最优化方法

授课教师：徐海涛

2024年秋

第三章 无约束优化问题

3.4 搜索法求极小值

基本概念：初始化、搜索方向和步长

- 绝大部分数值方法需要确定一个初始点$\boldsymbol{x}_0$，然后计算该点处的搜索方向$\boldsymbol{d}_0$
- 步长α_0按照使得函数f尽可能小的原则确定，下一个迭代点即为$\boldsymbol{x}_1 = \boldsymbol{x}_0 + \alpha_0\boldsymbol{d}_0$
- 以此类推，对$\boldsymbol{x}_1$执行同样的操作，可得到下一个迭代点
- 接下来要讨论的各种数值方法，其区别就体现在搜索方向和步长的选择方式上

3.4 基本概念：初始化、搜索方向和步长

- 对于包括多个局部极小点的非凸问题，利用本章给出的梯度方法只能获得起始点附近的局部极小点。因此，如果问题的规模较小，可以采用网格搜索的方法确定一个比较好的初始点。
- 具体方式为将每个决策变量在上下界内分为$p = (Nd - 1)$个区间，所有决策变量即可形成网格。计算每个网格点对应的函数值f 。
- 当n=2时、可得到p^2个网格点。n个决策变量对应着p^n个网格点。
- 此外，还可以利用试验来选择合适的网格点。最小的目标函数值f对应的网格点就可以作为梯度方法的初始点。

3.4 基本概念：初始化、搜索方向和步长

例 3.5　函数 $f=x_1^2+5x_2^2$，初始点 $\mathbf{x}_0=[3,1]^{\mathrm{T}}$，$f_0\equiv f(\mathbf{x}_0)=14$。

（i）在方向 $\mathbf{d}=[-3,-5]^{\mathrm{T}}$ 上构造函数 $f(\alpha)$，并绘制 f 与 $\alpha,\alpha\geqslant 0$ 的关系曲线。

（ii）求斜率 $\left.\dfrac{\mathrm{d}f(\alpha)}{\mathrm{d}\alpha}\right|_{\alpha=0}$，并证明其与 $\nabla f(\mathbf{x}_0)^{\mathrm{T}}\mathbf{d}$ 相等。

（iii）求 $f(\alpha)$ 的极小点 α_0，这就是初始点对应的步长。据此确定新点 $\mathbf{x}_1$ 及其对应的函数值 $f_1\equiv f(\mathbf{x}_1)$。

（iv）绘制函数的等值线，并标出方向 $\mathbf{d},\mathbf{x}_0$ 和 $\mathbf{x}_1$。

易知 $\mathbf{x}(\alpha)=[3-3\alpha,1-5\alpha]^{\mathrm{T}}$，可得 $f(\alpha)=(3-3\alpha)^2+5(1-5\alpha)^2$。为获得极小点，令导数 $\dfrac{\mathrm{d}f(\alpha)}{\mathrm{d}\alpha}=0$，得 $\alpha_0=0.253\,731$。利用二阶导数可断定其是极小点。故有 $\mathbf{x}_1=\mathbf{x}_0+\alpha_0\mathbf{d}=[2.238\,806,-0.268\,657]^{\mathrm{T}}$，对应的函数值为 $f_1=5.373\,14$，与初始点对应的函数 $f_0=14$ 相比的确有下降。$f(\alpha)$ 的初始斜率（$\alpha=0$ 对应的斜率）为 -68。点 $\mathbf{x}_0$ 处的梯度与搜索方向 $\mathbf{d}$ 之间的点乘 $\nabla f(\mathbf{x}_0)^{\mathrm{T}}\mathbf{d}=(6,10),[-3,-5]^{\mathrm{T}}=-68$。负值表示 $\mathbf{d}$ 是一个下降方向。利用 MATLAB 绘制了 $f(\alpha)$ 的曲线和 $f(x)$ 的等值线，分别如图 E3.5（a）和 E3.5（b）所示。注意点 $\mathbf{x}_1$ 与 $\mathbf{d}$ 相切，步长稍大或者稍小都将导致新点与更大的等值线相交。

3.4 基本概念：初始化、搜索方向和步长

例 3.5　函数 $f=x_1^2+5x_2^2$，初始点 $\mathbf{x}_0=[3,1]^{\mathrm{T}}$，$f_0\equiv f(\mathbf{x}_0)=14$。

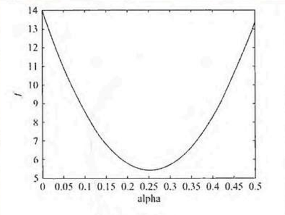

图 E3.5（a）

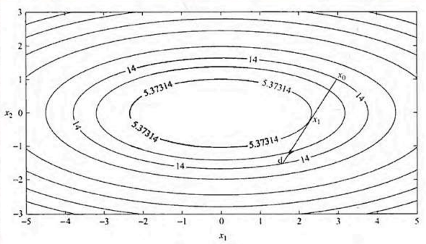

图 E3.5（b）

3.4 基本概念：初始化、搜索方向和步长

例 3.6　考虑如图 E3.6 所示的悬臂梁，其横截面为矩形，需要为其选择合适的宽 w 和高 h，故决策变量为 $\mathbf{x}=[w,h]^{\mathrm{T}}$，初始点选定为 $\mathbf{x}_0=[1,3]^{\mathrm{T}}$（单位：in），试求此时对应的弯曲应力。若初始点对应的搜索方向为 $\mathbf{d}_0=[-1/\sqrt{5},-2/\sqrt{5}]^{\mathrm{T}}$，步长为 $\alpha_0=0.2$，试求更新后的宽和高及其对应的弯曲应力。

当前的弯曲应力为

$$\sigma_0=\frac{6M}{wh^2}=\frac{6\times(2000\times 24)}{1\times 3^2}=32\,000\text{ psi}$$

更新后的设计参数为 $\mathbf{x}_1=\mathbf{x}_0+\alpha_0\mathbf{d}_0$，即

$$\mathbf{x}^1=\begin{pmatrix}1\\3\end{pmatrix}+0.2\begin{pmatrix}-1/\sqrt{5}\\-2/\sqrt{5}\end{pmatrix}=\begin{pmatrix}0.910\,56\\2.821\,11\end{pmatrix}$$

对应的弯曲应力为 $\sigma_1=71\,342$ psi。

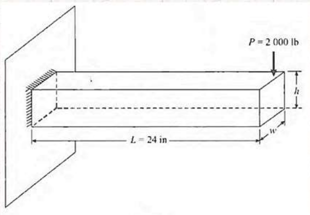

图 E3.6

3.5 最速下降法

从最速下降法开始介绍n元函数无约束极小点的数值求解方法

搜索方向

令 $\mathbf{x}_k$ 表示第 k 次搜索对应的迭代点，$k=0$ 对应的是初始点。要做的是确定一个合适的下降方向 $\mathbf{d}$ 以及对应的搜索步长 $\alpha>0$，使得新点 $\mathbf{x}_k+\alpha\mathbf{d}$ "更好"，即满足 $f(\mathbf{x}_k+\alpha\mathbf{d})<f(\mathbf{x}_k)$。为此，将函数 f 在点 $\mathbf{x}_k$ 处进行泰勒展开：

$$f(\mathbf{x}_k+\alpha\mathbf{d})=f(\mathbf{x}_k)+\alpha\nabla f(\mathbf{x}_k)^{\mathrm{T}}\mathbf{d}+\mathbf{O}(\alpha^2)$$

因此，函数值的波动 $\delta f=f(\mathbf{x}_k+\alpha\mathbf{d})-f(\mathbf{x}_k)$ 可表示为

$$\delta f=\alpha\nabla f(\mathbf{x}_k)^{\mathrm{T}}\mathbf{d}+O(\alpha^2)$$

3.5 最速下降法（搜索方向）

因此，函数值的波动 $\delta f = f(\mathbf{x}_k + \alpha\mathbf{d}) - f(\mathbf{x}_k)$ 可表示为

$$\delta f = \alpha \nabla f(\mathbf{x}_k)^{\mathrm{T}}\mathbf{d} + O(\alpha^2)$$

当 α 足够小时，$O(\alpha^2)$ 被 $\alpha \nabla f(\mathbf{x}_k)^{\mathrm{T}}\mathbf{d}$ 的高阶无穷小量，故有

$$\delta f \cong \alpha \nabla f(\mathbf{x}_k)^{\mathrm{T}}\mathbf{d}$$

要求函数值下降，即 $\delta f < 0$，这意味着 $\mathbf{d}$ 应该满足

$$\nabla f(\mathbf{x}_k)^{\mathrm{T}}\mathbf{d} < 0$$

此时，$\mathbf{d}$ 称为下降方向。

在最速下降法中，第 k 步搜索对应的搜索方向 $\mathbf{d}_k$ 按照如下方式选择：

$$\mathbf{d}_k = -\nabla f(\mathbf{x}_k)$$

有 $\nabla f(\mathbf{x}_k)^{\mathrm{T}}\mathbf{d}_k = -\|\nabla f(\mathbf{x}_k)\|^2 < 0$，因此，该方向能够满足式（3.10），称为最速下降方向。

3.5 梯度向量的几何意义

在二维空间，即 $x_1 - x_2$ 空间中，梯度向量 $\nabla f(\mathbf{x}_k)$ 与目标函数经过点 $\mathbf{x}_k$ 的等值线正交。更为一般的说法是在点 $\mathbf{x}_k$ 处构造一个与目标函数超平面相切的超平面，那么该点处的梯度向量与该超平面正交。为了证明这一结论，从点 $\mathbf{x}_k$ 出发，在目标函数的常值超平面任意构造一条曲线，令 s 表示曲线的弧长，这条曲线上的点都可以写为 $\mathbf{x} = \mathbf{x}(s)$，且 $\mathbf{x}(0) \equiv \mathbf{x}_k$，由于 $f(\mathbf{x}(s))$ 为常数，故

$$\frac{\mathrm{d}}{\mathrm{d}s} f(x(s))|_{s=0} = \nabla f(\mathbf{x}_k)^{\mathrm{T}}\dot{\mathbf{x}}(0) = 0$$

其中，$\dot{\mathbf{x}}(0) = \lim_{s\to 0}\frac{\mathbf{x}(s) - \mathbf{x}(0)}{s}$。由于 $\mathbf{x}(s) - \mathbf{x}(0)$ 表示连接 $\mathbf{x}_k$ 与 $\mathbf{x}(s)$ 的弦，当 $s \to 0$ 时，弦的极限为一个与目标函数常值超平面相切的向量。因此，$\dot{\mathbf{x}}(0)$ 就是切平面上的一个向量。由上式可知，$\nabla f(\mathbf{x}_k)$ 与该向量正交。由于这是针对任意一条从点 $\mathbf{x}_k$ 出发的曲线得出的结论，因此，可得 $\nabla f(\mathbf{x}_k)$ 必定与切平面正交，如图 3.3 所示。

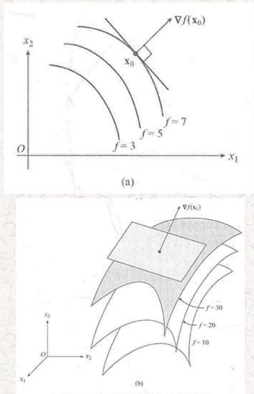

图3.3 二维和三维空间中梯度向量的几何意义

3.5 最速下降法（梯度向量例题）

例 3.7 函数为 $f = x_1x_2^2$，初始点为 $\mathbf{x}_0 = [1,2]^{\mathrm{T}}$。

（i）确定点 $\mathbf{x}_0$ 处的最速下降方向；

（ii）$\mathbf{d} = [-1,2]^{\mathrm{T}}$ 是不是下降方向？

函数梯度为 $\nabla f = [x_2^2, 2x_1x_2]^{\mathrm{T}}$。点 $\mathbf{x}_0$ 处的最速下降方向为梯度的负方向，即 $[-4, -4]^{\mathrm{T}}$。如式（3.10）所示，为了判断 $\mathbf{d} = [-1,2]^{\mathrm{T}}$ 是否是下降方向，需要计算 $\mathbf{d}$ 与梯度之间的点乘，有 $\nabla f(\mathbf{x}_0)^{\mathrm{T}}\mathbf{d} = (4,4)[-1,2]^{\mathrm{T}} = +4 > 0$，故 $\mathbf{d}$ 不是下降方向。

3.5 最速下降法（线性搜索）

- α_k 的意义：搜索步长
- 确定了点 x_k 处的下降方向 $\boldsymbol{d}_k$ 之后，接下来的问题就是确定沿着下降方向应该“前行多远”，即确定搜索的步长 $\boldsymbol{\alpha}_k$ 。
- 确定步长的过程实际上是一个一维极小化问题，因为只包括一个变量 $\boldsymbol{\alpha}$。
- 具体而言，就是在方向 $\boldsymbol{d}_k$ 上，决策变量和目标函数都只与 α 相关

$$\boldsymbol{x}_k + \boldsymbol{\alpha}\boldsymbol{d}_k \equiv \boldsymbol{x}(\boldsymbol{\alpha})$$

$$f(\boldsymbol{x}_k + \boldsymbol{\alpha}\boldsymbol{d}_k) \equiv f(\boldsymbol{\alpha})$$

3.5 最速下降法（线性搜索）

- 如何评价$\boldsymbol{d}_k$方向：函数$f(\alpha)$ 的斜率，或者导数$f'(\alpha) = df/d\alpha$ 称为沿着方向$\boldsymbol{d}$ 的方向导数

$$\frac{\mathrm{d}f(\hat{\alpha})}{\mathrm{d}\alpha} = \nabla f(\mathbf{x}_k + \hat{\alpha}\,\mathbf{d}_k)^{\mathrm{T}}\,\mathbf{d}_k$$

- 可以推导出$f(\alpha)$ 在$\alpha = 0$ 处的斜率为

$$\frac{\mathrm{d}f(0)}{\mathrm{d}\alpha} = [\nabla f(\mathbf{x}_k)]^{\mathrm{T}}\,[-\nabla f(\mathbf{x}_k)] = -\|\nabla f(\mathbf{x}_k)\|^2 < 0$$

- 在$\boldsymbol{x}_k$点沿着$\boldsymbol{d}_k$方向，斜率为负数，这意味着$-\nabla f(\boldsymbol{x}_k)$是一个下降方向

3.5 最速下降法（精确线性搜索）

- 构造并求解优化问题

$$\underset{\alpha\geq 0}{\text{minimize}}\;\; f(\alpha) \equiv f(\mathbf{x}_k + \alpha\,\mathbf{d}_k)$$

- 以获取步长。令α_k表示该问题的最优解。这种方式能够使得函数f沿着方向$\boldsymbol{d}_k$达到最小，称为精确线性搜索。

3.5 最速下降法（精确线性搜索）

在开展搜索之前，首先要获取“三点模式”。一旦确定了极小点所在的区间之后，就可以利用第 2 章中的二阶多项式拟合方法或其他方法求解极小点。考虑$f(\alpha)$的函数曲线，如图 3.4（a）所示，前面已经提及，$f(0) \equiv f(\mathbf{x}_k)$，由于$f(\alpha)$在$\alpha = 0$ 处的斜率为负值，因此，当α稍微增加一点，函数值就会递减，故可选择初始步长$\alpha_1 \equiv s$。如果$f(\alpha_1) < f(0)$说明方向正确，继续前进得到一个新点$\alpha_{i+1} = \alpha_i + \frac{s}{\tau^i}, i = 1, 2, \cdots$，其中，$\tau = 0.618\,034$ 为黄金分割率。持续执行这一过程，直到函数值开始增加，得到三个点满足$f(\alpha_{i+1}) \geqslant f(\alpha_i) < f(\alpha_{i-1})$，区间$[\alpha_{i-1}, \alpha_i, \alpha_{i+1}]$就是三点模式。

3.5 最速下降法（精确线性搜索）

反之，如果初始步长对应的函数值$f(\alpha_1) > f(0)$，如图 3.4(b)所示，说明极小点就位于$[0, \alpha_1]$中，按照$\alpha_{i+1} = \tau\alpha_i, i = 1, 2, \cdots$的方式构造新点，直到产生两个点满足$f(\alpha_i) > f(\alpha_{i+1}) \leqslant f(0)$，区间$[0, \alpha_{i+1}, \alpha_i]$就是三点模式。

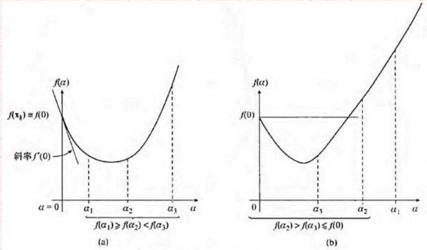

图 3.4 三点模式的获取

3.5 最速下降法（精确线性搜索）

- 一旦得到了三点模式，就可以利用二阶多项式拟合方法或其他方法（见第2章）来获取极小点α_k。
- 采用极小化问题式(3. 16) 的方式来开展精确线性搜索，可能需要求解很多次函数值。
- 采用”下降足够多”或Armijo 条件作为搜索的终止条件有时会更加高效，且能够极大地简化程序代码。具体细节将在3.9 节进行讨论，以免在此造成误导
- 本书中有关的程序代码使用的都是精确线性搜索策略，对于一般性的问题，这是最为稳健的方法。

3.5 最速下降法（停止规则）

- 在开展线性搜索之前，首先判断最优性条件是否满足：

$$\|\nabla f(\mathbf{x}_k)\| \leqslant \varepsilon_G$$

- 其中，ε_G是预先设定的梯度精度。如果上式所示的条件得到满足，迭代过程停止
- 需要注意的是，当接近局部极大点时，梯度同样会逐渐接近于0。但是，由于采用的是逐步下降策略而不是求根策略求极小点，即每次迭代函数值f都在下降，因此，收敛到局部极大点的可能性非常小

3.5 最速下降法（停止规则）

- 可以利用相邻两次迭代的函数值f之差作为停止规则

$$|f(\mathbf{x}_{k+1}) - f(\mathbf{x}_k)| \leqslant \varepsilon_A + \varepsilon_R |f(\mathbf{x}_k)|$$

- 判断上述条件是否满足，其中，ε_A表示函数值之差的绝对精度，ε_R为函数值之差的相对精度
- 此外，迭代次数也可以作为停止规则

3.5 最速下降法（算法）

1. 选择初始点 $\mathbf{x}_0$ 和精度参数 $\varepsilon_G, \varepsilon_A, \varepsilon_R$，设定迭代次数 $k=0$；
2. 计算$\nabla f(\mathbf{x}_k)$，如果 $\|\nabla f(\mathbf{x}_k)\| \leqslant \varepsilon_G$，算法停止；否则，设计归一化的搜索方向 $\mathbf{d}_k = -\nabla f(\mathbf{x}_k) / \|\nabla f(\mathbf{x}_k)\|$；
3. 利用精确或近似的线性搜索算法计算步长 α_k，更新迭代点 $\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha_k \mathbf{d}_k$；
4. 计算函数值$f(\mathbf{x}_{k+1})$，如果式（3. 18）满足，则算法停止；否则，令 $k=k+1$，返回第2步。

3.5 最速下降法（收敛特性）

最速下降法在向最优点迭代的路径呈现出“锯齿状”，这是因为相邻两次迭代的搜索方向正交。通过极小化$f(\mathbf{x}_k+\alpha\mathbf{d}_k)$来获取步长$\alpha_k$，就一定能够导致这一问题。具体而言，在$\alpha=\alpha_k$处，$\mathrm{d}/\mathrm{d}\alpha\{f(\mathbf{x}_k+\alpha\mathbf{d}_k)\}=0$，对其展开，可得$\nabla f(\mathbf{x}_k+\alpha_k\mathbf{d}_k)^{\mathrm{T}}\mathbf{d}_k=0$或$\nabla f(\mathbf{x}_{k+1})^{\mathrm{T}}\mathbf{d}_k=0$。由于$\mathbf{d}_{k+1}=-\nabla f(\mathbf{x}_{k+1})$可得$\mathbf{d}_{k+1}^{\mathrm{T}}\mathbf{d}_k=0$。注意到，只有采用完全精确线性搜索时，这一结论才成立。

明理，精工，笃学，致远 21

3.5 最速下降法（收敛特性）

$$f=x_1^2+ax_2^2$$

- 利用最速下降法求该函数的极小点，迭代路径如图3.5 所示
- α越大，f的等值线就越细长，迭代收敛速度就越慢
- 该方法的收敛速度与函数f黑塞矩阵的谱条件数有关
- 谱条件数κ定义为最大特征值与最小特征值之间的比值

$$\kappa=\frac{\lambda_{\max}}{\lambda_{\min}}$$

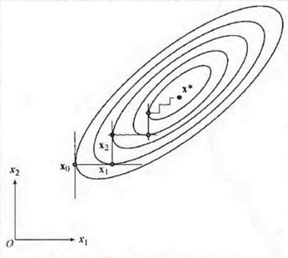

图 3.5 最速下降法的锯齿状迭代过程

明理，精工，笃学，致远 22

3.5 最速下降法（收敛特性）

- 函数的黑塞矩阵为 $\nabla^2 f=\begin{bmatrix}2 & 0\\ 0 & 2a\end{bmatrix}$

$$f=x_1^2+ax_2^2$$

- 条件数为$\kappa=a$。当$a=1$时，收敛速度最快。
- 对于非二次型函数，收敛速度与黑塞矩阵在最优点x^*处的条件数有关

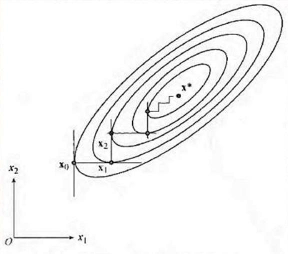

图 3.5 最速下降法的锯齿状迭代过程

明理，精工，笃学，致远 23

3.5 最速下降法（收敛特性）

- 在常态下，该算法对于任意初始点都是收敛的，可称为“全局收敛 ”，此处的全局并不是指能够收敛到全局极小点，该算法只能收敛到局部极小点
- 该算法具有线性收敛速度。如果最优点处的黑塞矩阵是正定的，谱条件数为κ，则迭代过程满足

$$f_{k+1}-f^*\approx\frac{(\kappa-1)^2}{(\kappa+1)^2}[f_k-f^*] \Rightarrow e_{k+1}\le Ce_k^r$$

- 最速下降法能够通过几次迭代就从距离最优点非常远的起始点收敛到最优点附近，即 “ 最优点区域”；但在最优点区域，可能需要经过几百次迭代才能够取得非常小的进展

明理，精工，笃学，致远 24

3.5 最速下降法（尺度变换）

- 如果对决策变量进行适当的尺度变换，可以使得黑塞矩阵变成良态
- 定义两个新变量y_1和y_2 $\quad y_1 = x_1,\ y_2 = \sqrt{a}x_2 \quad g = y_2^2 + y_2^2$

考虑函数$f(\mathbf{x})$，按照下式对变量进行尺度变换：

$$\mathbf{x} = \mathbf{T}\,\mathbf{y}$$

可得新函数为$g(\mathbf{y}) \equiv f(\mathbf{T}\,\mathbf{y})$，其梯度和黑塞矩阵为

$$\nabla g = \mathbf{T}^{\mathrm{T}} \nabla f$$

$$\nabla^2 g = \mathbf{T}^{\mathrm{T}} \nabla^2 f\,\mathbf{T}$$

通常，$\mathbf{T}$为对角矩阵。通过选择合适的$\mathbf{T}$，使得g的黑塞矩阵的条件数接近于1。

明理，精工，笃学，致远 25

3.5 最速下降法（例题）

例 3.8　函数为$f = (x_1 - 2)^4 + (x_1 - 2x_2)^2, \mathbf{x}_0 = [0,3]^{\mathrm{T}}$，利用最速下降法开展一次迭代。

有$f(\mathbf{x}_0) = 52, \nabla f(\mathbf{x}) = [4(x_1-2)^3 + 2(x_1 - 2x_2), -4(x_1 - 2x_2)]^{\mathrm{T}}$，因此，搜索方向为$\mathbf{d}_0 = -\nabla f(\mathbf{x}_0) = [44, -24]^{\mathrm{T}}$。将其归一化处理为一个单位向量$\mathbf{d}_0 = [0.8779, -0.4789]^{\mathrm{T}}$。利用线性搜索，可得函数$f(\alpha) = f(\mathbf{x}_0 + \alpha\mathbf{d}_0)$的极小点为$\alpha_0 = 3.0841$。因此，新点为$\mathbf{x}_1 = \mathbf{x}_0 + \alpha_0\mathbf{d}_0 = [2.707, 1.523]^{\mathrm{T}}$，对应的函数值为$f(\mathbf{x}_1) = 0.365$。计算该点的梯度$\nabla f(\mathbf{x}_{k+1})$，可开展新一轮的迭代。

明理，精工，笃学，致远 26

本节结束，谢谢

Thank you for listening

课程名称：最优化方法

明理，精工，笃学，致远

习题

作业题：P3.1 P3.3 P3.4 P3.10

思考题：P3.5 P3.6

明理，精工，笃学，致远 28