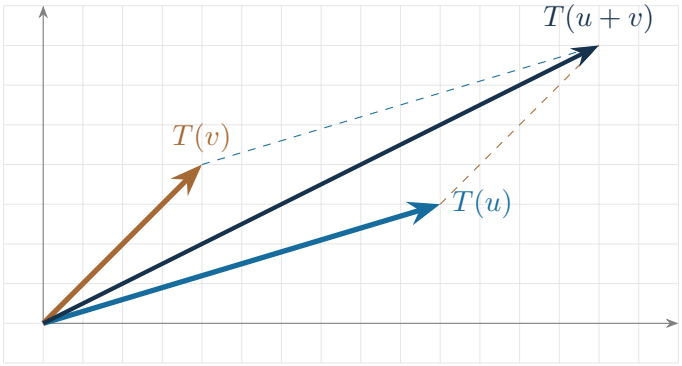
数学直觉 ⟶ 严格命题 ⟶ 机器校验

3Blue1Brown

中文导学教材

引导论述 · 动画实验 · Lean 4



6 组课程　66 个条目　从直觉到推导与验证

目录快照：2026 年 10 月 24 日

独立学习笔记；非 3Blue1Brown 官方出版物。

阅读说明

每节先提出一个具体问题，再用连续论述解释为什么引入概念、怎样选择表示、如何逐步得到公式。正文之后的要点回看、算例与自测用于整理和检查理解。建议先遮住答案独立预测，再观看动画或改变实验参数，最后核对 Lean 命题中的假设；不要让观看替代推导。

配套网页提供原课链接、61 个原视频播放器、9 组原创 Manim 动画和 18 类可操作实验。66 节各增加三段逐步讲解，共 198 段；宽屏页面左侧阅读、右侧操作，每一步可直接复位到对应演示情形。另 5 个目录条目未单列视频，直接链接官方图文与图示。PDF 共用逐步讲解与 18 组新增的定义、命题和纸笔证明，放入动画关键帧以及各课的原课、网页和实验入口；静态 PDF 不直接播放动画。网页地址：https://vitael04213.github.io/3blue1brown-notes/。观课提示定位应关注的概念；未核实的时间点不作标注。

本书按官网五个课程型主题目录整理：线性代数、微积分、神经网络、微分方程及 Lockdown Math；另外纳入有明确集数的“压缩与智能”两集。官网把补充视频也放入这些目录，因此包含散度与旋度、拉普拉斯变换、扩散模型等扩展课。交叉熵同时属于两组，只写一次正文，并在另一处交叉引用。公告页作为学习导读计入，不人为编造数学定理。

范围为该日期已经公开的上述课程序列，共 67 个目录位置、66 节独立课。未把官网“全部”中的每一个独立专题、演讲或外部频道视频算作课程，也不包含未来更新。逐条英文题名、发布日期、原课和视频链接见仓库的 `sources/catalog.json` 与 `COURSES.md`。

这些内容是围绕课程主题重新组织的中文伴读笔记，含原创计算例与练习，不是字幕全文翻译。对于官网提供正文的课程，参照其概念组织；对于只有标题、简介和视频入口的页面，依据已核实的主题独立推导标准数学内容。没有声称逐帧看完视频。页面获取记录见 `sources/retrieval.json`。公式中的实数是精确实数，数值近似会显式使用 $\approx$。

怎样理解 Lean 校验

每节末尾列出一个可在仓库中查找的 Lean 定理，并明确校验边界。函数可导、分母非零、概率为正、积分存在等条件会写进定理签名。命名空间统一加前缀 `ThreeBlueOneBrown`；例如 `Calculus.power_rule` 位于 `ThreeBlueOneBrown/Calculus.lean`。

有些 Lean 结果直接复用 mathlib 中已经证明的一般定理，有些是本书的坐标推导或具体例子。通过编译的含义是这些形式命题在 Lean 与 mathlib 的基础下成立；它不意味着视频的一切解释、统计经验或未形式化的分析结论都已经验证。没有使用 `sorry`、`admit`、自设公理或 `native_decide`。精确的形式化覆盖表见 `FORMALIZATION.md`。

先修路线：基础数学，然后线性代数与微积分，再读神经网络和微分方程；信息论适合在概率基础之后阅读。

目录

阅读说明 **i**

第一章　线性代数：把矩阵看成动作 **1**

1.1　向量究竟是什么 1
1.2　线性组合、张成与基 3
1.3　线性变换与矩阵 4
1.4　矩阵乘法就是复合 7
1.5　三维线性变换 8
1.6　行列式：有向体积倍率 10
1.7　逆矩阵、列空间与零空间 11
1.8　非方阵与维度变化 13
1.9　点积与对偶 14
1.10　叉积与有向面积 16
1.11　从线性测量重新认识叉积 18
1.12　克拉默法则 20
1.13　换基：改变坐标语言 21
1.14　特征向量与特征值 22
1.15　快速求二维特征值 24
1.16　抽象向量空间 26

第二章　微积分：局部变化与累积 **29**

2.1　从面积发现微积分 29
2.2　导数的“零除以零”悖论 30
2.3　幂法则的几何来源 33
2.4　三角函数的导数 34
2.5　乘积法则与链式法则 36

2.6 为什么自然常数 e 特别 . . . 37
2.7 隐函数求导 . . . 39
2.8 极限与导数的定义 . . . 40
2.9 把任意精度写清楚 . . . 41
2.10 洛必达法则：比较消失速度 . . . 43
2.11 积分与微积分基本定理 . . . 45
2.12 面积为什么与斜率相连 . . . 47
2.13 高阶导数 . . . 48
2.14 泰勒多项式与余项 . . . 50
2.15 泰勒展开的面积解释 . . . 52
2.16 导数作为局部变换 . . . 54
2.17 散度与旋度 . . . 55

第三章 神经网络：从像素到序列 58
3.1 神经网络的基本单元 . . . 58
3.2 梯度下降如何学习 . . . 59
3.3 检查网络究竟学到了什么 . . . 62
3.4 反向传播的直觉 . . . 63
3.5 反向传播的微积分 . . . 65
3.6 语言模型的简明图景 . . . 67
3.7 Transformer 的整体结构 . . . 68
3.8 注意力：按上下文加权信息 . . . 70
3.9 前馈层如何表达特征与关联 . . . 72
3.10 扩散模型：从噪声中逐步恢复 . . . 74

第四章 微分方程：从局部规律到整体演化 77
4.1 微分方程与向量场 . . . 77
4.2 偏微分方程与热扩散 . . . 79
4.3 用基本模式求解热方程 . . . 80
4.4 傅里叶级数：旋转与叠加 . . . 83
4.5 复指数与旋转的物理图像 . . . 84
4.6 拉普拉斯变换：带衰减的频率探测 . . . 86
4.7 把微分方程变成代数方程 . . . 87
4.8 动态理解欧拉恒等式 . . . 89

4.9 矩阵指数与线性动力系统 91
4.10 指数增长与饱和 92

第五章 Lockdown Math：重新组织基础数学 95

5.1 课程公告与学习路线 95
5.2 更简洁的二次公式 96
5.3 三角函数从单位圆出发 98
5.4 复数：缩放与旋转 99
5.5 欧拉公式在说什么 101
5.6 虚数利率与连续旋转 102
5.7 对数的基础结构 104
5.8 自然对数为何自然 105
5.9 幂塔与不动点 107
5.10 为什么 i 的 i 次方可以是实数 108
5.11 解题：找表示、找不变量、查边界 110

第六章 压缩与智能：熵和交叉熵 113

6.1 熵：平均要问多少个二选一问题 113
6.2 交叉熵：错误概率模型的编码代价 115

来源与复现 118

第一章　线性代数：把矩阵看成动作

本章固定列向量约定：Av 表示先给定向量 v，再作用矩阵 A；ABv 中 B 先作用。几何图像帮助发现结论，坐标与线性映射负责精确定义。

1.1　向量究竟是什么

官方课程：vectors

先问：我们究竟想记录什么？

假设你在地图上先向东走两格，再向南走一格。别人从另一个地点照做，到达的地点不同，但完成了同一个位移。向量首先记录这种可以搬动而不改变的动作。画箭头时，起点可以换，方向与长度却不能随意换。“从原点出发的箭头”只是方便的画法，并不是向量的全部含义。

把这段路写成 $(2,-1)$，两个数其实是两条指令：沿第一个基方向走两份，沿第二个基方向走负一份。坐标不是贴在箭头上的永久标签；换一套方向与单位，同一位移会得到另一组数。因此阅读坐标之前，要先问坐标轴怎样选、每格代表什么。若两个分量分别记录温度和质量，也不能直接赋予欧氏长度一个物理意义。

让运算来自动作

再接着走 $(-1,3)$，横向净移动一格、纵向净移动两格，所以相加可以逐坐标进行。把整条路线走两遍，每一步都扩大两倍；“先合并再扩大”与“分别扩大再合并”得到同一结果。分配律由此成为两种操作顺序的一致性要求。

先画首尾相接的箭头，再把它们搬到共同起点，补成平行四边形。两幅图为什么指向同一终点？回答这个问题，就把几何相加与坐标相加接了起来。后面遇到有几百个分量的数据向量，虽然画不出箭头，逐项加法与整体缩放仍保留这套结构。下一节将追问：只给少数几个方向，能拼出多少向量？

一步一步建立直观

1．先把向量看成一次位移　想象在一张地图上先走 u，再走 v。第二段路可以搬到第一段的终点，方向和长度却不改变。右图的虚线正是在做这个搬移；红色箭头只记录起点到最终位置的总变化。它不记录走过的路径，所以同一位移可以由不同路线实现。

2. 数乘改变的是位移的份数 把系数 s 改为二，相当于沿 u 连走两次，再走一次 v。总位移是 $su+v$，横纵两个分量分别累加。这里的加法并非任意规定：它来自每个坐标方向上净位移的相加，因此与先后分组无关。负系数则把相应位移反向。

3. 方向数目与独立信息 让两根箭头共线后，虽然仍可以填写两个系数，终点却只能沿一条直线移动。图形帮助我们分辨两个数字和两份独立信息。若两根箭头不共线，改变两个系数才能分别控制平面内的两个方向；这就是下一节讨论基的原因。

定义、命题与证明

定义：线性组合 给定向量 u,v 与标量 s,t，向量 $su+tv$ 称为它们的线性组合。系数记录每个方向参与的份额；负系数表示反向。

命题：向量平移不改变合位移 在同一向量空间中，先走 u 再走 v 的总位移为 $u+v$，与先走 v 再走 u 相同。

证明. 用坐标表示 $u=(u_1,u_2)$、$v=(v_1,v_2)$。第一条路径的终点是 (u_1+v_1,u_2+v_2)，第二条路径的终点是 (v_1+u_1,v_2+u_2)。

实数加法交换律使两点一致。这里比较的是终点位移；两条路径的中间位置可以不同，所以不能把结论改写为两次运动在每一时刻都相同。

□

> **观课提示** 先观察同一个箭头平移后哪些性质保持，再把向量相加解释成两段位移的合成。
>
> 原课程视频 本节网页与媒体 参数实验

要点回看

核心概念 向量可以看成位移、按顺序排列的特征，或满足加法与数乘规则的对象。坐标依赖基的选择，而向量本身不必依赖某一组坐标。

整理成数学语言 在 $\mathbb{R}^2$ 中，$(x,y)+(u,v)=(x+u,y+v)$，$a(x,y)=(ax,ay)$。数乘对加法分配：$a(u+v)=au+av$，这保证“先合并再缩放”和“先缩放再合并”一致。

完整算例 取 $u=(2,-1),v=(-1,3)$，则 $u+v=(1,2)$，$2(u+v)=(2,4)=2u+2v$。箭头首尾相接给出同一个结果。

易错点 点的位置与位移不是同一个概念；讨论位置向量时已经选了原点。列坐标的顺序也不能交换。

自测与答案 求 $3(1,-2)-2(-1,1)$。答案为 $(5,-8)$；交换数乘与相减顺序不会改变结果。

Lean 对照 `LinearAlgebra.vectors_distribute`

在实数对上完整验证数乘分配律；没有把几何直觉替代成浮点数计算。

1.2 线性组合、张成与基

官方课程：span

把系数看成两个可以调节的旋钮

给定两种移动指令 $u=(1,1)$ 与 $v=(1,-1)$，可以分别执行任意实数倍。先固定第二个旋钮，只改变第一个，终点沿直线移动；再改变第二个旋钮，这条直线整体平移。因为两方向不平行，这些直线扫过整个平面。“所有能到达的位置”就是张成空间。

为什么允许负数和非整数系数？若只准走整数步，得到的是格点；若只准非负步，通常得到一个扇形区域。张成允许任意实系数，正是为了让加法与数乘自由进行。定义中的量词决定了我们讨论哪个集合。

能到达与能唯一到达

要到达 $(4,2)$，横向必须满足 $a+b=4$，纵向必须满足 $a-b=2$。相加相减得到 $a=3,b=1$，也就是目标在新方向下的坐标。若另一对系数到达同一位置，两份表达相减会得到一个非平凡零组合；但两个不平行方向不能这样抵消，所以坐标唯一。

把 v 换成 $2u$。两个旋钮看似独立，实际上只改变同一条线上的位置；而 $u=1u+0v=-1u+1v$ 又说明描述出现冗余。基把两方面合在一起：既张成目标空间，又不给同一向量留下多套坐标。

看网格实验时，先预测两根基箭头靠近平行会怎样，再观察结果。网格塌缩将在行列式章节变成面积归零；坐标不再唯一会在零空间章节重新出现。它们不是互不相关的定义，而是同一现象的不同描述。

一步一步建立直观

1. 把系数当成两个运动旋钮 先固定两根方向箭头，只想象改变它们前面的系数。每一个系数组合产生一个终点，所有可能终点的集合才叫张成空间。右图显示的是其中一个组合；有限次拖动只能提供例子，不能代替对所有实系数的定义。

2. 可到达与唯一到达是两件事 若两个方向不共线，可以沿它们分别补足所需的两份位移。对于平面上的目标，联立两条坐标方程便得到系数；非零行列式保证答案唯一。基既要能到达全部目标，又要避免用不同配方描述同一目标。

3. 用退化情形检验条件 把 v 设成 $2u$，同一个终点可以通过增加一份 u、减少半份 v 保持不变。这个看得见的冗余就是线性相关。此时系数空间仍是二维，但输出只是一条线；区分输入参数数目与实际可达维数十分重要。

> **观课提示** 改变线性组合系数时，终点能到达哪里？特别留意两根基向量共线后的变化。
>
> 原课程视频 本节网页与媒体 参数实验

要点回看

核心概念 张成空间是“给定方向通过任意拉伸、相加能到达哪里”。基要求既能到达整个空间，又没有多余方向。

整理成数学语言 $\operatorname{span}(u,v)=\{au+bv : a,b\in\mathbb{R}\}$。标准基 $e_1=(1,0), e_2=(0,1)$ 满足 $xe_1+ye_2=(x,y)$。若 u,v 不共线，表达式 $au+bv$ 的系数唯一。

完整算例 对 $u=(1,1), v=(1,-1)$，解 $a+b=4, a-b=2$ 得 $(4,2)=3u+v$。若把 v 换成 $2u$，只能得到直线 $y=x$。

易错点 “有两个向量”不等于“有二维基”；零向量或共线向量不会增加维数。

自测与答案 $(3,5)$ 在上述 u,v 基下的坐标是什么？答案 $(4,-1)$，代回为 $(4,4)+(-1,1)$。

Lean 对照 `LinearAlgebra.coordinates_span`

验证标准基的张成等式。一般基的线性无关性与坐标唯一性在本节纸笔解释中给出，未在此命题中形式化。

1.3 线性变换与矩阵

官方课程：`linear-transformations`

用两个箭头描述整个平面的运动

设想写一个程序，让平面上的每个点移动。逐点列出目的地不现实，我们需要压缩信息的规则。线性规则承诺：一个向量若由两个向量按一定比例拼成，变换后仍按同样比例拼成。知道基向量的去向，就能恢复所有点的去向。

任意输入都能拆为 xe_1+ye_2。假如 e_1 被送到 $(2,0)$，e_2 被送到 $(1,3)$，只需取 x 份第一根新箭头与 y 份第二根新箭头，结果就是 $(2x+y,3y)$。矩阵把两条目的地记录并排存放，所以它们是列，而不是行。

公式从基向量的去向长出来

取 $x=1, y=2$，画一份 $(2,0)$ 和两份 $(1,3)$，终点为 $(4,6)$。再按坐标计算，会发现每一行都在计算输出的一个分量。列的解释告诉我们动作是什么，行的解释告诉我们输出数字怎么计算。

打开矩阵实验，先只改变一列。预测哪根基箭头会动、哪些输入受影响，再播放从单位网格到目标网格的变化。中间的连续变形只是展示方式；数学对象是指定的输入输出关系，不能把动画经过的每个阶段都当成同一个变换。

最后检查边界：平移虽然规整，却把零向量送到非零位置，因此不线性。只看水平线与竖直线是否笔直也不够，仍需检查任意向量的相加与缩放。下面的 Lean 命题精确表达的就是这个承诺，而不是对一张网格图作判断。

一步一步建立直观

1. 先识别什么是动作的输入 把左图的绿色箭头当成写在弹性网格上的一段位移。输入箭头改变时，矩阵保持固定，右图显示同一规则给出的输出。若是恒等矩阵，两边坐标完全一致。这一步把改变输入与改变变换区分开来，避免把两种操作混在一起。

2. 两根基箭头如何决定整张网格 剪切时第一根基箭头不动，第二根向右倾斜。任意输入 (x,y) 都由 x 份第一根箭头和 y 份第二根箭头组成，所以输出为 $(x+y,y)$。网格不是被逐点任意拉动：所有点共享同一个线性组合规则，平行线因而仍平行或塌缩。

3. 压扁揭示了线性规则的边界 把两列变成平行方向后，单位方形的面积归零，不同输入可以落到同一个输出。这不破坏线性，却破坏可逆性。纸笔证明应分别检查相加与数乘保持；不能因为动画看起来平滑，就断言动作线性，更不能进一步断言它可逆。

定义、命题与证明

定义：线性映射 映射 T 线性，是指它保持向量加法与标量乘法：$T(u+v)=T(u)+T(v)$，$T(cu)=cT(u)$。平移一般不满足这个定义。

命题：基向量的像决定整个映射 设 e_1, e_2 是平面的标准基。若 T 线性，则对任意坐标 x, y，有 $T(xe_1+ye_2)=xT(e_1)+yT(e_2)$。

证明. 先用加法保持性把两项分开，再分别把标量移出 T，就得到结论。把 $T(e_1)$ 与 $T(e_2)$ 写成矩阵的两列，右侧正是矩阵乘向量的列组合。

因此我们不必给平面上每个点单独记录去向。不过反过来只观察两根箭头，尚不能证明未知映射线性；线性是推论所需的前提。

□

观课提示 先盯住两根基向量，再预测任意格点的去向；解释矩阵为什么按列记录这些信息。

原课程视频 本节网页与媒体 参数实验

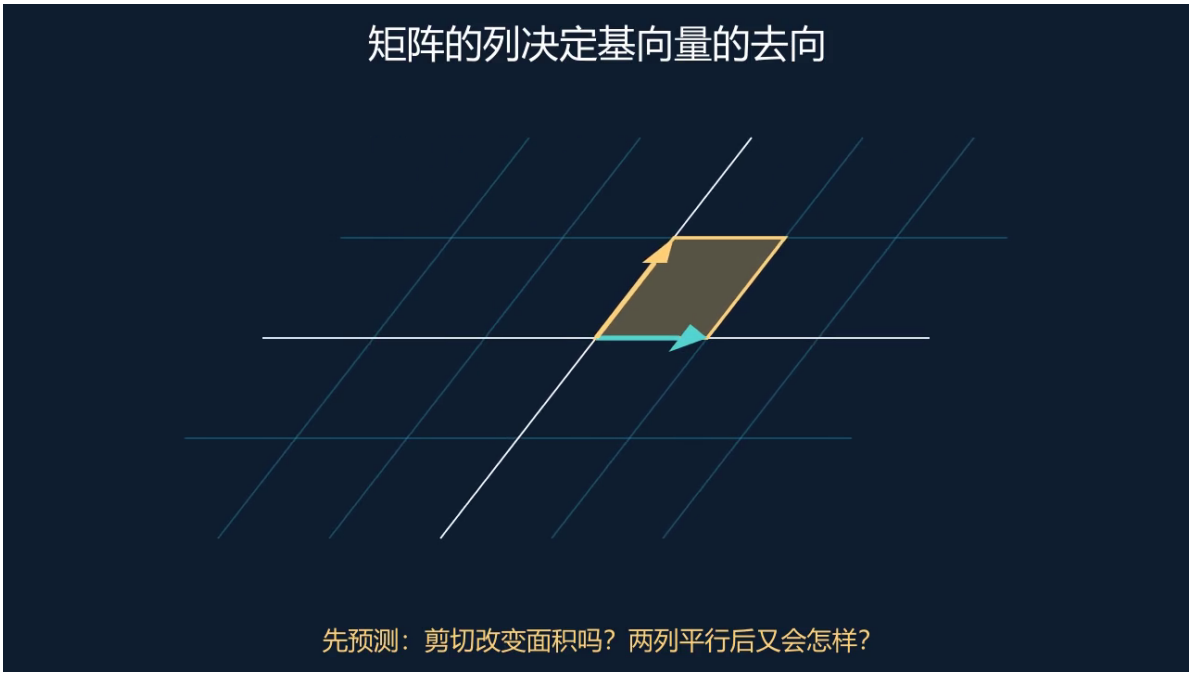


图 1.1: 剪切后的网格与单位方形。两列控制基向量去向；继续压扁时面积归零。本项目原创 Manim 动画关键帧；完整动画与操作入口见本节网页。

要点回看

核心概念 线性变换保持向量相加与缩放。平面上的网格可以被拉伸、剪切或压扁，但原点固定，平行的直线仍保持平行或塌缩。

整理成数学语言 $T(au+bv)=aT(u)+bT(v)$。矩阵的第 j 列就是 $T(e_j)$，所以

$$A=\begin{bmatrix}2 & 1\\ 0 & 3\end{bmatrix} \quad\Longrightarrow\quad A(x,y)=(2x+y,3y).$$

完整算例 $A(1,2)=(4,6)$，也等于 $T(e_1)+2T(e_2)=(2,0)+2(1,3)$。这从基向量图像直接导出矩阵乘法。

易错点 平移 $v\mapsto v+b$ 在 $b\neq 0$ 时不是线性变换；它是仿射变换。

自测与答案 $T(x,y)=(x,y+x^2)$ 是否线性？不是，例如 $T(2,0)=(2,4)\neq 2T(1,0)$。

Lean 对照 `LinearAlgebra.linear_combination`

对任意实向量空间间的线性映射验证线性组合保持律，形式化范围比二维例子更广。

1.4 矩阵乘法就是复合

官方课程：matrix-multiplication

从执行顺序推导乘法顺序

先把图形横向拉长两倍，再剪切，怎样用一个矩阵保存整个过程？必须知道每根标准基箭头经过两步后落在哪里。第一步的矩阵 B 已把位置存进自己的列；第二步对每一列作用 A。因此复合矩阵的列是 A 乘以 B 的各列，这就是矩阵乘法的来源。

ABv 的括号是 $A(Bv)$：书写从左到右，执行从最靠近输入的右端开始，与 $f(g(x))$ 一样。把乘法读成动作，还能判断不同形状的矩阵能否相乘：前一步的输出维度必须接上后一步的输入维度。

交换顺序为什么改变结果

拿 $(0,1)$ 做实验。先横向拉长时，它在纵轴上，不动；再剪切，横坐标变成 1。若先剪切，它先到 $(1,1)$，随后拉长会把刚产生的横坐标放大，结果是 $(2,1)$。第二步处理的是已经改变的对象，所以同样两步不必得到同一结果。

再看公式里的求和：第 i 个输出分量，要累加中间层各分量对它的贡献。下标 k 标记所有中间通道。神经网络层间计算与换基都会重用这个视角。

三个动作按 C, B, A 执行，先把哪两步打包，不改变实际过程，所以 $(AB)C = A(BC)$。结合律允许重新打包，却没有允许交换动作。以后读长矩阵公式时，给每个矩阵配一句实际指令，再决定能怎样移动括号。

一步一步建立直观

1. 每一幅图都是一次计算的状态 左图给出原始向量，中图是第一次动作的输出，右图才是最终结果。先让旋转角为零，只保留剪切，确认三幅图的含义。矩阵乘法不是把两个数表逐项相乘，而是把前一步产生的向量交给后一步处理。

2. 把中间状态代入第二条规则 令 B 是旋转，A 是剪切。先算 Bx，再将这个新向量当成 A 的输入，得到 $A(Bx)$。对每一根标准基箭头重复这个过程，就得到了乘积矩阵的各列。因此列的几何解释直接导出行列求和的计算公式。

3. 交换顺序是一项可检验的主张 切换为先剪切再旋转，中图首先发生改变，最后的红色向量通常也不同。原因是第二次操作面对的方向已由第一次改变。结合律只改变分组，动作队列不变；交换律却重排了队列，两者不能混为一谈。

定义、命题与证明

定义：映射的复合次序 若 B 把输入送到中间空间，A 再作用于中间结果，则复合映射写成 $A \circ B$，矩阵写成 AB。最右侧的矩阵先作用。

命题：矩阵乘法的列规则 复合矩阵 AB 的第 j 列是 A 作用在 B 的第 j 列所得的向量。各矩阵的输入输出维度须匹配。

证明. 把标准基向量 e_j 送入复合映射。第一步得到 Be_j，它就是 B 的第 j 列；第二步得到 $A(Be_j)$。

一个矩阵的第 j 列也等于它作用在 e_j 上的结果，所以复合矩阵的各列只能按此规则组成。交换 A、B 会改变动作顺序，通常也改变结果。

□

> **观课提示** 分别跟踪先剪切后缩放与先缩放后剪切，找出两种结果最早出现差异的步骤。
>
> 原课程视频 本节网页与媒体 参数实验

要点回看

核心概念 矩阵乘法的列，是后一动作的基向量像经过前一动作后的结果。因此乘法次序有实际意义。

整理成数学语言 $(AB)v = A(Bv)$，且 $(AB)_{ij} = \sum_k A_{ik}B_{kj}$。令

$$A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, \quad B = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}; \quad AB = \begin{bmatrix} 2 & 1 \\ 0 & 1 \end{bmatrix}, \quad BA = \begin{bmatrix} 2 & 2 \\ 0 & 1 \end{bmatrix}.$$

完整算例 对 $v = (0, 1)$，先 B 后 A 得 $(1, 1)$，反过来得到 $(2, 1)$。这是不交换性的直接见证。

易错点 AB 的右边先做；“每个位置对应相乘”是逐元素乘法，不是线性变换复合。

自测与答案 为什么 $(AB)C = A(BC)$？两边都对向量执行 C、B、A，或展开有限求和验证。

Lean 对照 `LinearAlgebra.matrix_composition`

复用 mathlib 的矩阵作用复合定理，在任意实 2×2 矩阵和向量上验证等式。

1.5 三维线性变换

官方课程：`3d-transformations`

先追踪三条坐标轴

二维的两个基向量变成三维的三个。想象透明方盒从同一角出发的三条棱，分别代表 e_1, e_2, e_3。线性变换把盒子变成平行六面体；记录三条棱的新方向，内部每点的位置就由同样的线性组合决定。九个矩阵元素依旧只是三份目的地记录。

绕竖直轴旋转四分之一圈时，横向箭头变为纵向，纵向箭头变为负横向，竖直箭头保持不动。把这三条记录排成列，旋转矩阵便自然出现。应用于 $(2, 1, 3)$ 时，组合三根新基箭头即可得到 $(-1, 2, 3)$。

寻找运动中没有改变的量

如何检查公式真像旋转？可以计算平方长度。变换后是 $(-y)^2 + x^2 + z^2$，恰等于原来的 $x^2 + y^2 + z^2$。一次数值算例不能证明一般结论，所以应保留任意的 x, y, z。Lean 在这里验证的就是这个一般恒等式。

三维还有一个新现象：不同轴的旋转通常不交换。拿一本书，先绕竖直方向再绕左右方向转动，随后试着倒过来做，会得到不同姿态。不能把二维平面内旋转的交换性推广到所有三维旋转。

本节不要求先掌握一整套旋转参数，只需学会从列读动作。后面遇到体积，回想这个盒子；遇到核，想象它压成面或线；遇到特征向量，寻找变换后仍留在原直线上的方向。三维只是多了一根基，核心推理仍是同一套。

一步一步建立直观

1. 从屏幕图像恢复空间对象 旋转观察视角，留意箭头在屏幕上变短并不意味着空间长度变小。屏幕只是三维对象的投影，下面的长度读数才使用三维坐标计算。操作参数改变数学对象，拖动空白处改变相机，这两件事必须明确区分。

2. 绕固定轴旋转改变哪些坐标 绕 z 轴转四分之一圈时，(x, y, z) 变为 $(-y, x, z)$。横纵分量互换并改变一个符号，高度保持。平方长度仍是 $x^2 + y^2 + z^2$，因此图像中的转动有一条可以独立核对的代数不变量。可以在任意角度先记录长度，再修改转角复核这个守恒关系。

3. 旋转与投影的本质差别 切换到投影模式，垂直虚线显示被丢弃的高度。所有只在 z 坐标上不同的点都有同一个平面像。旋转保留恢复原状态的信息，投影则一般做不到；能否求逆取决于这个信息差别，而不是画面是否仍有三维感。

> **观课提示** 观察三根基向量如何确定三维动作，并检查旋转时长度是否保持。
>
> 原课程视频 本节网页与媒体 参数实验

要点回看

核心概念 三维矩阵有三列，分别记录三个基向量去向。二维的“面积”将在三维变成“体积”。

整理成数学语言 绕 z 轴逆时针旋转 90° 的矩阵为

$$R = \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad R(x, y, z) = (-y, x, z).$$

完整算例 $(2, 1, 3)$ 变为 $(-1, 2, 3)$。两者平方长度均为 14，且 z 坐标保持不变；连续作用四次回到原向量。

易错点 三维旋转通常不交换；绕不同轴的两个旋转要按顺序处理。

自测与答案 对 $(1, 0, 0)$ 连续作用两次 R。答案 $(-1, 0, 0)$，不是原向量。

Lean 对照 `LinearAlgebra.rotation_preserves_square_norm`

验证该 90° 旋转保持平方长度。未宣称已经证明一般三维旋转的分类。

1.6 行列式：有向体积倍率

官方课程：`determinant`

用一个数描述整块区域的变化

已知每点去哪之后，可以问：面积为一的方形，变换后占多大地方？两列向量围成的平行四边形就是它的像。若两列为 (a, c) 与 (b, d)，计算有向面积会得到 $ad - bc$。这个数总结了变换对面积的一项整体作用。

剪切使上边向旁边滑动，但底与垂直高度不变，所以面积不变。水平反射也不改变面积大小，却翻转了两根基箭头的排列朝向。为了保存这个区别，我们让面积倍率带符号；负号表示定向翻转，不是实际土地具有负面积。

零值意味着信息丢失

让两列逐渐变得平行，平行四边形的高度消失，面积归零。此时不同输入压到同一条线上，许多位置无法区分。行列式为零由此预告了信息丢失，也解释了为什么无法通过逆变换找回所有输入。

在矩阵实验中把第二列设为第一列的两倍。先预测面积，再检查显示值；随后轻微偏转它，观察面积恢复。图形可以帮助理解压扁，但“严格等于零”仍依赖精确数值或符号关系，不能仅凭图形很窄来判断。

复合的倍率应该相乘：第一步把每块面积扩大两倍，第二步再扩大三倍，总倍率就是六。定向也遵守乘法规则，两次反射恢复原朝向。这给出了行列式乘法公式的意义；下面的代数展开检查它在一般二维坐标下成立。面积、可逆性与零空间将在下一节汇合。

一步一步建立直观

1. 先选一个可重复测量的面积单位 观察单位方形的四个顶点在同一矩阵下的去向。输出是由两列张成的平行四边形，它与原方形的面积之比不依赖方形放在哪里。行列式正是把这种统一的局部面积倍率压缩成一个数。

2. 剪切说明形状改变未必改变面积 剪切把顶部推向一侧，却保留底边和垂直高度，因此面积仍为一。只看两根边的长度会误判，因为夹角同时改变。坐标公式 $ad-bc$ 把两块有向矩形相减，恰好消除了这种倾斜造成的重复计算。

3. 用反射区分大小与定向 反射后方形面积仍为一，但两根边的顺时针或逆时针次序交换，因此行列式是负一。普通面积必须取绝对值。若继续把两列调成共线，面积为零；这又把几何退化与不能唯一恢复输入联系起来。

> **观课提示** 追踪单位方形的有向面积，区分剪切、反射和压扁这三种情况。
>
> 原课程视频 本节网页与媒体 参数实验

要点回看

核心概念 行列式记录面积或体积缩放，同时用符号记录定向是否翻转。压成更低维时，体积归零。

整理成数学语言 对 $A=\left[\begin{smallmatrix}a & b\\ c & d\end{smallmatrix}\right]$，$\det A=ad-bc$。复合的倍率相乘：$\det(AB)=\det A\det B$。

完整算例 剪切 $S=\left[\begin{smallmatrix}1 & 3\\ 0 & 1\end{smallmatrix}\right]$ 把单位方形推斜，面积仍为 1。反射 $F=\operatorname{diag}(-1,1)$ 的行列式为 -1，面积大小不变但方向反转。

易错点 行列式为负不表示几何面积为负；普通面积倍率是 $|\det A|$。非方阵没有这种同维体积行列式。

自测与答案 $A=\left[\begin{smallmatrix}2 & 1\\ 1 & 3\end{smallmatrix}\right]$ 的面积倍率？答案 5；A^2 为 25。

Lean 对照 `LinearAlgebra.determinant_product`

通过坐标展开严格证明两个一般实 2×2 矩阵的行列式乘法公式。

1.7 逆矩阵、列空间与零空间

官方课程：`inverse-matrices`

解方程是在寻找运动的起点

把 $Ax = b$ 读成“哪个输入经过 A 后落在 b”。所有可达位置是矩阵列向量的线性组合，组成列空间。若目标不在其中，怎样调输入都不会成功；目标可达之后，才轮到讨论起点是否唯一。

考虑 $(x, y) \mapsto (x + y, 2x + 2y)$。输出的第二分量总是第一分量的两倍，故 $(3, 7)$ 不可达，而 $(3, 6)$ 可达。对后者，$(3, 0)$ 和 $(2, 1)$ 都是起点；两输入的差 $(1, -1)$ 被送到了零。

零空间保存不可见的变化

沿 $(1, -1)$ 走任意距离，一坐标增加、另一坐标减少，和不变。核描述的正是输出无法看见的输入变化。若已找到一个解 x_0，加任意核向量仍是解；反过来，任意另一个解减去 x_0，由线性性也在核里。于是整个解集是一份平移过的零空间。

逆矩阵要求每个目标恰好对应一个起点：既没有不可达目标，也没有丢失方向。在同维有限维空间中，这两方面相联系；对方阵，又与非零行列式相联系。非方阵的两种要求却可能分别成立，下一节会把它们拆开。

练习时按这个顺序思考：先判断右端是否可达，再找一个特解，最后找所有不可见的变化。这样既能解释无解和多解，也能把计算与张成、基和面积压缩接起来，而不是试图把“除以矩阵”当成通用规则。

一步一步建立直观

1. 求解就是追问输出来自哪里 固定矩阵后，拖动左图输入并追踪右图红点。方程 $Ax = b$ 要求找到落到指定输出 b 的输入。对于非零行列式的方阵，每个目标恰有一个原像；逆矩阵记录这条反向寻找规则。

2. 压扁时同时出现不可达与不唯一 两列共线后，输出被限制在一条线上。线外目标没有解；线内目标通常有整条输入线作为原像。沿着被压成零的方向移动输入，输出不变，这就是核。不要把不可逆简单理解成每个方程都无解。

3. 把图像整理成解集的证明 若 x_0 已是一个解，且 $Av = 0$，线性性给出 $A(x_0 + v) = b$。反过来，任意另一个解减去 x_0 后都被送到零。因此解集恰为 x_0 加上核；这些平行原像族可以由这两步等式完整解释。保持压扁情形，沿核方向拖动输入，就能检查输出是否保持不变。

> **观课提示** 找出哪些输入会落到同一个输出，并用零空间解释这种信息丢失。
>
> 原课程视频 本节网页与媒体 参数实验

要点回看

核心概念 $Ax = b$ 是寻找能被送到 b 的输入。列空间描述可到达的输出，零空间描述被压成零的方向。

整理成数学语言 若 x_0 是一个解，则所有解为 $x_0 + \ker A$。方阵可逆当且仅当其零空间只有零向量，也等价于行列式非零。

完整算例 $A = [\frac{1}{2}\ \frac{1}{2}]$ 的列空间为 $\mathrm{span}(1,2)$，零空间为 $\mathrm{span}(1,-1)$。$Ax = (3,6)$ 的解是 $(3,0) + t(1,-1)$；右端若为 $(3,7)$ 则无解。

易错点 “不能求逆”不等于“没有解”，也可能有无穷多个解。

自测与答案 可逆矩阵的两个不同输入能有同一输出吗？不能；相减后得到一个非零零空间向量，与可逆性矛盾。

Lean 对照 `LinearAlgebra.kernel_translation`

证明给任意输入加上核中的向量不会改变输出；形式化的是解集结构的核心方向。

1.8 非方阵与维度变化

官方课程：nonsquare-matrices

输入与输出不必住在同一空间

删掉高度的线性投影 $P(x,y,z) = (x,y)$ 接受三个输入分量，却只给两个输出分量，所以矩阵有两行三列。把它想象成从三维场景保留一个平面视图，比先背矩阵尺寸规则更自然。

$m \times n$ 矩阵的每一列记录一个输入基方向在 m 维输出空间中的去向，一共需要记录 n 根输入基。因此行数是输出维度，列数是输入维度；检查乘法能否进行，也就是检查前一步输出是否能交给后一步。

“撤销”需要说明是哪一边

每个平面点 (a,b) 都有原像 $(a,b,0)$，所以 P 满射。但 $(a,b,5)$ 也投到同一点，所以它不单射。反向嵌入 $J(a,b) = (a,b,0)$ 只是约定一个高度，并没有找回输入原来的高度。

先嵌入再投影，平面点原样返回，得到 $PJ = I_2$。先投影再嵌入，三维点变成 $(x,y,0)$，所以 $JP \neq I_3$。一对矩阵能在一边互相撤销，却不能在另一边撤销，这就是左右逆要分别讨论的原因。

画几条平行于高度轴的线，把每条线看成一组无法区分的输入。这些线是核的平移。以后遇到数据降维、神经网络宽窄不同的层或欠定方程，先数维度，再问保留了什么、丢失了什么。维度分析不能独自解出问题，却经常能在计算前排除不可能的答案。

一步一步建立直观

1. 输入和输出可以属于不同空间 先观察三维箭头，再启用平面投影。输出只保留横纵两项，所以需要两行记录输出，却仍需三列说明三个输入方向的贡献。一个 2 行 3 列矩阵不是残缺的方阵，而是从三维到二维的完整规则。

2. 寻找消失的自由方向 保持横纵分量不动，只改变高度。平面输出不会移动，因此竖直轴属于核。这个观察说明为什么三维输入无法由二维投影唯一恢复；丢失的是一整份自由信息，而不是计算时偶然漏掉一个数。

3. 把观察与维数结论分开 每一个平面向量都可由高度为零的三维向量投过去，所以此投影是满射；它同时不是单射。维数论可以推广这种现象，但应先说明映射、定义域和值域。屏幕上的透视缩短只是相机效果，不等同于这个数学投影。

> **观课提示** 先数清输入与输出维度，再观察投影丢掉了哪一个方向。
>
> 原课程视频 本节网页与媒体 参数实验

要点回看

核心概念 $m \times n$ 矩阵把 n 维输入映到 m 维输出。行数与列数分别对应输出和输入维度。

整理成数学语言

$$P = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}, \quad P(x, y, z) = (x, y), \quad \ker P = \{(0, 0, z) : z \in \mathbb{R}\}.$$

它满射但不单射。相反，$J(x, y) = (x, y, 0)$ 单射但不是到整个 $\mathbb{R}^3$ 的满射，且 $PJ = I_2$。

完整算例 $(2, 5, 7)$ 和 $(2, 5, -1)$ 都被投到 $(2, 5)$。丢失的 z 信息无法从输出恢复。

易错点 有左逆或右逆并不一定同时有双侧逆；矩阵形状先于运算公式。

自测与答案 JP 是什么？它把 (x, y, z) 变为 $(x, y, 0)$，不是 I_3。

Lean 对照 `LinearAlgebra.projection_surjective`

完整证明这一坐标投影满射，构造每个输出 (x, y) 的原像 $(x, y, 0)$。

1.9 点积与对偶

官方课程：`dot-products`

把一个向量变成一把测量尺

若要回答“一个位移在某方向上走了多少”，输出应该是一个数。固定方向 u，把 v 沿它投影就是这种测量。反方向应给负数，垂直方向应给零，相接的两段位移应使测量结果相加。我们因此期待一个线性测量。

在标准正交坐标中，测量由它对基向量的读数决定。若对 e_1, e_2 的读数分别为 a, b，对 (x, y) 的读数只能是 $ax + by$。把读数收进向量 (a, b)，便得到点积公式。一端是被测对象，另一端是测量仪器，这是对偶空间的具体入口。

区分投影系数、长度与向量

取 $u = (1, 2), v = (3, 1)$，点积为 5，但投影长度不是 5，因为尺子不是单位长度。要找投影向量 cu，应要求残差 $v - cu$ 垂直于 u。令 $u \cdot (v - cu) = 0$，得到 $c = (u \cdot v)/(u \cdot u)$，本例恰为 1。

残差 $(2, -1)$ 与尺子点积为零，说明分解正确。这个推导也解释了分母和 $u \neq 0$ 的要求。向量的几何长度、投影向量、沿基方向的系数是三个对象，不能因为都出现一个数字就混为一谈。

若改用不垂直的基，逐坐标相乘相加一般不再表示同一几何内积，还需记录基之间的交叉贡献。后面的注意力虽然处理高维数据，仍沿用“用固定方向读出一个分数”的思想；但语义是否与某个方向对应，需要另作经验判断。

一步一步建立直观

1. 把点积当作沿某个方向的读数 把 u 看成测量方向，v 看成一次位移。投影回答位移沿尺子方向前进了多少；垂直于尺子的那部分不会贡献读数。若尺子没有归一化，点积还额外乘上尺子的长度，因此不能直接把它当作投影长度。

2. 垂直分量为什么消失 让待测向量与测量方向垂直，点积变为零。一般情形可以把 v 拆成平行部分与垂直部分；平行部分的系数必须是 $(u \cdot v)/(u \cdot u)$，因为减掉它以后与 u 的点积恰为零。这里要求 u 非零。

3. 负号表示相反的测量方向 把待测向量移到尺子的反方向，点积为负。这个符号表达朝向，并非说长度可以为负。在力学中，恒力与位移的点积描述功；垂直运动不做功、反向运动做负功，提供了同一个投影关系的物理解释。

定义、命题与证明

定义：沿一条非零方向的正交投影 对非零向量 u，定义 $\mathrm{proj}_u(v) = \dfrac{u \cdot v}{u \cdot u} u$。投影是向量，点积是标量，两者不能互相替代。

命题：投影后的剩余量与投影方向垂直 设 $u \neq 0$，令 $r = v - \mathrm{proj}_u(v)$。则 $u \cdot r = 0$。

证明. 由点积的线性性，$u \cdot r = u \cdot v - \dfrac{u \cdot v}{u \cdot u}(u \cdot u) = 0$。分母可约去，是因为实内积下非零向量满足 $u \cdot u > 0$。

因此 v 被分成沿 u 的部分与垂直于 u 的部分。若 $u = 0$，公式没有定义；图形窗口会报告这一退化情形，而不生成一个任意的投影方向。

□

观课提示 把点积看成投影测量，留意尺子的长度如何影响读数。

原课程视频 本节网页与媒体 参数实验

要点回看

核心概念 固定一个向量 u，点积 $v \mapsto u \cdot v$ 就是一个线性测量。几何上它把 v 沿 u 的投影长度乘以 $\|u\|$。

整理成数学语言 $u \cdot v = \sum_i u_i v_i = \|u\|\|v\| \cos\theta$。对 $u \neq 0$，投影为 $\operatorname{proj}_u v = \frac{u \cdot v}{u \cdot u} u$。

完整算例 $u = (1, 2), v = (3, 1)$ 的点积为 5，所以投影恰为 $(1, 2)$，残差 $(2, -1)$ 与 u 正交。

易错点 点积的坐标求和式默认标准正交基。一般基下需使用 Gram 矩阵 G，得到 $[u]^T G [v]$。

自测与答案 写出线性测量 $L(x, y) = 4x - y$ 对应的向量。标准内积下答案 $(4, -1)$。

Lean 对照 `LinearAlgebra.dot_linear`

验证二维点积对第二个输入可加，即固定向量所定义测量的线性结构之一。

1.10 叉积与有向面积

官方课程：`cross-products`

用一根箭头保存一块面的信息

两个三维向量围成一个平行四边形。要同时保存它的面积与朝向，单独一个数不够，而法向量能提供方向。我们约定输出长度等于面积，方向符合选定的右手规则，就得到叉积。它把一块有向面压缩成一根箭头。

沿 u 再沿 v 绕边界，与倒过来绕，围出同一块面，但绕行方向相反。所以交换输入次序，面积大小不变，法向量必须反向。两输入平行时，面塌缩，叉积为零。这两个极端情形可以用来检查任何候选公式。

让坐标公式接受几何检查

计算前先列要求：输出与两个输入都垂直；任意一边扩大两倍，输出也扩大两倍。坐标公式中的二阶行列式满足这些要求。与 u 点乘时，各项成对抵消，解释了垂直性，而不只是留下一个需要背的三分量式子。

取 $u=(1,2,0), v=(0,1,3)$，输出 $(6,-3,1)$。分别点乘检查与两边正交，再按右手方向检查符号。正交本身不能区别正反，也不能固定长度，因此几个检查必须合起来使用。

下一节会从体积测量反推这根法向量，把点积、行列式和叉积联系起来。那时也会看清这里依赖三维、内积和定向。抽象向量空间中有加法与数乘，并不意味着自动具有同一个叉积运算。

一步一步建立直观

1. 先看有向底面，再引入法向量　两根空间向量张成一个平行四边形。叉积把底面的两个信息合并成一根箭头：模长记录面积，方向垂直于底面并编码输入次序。旋转视角应当只改变屏幕外观，不应改变下面计算得到的面积。

2. 倾斜底面时法线也要跟着转　把第一根向量抬离平面，红色法向量随底面转动。检查它与两条边的点积都为零，就把“看起来垂直”变成了坐标验证。交换两条边的次序会反转法线，但不会改变平行四边形的普通面积。

3. 共线情形检验定义是否合理　当两根向量平行时，底面没有面积，叉积应为零。零向量没有唯一方向，因此此时不应继续讨论它指向哪个法线方向。叉积的三维构造把面积与定向联系起来，却不能不加条件地照搬到任意维数。

定义、命题与证明

定义：叉积　对三维向量 u, v，定义 $u\times v=(u_2v_3-u_3v_2,\ u_3v_1-u_1v_3,\ u_1v_2-u_2v_1)$，方向采用右手定向。

命题：叉积与两个输入都垂直　对任意 $u, v\in\mathbb{R}^3$，有 $u\cdot(u\times v)=0$ 且 $v\cdot(u\times v)=0$。叉积可能是零向量。

证明. 展开第一个点积，得到 $u_1u_2v_3-u_1u_3v_2+u_2u_3v_1-u_2u_1v_3+u_3u_1v_2-u_3u_2v_1$。各项两两抵消，因此和为零。

对第二个点积做同样展开即可。零向量也满足两个点积为零，所以该结论没有宣称平行输入仍产生一个有确定方向的非零法向量。

□

观课提示　同时检查叉积的垂直方向、面积大小和输入顺序改变后的符号。

原课程视频　本节网页与媒体　参数实验

要点回看

核心概念 在定向三维欧氏空间中，叉积垂直于两个输入向量；长度等于它们张成的平行四边形面积，方向由右手规则决定。

整理成数学语言

$$(a,b,c)\times(x,y,z)=(bz-cy,cx-az,ay-bx).$$

从坐标展开可得 $u\cdot(u\times v)=0$；交换顺序改变符号。

完整算例 $(1,2,0)\times(0,1,3)=(6,-3,1)$。与第一个向量点积为 $6-6=0$，与第二个点积也为 $-3+3=0$。

易错点 叉积不同于点积：一个输出向量，一个输出标量。其熟悉的三维形式需要内积和定向。

自测与答案 $(1,0,0)\times(0,1,0)$ 是什么？答案 $(0,0,1)$；反过来为 $(0,0,-1)$。

Lean 对照 `LinearAlgebra.cross_orthogonal`

完整验证一般三维实坐标下，叉积与第一个输入正交。

1.11 从线性测量重新认识叉积

官方课程：`cross-products-extended`

先有体积测量，再寻找代表它的箭头

固定底面边 u,v，让第三条边 w 变化。三边围成的有向体积是 w 的函数。把 w 拆成两段，两份有向体积相加；把它放大，体积同比例变化。因此体积是对 w 的线性测量，恰好落在点积与对偶的框架中。

标准三维内积中，每个线性测量都能写成某个固定向量 n 与 w 的点积。怎样找 n？分别输入三个标准基，读数就是它的三个坐标。展开这三个行列式，得到的正是叉积公式；复杂的分量来自一个简单的代表问题。

垂直与长度从同一要求产生

若 w 落在底面里，三条边共面，体积为零。所以 $n\cdot w=0$ 对底面内所有向量成立，迫使 n 垂直底面。若 w 是沿正法线的单位向量，体积等于底面积，所以 n 的长度也必须等于底面积。几何定义并非另加的装饰。

例如底边为 $(1,0,0),(0,2,0)$，第三边为 $(a,b,3)$。改变 a,b 只使顶面沿底面剪切，体积始终为 6。代表测量的 $n=(0,0,2)$ 点乘第三边，直接读出同一结果。

值得保留的方法是：遇到一个对某变量线性的复杂运算，先问它对基向量分别做什么。矩阵列、点积的测量向量、叉积法向量，都是用少量基上的信息描述全部输入。学会这条思路，比记住三张看似独立的公式表更能帮助迁移。

一步一步建立直观

1. 把底面测量与高度测量分开 先用右图的两根向量确定底面，并用叉积法线记录有向面积。若再给一根第三边 w，只有它沿法线的分量贡献高度。沿底面平移顶部只会剪切整个平行六面体，不会改变体积。

2. 两个测量合成一个标量 点积 $w \cdot (u \times v)$ 把第三边投到法线，同时乘上底面积。它的绝对值是体积，符号表示三根边的定向。这里的叉积因而可以理解成一种测量装置：输入 w，返回对应的有向体积。

3. 退化揭示三重积为零的原因 把底面两根边调成共线，不论第三边是什么，体积都为零。另一种零体积来自第三边落在非退化底面内。两种几何机制都被同一个标量三重积捕获；证明时应区分底面已退化与高度为零。

> **观课提示** 固定底面，改变第三条边，解释哪些变化影响有向体积。
>
> 原课程视频 本节网页与媒体 参数实验

要点回看

核心概念 固定 u, v，把第三个向量 w 送到有向体积的运算是线性的。这个线性测量可由唯一向量点积表示，而该向量就是 $u \times v$。

整理成数学语言 对以 u, v, w 为列的矩阵，$\det[u\ v\ w] = (u \times v) \cdot w$。这个标量三重积连接线性代数中的对偶、行列式与几何体积。

完整算例 $u = (1, 0, 0), v = (0, 2, 0), w = (a, b, 3)$，三重积为 6，与 a, b 无关：平行于底面的剪切不改变体积。

易错点 行向量和列向量约定混用会造成符号错误；交换任意两列会使有向体积反号。

自测与答案 若 $w = u + v$，体积是多少？为零，因为三个向量共面，也可由叉积正交性算出。

Lean 对照 `LinearAlgebra.scalar_triple_product`

展开并验证一般实坐标下的标量三重积等式；唯一表示的抽象论证未包含在该命题中。

1.12 克拉默法则

官方课程：cramers-rule

把未知系数藏进面积比

目标写作 $b = xu + yv$，其中 u, v 是矩阵的列。除了逐坐标解方程，还能寻找一个操作，只留下 x 而消去 y。以 b, v 为边计算有向面积：yv 与 v 平行，不贡献面积；剩余的 xu 使原面积乘以 x。

用行列式语言写出刚才的几何论证：

$$\det[b\ v] = \det[xu + yv\ v] = x \det[u\ v].$$

原面积不为零时，除过去便得到 x；类似地，以 u, b 为列得到 y。“换列后求行列式”其实是一种消去另一个未知数的测量。

计算之后仍要返回原问题

取 $u = (2, 1), v = (1, 3), b = (5, 5)$。原面积为 5，替换第一列后为 10，故第一系数为 2；替换第二列后为 5，另一系数为 1。最后代回 $2u + v = (5, 5)$，检查列序与符号。

若原面积为零，除法失效反映问题已经改变：两列只张成一条线，目标可能不可达，也可能有多套表示。应回到列空间与核，而不能把零分母约掉。几何解释恰好说明了公式条件为什么必要。

这套公式适合建立唯一解与有向面积的联系，并不意味着大规模数值求解都应采用它；效率与误差稳定性需要另外研究。练习时先画列向量，再把未知数解释为面积比，最后做坐标代回。这会把几何、代数和条件检查放在同一条推理链上。

一步一步建立直观

1. 把未知坐标看成两份面积贡献 设输出 $b = xu + yv$，其中 u, v 是矩阵两列。右图显示输出由这两根方向按系数组合得到。若只想求 x，就寻找一种测量，能让沿 v 的那份贡献自动消失，而保留沿 u 的贡献。这样的测量应该比较有向面积，而不是仅比较箭头长短。

2. 用替换列隔离一个系数 取 b 与 v 的有向面积，沿 v 的贡献因为共线而为零，于是面积为 x 倍原底面面积。因此 $x = \det(b, v)/\det(u, v)$；换另一个底边同理得到 y。这个推导说明公式为何替换整列，而非单独替换某个元素。

3. 分母条件对应可见的几何障碍 把两列调成共线后，分母为零。此时面积测量已经不能区分不同配方，继续做面积比没有意义。非零行列式不仅是为了避免除零，也是保证这套测量能够恢复两份坐标信息的实质条件。

观课提示 留意替换一列怎样消去另一份系数，并用面积比解释求解公式。

原课程视频 本节网页与媒体 参数实验

要点回看

核心概念 若矩阵可逆，把一列替换为目标向量，新的有向体积与原体积之比就是相应的坐标系数。

整理成数学语言 对 $ax+by=p, cx+dy=q$，在 $\Delta=ad-bc\neq 0$ 时，

$$x=\frac{pd-bq}{\Delta},\qquad y=\frac{aq-pc}{\Delta}.$$

完整算例 $2x+y=5, x+3y=5$ 给出 $\Delta=5$，故 $x=2, y=1$。几何上新列体积分别为 10 和 5。

易错点 $\Delta=0$ 时不能除以它；此时应改用秩、相容性或零空间判断。该公式也不总是大型数值求解的高效算法。

自测与答案 解 $x+2y=4, 3x+y=7$。$\Delta=-5$，答案 $x=2, y=1$。

Lean 对照 `LinearAlgebra.cramers_rule`

给定两条线性方程和非零行列式，完整推导这两个商公式。

1.13 换基：改变坐标语言

官方课程：`change-of-basis`

物体没有换，描述它的语言换了

两个人看同一箭头，一个用水平与竖直方向报坐标，一个用两根斜基报坐标。数字不同不代表向量不同。先约定 P 接收新坐标、输出旧坐标。因为新坐标 $(1,0)$ 就是第一根新基，P 的第一列必须是这根基在旧坐标下的描述。

已知动作 A 使用旧坐标，却收到新坐标输入，就要先翻译、再执行、再翻译回来。实际顺序是 P, A, P^{-1}，写成乘积是 $P^{-1}AP$。只要记清三句指令，逆矩阵的位置不必靠猜。

好语言让同一动作变简单

交换横纵坐标的动作，把 $(1,1)$ 留在原处，把 $(1,-1)$ 反向。选它们为新基，动作就从“交换两个坐标”变为“保留第一分量、取反第二分量”，因此出现对角表示。

例如 $(3,1)$ 的新坐标由 $a+b=3, a-b=1$ 得到 $(2,1)$。新矩阵把它变成 $(2,-1)$，再翻译回去得到 $(1,3)$，与原动作相符。走完这三步，往往比直接展开相似变换更容易发现方向约定错误。

换基改变描述，主动变换改变对象；两种解释共享矩阵形式，很容易混淆。阅读推导时要明确当前向量是物体、旧坐标还是新坐标。下一节寻找特征向量，就是寻找一套好语言，让重复作用变成独立方向上的简单伸缩。

一步一步建立直观

1. 换尺子不等于移动对象 绿色向量代表空间对象，坐标是用选定基向量给它写出的配方。换基只改变配方的记录方式，不应凭空移动箭头。右图选取实对称矩阵，虚线提供一组特别方便的基方向，作为一般换基的一个具体例子。

2. 为什么某些坐标更容易计算 当绿色箭头沿第一条特征方向时，矩阵作用只把它放大，不产生另一方向的分量。改用两条特征方向为基，每个坐标就各自乘上一个倍率。原本混合两个分量的计算，因此变成独立的缩放。

3. 按翻译顺序推导矩阵表达 若 P 的列是新基在旧坐标中的表达，新坐标先乘 P 才能交给旧矩阵 A，输出再乘 P^{-1} 译回。因此新矩阵是 $P^{-1}AP$。顺序来自坐标转换的职责；不是因为三块矩阵可以随意交换。

> **观课提示** 区分向量在空间中的位置与其坐标，按翻译、作用、翻译回来的顺序阅读公式。
>
> 原课程视频 本节网页与媒体 参数实验

要点回看

核心概念 换基改变的是向量的描述。同一个物理动作在不同坐标系统中有不同矩阵。

整理成数学语言 若 P 的列为新基在旧基下的坐标，则 $[v]_{\text{old}} = P[v]_{\text{new}}$，且 $[T]_{\text{new}} = P^{-1}[T]_{\text{old}}P$。

完整算例 对交换坐标的映射 $A = \left[\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right]$，选新基 $(1,1),(1,-1)$，得到对角表示 $\operatorname{diag}(1,-1)$。

易错点 P 的方向必须先定义清楚；如果定义为旧坐标到新坐标，公式中的逆就会对调。

自测与答案 上例的 $(3,1)$ 新坐标是什么？解 $a+b=3, a-b=1$，得 $(2,1)$。

Lean 对照 `LinearAlgebra.change_of_basis`

用可逆线性等价验证“换回旧坐标后得到原动作”的恒等式，明确要求坐标变换可逆。

1.14 特征向量与特征值

官方课程：`eigenvalues`

寻找没有偏离原直线的方向

一般向量经过矩阵后方向与长度都变，但某些方向只伸缩，仍在原直线上。找到这些方向，就可能把复杂动作拆成几种互不干扰的伸缩。特征向量是这些方向上的非零向量；倍率可以为负，表示反向，也可以为零，表示该方向被压掉。

对 $A = \left[\begin{smallmatrix} 2 & 1 \\ 1 & 2 \end{smallmatrix}\right]$，$(1,1)$ 变为 $(3,3)$，$(1,-1)$ 不动。任意输入都可用这两方向拼成，每作用一次只需把第一份系数乘以 3。多次作用后的主要增长方向因此一眼可见。

特征方程为什么出现

要求 $Av = \lambda v$ 且 $v \neq 0$，等价于要求 $A - \lambda I$ 有非零核。也就是从原动作中减去统一伸缩之后，能否压掉某个方向。对方阵，这要求行列式为零，所以保留方向的问题转成关于倍率的多项式方程。

求出根后仍需找向量。四分之一圈的实平面旋转没有实的保留直线，所以没有实特征向量；非平凡剪切可能只有一条保留直线，不足以组成基。能对角化要求足够多独立特征向量，不能只检查特征值存在。

在矩阵实验中选择对称示例，转动输入方向，预测哪两条线被保留。再把任意输入拆成这两个方向，预测四次作用的结果。计算应来自系数的幂，而非逐次展开四个矩阵；这是几何理解带来的实际收益。

一步一步建立直观

1. 一般输入会同时改变长度与方向 把绿色向量拖到普通方向，红色输出通常不与它共线。若反复作用，方向也会不断调整。寻找特征向量的动机，就是找出能把这种二维混合简化为一维倍率变化的特殊方向。

2. 保持的是直线，不一定是箭头朝向 当输入落在虚线上，输出与它共线，满足 $Av = \lambda v$。正倍率保留朝向，负倍率反转朝向，零倍率把非零向量压成零；三者都符合特征方程。定义要求 v 非零，否则任意数都会满足方程。

3. 用另一方向补全分解 右图的实对称例子有两条互相垂直的特征方向，可把任意输入分成这两份。反复作用时，两份坐标分别乘倍率的幂。这个例子的完整分解不能自动推广为“每个实矩阵都有两条实特征方向”，旋转矩阵就是需要注意的情形。

定义、命题与证明

定义：特征向量与特征值 若非零向量 v 满足 $Av = \lambda v$，称 v 为 A 的特征向量，λ 为对应特征值。允许 $\lambda = 0$；禁止把零向量当作特征向量。

命题：一例对称矩阵的两个特征方向 矩阵 $A=\begin{pmatrix}2&1\\1&2\end{pmatrix}$ 沿 $(1,1)$ 的倍率为三，沿 $(1,-1)$ 的倍率为一。

证明. 直接计算 $A(1,1)=(3,3)=3(1,1)$，以及 $A(1,-1)=(1,-1)$。两个输入都非零，故满足定义。

它们线性无关，任何平面向量都可分解为这两个方向的组合。在线性作用下，分别把两个分量乘以三和一再相加即可得到输出。这是此矩阵可对角化的具体解释，不代表所有矩阵都具有完整特征基。

□

观课提示 先预测哪些方向会保持在原直线上，再检查这些方向能否组成一组基。

原课程视频 本节网页与媒体 参数实验

要点回看

核心概念 某些方向经过线性变换后只被缩放而不偏离原直线。这些方向常使反复作用变得容易理解。

整理成数学语言 特征向量满足 $Av=\lambda v$ 且 $v\neq 0$。特征值由 $\det(A-\lambda I)=0$ 决定。若有足够多独立特征向量，可写 $A=PDP^{-1}$，进而 $A^n=PD^nP^{-1}$。

完整算例 $A=[\begin{smallmatrix}2&1\\1&2\end{smallmatrix}]$ 在 $(1,1)$ 上倍率为 3，在 $(1,-1)$ 上倍率为 1。

易错点 零向量虽满足方程却不是特征向量；实矩阵也可能没有实特征值，例如平面 90° 旋转。

自测与答案 上例 $A^4(1,1)$？答案 $(81,81)$，无需展开矩阵乘法。

Lean 对照 `LinearAlgebra.eigenvector_scaling`

证明特征方程在向量缩放后仍成立。若要称缩放结果为特征向量，还须另加原向量及缩放系数非零的条件。

1.15 快速求二维特征值

官方课程：`quick-eigen`

先找两个根的中心

二维特征多项式是 $\lambda^2-(a+d)\lambda+(ad-bc)$。不急着套二次公式，先把未知数移到两根的平均位置。令 $m=(a+d)/2$，写 $\lambda=m+r$，一次项便抵消，剩下的只是偏移距离 r。

整理得到 $r^2 = ((a-d)/2)^2 + bc$。对角元差和交叉作用的乘积共同决定偏移；迹的一半决定中心。配方因此揭示了结构，而不只是缩短算式。平方偏移的正负还告诉我们实根存在与否。

快速求根之后检查什么

对 $\left[\begin{smallmatrix}4 & 2\\ 1 & 3\end{smallmatrix}\right]$，中心是 3.5，平方偏移为 2.25，根是 $5, 2$。检查它们的和等于迹 7，积等于行列式 10。两项核对能发现不少算术或符号错误。

若根号内为负，在实数域没有根，在复数域仍有一对共轭根。旋转动作完全可以很自然，却没有实的保留方向；这个例子提醒我们区分“实数中无解”和“任何意义上都无解”。

若根重合，还必须看特征空间维数。单位矩阵保留所有方向，而非平凡剪切只有一条特征直线，虽然两者都能有重根。这个方法加快的是特征值计算，没有省掉判断可对角化的工作。把和、积、特征方向三种检查一起使用，快速计算才不会掩盖问题结构。

一步一步建立直观

1. 先找两个倍率的中心 二维矩阵的特征多项式是 $\lambda^2 - \mathrm{tr}(A)\lambda + \det(A)$。两根之和由迹决定，所以先把未知根写成中心加减偏移。右图对称例子的中心为两个对角项的平均数，这一步减少了未知信息。

2. 再让乘积决定偏移大小 两根相乘等于行列式，因此偏移的平方等于中心平方减去行列式。对实对称矩阵，这恰是 $((a-d)/2)^2 + b^2$，不会为负。可用右图改变非对角项，观察中心不变而两根向两侧分开的过程。

3. 速度不能代替适用条件 公式算出的根还应代回特征方程并检查方向。一般实矩阵的判别式可以为负，此时实平面中找不到非零的保持方向。快速方法的价值在于识别结构，并非免去对数域、重根和可对角化条件的判断。

> **观课提示** 把两根写成中心加减偏移，核对迹和行列式分别提供了什么信息。
>
> 原课程视频 本节网页与媒体 参数实验

要点回看

核心概念 两个根的平均值由迹的一半决定，偏离平均值的距离由判别式决定。先中心化，再处理平方根，比直接套长公式清楚。

整理成数学语言 令 $m = (a+d)/2$，则

$$\lambda_{\pm} = m \pm \sqrt{((a-d)/2)^2 + bc}.$$

根号内为负时，应在复数域解释。来源是特征多项式配方。

完整算例 $A=\left[\begin{smallmatrix}4&2\\1&3\end{smallmatrix}\right]$，$m=3.5$，根号内 $0.25+2=2.25$，故特征值 $5,2$。和为迹 7，积为行列式 10。

易错点 对角元平均值只是根的中心，通常不是其中一个根。

自测与答案 $\left[\begin{smallmatrix}0&-1\\1&0\end{smallmatrix}\right]$ 的根？根号内 -1，为 $\pm\mathrm{i}$。

Lean 对照 `LinearAlgebra.characteristic_centering`

验证配方后的特征多项式恒等式。平方根的取值与实根存在性在正文中单独讨论。

1.16 抽象向量空间

官方课程：abstract-vector-spaces

哪些推理其实没有用到箭头形状

回顾前面，许多证明只用了相加和数乘，并没有使用长度或角度。若另一类对象也支持相同规则，就可以搬用这些推理。函数相加按每个输入逐点相加，数乘把每个输出同比例缩放；多项式与矩阵也一样。

所有次数不超过二的实多项式，都唯一写成 $c_0+c_1x+c_2x^2$。所以 $1,x,x^2$ 扮演三根基，系数是坐标。这里“空间”首先是一种组织对象和操作的结构，不必真的把全部函数画进三维图中。

求导也能写成矩阵

求导把三根基依次送到 $0,1,2x$。把结果在原基下的坐标排成列，得到下面的求导矩阵。计算多项式导数就变成系数向量的矩阵变换。常数在核中，因为求导看不见它们；输出最多一次，所以像空间也受到限制。

这说明抽象并非远离应用：箭头上的基、核、线性映射，直接解释了函数运算。傅里叶级数会换用正弦和余弦作基本方向，热方程会研究各方向随时间的倍率，后续章节由此连接。

迁移结构时仍要核对封闭性。“恰好二次”的多项式相减可能成为一次或零，所以不组成这种向量空间；任意函数也不都可导，不能把求导当成整个函数空间上的算子。抽象的工作，是保留推理需要的条件，并明确哪些对象真正满足它们。

一步一步建立直观

1. 把坐标解释成函数的系数 在子空间 $\{a+bx:a,b\in\mathbb{R}\}$ 中，一个函数可以用系数对 (a,b) 表示。右图的箭头现在代表系数，不代表物体在平面中移动。函数逐点相加对应系数相加，因此同一个二维图可以帮助理解这个函数空间。

2. 抽象来自保留运算规则　把两份函数先相加再数乘，与分别数乘后相加，会在每一个输入点得到同一数值。这条分配律不依赖函数像不像箭头。向量空间的公理提取的正是这些可复用的运算关系，让几何语言能够进入函数、序列和多项式。

3. 选择基后才获得坐标图　在上述二维子空间中，1 与 x 可作基；若只保留 $1+x$ 的倍数，就像把右图限制在一条线上。坐标图依赖所选有限维子空间，不能由此宣称所有函数都只有两份坐标。抽象定义比这个示意图的适用范围更广。

> **观课提示**　把函数当成向量，跟踪求导对每个基函数的作用。
>
> 原课程视频　本节网页与媒体　参数实验

要点回看

核心概念　“向量”不必是箭头。多项式、函数、矩阵都可以成为向量，只要其加法与数乘满足同一组公理。

整理成数学语言　函数空间采用逐点运算：$(f+g)(x)=f(x)+g(x)$，$(af)(x)=af(x)$。微分算子 D 在可微函数组成的适当空间上满足 $D(af+bg)=aDf+bDg$。

完整算例　在基 $1, x, x^2$ 下，次数不超过二的多项式的求导矩阵为

$$\begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 2 \\ 0 & 0 & 0 \end{bmatrix}.$$

它把系数 (c_0, c_1, c_2) 送到 $(c_1, 2c_2, 0)$。

易错点　指定对象集合还不够；必须说明运算及封闭性。例如“恰好二次”的多项式相加可能不再恰好二次。

自测与答案　常数函数为什么在求导算子的核中？因为其导数恒为零。

Lean 对照　`LinearAlgebra.function_space_distribute`

验证整个实函数空间的数乘分配律。这里只形式化向量空间操作，未把任意函数都错误地视为可微。

章节综合：一次看懂反复作用

把本章的概念放进同一个问题：矩阵 $A = \left[\begin{smallmatrix} 2 & 1 \\ 1 & 2 \end{smallmatrix}\right]$ 反复作用于 $(3,1)$，怎样求第 n 次结果？先不要连乘矩阵。寻找被保留的方向 $u = (1,1)$、$v = (1,-1)$，它们的倍率分别为三和一，而且 $(3,1) = 2u + v$。

因此 $A^n(3,1) = 2 \cdot 3^n u + v = (2 \cdot 3^n + 1, 2 \cdot 3^n - 1)$。取 $n = 2$ 得 $(19,17)$，可用两次普通矩阵乘法交叉检查。换基把复合变成了对每份坐标单独乘倍率，特征向量的用途就在这里。

再问这个动作会丢失信息吗？两特征倍率都非零，行列式为三，面积扩大三倍且保持定向，所以可以求逆。若把第二倍率改为零，则所有沿 v 的变化都不可见，原来唯一的坐标恢复变成多解问题。这样，基、矩阵复合、行列式、核与特征值就在一个例子中汇合了。

完成本章后，试着分别用一幅图、一段文字和一组坐标式解释这个问题。三种表达应讲述同一个动作。这里的综合推导供纸笔检查；机器证明仍以各课列出的实际定理为准。

第二章　微积分：局部变化与累积

本章以实变量为主，角度使用弧度。导数是局部线性近似的系数，积分是有向累积；极限把近似推进为精确陈述。

2.1　从面积发现微积分

官方课程：essence-of-calculus

从一个熟悉但尚未解释的公式开始

圆面积是 πR^2，但若只背公式，就不知道它为什么与半径平方有关。把圆按距离中心的远近分成薄圆环：半径为 r 的那一层，长度约为周长 $2\pi r$，厚度为一个小增量。越外层越长，因此总面积是在累积不断增加的周长。

有限厚度的圆环并不等于长方形。外半径比内半径大 h 时，精确面积是 $2\pi rh+\pi h^2$；长方形近似少算了 πh^2。当我们把厚度切细，总误差能否趋零，才是从近似过渡到公式的关键。这也是极限需要出场的原因。

把累积的问题倒过来

现在假设面积函数已经知道，半径稍微增加时，面积增加多快？用新增面积除以 h，得到 $2\pi r+\pi h$，它趋向周长。于是“把周长累积起来”和“观察面积的瞬时变化”互相连接，这就是积分与导数关系的原型。

先在更简单的直线高度 $f(x)=x$ 上练习。到 x 为止的三角形面积是 $x^2/2$，多出宽度 h 时，新增矩形与小三角形的面积为 $xh+h^2/2$。除以宽度后，平均高度趋向边界高度 x。

看面积实验时，分别注意矩形总和与真实曲线间的差，不只看矩形数量增加。问题不是“小块很多所以看起来差不多”，而是能否让误差随切分变细而受控。后面的章节将先解释导数，再回到这一累积与变化的连接。

一步一步建立直观

1．先把累积当作一台计数器　把横轴看作时间，把曲线高度 x^2 看作某个时刻的流入速率。一个窄矩形的宽是持续时间，高是近似速率，因此面积是这段时间流入的量。先只放少量矩形：每

块都把一段变化中的速率当成常数，所以结果会偏离真正的累积。

2. 切得更细，究竟改善了什么 增加分割数时，每段内速率的变化变小，矩形顶部更贴近曲线。这里函数连续且递增，左端点和低估面积。关键不是图形越来越好看，而是可以把每段误差相加并控制总误差；无穷多小块只能在这样一个极限过程中得到明确含义。

3. 只移动终点，寻找新增的量 保持分割足够细，再改变右端点 b。原先已经累计的部分没有改变，新增部分是一条宽度约为 h、高度约为 b^2 的窄带。因此面积函数的增量约为 b^2h。将增量除以 h 后取极限，就把“累积的变化率”接回了原速率。

> **观课提示** 观察面积拆分中的小块近似，指出误差在哪里、为什么会随着切分变细而减小。
>
> 原课程视频 本节网页与媒体 参数实验

要点回看

核心概念 把圆切成许多薄圆环，半径 r 处的环面积近似为 $2\pi r\,\Delta r$。将所有圆环加起来，自然得到面积是周长的累积。

整理成数学语言 $A(R)=\int_0^R 2\pi r\,\mathrm{d}r=\pi R^2$，而 $A'(R)=2\pi R$。一般地，连续 f 的累积函数 $F(x)=\int_a^x f(t)\,\mathrm{d}t$ 满足 $F'=f$。

完整算例 对更简单的 $f(t)=t$，面积函数为 $F(x)=x^2/2$，因此增加小宽度 h 的面积是 $xh+h^2/2$；除以 h 后趋向高度 x。

易错点 薄圆环“长方形化”是极限近似，不能把有限厚度圆环直接当作精确矩形。

自测与答案 半径从 2 增到 $2+h$，圆面积多多少？精确值 $4\pi h+\pi h^2$，一阶近似 $4\pi h$。

Lean 对照 `Calculus.area_derivative`

完整验证简化面积函数 $x^2/2$ 的导数为 x；圆环几何分割未形式化。

2.2 导数的“零除以零”悖论

官方课程：`derivatives`

瞬间没有路程，怎么还能谈速度

速度表显示某个瞬间的速度，但若把时间间隔直接设成零，路程也为零，差商无法计算。解决方法并不是允许零除以零，而是保留非零的小时间间隔，观察平均速度是否逼近一个确定值。这个稳定趋势才是瞬时速度。

令位置函数为 $f(x)=x^2$，从 x 移动到 $x+h$，位移是 $2xh+h^2$。只要 $h\neq 0$，平均变化率就等于 $2x+h$。在 $x=3$ 处，小正步长给略大于 6 的斜率，小负步长给略小于 6 的斜率，两边向同一个数靠近。

切线是变化率的图像

在图上标出两点并连线，得到割线。让第二点靠近第一点，割线方向逐渐稳定为切线方向。动画中的切线不是凭目测贴在曲线上的装饰，而是差商极限的可见表示。把 h 从两侧接近零，可以检查左右是否一致。

再换成 $|x|$ 的尖角。零点右侧差商为 1，左侧为 -1，没有单一斜率能同时描述两边。图形没有断开，所以连续；却没有一个统一的局部线性变化率，所以不可导。这个例子把连续和可导的区别具体化了。

学习导数时要保留两句话：先计算非零间隔下的比值，再讨论其趋势。后面看见微小量记号时，也应能回到这条解释，而不把形式上的约分误读为对零进行除法。

一步一步建立直观

1. 先承认测量需要一段距离 把 $f(x)=x^2$ 当作位置读数。在时刻 a 和 $a+h$ 之间，平均速度由两点连成的割线斜率给出。拖动第二点，让 h 保持非零；此时分子、分母都来自真实的两次测量，没有任何一步要求除以零。

2. 缩小窗口，看什么保持稳定 展开平方可得割线斜率 $2a+h$。缩小 h 的同时，位移也在缩小，但它们的比值趋近 $2a$。右图把割线与切线同时保留，是为了把“一个仍有宽度的测量”和“测量结果的极限”区分开。导数描述后者。

3. 零斜率不等于函数不变化 把观察点移到 $a=0$。切线水平，而附近的函数仍向上弯曲。这里位置增量是 h^2，比一阶尺度 h 更小；所以局部一阶读数为零。判断一个量是否不变，还必须考察邻域、较高阶变化或额外条件，不能只看单点导数。

定义、命题与证明

定义：某一点的导数 若极限存在，定义 $f'(a)=\lim_{h\to 0,\ h\neq 0}\dfrac{f(a+h)-f(a)}{h}$。这里先在非零步长处计算，再讨论读数的极限。

命题：平方函数的差商与误差 对 $f(x)=x^2$ 及 $h\neq 0$，差商为 $2a+h$；因此 $f'(a)=2a$，差商的绝对误差恰为 $|h|$。

证明. 展开 $(a+h)^2-a^2=2ah+h^2$，因 $h\neq 0$，除以 h 得 $2a+h$。

与候选极限 $2a$ 相减后仅剩 h。对任意正容差，只要步长的绝对值更小，读数误差也更小；这给出了极限所需的控制，而非在 $h=0$ 处进行除法。

□

观课提示 从两侧让非零步长接近零，分别观察割线与切线。

原课程视频 本节网页与媒体 参数实验

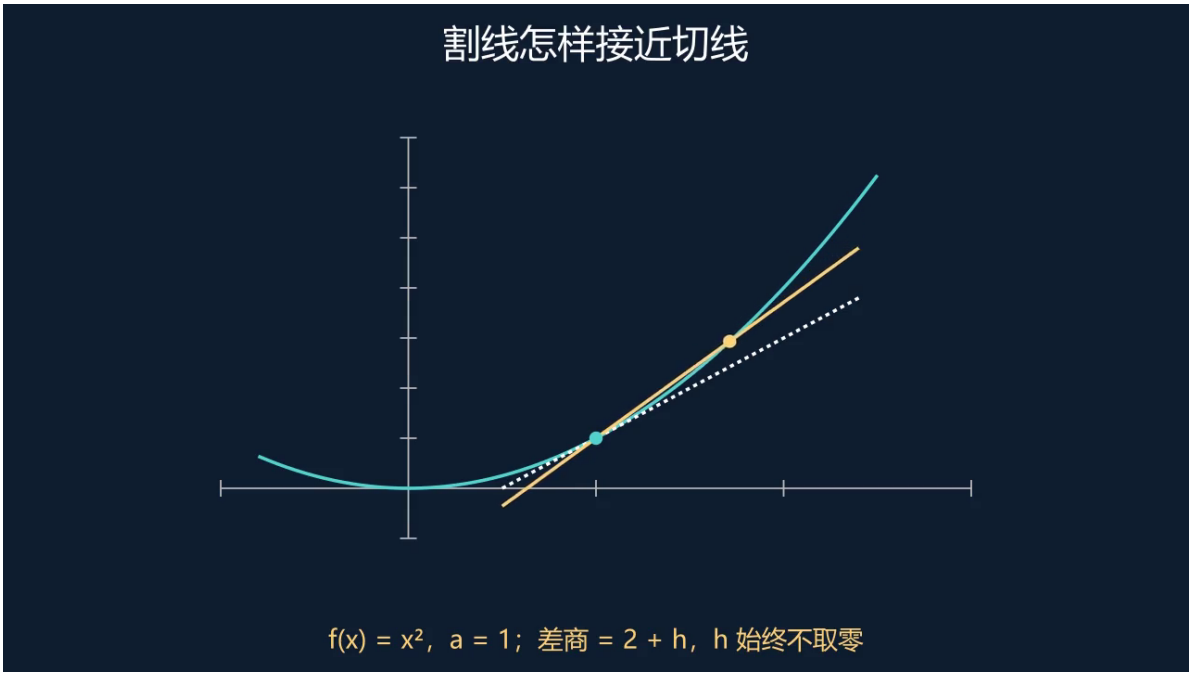


图 2.1: 平方函数的割线（黄）与切线（白色虚线）。非零步长缩小时，割线斜率趋向切线斜率。本项目原创 Manim 动画关键帧；完整动画与操作入口见本节网页。

要点回看

核心概念 瞬时速度不是把两个零相除，而是观察非零时间间隔的平均速度在间隔趋零时的趋势。

整理成数学语言 $f'(x)=\lim_{h\to 0,h\neq 0}\frac{f(x+h)-f(x)}{h}$。对 $f(x)=x^2$，

$$\frac{(x+h)^2-x^2}{h}=2x+h\quad(h\neq 0),\qquad f'(x)=2x.$$

完整算例 $x=3,h=0.01$ 时割线斜率为 6.01，它接近切线斜率 6，却并不等于它。

易错点 不应在约去 h 之前代入 $h=0$；可导也比连续更强，$|x|$ 在零点连续但不可导。

自测与答案 $|x|$ 在零点的左右斜率？右边 1，左边 -1，所以双侧导数不存在。

Lean 对照 `Calculus.square_secant`

验证非零步长下的差商恒等式，分母条件显式写为 $h \neq 0$。导数极限由下一节的形式化幂法则保证。

2.3 幂法则的几何来源

官方课程：`derivatives-power-rule`

为什么平方会带出一个二

边长为 x 的正方形，每边增加 h，新增区域由两条宽为 h 的长条和一个小角组成。两条长条总面积为 $2xh$，小角为 h^2。除以边长增量 h 后，长条贡献 $2x$，小角只贡献 h，趋零时消失。

这解释了二次幂导数中的系数二来自两条边，不是套公式的巧合。立方体则有三个主要薄层，每层体积约为 x^2h；交叠边角含更多个 h，在除以 h 后仍趋零，所以得到 $3x^2$。

从图形推广到代数

更高次幂不容易画，但展开乘积 $(x+h)^n$ 时可以数贡献。一阶项必须从 n 个因子中恰好选一个 h，其余选 x，一共有 n 种选法，所以系数是 nx^{n-1}。选两个及以上 h 的项都带 h^2 或更高次。

“高阶项小”必须说明与什么相比。导数计算比较的是它们相对于 h 的大小，而非只看绝对值。如果一个误差恰为 $3h$，它虽然趋零，除以 h 后仍留下 3，不能丢掉。

把 $x = 2, h = 0.1$ 代入立方增量，先算精确值，再算 $3x^2h$ 的预测，最后减出误差。这个小实验会把导数理解为有误差控制需求的局部预测器。自然数幂的规则可以统一证明，非整数幂则要重新核对定义域。

一步一步建立直观

1. 把立方体的变化拆成薄层 切换到立方函数，把边长 a 增加 h。三个面旁各添一层厚度 h 的薄片，主项合计为 $3a^2h$；棱边附近的重叠修正含 h^2，角上的小块含 h^3。体积图像解释了为什么系数是三，而不是一个需要记住的巧合。

2. 先除以步长，再比较阶数 对立方函数，差商是 $3a^2 + 3ah + h^2$。右图可先切回平方函数，比较其中只有一个一阶误差项的情形。两者都不是把高阶项直接删掉，而是在除以 h 之后，证明剩余误差随 h 趋零。然后再推广到整数次幂。

3. 把局部近似与大步长分开 将步长调大，割线明显偏离切线。幂法则给出的是观察点附近的一阶倍率，不能用它精确预测任意远处的变化。对于一般实数指数，还需要说明函数定义域；本窗口的平方、立方只是几何推导最直接的两个例子。

观课提示 找出新增面积或体积的一阶部分，并说明哪些角落在差商极限中消失。

原课程视频 本节网页与媒体 参数实验

要点回看

核心概念 边长增加 h 时，正方形多出两条窄带和一个小角；小角面积是二阶量。立方体的第一阶体积变化来自三个面。

整理成数学语言 对自然数 n，

$$(x+h)^n = x^n + nx^{n-1}h + O(h^2), \qquad (x^n)' = nx^{n-1}.$$

这里的余项表述对固定 x、有限多项式成立。实数非整数幂需要额外的定义域条件。

完整算例 x^3 的变化为 $3x^2h + 3xh^2 + h^3$。差商去极限得到 $3x^2$，在 $x = 2$ 处为 12。

易错点 “高阶项很小”必须相对于 h 比较；常数 $n = 0$ 的导数为零。

自测与答案 求 $f(x) = 5x^4 - 2x^2 + 7$ 的导数。答案 $20x^3 - 4x$。

Lean 对照 `Calculus.power_rule`

完整调用并验证自然数幂的一般导数定理；Lean 中自然数的 $n - 1$ 是截断减法，$n = 0$ 由前面的零系数正确处理。

2.4 三角函数的导数

官方课程：`derivatives-trig-functions`

从圆周运动读出两个变化率

单位圆上的位置写成 $(\cos\theta, \sin\theta)$。角度增加一点，点沿圆的切向移动；切向与半径垂直，相当于把位置方向转九十度，得到 $(-\sin\theta, \cos\theta)$。两坐标分别告诉我们横向与纵向变化速度。

在圆最右侧，纵向高度增长最快，正弦的变化率为一；来到最上方，高度到达峰值，一阶变化率为零；再向左上方走，高度开始下降。这样先预测符号，再看公式 $\cos\theta$，公式就变成了运动记录。

弧度不是可省略的单位约定

为什么速度向量恰好有单位长度？因为单位圆上的弧度增量等于弧长增量。若用“度”作为输入，一个单位角度只走 $\pi/180$ 的弧长，导数必须多一个换算因子。导数不仅关乎函数形状，也关乎输入单位。

几何论证背后有一个极限：小弧段与切线方向的直线段越来越接近，例如 $\sin h/h \to 1$。若用三角不等式或面积夹逼证明它，应在弧度约定下进行，不能把待证导数又拿来证明自身。

观看复平面动画时，同时观察位置箭头与垂直的速度箭头，再把它们分别投到横纵轴。你应能从图中解释正弦导数为何是余弦、余弦导数为何带负号。之后复合角度 $2x$ 的额外倍率，将由链式法则处理。

一步一步建立直观

1. 把函数值还原成圆上位置 在单位圆上，横坐标是余弦，纵坐标是正弦。先把角度调到四分之一圈：此时点在圆顶，纵坐标变化暂时最慢，而横坐标正向左快速变化。三角函数的导数因此可以先理解为圆周速度在两个坐标轴上的投影。

2. 速度为什么总与半径垂直 匀速转动时，速度朝切线方向，是位置向量逆时针转过直角后的方向。因此位置 $(\cos t, \sin t)$ 对应速度 $(-\sin t, \cos t)$。这里角度采用弧度，单位半径让角度增量和弧长一致；若采用度数，必须补上单位换算因子。

3. 改变半径，检查隐藏的假设 第三种情形加入径向增长，轨迹成为螺线。此时总速度同时包含切向分量和径向分量，不能再直接用“旋转九十度”表示全部速度。这个对照说明公式所依赖的是固定半径与单位角速度，而不是任意看起来像圆的运动。

> **观课提示** 把圆周运动的速度投影到两个坐标轴，解释导数的正负号。
>
> 官方图文课程 本节网页与媒体 参数实验

要点回看

核心概念 单位圆上匀速运动的切向量，把位置向量旋转 $90°$。因此横坐标 $\cos\theta$ 的速度是 $-\sin\theta$，纵坐标 $\sin\theta$ 的速度是 $\cos\theta$。

整理成数学语言

$$(\sin x)' = \cos x, \qquad (\cos x)' = -\sin x.$$

弧度让单位圆上的角度增量等于弧长增量，这是导数没有额外比例系数的原因。

完整算例 在 $x = \pi/3$ 处，正弦高度为 $\sqrt{3}/2$，变化率为 $1/2$。峰值 $x = \pi/2$ 处的变化率为零。

易错点 若输入用“度”，函数其实是 $\sin(\pi x/180)$，导数要乘 $\pi/180$。

自测与答案 $\sin(2x)$ 的导数是什么？答案 $2\cos(2x)$，需要链式法则。

Lean 对照 `Calculus.sine_derivative`

完整验证实正弦在任意实点的导数；采用 mathlib 的弧度三角函数。

2.5 乘积法则与链式法则

官方课程：`chain-rule-and-product-rule`

两个变化共同决定结果

面积 $f(x)g(x)$ 随 x 改变时，两条边都可能变化。新增面积包含一条 $g\,\Delta f$ 的带、一条 $f\,\Delta g$ 的带和角上的 $\Delta f\Delta g$。除以输入增量后，前两项分别成为 gf' 与 fg'。

角项为什么能消失？可导性意味着两条边的增量都是输入增量的一阶量，所以相乘是二阶量；再除以一次增量，仍会趋零。若没有这项数量级分析，“把角忽略”只是图像猜测。乘积法则计算的是两条独立贡献的和。

串行变化则把倍率相乘

复合函数的结构不同：输入先经过 f，中间输出再经过 g。附近小间距先放大约 $f'(x)$ 倍，再放大约 $g'(f(x))$ 倍，所以总倍率相乘。外层导数必须在实际收到的中间值处计算，不能把它误写成 $g'(x)$。

比较 $x^2\sin x$ 和 $\sin(x^2)$。前者同时改变两块因子，因此两条贡献相加；后者把平方结果送入正弦，因此倍率串联。先画计算关系，再选择法则，比看到两个符号便机械判断更可靠。

有时用差商约去中间增量来记忆链式法则很方便，但中间增量也可能为零，严格证明需要用局部线性近似处理。下一章的反向传播会重用同样结构：沿一条计算路径相乘，多条路径汇合时相加。

一步一步建立直观

1. 用放大倍率理解复合 设输入先经过 g，再经过 f。一个很小的输入位移先被放大为约 $g'(a)h$，随后又被放大为约 $f'(g(a))$ 倍。右图展示其中一台局部放大器：随着取样窗口缩小，曲线附近越来越像直线，而倍率就是切线斜率。

2. 为什么倍率相乘，却不总是常数 链式法则把两级倍率相乘，但第二级必须在第一级的输出位置测量。移动右图的观察点，切线斜率随位置改变，便能看到“先代入正确位置”不能省略。对每一级都使用同一个输入位置，是复合求导中常见的错误。

3. 乘积的变化来自两块薄条 乘积可以看作长方形面积。长度和宽度同时改变时，新增面积包含两块一阶薄条，还有一个二阶小角。除以输入步长后，小角项趋零，两块薄条分别留下 $f'g$ 与 fg'。右图仅演示差商极限；长方形分解才是乘积法则的几何理由。

观课提示 比较矩形两边同时变化与两级变化串联，辨认为什么一个相加、一个相乘。

原课程视频 本节网页与媒体 参数实验

要点回看

核心概念 面积 fg 的增量来自两条边的改变；复合函数的局部伸缩倍率则相乘。

整理成数学语言

$$(fg)' = f'g + fg', \qquad (g \circ f)'(x) = g'(f(x))f'(x).$$

乘积展开中还有 $\Delta f \Delta g$，可导性保证它除以 h 后趋零。链式法则中的外层导数要在内层输出处求值。

完整算例 $\frac{d}{dx}(x^2 \sin x) = 2x \sin x + x^2 \cos x$；$\frac{d}{dx} \sin(x^2) = 2x \cos(x^2)$。

易错点 乘积的导数一般不是导数的乘积；把外层求导点写成 x 也常出错。

自测与答案 e^{3x^2} 的导数？答案 $6x\mathrm{e}^{3x^2}$。

Lean 对照 `Calculus.chain_rule`

`chain_rule` 与同文件的 `product_rule` 完整保留两函数在相应点可导的假设。

2.6 为什么自然常数 e 特别

官方课程：eulers-number

增长速度为何取决于已有数量

设一个量每时每刻都按当前大小的一定比例增长。已有量越多，同样短时间里增加越多，数学关系是 $N' = rN$。这与每次增加固定数量不同；后者形成直线，前者的增加量自身也在增长。

对于 a^x，差商可拆成 $a^x(a^h - 1)/h$。当前位置的因子 a^x 被提了出来，其余极限只取决于底数。不同底数给不同的比例常数；选择使该常数等于一的底数，就得到自然底数 e。它特殊在于局部变化规则最简洁。

连续复利是一条通向同一对象的路

把单位时间分成 n 小段，每段按 $1/n$ 的比例增长，得到 $(1 + 1/n)^n$。分段越来越细时，它趋向 e。有限分段仍有自身的数值，不能把每次 “增长百分之一” 的离散模型直接当成连续模型。

要把一般正底数纳入同一语言，写 $a^x = \mathrm{e}^{x \ln a}$。内层把时间刻度缩放 $\ln a$ 倍，链式法则于是给出导数 $a^x \ln a$。额外因子可以理解为增长时钟的快慢，而不是一项无来由的常数。

例如希望一个量一单位时间翻倍，连续增长率应取 $\ln 2$。若取 $r = 1$，一单位时间的倍率是 e，并非二。练习时同时检查数量单位与时间单位；将来面对微分方程，指数函数将作为“变化率与自身成比例”的基本解再次出现。

一步一步建立直观

1. 寻找增长速度等于当前存量的刻度　选择指数函数。把纵轴读作某个不断增长的存量，切线斜率就是瞬时增长速度。自然指数的特殊性在于：把时间单位选定后，任一点的增长速度恰好等于该点的存量。这是一条局部演化规则，而不是从一个小数近似出发的定义故事。

2. 对照不同函数，辨认这一性质　切换回平方函数，比较读数中的函数高度和切线斜率。平方函数只有个别位置会让二者相等；指数函数则处处相等。要检验“处处”这类命题，有限幅图只能帮助发现规律，最终仍需导数公式或微分方程的论证。

3. 有限时间的增长不能直接线性外推　把步长调大，平均增长率不再等于起点的瞬时增长率。因为增长过程中存量本身持续增大，后面的增长也更快。复利中的细分极限把这种持续反馈纳入计算；因此一段时间的倍率是指数，而不是永远沿起点切线走。

> **观课提示**　把底数与局部相对增长率对应起来，区分离散增长和连续增长。
>
> 原课程视频　本节网页与媒体　参数实验

要点回看

核心概念　指数增长的变化速度与当前数量成比例。当这个比例常数为 1 时，对应的底数就是 e。

整理成数学语言　$\frac{d}{dx}\mathrm{e}^x = \mathrm{e}^x$，且对 $a > 0$，$a^x = \mathrm{e}^{x \ln a}$，所以 $(a^x)' = a^x \ln a$。

完整算例　连续增长模型 $N(t) = N_0 \mathrm{e}^{rt}$ 满足 $N' = rN$。当 $r = \ln 2$ 且时间单位为一时，$N(1) = 2N_0$。

易错点　增长率 r 的单位是时间的倒数；“每年增长 100%”的离散模型与 $r = 1$ 的连续模型不同。

自测与答案　2^x 在零点的斜率？答案 $\ln 2$，不是 1。

Lean 对照　`Calculus.exponential_derivative`

完整验证实指数在每个实点的导数；增长模型只作为数学例子。

2.7 隐函数求导

官方课程：implicit-differentiation

关系式已经足够描述曲线

圆由 $x^2+y^2=1$ 描述，不必先把 y 解成一个全局函数。沿圆移动时，x 和 y 不能自由独立变化，它们必须共同维持这个约束。若 x 增加一点，另一个坐标要怎样配合，才能仍在曲线上?

对关系作一阶变化，得到 $2x\,dx+2y\,dy=0$。这里的 dy 不是任意小量，而是曲线约束要求的响应。若 $y\neq 0$，解出 $dy/dx=-x/y$，就得到局部斜率，无须展开平方根。

斜率不存在，不代表切线不存在

在 $(3/5,4/5)$ 处，斜率为 $-3/4$。向右走一点必须略向下，才能补偿横坐标平方的增加。切线给出这种补偿的一阶近似；有限移动后的点通常不在圆上，因为曲线还存在二阶弯曲。

来到 $(1,0)$，公式不能除以 y。这并非圆突然不光滑，而是此处不能用单值的 $y(x)$ 斜率描述垂直切线。改用 x 作为 y 的局部函数，就能恢复合适的坐标语言。一个几何对象是否规则，与某种表示是否方便，是两个问题。

一般关系 $F(x,y)=0$ 中，沿曲线总变化为零，得到 $F_x+F_y y'=0$。使用它时仍须核对相关可微性及局部函数存在条件。Lean 对照只验证微分约束后的代数解法；不要把这一点误认为已经证明任意隐式关系都能解成光滑函数。

一步一步建立直观

1. 先把等式当作运动约束 令右图保持单位半径。圆方程 $x^2+y^2=1$ 要求点的两个坐标协同变化：一个变大时，另一个必须做出补偿。沿圆运动的小位移 (dx,dy) 在一阶上满足 $2x\,dx+2y\,dy=0$，这比先求出上下两个平方根分支更直接。

2. 梯度告诉我们不能朝哪里走 约束的梯度 $(2x,2y)$ 沿半径向外，切向位移与它垂直，所以一阶变化为零。转到圆的左端点，切线竖直；此时不能把附近的整段圆当成一个可微的 $y(x)$。隐式关系仍然有意义，某种坐标表达却可能失效。

3. 把可除的条件写出来 从约束式得到 $dy/dx=-x/y$ 时，必须假设 $y\neq 0$。回到圆顶，分母非零，斜率为零；回到左右端点，分母为零，提示应换用 $x(y)$ 或参数表示。公式里的限制并非技术障碍，而是图形姿态在坐标中的真实反映。

> **观课提示** 沿约束曲线移动，观察两个坐标必须如何互相补偿，并留意垂直切线。
>
> 原课程视频 本节网页与媒体 参数实验

要点回看

核心概念 曲线由关系式定义，不必先显式解出 y。沿曲线移动时，定义关系的总变化必须为零。

整理成数学语言 若 $F(x, y(x)) = 0$ 且相关函数可微，则 $F_x + F_y y' = 0$。对 $x^2 + y^2 = 1$，当 $y \neq 0$ 时 $y' = -x/y$。

完整算例 在 $(3/5, 4/5)$ 处，切线斜率为 $-3/4$，切线方程为 $3x + 4y = 5$。

易错点 圆不全局定义单值 $y(x)$；在 $(\pm 1, 0)$ 处上述商不能使用，切线垂直。

自测与答案 在同一点沿切线有 $dx = 0.04$，一阶估计 dy。答案 -0.03。

Lean 对照 `Calculus.circle_tangent`

从微分约束 $2x\,dx + 2y\,dy = 0$ 和 $y \neq 0$ 推导 $dy = -x\,dx/y$；隐函数存在性定理未形式化。

2.8 极限与导数的定义

官方课程：`limits`

关心附近的趋势，而不是点上的标签

把 $(x^2 - 1)/(x - 1)$ 的输入逐渐靠近一，会发现输出靠近二。原表达式在一处无定义，但这不妨碍附近的趋势稳定。极限问的是：输入足够接近时，输出能否被迫足够接近某个目标，而不是问在那个点代入后得到什么。

对 $x \neq 1$，因式分解允许约去 $x - 1$，表达式等于 $x + 1$。这说明邻近行为与一个简单连续函数相同。即使人为把原函数在一处的值设成七，极限仍是二；只有把该点值补成二，它才连续。

从数值观察走向保证

算几个小步长只能提示候选极限，不能证明所有更小步长都满足要求。一个函数可能在你没取到的点强烈振荡。证明需要把“足够近”写成一种对整片邻域有效的保证，下一节将把这个保证拆成输入和输出两种误差。

左右也必须分别检查。一个跳跃函数从左边靠近与从右边靠近可能指向不同值；单侧趋势各自稳定，不代表双侧极限存在。导数的差商同样需要这项检查，绝对值函数的尖角就是一个例子。

最后区分三个问题：函数在一点有没有值，附近有没有极限，极限是否等于点值。连续性把后二者连接，可导性再要求差商具有极限。把这些问题逐个问清，比笼统地说“函数很平滑”更精确。

一步一步建立直观

1. 极限关注附近，不强迫取到目标点 先观察非零步长下的割线。极限问题不是要求在 $h = 0$ 计算差商，而是问：当 h 从允许的非零值靠近零时，差商能否被逼近到某个固定数。右图保留两个点，让“接近”与“已经重合”在操作上分开。

2. 误差应当可以被事先控制 平方函数的差商与 $2a$ 相差恰好 h。因此只要要求 $|h|$ 小于给定误差，就能确保读数达到目标精度。这比观察小数位看起来稳定更强：它说明对任意要求的精度，都有一个明确的输入范围保证结果。

3. 检查两侧与定义域 改变步长的正负，从观察点两侧接近。两侧趋向同一值，才得到通常的双侧导数；若函数只在一侧有定义，应声明研究的是单侧极限。窗口提供的其他函数也有自己的定义域，尤其不能在实对数的非正输入处假装得到函数值。

> **观课提示** 区分函数在点上的值与附近的趋势，检查左右靠近是否得到同一结果。
>
> 原课程视频 本节网页与媒体 参数实验

要点回看

核心概念 极限描述输入接近某处时输出是否能被控制到任意精度，并不要求函数在该点已有定义。

整理成数学语言 连续性是 $\lim_{x\to a} f(x) = f(a)$；导数是另一函数“差商”的去心极限。多项式连续，所以 $\lim_{x\to a} x^2 = a^2$。

完整算例 $(x^2-1)/(x-1)$ 在 $x = 1$ 未定义，但对 $x \neq 1$ 等于 $x+1$，因此极限为 2，可用补值修复连续性。

易错点 极限存在并不意味着原函数在该点连续；左右极限也可能不同。

自测与答案 把上例在 $x = 1$ 处定义成 7，极限变了吗？没有，仍为 2，但不连续。

Lean 对照 `Calculus.square_limit`

完整验证平方函数的双侧极限；使用邻域过滤器表达“趋近”。

2.9 把任意精度写清楚

官方课程：`epsilon-delta`

把“再近一点”改成一场有规则的挑战

对方先指定输出误差 $\varepsilon > 0$，你必须给出一个输入范围 $\delta > 0$，保证之后任何落在范围内的输入都满足要求。关键是顺序：不能先挑一个有利的输入，再为它选择范围；也不能只展示几个成功的输入。

在线性函数 $f(x) = 3x + 2$ 中，输出误差恰为输入误差的三倍。要把输出误差压到 ε 以下，取 $\delta = \varepsilon/3$ 即可。这个选择展示了完整逻辑：从目标精度反推输入限制，然后对所有满足限制的输入验证。

非线性时先给变化倍率一个上界

平方函数的误差是 $|x^2 - a^2| = |x - a||x + a|$。第二个因子也随输入变化，不能当常数。先要求 $|x - a| < 1$，就有 $|x + a| < 2|a| + 1$；再取

$$\delta = \min\left(1, \frac{\varepsilon}{2|a| + 1}\right),$$

两个要求就同时成立。这个额外的一并不是答案本身，而是先把局部倍率圈住的工具。

你可以选择更小的 δ，定义不要求最佳范围。证明的目标是给出有效保证，而非求最大容许误差。这样的自由度常让推导简洁许多。

画一条包住目标值的水平细带，再找围绕输入点的竖直细带，使图像在竖带内都落进横带。图形解释了量词，但最终仍需不等式把所有点一起控制。这是从视觉直觉走向严格论证的一次完整练习。

一步一步建立直观

1. 先由读者提出允许误差 设你希望平方函数在 a 处的差商距离 $2a$ 不超过某个容差。先确定这个输出要求，再决定步长要缩到多小。极限定义中精度与邻域的先后次序，表达的正是“无论要求多严格，我都能给出足够近的范围”。

2. 把图形的窄带翻译成不等式 在本例中误差等于 $|h|$，因此可取与容差一样小的步长阈值。并非所有函数都如此简单：有时要先限制在一个较小邻域，再估计函数值的上界。证明的任务是找到能保证整段邻域都合格的条件，而非只试出几个漂亮样本。

3. 即使切线水平，量词仍然相同 把观察点移到零，目标斜率也变成零。此时仍需对所有足够小的非零步长给出控制，不能因为曲线在屏幕上显得平坦就结束论证。有限像素可能把小误差隐藏起来，严格定义则要求我们继续缩小任意指定的误差。

观课提示 先指定输出误差，再寻找统一有效的输入范围；不能只验证几个样本。

官方图文课程 本节网页与媒体 参数实验

要点回看

核心概念 对方先指定输出容忍误差 ε，我们再给输入误差上限 δ。顺序不能反过来，也不能让 δ 随后来选择的 x 改变。

整理成数学语言

$$\forall \varepsilon > 0\ \exists \delta > 0\ \forall x,\quad 0 < |x-a| < \delta \Rightarrow |f(x)-L| < \varepsilon.$$

完整算例 对 $f(x) = 3x + 2, L = 3a + 2$，选择 $\delta = \varepsilon/3$，因为输出误差恰为 $3|x-a|$。

易错点 证明极限不能只代几个小步长。对于平方函数，$|x^2 - a^2| = |x-a||x+a|$ 还需要控制第二个因子。

自测与答案 若 $f(x) = 5x - 1$，可取怎样的 δ？取 $\varepsilon/5$ 即可。

Lean 对照 `Calculus.epsilon_delta_linear`

显式构造并验证斜率 3 的函数的 ε–δ 证明；该结论允许 $x = a$，比去心版本更强。

2.10 洛必达法则：比较消失速度

官方课程：`l-hopitals-rule`

两个量都趋零，比例仍可能完全不同

x/x、x^2/x 与 x/x^2 的分子分母都趋零，比值却分别趋向一、零，或在零附近无界。零除以零不是一种答案，它只是提示直接代入没有揭示相对消失速度。需要进一步比较两个量的一阶或更高阶变化。

若 $f(a) = g(a) = 0$，且两者可导、$g'(a) \neq 0$，局部近似给出 $f(a+h) \approx f'(a)h$ 与 $g(a+h) \approx g'(a)h$，因此比例接近导数之比。这个简单版本提供了直觉，但一般洛必达法则可以处理更复杂的去心区间情况，需要明确的条件。

法则是有条件的工具

下面列出的条件要求相关区间可微、分母导数不为零，并且导数比值具有极限。使用时先判断原式的未定形式，再核对这些条件。仅仅看到一个分式，不意味着可以把分子分母各求一次导数。

例如 $(x+1)/(x+2)$ 在零附近没有未定性，直接连续代入得 $1/2$；随意求导反而得到一。又如 $(x^2-1)/(x-1)$ 在一处，可以因式分解直接求极限，没必要把简单结构藏到一般定理里。

如果一次求导后仍是未定式，再用一次也要重新核对条件。洛必达法则的假设不满足，并不能反推原极限不存在，只能说明这个工具暂时不给结论。解题时应先看能否约分、换元或利用已知局部近似，再考虑更一般的方法。

一步一步建立直观

1. 把消失的量看成一阶位移 当两个函数同时趋零时，它们的比值可能仍有极限。若在目标点附近分别近似为 $f'(a)h$ 和 $g'(a)h$，共同的小步长可以消去，留下局部变化率之比。选择正弦函数，观察原点附近正弦曲线与斜率为一的直线如何贴合。

2. 先弄清为什么允许比较导数 窗口切换到平方函数，展示差商的控制机制，但它本身不是一般洛必达法则的证明。一般法则还需要区间上的可导性、分母导数不为零及导数比值的极限等条件。其理由来自中值定理对邻域中变化率的联系，不能只靠画两条切线代替。

3. 若一阶项消失，检查更高阶结构 改用立方函数并把观察点移到零，一阶导数也为零。此时分子与分母可能以不同阶数消失，必须继续分析主导项，或在每次确认条件后再次使用法则。不要把任意分式直接替换成导数之比；只有符合条件的极限问题才有这种结论。

> **观课提示** 先识别未定式，再追踪分子分母的消失速度，记录适用条件。
>
> 官方图文课程 本节网页与媒体 参数实验

要点回看

核心概念 当分子分母同时趋零时，比值取决于它们接近零的速度。导数可揭示这个速度，但需要足够的可微性条件。

整理成数学语言 一个常用 0/0 版本要求：f, g 在去心区间可微，$g' \neq 0$，$f, g \to 0$，且 $f'/g' \to L$；在相应单侧区间条件下可推出 $f/g \to L$。双侧结论须两侧一致。

完整算例 $\lim_{x\to 1}(x^2-1)/(x-1) = 2$。可以用导数比值 $2x/1$，也可以先因式分解，后者在此更直接。

易错点 原比值不是未定式时不能随便求导。例如 $(x+1)/(x+2)$ 在零点的极限为 $1/2$，导数比值却为 1。

自测与答案 求 $\lim_{x\to 0}(\mathrm{e}^x - 1)/x$。答案 1，由指数函数在零点的导数定义即可得到。

Lean 对照 `Calculus.removable_limit`

完整验证上述有理函数的去心极限，方法是局部化简与连续性；没有把它标成一般洛必达法则的形式化证明。

2.11 积分与微积分基本定理

官方课程：integration

把变化率逐段累积起来

若知道一辆车每个时刻的速度，想求一段时间的净位移，可以把时间切成小段，每段用速度乘时长估计位移，再把贡献相加。速度变化时，有限切分存在误差；积分是切分变细后稳定的有向累积。

为什么叫有向？向前速度贡献正位移，倒退速度贡献负位移。净位移可能为零，但总路程并非零。同理，曲线在横轴下的部分在定积分中带负号，几何面积则需要另外处理绝对值。

原函数为什么能省掉求和

若速度本来就是某个位置函数 F 的导数，每小段的真实位移是两个端点的位置差。把相邻小段相加，中间端点一正一负地抵消，只剩 $F(b)-F(a)$。基本定理将这种望远镜式相消，与用导数近似小段变化的极限论证连接起来。

因此找到原函数是一种高效计算累积的方法，而不是积分唯一的定义。无法写出简单原函数时，定积分仍可能存在，也可通过数值求和逼近。反过来，形式上写出一个导数表达式，也应检查区间条件，不能跨过奇点随意用端点差。

在矩形实验中逐步加密切分，先看左端点和是否偏低，再解释原因。本例函数递增，每个矩形高度取在小段最左端，因而低于曲线。这样你不仅看到数值趋近，还知道误差的方向；这是数值计算与数学解释配合的方式。

一步一步建立直观

1. 从速率记录重建总量 如果只能读取瞬时流速，怎样恢复一段时间内的流入量？先用每个小区间起点的流速乘以持续时间，再把各段相加。右图的矩形把这个过程画出来。积分把“以离散记录估计累计”的方法推广为可控的极限。

2. 让分割变细，但不要混淆两种变量 增加矩形数时，终点不变，改变的是估计精度；移动终点时，问题本身的累计范围改变。前者通向定积分的定义，后者产生面积函数。把这两种操作分开，就能理解为什么积分既是一个数，也可以随上限变化成为函数。

3. 由新增薄带得到基本定理 将终点略向右移，新增面积约为端点函数值乘以移动距离。连续性保证这条窄带内部的高度不会突然偏离端点高度，于是面积函数的导数恢复原函数。再用任意原函数之差计算定积分，正是把这个局部事实连接到整个区间。

定义、命题与证明

定义：可变上限的面积函数　设 f 在区间上连续，固定左端点 a，定义 $F(x)=\int_a^x f(t)\,dt$。这是随终点变化的有向累积量。

命题：连续速率的累积恢复原速率　在区间内部，面积函数满足 $F'(x)=f(x)$。

证明. 把增量写成 $F(x+h)-F(x)=\int_x^{x+h} f(t)\,dt$。减去 $f(x)h$ 后，剩余量为 $\int_x^{x+h}(f(t)-f(x))\,dt$。

连续性保证足够短的区间内 $|f(t)-f(x)|$ 任意小。因此剩余量的绝对值至多为该小量乘 $|h|$。除以 $|h|$ 后取极限，差商与 $f(x)$ 的差趋零；正负步长都适用。

□

观课提示　观察小矩形累积以及误差方向，再把积分解释成端点变化。

原课程视频　本节网页与媒体　参数实验

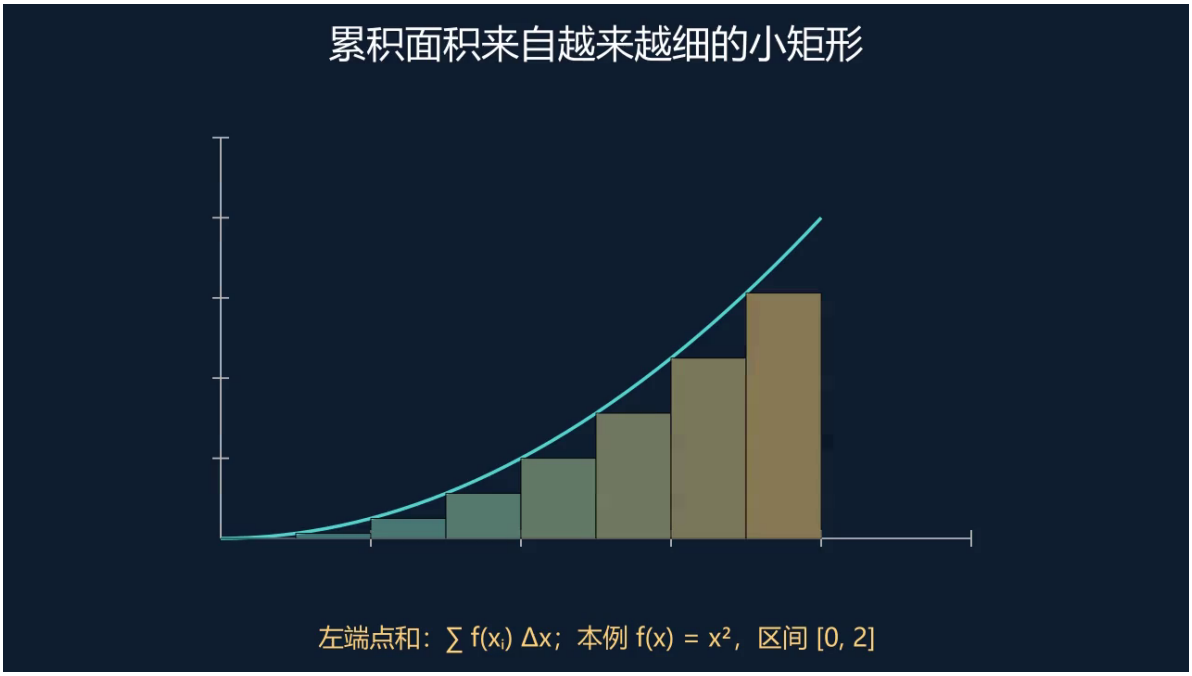


图 2.2: 平方函数的左端点矩形和。切分变细时，累积面积逐渐接近积分。本项目原创 Manim 动画关键帧；完整动画与操作入口见本节网页。

要点回看

核心概念　积分把每个位置的微小贡献累积起来。如果已知一个变化率的原函数，就能用端点差得到总变化。

整理成数学语言　若 F 在包含区间的范围内可导且 F' 在区间可积，则

$$\int_a^b F'(x)\,\mathrm{d}x = F(b)-F(a).$$

连续被积函数的原函数可由累积积分构造，这是基本定理的另一方向。

完整算例 $\int_1^3 2x\,\mathrm{d}x = 3^2 - 1^2 = 8$。把积分上下限交换，结果变成 -8。

易错点 定积分是有向累积，不总是几何面积；被积函数为负的部分带负号。原函数的常数在端点差中抵消。

自测与答案 $\int_{-1}^1 x\,\mathrm{d}x$？答案 0；两侧几何面积之和却是 1。

Lean 对照 `Calculus.fundamental_theorem`

调用完整的基本定理第二部分，保留区间上处处可导及导函数区间可积的精确假设。

2.12 面积为什么与斜率相连

官方课程：area-and-slope

边界向右移动，只有一条窄带发生变化

定义 $F(x)$ 为从固定起点累积到 x 的面积。比较 $F(x+h)$ 与 $F(x)$，大部分旧区域完全相同，相减后只剩边界附近的一条窄带。因此求面积函数的变化率，不需要重新计算整块面积。

新增面积除以宽度 h，就是这条带的平均高度。如果原函数在边界连续，带内所有高度在足够窄时都接近 $f(x)$，平均高度也被夹在附近，所以极限就是 $f(x)$。连续性在这里负责让“小区间里高度差不多”成为可靠判断。

把近似拆成可看见的两部分

取 $f(t) = t$，从 x 到 $x+h$ 的面积包含高度为 x 的矩形和高为 h 的小三角形。总增量 $xh + h^2/2$ 除以 h 后为 $x + h/2$，趋向边界高度。三角形就是有限步长近似的误差来源。

注意 $F(x)$ 的高度与 $F'(x)$ 的斜率不是同一种量：前者已经累积，后者描述新贡献进来的速度。某个点两者数值碰巧相同，不意味着两函数相同；检查单位也能发现这个区别。

反向移动边界时，窄带被减去，宽度也带负号，同样得到一致的导数。由此可见有向积分不是人为增加的麻烦，而是让累积与变化在两个方向都遵守同一规则。下一节将继续追问：变化速度自己又如何改变？

一步一步建立直观

1. 分别观察边界高度和已积面积 右端点处的曲线高度只描述“此刻增加得多快”，而此前曲线下的面积描述“到此刻已累积多少”。移动终点时，哪怕已积面积很大，新增长条的高度仍只由边界附近决定。导数关注局部增量，因此不会把整个历史面积直接拿来当斜率。

2. 粗糙估计也能先说明结构 减少矩形数后，估计会有明显误差，但每一块仍是高度乘宽度，物理单位也保持正确。若高度单位是米每秒、宽度是秒，面积单位就是米。单位检查让我们在正式求极限前，就能发现把速率与位移混在一起的错误。

3. 用误差夹逼支持直观 对递增连续函数，左端点和与右端点和分别从下方和上方夹住面积。分割细化使二者之差趋零，直观上的“贴紧曲线”就有了定量表达。本图只画左端点和；你可以用端点高度差与区间宽度自行估计另一侧的误差界。

> **观课提示** 移动面积边界，只看新添窄带，说明平均高度为什么趋向边界高度。
>
> 原课程视频 本节网页与媒体 参数实验

要点回看

核心概念 右边界移动 h，新增窄条的高度接近 $f(x)$。窄条面积除以宽度，便接近原图的高度。

整理成数学语言 $F(x+h)-F(x)=\int_x^{x+h} f(t)\,\mathrm{d}t$。若 f 在 x 连续，平均高度趋向 $f(x)$，所以 $F'(x)=f(x)$。

完整算例 $f(t)=t$ 时，$F(x)=\int_0^x t\,\mathrm{d}t=x^2/2$。在 $x=2$ 的累积面积为 2，面积增长速度也为 2，两者单位不同。

易错点 累积函数的高度与斜率是不同量；其数值偶然相同并无普遍意义。

自测与答案 从 $x=2$ 移到 $2+h$ 的精确面积增量？$2h+h^2/2$；一次近似忽略的部分是三角形面积。

Lean 对照 `Calculus.triangle_area`

完整验证 $\int_0^a x\,\mathrm{d}x=a^2/2$，也涵盖有向积分 $a<0$ 的情况。

2.13 高阶导数

官方课程：`higher-order-derivatives`

“正在下降”与“下降得越来越慢”可以同时成立

看位置函数的一阶导数，是在问速度；再对速度求导，则是在问加速度。两者的符号回答不同问题。一辆向左运动但正在减速的车，速度为负，加速度却可能为正。函数图像的下降与向上弯曲也可以同时发生。

以 x^2 在负半轴为例，斜率 $2x$ 为负，但随着 x 增加，它越来越接近零。图像仍下降，却下降得越来越慢，二阶导数为正。把“正二阶导数”误读为“函数递增”，就是把变化率与变化率的变化混在一起。

高阶信息用于更细的局部预测

只知道一点的值，可以给常数预测；知道一阶导数，可以加直线修正；知道二阶导数，就能预测斜率如何偏离初值，从而加上二次项。泰勒近似会把这套逐层修正系统化。

对 x^3，二阶导数 $6x$ 在零点两侧变号，所以凹凸改变。对 x^4，二阶导数 $12x^2$ 在零点为零，却不变号，所以“二阶导数等于零”并不足以判断拐点。需要检查邻域行为，而非只代一个点。

试着画函数、斜率函数和二阶导函数三幅图，并用同一竖线标出一个输入。每幅图的纵坐标分别回答“在哪里、往哪变、变化率怎么变”。把它们分开阅读，才不会在更复杂的近似或动力系统中混淆层次。

一步一步建立直观

1. 先读位置、速度，再读速度的变化 一次导数描述输出对输入的小变化有多敏感，二次导数则描述这种敏感度如何改变。把输入看成时间时，位置、速度、加速度依次出现。右图用指数函数的多项式近似演示逐阶匹配：一次项调整倾斜，二次项开始调整弯曲。

2. 同一条切线不能规定整条曲线 把观察位置移得更远，低阶多项式与原函数可能迅速分离。两条曲线在一点共享函数值与一阶导数，只保证它们在一阶尺度上接近；二阶导数及更高阶信息可以不同。因此切线相同不是函数相同，更不是远处预测可靠。

3. 高阶信息增加局部约束 增加阶数，检查读数中的实际误差。这里指数函数在零点的各阶导数都等于一，所以系数写成阶乘分母。一般函数需要分别计算各阶导数，并估计余项。即使已匹配许多阶，仍应同时注明近似中心、使用区间和允许误差。

> **观课提示** 分别解释函数值、斜率和斜率的变化，寻找下降但向上弯曲的例子。
>
> 原课程视频 本节网页与媒体 参数实验

要点回看

核心概念 一阶导数描述变化，二阶导数描述变化率怎样改变。运动中分别对应速度和加速度；函数图像中二阶导数控制局部凹凸。

整理成数学语言 对 $f(x)=x^3$，$f'=3x^2$，$f''=6x$，$f'''=6$。一般 $f''(a)>0$ 意味着附近斜率正在增加，但不保证 $f'(a)>0$。

完整算例 $f(x)=x^2$ 在 $x=-1$ 处仍在下降，因为 $f'=-2$，同时向上弯曲，因为 $f''=2$。

易错点 $f''(a)=0$ 不是拐点的充分条件；x^4 在零点二阶导数为零，但凹凸性不变。

自测与答案 $f(x)=\sin x$ 的四阶导数？答案仍为 $\sin x$，每四次循环。

Lean 对照 `Calculus.second_derivative_cubic`

完整计算三次函数的二阶导数为 $6x$，同时核实中间的一阶导函数。

2.14 泰勒多项式与余项

官方课程：taylor-series

用简单函数保留局部信息

假设要在零附近估算 e^x，直接计算指数不如多项式方便。先用常数一匹配函数值，再加 x 匹配斜率，再问二次项该是多少。若写 $1+x+cx^2$，它在零处的二阶导数为 $2c$，而指数的二阶导数为一，所以 $c=1/2$。

继续下去，每次求导都会带下一个整数因子，x^k 求 k 次导数后得到 $k!$。所以要匹配原函数的第 k 阶导数，系数必须除以 $k!$。泰勒公式由此来自逐阶匹配，而不是先出现一个无穷求和式。

多加一项不等于宣布全局正确

在动画里比较不同阶多项式与指数：中心附近先贴合，远处仍可能明显偏离。近似是否够用取决于输入范围和容许误差，不能只说“已经用了三阶”。若在区间上能界住下一阶导数，就能用余项估计把误差写成具体上界。

例如 $\mathrm{e}^{0.1}$ 的二阶近似为 1.105，而余项大小不超过 $\mathrm{e}^{0.1}0.1^3/6$。这里多出来的三次幂解释了小输入带来的精度；当输入很大时，同样阶数未必有用。

有限阶近似还不等于无穷级数表示。即使函数无穷可微，所有导数信息也可能不足以恢复邻域中的函数；下面的反例提醒我们核对解析性或余项趋零条件。Lean 对照给出一个精确多项式案例，不能替代一般余项定理。

一步一步建立直观

1. 把多项式做成局部仿制品 首先让近似函数在中心点拥有相同高度，再让斜率相同，然后匹配弯曲及更高阶变化。系数除以阶乘，是因为对幂函数重复求导会产生相应的阶乘。右图比较指数函数与在零点展开的多项式，每加一阶就增加一个局部匹配条件。

2. 近似好坏同时取决于阶数和距离 固定阶数而把观察点远离中心，误差通常变大。右图的数值读数比肉眼重合更可靠，但仍只是一个点的误差；要承诺整段区间的精度，还须对余项给出统一上界。近似式如果不附中心与误差范围，很容易被使用到不适合的地方。

3. 区分有限公式与无穷级数　增加阶数，在同一点比较误差是否缩小。指数函数的泰勒级数在每个实数点收敛到原函数，但这并不代表任意无穷可导函数都等于自己的泰勒级数。有限阶公式配合余项先成立；让阶数趋向无穷，还需要余项确实趋零。

定义、命题与证明

定义：指数函数在零点的泰勒多项式　定义 $P_n(x) = \sum_{k=0}^{n} x^k/k!$。每一项用于匹配指数函数在零点的一个更高阶导数。

命题：在固定区间内控制截断误差　若 $|x| \le R$，则 $|e^x - P_n(x)| \le e^R R^{n+1}/(n+1)!$。此界对区间中的所有点同时成立。

证明. 由带拉格朗日余项的泰勒公式，余项为 $e^\xi x^{n+1}/(n+1)!$，其中 ξ 位于零和 x 之间。指数函数具有公式所需的各阶导数。

因为 $|\xi| \le R$，有 $e^\xi \le e^R$；再用 $|x|^{n+1} \le R^{n+1}$ 得到上界。图中单点实际误差通常更小；统一上界的作用是事先保证整个使用区间的精度。

□

观课提示　逐次匹配值、斜率和更高导数，观察近似在离开中心后怎样变化。

原课程视频　本节网页与媒体　参数实验

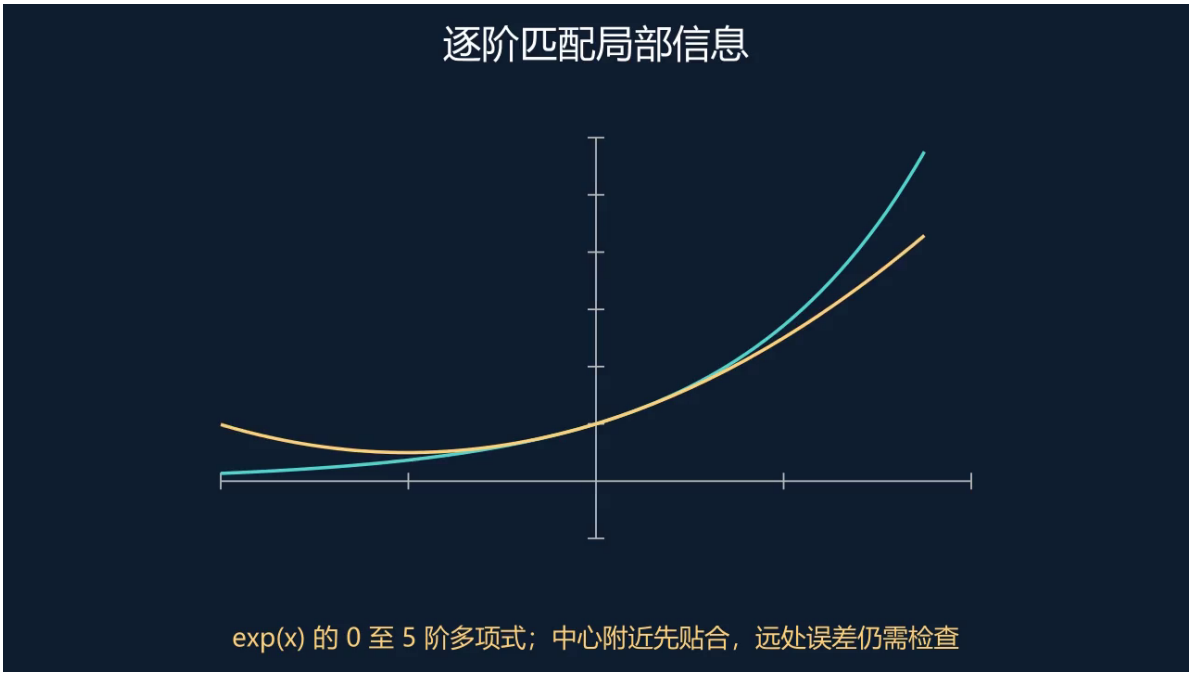


图 2.3: 指数函数（蓝）与逐阶泰勒多项式（黄）；精度需要结合离中心距离判断。本项目原创 Manim 动画关键帧；完整动画与操作入口见本节网页。

要点回看

核心概念 用多项式逐次匹配函数在一点的值、斜率、曲率等信息。匹配更多导数改善局部近似，但不自动保证无限级数等于原函数。

整理成数学语言

$$T_n(a+h)=\sum_{k=0}^{n}\frac{f^{(k)}(a)}{k!}h^k.$$

在相应区间有足够连续导数时，拉格朗日余项为 $f^{(n+1)}(\xi)h^{n+1}/(n+1)!$，其中 ξ 位于两端之间。

完整算例 对 $f(x)=x^3$，$f(a+h)=a^3+3a^2h+3ah^2+h^3$ 是精确等式；在 $a=1,h=0.1$ 得 1.331。

易错点 无穷可微不等于解析。把 e^{-1/x^2} 在零点补为零后，其零点泰勒级数恒零，却不等于邻域中的函数。

自测与答案 $\mathrm{e}^{0.1}$ 的二阶近似？$1+0.1+0.005=1.105$，须用余项界判断精度。

Lean 对照 `Calculus.cubic_taylor_exact`

验证三次多项式的精确泰勒恒等式；一般泰勒定理、级数收敛和余项界未纳入这一形式命题。

2.15 泰勒展开的面积解释

官方课程：taylor-series-geometric-view

二分之一从一块三角形出现

设 $F'=f$，想估计 $F(a+h)-F(a)$。第一步把高度固定为 $f(a)$，得到矩形面积 $f(a)h$。但若高度随着输入增加，它的变化也会贡献面积。把高度函数用直线近似，末端比起点高约 $f'(a)h$，新增区域近似三角形。

三角形底为 h、高为 $f'(a)h$，所以修正是 $\frac{1}{2}f'(a)h^2$。二阶项的二分之一因此可以看成变化高度的平均效果。若斜率为负，修正也带负号；有向面积让同一公式继续成立。

一次累积把幂次提高一级

高度函数的一阶变化形成累积函数的二阶修正。若继续近似高度本身的弯曲，再积分，就会产生三阶项。积分每次提高幂次并除以新的次数，这与泰勒系数里的阶乘相呼应。

对 $F(x) = x^2$，导数 $f(x) = 2x$ 本来就是直线，所以矩形加三角形已经精确，没有更高误差。对一般函数，三角形只模拟局部高度变化，还需要估计曲线与这条直线之间的面积差。

把本节与上一节对照：一种办法逐阶匹配导数，另一种从累积几何解释修正。两种路线得到同一局部表达，但适用的光滑条件仍要保留。看图时不要把“画起来像三角形”直接升级为精确等式，要先判断高度函数是否真的线性。

一步一步建立直观

1. 从导数的误差返回函数的误差 若导数只用常数近似，函数只得到一条切线；若把导数的线性变化也纳入，积累起来就得到二次修正。积分让一个关于斜率的局部误差变成面积，所以泰勒系数可以通过反复累计来理解，而不必只是代数配系数。

2. 反复积分为什么带来阶乘 在从零到 x 的区间上，常数积成 x，线性项积成 $x^2/2$，再次累计又引入新的分母。右图的指数函数各阶导数重复自身，这让 $1/n!$ 的结构特别明显。每一层累计都把先前的变化分摊到整段区间上。

3. 余项就是尚未解释的累积 把观察点移远，查看新增高阶项仍未覆盖的差距。积分形式的余项把它写成高阶导数与权重的积分，因而可以用导数上界控制。图形里的两条线看似重合，只表示当前显示尺度下误差小；余项估计才说明这种小有多可靠。

> **观课提示** 把二阶修正看成一块有向三角形，说明什么时候是精确面积。
>
> 官方图文课程 本节网页与媒体 参数实验

要点回看

核心概念 一阶近似对应窄条的矩形面积；若高度也在变化，再加入三角形面积，就出现二阶修正中的 $1/2$。

整理成数学语言 由 $F' = f$，在足够光滑时

$$F(a+h) - F(a) = f(a)h + \frac{1}{2}f'(a)h^2 + o(h^2).$$

它正是 F 的二阶泰勒式。

完整算例 取 $F(x) = x^2$，$f(x) = 2x$。矩形贡献为 $2ah$，三角形贡献为 h^2，且此时余项恰为零。

易错点 三角形近似依赖 f 的局部线性性；对一般曲线、有限大 h，不能把它当精确面积。

自测与答案 $F(x) = x^2/2$ 的一次近似误差？精确为 $h^2/2$，不随展开点 a 改变。

Lean 对照 `Calculus.quadratic_error`

精确验证平方函数减去一次近似后等于 h^2；这给出二阶修正的可核查例子。

2.16 导数作为局部变换

官方课程：`derivatives-and-transforms`

暂时不画函数曲线，改画两条数轴

把输入数轴上的小点送到输出数轴，函数就成为点的移动规则。在 a 附近放两个很近的点，它们的输出间距约是输入间距的 $f'(a)$ 倍。正导数保序，负导数反序，绝对值表示局部放大程度。

对平方函数，在三附近，输入间距大约被放大六倍；在负三附近，倍率是负六，输出次序反转。零点的导数为零，说明一阶间距被压掉，但二阶变化仍在，所以附近函数没有变成常数。

线性代数与微积分在这里接上

局部表达 $f(a+h) = f(a) + f'(a)h + o(h)$，把固定输出位置与输入微扰的线性响应分开。推广到多变量，数字 $f'(a)$ 会变成一个线性映射或雅可比矩阵；它描述附近小箭头怎样改变，正如前面的矩阵变换。

复合函数先通过内层局部变换，再通过外层局部变换，所以导数相乘。多变量情形使用矩阵乘法，顺序依旧是右边先做。链式法则因此不仅是一条求导技巧，也是一条局部动作的复合规律。

请在纸上画输入与输出两条平行数轴，用连线表示三个邻近输入的像，再根据间距预测导数正负。这个视角尤其适合解释函数图像难以直接画出的高维问题，也为神经网络中的敏感性分析准备了语言。

一步一步建立直观

1. 在输出中追踪一根小输入箭头 把观察点固定，在输入轴上画一段小位移 h，再比较输出变化。曲线通常不是线性变换，但在足够小的尺度上，输出变化可由一个固定倍率近似。这个倍率就是导数。右图的割线把有限位移联系起来，切线显示局部替代模型。

2. 缩小尺度，分离主项与误差 局部线性化应写成 $f(a+h) - f(a) = f'(a)h + o(h)$。余项不仅要趋零，还要相对于输入尺度 h 更小。平方函数的剩余项是 h^2，因此除以 h 后消失。这个相对误差条件才支持把导数视作真正的一阶变换。

3. 从一个倍率走向矩阵 移动观察点，局部倍率随位置更新。在多变量情形，小位移有多个方向，导数便成为作用在位移向量上的线性映射，用雅可比矩阵表示。本图只展示一维版本；链式法则在多维中相应变成矩阵复合，并须保持输入输出空间的次序。

观课提示 把函数看成数轴上点的移动，判断局部间距的放大与反向。

原课程视频 本节网页与媒体 参数实验

要点回看

核心概念 把函数看成移动数轴上的点。导数说明一点附近的间距被放大多少，以及是否反向；这种看法可自然推广为多变量的线性映射。

整理成数学语言 可导性也可写作 $f(a+h)=f(a)+f'(a)h+o(h)$。复合的线性部分相乘，正是链式法则。

完整算例 $x\mapsto x^2$ 在 $a=3$ 附近约放大 6 倍，在 $a=-3$ 附近反向并放大 6 倍。在零点导数为零，但函数没有在整个邻域变成常数。

易错点 局部倍率为零并不排除高阶变化；导数是某点的一阶信息。

自测与答案 先缩放 b 倍，再缩放 a 倍，总倍率？ab；若 a,b 均为负，复合保持方向。

Lean 对照 `Calculus.derivative_composition_scale`

完整验证两个线性缩放复合后的导数为 ab。一般非线性链式法则另见本章对应定理。

2.17 散度与旋度

官方课程：`divergence-and-curl`

流动场的一根箭头还不够

向量场在每个位置给出速度。要判断附近区域膨胀还是旋转，不能只看中心那根箭头，还要看周围箭头怎样变化。放入一个很小的方框，比较各边的流入与流出，就能测量局部净扩张趋势。

对 $F=(P,Q)$，左右边的水平流量差主要来自 P_x，上下边的垂直差来自 Q_y。除以小面积再取极限，得到散度 P_x+Q_y。它关心的是速度变化带来的净流出，而不是速度本身大不大。

环流与膨胀是不同测量

再沿小方框边界绕一圈，计算速度沿切向的贡献。相对边的差产生 Q_x-P_y，这是平面旋度。场 $(-y,x)$ 让小桨轮转动，却不膨胀，所以散度零、旋度非零。

场 $(2x, -2y)$ 则横向拉伸、纵向压缩，瞬时面积效应抵消，散度为零，也没有局部环流。零散度不是“什么都没发生”，而是两方向的面积效应恰好平衡。对这种线性场，追踪小方形会比看单根箭头清楚。

这些量都是局部一阶信息。把它们与整个区域的边界流量联系起来，需要散度定理或斯托克斯定理及其条件。这里只建立局部测量的意义；遇到全局势函数问题，还必须检查定义区域，而不能仅凭某一点或某个局部公式作结论。

一步一步建立直观

1. 先看小桨轮，而非单个粒子 选择纯旋转场。每个点沿圆周运动，速度处处与半径垂直；把一个很小的桨轮放入场中，它会旋转，但一小团粒子的面积不会改变。对于场 $F = (-\omega y, \omega x)$，散度为零，平面旋度为 2ω。旋度读数与角速度差一个因子二。

2. 再看小片区域是否膨胀 选择径向增长场，速度与位置成正比。小片区域的两条边都按同一比例增长，面积的瞬时相对增长率因此是 2λ。这解释散度为什么把不同坐标方向的伸缩率相加。单条轨线远离原点，并不单独构成散度的定义。

3. 旋转与压缩可以同时发生 第三种场把旋转与收缩叠加，粒子沿螺线向内运动。右图展示的是这一常系数场的轨线；散度与旋度分别检测局部面积变化和局部转动。一般空间变化的场需要在每个位置做局部计算，不能只看整条轨迹像不像旋涡。

> **观课提示** 同时放入小方框和小桨轮，分别判断净流出与局部环流。
>
> 原课程视频 本节网页与媒体 参数实验

要点回看

核心概念 散度衡量微小区域的净流出趋势，旋度衡量局部环流趋势。二者是不同的局部测量。

整理成数学语言 对平面向量场 $F = (P, Q)$，$\nabla \cdot F = P_x + Q_y$，标量旋度为 $Q_x - P_y$。对于 $F = (ax, by)$，散度为 $a + b$；对于 $F = (-y, x)$，散度为零、旋度为 2。

完整算例 场 (x, y) 让面积按瞬时相对速率 2 膨胀；旋转场 $(-y, x)$ 保持局部面积却产生旋转。

易错点 一个点散度为零只意味着局部净流出抵消，不表示场向量为零；旋度为零的全局势函数结论还与区域拓扑有关。

自测与答案 $F = (2x, -2y)$ 的散度和旋度？均为零，但它仍会拉伸一个方向并压缩另一个方向。

Lean 对照 `Calculus.planar_divergence`

计算对角线性向量场的两个偏方向导数之和；未形式化散度定理或斯托克斯定理。

章节综合：为一次近似负责

想估算 $f(1+h)=\mathrm{e}^{(1+h)^2}$，先问需要多高精度，而不是立即展开。链式法则给出 $f'(x)=2x\mathrm{e}^{x^2}$，再用乘积法则得到 $f''(x)=(2+4x^2)\mathrm{e}^{x^2}$。所以一点附近的二阶近似为

$$f(1+h)\approx \mathrm{e}+2\mathrm{e}h+3\mathrm{e}h^2.$$

当 $h=0.01$ 时，预测值为 $1.0203\,\mathrm{e}$。一次近似只保留 $1.02\,\mathrm{e}$，两者差 $0.0003\,\mathrm{e}$，正是二阶修正，而非某种计算器误差。

要声称准确到某位小数，还需给余项界。求得 $f'''(x)=(12x+8x^3)\mathrm{e}^{x^2}$ 后，在一到 1.01 的区间给它一个上界 M，误差即可控制在 $M|h|^3/6$ 内。局部匹配与可证明精度因此是相邻但不同的步骤。

请依次解释内层平方改变输入刻度、外层指数改变输出、二阶导数修正斜率变化、余项约束剩余误差。若能把这四句连起来，导数法则与泰勒公式就不再是分离的计算技巧。

第三章　神经网络：从像素到序列

本章把计算规则、训练目标和经验解释分开。公式能描述网络如何运算；“学到了怎样的概念”需要实验，不能仅凭结构图断言。

3.1　神经网络的基本单元

官方课程：neural-networks

从输入数字到可调整的判断

一张灰度小图可以按固定顺序展成向量，每个分量记录一个像素。要判断图中类别，最简单的尝试是给各像素分配权重，求加权和，再加一个偏置。正权重把某些像素的亮度当作支持证据，负权重当作反对证据，偏置调节整体阈值。

单个神经元接收这个分数，再通过非线性函数输出激活。以 $2x_1 - x_2 + 1$ 为例，输入 $(1, 3)$ 时分数为零，Sigmoid 输出一半。这里权重、输入、激活是三类对象：训练调整权重，样本提供输入，前向计算产生激活。

为什么堆很多层之前要插入非线性

如果每层只是仿射映射，前后层可以合并：第二层展开后，仍是一个新矩阵乘原输入再加新偏置。增加深度只把同一个仿射动作拆成多步，并没有得到新的非线性边界。激活函数打断这种合并，允许不同输入区域产生不同响应。

用图形理解时，可以把第一层看成建立一组可调整的测量，后续层组合这些测量。但“某个神经元识别一条笔画”只是可能的解释，不能仅凭网络结构认定训练必然得到这种分工。

本节先把计算规则写清楚，还没有解决权重从哪里来。下一节引入损失，把“判断好不好”压缩成可优化的数，再利用微积分修改参数。理解前向计算、评价结果、更新参数三步的区别，是后续所有模型的基础。

一步一步建立直观

1. 让一个单元先做一件能看清的事　输入向量不是一张图片的含义，而是被编码成数值的特征。右图只保留两个输入：权重决定各输入的影响方向和强度，偏置移动触发门槛。先计算加权和 z，

再经过 ReLU 得到输出，这样便能把“收集证据”与“决定是否激活”分成两个步骤。

2. 开关边界怎样切分输入空间 关闭单元时，加权和落在零以下，ReLU 把输出截成零。右图固定第二输入，只画输出随第一输入变化的曲线；折点对应 $z = 0$，两侧分别是关闭段和激活段。拖动输入点，观察跨过折点后的变化：单个单元只产生简单的分段线性结构，多层网络通过组合这些结构形成更复杂的表示。

3. 训练改变规则，推理使用规则 给定目标输出后，损失衡量当前结果与目标的差异。调权重或偏置是在改变这条输入输出规则；只拖动输入是在用既定规则计算新的样本。二者应明确区分。此处只有一个标量目标和一个单元，不代表真实分类网络的全部架构与训练过程。

> **观课提示** 区分输入、权重和激活，并解释为什么连续仿射层需要非线性才有新的表达能力。
>
> 原课程视频 本节网页与媒体 参数实验

要点回看

核心概念 把输入表示为向量后，每一层先做带偏置的线性组合，再经过非线性函数。参数控制哪些输入影响哪些输出。

整理成数学语言

$$z^{(\ell)} = W^{(\ell)} a^{(\ell-1)} + b^{(\ell)}, \qquad a^{(\ell)} = \sigma(z^{(\ell)}).$$

例如 $\sigma(z) = 1/(1 + \mathrm{e}^{-z})$。若输入维度为 n、输出为 m，权重形状为 $m \times n$，偏置有 m 个分量。

完整算例 单神经元 $z = 2x_1 - x_2 + 1$，输入 $(1, 3)$ 时 $z = 0$，Sigmoid 输出为 $1/2$。

易错点 没有非线性的多层仿射网络仍可合并成一个仿射映射，深度本身不会产生任意复杂决策边界。

自测与答案 $x \mapsto 3(2x + 1) - 4$ 可合并为什么？$6x - 1$；若中间插入 ReLU，一般不能这样整体合并。

Lean 对照 `NeuralNetworks.affine_composition`

精确验证两个标量仿射层的复合公式，说明非线性激活为何重要。

3.2 梯度下降如何学习

官方课程：`gradient-descent`

参数空间里的高度是什么

网络参数很多，每一组参数都给一组预测。用损失函数评价预测误差，就可以把参数想象成位置、损失想象成高度。学习过程希望移动到较低的位置；虽然高维地形画不出来，局部变化仍能用导数描述。

对小位移 d，损失的一阶变化约为 $\nabla L \cdot d$。若固定步长大小，这个点积在反梯度方向最小。因此选择 $d = -\eta \nabla L$，预测变化为 $-\eta \|\nabla L\|^2$。这个理由依赖局部近似，不是对任何大步都有效的保证。

用一个可精确求解的地形检验直觉

在 $L(w) = w^2$ 中，梯度为 $2w$，更新后 $w_+ = (1 - 2\eta)w$。损失每步乘以 $(1 - 2\eta)^2$。现在能精确看见学习率的作用：小步缓慢靠近，较大但合适的步长可能左右摆动着靠近，过大则越走越远。

在交互实验中先选 $\eta = 0.2$，再选 0.8，最后选 1.1。运行前分别预测序列的符号与绝对值变化；动画给出的应是公式的结果，而不是“训练总会下降”的预设故事。

真实网络的地形更复杂，可能有平坦方向、鞍点和噪声梯度。即使训练损失下降，也不直接保证对新数据有效。因此本节的二次模型用于理解更新机制与步长条件，不能被扩展成所有模型全局收敛的证明。

一步一步建立直观

1. 把损失想成脚下的地形 右图采用 $L(w) = w^2$，最低点在 $w = 0$。导数告诉我们在当前位置向右走会使损失增加还是减少，所以更新时沿导数的反方向迈步。步长还乘以学习率；它既不是导数本身，也不是每次必须走固定距离。

2. 跨过最低点不一定表示失败 把学习率提高后，轨迹可能左右交替。写出误差递推 $w_{k+1} = (1 - 2\eta)w_k$，就能看出只要这个倍率的绝对值小于一，交替仍会逐步缩小。图形里的来回摆动需要结合幅度判断，不能一见振荡就断言不收敛。

3. 过大的更新会放大误差 再提高学习率，误差在两侧反复放大。这里目标函数光滑而且只有一个最低点，算法仍然会失败，说明步长本身是关键条件。高维神经网络的曲率随位置变化，不能把这个玩具函数的稳定区间直接照搬成通用学习率建议。

定义、命题与证明

定义：梯度下降的一维更新 对可微损失 L，给定学习率 η，按 $w_{k+1} = w_k - \eta L'(w_k)$ 更新参数。学习率控制使用局部斜率的步幅。

命题：二次损失的精确稳定区间　对 $L(w)=w^2$，任意初值的迭代在 $0<\eta<1$ 时收敛到零；非零初值在 $\eta>1$ 时误差绝对值增长。

证明. 因为 $L'(w)=2w$，更新式成为 $w_{k+1}=(1-2\eta)w_k$，从而 $w_k=(1-2\eta)^k w_0$。

几何序列趋零当且仅当倍率的绝对值小于一（非零初值）。解 $|1-2\eta|<1$ 得 $0<\eta<1$。端点 $\eta=0$ 不更新，$\eta=1$ 对非零初值等幅往返，均不属于该收敛区间。

□

观课提示　先预测不同学习率下是靠近、振荡还是发散，再运行同一个二次损失实验。

原课程视频　本节网页与媒体　参数实验

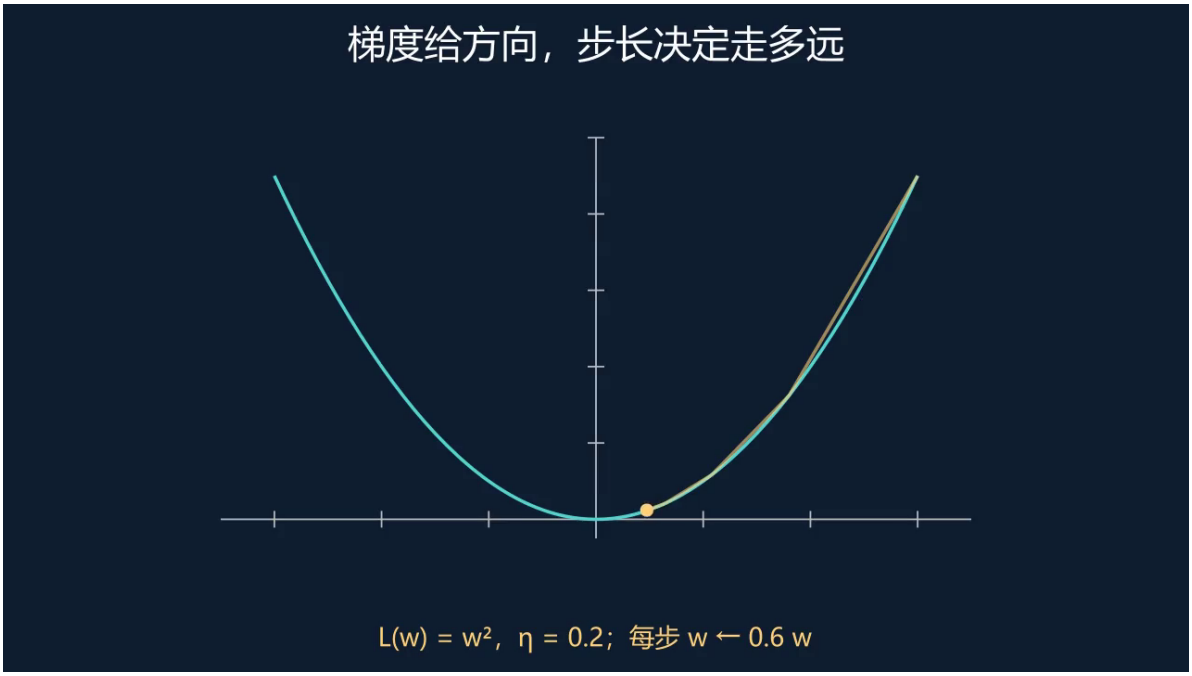


图 3.1: 二次损失上的更新轨迹；这里学习率为 0.2。网页可改用过大步长观察发散。本项目原创 Manim 动画关键帧；完整动画与操作入口见本节网页。

要点回看

核心概念　损失把预测误差压缩为一个标量。梯度指出局部增长最快的方向，向反方向走小步可望减少损失。

整理成数学语言　$\theta_{k+1}=\theta_k-\eta\nabla L(\theta_k)$。一阶近似给出 $\Delta L\approx-\eta\|\nabla L\|^2$，但它只在步长足够小时可信。

完整算例　对 $L(x)=x^2$，更新为 $x_+=(1-2\eta)x$。当 $0\le\eta\le1$ 时损失不增；若 $0<\eta<1$ 且 $x\neq0$，严格减小。$\eta=1$ 时只是来回翻号。

易错点 深网络的非凸损失一般不保证达到全局最优；训练损失低也不自动等于新数据上表现好。

自测与答案 $x = 3, \eta = 0.1$，更新后损失？$x_+ = 2.4$，损失 5.76，原先为 9。

Lean 对照 `NeuralNetworks.quadratic_gradient_step`

完整证明上述一维二次损失在闭区间步长条件下不增加；没有推广为所有网络的收敛保证。

3.3 检查网络究竟学到了什么

官方课程：`neural-network-analysis`

把好听的解释变成可以检验的假设

我们容易把隐藏神经元想象成边缘、笔画或概念检测器，但这种解释必须与实际参数和行为相符。先固定网络，找出什么输入能强烈激活某个单元，再查看看似相同的特征在不同背景下是否仍产生类似响应。一次漂亮的示例不足以建立稳定解释。

把输入稍微改变，观察输出怎样变，是更直接的实验。仿射层满足 $\Delta z = W\Delta x$：偏置在相减时消失，所以敏感性由权重和扰动方向共同决定。某方向的小变化可能被放大，另一些方向可能被忽略。

准确、稳定与可信不是同一个量

即使原图被正确分类，平移、旋转或加一点扰动后也可能变化；普通全连接层并不会自动知道这些变化应保持类别。高置信度也只是模型输出分布尖锐，并不保证它遇到的是熟悉数据，更不保证解释正确。

设计检查时可以分三步：保留独立测试样本观察泛化；做有控制的扰动检验稳定性；对置信度分组比较预测与实际频率。三步回答不同问题，不应让一个总体准确率代替所有分析。

本节的小线性算例让你能精确追踪一次扰动，但真实网络的解释还涉及非线性与数据分布。读课程中的可视化时，应同时问“这张图支持了什么结论”和“还排除了哪些替代解释”。数学公式与实验证据在这里承担不同角色。

一步一步建立直观

1. 把一个猜测变成可检查的扰动 如果猜测某个输入特征在控制输出，应固定其余输入，只改变这一项。右图的单个单元提供最简单的干预：在激活区域内，输出对第一输入的变化率就是第一权重。这个数值解释了局部敏感度，却还没有证明它在真实数据中代表某种语义。

2. 同一权重在不同输入处作用不同　把输入移到关闭区域。权重没有改变，输出对小扰动却暂时不敏感，因为激活函数将这一段截平。由此可见，只阅读权重大小不足以解释网络行为；还需要输入位置、激活状态，以及后续层如何使用这个单元的输出。

3. 解释也必须经受反例检验　调节偏置与目标，寻找能让原先解释失效的输入。一个“看起来检测到某特征”的单元，可能依赖背景或与该特征相关的另一信号。本窗口展示干预的逻辑，不承担真实模型的语义发现；更完整的分析还应比较样本、消融与输出变化。

> **观课提示**　把对神经元的解释当成待检验假设，观察输入扰动是否支持它。
>
> 官方图文课程　本节网页与媒体　参数实验

要点回看

核心概念　隐层激活不必对应人类想象的边缘或笔画。观察权重、输入扰动和分布外样本，才能检验解释是否与实际计算一致。

整理成数学语言　对仿射层，扰动满足 $\Delta z = W\Delta x$；在欧氏算子范数下 $\|\Delta z\| \le \|W\|\|\Delta x\|$。这是局部敏感性分析的起点。

完整算例　$z = 5x + b$，输入增加 0.02，激活前的数值增加 0.1。偏置不影响这一差值。对于像素图，平移会改变很多输入坐标，普通全连接层并不天然平移不变。

易错点　Softmax 的最大概率高，并不保证概率校准，也不意味着输入属于训练分布。

自测与答案　为什么只检查训练集准确率不够？模型可能记忆训练样本；还需独立测试及扰动检查。

Lean 对照　`NeuralNetworks.affine_input_sensitivity`

精确验证单输入仿射模型的扰动等式；范数界、校准与泛化的经验主张未形式化。

3.4　反向传播的直觉

官方课程：`backpropagation`

从最后的误差倒推每个中间量的影响

前向计算从输入走向预测，反向传播则从损失出发，询问每个中间量改变一点会让损失怎样变。输出端已经知道“希望预测往哪里改”，链式法则把这份敏感性逐层传回更早的节点。

沿一条路径，每经过一个运算都乘上它的局部导数；若一个中间量通过多条路径影响损失，要把贡献相加。这个规则来自复合与和的求导法则，并不是神经网络额外发明的数学。

共享参数为什么必须累加贡献

让同一参数出现在 $L(w)=(w-1)^2+(2w)^2$ 两个分支中。第一条路径贡献 $2(w-1)$，第二条贡献 $8w$。只看第一条会把最优点误判为一；两条相加得到 $10w-2$，真正最优点为 $1/5$。

反向计算的效率来自复用：同一个中间节点到最终损失的敏感性算一次，就能继续传给它的多个输入。若每个参数都从头展开整个损失，会重复大量共同计算。计算图把这些共享结构明确地记录下来。

一次反向传播给出梯度，还没有改变参数；优化器之后才使用梯度更新。若批量损失取平均，梯度也要平均，否则批量大小一变就暗中改变了更新尺度。下一节将把“沿路径乘、汇合时加”写成一个完整的小例子。

一步一步建立直观

1. 先沿计算方向记下中间量 从输入计算加权和，再计算激活输出，最后得到平方损失。反向传播需要这些前向中间量，因为每个局部导数都要在当前输入处求值。先在右图核对 z、输出和目标，才能解释梯度读数为什么具有当前的符号与大小。

2. 从输出误差追问谁该改变 损失先告诉输出应该增大还是减小，再通过激活函数传到加权和，最后乘以各输入，得到各权重应承担的变化率。这个连乘是链式法则。若一个参数通往损失的路径不止一条，各条路径的贡献还要相加，而不能任选其中一条。

3. 局部通路可能暂时关闭 当 ReLU 输入严格小于零时，这条通路的局部导数为零，即使当前损失很大，也不会通过它传回非零梯度。在零点本身，经典导数不存在，程序需选定实现约定。这里展示单个节点的机制；大型网络只是把同样的依赖关系按图重复计算。

> **观课提示** 沿每条影响路径相乘，在汇合处相加，留意共享参数不能漏掉分支。
>
> 原课程视频 本节网页与媒体 参数实验

要点回看

核心概念 从输出误差倒推每个中间量的责任：一个节点影响损失的每条路径都要计入，路径上的导数相乘，汇合处相加。

整理成数学语言 若 $L=L_1+L_2$，则 $\partial L/\partial w=\partial L_1/\partial w+\partial L_2/\partial w$。批量损失为平均值时，梯度也必须取平均。

完整算例 同一个参数 w 出现在 $L(w)=(w-1)^2+(2w)^2$ 中，导数为 $2(w-1)+8w=10w-2$。只沿第一条路径回传会漏掉 $8w$。

易错点 反向传播计算梯度；梯度下降使用梯度更新参数，两者是不同步骤。共享参数不能只保留最后一次出现的梯度。

自测与答案 上例最小点？解 $10w-2=0$ 得 $w=1/5$，二阶导数 $10>0$。

Lean 对照 `NeuralNetworks.gradient_accumulation`

完整证明两条可导损失分支相加时导数相加，作为共享梯度累积的标量核心。

3.5 反向传播的微积分

官方课程：backpropagation-calculus

把长表达式拆成三个短问题

考虑 $z=wa+b$、输出 $\hat{y}=\sigma(z)$、损失 $L=(\hat{y}-y)^2$。若直接对嵌套式求导，容易漏项。先分别问：损失对输出多敏感，输出对预激活多敏感，预激活对权重多敏感？答案是 $2(\hat{y}-y)$、$\sigma'(z)$、a。

把三个响应倍率按路径相乘，就得到权重梯度。输入激活 a 越大，同一权重变化越能改变预激活；激活函数若进入饱和区，$\sigma'(z)$ 很小，梯度会被削弱。公式因此也解释了训练现象，而不仅给出可执行运算。

先检查最简单的线性输出

让 σ 为恒等函数，取 $a=2,b=1,y=4,w=1$。预测为三，误差为负一，权重梯度为负四。梯度下降减去负数，使权重增加，预测向目标移动，符号与直觉一致。偏置梯度为负二，因为偏置改变一个单位只使预测改变一个单位。

对多层情形，上一层的一个单元通常影响下一层多个单元，因而误差信号需要求和；矩阵转置正好把下一层的影响汇集回上一层。权重梯度的外积则把“接收端敏感性”和“发送端激活”配成所有连接的贡献。

推导时先固定向量是行还是列，再检查每个矩阵形状。损失是否带二分之一、是否按样本平均，也会改变系数。用一个可手算的标量例子核对符号和尺度，往往比只盯着大矩阵更有效。

一步一步建立直观

1. 把复合函数拆成可求导的环节 记 $z=w_1x_1+w_2x_2+b$，输出为 $a=\max(0,z)$，损失为 $L=(a-y)^2$。对 w_1 求导时，依次经过 L、a、z 三个环节。右图同时给出中间量，方便逐项核对链式法则中的每个因子，而不是只看最后一个梯度数字。每一环的输入位置都来自同一次前向计算。

2. 用符号解释参数应该向哪边走 在激活区域，权重梯度为 $2(a-y)x_1$。如果输出过大且输入为正，梯度为正，梯度下降就减小这个权重；若输入为负，改变方向随之反转。这种符号推理是检查求导的有效方法。偏置对应的局部输入恒为一，所以少了输入因子。

3. 不可导点需要单独处理 把单元切到关闭侧，再靠近边界。严格负侧梯度为零，严格正侧使用前面的公式；边界两侧一般不一致。右图在零点明确提示折点，不给出通常导数；某些训练程序会约定传回零，但这不是证明 ReLU 在零点可导。形式化命题与手算推导的适用条件应按各自陈述逐一核对。

定义、命题与证明

定义：单个 ReLU 单元与平方损失 记 $z = w_1x_1 + w_2x_2 + b$、$a = \max(0, z)$、$L = (a - y)^2$。输入与目标固定时，损失是参数的复合函数。

命题：严格激活区域中的权重梯度 在 $z > 0$ 的邻域中，$\partial L/\partial w_1 = 2(a - y)x_1$。在 $z < 0$ 的邻域中，该梯度为零。

证明. 当 $z > 0$ 时，ReLU 在邻域内就是恒等函数，链式法则给出 $2(a - y) \cdot 1 \cdot x_1$。当 $z < 0$ 时，它在邻域内恒为零，局部损失因而不随该参数变化。

结论刻意避开 $z = 0$。ReLU 在零处没有经典导数，复合损失是否恰好可导还要另看具体参数与目标；程序采用的反向传播约定不能替代这项数学判断。

□

观课提示 分开计算损失、激活与权重之间的局部导数，最后检查符号和矩阵形状。

原课程视频 本节网页与媒体 参数实验

要点回看

核心概念 将损失拆成短小的可导计算，再用链式法则组合，避免对整个表达式盲目展开。

整理成数学语言 对 $z = wa + b, a_{\text{out}} = \sigma(z), L = (a_{\text{out}} - y)^2$，

$$\frac{\partial L}{\partial w} = 2(a_{\text{out}} - y)\sigma'(z)a.$$

多层网络写作 $\delta^{(\ell)} = ((W^{(\ell+1)})^T\delta^{(\ell+1)}) \odot \sigma'(z^{(\ell)})$，权重梯度为 $\delta^{(\ell)}(a^{(\ell-1)})^T$。

完整算例 线性输出时取 $a = 2, b = 1, y = 4, w = 1$，预测为 3，权重导数为 $2(3 - 4)2 = -4$；减去负梯度使权重上升。

易错点 若损失定义为 $\frac{1}{2}(\hat{y} - y)^2$，公式中的 2 消失；批量、均值、转置约定必须一致。

自测与答案 上例偏置导数是多少？$2(3 - 4) = -2$。

Lean 对照 `NeuralNetworks.squared_loss_weight_gradient`

完整证明线性输出的平方损失对权重的导数。非线性层及矩阵反传公式在正文推导，未由此单命题覆盖。

3.6 语言模型的简明图景

官方课程：`mini-llm`

把生成一句话拆成一连串预测

模型不必一次选出完整句子。给定当前上下文，它先为下一个 token 分配概率，选择一个 token 后放回上下文，再预测下一步。token 是编码单位，可能是词、字或词的一部分，不能总按完整单词理解。

一条序列的概率是沿这条路径的条件概率乘积。第二步的分布取决于第一步已经选了什么，所以不能把每步无条件概率相乘，更不能把概率相加。取负对数后，乘积转成和，这让每个位置对训练目标的贡献可以累积。

局部最可能不等于整句最可能

假设第一步选甲的概率为 0.6、选乙为 0.4；在甲之后最可能的下一步概率是 0.5，在乙之后则为 0.9。贪心先选甲，完整两步最多得 0.3；经过乙的路径却可达 0.36。因此每一步取最大值，不保证整段序列概率最大。

温度改变分数经 Softmax 后的尖锐程度。低温让高分候选占比更大，高温使分布更平缓；这会改变采样行为，却不凭空增加事实知识。概率较高描述模型的统计偏好，不等于内容一定真实。

接下来要理解的核心，是上下文如何变成这些概率。Transformer 先把 token 写成向量，再让它们交换信息，最后产生词表分数。先弄清输入、条件概率和采样的关系，后续矩阵细节才有明确任务。

一步一步建立直观

1. 先把上下文压缩成一次加权读取 完整语言模型要根据此前的词元预测下一个词元。右图只抽出其中的注意力读取：两个候选信息各带一个分数，归一化后决定各自参与输出的份额。分数相同就平均读取。这些份额不是整个词表的下一词概率，二者在模型中的位置不同。

2. 相关性改变了信息的去向 增加第二项相对分数，输出便更靠近它的值。可以把这一步理解为读者根据当前问题重新分配注意，而不是把所有上下文不加区分地平均。实际模型会从向量计算查询和键；本图直接调分数，用来隔离“分数如何影响权重”这一机制。

3. 集中读取不等于理解正确 降低温度后，权重更加集中，但若较高分数本来分配给错误线索，集中只会强化错误。语言模型还需要词元表示、多层变换、训练目标及最终概率输出。这个窗口帮助理解一个局部部件，不把输出偏向某个值称为已经完成语言理解。

> **观课提示** 追踪一次预测、采样和追加上下文，区分单步概率与整段序列概率。
>
> 原课程视频 本节网页与媒体 参数实验

要点回看

核心概念 语言模型为下一个 token 分配条件概率，采样一个结果后把它放回上下文，继续预测。token 可以是词的一部分，不能总当成完整单词。

整理成数学语言 概率链式分解为

$$P(x_1, \ldots, x_T) = \prod_{t=1}^{T} P(x_t \mid x_1, \ldots, x_{t-1}).$$

训练中常最小化正确 token 的负对数概率总和；推理中的采样温度改变概率分布的尖锐程度。

完整算例 两步条件概率分别为 $0.4, 0.25$ 时，这条具体序列的联合概率为 0.1，不是 0.65。

易错点 贪心地每次选最大概率 token，不一定得到全局概率最大的完整序列。

自测与答案 概率 $0.5, 0.2, 0.1$ 的三步序列概率？0.01，负对数概率将乘积变成和。

Lean 对照 `NeuralNetworks.conditional_product_bounds`

验证两个合法概率的乘积仍在 $[0, 1]$；完整条件概率的测度论链式法则未形式化。

3.7 Transformer 的整体结构

官方课程：gpt

沿着数据走一遍整体结构

一段文字先分成 token，每个 token 查表得到嵌入向量，再加入或结合位置信息。只有内容、没有顺序，就无法区分同一组词的不同排列。随后经过多层注意力和前馈计算，向量逐渐携带上下文信息，最后投影成词表上每个候选的分数。

分数不是概率：它可以为负，也不必相加为一。Softmax 先指数化使其为正，再用总和归一化。相对分数差决定概率偏好，同加一个常数只引入共同因子，除法会将其抵消。

从一个两分类例子看出数值实现的意义

分数 $(0, \ln 3)$ 指数化后是 $(1, 3)$，归一化得到 $(1/4, 3/4)$。两分数都减去 $\ln 3$ 后，虽然数值改变，概率不变。实现中常减去最大分数，避免指数过大；这个技巧基于精确恒等式，浮点误差仍需单独考虑。

嵌入空间的方向可能与某些语义变化相关，但不要把每个坐标轴都当成固定概念标签。向量表示允许特征分布在许多坐标中，网络也可能在不同层使用不同表示。

本节只建立全景路线：嵌入负责表示，注意力交换位置间信息，前馈层处理各位置的特征，输出投影把表示转为预测分数。真实架构还涉及归一化、残差等细节；理解它们时仍应沿数据形状和作用追踪，而不是只记一张盒子图。

一步一步建立直观

1. 在结构图中定位这一小块计算 Transformer 把词元变成向量，反复进行上下文信息交换与逐位置变换，再输出词表分数。右图展示信息交换中的一个头、一次查询和两个候选值。先用相同分数得到平均值，确认加权读取的基本算术，再放回完整网络结构中理解其作用。

2. 把归一化前后分清楚 注意力先得到实数分数，softmax 再把它们转成正权重并归一化。给所有分数加同一个常数不会改变结果，因为共同的指数因子会约去。本图只保留分数差，正是利用这一不变性。实际模型的缩放与掩码应在归一化前处理。

3. 明确预测时可见的信息范围 降低温度可以观察更集中的权重，但自回归预测还要用因果掩码禁止读取未来位置。权重是否集中与是否泄露未来是两个不同问题。此处只有两个可见候选，没有模拟掩码、多头、残差与完整词表；正文的架构说明负责连接这些部件。

> **观课提示** 沿着数据形状走过嵌入、注意力、前馈和输出分数，检查 Softmax 的归一化。
>
> 原课程视频 本节网页与媒体 参数实验

要点回看

核心概念 token 先嵌入为向量，再通过注意力与前馈层反复更新，最后映射为词表上的分数。位置信息使顺序能被模型感知。

整理成数学语言 logits z_i 转成概率：$p_i = \mathrm{e}^{z_i} / \sum_j \mathrm{e}^{z_j}$。对全部分数加同一常数 c 不改变概率，因为共同因子 e^{c} 会抵消。

完整算例 两个 logits 为 $(0, \ln 3)$，得到概率 $(1/4, 3/4)$。把两者都减去 $\ln 3$ 后概率相同，这也是数值稳定 Softmax 的依据。

易错点 embedding 的某个方向可在特定实验中呈现语义关联，但不能把每一维机械地理解成一个固定概念。

自测与答案 $z = (2, 2)$ 的概率？$(1/2, 1/2)$；这与 logits 的绝对大小无关。

Lean 对照 `NeuralNetworks.softmax_shift`

完整证明双分类 Softmax 的平移不变性。定理使用精确实指数；浮点稳定性仍属于数值实现问题。

3.8 注意力：按上下文加权信息

官方课程：`attention`

同一个词为什么需要不同上下文表示

想象“银行”出现在取款句子与河岸翻译任务中。初始 token 表示本身不足以决定当前语境，需要从周围位置读取信息。注意力为每个位置建立一份按上下文改变的加权汇总，而不是给所有位置固定同一平均值。

查询负责提出匹配需求，键负责供匹配，值负责携带真正传递的内容。查询与键点积得到分数，再通过归一化产生权重。键与值分开，使“怎样找到信息”和“取回什么信息”能使用不同变换。

用两个值理解汇总，再放回矩阵

只有两个标量值二与十时，权重 $(0.75, 0.25)$ 给出四；把第二权重增大，输出向十移动。非负且和为一的权重保证汇总位于两值之间。交互实验会同时展示分数、温度、权重和输出，便于区分这四个对象。

真实单头中，每个查询都对多个键计算一行分数，所以出现 QK^T。除以 $\sqrt{d_k}$ 是缩放分数的设计；因果掩码让当前位置不能偷看未来 token。对每行独立归一化，再乘 V，便批量完成所有查询的汇总。

凸包解释只约束这一步的值平均，后续输出投影、多个头和残差连接都会改变最终位置。查询“寻找某类词”的比喻有助于理解可能行为，但某个训练模型究竟学了什么，仍须查看参数和实验，不能从命名直接断言。

一步一步建立直观

1. 从一个问题出发选择要读取的信息 把查询看成当前要找的特征，把键看成可供匹配的索引，把值看成真正需要传递的内容。右图直接调两个匹配分数的差，省去向量点积，只观察匹配之后的读取阶段。分数相等时，两个值具有相同权重，输出落在它们的中点。

2. 权重移动，输出也随之移动 提升第二项分数后，它的指数权重增大，输出向第二个值移动。由于权重非负且总和为一，单次标量加权结果始终落在两个值之间。读数与图形共同显示这个

凸组合约束；经过后续线性层或残差相加后，整体输出则不再受同一界限约束。

3. 温度改变选择的锐利程度 固定分数差而降低温度，较大分数占据更大份额；提高温度则趋向平均。softmax 仍然是连续加权，不是简单地取最大值。非常小的温度会使数值敏感，程序还应使用稳定计算方法。这里的两个标量值只是便于检查机制的缩小模型。

定义、命题与证明

定义：归一化指数权重 给定实数分数 s_j 与正温度 T，令 $a_j = e^{s_j/T} / \sum_k e^{s_k/T}$。各权重为正，且总和为一。

命题：一次注意力读取是凸组合 若标量值都在区间 $[m, M]$ 中，则加权输出 $\sum_j a_j v_j$ 也在 $[m, M]$ 中。

证明. 逐项有 $a_j m \le a_j v_j \le a_j M$，因为各权重非负。求和后得到 $m \sum_j a_j \le \sum_j a_j v_j \le M \sum_j a_j$。

使用权重和为一便得到结论。本命题只约束这一次加权汇总，未约束接下来的线性输出变换、残差相加或整个网络的数值范围。

□

观课提示 分别追踪查询、键和值，观察分数如何变成每行和为一的权重。

原课程视频 本节网页与媒体 参数实验

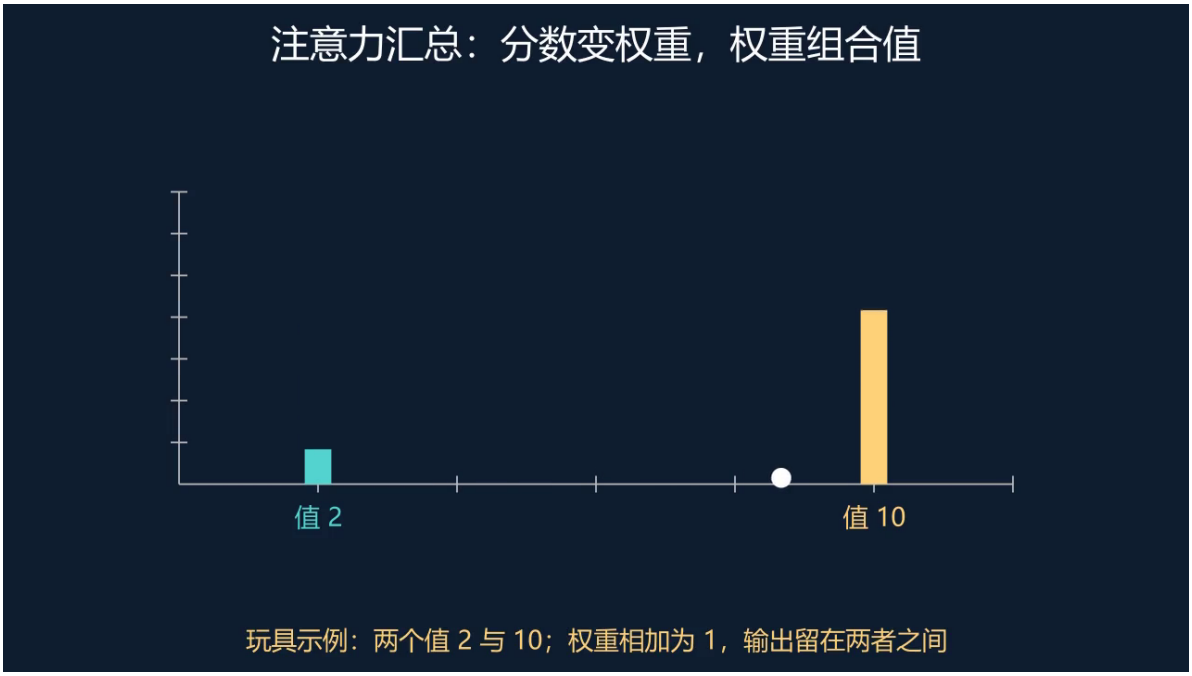


图 3.2: 两个值的加权汇总。柱高表示权重，白点表示输出；这是标量玩具模型。本项目原创 Manim 动画关键帧；完整动画与操作入口见本节网页。

要点回看

核心概念 查询向量表达“正在寻找什么”，键向量用于匹配，值向量提供被汇总的信息。每个位置根据自身查询选择不同的权重。

整理成数学语言 以 token 为行：$Q = XW_Q, K = XW_K, V = XW_V$，

$$\mathrm{Attention}(Q, K, V) = \mathrm{softmax}_{\mathrm{row}}\left(\frac{QK^T}{\sqrt{d_k}} + M\right)V.$$

因果掩码 M 禁止关注未来位置。每一行的有限权重非负且和为 1。

完整算例 两个标量值 $u = 2, v = 10$，权重 $(0.75, 0.25)$ 得输出 4，位于 $[2, 10]$ 内。

易错点 加权平均的凸包约束只适用于单头的值汇总；输出投影、多个头及残差连接之后，最终向量不必仍在原值凸包内。

自测与答案 权重 $(0.2, 0.8)$ 时结果？8.4。查询与键的维数相同，值维数可以不同。

Lean 对照 `NeuralNetworks.attention_convex`

完整证明两标量值的凸组合处在二者之间，保留权重 $0 \le w \le 1$ 的条件。

3.9 前馈层如何表达特征与关联

官方课程：`mlp`

先检测，再写回

前馈层可以看成两阶段：第一组权重对输入做许多测量，非线性决定这些测量怎样激活，第二组权重把激活写回表示空间。一个简化单元先计算 $k^T x$，再沿方向 v 输出相应强度，贡献为 $v(k^T x)$。

取 $k = (1, 2), v = (3, -1), x = (2, 1)$，检测值为四，写回向量是 $(12, -4)$。先看检测方向，再看写回方向，可以避免把同一神经元的输入权重和输出权重误认成同一种角色。

外积显示单个通道能做多大的事

把表达式重新加括号，是 $(vk^T)x$。外积矩阵的每列都与 v 平行，所以秩至多一：单个线性通道只能沿一条输出方向写信息。多个通道相加能形成更丰富的作用，而非线性又使不同输入激活不同组合。

在 Transformer 中，前馈层通常逐位置处理表示，位置之间的通信主要由注意力承担；不过每个位置的表示此前已经包含上下文，所以“逐位置”不等于“完全不知上下文”。

把通道解释成某个特征或关联的检测器，是有用的构造性图景，却不是对真实模型的唯一解读。特征可能分散在多个单元，也可能在一个单元中叠加。练习时先精确核对外积的计算，再把“网络可以实现什么”与“训练后实际实现什么”分开。

一步一步建立直观

1. 把单元看成一条条件化的贡献 前馈层先把输入投到许多方向，经过非线性，再把各单元贡献合成输出。右图保留一个方向和一个 ReLU：激活强度说明这条方向上的证据有多强。实际向量输出还会乘以一个写出方向，所以一个单元可以理解为“读取一个方向，再写入另一个方向”。

2. 非线性让贡献具有选择性 当输入位于关闭一侧，当前单元不贡献输出；跨过边界后，贡献随距离线性增长。许多这样的边界可以组合出更复杂的分区。若去掉所有非线性，即便叠加很多仿射层，整体仍只是一个仿射变换，不会自动得到新的表达层次。

3. 关联解释需要结合整体网络 改变权重和偏置会移动读取方向与门槛，但某个神经元的激活不必对应单一、稳定、易命名的事实。多个特征可能共用方向，一个特征也可能分布在许多单元中。本图用于拆解计算结构；对真实模型的“知识存储位置”仍须做实证检验。

> **观课提示** 区分检测方向与写回方向，用一个通道的外积理解两次线性作用。
>
> 原课程视频 本节网页与媒体 参数实验

要点回看

核心概念 前馈层可以先检测输入中的某个方向，再沿另一个方向写回信息。这样能帮助理解关联的存储，但真实网络常使用分布式和叠加表示。

整理成数学语言 $\mathrm{MLP}(x) = W_2\sigma(W_1x + b_1) + b_2$。忽略偏置并在激活处于线性区时，单神经元贡献为 $v(k^Tx)$，等于秩至多为一的矩阵 vk^T 作用于 x。

完整算例 $k = (1, 2), v = (3, -1), x = (2, 1)$，检测值为 4，写回向量为 $(12, -4)$。

易错点 这种构造说明网络“可以如何”表达一个关联，并不证明某个真实模型恰好如此，也不意味着一个神经元只存一个事实。

自测与答案 写出上例的 vk^T。答案 $\left[\begin{smallmatrix} 3 & 6 \\ -1 & -2 \end{smallmatrix}\right]$，行列式为零。

Lean 对照 `NeuralNetworks.mlp_rank_one`

验证二维检测后写回等于相应外积矩阵作用的坐标公式；不是对模型知识存储机制的经验证明。

延伸阅读：与压缩课程共享的交叉熵

这一课在官网同时出现在神经网络与“压缩与智能”序列。完整笔记、例子、KL 分解及 Lean 对照统一放在第 6.2 节，以免出现两份相互漂移的版本。

3.10 扩散模型：从噪声中逐步恢复

官方课程：`diffusion-models`

先构造一个容易的破坏过程

直接生成复杂图像很难，但给图像加噪很容易。扩散方法规定一系列噪声水平，把数据逐渐混入随机噪声；训练时知道原图、噪声和阶段，就能构造监督任务，让模型估计噪声或干净信号。反向生成尝试从噪声出发，逐步执行学习到的更新。

正向公式中的两个系数分别控制信号和噪声强度。它描述的是已知阶段的边际采样，不等于完整算法的所有步骤。模型还需要时间或噪声水平信息，因为同一个观测在不同噪声强度下应有不同解释。

代数可逆与统计可恢复不同

在标量玩具模型 $x_t = ax_0 + b\varepsilon$ 中，若知道真噪声且 $a \neq 0$，减掉噪声再除以 a 就能恢复原信号。这里的代数完全确定，正是 Lean 校验的内容。但实际只看到 x_t，真噪声未知，多组信号噪声都可能产生同一观测。

若估计噪声误差为 δ，重建信号误差是 $-(b/a)\delta$。当信号系数很小，同样噪声误差会被显著放大。这解释了为什么不能把一次代数反演当成真实生成过程的完整答案。

文本条件影响模型更倾向恢复哪类结构，逐步采样则涉及随机性与具体更新规则。本节建立“容易加噪、学习条件恢复”的思想；训练有效性、生成质量和完整随机过程需要额外分析，不能由一条恢复恒等式推出。

一步一步建立直观

1. 先明确加噪是怎样生成观测的 右图用一个标量替代整幅图像，按照 $x_t = \sqrt{\alpha}x_0 + \sqrt{1-\alpha}\,\epsilon$ 混合信号与噪声。改变保留比例，就能看出同一个观测可能主要由信号或噪声主导。这一步是已知的前向过程，还没有涉及学习如何生成图像。

2. 去噪依赖估计，不能凭空恢复 若噪声估计准确，可由代数关系解出原信号；若估计偏差为 $\Delta\epsilon$，恢复误差会乘以与噪声强度有关的倍率。第二步故意用错误估计，让你比较真值与恢复值。真实模型并不知道当次噪声，需要依据训练获得的统计规律进行估计。

3. 为什么强噪声阶段更依赖统计结构 当保留比例很小，恢复公式里的信号系数也很小，噪声估计误差因此被放大。图中可精确比较误差，但不代表高维去噪能唯一找回原图。扩散生成通过多步反向演化利用分布信息；这里验证的是单步代数关系，不能替代完整采样器。

定义、命题与证明

定义：单步信号与噪声混合 取 $0 < \alpha < 1$，设 $x_t = \sqrt{\alpha}x_0 + \sqrt{1-\alpha}\epsilon$。给定噪声估计 $\widehat{\epsilon}$，按此等式解出重建值 $\widehat{x}_0$。

命题：噪声估计误差的传播 重建误差满足 $\widehat{x}_0 - x_0 = -\sqrt{(1-\alpha)/\alpha}\,(\widehat{\epsilon} - \epsilon)$。

证明. 将观测的定义代入 $\widehat{x}_0 = (x_t - \sqrt{1-\alpha}\widehat{\epsilon})/\sqrt{\alpha}$，原信号项直接留下 x_0，剩余两项合并即可。

分母合法来自 $\alpha > 0$。当 α 很小时，误差倍率很大；即使代数恒等式正确，也不能由它推出噪声估计足够准确或整套生成过程收敛。

□

> **观课提示** 区分已知噪声时的代数反演与实际模型估计噪声的统计任务。
>
> 原课程视频 本节网页与媒体 参数实验

要点回看

核心概念 正向过程逐渐给数据加噪，反向模型学习在噪声水平及条件信息已知时估计噪声或干净数据。文本条件可通过联合表示和注意力影响生成。

整理成数学语言 常见正向边际为

$$x_t = \sqrt{\bar{\alpha}_t}\,x_0 + \sqrt{1-\bar{\alpha}_t}\,\varepsilon, \qquad \varepsilon \sim \mathcal{N}(0, I).$$

当 $\bar{\alpha}_t > 0$，若噪声估计为 $\hat{\varepsilon}$，可构造 $\hat{x}_0 = (x_t - \sqrt{1-\bar{\alpha}_t}\hat{\varepsilon})/\sqrt{\bar{\alpha}_t}$。

完整算例 用简化标量系数 $a = 0.8, b = 0.6, x_0 = 2, \varepsilon = -1$，得到 $x_t = 1$；已知真噪声时 $(1 - 0.6(-1))/0.8 = 2$。

易错点 上述代数恢复要求知道噪声；模型实际只是估计它。单个带噪样本通常不能唯一确定原始图像。

自测与答案 若噪声估计多了 δ，恢复误差是什么？$-(b/a)\delta$，说明 a 很小时误差会被放大。

Lean 对照 `NeuralNetworks.denoising_reconstruction`

完整验证非零信号系数下的代数恢复式；未形式化随机扩散过程、训练有效性或生成质量。客座原课署名 Stephen Welch。

章节综合：完整走一次学习循环

用最小网络 $\hat{y} = wa + b$ 拟合一个样本 $a = 2, y = 4$。初始 $w = 1, b = 1$，前向得到预测三，平方损失为一。反向计算得到权重梯度负四、偏置梯度负二。使用学习率 0.1 同时更新，得到 $w = 1.4, b = 1.2$，新预测恰好为四。

这个幸运结果不意味着训练通常一步完成。它与本例只有一条线性约束、损失为二次型及特定步长有关。实际上，任意满足 $2w + b = 4$ 的参数都能让这个样本损失为零，因此一个样本无法唯一确定两个参数。这与线性代数中的多解和不可见方向完全相通。

如果再加入不同输入的样本，参数必须同时兼顾多个约束。若数据存在噪声，可能没有一组参数把所有误差都消成零，优化转为寻找合适折中。即使训练误差很小，也还需要独立样本评估。

复习更大模型时，同样按输入表示、前向计算、损失、反向敏感性、更新、外部评估这六步追踪。注意力或扩散改变了内部计算结构，却没有消除这些需要分别说明的问题。

第四章　微分方程：从局部规律到整体演化

微分方程给出状态的变化规则，初始条件选择具体轨迹，边界条件限制空间上的可行行为。一个显式解公式不是理解系统的唯一方式。

4.1　微分方程与向量场

官方课程：`differential-equations`

先知道变化规则，也能研究运动

很多问题先给出的不是位置公式，而是当前位置决定的速度。例如 $x' = kx$ 说“量越大，单位时间的变化越大”。微分方程的未知对象是一条函数轨迹；初值告诉我们从哪里出发，变化规则告诉我们之后往哪里走。

在每个状态处画一根速度箭头，解曲线就必须沿箭头前进。即使没有简单显式解，也可观察平衡点、增长方向或长期趋势。二阶摆方程同时依赖角度和角速度，把它们合成状态 (θ, ω)，就能在相平面中画一阶系统的箭头。

公式解、几何观察与数值近似各做什么

猜测 $x = x_0 \mathrm{e}^{kt}$ 后，对它求导并代回，能验证它满足方程与初值。但“找到了一个解”与“证明只有这一个解”仍是两件事，唯一性需要相应定理条件。

欧拉法在当前点沿速度走一个有限小步，再重新估计速度。它把局部直线近似串起来，并不精确遵循曲线；步长过大可能改变定性行为。减小步长并比较结果，是检查数值可信度的基本方式。

读本章时坚持三个层次：规则是什么，哪些轨迹符合规则，从初值能否确定唯一轨迹。方向场提供直觉，代回提供局部验证，分析定理给存在与唯一性保证。后续热方程会把有限维状态变成整条温度分布，但这种提问方式不变。

一步一步建立直观

1. 规则先告诉你下一瞬间往哪里走　微分方程给出当前位置对应的速度，而不是直接给出整条路线。右图的小箭头标示当地速度方向；选定一个初始点后，轨线必须处处沿着这些方向前进。

先选择纯旋转场，比较沿轨线的切向方向与经过位置上的速度箭头是否一致。

2. 同一旋转可以叠加径向收缩　加入负的径向增长率，轨线向内盘旋。角度继续变化，半径却按指数缩小，因此“在转动”和“向平衡点靠近”并不冲突。这里使用有显式解的线性系统，便于把局部方程、整体轨迹和实时坐标三者直接核对。

3. 初值与稳定性是不同问题　把径向增长率改为正，同样的旋转伴随向外增长。拖动初始点可得到同一方程的另一条解；修改增长率则已经改变方程本身。靠近原点的初始状态能否始终保持接近，是稳定性问题，不能仅凭某个初值恰好静止来判断。

定义、命题与证明

定义：线性平面速度场　考虑 $x' = \lambda x - \omega y$、$y' = \omega x + \lambda y$，其中 λ, ω 为常数。前者控制径向伸缩，后者控制角向转动。

命题：长度平方的增长只由径向项控制　沿任意可微解，$\dfrac{d}{dt}(x^2 + y^2) = 2\lambda(x^2 + y^2)$。特别地，$\lambda = 0$ 时长度不变。

证明. 由乘积法则，导数为 $2xx' + 2yy'$。代入速度场，两个交叉项 $-2\omega xy$ 与 $2\omega xy$ 抵消，剩下 $2\lambda(x^2 + y^2)$。

因此转动不会单独改变长度，正负径向项则分别使长度平方增长或衰减。这是本线性系统的精确计算，并非任意速度场都可如此分离。

□

> **观课提示**　先从状态画速度箭头，再讨论轨迹；不要把方向场当成唯一性证明。
>
> 原课程视频　本节网页与媒体　参数实验

要点回看

核心概念　与其先寻找闭式解，不如在每个状态处画出速度。解曲线必须始终沿着这些箭头前进；相空间把位置和速度合并为状态。

整理成数学语言　一阶系统写成 $x' = F(t, x)$。二阶摆方程 $\theta'' = -(g/L)\sin\theta$ 可改写为 $\theta' = \omega$，$\omega' = -(g/L)\sin\theta$。

完整算例　最简单的 $x' = kx$ 有解 $x(t) = x_0 \mathrm{e}^{kt}$。若 $k < 0$，它趋近零；若 $k > 0$，非零初值的绝对值增大。

易错点　画出方向场不自动保证唯一解；存在唯一性定理需要连续性、局部 Lipschitz 等相应条件。欧拉法也只是近似。

自测与答案　$x' = 2x, x(0) = 3$ 的解？$3\mathrm{e}^{2t}$，代回可直接检查。

Lean 对照　`DifferentialEquations.exponential_solution`

完整验证 e^{kt} 的导数等于 $k\mathrm{e}^{kt}$；不是对所有微分方程的存在唯一性证明。

4.2　偏微分方程与热扩散

官方课程：`pdes`

状态不再是几个数，而是一整条曲线

一根杆在各位置有不同温度，所以状态是温度分布 $u(x,t)$。时间改变时，整条曲线都变。偏导数让我们分别固定位置观察时间变化，或固定时间观察空间变化，两者承担不同物理角色。

为什么热方程出现空间二阶导数，而不是温度本身？一个位置比两侧邻居都热，就会向两侧散热；比邻居冷则吸热。决定变化的是中心相对于邻域平均的偏离。三点模型给出 $u(x-h) - 2u(x) + u(x+h)$，除以 h^2 后趋向曲率。

空间斜率与温度变化之间隔着一层差分

在常系数模型中，热通量与负温度梯度成正比；某点温度变化来自流入与流出的差，所以对通量再取一次空间变化，便得到二阶导数。线性温度分布虽然有热流，但各小段流入流出平衡，内部温度可以暂时不变。

端点怎样处理会改变整个问题。保持端点温度固定，与隔热端点无热流，是不同边界条件；同一个初始曲线在不同边界下会演化不同。只给方程而不给初始和边界信息，通常不足以选出具体解。

若做离散模拟，还要检查时间步稳定性。把空间分得更细却不相应调整时间步，可能让数值解振荡甚至爆炸，那不一定是物理现象。接下来我们寻找一些形状简单、能精确追踪的模式，用来解释也用来校验模拟。

一步一步建立直观

1. 每个位置都带着一个待演化的量　常微分方程追踪有限个状态，热方程则追踪每个空间位置的温度。右图的横轴是位置，纵轴可看作相对于平均温度的偏差，负值表示低于平均。时间推进时，整条温度曲线改变，空间形状本身成为系统状态。

2. 温度的弯曲决定当地升温还是降温　一个温度尖峰比两侧更热，热流向外，峰顶就会下降；谷底则因邻域更热而上升。二阶空间导数把这种邻域不均衡写成局部量。本例的方程是 $u_t = u_{xx}$，且区间首尾相接；边界条件不同，适用的基本模式也会改变。

3. 更细的起伏更快被抹平　将高频模式的频率加大，而保持观察时间相同。高频部分的曲率更大，所以衰减更快，速度含频率的平方。图形显示的是两个正弦模式的精确演化，不是任意材料与边界的完整仿真；物理单位和扩散系数在这里已经归一化。

> **观课提示**　观察中心相对邻居的高低，解释为什么温度变化由空间二阶差分决定。
>
> 原课程视频　本节网页与媒体　参数实验

要点回看

核心概念　温度同时依赖位置和时间。热从较热处流向较冷处，局部的“比邻居更高还是更低”决定温度怎样变化。

整理成数学语言　一维均匀介质模型为 $u_t = \kappa u_{xx}$，$\kappa > 0$。离散空间中，

$$u_{xx}(x) \approx \frac{u(x-h) - 2u(x) + u(x+h)}{h^2}.$$

这个二阶差分是局部平均恢复趋势的代数版本。

完整算例　邻点温度为 $4, 10, 6$，中心二阶差分分子为 -10，所以中心倾向降温。若为 $6, 2, 8$，分子为 10，中心倾向升温。

易错点　需要给初始温度分布及边界条件；固定端温度与绝热边界是不同问题。显式数值时间步也受稳定性限制。

自测与答案　为什么两个解相加仍满足齐次热方程？时间导数与空间二阶导数都线性。

Lean 对照　`DifferentialEquations.heat_stencil`

验证三点二阶差分算子对相加的线性性；连续 PDE 的分析及离散极限未由此命题覆盖。

4.3　用基本模式求解热方程

官方课程：`heat-equation`

寻找扩散时不改变形状的波

任意温度曲线难追踪，但正弦有一个特殊性质：对空间求两次导数，仍是自身乘一个负常数。若初始形状是 $\sin(kx)$，热方程只要求振幅随时间满足 $A' = -\kappa k^2 A$，所以振幅指数衰减，形状保持。

这与线性代数的特征方向相似。求二阶导是一个线性算子，正弦是它的特征模式；热演化在每个模式上只改变系数。对固定零温边界，频率还必须使正弦在两个端点都为零，因此允许的 k 不是任意数。

高频为什么先消失

频率增为三倍，二阶导倍率增为九倍，因此高频振幅衰减得更快。温度曲线上细小急促的起伏先被抹平，宽缓结构保留更久；“热让分布平滑”由此有了可计算的原因。

动画采用周期边界的一低一高两个模式：蓝色是低频，黄色是高频，白色是总和。先在暂停时预测哪部分会先消失，再推进时间，观察总曲线逐渐接近低频形状。这里的边界选择与正文零温杆示例不同，但模式衰减机制相同。

有限个模式相加可以直接利用线性性；无限级数则需要额外条件，才能交换求和、求导并恢复初值。把每个模式代回方程是必要检查，却还没有自动完成一般解理论。这一点将在傅里叶章节继续讨论。

一步一步建立直观

1. 选择在求导后仍保持形状的模式 正弦函数二次求导后仍是同一个正弦形状，只多了负的频率平方。于是试设温度为一个正弦模式乘以随时间变化的振幅，偏微分方程就变成振幅的常微分方程。右图蓝线与橙线分别承担一个模式，紫线是二者相加。

2. 把空间弯曲转成振幅衰减 频率为 n 的模式满足振幅变化率等于 $-n^2$ 倍当前振幅，所以衰减因子为 e^{-n^2t}。观察同一时刻的两个振幅读数，就能把“尖细起伏消失得快”的图像翻译成定量关系。这里扩散系数取一，其他系数会相应改变时间尺度。

3. 叠加有效，是因为方程线性 加大高频频率后，紫线更快接近低频模式。能够逐个演化再相加，依赖热方程的线性。若把有限个模式进一步推广成无限傅里叶级数，还需说明级数与求导、取极限之间的交换条件；两个模式的图示不能自动完成这些分析论证。

定义、命题与证明

定义：周期热方程的一个模式 在首尾相接的区间 $[0, 2\pi]$ 上，取整数频率 n，考虑 $u(x,t) = Ae^{-n^2t}\sin(nx)$。扩散系数归一化为一。

命题：正弦模式满足热方程 上述模式满足 $u_t = u_{xx}$，且初值为 $u(x,0) = A\sin(nx)$。有限个此类模式相加仍满足该方程。

证明. 对时间求导得到 $-n^2Ae^{-n^2t}\sin(nx)$；对空间求两次导数得到同一个表达式。整数频率保证周期边界匹配。

对有限和可以逐项求导，故线性组合仍满足方程。推广到无限级数时，逐项求导需要额外收敛条件；有限叠加的验证不自动证明任意初值问题的解存在。

□

观课提示 比较低频和高频模式，预测哪一项先衰减，再核对边界条件。

原课程视频　本节网页与媒体　参数实验

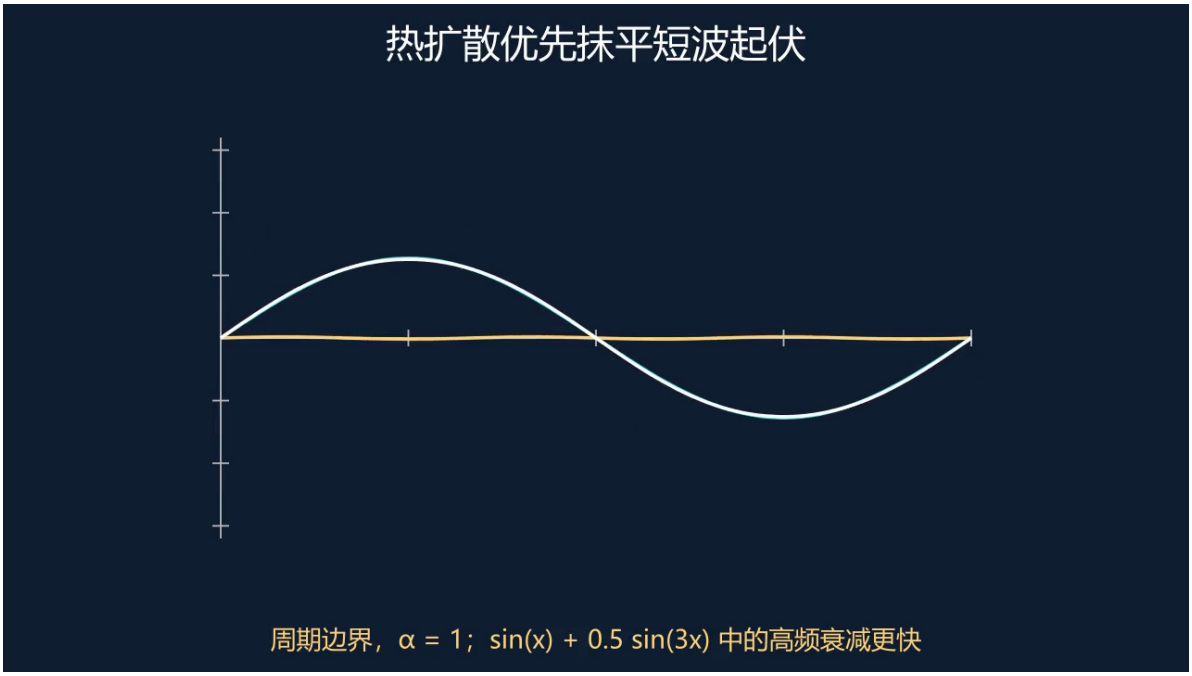


图 4.1: 周期边界下，低频（蓝）与高频（黄）以不同速度衰减，白线为总和。本项目原创 Manim 动画关键帧；完整动画与操作入口见本节网页。

要点回看

核心概念　找到形状在扩散中保持、只改变振幅的波形，再将一般初始数据组合成这些模式。高频因弯曲更强而衰减更快。

整理成数学语言　对 $0 < x < L$ 的零温边界，模式为

$$u_n(x,t) = b_n \sin\frac{n\pi x}{L}\,\exp\left[-\kappa\left(\frac{n\pi}{L}\right)^2 t\right].$$

因正弦的二阶导数为频率平方的负倍数，代入后两边相等。

完整算例　取 $L=\pi, \kappa=1$，$u = \sin x\mathrm{e}^{-t} + 2\sin(3x)\mathrm{e}^{-9t}$。第三模式的衰减指数是第一模式的九倍。

易错点　无限求和与求导交换需要收敛或正则性条件；边界不是零温时需换模式或先减去稳态。

自测与答案　初值仅为 $\sin(2x)$，同样参数下解是什么？$\sin(2x)\mathrm{e}^{-4t}$。

Lean 对照 `DifferentialEquations.heat_mode_decay`

完整验证模式振幅的时间导数。空间正弦导数可由微积分模块组合；无限级数解与边界条件未形式化。

4.4 傅里叶级数：旋转与叠加

官方课程：`fourier-series`

把复杂周期形状拆成简单运动

一条复杂周期曲线，可以尝试写成许多不同频率的正弦、余弦之和。每个频率只承担一种规则振动，幅度与相位说明它贡献多少、从哪里开始。复指数把一对正弦余弦合成平面旋转，使正负频率能用统一记号表达。

关键问题是怎样测出某个频率的系数。把信号乘上反向旋转的目标频率：与目标同频的部分变成常量，其他整倍频仍在一周期内转完若干圈，平均为零。因此积分能够挑出匹配部分，这就是正交性的作用。

系数公式之后还有收敛问题

对 $3+2\cos t$，常量三不旋转，余弦可拆成两个幅度为一、方向相反的旋转。这是可逐项验证的有限例子。先把它算明白，再处理一般系数公式，会更容易看清归一化因子和正负频率。

在热实验中改变高频振幅，可以看到同一个低频骨架怎样叠加出细节；随后推进时间，高频先衰减。傅里叶分解因而不仅是换一种画法，还把原本耦合的演化拆成独立系数的变化。

并非写出形式级数就证明它处处等于原函数。对跳跃点，常见分段光滑条件下级数趋向左右极限平均，还会出现局部振铃。讨论近似时必须说明点态、均方或其他意义。这里先掌握频率测量机制，把完备性和收敛留作明确的分析问题。

一步一步建立直观

1. 把复杂曲线拆成可单独追踪的成分 先在初始时刻观察两条正弦曲线及其和。局部看起来复杂的起伏，可能只是少数简单模式叠加。傅里叶展开把这种想法推广到一组频率，并用投影计算系数。右图只取两个分量，目的是让每个成分的贡献可以被直接辨认。

2. 同一种分解可以服务不同问题 推进时间后，每个系数按热方程的规则衰减，基函数本身没有更换。若研究声音，则常把系数看成不同频率成分；若研究旋转向量，则用复指数表示同样的频率结构。本窗口展示热扩散用途，并不把时间衰减当作所有傅里叶问题的固有属性。

3. 增加频率并不等于增加任意细节 改变第二个模式的频率，比较它与总曲线的关系。一个固定频率、振幅和相位只能添加特定的周期形状；重建一般目标函数需要很多系数，并受收敛方式

影响。跳跃附近的行为尤其要单独分析，不能把有限平滑曲线的视觉效果直接推广到所有函数。

观课提示　把信号与反向旋转相乘，观察匹配频率如何留下非零平均。

原课程视频　本节网页与媒体　参数实验

要点回看

核心概念　正弦、余弦可以合并成复平面上的旋转。不同频率的旋转向量相加，能表示复杂的周期变化。

整理成数学语言　对周期 T，令 $\omega_0 = 2\pi/T$，

$$f(t) \sim \sum_{n\in\mathbb{Z}} c_n \mathrm{e}^{\mathrm{i}n\omega_0 t}, \qquad c_n = \frac{1}{T}\int_0^T f(t)\mathrm{e}^{-\mathrm{i}n\omega_0 t}\,\mathrm{d}t.$$

正交性使积分能够挑出对应频率。

完整算例　$f(t) = 3 + 2\cos t$ 的非零复系数为 $c_0 = 3, c_1 = c_{-1} = 1$。常数项不转，其余两个方向相反地旋转。

易错点　对不连续周期函数，点态收敛需条件；常见分段光滑情形在跳跃点趋向左右极限的平均值，且会出现 Gibbs 振铃。

自测与答案　$\sin t$ 的复系数？$c_1 = 1/(2\mathrm{i}), c_{-1} = -1/(2\mathrm{i})$。

Lean 对照　`DifferentialEquations.fourier_superposition`

验证单频复指数模式的线性叠加恒等式；不将它冒充傅里叶完备性或收敛定理。

4.5　复指数与旋转的物理图像

官方课程：`complex-exponents`

把两个实分量装进一个复数

平面位置 (x, y) 可以写成 $z = x + \mathrm{i}y$。乘以 i 后得到 $-y + \mathrm{i}x$，恰是把位置旋转九十度。若速度满足 $z' = \mathrm{i}\omega z$，它始终与位置垂直，所以位置沿圆周运动，角速度由 ω 控制。

再加上实系数 σ，速度多出一份沿径向的 σz：正值向外、负值向内。因此 $z' = (\sigma + \mathrm{i}\omega)z$ 同时包含缩放与旋转，解呈现向外或向内的螺旋。

实部与虚部分别控制什么

解中的 $\mathrm{e}^{\sigma t}$ 决定模长倍率，$\mathrm{e}^{\mathrm{i}\omega t}$ 决定转过的角度。对于 $-1+2\mathrm{i}$，模长按 e^{-t} 衰减，角度每单位时间增加二弧度。这两个独立作用相乘，比用两个三角方程追踪更紧凑。

观察复平面动画时，蓝箭头是位置，黄箭头是速度方向。先预测转过四分之一圈后两者指向哪里，再播放。纯虚系数不改变模长，有限数值步进却可能产生半径漂移，应区别精确轨迹与近似算法。

复数在这里是一种高效的二维表示，不是额外虚构的物理空间。实际测量可以取实部，另一分量负责保存相位关系。这种表示将继续进入傅里叶与拉普拉斯变换，用统一的指数核记录振动与衰减。

一步一步建立直观

1. 把乘以虚数看成速度规则 对复平面上的位置 z，乘以 i 会把它逆时针转过直角。如果规定速度为 $z' = iz$，速度就始终与半径垂直，且大小与半径相等。右图的蓝色短箭头表示这个方向，位置点因此沿圆运动，而不是沿固定直线前进。

2. 连续执行局部规则得到整体旋转 转到半圈时，单位初值到达负实轴。复指数把整段演化记成 e^{it}：指数中的参数是持续执行局部旋转规则的时间。它与实指数的增长共享同一种微分结构，只是实数倍率改变长度，虚数倍率改变方向。

3. 实部与虚部各负责一种变化 加入正的实部增长率后，速度既有径向分量，也有切向分量，轨迹变为螺线。公式 $e^{(\sigma+i)t} = e^{\sigma t}e^{it}$ 正好把这两种作用分开。窗口固定角速度为一；一般角速度应写入虚部，并与实际时间单位保持一致。

> **观课提示** 同时观察位置和速度方向，分别解释指数实部和虚部的作用。
>
> 原课程视频 本节网页与媒体 参数实验

要点回看

核心概念 乘以 i 表示旋转 90°。若速度始终是位置旋转后的向量，运动便沿圆周持续进行。

整理成数学语言 $z' = (\sigma + \mathrm{i}\omega)z$ 的解为 $z(t) = z_0\mathrm{e}^{\sigma t}\mathrm{e}^{\mathrm{i}\omega t}$。实部 σ 控制增长或衰减，虚部 ω 控制角速度。

完整算例 $z' = (-1+2\mathrm{i})z$ 产生向内螺旋：模长按 e^{-t} 衰减，角度每单位时间增加 2 弧度。

易错点 复数在这里把两个实维度的系统压缩为一个表达式，并不引入“虚假的”物理量；测量量常取实部。

自测与答案 $\mathrm{e}^{3\mathrm{i}t}$ 的模长和周期？模长 1，周期 $2\pi/3$。

Lean 对照 `DifferentialEquations.rotating_exponential_norm`

完整验证任意实角度的纯旋转复指数模长为一。

4.6 拉普拉斯变换：带衰减的频率探测

官方课程：laplace-transform

给频率测量增加一个衰减旋钮

傅里叶测量把信号乘上反向旋转，再累积。如果信号在无穷远持续增长，积分可能不收敛。拉普拉斯核加入指数衰减：写 $s=\sigma+\mathrm{i}\omega$，核为 $\mathrm{e}^{-\sigma t}\mathrm{e}^{-\mathrm{i}\omega t}$。一个旋钮控制遗忘远处的速度，另一个控制匹配频率。

对 $f(t)=\mathrm{e}^{at}$，乘核后是 $\mathrm{e}^{-(s-a)t}$。只有 $\operatorname{Re}s>a$，衰减才压过增长，积分得到 $1/(s-a)$。收敛条件来自积分的尾部，不是公式之外可有可无的小注释。

一张有理式不能代替全部变换信息

令 $a=2$，在实数 $s=3$ 处，原积分收敛且为一；在 $s=1$，有理式形式上给负一，但原积分发散。这个例子说明解析表达式延伸到某处有值，不意味着它仍代表定义积分。

单边变换从时间零开始，适合记录带初值的过程。选择单边或双边、选择何种收敛域，会影响反演与解释；因此每次使用变换表，都应同时读变换约定和适用区域。

这一节先把变换理解成一族带权测量。下一节真正有用的地方是：求导在这种测量下转为乘以 s，并留下起点信息。原本关于时间函数的微分方程，因此可能变成关于变换值的代数关系。

一步一步建立直观

1. 先辨认信号是在衰减还是增长 右图的二维轨线也可以看成复信号 $z(t)=e^{(\lambda+i\omega)t}z(0)$。选择衰减情形，幅度逐渐减小。拉普拉斯变换给信号再乘一个指数权重后积分，让我们用一个复参数同时探测衰减与振荡，而不仅是读取某个时刻的值。

2. 探测频率与信号频率相互比较 纯旋转信号幅度不衰减；若只无限累计它，普通意义的收敛就需要格外小心。乘上探测核 e^{-st} 后，s 的虚部比较旋转频率，实部提供衰减。右图展示被探测的信号，不直接画积分；应把信号轨迹与变换结果这两个对象区分开。

3. 收敛域是计算的一部分 改成增长信号。若其增长率为 λ，要使指数加权后的绝对积分收敛，需要 $\operatorname{Re}s>\lambda$。这一条件说明“给函数写一个变换式”不能脱离参数范围。图像可以提示增长趋势，最终的收敛判断仍由积分估计完成。

观课提示　比较旋转与衰减两个参数，留意公式的收敛域不能省略。

原课程视频　本节网页与媒体　参数实验

要点回看

核心概念　傅里叶核只旋转；拉普拉斯核同时旋转和指数加权。加入衰减可让原本无法绝对积分的函数变得可处理。

整理成数学语言　单边变换为 $F(s)=\int_0^\infty f(t)\mathrm{e}^{-st}\,\mathrm{d}t$，$s=\sigma+\mathrm{i}\omega$。对实 a，

$$\mathcal{L}\{\mathrm{e}^{at}\}(s)=\frac{1}{s-a},\qquad \operatorname{Re}s>a.$$

这个条件保证积分在无穷远收敛。

完整算例　$f(t)=\mathrm{e}^{2t}$ 的变换在实 $s=3$ 处为 1，在 $s=1$ 处原积分发散，不能仅把形式有理式代进去当作积分值。

易错点　代数表达式与收敛域共同构成变换信息；有理式在某点有值不等于定义积分收敛。

自测与答案　常数函数 1 的单边变换？$1/s$，要求 $\operatorname{Re}s>0$。

Lean 对照　`DifferentialEquations.laplace_kernel`

验证实指数与拉普拉斯核相乘后合并为 $\mathrm{e}^{-(s-a)t}$ 的恒等式；不声称该恒等式本身证明了反演定理。

4.7　把微分方程变成代数方程

官方课程：`laplace-for-odes`

求导为何变成乘法

从 $\int_0^\infty y'(t)\mathrm{e}^{-st}\,dt$ 出发，对乘积做分部积分。导数从 y 移到指数核时带出 s，下端边界留下 $-y(0)$，无穷远边界在适当增长与收敛条件下消失。于是得到 $\mathcal{L}(y')=sY-y(0)$。

再做一次，就得到二阶导数的变换，除 s^2Y 外，还必须保留初始位置和初始速度。初值并没有因换到变换域而消失，而是以边界项的形式编码进去。

把一条完整求解链走完

考虑 $y'+2y=0, y(0)=3$。变换后是 $sY-3+2Y=0$，解得 $Y=3/(s+2)$，再从已知指数变换识别 $y=3\mathrm{e}^{-2t}$。最后求导代回并检查零时刻值，完成闭环验证。

二阶受迫振子类似：先整理成 $(s^2+\omega^2)Y = F + sy_0 + v_0$，再除以分母。每一项都有来源，若漏了初始值，得到的就会是另一个问题。分母中的根提示系统模态，但还需结合收敛域和约定理解。

变换法把困难移到了另一个表示中，并没有免除存在性和反演条件。Lean 在这里核验的是已经得到代数方程之后的求商步骤；积分分部与函数空间条件仍属于正文分析。读证明时要知道每一步解决了哪一部分，而不是把整个方法看成一条自动规则。

一步一步建立直观

1. 把一个初值问题看成待求的整条信号　选定初始状态后，局部速度规则产生整条未来轨迹。拉普拉斯方法将这条未知信号整体送到参数域，在那里把微分操作转成乘法，同时留下初始值项。右图的初值拖动可以提醒我们：初始条件改变，解就改变，变换计算不能把它丢掉。

2. 为什么导数变换带着边界项　对加权积分做分部积分，导数从未知信号转移到指数核上，因而出现乘以 s 的项。下端点则贡献初始值，上端点需借助衰减条件消失。因此公式不是单纯的“把求导替换成乘法”；边界与收敛假设同样参与计算。

3. 代数解仍要回到原问题核对　改成增长轨迹，观察解的长期行为怎样变化。在参数域求出的分式必须在合适的收敛域内解释，再还原时间函数，最后检查微分方程与初值。右图不是拉普拉斯计算器，而是用于核对所得解的轨迹、初值与稳定性是否合理的模型。

> **观课提示**　逐步记录分部积分产生的初值项，最后将反演结果代回原方程。
>
> 原课程视频　本节网页与媒体　参数实验

要点回看

核心概念　积分分部把求导转换为乘 s，同时保留初始值。于是线性常系数方程可以在变换域中求解。

整理成数学语言　在所需光滑、增长及收敛条件下，$\mathcal{L}(y') = sY - y(0)$，$\mathcal{L}(y'') = s^2Y - sy(0) - y'(0)$。若 $y'' + \omega^2 y = f$，则

$$Y(s) = \frac{F(s) + sy_0 + v_0}{s^2 + \omega^2}.$$

完整算例　$f = 0, y_0 = 0, v_0 = 1$ 给 $Y = 1/(s^2+\omega^2)$，当 $\omega > 0$ 时对应 $y(t) = \sin(\omega t)/\omega$。

易错点　漏掉初始值项会改变所求解。极点描述模态，但只有结合收敛域与变换约定才能正确反演。

自测与答案　$y' + 2y = 0, y(0) = 3$ 的变换与解？$Y = 3/(s+2)$，$y = 3\mathrm{e}^{-2t}$。

Lean 对照 `DifferentialEquations.transformed_oscillator`

从已得到的变换域等式及非零分母推导商公式。积分分部、拉普拉斯存在性和反演不在该形式命题中。

4.8 动态理解欧拉恒等式

官方课程：`eulers-formula-dynamically`

用运动规律识别一个函数

从复平面的点一出发，让速度永远等于当前位置乘以 i。速度与半径垂直，长度与半径相同，所以它沿单位圆以单位角速度运动。走过时间 t 后的位置应是 $\cos t + \mathrm{i}\sin t$。

另一边，指数函数 $\mathrm{e}^{\mathrm{i}t}$ 对时间的导数是 $\mathrm{i}\mathrm{e}^{\mathrm{i}t}$，起点也是一。两个表达式满足同一微分方程和初值，在相应唯一性结论下，它们就是同一条轨迹。欧拉公式因此可被理解为两种描述相遇。

特殊角度是整个运动的路标

时间 $\pi/2$ 到达 i，时间 π 到达负一，时间 2π 回到一。著名恒等式是半圈位置的记录；若只背五个常数排在一起，就会丢失背后的完整连续运动。

在动画中暂停于四分之一圈和半圈，分别读横纵分量，再把它们与余弦正弦对应。改变转动方向只改变到达路径，走正半圈或负半圈都到负一，这也展示了指数的周期性。

周期性将在复对数中变成多值性：同一个终点可以对应许多角度，所以从终点反推指数不能保持全局唯一。欧拉公式既统一了指数与三角，也提醒我们把实数反函数的习惯推广到复数时需要重新检查。

一步一步建立直观

1. 用同一个初值锁定两种表达 余弦和正弦组成的位置 $(\cos t, \sin t)$ 从 $(1, 0)$ 出发，速度是位置旋转直角后的向量。复指数 e^{it} 满足同样的局部规则与初值。若调用相应微分方程的唯一性，两种表达就描述同一运动，而不只是偶然在几个角度上数值相同。

2. 半圈的位置给出恒等式 当时间走到半圈，横坐标为负一、纵坐标为零，因此 $e^{i\pi} = -1$。右图保留速度箭头，强调等式中的常数来自整条连续运动在特殊时刻的位置。它不是把五个符号强行拼在一起，而是指数结构与圆周几何的交汇。

3. 检查哪一种修改会破坏单位圆 把实部增长率调正，点仍旋转，但半径不再固定。此时半圈位置不再是负一，而是负实轴上一个更远的点。这个对照帮助检验欧拉公式的准确条件：纯虚指数只改变相位，加入实部还会乘上实指数的模长因子。

观课提示　暂停在四分之一圈和半圈，读取复平面坐标并解释欧拉恒等式。

原课程视频　本节网页与媒体　参数实验

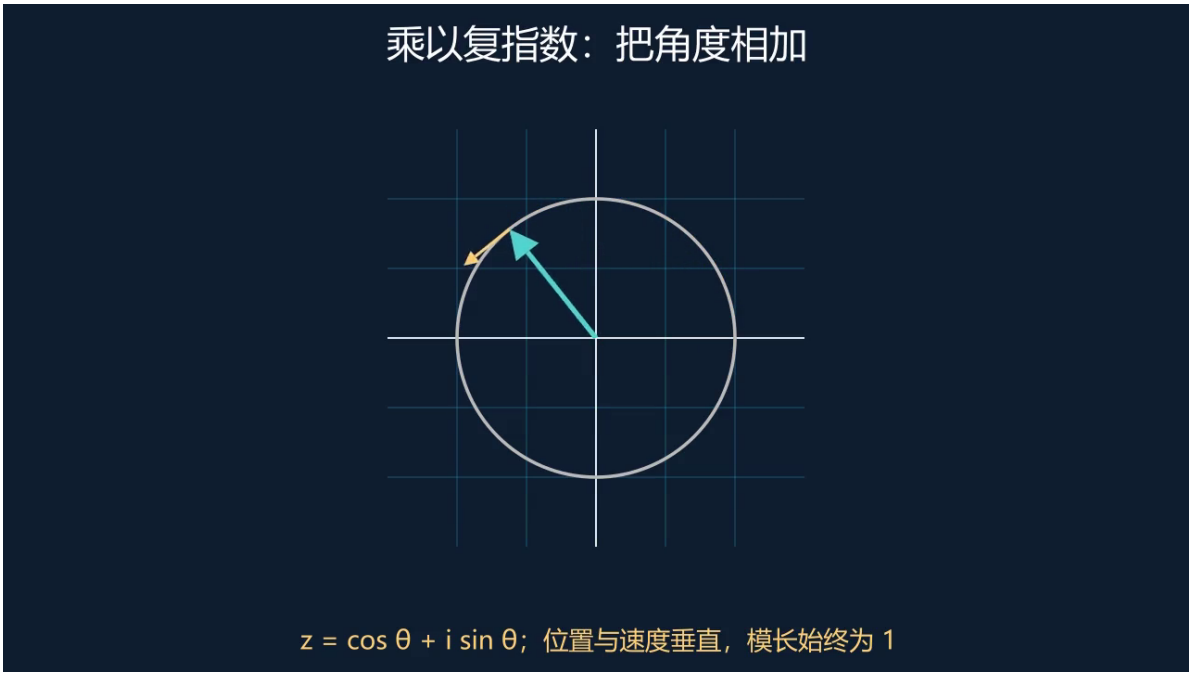


图 4.2: 单位圆上的位置（蓝）与切向速度方向（黄）；纯旋转保持模长。本项目原创 Manim 动画关键帧；完整动画与操作入口见本节网页。

要点回看

核心概念　复指数 $\mathrm{e}^{\mathrm{i}t}$ 从 1 出发沿单位圆运动，走过 π 弧度后到达 -1。

整理成数学语言　$\mathrm{e}^{\mathrm{i}t} = \cos t + \mathrm{i}\sin t$，所以 $\mathrm{e}^{\mathrm{i}\pi} + 1 = 0$。也可比较两边的微分方程 $z' = \mathrm{i}z$ 与初始值 $z(0) = 1$，在唯一性定理下识别二者。

完整算例　$t = \pi/2$ 时值为 i，$t = 2\pi$ 时回到 1。四个四分之一圈的乘积为 $\mathrm{i}^4 = 1$。

易错点　$\mathrm{e}^{z+w} = \mathrm{e}^{z}\mathrm{e}^{w}$ 在复数上成立，但取复对数后要考虑分支，不能直接取消周期性。

自测与答案　$\mathrm{e}^{-\mathrm{i}\pi}$ 是多少？同样为 -1；方向不同，终点相同。

Lean 对照　`DifferentialEquations.euler_identity`

完整验证复指数的欧拉恒等式，复用 mathlib 已证明的结论。

4.9　矩阵指数与线性动力系统

官方课程：matrix-exponents

把比例增长推广成向量之间的耦合

标量方程 $x' = ax$ 用指数描述。若状态包含多个分量，每个分量的速度都由全部当前分量的线性组合决定，就得到 $x' = Ax$。需要寻找一个随时间变化的矩阵，把初值送到后续状态，矩阵指数承担这个任务。

把标量指数的幂级数中数字 a 换成矩阵 A，得到 e^{tA}。它不是对每个元素单独取指数，因为矩阵乘法描述分量间反复耦合，A^2, A^3 保留了这种传递结构。

从能截断的例子建立信心

取 $A = \left[\begin{smallmatrix} 0 & 1 \\ 0 & 0 \end{smallmatrix}\right]$，对应方程 $x' = y, y' = 0$。第二分量固定，第一分量以固定速度移动，所以解应是 $(x_0 + ty_0, y_0)$。由于 $A^2 = 0$，指数级数恰好截断为 $I + tA$，给出同一结果。

先演化 s 再演化 t，等于演化 $s + t$。在本例中展开乘积，额外的 stA^2 为零，便直接验证了这个性质。它表达的是同一个自治系统连续运行的时间相加。

但不同矩阵代表不同耦合规则，通常不能把 e^{A+B} 拆成 $\mathrm{e}^{A}\mathrm{e}^{B}$，除非有相应交换条件。与矩阵乘法一样，动作顺序不可随意忽略。这个简单系统是入口，一般矩阵指数仍需要级数收敛和求导的分析基础。

一步一步建立直观

1. 让矩阵成为局部速度的生成规则　对线性系统 $x' = Ax$，矩阵把当前位置变成速度，而不是直接把当前位置变成下一时刻的位置。右图用旋转生成矩阵演示这一点：小箭头是 Ax，整段时间后的状态则由 $e^{tA}x(0)$ 给出。二者一个是瞬时规则，一个是累计演化。

2. 从很小的时间步想象指数　很短时间内，可以用 $(I + hA)x$ 近似新位置；不断缩小时间步并重复作用，便引出矩阵指数。加入负的径向项时，每步既旋转又收缩。本图使用这一特殊矩阵的精确闭式解，避免把数值步长造成的误差误认成方程本身的性质。

3. 特征方向解释增长，不能包办所有矩阵　把径向项改正，模长随时间增长。当前矩阵由伸缩与平面旋转组成，结构特别简单；一般矩阵可能有剪切、不可对角化等情况，需要利用级数、分块或其他方法。本节 Lean 示例也只覆盖正文声明的具体矩阵恒等式，不证明全部谱理论。

> **观课提示**　把矩阵理解为速度规则，使用平方零例子区分矩阵指数与逐元素指数。
>
> 原课程视频　本节网页与媒体　参数实验

要点回看

核心概念 标量指数把“按当前大小增长”变成演化；矩阵指数把“速度由当前向量线性决定”变成多维演化。

整理成数学语言 $\mathrm{e}^{tA}=\sum_{n\ge 0}t^nA^n/n!$，线性系统 $x'=Ax$ 的解为 $x(t)=\mathrm{e}^{tA}x_0$。若 $A^2=0$，级数截断为 $I+tA$。

完整算例 $A=\left[\begin{smallmatrix}0&1\\0&0\end{smallmatrix}\right]$ 满足 $A^2=0$，所以 $\mathrm{e}^{tA}=\left[\begin{smallmatrix}1&t\\0&1\end{smallmatrix}\right]$。初值 $(0,2)$ 演化为 $(2t,2)$。

易错点 一般 $\mathrm{e}^{A+B}\neq\mathrm{e}^A\mathrm{e}^B$，交换条件至关重要。逐元素取指数也不是矩阵指数。

自测与答案 为什么上例先演化 s 再演化 t 等于演化 $s+t$？$(I+sA)(I+tA)=I+(s+t)A$，因为 $A^2=0$。

Lean 对照 `DifferentialEquations.nilpotent_exponential_add`

完整证明平方零矩阵的上述半群恒等式；矩阵指数级数的收敛与求导由正文解释，未包含于此定理。

4.10 指数增长与饱和

官方课程：exponential-and-epidemics

先识别固定相对增长率的含义

指数模型 $N'=rN$ 说的是变化量与当前数量成比例，所以每过同样时间，数量乘上同样倍率。若 $r>0$，翻倍所需时间满足 $\mathrm{e}^{rT}=2$，得到 $T=\ln 2/r$，与当前绝对数量无关。

这解释了早期看似缓慢、之后迅速增加的曲线形状，但形状本身不能证明背后机制。任何固定相对增长率模型，都暗含条件在研究期间保持足够稳定；当资源或可参与群体有限时，这个假设会逐渐失效。

给增长规则加入容量限制

一个教学修正是 Logistic 方程 $N'=rN(1-N/K)$。数量远小于 K 时，括号接近一，恢复指数近似；数量接近 K 时，相对增长率减小，最终趋向平衡。这里 K 是模型设定的容量，不能不经估计就赋给真实系统。

把增长率看成 N 的函数，它是开口向下的二次式，在 $K/2$ 最大。于是总数量还在增加时，增加速度已经可以开始下降。区分状态与变化率，正是高阶导数章节中的同一个主题。

本节用于理解数学模型，不给实际疫情预测。真实观测还受到报告延迟、检测和行为变化等因素影响。合理的学习顺序是先从假设推导后果，再对照数据检验假设；不能因为一条曲线像指数或 S 形，就认定机制已经确定。

一步一步建立直观

1. 正反馈在资源充足时近似指数　把角速度设为零，只保留径向增长。在固定方向上，状态的大小满足“变化率与当前大小成正比”，于是呈指数增长。这可用来理解传播初期的简化机制：参与传播的人越多，单位时间新增的潜在来源也越多。图中数值没有真实人群的单位。

2. 区分数学相似与完整模型　第二种情形同时带有旋转；它说明相同的指数幅度可以配上完全不同的其他机制，却不代表流行病会按这条平面螺线演化。人口流动、接触结构与易感者消耗都需要额外状态变量。本窗口仅帮助辨认线性增长因子，不能充当预测工具。

3. 饱和需要改变方程，而非缩放画面　负的增长率使幅度下降，但这仍是线性指数模型。逻辑斯蒂增长则在方程中加入随存量降低的有效增长率，因而能趋近承载量。把坐标轴缩放得更大不会产生饱和；要解释增长变慢，必须指出模型中哪一项反馈发生了变化。

> **观课提示**　区分数量、增长量和相对增长率，检查指数模型何时需要容量限制。
>
> 原课程视频　本节网页与媒体　参数实验

要点回看

核心概念　若增长速度正比于当前数量，就会出现指数增长。资源或易感群体有限时，固定相对增长率通常只适合早期阶段。

整理成数学语言　指数模型 $N' = rN$ 的倍增时间为 $\ln 2/r$（$r > 0$）。一个简化饱和模型是 $N' = rN(1 - N/K)$，其中 $K > 0$；在 $0 < N < K$ 时仍增长，但相对增长率逐渐降低。

完整算例　若相对增长率从 r 降为 $r/2$，在指数近似下倍增时间翻倍。对 Logistic 模型，$N = K/2$ 时增长率为 $rK/4$。

易错点　这是数学教学模型；观测数据、报告延迟和行为改变均可能使真实时间序列偏离模型，不能仅凭曲线外形确认机制。

自测与答案　Logistic 模型的平衡点有哪些？$N = 0$ 和 $N = K$；对 $r > 0$，区间 $(0, K)$ 内导数为正。

Lean 对照 `DifferentialEquations.logistic_threshold`

验证正增长率、正容量且状态处于 $(0, K)$ 时导数严格为正；未形式化流行病预测或参数估计。

章节综合：先分模式，再读时间尺度

取周期边界上的初值 $u(x,0) = \sin x + 2\sin(3x)$，扩散系数为一。两个模式分别是空间二阶导算子的特征函数，所以解为

$$u(x,t) = \mathrm{e}^{-t}\sin x + 2\mathrm{e}^{-9t}\sin(3x).$$

在 $t = \ln 2$ 时，第一振幅为 $1/2$，第二振幅为 $2/2^9 = 1/256$。总曲线几乎由低频决定，这个判断来自明确倍率，而不仅是动画看起来变平滑。

把每一项分别对时间求一次导、对空间求两次导，便能检查方程两边相等；再设 $t = 0$ 检查初值；最后检查区间两端的周期匹配。三个检查分别对应方程、起点和边界，不能互相代替。

本例只含有限两项，线性叠加没有交换无限求和与导数的困难。若想推广到任意初始曲线，仍需傅里叶展开与收敛理论。理解一个解的可靠方式，是同时说清计算为何有效，以及推广时新增了哪些问题。

第五章　Lockdown Math：重新组织基础数学

这一组从高中代数、三角函数和复数出发，把看似独立的公式联成一条学习路径。公告是导读，其余十节有数学正文与对应 Lean 结果。

5.1　课程公告与学习路线

官方课程：`lockdown-math-announcement`

把这组课程当成一条重新学习的路线

这条目录位置是课程公告，不是一节包含新定理的数学课。它在教材中的作用，是说明后续基础主题怎样连接：二次方程教我们用中心化简结构，单位圆教我们用运动解释三角，复数把旋转加入乘法，对数再把乘法翻译成加法。

可以先读本组，再进入线性代数与微积分；也可以遇到复指数或对数困惑时回来补基础。学习顺序不应只按公式难度排，而应看后一概念需要借用前一概念的哪些解释。例如复数幂需要先知道指数的周期性与对数分支，否则容易只剩形式计算。

怎样把观看变成能独立复述的理解

每次暂停后，用自己的例子解释正在解决的问题。不要先问“这一页有哪些公式”，而要问“为什么需要这个新表示，上一种表示哪里不够方便”。如果无法说出引入符号的原因，通常说明还没有把推导连接起来。

做练习时先保留预测，再看答案。错误预测可以告诉你混淆了哪些对象：把坐标当向量，把有限步近似当极限，或把必要条件当充分条件。修正这类误解比多抄一遍结论更有价值。

本书把连续讲解放在前面，把定义回顾、算例、易错点和 Lean 边界放在后面。先沿着叙述重建思路，再用回顾定位概念，最后核对证明中的假设。公告页本身没有机器定理，这种区分也是认真记录学习范围的一部分。

一步一步建立直观

1. 用一个能亲手改变的对象开始　先把两根箭头当作地图上的两次移动，拖动它们的端点，描述合位移如何变化。此处不需要预先知道向量空间的定义；目标是先形成一个明确问题：哪些变

化来自移动的次序，哪些来自方向和长度？这会成为后面比较不同表示方法的练习。

2. 每次只改变一件事，再解释结果　保持两根箭头不动，只改变其中一根的系数。先写下你预计的方向和长度，再查看结果。学习数学时，操作的价值在于让预测暴露出来，而不是替代思考。若结果不符，应追问是计算错误、坐标误解，还是原先使用的规则缺少条件。

3. 把特殊情形当作检查工具　让两根向量共线，观察原先能覆盖平面的组合怎样退化到一条线。零、对称、相等和极端情形经常能揭示一般公式隐藏的条件。本条目是学习路线说明；右图只是这一学习方法的示例，不把课程公告改写成一条不存在的原课数学定理。

> **观课提示**　把公告作为学习路线导读，记录后续每个新概念解决了什么问题。
>
> 原课程视频　本节网页与媒体　参数实验

要点回看

核心概念　这是一组由基础概念逐步深入的直播课程。学习时可以把“定义”“例子”“猜想”“证明”分别记下；直播中的探索过程不应被压缩成只有结论的公式表。

整理成数学语言　推荐先掌握二次方程和单位圆，再学复数、复指数与对数，最后处理幂塔和复数幂。每一步都检查新符号的定义域与旧代数规律是否仍成立。

完整算例　将 $\sqrt{x^2}=x$ 写入笔记前，应代入 $x=-1$。反例提示正确实数恒等式是 $\sqrt{x^2}=|x|$。

易错点　看懂讲解不等于能独立重建推导。复习时先遮住答案，画图、做特殊例、检查边界，再写一般论证。

自测与答案　用自己的话区分“例子支持猜想”和“证明结论”。答案要点：有限个例子不能穷尽一般情形，但一个反例足以否定全称命题。

Lean 对照　`not-applicable`

不适用：这是公告和学习方法页，没有需要机器证明的课程命题。仓库的覆盖统计对此显式豁免。

5.2　更简洁的二次公式

官方课程：`ldm-quadratic`

不要先追两根，先追它们的中心

若两个数的和已知，就知道它们的平均值。把两根写成 $m-d, m+d$，和自动成为 $2m$，只剩一个未知偏移 d。它们的积是 m^2-d^2，因此给定和与积之后，问题只剩求一个平方。

对 $x^2-10x+21=0$，两根中心是五，偏移平方是 $25-21=4$，于是根为三和七。这里不是省略了求根原理，而是选择了让对称性直接显现的变量。配方 $(x-5)^2=4$ 是同一想法的另一种写法。

从特殊形式回到一般二次式

面对 $Ax^2+Bx+C=0$，先确认 $A\neq 0$，再除以它，把二次项系数变为一。中心是 $-B/(2A)$，剩下的判别信息来自偏移平方。若它为零，两根重合；若为负，实数中无根，但复数中仍可解释。

还可以从抛物线看：平移到对称轴后，一次项消失，左右两交点围绕中心成对出现。方程的代数对称性和图像对称轴于是对应起来。

解完后检查两根的和与积，能快速发现符号错误。这个方法以后还会出现在二维特征值计算中：迹给根的和，行列式给根的积。基础代数不是独立的技巧仓库，它会在更高层问题里反复提供结构。

一步一步建立直观

1. 从中心开始看两个根的对称性　对一元二次式，配方把一次项吸收到平移后的平方中，两个实根便围绕同一个中心对称。右图的对称矩阵提供一个相关例子：其特征值满足二次特征方程，两个根的中点是迹的一半。这里先读矩阵，再把它当作检验配方结果的具体对象。

2. 中心加上偏移得到一个根　当前矩阵的对角元相等，特征方向之一沿两坐标同号的对角线，放大倍率为三。把特征方程写成围绕中心二的平方关系，就得到向右偏移一的根。一般配方的中心与偏移应从系数计算，而不是凭图上看起来对称直接猜出。

3. 另一个根来自相反的偏移　转到另一条特征方向，放大倍率为一，恰好在中心二的另一侧。这个例子有两个实根，因为实对称矩阵的特征值为实数；普通二次方程还可能出现负判别式。图示说明配方的对称结构，不能把本例的实根性质推广到任意系数。

> **观课提示**　先找两根的平均值，再求偏移；用根的和与积核对答案。
>
> 原课程视频　本节网页与媒体　参数实验

要点回看

核心概念　把两个根看成“中心 m 加减偏移 d”，和与积就变为 $2m$ 与 m^2-d^2。

整理成数学语言　对 $x^2-2mx+p=0$，配方得到 $(x-m)^2=m^2-p$，所以 $x=m\pm\sqrt{m^2-p}$。一般二次式须先除以非零二次项系数。

完整算例 $x^2 - 10x + 21 = 0$，根的中心为 5，偏移平方为 $25 - 21 = 4$，故根为 $3, 7$。

易错点 根号内为负时没有实根，但可以有复根；使用中心形式并没有免除域的选择。

自测与答案 根的和为 8、积为 12，求根。中心 4，偏移 2，答案 $2, 6$。

Lean 对照 `Lockdown.quadratic_centering`

完整验证配方恒等式；从平方等式分类实根或复根需另行处理。

5.3 三角函数从单位圆出发

官方课程：`ldm-trigonometry`

从直角三角形走到整圈运动

锐角三角比只覆盖部分角度，单位圆则能连续描述整圈。点在圆上转到角度 θ 时，横坐标定义为余弦，纵坐标定义为正弦。负角、超过一圈的角和各象限符号，都由点的位置统一解释。

半径为一，所以两坐标平方和为一。这不是一条需要孤立记住的恒等式，而是点始终在单位圆上的坐标表达。关于纵轴或横轴反射，也会直接给出一批符号关系。

和角公式来自两次旋转

先旋转 β，再旋转 α，总角度是 $\alpha + \beta$。把每次旋转写成矩阵，乘积的第一列记录标准横向箭头最终落点；与总旋转矩阵比较，就得到正弦余弦的和角公式。

这条路线解释了为什么正弦和角不是简单相加：第二次旋转会同时混合前一步的横纵分量。公式中的乘积项记录这种投影与混合，而非人为增加的复杂性。

在复平面实验里，先暂停在某个角度，预测再转四分之一圈后的横纵坐标，再改变滑块验证。学习三角时应能在角度、点坐标、旋转动作三种语言之间转换。后面复数乘法会把同一个旋转矩阵压缩成一个复数。

一步一步建立直观

1. 先读圆上的坐标，再给它们命名 把单位圆上一点投影到横轴和纵轴，两个读数就是余弦与正弦。角度决定的是位置，不是单独某条三角形边的长度。这个表示在越过直角后仍然有效，坐标符号自然改变，所以三角函数不必被限制在锐角直角三角形中。

2. 用对称性检查符号和周期 转到半圈，横坐标变为负一，纵坐标回到零。继续移动点，可以通过圆的对称性推断其他象限的符号与数值关系。因为横纵坐标来自同一根单位向量，它们的平方和始终为一；这条恒等式有直接的长度解释。

3. 区分单位圆与一般极坐标　加入径向增长后，点不再位于单位圆。方向仍用角度描述，坐标却要乘以当前半径，平方和也变成半径的平方。这个对照帮助避免把 $x = \cos\theta$、$y = \sin\theta$ 无条件套到任意平面点上；一般情况还必须记录模长。

> **观课提示**　从单位圆读出符号和投影，把和角公式理解成两次旋转。
>
> 原课程视频　本节网页与媒体　参数实验

要点回看

核心概念　单位圆将角度和坐标联系起来：$\cos\theta$ 是横坐标，$\sin\theta$ 是纵坐标。三角恒等式可由旋转、投影和勾股关系产生。

整理成数学语言　$\sin^2\theta + \cos^2\theta = 1$，且旋转矩阵为

$$R_\theta = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}.$$

$R_\alpha R_\beta = R_{\alpha+\beta}$ 可导出和角公式。

完整算例　对 $\theta = \pi/3$，坐标为 $(1/2, \sqrt{3}/2)$，平方和为 $1/4 + 3/4 = 1$。

易错点　$\sin(\alpha+\beta)$ 通常不等于 $\sin\alpha + \sin\beta$。角度单位改变时，涉及导数的公式需相应换算。

自测与答案　$\cos(\pi-\theta)$ 和 $\sin(\pi-\theta)$？分别为 $-\cos\theta$、$\sin\theta$。

Lean 对照　`Lockdown.unit_circle`

完整验证实数任意角度的单位圆恒等式。

5.4　复数：缩放与旋转

官方课程：`ldm-complex-numbers`

扩展数系，也扩展乘法能表达的动作

实数乘法能沿数轴缩放或反向，却无法表示平面中的四分之一圈。引入 i 并规定 $\mathrm{i}^2 = -1$，可以把实数对写成 $a + b\mathrm{i}$。乘以 i 得到 $-b + a\mathrm{i}$，正是把平面坐标转九十度。

两次四分之一圈合为半圈，所以 $\mathrm{i}^2 = -1$ 的几何解释与代数规则一致。这个扩展不是让原有规则随意失效，而是在保留加法乘法基本规律的同时，让原本无实数解的方程得到新的表示。

同一个乘法有两种阅读方式

直角坐标形式适合展开：按分配律相乘，再把 i^2 换成负一。极坐标形式适合看整体动作：模长相乘、角度相加。对 $(1+\mathrm{i})^2$，展开得 $2\mathrm{i}$；几何上是两次四十五度旋转，模长 $\sqrt{2}$ 也相乘为二。

求倒数则需要撤销这两个动作：模长取倒数、角度取负。乘共轭能在坐标中实现同一件事，因为分母变成模长平方。请把 $1/(1+\mathrm{i})$ 先画成反向旋转加缩小，再算出 $(1-\mathrm{i})/2$。

复数有自然的长度和角度，却没有与实数完全相同的有序大小关系。谈“更大”时要说明是模长、实部还是别的量。下一节指数语言会把这套旋转与缩放规则进一步统一。

一步一步建立直观

1. 把一对实数当作一个可相乘的对象　复数 $a+bi$ 对应平面中的点，但关键增加的是乘法规则。乘以 i 把实轴正方向转到虚轴正方向，继续乘一次就到负实轴。右图先停在四分之一圈，用位置而非符号记忆来理解这个动作。

2. 两次直角旋转解释平方为负一　半圈对应两次乘以 i，因此 $i^2=-1$ 与平面旋转相容。一般复数乘法则把模长相乘、辐角相加。这种结构让复数适合同时记录缩放和旋转；它不是在实数轴上寻找一个平方为负数的普通实数。

3. 连续的缩放与转动可以合成　在螺线情形中，每一小段运动同时改变长度和方向。用复数相乘，两个动作可以用一个数统一表示。零复数没有确定辐角，取辐角还存在整圈的不唯一性；这些边界在研究复对数时会再次出现，不能被圆上的直观省略。

定义、命题与证明

定义：复数的模长　对 $z=a+bi$，定义 $|z|=\sqrt{a^2+b^2}$。它等于复平面中从原点到该点的距离。

命题：乘以 i 保持长度并旋转直角　若 $z=a+bi$，则 $iz=-b+ai$，且 $|iz|=|z|$；对非零 z，该操作是逆时针直角旋转。

证明. 由 $i^2=-1$，有 $i(a+bi)=ai-b$。坐标变换为 $(a,b)\mapsto(-b,a)$，新旧长度平方均为 a^2+b^2，点积则为 $-ab+ab=0$。

再检查 $(1,0)$ 被送到 $(0,1)$，就确定了旋转方向。零向量同样保持长度，但没有可供比较的辐角。

□

观课提示　比较坐标展开与旋转缩放两种乘法解释，检查倒数如何撤销动作。

原课程视频　本节网页与媒体　参数实验

要点回看

核心概念 实数乘法能缩放数轴，复数乘法还能旋转平面。新单位 i 满足 $\mathrm{i}^2 = -1$，与二维坐标运算一致。

整理成数学语言 $(a+b\mathrm{i})(c+d\mathrm{i}) = (ac-bd)+(ad+bc)\mathrm{i}$。若写成极坐标 $r\mathrm{e}^{\mathrm{i}\theta}$，相乘时模长相乘、角度相加。

完整算例 $(1+\mathrm{i})^2 = 2\mathrm{i}$：两个 45° 旋转合为 90°，模长从 $\sqrt{2}$ 平方为 2。

易错点 复数不能像实数那样建立兼容乘法的全序；“更大”必须说明比较模长、实部或其他量。

自测与答案 $1/(1+\mathrm{i})$ 是什么？乘共轭得 $(1-\mathrm{i})/2$。

Lean 对照 `Lockdown.complex_square_i`

验证复数单位的核心关系 $\mathrm{i}^2 = -1$；其余运算公式可通过复数域的代数规则推出。

5.5 欧拉公式在说什么

官方课程：`ldm-eulers-formula`

指数把相加的参数变成复合的动作

实指数满足参数相加、结果相乘。旋转也有同样结构：两个角度相加，对应两个旋转相继执行。因此纯虚指数与单位圆之间可能有联系，但结构相似还不是证明，需要检查它们确实满足同一数学定义或规律。

一条证明路线从幂级数出发：偶数次 i 的幂给实部交错项，奇数次给虚部交错项，分别组成余弦与正弦。另一条路线从微分方程出发，比较速度与初值。本书的动态章节展开后一种解释。

欧拉公式让三角关系变成指数运算

有了 $\mathrm{e}^{\mathrm{i}\theta} = \cos\theta + \mathrm{i}\sin\theta$，两次旋转相乘就能写成一个指数。比较结果的实部和虚部，会重新得到和角公式。指数规则与三角恒等式不再是两套互不相干的记忆任务。

进一步写 $\mathrm{e}^{x+\mathrm{i}y} = \mathrm{e}^x\mathrm{e}^{\mathrm{i}y}$，实部参数控制缩放，虚部参数控制角度。这里“实部参数”指指数里的 x，不要误认为结果的实部就是 e^x；结果实部还要乘 $\cos y$。

角度绕整圈后回到同一点，因此复指数不单射。下一步取对数时必须面对这个事实，不能简单把指数符号取消。一个好公式既能简化计算，也会暴露哪些旧习惯需要修正。

一步一步建立直观

1. 从乘法结构寻找圆上的指数　实指数把相加的参数变成相乘的倍率；圆周旋转也有同样结构，角度相加等于旋转连续执行。因此用 $e^{i\theta}$ 表示单位圆位置，与指数的组合规律相容。右图让你先把这个结构看成动作，再与正弦余弦坐标对应。

2. 特殊角度只是一般规律的取样　在半圈处读到负一，在整圈处回到一。这些值帮助检查符号，却不能独自证明所有角度下的欧拉公式。可以通过幂级数、微分方程等方法把一般规律建立起来；本节正文采用的论证应与它依赖的定义相互配套。

3. 指数的实部负责模长　第三步让指数同时具有实部和虚部，轨迹离开单位圆。实部带来指数缩放，虚部带来旋转。理解了这种分工，就不会把复指数的“周期性”错误地理解成所有复方向都周期；周期出现在纯虚方向的整圈平移中。

> **观课提示**　观察指数的加法规则怎样对应旋转复合，同时记住整圈周期性。
>
> 原课程视频　本节网页与媒体　参数实验

要点回看

核心概念　指数将加法转换为乘法；纯虚数指数把角度相加转换为旋转复合。这统一了三角恒等式与指数规律。

整理成数学语言　$\mathrm{e}^{\mathrm{i}\theta}=\cos\theta+\mathrm{i}\sin\theta$，并且 $\mathrm{e}^{z+w}=\mathrm{e}^{z}\mathrm{e}^{w}$。比较实部与虚部便得到和角公式。

完整算例　$\mathrm{e}^{\mathrm{i}\pi/3}\mathrm{e}^{\mathrm{i}\pi/6}=\mathrm{e}^{\mathrm{i}\pi/2}=\mathrm{i}$；这比逐项展开两个三角表达式更短。

易错点　欧拉公式不是把实指数的图像“转过去”就自动成立；可以由幂级数或微分方程严谨建立。

自测与答案　$\mathrm{e}^{x+\mathrm{i}y}$ 的模长是多少？e^{x}，角度在模 2π 意义下为 y。

Lean 对照　`Lockdown.exponential_add`

完整验证复指数的加法到乘法规则。欧拉特值另由微分方程模块中的 `euler_identity` 验证。

5.6　虚数利率与连续旋转

官方课程：`ldm-imaginary-interest`

把复利的小步改成切向小步

普通复利每步按当前量增加一小部分。若把增量乘上 i，新增部分不再沿原方向，而是垂直于原方向。一步变成乘 $1+\mathrm{i}h$，不断重复，就得到一条绕原点前进的折线轨迹。

每步到底是不是旋转？它的模长为 $\sqrt{1+h^2}>1$，所以还夹带一点向外扩张。图像若只强调绕圈，容易隐藏这个有限步误差。精确的圆周运动来自步长趋零后的极限，而不是任意有限步都在圆上。

为什么误差在连续极限消失

固定总时间 t，分成 n 步，模长倍率是 $(1+t^2/n^2)^{n/2}$。单步误差是二阶的，总步数只是一阶增加；取对数后约为 $t^2/(2n)$，趋向零，所以总模长趋向一。

角度每步增加 $\arctan(t/n)$，总角度为 $n\arctan(t/n)$，趋向 t。模长和角度两个条件合起来，解释了极限为何是 $\mathrm{e}^{\mathrm{i}t}$。这里用到的局部近似必须在小步长下理解。

这个例子也说明欧拉数值法可能产生什么误差：真实系统守恒半径，有限切向步却慢慢漂离。把步数加倍，观察偏差变化，会比单看一条漂亮轨迹更有教学价值。Lean 只核验单步模长恒等式，收敛论证仍需另外完成。

一步一步建立直观

1. 把每次增加一点理解为乘法更新　实数复利每一步乘以 $1+r\Delta t$。若把增长率换成 i，一步变成乘以 $1+i\Delta t$，它近似一次小旋转，却还带着略大于一的模长。右图显示连续极限的圆周轨迹，而不是直接播放有限次复利步骤。

2. 为什么细分后多余增长会消失　单步模长为 $\sqrt{1+\Delta t^2}$，偏离一的主要量是二阶；完成固定总时间需要的一阶数量级步数，不足以留下非零的二阶累计模长误差。另一方面，每步的一阶转角会积成有限总角度。这是旋转留下、额外膨胀消失的原因。

3. 真实径向增长与离散误差不同　螺线情形故意加入非零实部，表示方程中真实存在的径向增长。它不会因为时间步细分而消失；有限步虚数复利的模长偏差则会在极限中消失。两幅外观相似的螺线可能来自不同机制，必须根据更新规则区分。

> **观课提示**　仔细区分有限切向步与精确圆周运动，追踪有限步的模长误差。
>
> 原课程视频　本节网页与媒体　参数实验

要点回看

核心概念　把复利每一步的增长率改成纯虚数，每个小步就增加一个切向位移；步长越来越小时，轨迹趋于圆周运动。

整理成数学语言 离散近似为 $(1+it/n)^n$。单步模长平方为 $1+(t/n)^2$，整个模长为 $(1+t^2/n^2)^{n/2}$，其极限为 1。

完整算例 单步取 $h=0.1$，$1+0.1\mathrm{i}$ 的模长是 $\sqrt{1.01}>1$；有限步的轨迹略向外扩张，这个误差在连续极限消失。

易错点 每一个离散步骤都不是精确旋转；其模长大于一。不能把极限性质反过来当成有限步恒等式。

自测与答案 $h=1$ 时，乘 $1+\mathrm{i}$ 后角度和模长怎样变化？增加 $\pi/4$，模长乘 $\sqrt{2}$。

Lean 对照 `Lockdown.imaginary_step_norm`

验证单步复数的模长平方公式。复利序列收敛到复指数的极限定理未在此命题中形式化。

5.7 对数的基础结构

官方课程：`ldm-logarithms`

问“乘了几次”，而不是再发明一种运算口诀

指数把次数送到倍率，对数反过来问某个正数对应多少次增长。$\log_2 8=3$ 的内容就是 $2^3=8$。当次数允许实数，对数成为连续的尺度转换，而不限于整数乘法次数。

若 $a=c^u, b=c^v$，乘积是 c^{u+v}，所以乘积的对数是两对数之和。这个法则直接来自指数的加法规则。商、幂的对数关系也可从同一来源推导，不必各自背一遍。

换底只是换了一把刻度尺

底数二按翻倍计数，底数十按十倍计数，自然底数按连续增长计数。同一倍率在不同刻度中数值不同，彼此只差一个固定比例，因而出现换底公式。

输入为 $1/9$ 的三底对数是负二，负号表示反向增长或缩小；真数依旧为正。实对数的定义域不能因结果允许负数就扩展到负真数。底数一又无法区分任何次数，因此不能作为对数底。

对数尤其适合把很大的倍率范围压成可读的加性刻度，但必须知道图上的等距离表示等比例，而非等绝对差。后面的自然对数将把这条思想与微小相对变化联系起来。

一步一步建立直观

1. 先把对数读作倍率的坐标 对数回答“要使用多少指数参数，才能得到这个正倍率”。乘法在原尺度上组合倍率，在对数尺度上则变成加法。右图选择自然对数，用曲线展示这种坐标压缩：相同的绝对输入增量，作用在较大基数上时代表较小的相对变化。

2. 用其他函数作对照，辨认曲率　切换到平方函数，看到输入越大，局部变化率反而越大。这与对数曲线越来越平缓形成对照。函数图像的弯曲不是装饰，它反映输出更关心绝对变化还是相对变化。不过乘法变加法的恒等式仍需由指数与对数定义推导，不能只凭曲线形状断言。

3. 定义域是意义的一部分　回到对数函数并移动观察点。实对数要求输入为正，因为它在这里是正倍率的指数坐标；接近零时导数变大，非正位置则没有实函数值。复数对数可以扩展讨论，却需要处理辐角分支，不能将它与实对数公式不加说明地混用。

> **观课提示**　把对数读作倍率对应的指数，并检查底数与真数条件。
>
> 原课程视频　本节网页与媒体　参数实验

要点回看

核心概念　对数把乘法关系改写为加法关系。它是指数函数的反函数，因此必须清楚底数和定义域。

整理成数学语言　对实数 $a, b > 0$，$\ln(ab) = \ln a + \ln b$；对底数 $c > 0, c \neq 1$，$\log_c a = \ln a / \ln c$。

完整算例　$\log_2 8 = 3$，因为 $2^3 = 8$。$\log_2(8 \cdot 4) = 3 + 2 = 5$，乘法变成指数相加。

易错点　实对数不能直接输入负数；$\log(a + b)$ 没有一般的加法拆分式。复对数的乘法公式还受分支影响。

自测与答案　$\log_3(1/9)$？答案 -2。负的是对数值，真数仍为正。

Lean 对照　`Lockdown.logarithm_product`

完整验证正实数乘积的自然对数公式，并显式要求两个真数为正。

5.8　自然对数为何自然

官方课程：`ldm-natural-logs`

一种对相同比例变化给相同读数的刻度

从一百增加到一百零一，与从一千增加到一千零一，绝对变化相同，但相对变化不同。自然对数在局部测量的是 dx/x，因此更接近按比例的变化尺度。相同百分比增幅产生相同对数增量。

因为 ln 与指数互逆，对 $e^{\ln x} = x$ 求导，得到 $x(\ln x)' = 1$，所以导数为 $1/x$。定义域要求 $x > 0$，分母也说明不能在零点套用这个推导。

用面积把乘法法则重新看一遍

曲线 $1/t$ 从一到 x 的有向面积等于 $\ln x$。区间同比例缩放时，高度相应反比例变化，面积保持；这为乘法转换成加法提供了另一种几何解释。

当 x 从一百变为一百零一，对数增量为 $\ln(1.01)$，一阶近似是 0.01。但若从一变为二，直接把对数增量当一就不准确，因为变化已不够小。局部相对变化近似不能当成全局等式。

读对数图时也要区分“增长量下降”和“相对增长率下降”。指数曲线在对数坐标中成为直线，因为它每单位时间增加相同的对数值。这个表示很适合检查比例增长假设，但图形近似直线仍不足以证明具体机制。

一步一步建立直观

1．相对变化给出自然的局部单位　若存量为 $a>0$，很小的绝对增加 h 对应相对增加约 h/a。自然对数把这种相对增量变成输出增量，所以局部斜率是 $1/a$。右图在 $a=1$ 附近的切线斜率为一，正好给出一套没有额外比例常数的刻度。

2．不是每个函数都测量相对变化　切换到平方函数，比较同样的观察位置与步长。它的斜率是 $2a$，随存量增大而增大，与对数恰好相反。这一步用于突出自然对数的特殊尺度；右图切换函数后的读数属于新函数，不能把它继续解释成对数导数。

3．从导数回到积累的解释　回到对数曲线，试着解释从一增长到 x 的对数值：它等于沿途累计 $1/t$ 的面积。因为每段按当前存量归一化，同样的比例增长产生相同的对数增量。将起点与终点保持为正，面积解释便与对数的乘法规律互相支持。

> **观课提示**　比较相同绝对增长与相同比例增长，用倒数面积解释对数尺度。
>
> 原课程视频　本节网页与媒体　参数实验

要点回看

核心概念　$\ln$ 把相对变化变成加性变化：若 x 小幅增加 dx，则 $d\ln x \approx dx/x$。这让相同比例的增长具有相同对数增量。

整理成数学语言　对 $x>0$，$(\ln x)' = 1/x$，且 $\ln x = \int_1^x dt/t$。与 e^x 互逆给出 $\ln(\mathrm{e}^x) = x$。

完整算例　x 从 100 增为 101，对数增量 $\ln 1.01 \approx 0.00995$，一阶估计为 0.01。

易错点　近似 $\ln(1+h) \approx h$ 只适用于 h 接近零且 $1+h>0$；它不是全局恒等式。

自测与答案　$\ln(x^2)$ 在 $x>0$ 时的导数？$2/x$；也可先写成 $2\ln x$。

Lean 对照 `Lockdown.natural_log_derivative`

完整验证正实数处的自然对数导数，写明 $x > 0$ 避免除零和实定义域误解。

5.9 幂塔与不动点

官方课程：`ldm-power-towers`

先说明无限符号代表哪个过程

写下无限幂塔之前，先定义有限迭代 $x_0 = 1, x_{n+1} = a^{x_n}$。无限表达式若要有意义，应是这个序列的极限，而不是允许任意剪掉顶部、移动括号的代数对象。

如果序列确实收敛到 L，连续性允许把极限送入函数，得到 $L = a^L$。这只是极限必须满足的条件，还没有保证序列收敛，也没保证方程的每个根都能从当前初值到达。

不动点还要分稳定与不稳定

取 $a = \sqrt{2}$，二与四都满足不动点方程。附近的微小偏差经过一轮迭代，大约乘以 $f'(L) = L \ln a$。在二处倍率小于一，偏差局部缩小；在四处倍率大于一，偏差局部放大。

把图 $y = a^x$ 与 $y = x$ 画在同一坐标系，从当前值竖直走到函数曲线，再水平走回对角线，可以追踪迭代。这样的蛛网图帮助观察靠近、远离或振荡，但严格证明仍需单调、有界或压缩等条件。

一个重要解题习惯在这里非常清楚：从“假设存在极限”推出方程之后，必须回头检查存在性。否则会把必要条件误当充分条件。Lean 定理也明确把收敛写成假设，保留了这道不能被符号操作跳过的门槛。

一步一步建立直观

1. 先写出递推，而非直接画无穷高 把有限幂塔看作迭代 $x_{k+1} = b^{x_k}$。右图的蜘蛛网先竖直到达函数曲线，再横着回到对角线，把输出作为下一次输入。若序列真的收敛且指数函数连续，极限必满足 $x = b^x$；这一步给出必要条件，还没有保证收敛。

2. 交替接近也可能收敛 当底数小于一时，指数函数递减，迭代可能在不动点两侧往返。检查附近导数的绝对值是否小于一，有助于理解小误差是被缩小还是放大。图中几次迭代看似稳定只是线索；局部收缩还需结合进入该邻域的条件。

3. 不动点方程不能替代收敛证明 将底数提高后，轨迹越走越大，最终可能超出显示范围。程序会明确报告溢出范围，而不会把屏幕边缘当成极限。无穷表达式若没有先定义为一个收敛过程，就不能随意把“顶端的无穷幂塔”替换为自身并当作求值完成。

定义、命题与证明

定义：幂塔的有限迭代　给定底数 $b>0$ 和起点 x_0，定义 $x_{k+1}=b^{x_k}$。所谓无限幂塔的取值，需要先说明这个序列的极限是否存在。

命题：收敛的迭代必到达不动点　若 x_k 收敛到有限实数 L，则 $L=b^L$。该方程只是收敛的必要条件。

证明. 收敛序列平移一项仍趋于 L，所以 $x_{k+1}\to L$。函数 $x\mapsto b^x$ 在实数上连续，故 $b^{x_k}\to b^L$。

迭代两边逐项相等，极限的唯一性给出 $L=b^L$。证明从“序列收敛”开始，不能倒过来用同一个推理宣称只要方程有解，所有初值的迭代就会收敛。

□

> **观课提示**　先定义有限迭代，再检查收敛；区分不动点方程与稳定性。
>
> 原课程视频　本节网页与媒体　参数实验

要点回看

核心概念　无限幂塔应先定义为有限迭代，再讨论其极限。如果迭代确实收敛，极限才必须满足不动点方程。

整理成数学语言　对 $a>0$，令 $x_0=1, x_{n+1}=a^{x_n}$。若 $x_n\to L$，由连续性得到 $L=a^L$。局部稳定性的常用线性判据是 $|a^L\ln a|=|L\ln a|<1$。

完整算例　$a=\sqrt{2}$ 时 $L=2$ 和 $L=4$ 都满足方程；在前者的导数为 $\ln 2<1$，后者为 $2\ln 2>1$。不动点不一定是从给定初值可达的稳定极限。

易错点　解出 $L=a^L$ 仅给必要条件，不能用它证明迭代收敛。完整参数范围还需要单调性、有界性等分析。

自测与答案　$a=1$ 时该迭代的每一项与极限？都为 1。

Lean 对照　`Lockdown.tower_limit_fixed_point`

完整证明一般连续迭代在已知收敛时，其极限为不动点；收敛是假设，绝不隐含在代数方程里。

5.10　为什么 i 的 i 次方可以是实数

官方课程：`ldm-i-to-i`

幂运算先需要一个定义

实数正底数的幂可以写成 $a^b = \exp(b \ln a)$。把底数扩展到复数，首先要解决复对数的问题。复指数绕整圈回到同一点，因此 i 对应角度 $\pi/2 + 2\pi k$，而非只有一个角度。

选择其中一个对数值 $\log \mathrm{i} = \mathrm{i}(\pi/2 + 2\pi k)$，再乘以指数 i，得到实数 $-(\pi/2 + 2\pi k)$。对它取指数，结果便是正实数。所谓“虚数的虚数次方是实数”，关键并非神秘抵消，而是两次乘 i 带来负号。

主值是约定，多值是结构

选择主对数对应 $k = 0$，得到 $\mathrm{e}^{-\pi/2}$。允许所有分支，则得到一族正实值。两种说法都能使用，但必须说明当前约定；不能先采用多值对数，再在计算中悄悄把它当唯一函数。

幂的复合规则也因此需要谨慎。实数正底数下熟悉的某些恒等式，推广后可能因为中间重新选取分支而失败。每次消去指数与对数之前，都应检查是否丢失整圈信息。

请把推导分成三行：选对数值、乘指数、再取指数函数。这样每一步都有明确对象与约定，也能看见 Lean 验证的是其中的代数核心，而不是建立一个没有分支问题的全局复数幂系统。

一步一步建立直观

1. 先为复数选择一个角度　单位虚数 i 位于单位圆的上端，主辐角为 $\pi/2$，但加上任意整数圈仍表示同一个点。右图位置只能显示点本身，不能区分绕了多少圈。复对数必须记录所选辐角，因此同一个非零复数会对应不同的对数值。

2. 把指数定义中的分支写出来　若取主值，$\log i = i\pi/2$，于是 i^i 按 $e^{i \log i}$ 计算，指数变成实数 $-\pi/2$。右图的半圈是几何参照，不是 i^i 的输出位置；本窗口演示辐角与分支，实际幂值需要用指数定义计算。选定分支后，再给出一个确定数值才具有明确含义。

3. 每多记一圈，结果都会改变　取其他对数分支时，$\log i = i(\pi/2 + 2k\pi)$，相应幂值为 $e^{-\pi/2 - 2k\pi}$。它们都是正实数，但彼此不同。因此“它是实数”并不意味着“没有分支选择”；给出一个数值答案前，必须明确采用主值还是多值约定。

> **观课提示**　写清选择的复对数分支，再进行指数运算，追踪整圈角度的信息。
>
> 原课程视频　本节网页与媒体　参数实验

要点回看

核心概念　复数幂通过复对数定义。复对数记录角度，而角度相差整圈时对应相同复数，因此必须指定分支。

整理成数学语言 $\log \mathrm{i} = \mathrm{i}(\pi/2 + 2\pi k)$，$k \in \mathbb{Z}$。于是多值解释下

$$\mathrm{i}^{\mathrm{i}} = \exp\bigl[-(\pi/2 + 2\pi k)\bigr].$$

选择主对数时 $k = 0$，主值为 $\mathrm{e}^{-\pi/2} \approx 0.20788$。

完整算例 关键代数是 $\mathrm{i} \cdot (\mathrm{i}\theta) = -\theta$，使指数变成实数。不同 k 给出不同正实值，而非单个必然唯一的答案。

易错点 不指定对数分支就使用 $(z^a)^b = z^{ab}$ 可能出错；该实数幂习惯不能原样推广。

自测与答案 主值 $\mathrm{i}^{2\mathrm{i}}$ 是多少？$\mathrm{e}^{-\pi}$，使用相同主对数定义。

Lean 对照 `Lockdown.i_log_branch_exponent`

验证任意实角度的关键指数恒等式 $\mathrm{i}(\theta\mathrm{i}) = -\theta$；并未将多值复幂误写成全局单值定理。

5.11 解题：找表示、找不变量、查边界

官方课程：`ldm-tips-to-problem-solving`

先减少未知信息，再增加计算

已知 $x + y = 7, xy = 10$，要求 $x^2 + y^2$，并不一定要先求出两个数。目标对交换 x, y 不变，提示我们改用和与积。展开 $(x + y)^2$ 后立即得到答案 $49 - 20 = 29$。

这不是偶然的巧招，而是表示选择：若问题不关心两个数谁是谁，就没必要保留这份区分。二次方程的中心、矩阵的迹和行列式也在压缩类似的对称信息。

例子、猜想与证明分工不同

做小例子可以发现规律，极端例子可以暴露漏洞，图形可以提示不变量；但它们通常还没有覆盖全部情形。猜想形成后，要把观察变成对任意允许输入有效的推理，并记录用了哪些条件。

例如给定和 s 与积 p，实数解的存在要求 $s^2 - 4p \geq 0$。即使目标表达式能顺利算出，原问题也可能根本没有实数对象。计算简化之后仍需回头检查可行性。

建立自己的解题记录时，可以写三段：原表示哪里困难，新表示保留了什么，如何把结果返回原问题并检查边界。能解释为什么换表示，就更容易把方法迁移到下一道题。Lean 可以核验具体恒等式，但不会替你决定哪个表示最有启发性。

一步一步建立直观

1. 把抽象条件换成可追踪的动作 面对线性方程，先把未知数看作输入向量，把方程右侧看作期望输出。右图让矩阵两列成为基向量的去向，于是求解问题变成“哪一个输入能到达目标”。选择合适表示常常比立即展开计算更重要，因为它能先暴露问题的结构。

2. 检查退化情形，寻找缺失的条件 把两列调到共线，输出平面压成直线。此时某些目标不可达，某些目标又有许多原像，立即说明唯一解需要额外条件。一般公式中的非零分母常常对应这种几何退化；检查它，比在算完后才发现除以零更有效。

3. 用不变量与简单例子回查答案 切到剪切，形状改变但面积倍率为一。若计算结果声称面积翻倍，就可以迅速定位错误。解题结束不只是写出一个表达式，还应代回原条件，核对单位、符号和特殊值，并说明结论在哪些情况下失效。图形提供检验线索，证明负责说明一般性。

> **观课提示** 寻找能减少未知信息的表示，并在得到答案后返回原问题检查可行性。
>
> 原课程视频 本节网页与媒体 参数实验

要点回看

核心概念 遇到困难时可以先做小例子、画图、改变表示，或寻找在变换下不变的量。目标是减少独立信息，而不是堆积运算。

整理成数学语言 对称表达式常改用和 $s = x + y$ 与积 $p = xy$，因为 $x^2 + y^2 = s^2 - 2p$。这种替换也连接到二次方程与矩阵的迹、行列式。

完整算例 已知 $x + y = 7, xy = 10$，不必先求 x, y，就能得到 $x^2 + y^2 = 49 - 20 = 29$。再求根会得到 $2, 5$，作为交叉检查。

易错点 一种表示简化目标后，还需保留可行性条件。例如实数 x, y 要满足 $s^2 - 4p \geq 0$。

自测与答案 已知 $x + y = 3, xy = 1$，求 $x^3 + y^3$。用 $s^3 - 3ps$ 得 $27 - 9 = 18$。

Lean 对照 `Lockdown.symmetric_problem_reduction`

完整验证对称二次式的代数变换。一般解题策略不是数学全称定理，不声称已被形式化证明。

章节综合：用不同语言解同一问题

求复方程 $z^2-2z+2=0$。中心化后是 $(z-1)^2=-1$，所以两根为 $1+\mathrm{i}$ 与 $1-\mathrm{i}$。到这里，二次公式与复数扩展已经合在一起；接着用单位圆语言看它们，模长都是 $\sqrt{2}$，角度分别是正负四分之一的直角。

因此 $(1+\mathrm{i})^4$ 无需展开四次：模长变成四，角度变成 π，结果为负四。用代数平方两次也得到同一结果，提供了独立检查。极坐标适合乘方，直角坐标适合加法，选择表示可以显著减少运算。若继续取对数，主值为 $\frac{1}{2}\ln 2+\mathrm{i}\pi/4$，但所有对数还允许虚部增加 $2\pi k$。从乘方返回对数时必须保留这份角度信息，而不能把实数中的消去规则无条件搬过来。

复习本组时，给每个答案补一句“我现在使用哪一种表示、定义域和分支约定是什么”。这个习惯能把基础公式变成可迁移的推理工具。

第六章　压缩与智能：熵和交叉熵

官网截至目录快照日公开的这组编号课程有两集。本章先用二进制位解释编码，再用自然对数写 Lean 公式；两种单位只相差常数 $\ln 2$。

6.1　熵：平均要问多少个二选一问题

官方课程：`entropy`

怎样把常见消息写短一点

假设要传递三个符号，出现概率为一半、四分之一、四分之一。统一用两位能表达，但较常出现的符号值得更短的编码。选择 $0, 10, 11$，读取时不会把某个完整码当成另一码的开头，因而可无歧义地逐个解码。

平均长度是每个码长乘出现概率再求和，本例为一点五位。这里没有任何一个符号使用“一点五位”；分数来自长期平均。这一区别对理解熵与实际整数码长非常重要。

稀有程度为什么变成负对数

一个公平二选一结果带来一位信息，两个独立结果有四种可能，信息量应相加。概率相乘而信息量相加，提示使用对数；概率越小信息越多，因此取负号。理想长度 $-\log_2 p$ 与这种组合规律一致。

熵把各事件的信息量按真实频率平均。二元实验里，接近确定的结果几乎没有意外，公平的一半一半则平均最难预测，所以曲线在中间最高。改变概率前先预测熵的方向，再看动画与数值。

熵给出概率模型下的平均信息尺度，不是某段文字对人的意义，也不是单个字符串必然能压到的长度。单符号前缀码受整数码长约束，长块编码才可逼近更细的平均值；若序列有相关性，还要利用条件分布。下一节问：若编码使用了错误概率，平均要多付多少代价？

一步一步建立直观

1．把信息量看成排除可能性的成本　若两种结果等可能，一个二选一问题就能区分它们。若其中一种很罕见，常见结果可以在平均意义上花更少的描述成本，罕见结果则需要更长描述。右图

从公平二元分布出发，基准熵为一比特；对数中的底二对应二进制问题的单位。

2. 平均代价要用真实发生概率加权　把真实分布改为不均匀。罕见结果的单次信息量变大，但它出现得少，所以熵要把每种信息量乘以真实概率再相加。不能只看最大码长就判断平均压缩难度。这里显示的是理想平均编码代价，有限样本的整数码长还受具体编码方案影响。

3. 分布的不确定性与模型失配不同　保持真实分布固定，改用偏斜的模型。蓝色曲线升高，橙色熵基线不变，说明额外成本来自模型失配，而不是数据本身突然更随机。熵与交叉熵应分开理解；前者由真实分布决定，后者还依赖你拿来编码或预测的模型。

> **观课提示**　对比确定来源与公平二元来源，区分理想平均长度和单符号整数码长。
>
> 原课程视频　本节网页与媒体　参数实验

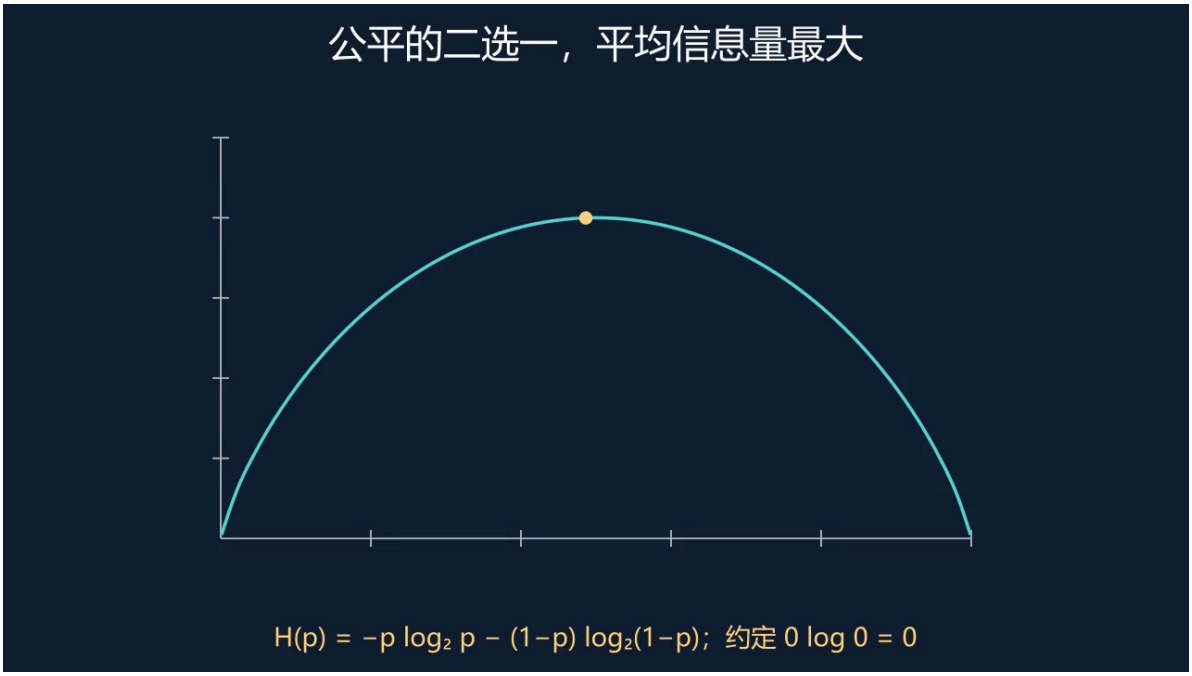


图 6.1: 二元熵在公平概率处达到一位；接近确定事件时下降。本项目原创 Manim 动画关键帧；完整动画与操作入口见本节网页。

要点回看

核心概念　常见事件可以分配短码，罕见事件用长码。编码的平均长度取决于整个概率分布，而不只是符号总数。

整理成数学语言　对有限分布 $p_i \geq 0, \sum_i p_i = 1$，Shannon 熵为

$$H_2(p) = -\sum_i p_i \log_2 p_i, \qquad 0 \log_2 0 := 0.$$

理想信息长度为 $-\log_2 p_i$。独立同分布源的长块编码能让平均每符号长度逼近熵；单符号前缀码的整数长度通常有额外开销。

完整算例 分布 $(1/2, 1/4, 1/4)$ 可用前缀码 $0, 10, 11$，平均长度为 $\frac{1}{2}\cdot 1 + \frac{1}{4}\cdot 2 + \frac{1}{4}\cdot 2 = 1.5$ 位，恰好等于熵。

易错点 熵不是某个特定字符串的长度，也不直接等同于人类语义。对有相关性的序列，要用条件熵或熵率，而不能只看单个符号频率。

自测与答案 四个等概率符号的熵是多少？2 位；一个概率为 1 的确定符号的熵为零。

Lean 对照 `Information.entropy_term_nonnegative`

验证正概率 $p \le 1$ 的单项 $-p\ln p$ 非负。代码还定义了有限分布的熵；编码下界、可达性和熵率并未形式化。

6.2 交叉熵：错误概率模型的编码代价

官方课程：`cross-entropy`

真实出现频率与编码假设可以不同

假设真实来源是公平二元分布，但编码或预测模型认为第一类占四分之三。模型因此给第一类较短的理想长度，给第二类较长的长度；实际使用频率仍由真实分布决定。平均时用真实概率加权模型码长，得到交叉熵。

这解释了公式里为什么同时出现 p 和 q：p 决定现实中哪种事件出现多少次，q 决定每次出现时付多大代价。若把两个位置交换，就换了问题，不能期待相同答案。

把总代价分成不可避免与模型失配

在每项里加减 $\ln p_i$，交叉熵分解为真实熵加 KL 散度。前者是来源自身的不确定性，后者衡量采用模型分布多出的平均对数代价。在两分布归一化且严格为正时，用 $\ln u \le u - 1$ 可证明额外代价非负。

在交互实验中固定真实概率 p，只移动模型概率 q。预测交叉熵在哪里最低，再观察 $q = p$ 时曲线达到真实熵。把 q 推近零或一，会看到模型对实际可能事件过度否定时，代价迅速增加。

分类训练中，单样本真实标签只在正确类别上取一，所以损失变成 $-\ln q_j$。这让“提高正确类概率”与“减少编码代价”连接起来。但训练数据上的低损失不自动保证新数据上的校准或可靠性。数学分解、统计估计与模型泛化仍需分别论证。

一步一步建立直观

1. 先固定数据来源，再评价一种模型 真实分布 p 决定结果出现的频率，模型分布 q 决定各结果被赋予的理想码长。交叉熵因此是按 p 求平均、对 q 取负对数。右图让两者分别可调，避免把“模型认为罕见”和“现实中确实罕见”混为一谈。

2. 错把常见结果当罕见，会付出更多代价 保持真实概率不变，拖动模型点偏离它。某些常出现的结果被分配到更低概率，负对数损失随之变大。蓝线高出熵基线的部分是 KL 散度；在正概率条件下，它非负，模型等于真实分布时为零。这是平均陈述，不保证每个单独样本损失都最小。

3. 更低的最优损失不总说明模型更强 把真实分布改为偏斜分布，并让模型准确匹配。最优交叉熵随真实熵一起下降，因为数据更容易预测，而非模型结构发生了进步。比较不同任务的损失时，应考虑数据分布、单位与分词方式。窗口只验证二元分布关系，不直接评价大型语言模型的能力。

定义、命题与证明

定义：有限分布的交叉熵与 KL 散度 对正概率分布 p, q，定义 $H(p,q) = -\sum_i p_i \log_2 q_i$，$H(p) = -\sum_i p_i \log_2 p_i$，$D(p\|q) = \sum_i p_i \log_2(p_i/q_i)$。

命题：模型失配增加平均编码代价 有 $H(p,q) = H(p) + D(p\|q)$，且 $D(p\|q) \geq 0$；在此正概率情形，等号当且仅当 $p = q$。

证明. 把 $\log_2(p_i/q_i)$ 拆成两个对数并求和，立即得到分解。为证明非负，用 $\ln u \leq u - 1$，代入 $u = q_i/p_i$，得到 $p_i \ln(p_i/q_i) \geq p_i - q_i$。

求和后右侧为零，再除以正数 $\ln 2$ 即得结论。对数不等式仅在 $u = 1$ 时取等，因此所有正权重项同时取等恰好要求 $p_i = q_i$。含零概率时需另外约定边界与无穷代价。

□

> **观课提示** 固定真实分布，只改变模型分布，观察多出的平均编码代价。
>
> 原课程视频 本节网页与媒体 参数实验

要点回看

核心概念 数据真实分布是 p，编码或预测使用的模型是 q。模型给真实常见事件过小的概率，就会在这些事件上支付更长的编码代价。

整理成数学语言 在 $p_i, q_i > 0$ 时，

$$H(p,q) = -\sum_i p_i \ln q_i,$$
$$D_{\mathrm{KL}}(p\|q) = \sum_i p_i \ln \frac{p_i}{q_i},$$
$$H(p,q) = H(p) + D_{\mathrm{KL}}(p\|q).$$

若 p, q 均归一化，则 $D_{\mathrm{KL}} \geq 0$。用 $\ln u \leq u - 1$，对 $u = q_i/p_i$ 逐项乘以 p_i 并求和，即可证明该不等式。

完整算例 真实分布为 $(1/2, 1/2)$，模型为 $(3/4, 1/4)$。以位为单位，真实熵为 1，交叉熵为 $-\frac{1}{2}\log_2(3/4) - \frac{1}{2}\log_2(1/4) \approx 1.20752$，多付约 0.20752 位。

易错点 $D_{\mathrm{KL}}(p\|q)$ 一般不对称，不是距离。若 $p_i > 0$ 而 $q_i = 0$，扩展意义下代价为无穷，不能直接套普通实数对数代数。

自测与答案 正确标签是类别 j 时，单样本交叉熵为何为 $-\ln q_j$？因为 one-hot 标签使其余求和系数均为零。$q_j = 1$ 时损失为零。

Lean 对照 `Information.cross_entropy_decomposition`

完整验证严格正有限向量的熵、交叉熵与 KL 代数分解；此分解本身不要求归一化。同文件的 `kl_nonnegative` 进一步在两分布归一化时证明 KL 非负；零概率与无穷代价情形不在这两个定理的覆盖范围内。

章节综合：预测损失也是一种编码代价

假设样本中第一类占四分之一，用二元模型概率 q 来编码或预测。经验平均交叉熵为

$$C(q) = -\tfrac{1}{4}\log_2 q - \tfrac{3}{4}\log_2(1-q), \qquad 0 < q < 1.$$

求导得到 $C'(q) = (q - \frac{1}{4})/(q(1-q)\ln 2)$。分母为正，所以 $q < 1/4$ 时下降，$q > 1/4$ 时上升，最小点恰在经验频率处。

这说明在当前二元、独立样本模型中，匹配频率会最小化经验编码代价。它没有保证未来样本仍有相同频率，也没有处理参数受复杂网络约束的情况。经验最优与总体最优的区别，需要统计推断补足。

在实验室把 p 固定为 0.25，拖动 q，核对导数符号所预测的升降与最低点。再检查最低值等于真实熵，而多付的部分等于 KL。通过概率、编码长度、优化导数三种语言理解同一问题，便完成了这两课的连接。

来源与复现

原课程由 Grant Sanderson 及其署名合作者创作；扩散模型客座课程署名 Stephen Welch / Welch Labs。本书未附带原视频、原图、字幕全文或原网站正文。所有示意图均为本项目重绘的数学图形。

- 官方目录：https://www.3blue1brown.com/
- 官方频道：https://www.youtube.com/@3blue1brown
- Lean：https://lean-lang.org/
- mathlib：https://github.com/leanprover-community/mathlib4

本项目固定 Lean 4.19.0 和 mathlib 的同版本发布提交。构建命令与验证报告见仓库 README；PDF 是相同 LaTeX 源码的构建产物。更新目录时，先比较新增或删除的课程，再逐项补正文和形式化覆盖说明，不能只修改完成计数。