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

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综述

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

在模型驱动的人形机器人全身控制方法中，任务优先级通常以显式层级结构进行刻画。早期研究主要基于解析方法，通过雅可比矩阵投影或零空间分解实现多任务协调，使低优先级任务在不干扰高优先级任务的前提下执行。这类方法结构清晰、物理意义明确，在人形机器人站立控制、姿态调节以及简单运动生成中得到了广泛应用。

基于零空间的多任务优先级控制思想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，系统存在无限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矩阵的零空间结构进行分层控制。Liegeois[49] 首次系统性提出在主任务解的零空间中嵌入次级优化目标；随后 Nakamura[50] 对冗余系统的运动学分解与优先级结构进行了系统化分析，为多任务控制理论奠定了基础。

在此基础上，Siciliano 与 Slotine[51] 提出了较为完整的任务优先级控制框架，通过零空间递归投影实现多任务分层控制，该思想随后在“Stack-of-Tasks”结构中得到形式化表达，并逐渐推广至复杂机器人系统。随着研究对象从单臂系统扩展至高自由度的人形机器人，零空间优先级思想被自然引入全身控制框架。Khatib 提出的操作空间控制理论（Operational Space Control）[52-53] 将任务空间动力学与关节空间控制统一建模；Sentis 等进一步在动力学一致性框架下推广该思想[54]，实现了多接触与多任务分层控制的系统化建模。由此，零空间分解逐渐成为人形机器人腿臂协同控制的基础理论框架。

其核心思想可概括为：通过零空间递归投影构建严格的任务层级关系。设机器人关节向量为 $q \in \mathbb{R}^n$，任务空间变量为 $x \in \mathbb{R}^m$，其速度映射关系为

$$\dot{x} = J(q)\dot{q}, \tag{1.1}$$

其中 $J(q)$ 为任务雅可比矩阵。当系统存在冗余（$n > m$）时，关节速度的一般解可表示为：

$$\dot{q} = J^{\dagger}\dot{x} + (I - J^{\dagger}J)\dot{q}_0, \tag{1.2}$$

其中 $J^{\dagger}$ 为 Moore–Penrose 伪逆，$(I - J^{\dagger}J)$ 为零空间投影矩阵，$\dot{q}_0$ 为任意零空间速度。

对于多任务系统，若任务 x_1 优先级高于 x_2，则其递归形式可表示为：

$$\dot{q} = J_1^{\dagger}\dot{x}_1 + N_1 J_2^{\dagger}\dot{x}_2 + N_1 N_2 \dot{q}_0, \tag{1.3}$$

其中 $N_i = I - J_i^{\dagger}J_i$ 为第 i 个任务的零空间投影矩阵。

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。

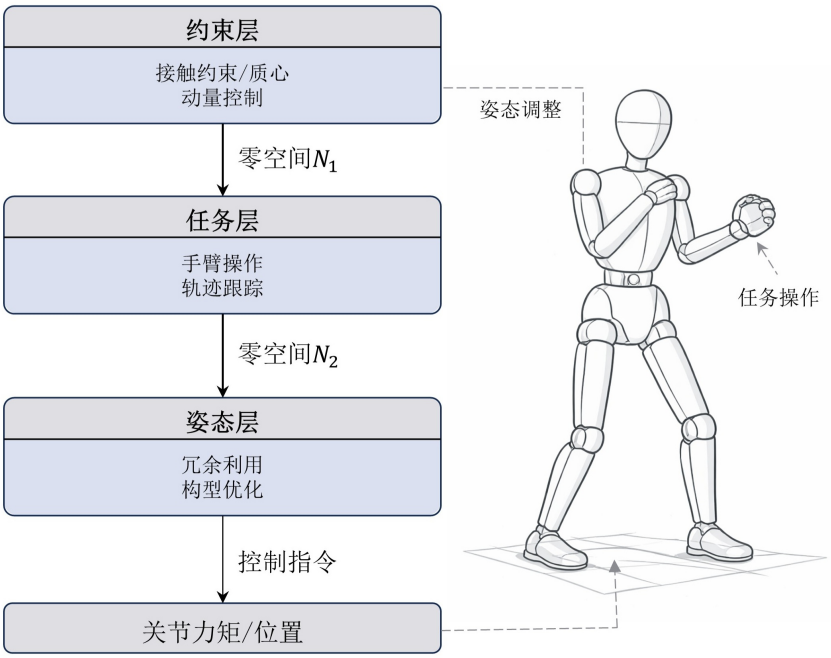

图 1-1　基于零空间分解的任务优先级全身控制框架示意图。

如图所示 1-1，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）。

该算法的核心思想在于：通过零空间递归投影建立严格的任务层级关系——高优先级任务被优先满足，低优先级任务仅在其零空间内执行，而剩余冗余自由度用于姿态优化或次级性能提升。

在腿臂协同控制场景 1-1中，这种层级结构通常体现为：支撑与质心稳定任务处于最高优先级，手臂操作任务嵌入稳定性零空间中执行，而构形与姿态调节则利用剩余冗余自由度进行优化。由此实现“稳定优先、操作嵌入、姿态优化”的分层协同控制机制。

从方法特性来看，基于零空间分解的任务优先级控制方法在腿臂协同问题中具有显著优势。首先，该方法能够通过严格的零空间投影机制保证高优先级任务的精确满足，使支撑稳定性、质心控制等关键任务在数学上得到明确保护。这种严格的层级结构具有清晰的物理含义，使不同任务之间的干扰关系可以通过投影矩阵形式直观刻画，因而具有良好的可解释性与结构透明性。其次，零空间方法通常基于解析形式推导，其计算结构相对紧凑，实时性较好，适用于早期计算资源受限的人形机器人控制系统。

然而，随着控制任务复杂度的提升，该方法的局限性逐渐显现。其一，零空间投影本质上基于等式约束构建，对于摩擦锥、关节力矩限制、接触不等式等物理约束的处理能力有限，往往需要额外近似或线性化处理。其二，在多接触与强非线性动态场景下，连续零空间递归可能导致数值不稳定或解空间退化问题。其三，任务优先级通常在设计阶段固定，一旦高优先级任务过于严格，低优先级任务的可行空间可能被过度压缩，从而限制手臂与腿部之间的协调自由度。在高速动态运动或频繁接触切换过程中，这种刚性层级结构可能难以兼顾稳定性与灵活性。

因此，尽管零空间方法为多任务优先级控制提供了重要理论基础，并在人形机器人早期全身控制中发挥了关键作用，但在复杂动态腿臂协同场景下，其对不等式约束的处理能力、数值鲁棒性以及结构自适应能力仍存在一定局限。这也促使后续研究逐步向基于优化的分层控制框架发展。

1.2.2　基于分层二次规划的腿臂协同控制

随着人形机器人控制任务复杂度的不断提升，单纯依赖零空间投影的解析式优先级控制方法逐渐暴露出其在不等式约束表达、多接触建模以及数值稳定性方面的局限。为克服上述问题，研究者开始将多任务控制问题统一转化为带约束的优化问题，通过二次规划（Quadratic Programming, QP）框架在满足系统动力学方程的同时处理多种等式与不等式约束。

早期工作中，Mansard 等[55] 提出了将任务优先级嵌入二次规划结构的方法，通

过构造分层优化问题实现严格优先级的保持，为多任务控制问题提供了统一的优化表达形式。随后，Escande 等[56] 系统化提出分层二次规划（Hierarchical Quadratic Programming, HQP）框架，通过递归构建一系列带约束的二次规划子问题，实现任务间的严格优先级分离，同时保证数值稳定性和计算可行性。这一工作在理论层面上将“Stack-of-Tasks”结构推广至优化框架，并为复杂约束处理提供了严谨的数学基础。在动力学层面，Righetti 等[57] 将逆动力学控制问题表述为二次规划问题，将关节加速度、力矩以及接触力作为决策变量，从而在满足动力学一致性的前提下显式引入摩擦锥约束和力矩限制，为全身动力学控制提供了可扩展的优化求解结构。Herzog 等[58] 在此基础上提出基于逆动力学的实时 QP 控制框架，强调在浮动基系统中直接将接触力作为优化变量处理，从而避免显式求解约束动力学逆问题，提高了实时性与数值鲁棒性。

在人形机器人系统中，HQP 方法逐渐成为主流全身控制实现方式。Kim 等**Kim2019** 提出 Whole-Body Locomotion Controller（WBLC），采用分层优化结构统一处理质心控制、摆动足轨迹与接触约束，并通过软约束与权重调节机制实现平滑接触切换，为动态步行提供了实验验证。近年来，Paredes 等**Paredes2023** 在逆动力学 QP 框架中引入加速度形式的指数控制屏障函数（Acceleration-based Exponential Control Barrier Function, A-ECBF），在保持任务性能的同时显式保证安全约束不被破坏，进一步拓展了 HQP 在安全全身控制中的应用范围。

总体而言，分层二次规划方法在理论上继承了任务优先级控制的层级思想，在工程实现上则通过优化框架自然融合动力学约束、接触条件与不等式限制，使其成为现代人形机器人腿臂协同控制的重要技术路径。

在分层二次规划（HQP）框架中，多任务控制被统一表示为带约束的优化问题。对于单一任务，其基本形式为：

$$\min_{z} \|Az - b\|^2, \tag{1.4}$$

其中 z 为优化变量（如 $\dot{q}$ 或 $\ddot{q}$），A 为任务雅可比矩阵，b 为期望任务量。

对于分层结构，设第 k 层任务为 (A_k, b_k)，则 HQP 递归形式为：

$$z_k^* = \arg\min_{z} \|A_k z - b_k\|^2, \quad \text{s.t. } z \in \mathcal{S}_{k-1}, \tag{1.5}$$

其中 $\mathcal{S}_{k-1}$ 表示满足第 1 至 $k-1$ 层任务最优解的可行解集合。该结构保证高优先级任务严格保持，而低优先级任务仅在其零干扰空间内优化。

在全身控制中，系统动力学方程作为等式约束引入：

$$M(q)\ddot{q} + h(q,\dot{q}) = S^{\mathsf{T}}\tau + J_c^{\mathsf{T}} f_c, \tag{1.6}$$

并可进一步加入接触与摩擦等不等式约束，从而形成完整的分层逆动力学优化框架。

综合近年来分层二次规划方法在人形机器人中的发展与应用，可以看到其在腿臂协同控制问题中形成了一种较为成熟的结构范式。该类方法通常以质心或动量控制、接触一致性作为最高优先级任务，在此基础上逐层嵌入躯干姿态调节与手臂末端操作任务，并通过不等式约束显式处理关节限位与力矩限制。这种分层逆动力学结构在多个扭矩控制人形机器人平台上得到验证，成为当前模型驱动全身控制的主流实现方式。

从方法特性来看，基于 HQP 的全身控制框架具有若干显著优势。首先，其优化结构能够在统一框架内同时处理等式与不等式约束，使接触力、摩擦锥、零力矩点（ZMP）以及关节力矩限制等物理约束得到显式表达，从而提高系统的安全性与物理一致性。其次，严格的层级优化机制保证高优先级稳定性任务不被低优先级操作任务破坏，使腿部支撑与手臂操作之间形成清晰的协调关系。此外，逆动力学形式下的优化变量通常包含关节加速度与接触力，使力控制与运动控制能够在同一求解框架中实现统一协调，这对于多接触、高冗余的人形系统尤为重要。

然而，该方法同样存在一定局限性。其一，HQP 框架高度依赖精确的动力学模型与接触建模，当模型误差、结构柔性或环境不确定性较大时，控制性能可能显著下降。其二，多层优化问题的实时求解带来较高的计算开销，随着自由度与约束数量增加，计算复杂度迅速提升，对控制频率与硬件算力提出较高要求。其三，权重矩阵与松弛变量的设计往往依赖人工经验，不同任务组合与场景下需要反复调节参数，增加了工程实现复杂度。其四，优先级结构通常在设计阶段固定，缺乏在线自适应调整能力，在高度动态或接触频繁切换的场景中可能出现层级冲突或解空间收缩现象。尤其在腿臂强耦合的高速运动过程中，固定层级结构可能限制全身协调的灵活性。

因此，尽管 HQP 方法在结构化建模与物理约束表达方面具有显著优势，是当前模型驱动全身控制的重要技术路径，但其在模型依赖性、参数调节复杂性以及泛化能力方面仍存在改进空间。这些局限性也为后续结合数据驱动与学习方法的研究提供了重要动机，使控制策略能够在保持物理一致性的同时具备更强的自适应与泛化能力。

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[59]。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行走扩展至全身控制。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[60]。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，逐渐形成了从单一行走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[59] 提出统一奖励结构下的多步态学习方法，通过在同一策略中嵌入速度指令与步态调度变量，使机器人在无显式轨迹参考的情况下实现站立、行走、跑步与跳跃等多种步态的连续切换。该工作证明了强化学习策略能够在单一网络结构中内化多种动态模式，实现跨步态的统一控制。在高动态动作学习方面，Xie 等提出 PBHC（Phase-Based Hybrid Control）框架[61]，通过在强化学习中引入相位变量与混合控制结构，增强策略对高速摆动与冲击接触的建模能力，从而提升翻滚、腾空等高动态全身动作的稳定性。该方法在保留学习灵活性的同时，引入结构化先验以缓解训练难度，体现出“结构约束 + 数据驱动”结合的发展趋势。针对复杂多接触任务，Zhang 等提出 WoCoCo（Whole-Body Contact Control）框架[62]，通过在强化学习中显式引入接触模式编码与任务阶段划分，使策略能够在多接触序列（如攀爬、支撑转换等）中学习稳定的接触力分配与全身协调机制。该工作拓展了强化学习在多接触操作与复杂地形交互中的应用范围。Fu 等提出 HumanPlus 系统[63]，通过域随机化、残差学习与控制器融合机制，实现高自由度类人机器人在真实环境中的稳定全身控制。该系统验证了强化学习策略在复杂机械结构与实际传感噪声条件下的可迁移性与工程可行性。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。

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（Markov Decision Process, MDP）：

$$\mathcal{M} = \langle \mathcal{S}, \mathcal{A}, \mathcal{T}, r, \gamma \rangle, \tag{1.7}$$

其中，$\mathcal{S}$ 为状态空间，$\mathcal{A}$ 为动作空间，$\mathcal{T}$ 为状态转移概率函数，r 为即时奖励函数，$\gamma \in (0,1)$ 为折扣因子。

在离散时间步 t，机器人处于状态 $s_t \in \mathcal{S}$，根据策略 $\pi_\theta(a_t \mid s_t)$ 选择动作 $a_t \in \mathcal{A}$，环境根据转移概率 $\mathcal{T}$ 演化至下一个状态 s_{t+1}，并反馈奖励 $r_t = r(s_t, a_t)$。策略参数 θ 通过最大化期望累积折扣回报进行优化：

$$J(\theta) = \mathbb{E}_{\pi_\theta}\left[\sum_{t=0}^{\infty} \gamma^t r_t\right], \tag{1.8}$$

其中期望是对策略 π_θ 下产生的轨迹分布进行统计得到的。

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组合而成：

$$r_t = \sum_i w_i \, r_t^{(i)}, \tag{1.9}$$

其中 $r_t^{(i)}$ 表示第 i 个子奖励项，用于刻画特定控制目标（如速度跟踪、姿态稳定、能量消耗或接触约束等），$w_i \in \mathbb{R}$ 为对应的权重系数，用于调节各子目标在总体奖励中的相对重要性。权重 w_i 通常由人工设计或通过经验调节确定。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。

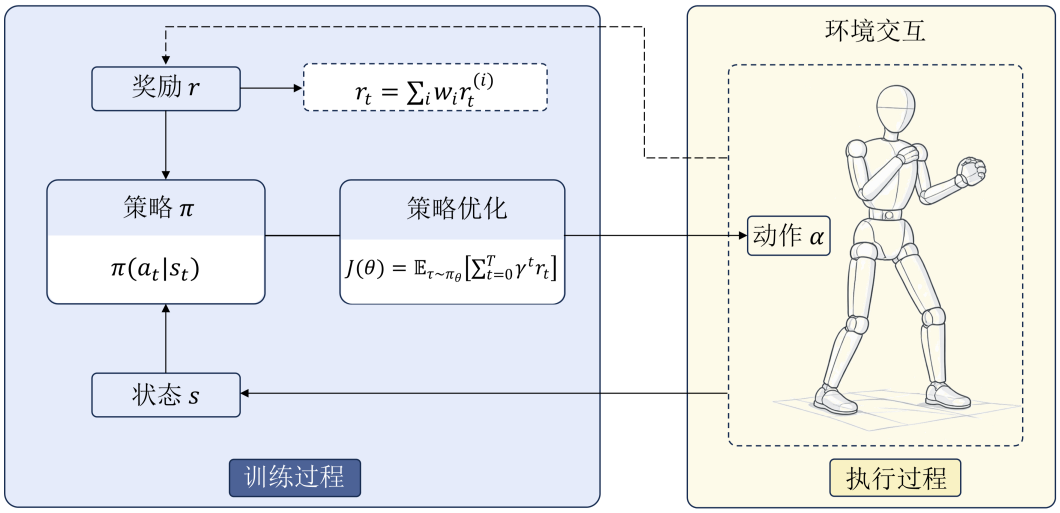

图 1-2　基于强化学习的腿臂协同控制框架示意图

综合现有研究可以发现，基于强化学习的腿臂协同控制方法在方法范式上呈现出明显区别于解析优先级控制框架的技术特征。其核心思想在于通过奖励函数的加权组合，在统一的优化目标下对稳定性、任务性能与能耗等因素进行整体建模，使不同控制目标在同一策略参数空间内协同优化。这种“隐式优先级”机制并非通过显式层级结构实现，而是通过长期回报最大化过程在高维动力学系统中自动形成腿部支撑与手臂动作之间的协调关系。

从工程实践角度看，该范式具有若干显著优势。首先，强化学习能够直接在高维、强非线性的动力学系统中搜索控制策略，对复杂接触序列与频繁支撑切换具有较强适应能力。在多接触或高动态动作场景中，策略可以通过大量仿真交互学习到稳定的全身协调模式，而无需显式推导接触力分配规则。其次，依托大规模并行仿真与域随机化技术，策略在训练阶段可暴露于多种扰动与参数变化，从而获得一定程度的鲁棒性与跨平台迁移能力。此外，通过奖励结构设计与条件变量引入，单一策略网络可以统一编码多步态与多技能行为，使全身动作生成在形式上实现一体化。

然而，强化学习方法同样存在不可忽视的局限性。其一，不同控制目标被压缩为奖励加权结构后，其相对重要性依赖人工权重设计，缺乏类似零空间或 HQP 方法所具有的严格层级保障。当奖励项之间存在潜在冲突时，策略可能在训练过程中形成不可预测的折中行为。其二，强化学习训练通常需要大量仿真交互与计算资源，尤其在高自由度类人系统中，训练成本与调参复杂度较高。其三，硬约束（如接触力上界、关节力矩限制或安全边界）难以通过纯奖励形式严格保证，在安

全性要求较高的实际部署场景中，往往仍需结合模型约束或安全过滤机制进行补偿。其四，由于协同关系以隐式形式存在于策略参数中，其结构解释性与理论可证明性相对较弱，难以像解析方法那样给出明确的优先级与稳定性保证。

因此，基于强化学习的腿臂协同控制方法在复杂动力学建模能力与自适应性方面具有显著优势，但在约束可控性、训练效率以及结构可解释性方面仍存在一定挑战。这些特性也推动近年来研究逐渐探索模型驱动方法与数据驱动方法的融合框架，以期在稳定性保障与自适应能力之间取得平衡。

1.2.4　基于模仿学习的腿臂协同控制

模仿学习（Imitation Learning, IL）的思想最早来源于机器人学习与控制中的示教学习问题[64]。其核心目标是在不给定明确代价函数或系统模型的情况下，通过专家示范数据直接学习控制策略。早期方法主要基于监督学习框架，通过回归或分类模型逼近专家策略，即行为克隆[65-66]。在该阶段，模仿学习主要应用于机械臂轨迹复现与简单运动控制问题，其优势在于实现简单、训练稳定，但在长时序控制与状态分布偏移问题上存在明显局限。

随着深度学习的发展以及大规模动作数据的获取，模仿学习逐渐从“轨迹复制”扩展至“运动分布建模”。Ho 与 Ermon 提出的生成对抗模仿学习（Generative Adversarial Imitation Learning, GAIL）[67] 将策略学习转化为专家与策略轨迹分布之间的对抗匹配问题，为后续对抗式运动先验方法奠定了理论基础。在物理仿真角色控制领域，Peng 等提出的 Adversarial Motion Priors（AMP）[68] 首次将对抗式分布约束引入物理仿真控制框架，通过训练判别器学习人类运动分布，并将其输出转化为奖励信号，使策略在强化学习过程中逐步逼近人类动作统计特征。该方法证明，在不显式构造复杂奖励函数的情况下，机器人可以生成具有自然节律与协调结构的全身运动。

随后，Adversarial Skill Embedding(ASE)[69] 将对抗式运动先验扩展至多技能统一控制框架，通过构建共享潜在技能空间，使单一策略能够编码多种运动模式。这一阶段的重要进展在于将“模仿单一动作”拓展为“学习可组合的运动分布”，为复杂全身行为的统一表达提供了结构化表示。

在人形机器人领域，模仿学习逐渐与强化学习相结合，形成“模仿 + 强化”的混合训练范式。一方面，Adaptive Humanoid Control（AHC）等多行为蒸馏框架[70] 通过多专家策略训练与蒸馏融合，实现起身、行走等多技能统一控制，并在蒸馏过程中保留不同动作中的腿臂协同模式；另一方面，基于生成模型的运动潜空间方法，如 Motion Variational Autoencoder（MVAE）[71]，通过变分自编码器（Variational Autoencoder, VAE）学习低维运动潜在空间，使控制问题在结构化动作空间中进行

优化，从而在保证运动自然性的同时提高训练稳定性与可控性。

随着高精度动作捕捉数据集（如 AMASS）与高保真物理仿真平台的发展，模仿学习方法逐渐能够在复杂接触、多阶段动作与高动态场景下实现全身控制策略训练，并成功部署于真实人形机器人平台。这一阶段的核心贡献在于：将人类运动分布作为结构先验引入控制问题，使腿部推进、躯干稳定与手臂辅助之间的协同关系通过数据分布本身得到编码，而无需显式构造耦合规则。

从发展脉络来看，模仿学习在人形机器人腿臂协同控制中的演进可以概括为三个阶段：第一阶段为基于监督回归的动作复制；第二阶段为基于分布匹配的对抗式运动先验；第三阶段为基于潜空间建模的生成式运动表示。在这一过程中，腿臂协同结构逐渐从“被动复制的轨迹模式”演化为“可泛化的结构化运动分布表示”，为复杂全身控制提供了重要的数据驱动支撑。

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 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，其基本形式为：

$$\mathcal{L}_{BC}(\theta) = \mathbb{E}_{(s,a^*)\sim\mathcal{D}}\left[\|\pi_\theta(s) - a^*\|^2\right], \tag{1.10}$$

其中 $\mathcal{D}$ 为专家数据集，a^* 为专家动作。行为克隆的核心思想是通过函数逼近直接复制专家策略。该方法实现简单、训练稳定，但由于仅在专家状态分布上优化，当执行过程中状态分布发生偏移时，误差可能逐步累积，影响长期控制性能。

2）对抗式模仿学习 对抗式模仿学习通过匹配专家轨迹分布与策略生成轨迹分布进行训练。典型方法如 AMP[68] 引入判别器 D_ϕ，并构造如下对抗目标：

$$\mathcal{L}_D = \mathbb{E}_{s_E}[\log D_\phi(s_E)] + \mathbb{E}_{s_\pi}[\log(1 - D_\phi(s_\pi))]. \tag{1.11}$$

判别器用于区分专家状态 s_E 与策略生成状态 s_π，并将判别结果转化为训练信号。其核心思想并非逐帧回归专家动作，而是在分布层面对齐专家与策略行为，使策略在统计意义上继承人类运动结构，从而在物理仿真环境中自然形成腿臂协同模式。

3）生成式运动先验　生成式运动先验方法通过学习低维潜在空间对复杂运动分布进行结构化建模。以基于变分自编码器的 MVAE[71] 为例，其运动生成过程可表示为：

$$p_\theta(x_{t+1} \mid x_t, z), \tag{1.12}$$

其中 z 为潜变量，用于调节运动演化方向。通过在潜在空间中进行控制或优化，原本高维复杂的动作生成问题被转化为对潜变量的调节过程。由于人类运动中的腿臂协同关系已在潜空间中被编码，控制过程中无需显式构造耦合约束即可生成自然协调的全身动作。

为更加系统地分析模仿学习在腿臂协同控制中的技术特征，有必要从结构继承能力、泛化性能以及约束表达能力等方面进行综合讨论。

首先，在动作自然性与协同结构形成方面，模仿学习具有显著优势。由于人类运动数据中天然包含腿部推进、躯干稳定与手臂辅助之间的动力学耦合关系，策略在学习运动分布的过程中能够隐式继承这种协调结构，而无需显式构建腿臂耦合规则或复杂奖励函数。相比纯强化学习依赖人工奖励调节的方式，模仿学习在动作节律保持、动量分配一致性以及整体姿态连贯性方面通常表现更加自然与稳定。此外，生成式潜空间方法通过构建结构化运动表示，使高维动作生成过程受限于低维潜变量调节，从而提升训练稳定性并减少策略搜索空间的复杂度。

其次，在训练效率与工程实现层面，模仿学习能够利用离线运动数据进行预训练，降低在线探索成本，并可作为强化学习的运动先验模块使用。这种“模仿 + 强化”的混合范式在复杂全身控制任务中具有较强实用价值，有助于缩短训练周期并提升策略初始稳定性。

然而，模仿学习方法同样存在若干局限。其一，方法性能高度依赖示范数据的质量与覆盖范围，当目标任务超出数据分布支持区域时，策略可能出现退化或失稳。其二，由于模仿学习主要关注运动分布匹配，动力学约束与接触一致性通常未被显式建模，在复杂多接触或强扰动环境中可能缺乏严格的稳定性保证。其三，虽然腿臂协同关系可以在数据中被继承，但其物理形成机制并未显式刻画，协同结构缺乏明确的理论可解释性与层级控制结构。

综合来看，模仿学习在腿臂协同控制中的核心价值在于结构继承，即通过数据驱动方式将人类运动中的协调模式直接编码于策略之中。然而，在泛化能力、约束显式建模以及可解释性方面仍存在改进空间。

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—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. 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.

[53] 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.

[54] Sentis L. Synthesis and control of whole-body behaviors in humanoid systems[D]. Stanford University, 2007.

[55] 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.

[56] Escande A, Mansard N, Wieber P B. Hierarchical quadratic programming: fast online humanoid-robot motion generation[J]. The International Journal of Robotics Research, 2014, 33(7): 1006-1028.

[57] 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.

[58] 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.

[59] Siekmann J, et al. Learning diverse bipedal gaits from scratch[J]. Science Robotics, 2021.

[60] Gu X, et al. Humanoid-gym: reinforcement learning for humanoid robots[J], 2024.

[61] Xie Z, et al. Pbhc: physics-based humanoid control[J], 2025.

[62] Zhang Y, et al. Wococo: whole-body contact control via reinforcement learning[J], 2024.

[63] Fu Y, et al. Humanplus: humanoid whole-body control via reinforcement learning[J], 2024.

[64] Argall B D, Chernova S, Veloso M, et al. A survey of robot learning from demonstration[J]. Robotics and Autonomous Systems, 2009, 57(5): 469-483.

[65] Pomerleau D A. Alvinn: an autonomous land vehicle in a neural network[C]. //Advances in Neural Information Processing Systems (NeurIPS). 1989.

[66] Bain M. Experiments in imitation learning[D]. University of Cambridge, 1995.

[67] Ho J, Ermon S. Generative adversarial imitation learning[C]. //Advances in Neural Information Processing Systems (NeurIPS). 2016.

[68] 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.

[69] 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.

[70] 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.

[71] 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.