OPTIMIZED METHODS IN COMMUNICATION ENGINEERING

通信工程中的最优化方法

郑宝玉 糜正琨 王良元 编著

欧阳琨 主审

北京邮电大学出版社

通信工程中的最优化方法

北京邮电大学出版社

通信工程中的最优化方法

郑宝玉　糜正琨　王良元　编著

欧阳琨　主审

北京邮电大学出版社

图书在版编目(CIP)数据

通信工程中的最优化方法/郑宝玉,糜正琨,王良元编著. 一北京:北京邮电大学出版社,1996.1

ISBN 7-5635-0204-1

Ⅰ.通… Ⅱ.①郑… ②糜… ③王… Ⅲ.通信技术-最优化算法 Ⅳ.TN911.1

中国版本图书馆 CIP 数据核字(96)第 18219 号

内 容 提 要

本书主要阐述最优化方法的原理和算法及其在通信工程中的应用。全书共分八章,内容分别为最优化问题的数学模型和最优解的数学基础、无约束优化问题的求解方法、求解线性规划的单纯形算法和卡玛卡算法、非线性规划的求解策略和主要算法、多目标规划和动态规划的求解原理和方法,最后一章综合了各种算法,较为详细地分析和介绍最优化方法在通信工程中的七类典型应用实例。

通信工程中的最优化方法

编　　著　郑宝玉　糜正琨　王良元

主　　审　欧阳琨

责任编辑　王守平

北京邮电大学出版社出版

新华书店北京发行所发行　各地新华书店经售

通县向阳印刷厂印刷

*

787×1092 毫米　1/16　印张 20.625　字数 509 千字

1996 年 1 月第一版　1996 年 1 月第一次印刷

印数:1—2000 册

ISBN 7-5635-0204-1/TN·75　定价:22.50 元

前　言

随着通信技术的日益复杂和计算机科学的不断发展，以最优化方法为基础的计算机辅助设计已经日益广泛地应用于通信工程设计。为了能向高校师生和科技人员提供一本具有通信专业特色、面向工程实际的较为系统的最优化方法著作，笔者结合多年的科研成果编就此书。

本书力求以简炼的方法阐明最优化理论的基本概念，尽可能避免复杂的数学推导。全书以实用算法为主线，全面介绍了通信工程中常用的最优化技术，并在保证系统性的前提下，适当加强了在理论上和实际上都具有重要意义的有约束最优化的篇幅。我们在每一章都编有典型设计实例，除此以外，本书还以一整章的篇幅介绍了最优化在通信工程领域中的一些应用实例，给出各个应用的技术背景、数学模型、算法选择和优化过程，期望藉此加强最优化理论和实践的统一性，提高读者对最优化方法的应用能力。

本书作为通信专业研究生教材，推荐学时为 60 课时，也可以无约束优化部分为主体，作为本科生的选修课教材，推荐学时为 32 学时，同时还可作为通信专业科研设计人员的技术参考书。

本书第二章、第三章、第四章和第八章的三、四、五、六小节由郑宝玉编写，第一章、第五章、第六章、第七章的第一小节和第八章的一、二小节由糜正琨编写，第七章的第二小节和第八章的第七小节由王良元编写。全书由欧阳珉教授主审。

由于编者水平有限，书中难免有不妥之处，衷心希望广大读者批评指正。

编者

一九九四年七月

目　　录

第四章 吕美兹(Remez)算法

第五章 线性规划

第六章 非线性规划

第七章 多目标规划和动态规划

第八章 最优化方法在通信工程中的应用

第一章　最优化问题和最优性条件

本章首先从具体实例出发，引入最优化问题的概念及其数学模型，然后从数学角度论证各类问题最优解的必要条件和充分条件，最后给出最优化算法的一般特征和构造原则。

1.1　最优化问题

1.1.1　最优化问题的数学模型

什么是最优化问题？对于通信工程来说，是指最优工程设计问题，即在给定设计指标和元件、参数的允许取值范围条件下，确定一组独立的设计参数，使系统达到最佳技术经济性能。系统性能的优劣通常用一个关于设计参数的函数来描述，该函数即称为“目标函数”，待定的设计参数称为“优化变量”，而参数范围和未包含在目标函数中的一些设计指标即构成优化变量的“约束条件”。寻求系统的最佳性能，在数学上通常就是最小化或最大化目标函数。例如最小误差、最低成本或最大增益、最高利润等。因此，最优化问题的数学模型可表示为：

$$
\begin{aligned}
&\min_{X\in R^n}\ f(\boldsymbol{X}) \\
\text{subject to}\quad & C_i(\boldsymbol{X}) = 0 \quad i = 1,2,\cdots,m \\
& C_i(\boldsymbol{X}) \leqslant 0 \quad i = m+1,\cdots,m+p
\end{aligned}
\tag{1.1}
$$

式中，$\boldsymbol{X} = [x_1, x_2, \cdots, x_n]$：优化变量

$f(\boldsymbol{X})$：目标函数

$C_i(\boldsymbol{X})$：约束函数

并记，$\boldsymbol{X}^*$、$f(\boldsymbol{X}^*)$ 为最优解和最优目标函数值。

由于 $\max f(\boldsymbol{X}) \Leftrightarrow \min[-f((\boldsymbol{X})]$，$C_i(\boldsymbol{X}) \geqslant 0 \Leftrightarrow -C_i(\boldsymbol{X}) \leqslant 0$，不难看出(1.1)式是适用于任何优化形式、任何约束条件的最优化问题的一般数学描述。此外，还应该指出，上述分析不仅适用于最优化工程设计，也适用于应用科学、数学、医学、经济学、统计学等所有领域。这些领域中的数据分析、优选、决策等都可以归结为不同形式的最优化问题，并由其相应的目标函数和约束条件来描述。

因 此，(1.1)式即为本书采用的最优化问题的标准形式。下面考察几个通信工程中的实例，藉此进一步理解最优化问题的含义及其数学模型的构造。

1.1.2 通信工程最优化问题实例

一、曲线拟合问题

通信工程中经常会遇到曲线拟合问题。例如给定线路的频率特性为 $F(\omega)$，要求在给定的频带$[\omega_L, \omega_H]$上设计一个接近于 $F(\omega)$ 特性的仿真线。这里，$F(\omega)$ 可以是衰减特性，也可以是

群时延特性。记仿真线的元件参数为 $x_1, x_2, \cdots, x_n$，依照网络理论，可以求出仿真线的实际频率特性为 $F_a(\omega, x_1, \cdots, x_n) \triangleq F_a(\omega, \boldsymbol{X})$。故设计目的就是调整 $\boldsymbol{X} = [x_1, x_2, \cdots, x_n]^T$ 各分量的值，使得在整个频带 $[\omega_L, \omega_H]$ 上，$F_a(\omega, \boldsymbol{X})$ 与 $F(\omega)$ 的偏差为最小，即曲线 $F_a(\omega, \boldsymbol{X})$ 与 $F(\omega)$ 相拟合。

为了衡量偏差的大小，我们首先定义误差函数：

$$e(\omega, \boldsymbol{X}) \triangleq F(\omega) - F_a(\omega, \boldsymbol{X}) \tag{1.2}$$

它表征了在频率 ω 处，实际特性和给定特性间的绝对误差。

若以适当方式定义一个表征在整个频带上 $F_a - F$ 拟合误差的函数 $E(\boldsymbol{X}) = f[e(\omega_i, \boldsymbol{X})]$，$\omega_i \in [\omega_L, \omega_H]$，则上述设计可归结为如下的最优化问题：

$$\min_{\boldsymbol{X}} \quad E(\boldsymbol{X}) \tag{1.3}$$

根据网络设计中不同的误差判别准则，可以定义不同的 $E(\boldsymbol{X})$，从而构造出不同的最优化目标函数。

1. 极大 — 极小误差准则

定义：

$$E_\infty(\boldsymbol{X}) \triangleq \max_{\omega \in [\omega_L, \omega_H]} |e(\omega, \boldsymbol{X})| \tag{1.4}$$

这时，上述设计问题即为：

$$\min \max_{\omega \in [\omega_L, \omega_H]} |e(\omega, \boldsymbol{X})| \tag{1.5}$$

目标函数(1.4)式通常称为 min-max 目标函数。一般，我们在 $[\omega_L, \omega_H]$ 中取一些离散样点 ω_1，$\omega_2, \cdots, \omega_m$，以下述形式近似 $E_\infty(\boldsymbol{X})$：

$$E_{\infty i}(\boldsymbol{X}) \triangleq \max_{\omega_L \leqslant \omega_i \leqslant \omega_H} |e(\omega_i, \boldsymbol{X})| \tag{1.6}$$

为了使 $E_{\infty i}(\boldsymbol{X})$ 尽可能和 $E_\infty(\boldsymbol{X})$ 接近，频率取样点数 m 应足够大。此外在 $F(\omega)$ 变化剧烈的区间取样点分布应稠密，而 $F(\omega)$ 平坦的区间取样点可以较少。

设问题(1.5)式的最优解点为 $\boldsymbol{X}^*$，则称函数 $Fa(\omega, \boldsymbol{X}^*)$ 最优一致逼近于 $F(\omega)$，也称作以契比雪夫方式逼近于 $F(\omega)$。

2. 最小二乘方误差准则

定义：

$$E_2 \triangleq \int_{\omega_L}^{\omega_H} |e(\omega, \boldsymbol{X})|^2 \mathrm{d}\omega \tag{1.7}$$

显然，当且仅当 $E_2 = 0$ 时，$F_a(\omega, \boldsymbol{X})$ 与 $F(\omega)$ 在 $[\omega_L, \omega_H]$ 区间内点点相同。这时，网络设计归结为以 E_2 为目标函数的最优化问题，称作最小二乘问题。

同样，可在 $[\omega_L, \omega_H]$ 区间取 m 个样点频率 $\omega_1, \omega_2, \cdots, \omega_m$，然后取各个样点的误差函数平方和，作为最小二乘方拟合误差函数的定义，即：

$$E_2 \triangleq \sum_{i=1}^{m} [e(\omega_i, \boldsymbol{X})]^2 \qquad \omega_L \leqslant \omega_i \leqslant \omega_H \tag{1.8}$$

(1.8)式可以进一步推广到更一般的情况：

$$E_p \triangleq \sum_{i=1}^{m} [e(\omega_i, \boldsymbol{X})]^p \qquad \omega_L \leqslant \omega_i \leqslant \omega_H \tag{1.9}$$

式中，p 为偶数，一般取为 $4 \sim 10$。E_p 称之为最小 p 乘方目标函数。可以证明，当 $p \to \infty$ 时，E_p 就是 min-max 目标函数。

一般来说，最小二乘方问题比 min max 问题容易求解，有比较成熟而有效的算法。但当 $F(\omega)$ 曲线含有间断点时，用 E_∞ 作为目标函数，其拟合效果要好一些。对于网络设计来说，在截止频率附近，网络特性从阻带过渡到通带，变化往往比较剧烈，因此在通信工程中，契比雪夫逼近应用比较广泛，我们将在第四章对其算法作专门的介绍。

3. 加权误差函数

工程实际往往对 $[\omega_L, \omega_H]$ 频带中不同的频率点有不同的误差要求。对一些重要频率点，允许其误差小，即在这些频率点要求实际特性充分接近理想特性。为此，我们定义加权误差函数为：

$$e_w(\omega, X) \triangleq W(\omega) \cdot e(\omega, X) \tag{1.10}$$

其中，$W(\omega)$ 称为权函数，它是一个取值为正的有限函数。$W(\omega)$ 随频率的变化，反映了不同频段的指标要求。

例如，图 1.1 示出在 0 ～ 0.5 范围内数字带通滤波器的幅频特性要求，其通带为 0.2 ～ 0.35，下阻带为 0 ～ 0.1，上阻带为 0.425 ～ 0.5，其余为过渡带。通带的理想幅度值为 1，阻带的理想幅度值为 0。给定的待设计滤波器在通带、上下阻带的幅度容差分别为 δ_1、δ_2 和 δ_3。显然，允许容差越小，表示对该频段的特性要求越严格，因此在该频段的频率点上的误差函数在目标函数中占的比重应该比较大。据此，权函数的一种简单的取法是：

$$W(\omega) = \begin{cases} 1/\delta_2 & \omega \in \text{下阻带} \\ 1/\delta_1 & \omega \in \text{通带} \\ 1/\delta_3 & \omega \in \text{上阻带} \end{cases} \tag{1.11}$$

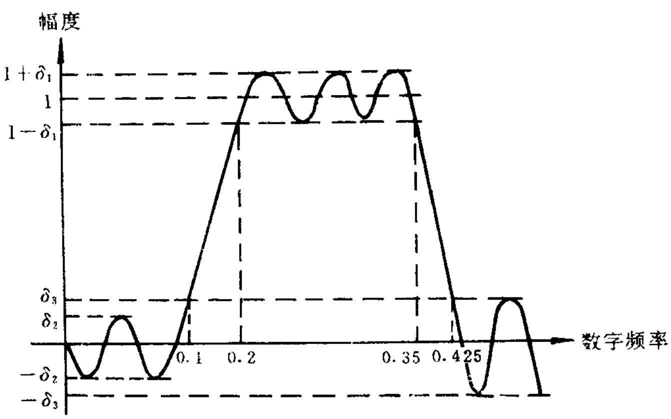

图 1.1

以(1.10)式定义的加权误差函数 $e_W(\omega, X)$ 代替(1.8)、(1.6)式中的 $e(\omega, X)$，则得相应加权后的最小二乘方和 min max 目标函数。

4. 复合目标函数

上述目标函数都是针对某一个指标要求而建立起来的，实际设计中常常可能要求系统同时满足两个或多个特性指标。这时，我们可以将关于各个特性指标的目标函数加权组合后构成一个复合目标函数：

$$E = W_1 E_1 + W_2 E_2 + \cdots + W_p E_p \tag{1.12}$$

式中，E_i：关于第 i 个指标的目标函数；

W_i：第 i 个权因子。

E_i 可以是根据任意一种误差准则建立起来的、加权或不加权的目标函数。调整 W_i 值即可改变

各个指标要求在整体设计中所占的比重，对于重要的指标，赋予大的权因子。这样，最小化复合目标函数 E 即可获得兼顾各项指标要求的折衷的最优设计。

例如，要求设计一个滤波器，使它的幅度特性与群时延特性分别在给定容差范围内逼近给定的幅度特性 $F_1(\omega)$ 与群时延特性 $F_2(\omega)$。设采用最小二乘方误差准则，则当只考虑幅度特性时，目标函数为：

$$E_1 = \sum_{i=1}^{m}[W_1(\omega_i)(F_1(\omega_i) - F_{a_1}(\omega_i, \boldsymbol{X})]^2 \tag{1.13}$$

其中，F_{a_1} 为待设计滤波器在各个频率点 ω_i 处的实际幅度特性。从最小化 E_1，求得最优点及最优目标函数值为 $\boldsymbol{X}_1^*$ 和 $E_1^* = E_1(\boldsymbol{X}_1^*)$。

当只考虑群时延特性时，目标函数为：

$$E_2 = \sum_{i=1}^{m}[W_1(\omega_i)(F_2(\omega_i) - F_{a2}(\omega_i, \boldsymbol{X}))]^2 \tag{1.14}$$

其中，F_{a2} 为待设计滤波器的实际群时延特性。从小化 E_2，求得最优点 $\boldsymbol{X}_2^*$，相应的目标函数值为 $E_2^* = E_2(\boldsymbol{X}_2^*)$。

当同时考虑幅度特性和群时延特性时，一般不可能使 E_1、E_2 同时达到 E_1^*、E_2^*，而且当 $E_1^* \ll E_1(\boldsymbol{X}_2^*)$，$E_2^* \ll E_2(\boldsymbol{X}_1^*)$ 时，无论是 $\boldsymbol{X}_1^*$ 还是 $\boldsymbol{X}_2^*$ 都不能作为同时考虑幅度特性和群时延特性的最优设计点。如果取 $\boldsymbol{X} = \boldsymbol{X}_1^*$，则幅度特性满足要求，但因 $E_2(\boldsymbol{X}_1^*) \gg E_2^*$，群时延特性误差将很大；反之，若取 $\boldsymbol{X} = \boldsymbol{X}_2^*$，则幅度特性误差将很大。因此，为了兼顾两者，必须定义如下的复合目标函数：

$$E = W_1(\omega)E_1 + W_2(\omega)E_2 \tag{1.15}$$

注意，这里权因子 W_1、W_2 可以是频率的函数，表示在某些频率点上，可以幅度特性为主，某些频率点则以群时延特性为主，或者两者并重。通信工程中类似的例子还有增益与通频带宽度均衡优化问题。

二、最小成本问题

在曲线拟合问题中，我们是根据技术特性要求构造目标函数的，但是实际设计不但要考虑技术指标，而且必须考虑经济指标，即系统的成本和效益。一个好的设计应该是在满足技术指标的前提下，使待设计部件或系统的成本越低越好。为此，我们可将设计归结为如下的最优化问题：

$$\begin{aligned} &\min_{\boldsymbol{X}} \; C(\boldsymbol{X}) \\ &\text{s.t.} \;\; \boldsymbol{X}_L \leqslant \boldsymbol{X} \leqslant \boldsymbol{X}_H \\ &\qquad |S_i - S_{ai}(\boldsymbol{X})| \leqslant \varepsilon_i \quad i = 1, 2, \cdots, p \end{aligned} \tag{1.16}$$

其中，$C(\boldsymbol{X})$ 为待设计系统的成本函数，即优化问题的目标函数；

$\boldsymbol{X}_L \leqslant \boldsymbol{X} \leqslant \boldsymbol{X}_H$ 为边界（几何）约束条件，它规定了设计参数的允许取值范围；

$|S_i - S_{ai}(\boldsymbol{X})| \leqslant \varepsilon_i$ 为性能约束条件，它规定了各项技术指标及其容差范围；

S_i 为第 i 项技术指标要求；

$S_{ai}(\boldsymbol{X})$ 为待设计系统的实际指标（第 i 项）。

边界约束条件和性能约束条件确定了设计变量的可行域，不同类型的性能约束条件规定了不同类型的最小成本设计问题。下面举两个例子予以说明。

1. 电路容差设计

众所周知，电路元件参数包括标称值和容差。按照曲线拟合方式构造目标函数进行优化设计，可以求得元件的标称值。若所有元件均取该标称值，则电路的各项技术指标都得到满足。但由于容差的存在，实际电路的元件将偏离标称值。在最坏情况下，各元件值的偏离所引起的电路指标的偏离互相加强，以至于有可能使电路的某项指标遭受破坏，特别是批量生产时，会导致产品合格率的下降。因此，我们不但要求标称电路的各项指标符合设计要求，而且当各元件在容差范围内变化时，电路指标仍然满足要求。

一个最简单的方法是全部元件都采用容差极小的精密元件，这样就可将容差的影响减到最低限度，甚至可以毋需考虑，但是其代价是电路成本将大大增加。实际上，各元件值的变化对电路指标的影响大小是不一样的，因为电路响应对各元件的灵敏度不同。对于灵敏度小的元件来说，其容差大小对电路指标影响不大，故可以取较大容差的元件，只有灵敏度大的元件才需要采用精密元件。也就是说，我们应对各元件的容差进行合理分配，使得在保证电路指标的前提下，尽量降低电路的成本，即求解最小成本问题。

(1) 成本函数

根据各元件的容差，可取成本函数为：

$$C = \sum_{i=1}^{n} C_i = \sum_{i=1}^{n} (a_i + b_i/\varepsilon_i) \tag{1.17}$$

式中，C_i 为构成电路的各元件的成本；

a_i 为元件的固定成本(如材料等)；

b_i/ε_i 为元件的可变成本，它取决于元件的精密度；

若进一步考虑到批量生产时，在全部 S 个产品中，仅有 S_p 个产品合格，则成本函数为：

$$C = S\sum_{i=1}^{n} C_i/S_p = \sum_{i=1}^{n} C_i/Y \tag{1.18}$$

式中，$Y = S_p/S$ 为该电路的合格率，显然它也是元件容差的函数。

(2) 性能约束条件

对于重要产品，如用于军事、医疗、宇航等方面的电路，我们要求其合格率为100%，即电路元件在其容差范围内不论取什么值，电路各项指标均满足要求。这时，性能约束条件的形式和意义如(1.16)式所示，相应的成本函数取为(1.17)式，称之为“最坏情况设计”。

对于一些民用产品，为了降低成本，放宽对元件容差的要求，我们只要求有一定的合格率，这时性能约束条件可为：

$$Y \geqslant Y_0 \tag{1.19}$$

式中，Y_0 为要求的合格率，相应的成本函数取为(1.18)式。

(3) 设计变量

在上述分析中，我们假定各元件的标称值已确定，然后分配各元件的容差，使电路成本最低。这时设计变量就是各元件的容差，即：

$$\boldsymbol{X} = [\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_n]^T \tag{1.20}$$

进一步考察，发现元件允许误差与标称值的选定很有关系。图1.2示出二元件的情况。S 为由性能约束条件确定的可行域，$\boldsymbol{P}_0^1$、$\boldsymbol{P}_0^2$ 为选定的二个标称值点(中心值点)。由图显见，由于 $\boldsymbol{P}_0^1$ 靠近 $\boldsymbol{S}$ 的边界，因此这时二元件的允许最大容差很小，而 $\boldsymbol{P}_0^2$ 位于 $\boldsymbol{S}$ 的中心，这时元件容差可以很大，电路成本相应地也下降了。所以，在电路容差设计中，我们可以将元件标称值及容差一

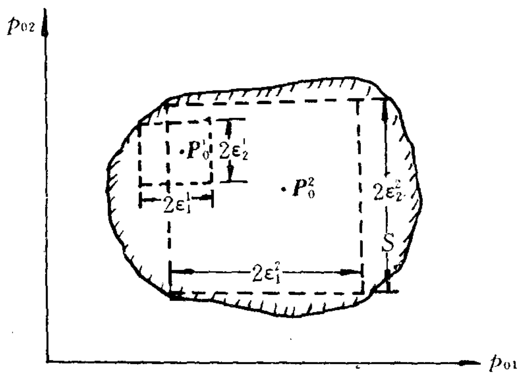

图 1.2

起作为设计变量，即：

$$X = [p_{01}, \cdots, p_{0n}, \varepsilon_1, \cdots, \varepsilon_n]^T \quad (1.21)$$

将这称为电路的“中心值设定及容差分配问题”。

用最优化方法求解电路容差设计问题，关键在于如何有效地计算电路合格率 Y 及其对设计变量的梯度。在数学上，一个 n 元函数 $f(X)$ 的梯度定义为：$\nabla f(X) = \left[\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \cdots, \frac{\partial f}{\partial x_n}\right]^T$。这里，$Y$ 对设计变量的梯度即为：

$$\nabla Y = \left[\frac{\partial Y}{\partial p_{01}}, \cdots, \frac{\partial Y}{\partial p_{0n}}, \frac{\partial Y}{\partial \varepsilon_1}, \cdots, \frac{\partial Y}{\partial \varepsilon_n}\right]^T$$

一般，Y 值采用统计试验法不难确定，而 Y 梯度的计算则是容差问题的难点，也是电路 CAD（计算机辅助设计）的一个重要研究课题。

2. 系统可靠性问题

通信系统是否可靠是关系到通信质量的重要问题。系统可靠性指的是系统正常工作的概率，它取决于构成该系统的部件的可靠性以及系统本身的结构方式。

主备用结构是提高系统可靠性的一种常用方法。图 1.3 就是一个有备用部件的复杂通信系统的系统模型。其中每个方框表示由许多个元器件组成的子系统，N 个子系统串联组成一个可独立工作的系统。图中用虚线方框表示的一个串联系统为主用系统，其余 $M-1$ 个串联系统则为备用系统。显然，只要这 M 个系统不同时发生故障，该通信系统就能正常工作。

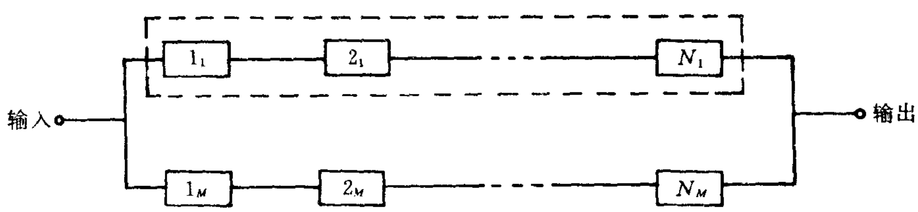

图 1.3

对于任一串联系统来说，其可靠性等于各个子系统可靠性的乘积（假设各个子系统相互独立），即：

$$R_N = \prod_{j=1}^{N} R_j(X_j) \quad (1.22)$$

式中，R_N 为串联系统的可靠性；

$R_j(X_j)$ 为第 j 个子系统的可靠性；

X_j 为第 j 个子系统包含的元件。

如果组成子系统的各元件不出现故障的事件相互独立，则子系统可靠性 $R_j(X_j)$ 等于各元件可靠性的乘积。

对于全系统来说，系统可靠的概率应是 M 个串联系统至少有一个可以正常工作的概率。记全系统不出故障的概率为 R_S，很容易求得 M 个主、备系统全部出故障的概率为 $(1-R_N)^M$，从而全系统可靠的概率为

$$R_S = 1 - (1 - R_N)^M \tag{1.23}$$

以 $M = 2$ 为例，设已求得 $R_N = 0.8$，则 $R_S = 0.96$，显然可靠性有较多的提高。考虑到串联系统中有些子系统的可靠性高，有些子系统可靠性低，显然，从降低成本出发，完全没有必要一视同仁地都备用 M 套，有些子系统甚至可以不设置备份。这时系统可靠性可按下式计算：

$$R_{sj} = 1 - (1 - R_j(\boldsymbol{X}_j))^{M_j} \qquad j = 1,2,\cdots,N \tag{1.24}$$

$$R_S = \prod_{j=1}^{N} R_{sj}(\boldsymbol{X}_j) \tag{1.25}$$

为了在确保通信质量的前提下，尽量降低系统成本，我们可构造如下的最小成本问题：

$$\begin{aligned} \min \quad & C_S = \sum_{j=1}^{N} C_j(\boldsymbol{X}_j) \\ \text{s.t.} \quad & R_S = \prod_{j=1}^{N} R_{sj}(\boldsymbol{X}_j) \geqslant R_0 \end{aligned} \tag{1.26}$$

式中，C_S 为全系统的总费用；

C_j 为第 j 级子系统的费用，它是元器件 X_j 及子系统备份数 M_j 的函数；

R_0 为要求的系统可靠性。

求解最优化问题(1.26)即可获得兼顾系统可靠性和系统成本的最优设计。

三、最大效益问题

在邮电企业管理中，许多问题的决策基于如何获得最大的经济效益。例如，长途线路工程和市话工程的材料分配和运输问题；市话电缆的扩容问题；通信网布局问题；工程劳务安排问题等，它们常常可以归结为如下的最优化问题：

$$\begin{aligned} \max \quad & B(\boldsymbol{X}) = \sum_{i=1}^{N} C_i x_i + b \\ \text{s.t.} \quad & \sum_{j=1}^{N} a_{ij} x_j \leqslant S_i \qquad i = 1,2,\cdots,m \end{aligned} \tag{1.27}$$

式中，$B(\boldsymbol{X})$ 为效益函数；

S_i 为第 i 项可用资源总量；

$\boldsymbol{X}$ 为决策变量。

求解上述问题，可以获得在给定的资源条件下，以期达到最大经济效益的最佳方案。

下面以市话电缆扩容为例，说明如何建立最大效益问题的数学模型。

[**例 1.1**] 某市话局为解决用户装机问题，准备铺设新市话电缆扩容。据调查：向东扩容平均每公里可增加 60 门电话，向南平均每公里可增 45 门，向西每公里可增 30 门，向北每公里可增加 25 门。仓库现有扩容材料及各方向扩容所需材料(含费用)如表 1-1 所示。如何选择最佳扩容方案，使在现有资源条件下扩展的电话门数最多。

表 1-1

工程用料名称	现有总数	每公里扩容所需材料或费用			
		东	南	西	北
管道(个)	100000	2000	1000	1000	0
电缆(m)	11000	1000	1000	800	1000
其他材料及劳务费用(元)	20000	2000	2500	1000	1500

设选定的方案为向东扩容x_1公里，向南扩容x_2公里，向西扩容x_3公里，向北扩容x_4公里，则根据已知条件建立数学模型如下：

$$
\begin{aligned}
\max\ & B = 60x_1 + 45x_2 + 30x_3 + 25x_4 \\
\text{s.t.}\ & 2000x_1 + 1000x_2 + 1000x_3 \leqslant 10000 \\
& 1000x_1 + 1000x_2 + 800x_3 + 1000x_4 \leqslant 11000 \\
& 2000x_1 + 2500x_2 + 1000x_3 + 1500x_4 \leqslant 20000
\end{aligned}
\tag{1.28}
$$

求解问题(1.28)可得最佳方案为向东扩容3公里，向南扩容4公里，向北扩容4公里，向西不扩容，总共可扩装电话460门。

1.1.3 最优化问题分类

前已述及，最优化问题均可由(1.1)式描述。但是，为了充分利用不同问题的特性，导出适用于该类问题的有效算法，有必要进一步对最优化问题进行分类。通常可根据目标函数和约束函数的形式划分最优化问题的类型。划分方法如下：

$$
\text{最优化问题}\begin{cases}
\text{无约束问题}\begin{cases}\text{一维问题}\\ \text{多维问题}\end{cases}\\
\text{有约束问题}\begin{cases}\text{线性规划}\\ \text{非线性规划}\end{cases}\\
\text{多目标问题}\\
\text{动态问题}
\end{cases}
$$

没有任何约束条件的最优化问题，如(1.3)式所示的曲线拟合问题，称之为无约束问题。严格说来，绝大多数工程设计问题都是有约束的。例如在网络设计中，至少要求元件参数取正值。但是这样的平凡约束条件，只要初始点选择合理，总是能自然地得到满足。因此，许多设计问题都可以简单地视作为一个无约束优化问题。

如果优化变量数只有一个，则就是一维问题，称之为一维寻查或线性寻查。最优化算法一般都是包含多次一维寻查的迭代过程，因此，一维寻查的效率将直接影响所有最优化算法的效率。多维问题则是无约束问题的核心，它包括一般目标函数优化以及非线性最小二乘问题和 min max 最优一致逼近问题等特殊目标函数的优化。

对于工程设计来说，有约束问题具有更为实际的意义。但是无论在理论上还是在算法实现上，有约束问题都远比无约束问题困难得多，它是目前最优化研究中的重要课题。

若目标函数和约束函数均为优化变量的线性函数，如(1.27)式所示最大效益问题，就称为线性规划。若目标函数和约束函数中至少有一个是优化变量的非线性函数，如(1.16)式所示的最小成本问题，就称为非线性规划。特别地，当目标函数为二次函数，约束函数均为线性函数时，则称为二次规划。

多目标问题的实际背景是工程设计中有多个需同时满足的设计指标，其目的是获得一个兼顾所有指标的最佳设计方案。(1.12)式所示曲线拟合中的复合目标函数法以及(1.26)式所示系统可靠性设计约束优化法就是解决多目标优化问题的二种简单的方法。

动态问题即多阶段决策问题，它具有时间上的顺序性和阶段性，因此称求解这类决策问题的方法为动态规划。相应地，我们称前述与时间无关的最优化问题为静态问题。动态规划就是把原始的动态问题分解为许多静态子问题(阶段)，然后顺序地求解。最后一个静态子问题的

解，就是原问题的最优解。

1.2 最优性条件

本节讨论最优解的必要条件和充分条件，即最优条件，并给出其几何意义。由最优化数学模型(1.1)式可知，最优解就是目标函数的最小值点，因此我们将从函数的极值理论出发进行研究。

1.2.1 最优化过程的几何意义

一、等值面

在 n 维空间中，曲面 $f(\boldsymbol{X}) = C$ 称为目标函数的等值面。式中，C 为任意常数。取不同的 C 值，就可得到一簇等值面。等值面的法线方向定义为函数值增加的方向，它就是目标函数在该点的梯度方向 $\boldsymbol{g}$：

$$\boldsymbol{g} \triangleq \nabla f(\boldsymbol{X}) = \left[\frac{\partial f}{\partial x_1}, \cdots, \frac{\partial f}{\partial x_n}\right]^T \tag{1.29}$$

图 1.4 示出函数

$$f(\boldsymbol{X}) = x_1{}^4 - 2x_2x_1{}^2 + x_2{}^2 + x_1{}^2 - 2x_1 + 5 \tag{1.30}$$

的等值面，$\boldsymbol{n}$ 为等值面 $f(\boldsymbol{X}) = 5$ 在点 $\boldsymbol{X}_k$ 处的法线方向，它指向函数值增加的方向，$\boldsymbol{X}^*$ 为 $f(\boldsymbol{X})$ 的最小值点。

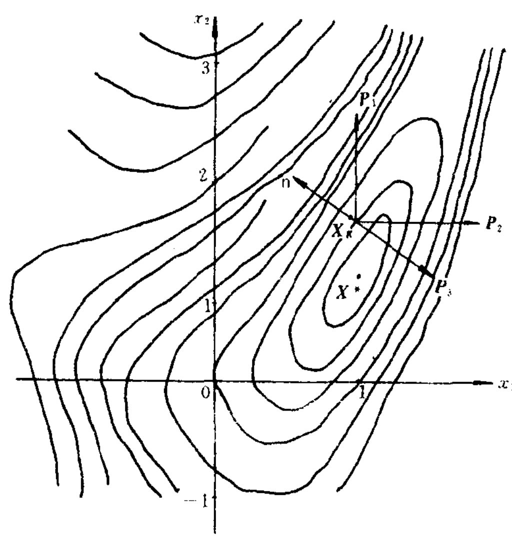

图 1.4

二、下降方向

由数学分析理论知，函数沿方向 $\boldsymbol{p}$ 的变化率可由其在该方向的方向导数来度量，即

$$\frac{\partial f}{\partial \boldsymbol{p}} = [\nabla f]^T \boldsymbol{p}^0 \tag{1.31}$$

式中，$\boldsymbol{p}^0$ 为 $\boldsymbol{p}$ 方向的单位向量。

因此，对于给定的点 $\boldsymbol{X}_k$，我们可以根据方向导数的符号，判定通过该点的方向 $\boldsymbol{p}_k$ 是函数的上升方向还是下降方向：

$$\nabla f(\boldsymbol{X}_k)^T \boldsymbol{p}_k \begin{cases} < 0 & \boldsymbol{p}_k \text{ 为下降方向} \\ > 0 & \boldsymbol{p}_k \text{ 为上升方向} \\ = 0 & \boldsymbol{p}_k \text{ 位于等值面的切面上} \end{cases} \tag{1.32}$$

显然，当

$$\boldsymbol{p}_k = -\nabla f(\boldsymbol{X}_k) \tag{1.33}$$

时，$\nabla f(\boldsymbol{X}_k)^T \boldsymbol{p}_k^0 = \min$，即负梯度方向是目标函数的最速下降方向。在图 1.4 中，$\boldsymbol{p}_1$、$\boldsymbol{p}_2$、$\boldsymbol{p}_3$ 分别为函数 $f(\boldsymbol{X})$ 在 $\boldsymbol{X} = \boldsymbol{X}_k$ 处的上升方向、下降方向和最速下降方向。