文再文老师的最优化课讲的很不错。

最小二乘的思想非常直观：若方程 (3.1.1) 存在解，则求解问题 (3.1.2) 的全局最优解就相当于求出了方程的解；当方程的解不存在时，问题(3.1.2) 实际给出了某种程度上误差最小的解. 读者可能注意到，最小二乘法使用了 ℓ_2 范数来度量误差大小，其主要优点有两个：第一，ℓ_2 范数平方是光滑可微的，它会给目标函数带来较好的性质；第二，ℓ_2 范数对于某种误差的处理有最优性，这一点我们在第 3.2 节中会进一步给出解答.

最小二乘并不总是最合理的. 除了构造最小二乘问题外，根据实际问题

最大似然估计则是极大化对数似然函数，去除掉常数项之后我们得到了如下最小二乘问题：

$$\min_{x\in\mathbb{R}^n} \quad \frac{1}{2}\|Ax-b\|_2^2. \tag{3.2.4}$$

注意，在构建最大似然估计时不需要知道 ε_i 的方差 σ^2. 以上变形的最重要的意义在于它建立了最小二乘法和回归分析的联系：当假设误差是高斯白噪声时，最小二乘解就是线性回归模型的最大似然解.

当 ε_i 不是高斯白噪声时，求解线性回归模型 (3.2.3) 和最小二乘模型 (3.2.4) 并不等价，因此我们需要借助似然函数来构造其他噪声所对应的目标函数. 例如，在某些噪声下构造出的模型实际上为最小一乘问题：

$$\min_{x\in\mathbb{R}^n} \quad \|Ax-b\|_1. \tag{3.2.5}$$

在习题 **3.7** 中也给出了在其他噪声假设下相应的目标函数.