基于物理的仿真

Physics-Based Simulation

学习与中文翻译笔记 | Version 1.0.3

原著作者

Minchen Li (CMU) Chenfanfu Jiang (UCLA)
Zhaofeng Luo (CMU) Wenxin Du (UCLA) Chang Yu (UCLA)
Žiga Kovačič (Cornell) Tianyi Xie (UCLA)

原书地址

https://phys-sim-book.github.io/

译者 / 笔记整理

Fanyi Pu Gemini 4 Neon

许可协议

CC BY-NC-SA 4.0（署名-非商业性使用-相同方式共享 4.0 国际）

2026 年 11 月 22 日

目录

前言 **ii**

第一部分　基于优化的仿真 **1**

第 1 章　离散空间与时间 **2**
1.1　固体几何的表示方法 2
1.2　牛顿第二定律 4
1.3　时间积分 5
1.4　显式时间积分 6
1.4.1　前向欧拉法 6
1.4.2　辛欧拉法 6
1.5　隐式时间积分 7
1.6　小结 8

第 2 章　优化框架 **9**
2.1　优化时间积分器 9
2.2　狄利克雷边界条件 10
2.3　接触 10
2.4　摩擦 12
2.5　小结 12

参考文献 **14**

前言

本文档为开源在线教材 *Physics-Based Simulation* [1]（Version 1.0.3）的中文学习与翻译笔记。原书由 Minchen Li、Chenfanfu Jiang、Zhaofeng Luo、Wenxin Du、Chang Yu、Žiga Kovačič 与 Tianyi Xie 共同编写，在线版本发布于 `https://phys-sim-book.github.io/`。

版权与许可说明 原书 *Physics-Based Simulation* 依据 CC BY-NC-SA 4.0（知识共享署名-非商业性使用-相同方式共享 4.0 国际）许可协议发布（原书许可文件见 `https://github.com/phys-sim-book/mdbook-src/blob/main/LICENSE`）。本文档作为基于原书的中文翻译与学习笔记（演绎作品），遵循该协议的署名、非商业性使用与相同方式共享条款，同样以 CC BY-NC-SA 4.0 协议发布，仅供非商业性学习与交流使用。

> **翻译与笔记约定**
>
> - 保留常见且约定俗成的英文专业术语（如 *Implicit Euler*、*Projected Newton*、*Incremental Potential*），首次出现时使用 `\term{中文}{English}` 标注。
> - 数学符号尽量与原书保持一致，常用向量、矩阵与算子统一定义在 `macros.tex` 中。
> - 个人推导、直觉补充与疑问放入 `studynote` 环境，术语说明与勘误放入 `transnote` 环境，以便与原书正文清晰区分。

第一部分

基于优化的仿真

第 1 章

离散空间与时间

在这一章中，我们将探讨可变形固体（deformable solids）的仿真，目标是构建一个离散且可在计算机上求解的问题。本章的主要目的是引出该问题所固有的抽象代数概念。我们将从自顶向下（top-down）的架构视角切入弹性仿真，把数学建模置于核心位置。

经典弹性固体物理学的研究大多围绕偏微分方程（Partial Differential Equations, PDEs）展开。在连续介质力学与有限元分析的传统文献中，常见做法是先推导这些 PDE 的连续形式、详细阐明每一项的物理来源，再将其适配到离散的编程语言中。这种离散适配往往出现在较后的章节，让读者在前面一直处于等待状态。

然而，本书采取了不同的路线：在需要之时再将连续介质力学与 PDE 融入讨论，使这些主题均匀分布，避免一开始就让读者感到不堪重负。这种方式循序渐进地将理论与实践联系起来，从而加深理解。

我们会尽早引入核心问题的数学表述，并概述其数值求解方法。这让读者在起步阶段就能获得完整的全局视野，激发好奇心并推动后续章节的深入探索。这一策略使学习过程更加顺畅直观，帮助读者轻松地将复杂概念串联起来，快速把握学科核心。

我们的目标是对可变形固体仿真提供全面、深入且引人入胜的探讨，使之对初学该领域的学生与经验丰富的研究者都具有价值。

1.1 固体几何的表示方法

在日常生活中，固体被感知为连续的物体。但在计算机的数字世界中，我们使用离散的数值进行表示，由此衍生出一系列有趣的表示方法。

一种方法是**参数化**（parametrization）。以三维球体为例，以点 $\boldsymbol{c} \in \mathbb{R}^3$ 为球心、半径为 $r > 0$ 的球体可描述为

$$\{\boldsymbol{x} \in \mathbb{R}^3 \mid \|\boldsymbol{x} - \boldsymbol{c}\| \leq r,\ \boldsymbol{c} \in \mathbb{R}^3,\ r > 0\}.$$

这一方法可扩展到球体之外的其他几何形状，例如半空间、长方体、椭球、圆环面等，通常借助有向距离（signed distances）等函数来刻画其内部。然而，在面对现实场景中频繁出现的复杂几何体时，参数化方法往往会遇到困难。这一局限的一个新兴例外是基于神经网络的先进神经表示（neural representations），这类新方法在有效表达更复杂的几何形态方面展现出了良好前景。

另一种替代方案是**采样表示**（representing with sampling），即在物体表面及内部选取若干采样点。但仅有离散点是不够的：为了用于渲染、3D 打印等应用，我们通常需要在这些点之间建立连接关系以确定物体的边界；追踪一组采样点随时间的变化也有助于度量形变。

在连续介质力学中，物体被视为具有连续的密度场。在数字表示中，这种连续性必须以离散方式表达，通常正是通过定义固体几何的**连接关系**（connectivity）来实现。

注记 1.1.1 (其他固体表示方法). 还有其他表示固体几何的方法，例如基于体素（voxel）的方法。这类方法将空间划分为由小方盒（即体素）组成的三维网格，每个体素代表物体的一部分，类似于二维图像中的像素。基于体素的方法有若干优势。首先，它们可以作为离散的水平集（level set）表示，能够建模复杂几何并追踪其随时间的演化。每个体素都包含其相对于物体表面的位置信息，从而为连续的水平集函数提供高效的离散近似。这对于涉及表面演化、形状优化与碰撞检测的算法非常有利。其次，基于体素的方法有助于执行构造实体几何（Constructive Solid Geometry, CSG）操作。这种实体建模技术利用布尔算子将较简单的形状组合成复杂的三维模型。体素化框架允许在体素网格上直接且高效地执行并、交、差等运算，从而便于创建和修改复杂形状。

例 1.1.1 (网格). 直接用边或三角形将点连接起来以创建**网格**（mesh），是计算几何中一种常用的技术。如图 1.1.1 所示，左图与中图展示了两个不同的网格。值得注意的是，尽管这两个网格使用了相同的采样点（或节点），但它们具有不同的连接关系，因而形成了不同的形状。图中最右侧的网格展示了从一种形状到另一种形状的变换：该网格是中间网格发生形变的结果，通过将其上半部分沿竖直方向压缩而得到。

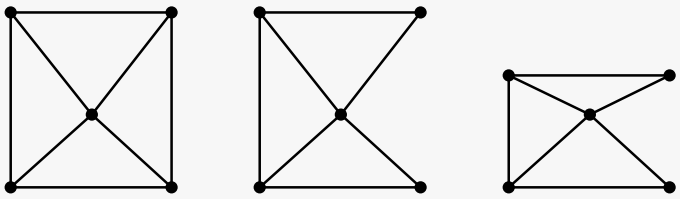


图 1.1.1: 网格

例 1.1.2 (粒子与网格). 通过在空间表示中引入均匀网格结构，我们可以在每个节点位置记录固体物质的多少。这使我们能够利用采样点计算每个网格节点处的固体密度。该方法有助于量化固体在网格中的分布，并在原始采样点之间建立连接网络。这一概念的直观示意如图 1.1.2 所示。在图中，采样点以绿色圆点表示；记录固体密度的网格节点以黑色圆圈表示；这些节点通过网格相互连接，以蓝色线段表示。

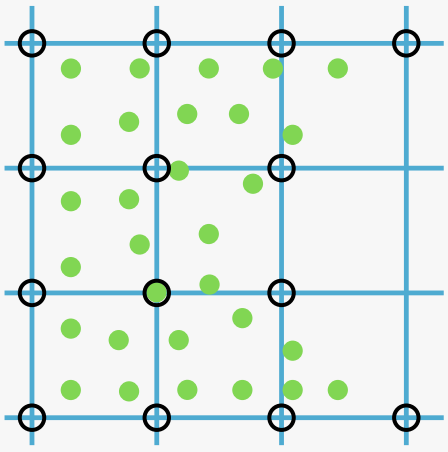


图 1.1.2: 粒子与网格

在现代固体仿真领域，上述两种定义连接关系的方法都至关重要。第一种方法通过边或

三角形构成的网格建立连接，是有限元方法（Finite Element Method, FEM）仿真器的基础；第二种方法利用均匀网格计算固体密度并建立连接，则是物质点法（Material Point Method, MPM）仿真器的核心 [2]。本书主要聚焦于前者，深入讨论 FEM 的细节。基于网格的 FEM 结构通过将复杂求解域划分为更简单的单元，能够特别有效地处理复杂区域。这使得 FEM 成为研究与仿真可变形固体的重要工具，理解其细节对从事该领域学习的人至关重要。

乍看之下，MPM 同时使用两种固体几何表示似乎显得冗余。然而，这种双重方法赋予了 MPM 显著优势，尤其是在模拟固体断裂（fracture）等动态事件时。在这类场景中，FEM 需要精细地修改定义原有连接关系的边与单元，才能准确刻画破坏过程；相比之下，MPM 能高效地处理这些情况。均匀网格能够自然地容纳断裂过程中物体各部分的分离，因为断裂节点处缺乏物质会导致相邻网格节点自动断开连接。这一特性使 MPM 擅长处理固体拓扑结构的变化。

然而，在仿真精度控制方面，有限元方法（FEM）表现更出色。FEM 直接在网格上求解，免去了持续的信息传递过程，从而保证了更高的精度。这种精度水平使 FEM 成为精确仿真可变形固体的宝贵工具，也是本书的主要侧重点。

在连接关系保持不变的前提下，将每个采样点的坐标合并为一个扩展向量 $\boldsymbol{x} \in \mathbb{R}^{dn}$，如图 1.1.3 所示，为描述特定的几何构型提供了一种有效手段。在这一表示中，d 表示空间维数（1、2 或 3），n 表示采样点的总数。类似地，每个采样点的速度、加速度和受力等属性也可以合并为相应的扩展向量，分别记为 $\boldsymbol{v}$、$\boldsymbol{a}$ 与 $\boldsymbol{f}$。这种有条理的数据组织方式不仅有助于全面理解各参数及其相互关系，也简化了仿真过程的数学表述。

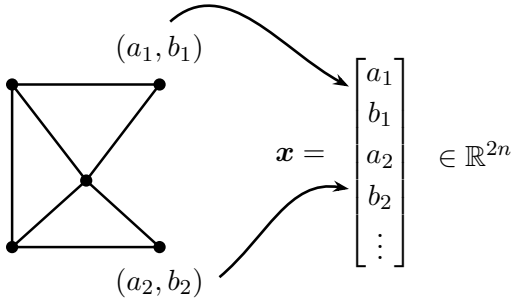


图 1.1.3: 堆叠位置向量

1.2　牛顿第二定律

在定义了固体几何在某一时刻的表示方法之后，我们现在面临的挑战是预测固体随时间的运动与形变。这一预测是实现精确仿真的关键组成部分。

牛顿第二定律写作 $\boldsymbol{f} = m\boldsymbol{a}$，表明力 $\boldsymbol{f}$ 是引起速度变化（由加速度 $\boldsymbol{a}$ 表示）的主要原因。需要理解的是，当固体的位移场超出单纯的平移、旋转或其线性组合时，即表明发生了形变。通过将牛顿第二定律应用于每一个采样点，我们可以有效地预测固体的运动与形变。这一概念可以简洁地用向量形式表示为：

$$\begin{aligned} \frac{\mathrm{d}\boldsymbol{x}}{\mathrm{d}t} &= \boldsymbol{v}, \\ M\frac{\mathrm{d}\boldsymbol{v}}{\mathrm{d}t} &= \boldsymbol{f}. \end{aligned} \tag{1.2.1}$$

在这一表示中，$M \in \mathbb{R}^{dn\times dn}$ 是质量矩阵（mass matrix），$\boldsymbol{x}$、$\boldsymbol{v}$ 和 $\boldsymbol{f}$ 分别是堆叠了位置、速度和受力的列向量。该方法将离散空间与连续时间中的运动定律结合起来，为我们的可变形

固体仿真奠定了基础。

注记 1.2.1 (堆叠变量). 尽管理论上质量矩阵 M 不一定是对角矩阵，但在实际应用中常将其简化为对角矩阵。由此得到集中质量矩阵（lumped mass matrix），它表示一个离散质点系统，为处理复杂系统提供了高效途径。下面以二维空间中的两质点系统为例加以说明。该系统的集中质量矩阵表示为：

$$M = \begin{pmatrix} m_1 & & & \\ & m_1 & & \\ & & m_2 & \\ & & & m_2 \end{pmatrix},$$

这里我们假设向量 $\boldsymbol{v}$（以及 $\boldsymbol{x}$ 和 $\boldsymbol{f}$）按特定顺序堆叠：

$$\boldsymbol{v} = (v_{11}, v_{12}, v_{21}, v_{22})^{\mathsf{T}},$$

其中 $v_{i\alpha}$ 表示第 i 个质点的速度 $\boldsymbol{v}_i$ 的第 α 个分量。这种有序的结构显著简化了计算，也有助于理解系统的动力学特性。

1.3　时间积分

牛顿第二定律为一组连续形式的常微分方程（Ordinary Differential Equations, ODEs）奠定了基础。这与前文在空间上使用采样点离散表示连续几何的做法类似；现在我们在时间维度上采取同样的方法。通过对时间点进行采样，我们可以有效地表示时间导数 $\frac{\mathrm{d}\boldsymbol{x}}{\mathrm{d}t}$ 与 $\frac{\mathrm{d}\boldsymbol{v}}{\mathrm{d}t}$。

定义 1.3.1 (时间积分). 将时间离散为固定长度的小区间时，记第 n 步的时刻为 t^n，通常称为一个**时间步**（timestep）。该区间的长度，即**时间步长**（timestep size），由 $\Delta t = t^{n+1} - t^n$ 给出。时间步计数 n 通常是从零开始的整数，使得 $t^0 = 0\,\mathrm{s}$ 成为仿真的起始点。

时间步概念的引入使我们能够用符号 $\boldsymbol{x}^n$、$\boldsymbol{v}^n$ 和 $\boldsymbol{f}^n$ 分别表示第 n 个时间步上各节点的位置、速度和受力。术语**时间步进**（timestepping）或时间积分（time integration），指的是在每个递增的时间步 $n = 0, 1, 2, \ldots$ 上，由 $\boldsymbol{x}^n, \boldsymbol{v}^n$ 计算 $\boldsymbol{x}^{n+1}, \boldsymbol{v}^{n+1}$ 的过程。作为直观展示，考虑图 1.3.1 中的犰狳（Armadillo）弹弓动画：从左到右，动画中的每一帧都是逐步计算得到的。在这种语境下，时间步进如同逐帧推进的电影，一步步展现出系统不断演化的动力学过程。

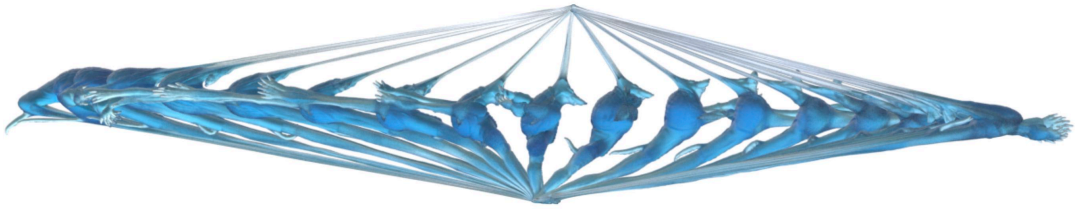


图 1.3.1: 犰狳弹弓逐帧动画

在本书及我们所考察的仿真场景中，必须强调一项关键假设：在仿真开始时，我们总是

精确已知初始值 $\boldsymbol{x}^0$ 和 $\boldsymbol{v}^0$。此外，对于每个时间步，我们要么有办法基于物理模型计算 $\boldsymbol{f}^n$，要么直接已知其精确值（例如重力这类恒定外力）。这一假设是我们方法的基础，确保仿真建立在已知的初始条件与受力之上，从而获得更准确可靠的结果。

1.4　显式时间积分

显式时间积分（explicit time integration）格式通过将已知量代入简单公式直接计算 $\boldsymbol{x}^{n+1}$ 和 $\boldsymbol{v}^{n+1}$，这正是它们被称为显式（explicit）的原因。本节重点介绍两种基础的显式格式：前向欧拉法（Forward Euler）与辛欧拉法（Symplectic Euler）。

1.4.1　前向欧拉法

为了将连续时间系统转化为离散形式，我们采用向前差分（forward difference）近似。在这种近似中，导数 $\left(\frac{\mathrm{d}\boldsymbol{x}}{\mathrm{d}t}\right)^n$ 估计为 $\frac{\boldsymbol{x}^{n+1}-\boldsymbol{x}^n}{\Delta t}$，同理 $\left(\frac{\mathrm{d}\boldsymbol{v}}{\mathrm{d}t}\right)^n$ 估计为 $\frac{\boldsymbol{v}^{n+1}-\boldsymbol{v}^n}{\Delta t}$。上标 n 表示第 n 个时间步的状态变量。因此，系统的离散形式表示为：

$$\begin{aligned}\frac{\boldsymbol{x}^{n+1}-\boldsymbol{x}^n}{\Delta t}&=\boldsymbol{v}^n,\\ M\frac{\boldsymbol{v}^{n+1}-\boldsymbol{v}^n}{\Delta t}&=\boldsymbol{f}^n.\end{aligned}\tag{1.4.1}$$

假设质量不随时间变化，这组方程为更新状态变量提供了清晰的机制：已知第 n 个时间步的当前值 $\boldsymbol{x}^n$、$\boldsymbol{v}^n$ 和 $\boldsymbol{f}^n$，便可以直接确定下一时间步 $n+1$ 的值。

方法 1.4.1（牛顿第二定律的前向欧拉时间积分）. 在前向欧拉法中，下一时间步 $n+1$ 的状态变量 $\boldsymbol{x}^{n+1}$ 和 $\boldsymbol{v}^{n+1}$ 根据当前值 $\boldsymbol{x}^n$ 和 $\boldsymbol{v}^n$ 计算得到，更新规则为：

$$\begin{aligned}\boldsymbol{x}^{n+1}&=\boldsymbol{x}^n+\Delta t\boldsymbol{v}^n,\\ \boldsymbol{v}^{n+1}&=\boldsymbol{v}^n+\Delta tM^{-1}\boldsymbol{f}^n.\end{aligned}\tag{1.4.2}$$

这里，Δt 表示时间步长，M 是质量矩阵，$\boldsymbol{f}^n$ 是当前时间步 n 的受力。

前向欧拉法被认为是无条件不稳定（unconditionally unstable）的，这意味着无论所选的时间步长 Δt 多么小，对于 $\boldsymbol{f}$ 非恒定的方程，数值解最终都会显著增长（发散爆炸），而精确解本身不受影响（参见图 1.4.1 左图）。

1.4.2　辛欧拉法

如果在位置导数的离散中给 $\boldsymbol{v}$ 加上上标 $n+1$，同时保持速度导数的离散形式不变，我们就会得到一种新的更新规则：

方法 1.4.2（牛顿第二定律的辛欧拉时间积分）. 给定当前状态变量、质量矩阵以及从 t^n 到 t^{n+1} 的时间步长，有：

$$\begin{aligned}\boldsymbol{x}^{n+1}&=\boldsymbol{x}^n+\Delta t\boldsymbol{v}^{n+1},\\ \boldsymbol{v}^{n+1}&=\boldsymbol{v}^n+\Delta tM^{-1}\boldsymbol{f}^n,\end{aligned}\tag{1.4.3}$$

其中 $n=0,1,2,\ldots$。

仅这一处微小改动，积分便成为**条件稳定**（conditionally stable）的。这意味着只要 Δt 保持在与具体问题相关的上限之内，就能有效限制解的数值误差。此外，辛欧拉法还展现出保持系统能量的良好特性，如图 1.4.1 中图所示。

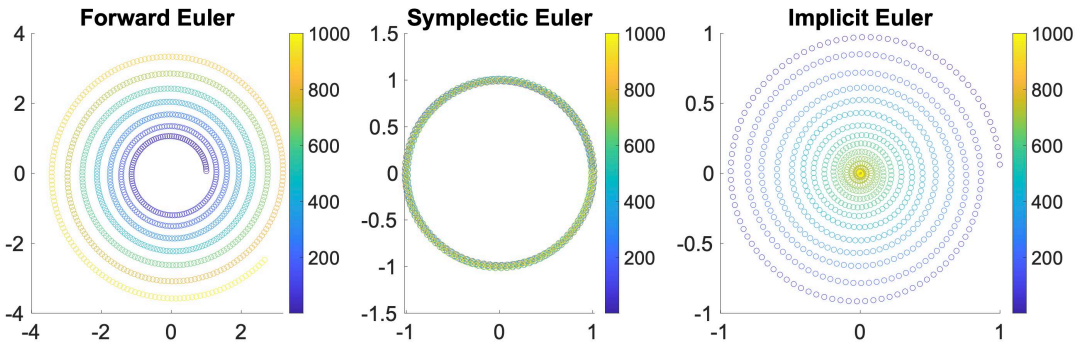


图 1.4.1: 时间积分器的稳定性。图示展示了一个做匀速圆周运动的质点，从左到右分别使用前向欧拉法、辛欧拉法和隐式欧拉法模拟，颜色变化表示时间的推进。每种方法在仿真中都呈现出显著不同的特征：前向欧拉法最终发生不稳定发散，辛欧拉法紧密贴合理论轨迹，而隐式欧拉法在保持稳定的同时逐渐使运动停止。

1.5　隐式时间积分

与显式时间积分不同，隐式时间积分（implicit time integration）需要求解方程组来确定 $\boldsymbol{x}^{n+1}$ 和 $\boldsymbol{v}^{n+1}$ 的值。这种方法的一个显著优势在于其有望大幅提升稳定性。下面介绍最简单的隐式积分形式——后向欧拉法（Backward Euler）。

方法 1.5.1 (牛顿第二定律的后向欧拉时间积分). 给定当前状态变量、质量矩阵以及从 t^n 到 t^{n+1} 的时间间隔，更新规则如下：

$$\begin{aligned} \boldsymbol{x}^{n+1} &= \boldsymbol{x}^n + \Delta t \boldsymbol{v}^{n+1}, \\ \boldsymbol{v}^{n+1} &= \boldsymbol{v}^n + \Delta t M^{-1} \boldsymbol{f}^{n+1}, \end{aligned} \tag{1.5.1}$$

其中 n 取值为 $0, 1, 2, \ldots$。

在本书讨论的许多场景中，力由位置向量 $\boldsymbol{x}$ 导出，因此可以表示为 $\boldsymbol{f}^{n+1} = \boldsymbol{f}(\boldsymbol{x}^{n+1})$。需要认识到，$\boldsymbol{x}^{n+1}$ 的更新依赖于已知的 $\boldsymbol{v}^{n+1}$，但 $\boldsymbol{v}^{n+1}$ 的计算又取决于 $\boldsymbol{x}^{n+1}$。这种相互依赖造成了循环耦合，必须通过求解方程组来准确求出 $\boldsymbol{x}^{n+1}$ 和 $\boldsymbol{v}^{n+1}$。将 $\boldsymbol{v}^{n+1} = (\boldsymbol{x}^{n+1} - \boldsymbol{x}^n)/\Delta t$ 代入，式 (1.5.1) 可改写为：

$$M\big(\boldsymbol{x}^{n+1} - (\boldsymbol{x}^n + \Delta t \boldsymbol{v}^n)\big) - \Delta t^2 \boldsymbol{f}(\boldsymbol{x}^{n+1}) = \boldsymbol{0}. \tag{1.5.2}$$

鉴于力 $\boldsymbol{f}$ 通常关于位置 $\boldsymbol{x}$ 呈现非线性，式 (1.5.2) 一般是非线性的，需要使用**牛顿法**（Newton's method）等非线性求根技术来求解。

方法 1.5.2 (用于后向欧拉时间积分的牛顿法). 如下方算法 1.5.1 所述，牛顿法是一种从解的初始估计 $\boldsymbol{x}^i$ 出发的迭代技术。在当前迭代点 $\boldsymbol{x}^i$ 处，它对力做线性近似 $\boldsymbol{f}(\boldsymbol{x}^{n+1}) \approx \boldsymbol{f}(\boldsymbol{x}^i) + \nabla \boldsymbol{f}(\boldsymbol{x}^i)(\boldsymbol{x}^{n+1} - \boldsymbol{x}^i)$，然后求解一个线性方程组并更新迭代点。重复

这一过程，直至达到令人满意的收敛程度。

算法 1.5.1: 用于后向欧拉时间积分的牛顿法

Result: $\boldsymbol{x}^{n+1}$, $\boldsymbol{v}^{n+1}$

$\boldsymbol{x}^i \leftarrow \boldsymbol{x}^n$;

while $\|M(\boldsymbol{x}^i - (\boldsymbol{x}^n + \Delta t \boldsymbol{v}^n)) - \Delta t^2 \boldsymbol{f}(\boldsymbol{x}^i)\| > \epsilon$ **do**

  solve $M(\boldsymbol{x} - (\boldsymbol{x}^n + \Delta t \boldsymbol{v}^n)) - \Delta t^2 \big(\boldsymbol{f}(\boldsymbol{x}^i) + \nabla \boldsymbol{f}(\boldsymbol{x}^i)(\boldsymbol{x} - \boldsymbol{x}^i)\big) = \boldsymbol{0}$ for $\boldsymbol{x}$;

  $\boldsymbol{x}^i \leftarrow \boldsymbol{x}$;

$\boldsymbol{x}^{n+1} \leftarrow \boldsymbol{x}^i$;

$\boldsymbol{v}^{n+1} \leftarrow (\boldsymbol{x}^{n+1} - \boldsymbol{x}^n)/\Delta t$;

尽管后向欧拉法即使对于较大的 Δt 也能保证**无条件稳定**（unconditional stability），但必须认识到，增大 Δt 可能会导致系统的条件数变差，从而使线性方程组的求解更加困难。此外，力的线性化本质上只是一种近似：如果解的初始估计距离真实解较远，标准牛顿法迭代可能无法收敛，甚至可能发散。

在后续讨论中，我们将介绍一种改进版的牛顿法。该改进方法旨在针对特定类型的问题保证收敛性，而不论初始估计如何或 Δt 的大小如何。

1.6　小结

仿真固体涉及预测其位置与形状随时间的变化。要在计算机上实现这一点，几何与时间都必须以离散方式表示。

几何通常使用以特定方式相互连接的采样点来表示：

- **有限元方法（Finite Element Methods, FEM）**通过非结构网格连接采样点；
- **物质点法（Material Point Methods, MPM）**利用均匀笛卡尔网格连接采样点。

FEM 擅长提供高精度结果，而 MPM 在处理拓扑变化方面具有优势。本书主要聚焦于 FEM。

时间被离散为若干离散时刻，并结合牛顿第二定律，采用有限差分法计算物理量的时间导数。

前向欧拉法由于无条件不稳定通常避免使用。相比之下，辛欧拉法是显式且条件稳定的，常用于良态问题。对于刚性（stiff）问题，虽然后向欧拉法计算开销较大且可能出现数值不稳定，但因其无条件稳定且仅需配合非线性方程组求解，仍然是被广泛采用的选择。

在下一章中，我们将探讨隐式时间积分的优化视角，为稳健求解这类问题提供方法。

第 2 章

优化框架

2.1 优化时间积分器

在后向欧拉法中，每个时间步都需要求解一个非线性方程组，如式 (1.5.2) 所示。实际上，这等价于求解如下优化问题：

$$\begin{aligned} \boldsymbol{x}^{n+1} &= \arg\min_{\boldsymbol{x}} E(\boldsymbol{x}), \\ \text{其中} \quad E(\boldsymbol{x}) &= \frac{1}{2}\|\boldsymbol{x}-\tilde{\boldsymbol{x}}^n\|_M^2 + \Delta t^2 P(\boldsymbol{x}). \end{aligned} \tag{2.1.1}$$

这里，$\tilde{\boldsymbol{x}}^n = \boldsymbol{x}^n + \Delta t \boldsymbol{v}^n$，$\frac{1}{2}\|\boldsymbol{x}-\tilde{\boldsymbol{x}}^n\|_M^2 = \frac{1}{2}(\boldsymbol{x}-\tilde{\boldsymbol{x}}^n)^\mathsf{T} M(\boldsymbol{x}-\tilde{\boldsymbol{x}}^n)$ 表示惯性项（inertia term），$P(\boldsymbol{x})$ 表示力 $\boldsymbol{f}(\boldsymbol{x})$ 对应的势能（potential energy），满足 $\frac{\partial P}{\partial \boldsymbol{x}}(\boldsymbol{x}) = -\boldsymbol{f}(\boldsymbol{x})$，而 $E(\boldsymbol{x})$ 被称为**增量势**（Incremental Potential）。在 $E(\boldsymbol{x})$ 的局部极小值处，有 $\frac{\partial E}{\partial \boldsymbol{x}}(\boldsymbol{x}^{n+1}) = \boldsymbol{0}$，恰好对应式 (1.5.2)。

将时间积分视为一个优化问题，使我们能够运用成熟的优化方法稳健地求得解，同时也为建模更复杂的物理现象提供了一个统一的框架。

定义 2.1.1 (保守力). 若存在势能 $P(\boldsymbol{x})$ 使得力 $\boldsymbol{f}(\boldsymbol{x})$ 满足 $\frac{\partial P(\boldsymbol{x})}{\partial \boldsymbol{x}} = -\boldsymbol{f}(\boldsymbol{x})$，则称 $\boldsymbol{f}(\boldsymbol{x})$ 为**保守力**（conservative forces）。常见的弹性力以及重力等体力都是保守力的例子。只需将相应的势能项加入增量势中，即可将它们轻松纳入优化框架。

注记 2.1.1 (重力). 作用在质量为 m、位于地球表面上方高度 h 处物体上的重力（表示为力 $\boldsymbol{F} = -mg\boldsymbol{z}$，其中 g 为重力加速度，$\boldsymbol{z}$ 为指向正上方的单位向量）对应于重力势能 $U = mgh$。这里，U 是克服重力将物体从参考点（$h = 0$ 处）移动至高度 h 所做的功。力是能量关于位置的负梯度（数学上写作 $\boldsymbol{F} = -\nabla U$），这印证了能量守恒原理。将 U 关于 h 求导，可得 $\nabla U = mg\boldsymbol{z}$，从而 $\boldsymbol{F} = -\nabla U = -mg\boldsymbol{z}$，与我们最初给出的力的表达式一致。

注记 2.1.2 (弹性). 弹性（elasticity）是固体在外力作用下保持其静止形状的能力。在弹性的作用下，同一固体上的采样点在仿真过程中会彼此绑定在一起。越刚硬的固体具有越高的弹性能量刚度，在相同形变程度下会提供更大的弹性力，从而有助于恢复静止形状。犰狳弹弓的例子（图 1.3.1）展示了典型的弹性效应。弹性是所有固体的共同属性，无论其几何形态如何，也无论在直觉上属于刚体还是非刚体，例如金属、木材、软组织、橡胶、布料、毛发、沙子等。

2.2　狄利克雷边界条件

势能并非建模物理现象的唯一手段，约束同样至关重要。让我们先从最简单的形式——**线性等式约束**（linear equality constraints）开始。该约束优化问题定义如下：

$$\min_{\boldsymbol{x}} E(\boldsymbol{x}) \quad \text{s.t.} \quad A\boldsymbol{x} = \boldsymbol{b}, \tag{2.2.1}$$

其中 $A \in \mathbb{R}^{m\times dn}$ 和 $\boldsymbol{b} \in \mathbb{R}^m$ 表示 m 个线性等式约束。

在仿真过程中，我们常常需要在每个时间步控制固体上某些点的位置。这可能涉及固定一组节点以建模不可移动的物体（如地面或障碍物），或者通过让特定节点沿预定轨迹运动来引导固体的运动。例如，在弹弓场景（图 1.3.1）中，犰狳的脚部与耳部保持静止。这类控制被称为**狄利克雷边界条件**（Dirichlet boundary conditions, BC）。在优化时间积分器框架中，这些条件可以表示为线性等式约束。

具体而言，式 (2.2.1) 中的矩阵 A 通常是一个 $m \times dn$ 矩阵（满足 $m \le dn$ 且 $m \bmod d = 0$），用于选取边界条件点的坐标；相应地，$\boldsymbol{b}$ 是一个 $m \times 1$ 向量，用于定义指定的目标位置。通过求解该优化问题，所选点便被固定在这些指定位置上，且这些位置可以随时间步变化。

在式 (2.2.1) 的局部极小值处，满足 **KKT 条件**：

$$\begin{aligned} \nabla E(\boldsymbol{x}) - A^{\mathsf{T}}\boldsymbol{\lambda} &= \boldsymbol{0}, \\ A\boldsymbol{x} &= \boldsymbol{b}, \end{aligned} \tag{2.2.2}$$

其中 $\boldsymbol{\lambda} \in \mathbb{R}^m$ 表示**拉格朗日乘子**（Lagrange multiplier）向量，包含所有的拉格朗日乘子。

注记 2.2.1 (求解 KKT 系统). 带等式约束的非线性优化问题可以通过直接求解如式 (2.2.2) 所示的非线性 KKT（Karush-Kuhn-Tucker）系统来解决，通常采用牛顿法等方法。然而，这种做法的计算开销可能很大。针对边界条件，我们可以利用矩阵 A 的特殊结构，以无约束的方式求解这一约束问题。相关技术将在后续章节中介绍。

2.3　接触

为了准确仿真固体，必须确保它们不会相互穿透（interpenetrate），如下方图 2.3.1 左侧所示。一种有效方法是对时间步长施加 **CFL**（Courant-Friedrichs-Lewy 条件）上限，这在 MPM 等方法中尤为常见。而在有限元方法（FEM）中，这需要对接触力进行精确建模。然而，精确建模接触充满挑战：接触本质上是一个非光滑过程，在固体相互碰触的瞬间突然发生，并不存在某种势能形式能够精确刻画这一现象。

在实际应用中，判断两个物体是否发生碰撞通常依赖于直观评估它们之间的接近程度：当两者之间的距离不为零时，说明它们之间仍有空隙，尚未发生碰撞。这一概念在计算环境中建模物体交互时至关重要。

为避免碰撞或穿透，我们可以确保运动物体表面之间的距离永远不会减小到零。这一思路在计算仿真的时间积分问题中尤为有用。我们采用不等式约束来建模这一情形；将其与边界条件结合后，时间积分问题可表述如下：

$$\min_{\boldsymbol{x}} E(\boldsymbol{x}) \quad \text{s.t.} \quad A\boldsymbol{x} = \boldsymbol{b} \text{ 且 } \forall k,\ c_k(\boldsymbol{x}) \ge \epsilon. \tag{2.3.1}$$

这里，c_k 度量固体表面上特定成对区域之间的距离，$\epsilon \to 0$ 是一个极小的正数，用以保证 $c_k(\boldsymbol{x})$ 严格为正。



图 2.3.1: 接触与摩擦的仿真示例。左图展示了一段有趣的仿真：四个角色落入漏斗后被一个运动平面挤压排出。在整个复杂交互过程中完全没有出现任何穿透，完美体现了所用模型的精度。右图展示了在不同拉动速度下对经典抽桌布魔术的仿真。当桌布被抽走而桌面餐具不受干扰时——正如现实生活中所预期的那样——该仿真的真实感（尤其是对摩擦的精确刻画）便一目了然。这些仿真展现了当代计算模型在模拟接触方面的强大能力与精度，将抽象的物理概念生动而引人入胜地呈现出来。

在式 (2.3.1) 的局部极小值处，满足如下 Karush-Kuhn-Tucker（KKT）条件：

$$\begin{aligned} \nabla E(\boldsymbol{x}) - A^{\mathsf{T}}\boldsymbol{\lambda} - \sum_k \gamma_k \nabla c_k(\boldsymbol{x}) &= \mathbf{0}, \\ A\boldsymbol{x} &= \boldsymbol{b}, \\ \forall k,\ c_k(\boldsymbol{x}) - \epsilon \geq 0,\ \gamma_k \geq 0,\ \gamma_k\big(c_k(\boldsymbol{x}) - \epsilon\big) &= 0. \end{aligned} \tag{2.3.2}$$

在该条件中，γ_k 是约束 $c_k(\boldsymbol{x}) \geq \epsilon$ 对应的拉格朗日乘子。具体来说，$\nabla c_k(\boldsymbol{x})$ 指向第 k 对接触区域的接触力方向，将该方向与 γ_k 所表示的大小相结合，便得到该点处实际的接触力。

注记 2.3.1 (互补松弛条件). 互补松弛条件（complementarity slackness condition）$\gamma_k(c_k(\boldsymbol{x}) - \epsilon) = 0$ 起着关键作用：它确保接触力仅在固体相互接触（$c_k(\boldsymbol{x}) = \epsilon$）时才存在（$\gamma_k \neq 0$）；相反，当固体未接触（$c_k(\boldsymbol{x}) > \epsilon$）时，不应存在接触力（$\gamma_k = 0$）。

定义 2.3.1 (有效集). 在带有如下不等式约束的优化问题中：

$$\forall k,\ c_k(\boldsymbol{x}) \geq 0,$$

有效集（active set）定义为

$$\{l \mid c_l(\boldsymbol{x}^*) = 0\},$$

其中 $\boldsymbol{x}^*$ 为该问题的一个局部最优解。

注记 2.3.2 (组合困难性). 互补松弛条件揭示出：在最优解处，只有属于有效集的约束才会具有非零的拉格朗日乘子 γ_k。这意味着，与等式约束不同，不等式约束不仅要求解出拉格朗日乘子的数值，还需要辨识哪些 γ_k 应当取 0。这带来了组合意义上的困难性。

针对带不等式约束的优化问题，已有多种技术可供选择。每种方法都引入了不同的思路，有效针对问题的不同侧面：

- **原始-对偶方法（Primal-Dual Methods）：**这类方法同时求解原始问题（原优化问题）及其对偶问题。对偶问题往往能为原始问题的解提供有价值的洞察，因此该方法颇具吸引力。这类方法通过迭代，利用原始问题与对偶问题之间的联系逐步改进初始解。然而，设计与实现原始-对偶算法可能相当复杂，需要在两类问题之间精细平衡。尽管有效，但对于复杂的高维问题，这类方法未必高效或直观。
- **投影最速下降法（Projected Steepest Descent Methods）：**这是经典最速下降法针对约束问题的改进。在每次迭代中，算法沿最速下降方向移动，若因约束而偏离可行域，则将其投影回可行集。该方法因简单直观而广受欢迎，但在面对收敛缓慢的病态问题或难以投影的约束时可能会陷入困境。
- **内点法（Interior-Point Methods）：**又称障碍法（barrier methods），这类技术引入一个障碍函数来惩罚不可行解，从而将解引导至可行域内部。这种方法有效地将约束问题转化为无约束问题，使其可用常规技术求解。然而，障碍函数的选择会显著影响方法的性能。尽管对特定类型的问题非常高效，但当可行域难以定义或缺乏简单内部时，这类方法可能会表现不佳。

尽管上述每种方法各有优劣，但我们的主要关注点将放在一种稳健且精确的接触建模方法上，即**增量势接触**（Incremental Potential Contact, IPC）。IPC 的独特之处在于利用光滑势能来近似接触过程。这一转化有效地将问题变为无约束问题，从而便于应用各种高效且稳健的优化技术。IPC 的一项核心特性是能够将相对于非光滑表述的近似误差控制在预设界限之内。这一特点为该方法增添了稳健性与可靠性，使其成为解决当前问题极具前景的方案。

2.4　摩擦

摩擦（friction）是涉及运动的物理交互中的关键要素，常常会显著影响仿真结果。因此，对其精确建模对于实现真实可靠的仿真至关重要。可参见图 2.3.1 右侧，它展示了一个需要精确刻画摩擦的场景。

应用最广泛的摩擦模型之一是**库仑摩擦模型**（Coulomb Friction model）。该模型建立在**最大耗散原理**（Maximal Dissipation Principle, MDP）之上，能够有效刻画静摩擦与动摩擦之间的非光滑转变。静摩擦（static friction）是阻止物体开始运动的力，而动摩擦（dynamic friction，又称滑动摩擦 kinetic friction）则阻碍运动物体的运动。库仑摩擦模型精确地描述了这两种摩擦类型之间的关键转变。

在标准物质点法（MPM）中，摩擦由背景网格隐式建模。然而，这种做法存在缺点，尤其是摩擦系数不可控且大得不切实际。

在有限元方法（FEM）中，可以通过增量势接触（IPC）模型中的近似势能，更真实且可控地表示摩擦。这与我们的优化时间积分框架高度契合。通过使用势能来近似摩擦，我们不仅保持了仿真的稳健性，还获得了对摩擦模型精度的控制能力。

在后续章节中，我们将深入探讨 IPC 模型中用于表示摩擦力的具体技术与方法，以及它们在提升仿真精度与真实感方面的作用。

2.5　小结

至此，我们讨论的目标一直在于为无条件稳定的隐式时间积分问题设计可靠的求解方案，旨在解决由截断误差引发的迭代不收敛问题。为此，我们将时间积分问题重新表述为一个极小化问题。这一表述不仅使我们能够应用成熟的优化技术，也为不同物理现象提供了统一的建模框架。

以下简要总结该框架内用于建模各类物理现象的技术：

- 对于重力和弹性等保守力，我们使用势能，并将其整合进目标函数中，以准确表征所涉及的力。
- 用于指定系统约束的边界条件采用简单的线性等式约束建模，从而在仿真过程中将系统限制在可行状态内。
- 为防止固体在仿真过程中相互穿透，我们使用不等式约束来建模接触与摩擦。这些约束确保物体保持自身的物理完整性，并在相互接触时表现出符合预期的行为。

这里值得注意的一点是，我们可以利用边界条件的独特结构，以无约束的方式施加等式约束，这将显著降低计算复杂度。

此外，我们还介绍了增量势接触（IPC）方法的概念。IPC 方法以可控的精度将接触与摩擦建模为光滑势能，从而确保对固体对象进行稳健、精确且无穿透的仿真。

接下来，在下一章中，我们将深入学习用于求解无约束优化问题的投影牛顿法（projected Newton method）。该方法的优势在于具备全局收敛性，即只要初始构型可行，无论初始状态如何，方法都能保证收敛。这一特性对于复杂仿真极具价值，有助于使整个方法更加稳健可靠。

参考文献

[1] Minchen Li et al. *Physics-Based Simulation*. Open-source online book. Live version available at `https://phys-sim-book.github.io/`. Mar. 2026. DOI: `10.5281/zenodo.20597655`. URL: `https://doi.org/10.5281/zenodo.20597655`.

[2] Chenfanfu Jiang et al. "The Material Point Method for Simulating Continuum Materials". In: *ACM SIGGRAPH 2016 Courses*. 2016.