualization and computer graphics*, 2012, 19(8): 1405-1414.
- [27] Gehring C, Coros S, Hutter M, et al. Practice makes perfect: an optimization-based approach to controlling agile motions for a quadruped robot[J]. *IEEE Robotics & Automation Magazine*, 2016, 23(1): 34-43.
- [28] Dogar M, Srinivasa S. A framework for push-grasping in clutter[J]. *Robotics: Science and systems VII*, 2011, 1: 65-72.
- [29] Apgar T, Clary P, Green K, et al. Fast online trajectory optimization for the bipedal robot cassie.[C]. //Robotics: Science and Systems: vol. 101. 2018: 14.
- [30] Van de Panne M, Kim R, Fiume E. Virtual wind-up toys for animation[C]. //Graphics Interface. 1994: 208-208.
- [31] Kohl N, Stone P. Policy gradient reinforcement learning for fast quadrupedal locomotion[C]. //IEEE International Conference on Robotics and Automation, 2004. Proceedings. ICRA'04. 2004: vol. 3. 2004: 2619-2624.
- [32] Tedrake R, Zhang T W, Seung H S. Stochastic policy gradient reinforcement learning on a simple 3d biped[C]. //2004 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)(IEEE Cat. No. 04CH37566): vol. 3. 2004: 2849-2854.
- [33] Endo G, Morimoto J, Matsubara T, et al. Learning cpg sensory feedback with policy gradient for biped locomotion for a full-body humanoid[C]. //AAAI. 2005: 1267-1273.
- [34] Coros S, Beaudoin P, Van de Panne M. Robust task-based control policies for physics-based characters[B]. //ACM SIGGRAPH Asia 2009 papers. 2009: 1-9.
- [35] Tan J, Zhang T, Coumans E, et al. Sim-to-real: learning agile locomotion for quadruped robots[J]. *arXiv preprint arXiv:1804.10332*, 2018.
- [36] Haarnoja T, Ha S, Zhou A, et al. Learning to walk via deep reinforcement learning[J]. *arXiv preprint arXiv:1812.11103*, 2018.

-
- [37] Hwangbo J, Lee J, Dosovitskiy A, et al. Learning agile and dynamic motor skills for legged robots[J]. *Science Robotics*, 2019, 4(26): eaau5872.
 - [38] Wang J M, Hamner S R, Delp S L, et al. Optimizing locomotion controllers using biologically-based actuators and objectives[J]. *ACM Transactions on Graphics (TOG)*, 2012, 31(4): 1-11.
 - [39] Tan J, Gu Y, Liu C K, et al. Learning bicycle stunts[J]. *ACM Transactions on Graphics (TOG)*, 2014, 33(4): 1-12.
 - [40] Levine S, Finn C, Darrell T, et al. End-to-end training of deep visuomotor policies[J]. *Journal of Machine Learning Research*, 2016, 17(39): 1-40.
 - [41] Peng X B, Berseth G, Van de Panne M. Terrain-adaptive locomotion skills using deep reinforcement learning[J]. *ACM Transactions on Graphics (TOG)*, 2016, 35(4): 1-12.
 - [42] Gu S, Holly E, Lillicrap T, et al. Deep reinforcement learning for robotic manipulation with asynchronous off-policy updates[C]. //2017 IEEE international conference on robotics and automation (ICRA). 2017: 3389-3396.
 - [43] Andrychowicz O M, Baker B, Chociej M, et al. Learning dexterous in-hand manipulation[J]. *The International Journal of Robotics Research*, 2020, 39(1): 3-20.
 - [44] Wang J M, Fleet D J, Hertzmann A. Optimizing walking controllers[B]. //ACM SIGGRAPH Asia 2009 papers. 2009: 1-8.
 - [45] Geijtenbeek T, Van De Panne M, Van Der Stappen A F. Flexible muscle-based locomotion for bipedal creatures[J]. *ACM Transactions on Graphics (TOG)*, 2013, 32(6): 1-11.
 - [46] Clegg A, Yu W, Tan J, et al. Learning to dress: synthesizing human dressing motion via deep reinforcement learning[J]. *ACM Transactions on Graphics (TOG)*, 2018, 37(6): 1-10.
 - [47] Yu W, Turk G, Liu C K. Learning symmetry and low-energy locomotion[J]. *ArXiv e-prints*.
 - [48] Abdolhosseini F, Ling H Y, Xie Z, et al. On learning symmetric locomotion[C]. //Proceedings of the 12th ACM SIGGRAPH Conference on Motion, Interaction and Games. 2019: 1-10.

- [49] Liegeois A. Automatic supervisory control of the configuration and behavior of multibody mechanisms[J]. IEEE Transactions on Systems, Man, and Cybernetics, 1977, 7(12): 868-871.
- [50] Nakamura Y. Advanced robotics: redundancy and optimization[B]. Addison-Wesley, 1987.
- [51] Siciliano B, Slotine J J E. A general framework for managing multiple tasks in highly redundant robotic systems[C]. //Proceedings of the 5th International Conference on Advanced Robotics. 1991: 1211-1216.
- [52] Khatib O, Sentis L, Park J Y, et al. Whole-body dynamic behavior and control of human-like robots[J]. International Journal of Humanoid Robotics, 2004, 1(1): 29-43.
- [53] Sentis L. Synthesis and control of whole-body behaviors in humanoid systems[D]. Stanford University, 2007.
- [54] Khatib O. A unified approach for motion and force control of robot manipulators: the operational space formulation[J]. IEEE Journal on Robotics and Automation, 1987, 3(1): 43-53.
- [55] Ott C, Dietrich A, Albu-Schäffer A. Prioritized multi-task compliance control of redundant manipulators[J]. Automatica, 2015, 53: 416-423.
- [56] Herzog A, Rotella N, Mason S, et al. Momentum control with hierarchical inverse dynamics on a torque-controlled humanoid[J]. Autonomous Robots, 2016, 40(3): 473-491.
- [57] Mansard N, Khatib O, Kheddar A. A unified approach to integrate unilateral constraints in the stack of tasks[J]. IEEE Transactions on Robotics, 2009, 25(3): 670-685.
- [58] Escande A, Mansard N, Wieber P B. Hierarchical quadratic programming: fast on-line humanoid-robot motion generation[J]. The International Journal of Robotics Research, 2014, 33(7): 1006-1028.
- [59] Righetti L, Buchli J, Mistry M, et al. Inverse dynamics control of floating-base robots with external constraints: a unified view[C]. //Proceedings of the IEEE International Conference on Robotics and Automation (ICRA). 2011: 1085-1090.

- [60] Siekmann J, et al. Learning diverse bipedal gaits from scratch[J]. Science Robotics, 2021.
- [61] Gu X, et al. Humanoid-gym: reinforcement learning for humanoid robots[J], 2024.
- [62] Xie Z, et al. Pbhc: physics-based humanoid control[J], 2025.
- [63] Zhang Y, et al. Wococo: whole-body contact control via reinforcement learning[J], 2024.
- [64] Fu Y, et al. Humanplus: humanoid whole-body control via reinforcement learning[J], 2024.
- [65] Peng X B, Kumar A, Zhang G, et al. Amp: adversarial motion priors for stylized physics-based character control[J]. ACM Transactions on Graphics (SIGGRAPH), 2021, 40(4): 1-15.
- [66] Ling H Y, Zinno F, Cheng G, et al. Character controllers using motion vaes[J]. ACM Transactions on Graphics, 2020, 39(4): 40:1-40:12.
- [67] Peng X B, Zhang Z, Yu W, et al. Ase: large-scale reusable adversarial skills[J]. ACM Transactions on Graphics (SIGGRAPH), 2022, 41(4): 1-15.
- [68] Zhao Y, Wang X, Wang D, et al. Towards adaptive humanoid control via multi-behavior distillation and reinforced fine-tuning[J]. arXiv preprint arXiv:2511.06371, 2025.