强化学习与机器人控制

[墨] 余文(Wen Yu)
阿道夫·佩鲁斯基亚(Adolfo Perrusquía) 著

刘晓骏 译

清華大學出版社
北 京

北京市版权局著作权合同登记号　图字：01-2023-4476

图书在版编目(CIP)数据

强化学习与机器人控制 /(墨) 余文 (Wen Yu)，(墨) 阿道夫 • 佩鲁斯基亚 (Adolfo Perrusquía) 著；刘晓骏译. —北京：清华大学出版社，2023.7
书名原文：Human-Robot Interaction Control Using Reinforcement Learning
ISBN 978-7-302-63740-0

I. ①强… II. ①余… ②阿… ③刘… III. ①机器人控制 IV. ①TP24

中国国家版本馆 CIP 数据核字(2023)第 111020 号

责任编辑：王　军
装帧设计：孔祥峰
责任校对：成凤进
责任印制：沈　露

出版发行：清华大学出版社
网　　址：http://www.tup.com.cn，http://www.wqbook.com
地　　址：北京清华大学学研大厦 A 座　　邮　　编：100084
社 总 机：010-83470000　　邮　　购：010-62786544
投稿与读者服务：010-62776969，c-service@tup.tsinghua.edu.cn
质 量 反 馈：010-62772015，zhiliang@tup.tsinghua.edu.cn
印 装 者：三河市东方印刷有限公司
经　　销：全国新华书店
开　　本：148mm×210mm　　**印　　张**：8.125　　**字　　数**：256 千字
版　　次：2023 年 9 月第 1 版　　**印　　次**：2023 年 9 月第 1 次印刷
定　　价：98.00 元

产品编号：098661-01

作者简介

Wen Yu 于 1990 年获得清华大学自动控制学士学位，于 1992 年和 1995 年在东北大学分别获得电气工程硕士学位和博士学位。1995—1996 年，他在东北大学自动控制系担任讲师。自 1996 年以来，他一直在墨西哥国立理工学院(CINVESTAV-IPN)工作，目前在该校担任自动化控制系教授。2002—2003 年，他在墨西哥石油研究所担任研究职位。2006—2007 年，他是英国贝尔法斯特女王大学的高级客座研究员。2009—2010 年，他担任加州大学圣克鲁斯分校的客座副教授。2006 年，他还担任东北大学的客座教授。另外，Wen 博士还是 IEEE 控制论、神经计算学报，以及智能和模糊系统杂志的副编辑，是墨西哥科学院的成员。

Adolfo Perrusquía(IEEE 成员)于 2014 年获得墨西哥国立理工学院(UPIITA-IPN)工程与先进技术跨学科专业机电一体化工程学士学位。他分别于 2016 年和 2020 年获得墨西哥国立理工学院(CINVESTAV-IPN)研究与高级研究中心自动控制系的硕士和博士学位。目前，他是克兰菲尔德大学的研究员，是 IEEE 计算智能协会的成员。他的主要研究方向是机器人学、机械、机器学习、强化学习、非线性控制、系统建模和系统识别。

前　言

机器人控制是控制理论和应用领域的一个热门话题，主要的理论贡献是利用线性和非线性方法，使机器人能够执行一些特定的任务。机器人交互控制是科学研究和工程应用领域的一个热门课题。机器人交互控制方案的主要目标是实现机器人与环境之间的预期性能，并能够安全、精确地运动。环境可以是机器人外部的任何材料或系统，如操作人员。机器人交互控制器可以根据位置、受力或两者结合进行设计。

最近，通过利用动态规划理论，强化学习技术被应用于最优控制和鲁棒控制。它们不需要具有系统动力学基础并且能够进行内部和外部更改。

从 2013 年开始，作者及其团队开始使用神经网络和模糊系统等智能技术研究人机交互控制。2016 年，作者将更多注意力放在如何利用强化学习解决人机交互问题上。经过四年的工作，他们在关节空间和任务空间中提出了基于模型和无模型的阻抗和导纳控制的结果，还分析了闭环系统，并且讨论了无模型最优机器人交互控制和基于强化学习的位置/受力控制设计。他们研究了庞大的离散时间空间和连续时间空间中的强化学习方法。对于冗余机器人的控制，他们使用多智能体强化学习来解决，并分析强化学习的收敛性。将最坏情况下不确定性的鲁棒人机交互控制转化为“$\mathscr{H}_2/\mathscr{H}_\infty$问题”，采用强化学习和神经网络设计并实现最优控制器。

本书假设读者熟悉基于经典和高级控制器进行机器人交互控制的一些应用，将进一步对系统识别、基于模型和无模型的机器人交互控制器进行系统性分析。本书适用于研究生以及执业工程师。阅读本书需要掌握的先决知识是：机器人控制、非线性系统分析，特别是 Lyapunov 方法、神经网络、优化技术和机器学习。本书还适用于许多对机器人和控制感兴趣的研究人员和工程师。

许多人对本书做出了贡献。第一作者要感谢 CONACYT 基金项目 CONACYT-A1-S-8216、CINVESTAV 基金项目 SEP-CINVESTAV-62 和 CNR-CINVESTAV 基金项目提供的财政支持；他还要感谢妻子 Xiaoou，她为本书投入了大量的时间和精力，没有她的帮助本书不可能完成。第二作者要衷心感谢他的导师 Prof. Wen Yu 对其博士研究的不断支持，感谢他给予的耐心、积极、热情和渊博的知识，在他的悉心帮助指导下才得以完成本书。此外，他还要感谢 Prof. Alberto Soria、Prof. Rubén Garrido、Ing. José de Jesús Meza。最后，第二作者还要感谢他的父母 Adolfo 和 Graciela，他们为本书花费了许多时间和心血，没有他们，本书不可能顺利出版。

在此要说明的是，本书各章正文在涉及参考文献时，采用的是中括号内加数字的形式，如[3]、[3,8]、[3-10]这三种形式分别表示该章中的第 3 个参考文献、第 3 个和第 8 个参考文献、第 3 个到第 10 个参考文献。本书各章中的参考文献我们采用线上形式提供，读者可通过扫描封底二维码下载得到。

目　录

第 I 部分
人机交互控制

第Ⅰ部分
人机交互控制

第1章
介　绍

1.1　人机交互控制

如果已经了解了机器人动力学的相关知识，就可以设计基于模型的控制器(见图 1.1)。著名的线性控制器包括比例微分器(Proportional-Derivative，PD)[1]、线性二次调节器(Linear Quadratic Regulator，LQR)和比例积分微分器(Proportional-Integral-Derivative，PID)[2]，它们基于线性系统理论，需要在某些操作点对机器人动力学进行线性化。LQR[3-5]控制已被用作强化学习方法设计的基础[6]。

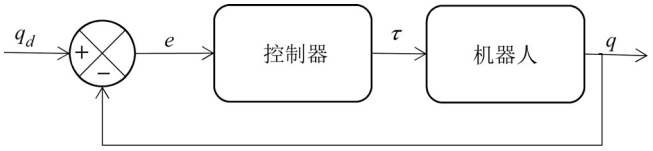
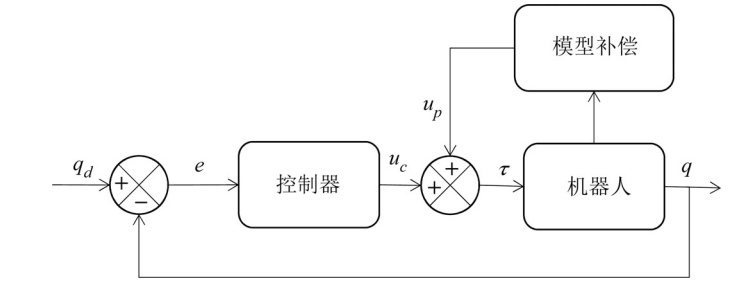
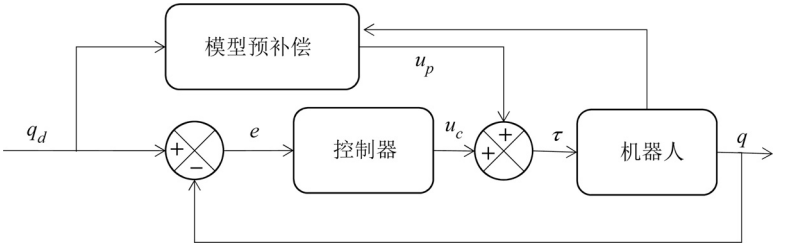


图 1.1　经典的机器人控制

经典控制器使用机器人动力学的全部或部分知识。在这些情况下(不考虑干扰)，可以设计具有完美跟踪性能的控制器。通过使用补偿或预补偿技术来代替机器人动力学，并建立了更简单的期望动力学[7-9]。在关节空间中具有模型补偿或预补偿的控制方案如图 1.2 所示。这里 q_d 是所需的参考值，q 是机器人的关节位置，$e = q_d - q$ 是关节误差，u_p 是动力学的补偿器或预补偿器，u_c 是来自控制器的控制，而 $\tau = u_p + u_c$ 是控制扭矩。典型的模型补偿控制是带有重力补偿的比例微分(PD)控制器，这有助于减少由机器人动力学的重力项而引起的稳态误差。



(a) 模型补偿



(b) 模型预补偿

图 1.2　模型补偿控制

当没有熟练掌握动力学知识时，就不可能设计以前的控制器。因此需要使用无模型控制器。一些著名的无模型控制器有 PID 控制[10, 11]、滑动模型控制[2, 12]和神经控制[13]。这些控制器在某些条件下(干扰、摩擦、参数)根据特定的设备进行调整。当新的情况出现时，控制器就会出现不同的行为，甚至达不到稳定状态。无模型控制器能胜任不同的任务，并且相对容易调整；然而当改变机器人参数或施加干扰时，它们不能保证最佳性能，并且需要重新调整控制增益。

上述所有控制器均用于位置控制的设计，不考虑与环境的交互。与交互作用相关的任务有很多种，如刚度控制、力控制、混合控制和阻抗控制[14]。可以使用 P(刚度控制)、PD 和 PID 力控制器来调节交互力[15]。位置控制也可以使用力控制来执行位置跟踪和速度跟踪[16, 17](见图 1.3)。此处，f_d 为所需力，f_e 为接触力，$e_f = f_d - f_e$ 是力的误差，x_r 是力控制器的输出，$e_r = x_r - x_d$ 是任务空间中的位置误差。力/位置控制是使用力进行补偿的[17]，它还可以使用全动态来线性化闭环系统，以实现完美跟踪[18]。

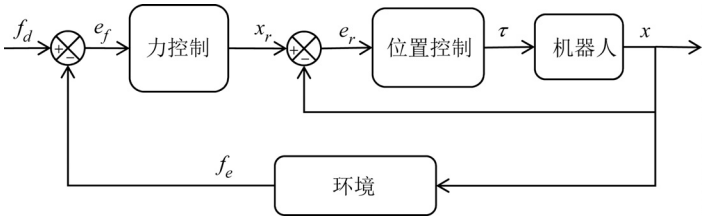


图 1.3　位置/力控制

阻抗控制[7]解决了机器人终端执行器与外部环境接触时如何移动的问题。它使用所需的动态模型(也称为机械阻抗)来设计控制。最简单的阻抗控制是刚度控制，其中机器人的刚度和环境具有比例交互作用[19]。

传统的阻抗控制是通过假设机器人模型精确来线性化系统[20-22]。这些算法需要强有力的假设，即精确的机器人动力学是已知的[23]。控制的鲁棒性在于模型的补偿。

大多数阻抗控制器假设阻抗模型的所需惯量等于机器人惯量。因此，我们只有刚度和阻尼项，这相当于 PD 控制律[8, 21, 24]。解决动态模型补偿不准确的方法是使用自适应算法、神经网络或其他智能方法[9, 25-31]。

阻抗控制有几种实现方式。在[32]中，阻抗控制使用人体特征来获得所需阻抗的惯性、阻尼和刚度分量。对于位置控制，它使用了 PID 控制，这有利于省略模型补偿。避免使用模型或在不完全了解模型的情况下继续操作的另一种方法是利用系统特性，即高传动比减速，这会导致非线性元件变得非常小，系统从而变得解耦[33]。

在机械系统中，特别是在触觉领域，导纳是从力到运动的动态映射。输入力“允许”一定量的运动[11]。基于阻抗或导纳的位置控制需要通过逆阻抗模型来获得参考位置[34–38]。这种方案更完整，因为有一个双控制回路，可以更直接地使用与环境的交互。

阻抗/导纳控制的应用相当广泛，例如，人类操作员所使用的外骨骼。为了保护人身安全需要低机械阻抗，同时跟踪控制需要高阻抗来抑制干扰。因此，还有不同的解决方案，例如频率建模和使用系统的极点和零点来降低机械阻抗[39, 40]。

基于模型的阻抗/导纳控制对建模误差较敏感。这里对经典阻抗/导纳控制器进行了一些修改，例如基于位置的阻抗控制，它使用内部位置控制回路提高了存在建模误差时的鲁棒性[21]。

1.2　控制强化学习

图 1.4 显示了强化学习的控制方案。与图 1.1 中无模型控制器的主要区别在于，强化学习在每个步骤中使用跟踪误差和控制扭矩来更新其值。

强化学习方案首先用于具有离散输入空间的离散时间系统[6, 41]。最著名的方法有 Monte Carlo[42]、Q 学习[43]、Sarsa[44]和评论家算法[45]。

如果输入空间较大或者连续，由于计算成本，经典的强化学习算法无法直接实现，并且在大多数情况下，算法不会收敛[41, 46]，这个问题被称为机器学习的维数灾难。对于机器人控制，维数灾难会增加，因为存在不同的自由度，每个自由度都需要各自的输入空间[47, 48]。另一个使维数问题更加突出的因素是扰动，因为必须考虑新的状态和控制。

为了解决维数灾难问题，可以将基于模型的技术应用于强化学习[49-51]。这些学习方法非常流行，其中一些算法被称为“策略搜索”[52-59]。然而，这些方法需要利用模型的知识来降低输入空间的维数。

与离散时间算法类似，无模型算法也有多种类型。这些算法的主要思想是设计足够多的奖励和逼近器，从而在具有较大或连续输入空间的情况下降低计算成本。

减小输入空间最简单的逼近器就是手工方法[60-65]。这种方法是通过寻找回报最小化/最大化的区域来加快学习时间。[66, 67]使用来自输入数据的学习方法，这类似于离散时间学习算法，但学习时间会增加。其他技术基于先前一系列相关方法所实现的动作，也就是说，每个时刻必须实现的动作被定义为自我完成的一项简单任务[68-72]。这些方法的主要问题是需要专家知识来获得最佳区域并且需要设置预定义动作。

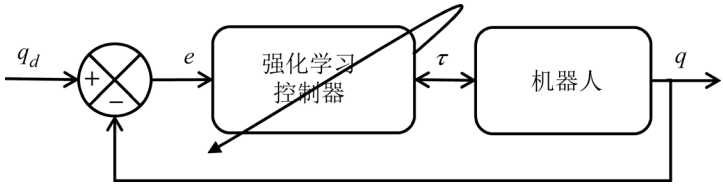


图 1.4　强化学习控制方案

逼近器的线性组合从输入数据中学习，无须专家干预。机器人控制中最广泛使用的逼近器受到人类形态学[73, 74]、神经网络[75-77]、局部模型[74, 78]和高斯回归过程[79-82]的启发。这些逼近器的成功归功于其参数和超参数的充分选择。

较差的奖励设计可能涉及较长的学习时间、收敛到错误的解，或者算法永远不会收敛到任何解。另一方面，适当的奖励设计有助于算法更快地在每个时刻找到最佳解。这个问题被称为“奖励设计的诅咒”[83]。

当使用无模型方法时，奖励的设计应使其适应系统的变化和可能的误差，这在鲁棒控制问题中非常有用，其中控制器需要能够补偿干扰或限制干扰以获得最佳性能。

1.3　本书的结构安排

本书由两个主要部分组成：

- 第 I 部分涉及不同环境下人机交互控制的设计(第 2~5 章)。
- 第 II 部分讨论了机器人交互控制的强化学习(第 6~10 章)。

第 I 部分

第 2 章：讨论机械和电气意义上机器人交互控制的一些重要概念。阻抗和导纳的概念在机器人交互控制设计和环境建模中起着重要作用。介绍典型的环境模型和一些最著名的环境模型参数估计识别技术。

第 3 章：讨论第一个使用阻抗和导纳控制器的机器人交互方案。经典控制器基于反馈线性化控制律的设计。基于所提出的阻抗模型，将闭环动力学简化为所需动力学，该阻抗模型被设计为二阶线性系统。详细解释经典阻抗和导纳

控制的精度和鲁棒性问题。通过在两种不同环境中的仿真，说明这些控制器的适用性。

第 4 章：研究一些不需要完全了解机器人动力学的无模型控制器。为导纳控制方案设计无模型控制器。交互由导纳模型控制，而位置控制器使用自适应控制、PID 控制或滑动模型控制。通过 Lyapunov 稳定性理论证明了这些控制器的稳定性。通过使用不同环境和机器人进行仿真和实验，证明了这些算法的适用性。

第 5 章：提出一种新的机器人交互控制方案，称为闭环控制。本章中，环境是人类的操作者。人类与机器人没有接触。该方法使用操作员的输入力/扭矩，并通过导纳模型将其映射到终端执行器的位置/方向。由于人处于控制回路中，并不知道施加的力/扭矩是否会依赖于奇异点位置，因此这对实际应用是危险的。因此，对前几章中的导纳控制器进行了修改，以避免逆运动，并使用欧拉角修改 Jacobian 矩阵。实验证明了该方法在关节空间和任务空间的有效性。

第 II 部分

第 6 章：前几章使用所需的阻抗/导纳模型来实现所需的机器人-环境交互。在大多数情况下，这些交互并不具有最优性能，存在相对较高的接触力或较高的位置误差，因为它们需要基于环境和机器人动力学。本章讨论离散时间位置/力控制的强化学习方法。强化学习技术可以实现次优的机器人-环境交互。

最优阻抗模型通过两种不同的方法实现：使用线性二次调节器的动态规划和强化学习。第一种是基于模型的控制律，第二种是基于无模型的。为了加速强化学习方法的收敛，使用了资格迹和时间差分方法。本章讨论了强化学习算法的收敛性。利用 Lyapunov-like 分析法分析了位置和力控制的稳定性。仿真和实验验证了该方法在不同环境下的有效性。

第 7 章：本章涉及的内容与第 6 章中讨论的强化学习方法的大规模连续时间相对应。由于我们考虑的是较大的输入空间，因此经典的强化学习方法无法处理最优解问题，并且可能无法收敛。这需要使用逼近器来减少计算工作量，并且获得可靠的最优或逼近最优解。本章讨论离散和连续时间中的维数问题。

本书使用基于正则化径向基函数的参数逼近器。通过 K 均值聚类算法和随机聚类获得每个径向基函数的中心。利用压缩性质和 Lyapunov-like 分析，分析了强化学习逼近的离散和连续时间版本的收敛性。

本章提出了一种利用离散和连续时间版本的混合强化学习控制器，通过仿真和实验验证了算法在不同环境下的位置/力控制任务中的性能。

第 8 章：在最坏情况不确定性下基于改进的强化学习来设计鲁棒控制器，该控制器具有鲁棒性，并提供最优或逼近最优解。强化学习方法分为离散时间和连续时间两种。这两种方法都将奖励设计为约束下的优化问题。

在离散时间内，使用 k 近邻和双估计器技术修改强化学习算法，可以避免出现运动的高估计值。为此，开发了两种算法：大型状态和离散动作的情况以及大型状态-动作的情况。利用收缩特性分析了算法的收敛性。在连续时间内，我们在修改后的奖励下使用了与第 7 章相同的算法。通过仿真和实验证明了鲁棒控制器的有效性。

第 9 章：对于某些类型的机器人，如冗余度机器人，由于它具有奇异性，不可能计算逆运动或使用 Jacobian 矩阵。本章仅结合机器人正向运动知识，使用多智能体强化学习方法来处理此问题。本章讨论了机器人和冗余度机器人中逆运动和速度运动的解决方法。

为了确保可控性并避免奇异解或多数解，本书使用了多智能体强化学习和双值函数方法。运动方法用于避免维数灾难。我们使用小关节位移作为控制输入，直到达到所需的参考值。此外，还讨论了算法的收敛性。仿真和实验表明，与标准的 actor-critic 方法相比，该方法具有令人满意的结果。

第 10 章：我们在前几章使用的强化学习方法是从头开始学习最优控制策略，这意味着需要大量的学习时间。为了给控制器提供先前的知识，本章给出了一个$\mathcal{H}_2$在离散时间和连续时间使用强化学习的神经控制。控制器使用学习得到的动力学知识来计算最优控制器。利用收缩特性和 Lyapunov-like 分析对所提出的神经控制的收敛性进行了分析。通过仿真验证了控制器的优化和鲁棒性。

附录

附录 A：讨论机器人运动和动力学模型的一些基本概念和特性。动力学表

达在关节空间和任务空间中。我们还通过 Denavit-Hartenberg 约定和 Euler-Lagrange 公式给出本书中使用的机器人和系统的运动和动力学模型。

附录 B：给出强化学习的基本理论和一些著名的控制器设计算法。讨论强化学习方法的收敛性。

第 2 章
人机交互的环境模型

2.1 阻抗和导纳

在电场中通常使用阻抗概念。图 2.1 所示是一个串行 RLC 电路，其中 V 是电源，$I(t)$是流经的电流，R 是电阻器，L 是电感器，C 是电容器。阻抗是电压相量和电流相量之间的商，即它遵循欧姆定律。

$$Z = \mathbb{V}/\mathbb{I} \tag{2.1}$$

其中$\mathbb{V}$是电压相量，$\mathbb{I}$是电流相量。

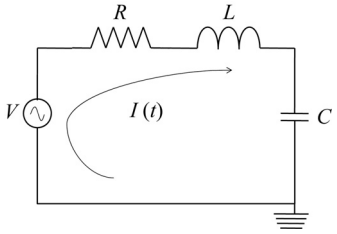


图 2.1 RLC 电路

将 Z_R、Z_L 和 Z_C 分别定义为电阻器、电感器和电容器的阻抗。Z_R 测量电阻元件阻碍通过电路的电流的程度(电阻是耗散元件)，Z_L 测量电感元件阻碍通过网络的电流水平的程度(电感元件是存储设备)，Z_C 测量电容元件对通过网络的电流水平的阻碍程度(电容元件是存储设备)[1]。

图 2.1 中 RLC 电路的动力学方程为：

$$\mathbb{V} = \left(Z_R + Z_L + Z_C\right)\mathbb{I} \tag{2.1}$$

在上面的方程中，可以通过电阻找到电流和电压之间的关系。RLC 电路的

动力学方程也可以从 Kirchhoff 定律中获得，如下所示：

$$L\frac{\mathrm{d}I(t)}{\mathrm{d}t}+RI(t)+\frac{1}{C}\int_0^t I(\tau)\mathrm{d}\tau=V \tag{2.2}$$

如果将拉普拉斯变换应用于初始条件为零的微分方程(2.2)，则为：

$$\left(Ls+R+\frac{1}{Cs}\right)I(s)=V(s),\quad \text{或}\quad \left(Ls^2+Rs+\frac{1}{C}\right)q(s)=V(s) \tag{2.3}$$

其中 q 是电荷量。

第一个表达式显示了电压和电路电流之间的阻抗关系，而第二个表达式给出了电压和电荷之间的关系，也称为阻抗滤波器：

$$Z(s)=Ls^2+Rs+1/C$$

对于机械系统，可以写出类似的关系式。考虑图 2.2 中的质量-弹簧-阻尼器系统，其中 m 是汽车质量，k 是弹簧刚度，c 是阻尼系数，F 是作用力。汽车只有水平运动，弹簧和减振器有线性运动。表示质量-弹簧-阻尼器系统的动力学方程为

$$m\ddot{x}+c\dot{x}+kx=F \tag{2.4}$$

式(2.4)的拉普拉斯变换是

$$\left(ms^2+cs+k\right)x(s)=F(s) \tag{2.5}$$

机械阻抗或阻抗滤波器为

$$Z(s)=ms^2+cs+k$$

每个元素都有其各自的作用。弹簧存储势能(相当于电容器)，阻尼器耗散动能(相当于电阻器)，质量阻碍汽车可以获得的速度(相当于电感)。

阻抗的倒数叫作导纳。在这里，当施加力时，导纳允许一定量的运动。在机械系统中，导纳可以表示为：

$$Y(s)=Z^{-1}(s)=\frac{1}{ms^2+cs+k} \tag{2.6}$$

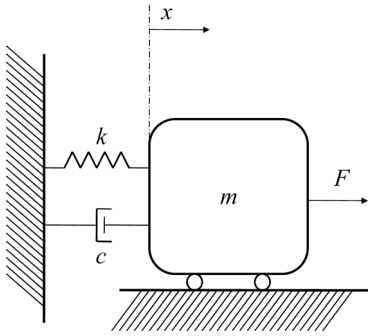


图 2.2　质量-弹簧-阻尼器系统

机械阻抗和导纳模型在环境交互应用中非常有用。环境模型是力控制、人机交互和医疗机器人的组成部分，它们都是交互控制策略的示例。如果环境不随时间变化，通常就会和线性弹簧 k 一起建模，有时还会与阻尼器 c 一起建模。这两个元素都是常量。

对于线性环境，阻抗由作用力和流的拉普拉斯变换的商定义。在电力系统中，作用力等于电压，流等于电流。在机械系统中，作用力是力或扭矩，而流是线速度或角速度。对于任何给定频率 ω, 阻抗是具有实部 $R(\omega)$和虚部 $X(\omega)$的复数。

$$Z(\omega) = R(\omega) + \mathrm{j}X(\omega) \tag{2.7}$$

当 ω 接近零时，环境阻抗的幅值可能具有以下值：幅值可以接近无穷大、有限值、某个非零值，或者可以接近零。介绍以下定义：

定义 2.1　当且仅当$|Z(0)| = 0$ 时，具有(2.7)阻抗性质的系统称为惯性系统。

定义 2.2　当且仅当$|Z(0)| = a$ 时，且对于某些正数$a \in (0, \infty)$ 成立时，具有(2.7)阻抗性质的系统是具有电阻特性的。

定义 2.3　当且仅当$|Z(0)| = \infty$ 时，具有(2.7)阻抗性质的系统是具有电容特性的。

电容和惯性环境是对偶表示，即电容系统的逆是惯性的，惯性系统的逆是电容的。电阻特性的环境是自对偶的。为了表示系统的对偶性，我们使用了 Norton 和 Thèvenin[2]提出的等效电路。

Norton 等效电路由与流并联的阻抗组成。Thèvenin 等效电路由与作用源串

联的阻抗组成。Norton 等效电路表示电容系统。Thèvenin 等效电路表示惯性系统。这两种等效电路都可以用来表示电阻系统。

控制设计的主要基础是阶跃控制输入的稳态误差必须为零。这可以通过以下原则来实现。

对偶原理：必须控制机器人操纵器，使其能够响应环境的对偶。

这个原理可以容易地使用 Norton 和 Thèvenin 等效电路来解释。当环境具有电容性质时，它由与流并联的阻抗表示。机器人操纵器(对偶)必须由带有非电容阻抗(惯性或电阻)的作用源串联表示(见图 2.3)。

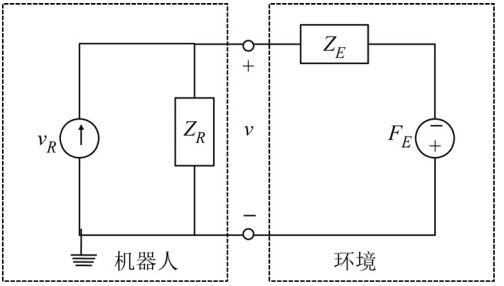


图 2.3　位置控制

当环境具有惯性特性时，它由与作用源串联的阻抗表示，机器人必须由与非惯性阻抗(电容或电阻)并联的流源表示(见图 2.4)。当环境具备电阻特性时，它可以使用任何等效电路，但操纵器的阻抗必须是非电阻特性。总之，电容特性的环境需要由力控制的机器人，惯性特性的环境需要由位置控制的机器人，电阻特性的环境则需要由位置和力两者共同控制的机器人。

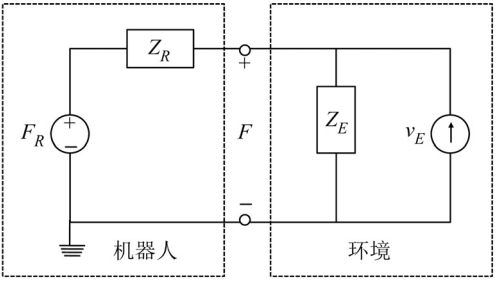


图 2.4　力控制

现在已经证明了对偶原理保证了在环境中不存在输入的阶跃控制的零稳态误差。首先，假设环境是惯性特性的，因此

$$Z_E(0) = 0$$

其中下标 E 表示环境。图 2.3 显示了环境及其各自的操纵器，其中 Z_E 是环境的阻抗，Z_R 是机器人操纵器的阻抗，v_R 是输入流，v 是在环境和机器人之间测量出来的流。流的输入-输出传递函数为

$$\frac{v}{v_R} = \frac{Z_R(s)}{Z_R(s) + Z_E(s)} \tag{2.8}$$

如果极点位于复平面的左半平面，那么可由终值定理得出，阶跃控制输入的稳态误差为 1/s：

$$e_{ss} = \lim_{t\to\infty}(v - v_R) = \lim_{s\to 0} s(v(s) - v_R(s)) = \frac{-Z_E(0)}{Z_R(0) + Z_E(0)} = 0 \tag{2.9}$$

此处 $Z_R(0) \neq 0$，即操纵器的阻抗是非惯性特性。

当环境具有电容特性时，$Z_E = \infty$。图 2.4 显示了环境及其各自的对偶操纵器，其中 F_R 是机器人作用力的输入。力 F 的输入/输出转换函数为：

$$\frac{F}{F_R} = \frac{Z_E(s)}{Z_R(s) + Z_E(s)} \tag{2.10}$$

阶跃输入的稳态误差为：

$$e_{ss} = \lim_{t\to\infty}(F - F_R) = \lim_{s\to 0} s(F(s) - F_R(s)) = \frac{-Z_R(0)}{Z_R(0) + Z_E(0)} = 0 \tag{2.11}$$

其中 $Z_R(0)$是有限的，即机器人阻抗是具有非电容特性的。

对于电阻环境，满足零稳态误差，$Z_R(0) = 0$ 且操纵器为力控制，或 $Z_R(0) = \infty$ 且操纵器是位置控制的。

二元性原则表明，在关节网络里不能同时维持两种不同的流或两种不同的作用力。跟踪位置轨迹和跟踪环境轨迹是不一致的。然而，Norton 流源和 Thèvenin 作用源的双重组合可以同时存在[2]。

2.2　人机交互阻抗模型

由阻抗控制的机器人基于二阶动力学[3]，它给出了终端执行器的位置和外力之间的关系。这种关系的特征由所需的阻抗值 $\boldsymbol{M}_d$、$\boldsymbol{B}_d$ 和 $\boldsymbol{K}_d$ 决定。它们由用户根据所需的动态性能选择得到[3]。

这里 $\boldsymbol{M}_d \in \mathbb{R}^{m\times m}$ 是所需质量矩阵，$\boldsymbol{B}_d \in \mathbb{R}^{m\times m}$ 是所需阻尼矩阵，$\boldsymbol{K}_d \in \mathbb{R}^{m\times m}$ 是所需刚度矩阵。假设环境是具有刚度和阻尼参数($\boldsymbol{K}_e$、$\boldsymbol{C}_e$)的线性阻抗特性。

通常要独立处理每个Cartesian变量，即假设不同轴上的环境阻抗是解耦的。因此，所需阻抗模型由以下 m 个微分方程给出：

$$m_{d_i}(\ddot{x}_i - \ddot{x}_{d_i}) + b_{d_i}(\dot{x}_i - \dot{x}_{d_i}) + k_{d_i}(x_i - x_{d_i}) = f_{e_i} \tag{2.12}$$

其中 $i = 1, \cdots, m, x_d, \dot{x}_d, \ddot{x}_d \in \mathbb{R}^m$ 是所需的位置、速度和加速度，以及 x，$\dot{x}$，$\ddot{x} \in \mathbb{R}^m$ 是机器人终端执行器的位置、速度和加速度。

以矩阵形式表示为：

$$\boldsymbol{M}_d(\ddot{x} - \ddot{x}_d) + \boldsymbol{B}_d(\dot{x} - \dot{x}_d) + \boldsymbol{K}_d(x - x_d) = \boldsymbol{f}_e \tag{2.13}$$

其中 $\boldsymbol{f}_e = [F_x, F_y, F_z, \tau_x, \tau_y, \tau_z]^\top$ 是力/扭矩矢量。所需阻抗矩阵为：

$$\boldsymbol{M}_d = \begin{bmatrix} m_{d_1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & m_{d_n} \end{bmatrix}, \boldsymbol{B}_d = \begin{bmatrix} b_{d_1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & b_{d_n} \end{bmatrix}, \boldsymbol{K}_d = \begin{bmatrix} k_{d_1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & k_{d_n} \end{bmatrix} \tag{2.14}$$

环境阻抗矩阵为：

$$\boldsymbol{C}_e = \begin{bmatrix} c_{e_1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & c_{e_m} \end{bmatrix}, \quad \boldsymbol{K}_e = \begin{bmatrix} k_{e_1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & k_{e_m} \end{bmatrix} \tag{2.15}$$

对于单自由度机器人-环境交互，机器人和环境有接触(见图 2.5)，得到了由控制器和环境的阻抗特性组成的二阶系统。如果可以确定组合系统(阻抗控制器和环境)的阻抗特征，则可以计算环境特征。

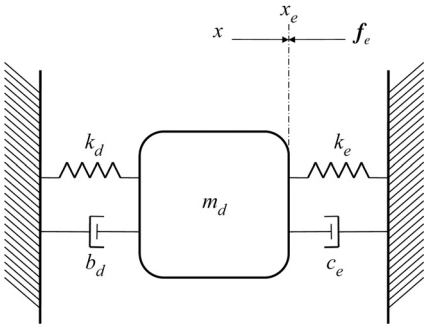


图 2.5　环境和机器人的二阶系统

如果环境位于 x_e 位置，则图 2.5 中所展示的交互动力状态为：

$$m_d(\ddot{x}-\ddot{x}_d)+b_d(\dot{x}-\dot{x}_d)+k_d(x-x_d)=-c_e\dot{x}+k_e(x_e-x) \tag{2.16}$$

其中 $x_e \leqslant x$。如果 $x_e=0$，上述系统的传递函数为：

$$x(s)=x_d(s)\left[1-\frac{c_e s+k_e}{m_d s^2+(b_d+c_e)s+k_d+k_e}\right] \tag{2.17}$$

当没有交互作用时，接触力为零，因此机器人位置 $x(t)$等于所需位置 $x_d(t)$。

机器人-环境阻抗的特性可以通过其对阶跃输入的响应来确定。这可以通过在机器人终端执行器处应用阶跃输入并测量接触力来实现。对于欠阻尼响应，阻尼固有频率 ω_d 可以使用测量接触力的快速傅里叶变换(FFT)确定。阻尼比 ζ 由稳定时间 T_s 得出，其给出稳态值 5%以内的收敛时间：

$$\exp^{-\zeta\omega_n T_s}=0.05 \tag{2.18}$$

其中 ω_n 是无阻尼频率。稳定时间由以下公式得出：

$$T_s=\frac{2.996}{\zeta\omega_n} \tag{2.19}$$

稳定时间之前的循环次数为：

$$\#\text{cycles}=\frac{2.996\sqrt{1-\zeta^2}}{2\pi\zeta} \tag{2.20}$$

阻尼比通过以下方式获得：

$$\zeta = \frac{0.4768}{\sqrt{\#\text{cycles}^2 + 0.2274}} \tag{2.21}$$

ζ是在接触力收敛到稳态误差的 5%以内之前的可见循环数。

环境刚度和阻尼可以通过 ω_d 和 ζ 得到。如图 2.5 所示，在机器人-环境系统下的等效刚度、阻尼和质量为：

$$\begin{aligned} k_{eq} &= k_d + k_e \\ b_{eq} &= b_d + c_e \\ m_{eq} &= m_d \end{aligned} \tag{2.22}$$

根据等效值，可以将固有频率和阻尼比写成

$$\begin{aligned} \omega_n &= \frac{\omega_d}{\sqrt{1-\zeta^2}} = \sqrt{\frac{k_{eq}}{m_{eq}}} \\ \text{和}\quad \zeta &= \frac{b_{eq}}{2\sqrt{k_{eq} m_{eq}}} \end{aligned} \tag{2.23}$$

利用 ω_d 和 ζ 所产生的力的响应，以及式(2.22)和式(2.23)，可以计算环境参数为：

$$\begin{aligned} k_e &= \omega_n^2 m_d - k_d \\ c_e &= 2\zeta\sqrt{(k_d + k_e) m_d} - b_d \end{aligned}$$

这种方法的主要优点是需要的数据很少。它只需测量接触力，在大多数情况下适用于机器人交互控制方案。这是一种离线方法，不需要测量环境的挠曲度和速度。其缺点是力的响应必须是欠阻尼的，以便可以获得 ω_d 和 ζ。 这意味着必须仔细选择所需阻抗值，以获得所需的性能。此外，该方法不易应用于更复杂的环境和多点接触的情况[4, 5]。