深圳北理莫斯科大学
УНИВЕРСИТЕТ МГУ-ППИ В ШЭНЬЧЖЭНЕ
SHENZHEN MSU-BIT UNIVERSITY

Математическое моделирование и исследование моделей с помощью математических программ

数学建模及数学软件的使用

Лекция № 7a (优化)

张晔

ye.zhang@smbu.edu.cn

- 最优化问题: 在数学与计算机科学领域中，是从所有可行解中查找最优良的解的问题。
- Оптимизация/ Optimization/ Экстремальные задачи
- 根据变量是连续的或离散的，最优化问题可分为两类：连续最优化问题 与 组合优化。
- 相对于决策问题(Decision problem)、功能性问题(Function problem)，最优化问题是：从问题的多个解中，求出最佳解。
- 袁亚湘 院士的ppt

引子：瞎子爬山

瞎子与计算机

瞎子　－－－　知道脚底下情况，
但看不见周围的东西

计算机　－－－　给一个点x，可计算：
$f(x)$, $\partial f(x)$, … …
但对于x 附近的其他y,
不知道 $f(y)$

瞎子爬山 vs 优化方法

瞎子和计算机谁快?

瞎子和计算机谁聪明?

瞎子会如何“看” “瞎子爬山法”呢?

优化方法的特征

- 基于极小（ min f(x)）
- 基于KKT 条件（ $\nabla L(x, \lambda) = 0$ ）
- 迭代

$$x_{k+1} = U(x_k, \dots, \dots,)$$

KKT条件/Karush-Kuhn-Tucker Conditions

- В теории оптимизации условия Каруша — Куна — Таккера: необходимые условия решения задачи нелинейного программирования. Чтобы решение было оптимальным, должны быть выполнены некоторые условия регулярности. Метод является обобщением метода множителей Лагранжа. В отличие от него, ограничения, накладываемые на переменные, представляют собой не уравнения, а неравенства.

$$\min_{x} \quad f(x)$$

Subject to:

$$g_i(x) \leq 0, h_j(x) = 0$$

$f(x)$是需要最小化的函数，$g_i(x)\ (i=1,\ldots,m)$是不等式约束，$h_j(x)\ (j=1,\ldots,l)$是等式约束，m和l分别为不等式约束和等式约束的数量。

必要条件

假设有目标函数，即是要被最小化的函数$f: \mathbb{R}^n \to \mathbb{R}$，约束函数$g_i: \mathbb{R}^n \to \mathbb{R}$及$h_j: \mathbb{R}^n \to \mathbb{R}$。再者，假设他们都是于$x^*$这点是连续可微的，如果$x^*$是一局部极小值，那么将会存在一组所谓乘子的常数$\lambda \geq 0,\ \mu_i \geq 0\ (i=1,\ldots,m)$及$\nu_j\ (j=1,\ldots,l)$令到

$$\lambda + \sum_{i=1}^{m} \mu_i + \sum_{j=1}^{l} |\nu_j| > 0,$$

$$\lambda \nabla f(x^*) + \sum_{i=1}^{m} \mu_i \nabla g_i(x^*) + \sum_{j=1}^{l} \nu_j \nabla h_j(x^*) = 0,$$

$$\mu_i g_i(x^*) = 0 \text{ for all } i = 1, \ldots, m.$$

充分条件

假设目标函数$f: \mathbb{R}^n \to \mathbb{R}$及约束函数$g_i: \mathbb{R}^n \to \mathbb{R}$皆为凸函数，而$h_j: \mathbb{R}^n \to \mathbb{R}$是一仿射函数，假设有一可行点$x^*$，如果有常数$\mu_i \geq 0\ (i=1,\ldots,m)$及$\nu_j\ (j=1,\ldots,l)$令到

$$\nabla f(x^*) + \sum_{i=1}^{m} \mu_i \nabla g_i(x^*) + \sum_{j=1}^{l} \nu_j \nabla h_j(x^*) = 0$$

$$\mu_i g_i(x^*) = 0 \text{ for all } i = 1, \ldots, m,$$

那么x^*这点是一全局极小值。

- 对偶：Duality / Двойственность

- 优化角度：primal problem/ dual problem
  - прямая задача / двойственняя задача

- **弱对偶 vs 强对偶**：“ $\leq$ ” **vs** “ $=$ ”

- **Min-Max问题（Bilevel optimization）：**

$$\max_x \min_y f(x,y) \leq \min_y \max_x f(x,y)$$

- **“宁做鸡头,不做凤尾”：min = 鸡；max=凤**

- 线性问题/二次规划 => **强对偶** => “ $=$ ”：等价求解比较容易的对偶问题

- 强对偶关系 ⇔ 满足KKT条件

(Lagrangian) duality

➢ 如何找到问题(1) 的最优值的一个最好的下界?

$$\min\{f(\mathbf{x}) : g_i(\mathbf{x}) \leq 0,\ i = 1, 2, \ldots, m\} \quad (1)$$

- 若方程组 $\begin{array}{l} f(\mathbf{x}) < v \\ g_i(\mathbf{x}) \leq 0, i = 1, 2, \ldots, m \end{array}$ (2) 无解，则 v 是问题(1)的一个下界。
- 注意到方程组(2)有解可以推出对于任意的 λ≥ 0, 方程(3)有解

$$f(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i g_i(\mathbf{x}) < v \quad (3)$$

- 根据逆否命题，方程组(2)无解的充分条件是存在 λ≥ 0 ，让方程(3)无解。

  方程(3)无解的充要条件是 $\min_{\mathbf{x}} f(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i g_i(\mathbf{x}) \geq v \quad (4)$

- 因为我们要找最好的下界，所以这个时候的 v 和 λ 应该取

$$v = \max_{\lambda \geq 0} \min_{\mathbf{x}} f(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i g_i(\mathbf{x}) \quad (5)$$

- 由此引入了dual problem. 证明逻辑是根据式(5)取 v 和 λ, 则(4)成立，从而导出(3)无解，然后可以知道(2)无解，因此 v 是问题(1)的下界。

迭代　–　计算的基本

千里之行始于足下　－－－老子

几个经典优化方法

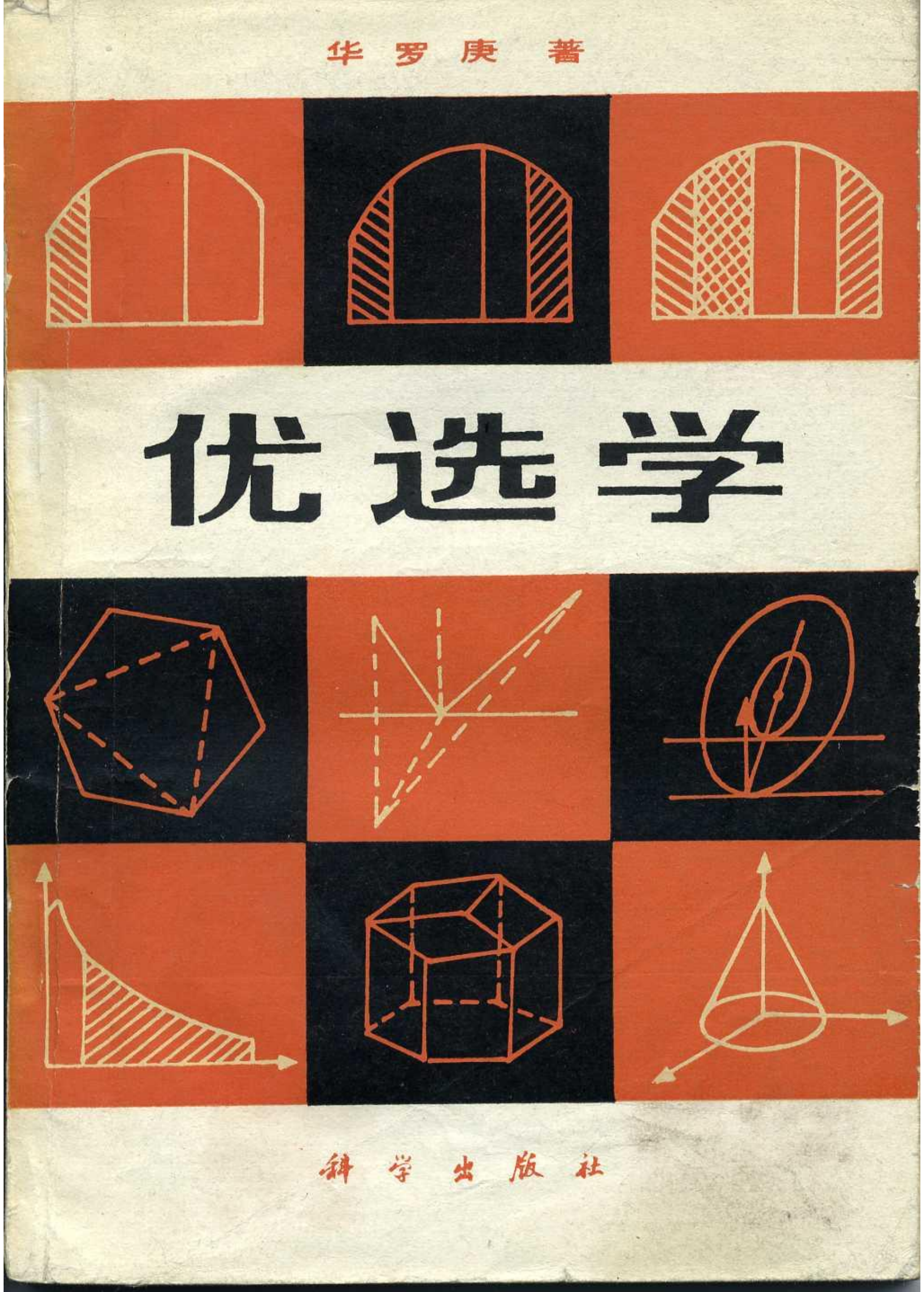
- **黄金分割法**
- **最速下降法**
- **共轭梯度法**
- **牛顿法与拟牛顿法**
- **高斯牛顿法与信赖域方法**
- **单纯形法与内点法**
- **交替方向法**

黄金分割法/Золотое сечение/ Golden ratio

目　录

第一部分　单因素优选法

第二部分　多因素优选法



华罗庚（1910－1985）

华罗庚在农村推广优选法

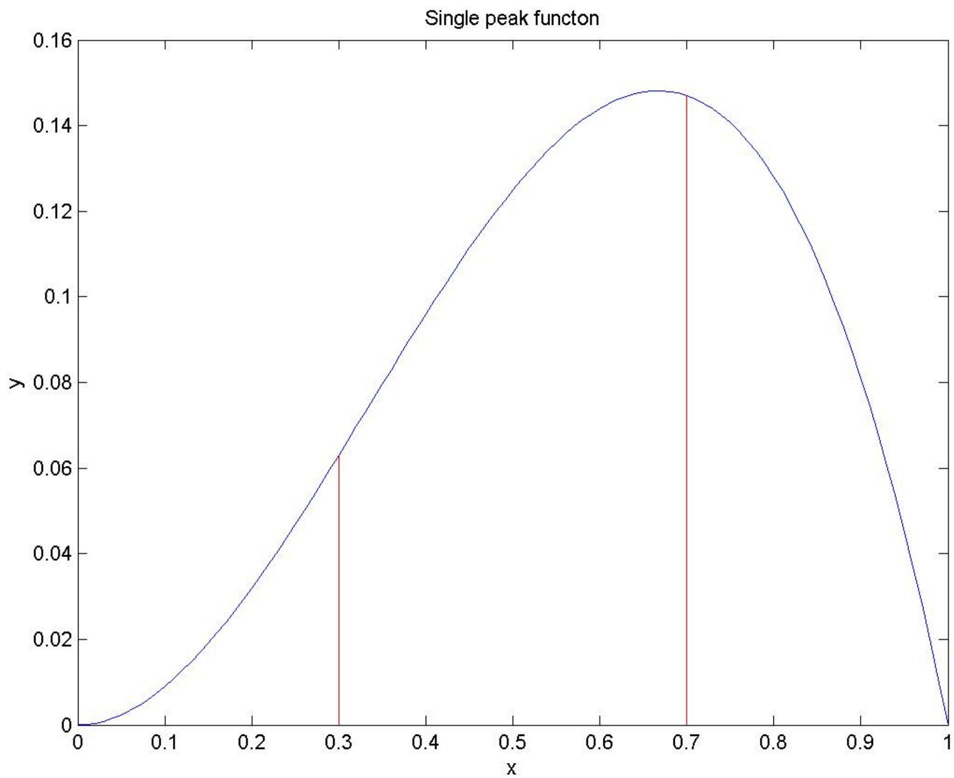
华罗庚在大庆油田讲优选法

华罗庚在矿山推广优选法

华罗庚在工厂、车间

max f(x)

- [a, b] 上的连续函数 f(x) 是单峰的（只有一个最大值点），求解 max f(x)
- 任取 a<c<d<b，如果 f(c) < f(d)，则我们只需在 [c, b] 上求 max f(x)



如何 选取 c，d ？

最优的 c, d

1） $\max[b-c, d-a]$ 达到最小

→ $c \approx d = (a+b)/2$!

2) Repeat the procedure ?

Once → $c=1/3$, $d=2/3$

Twice → $c=2/5$, $d=3/5$?

一般的 c, d

$F_0=1$, $F_1=1$,

$F_{k+1}= F_k + F_{k-1}$

General case

$c=F_{k-1}/F_{k+1}$ $d=F_k/F_{k+1}$

$$\lim_{k\to\infty}\frac{F_k}{F_{k+1}}=\frac{\sqrt{5}-1}{2}\approx 0.618$$

Fibonacci(1170-1250)

斐波那契数列又因数学家列昂纳多·斐波那契以兔子繁殖为例子而引入，故又称为“兔子数列”。

一般而言，兔子在出生两个月后，就有繁殖能力，一对兔子每个月能生出一对小兔子来。如果所有兔子都不死，那么一年以后可以繁殖多少对兔子？

我们不妨拿新出生的一对小兔子分析一下：

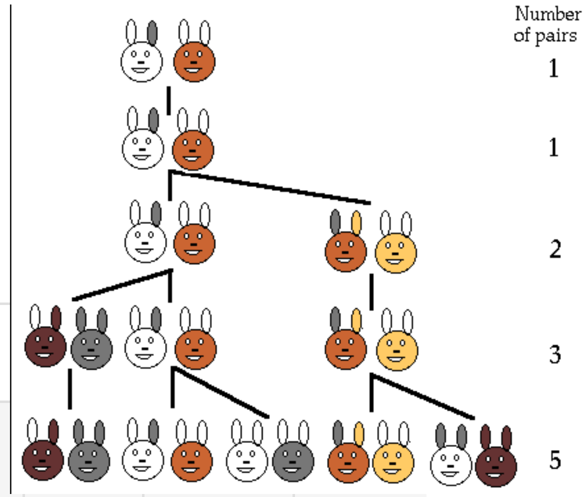
第一个月小兔子没有繁殖能力，所以还是一对

两个月后，生下一对小兔对数共有两对

三个月以后，老兔子又生下一对，因为小兔子还没有繁殖能力，所以一共是三对

——————

依次类推可以列出下表：



经过月数	0	1	2	3	4	5	6	7	8	9			
幼仔对数	1	0	1	1	2	3	5	8	13	21			
成兔对数	0	1	1	2	3	5	8	13	21	34	55	89	144
总体对数	1	1	2	3	5	8	13	21	34	55	89	144	233

幼仔对数=前月成兔对数

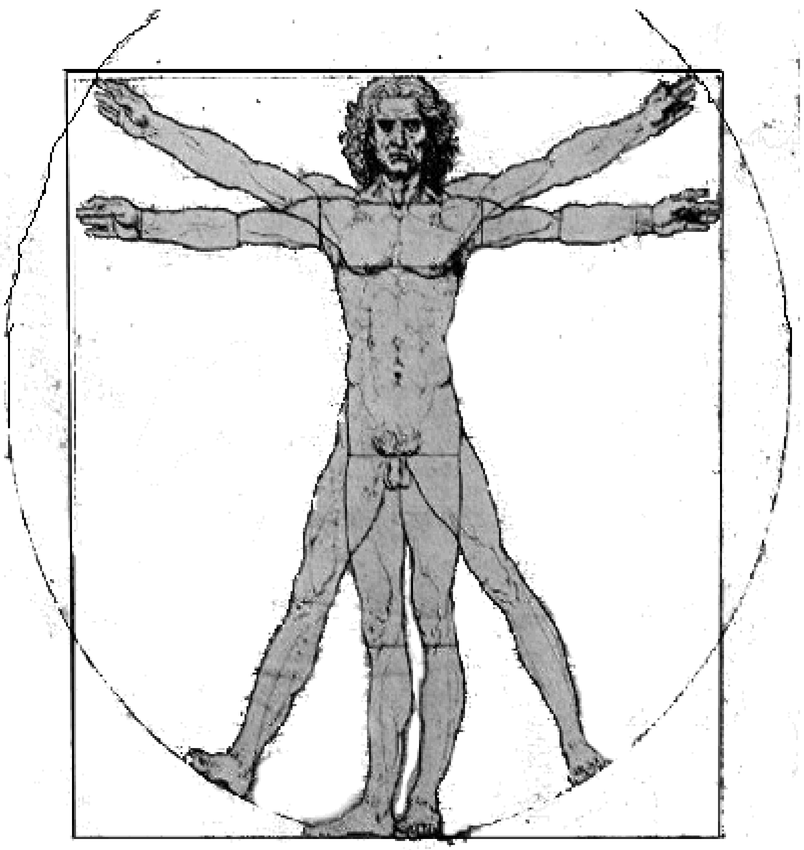
成兔对数=前月成兔对数+前月幼仔对数

总体对数=本月成兔对数+本月幼仔对数

可以看出幼仔对数、成兔对数、总体对数都构成了一个数列。这个数列有关十分明显的特点，那是：前面相邻两项之和，构成了后一项。

这个数列是意大利中世纪数学家斐波那契在<算盘全书>中提出的，这个级数的通项公式，除了具有a(n+2)=an+a(n+1）的性质外，还可以证明通项公式为：an=(1/√5)*{[(1+√5)/2]^n-[(1-√5)/2]^n}（n=1,2,3.....）

达.芬奇与黄金分割



- 黄金分割法：

- 给出[0， 1]：
- X=0.382
- Y=0.618

- 新区间：
- [0， 0.618] or
- [0.382，1]

Leonardo da Vince (1452-1519)

article | discussion | view source | history

WIKIPEDIA
The Free Encyclopedia

navigation

- Main page
- Contents
- Featured content
- Current events
- Random article

search

Go | Search

Leonardo da Vinci

From Wikipedia, the free encyclopedia

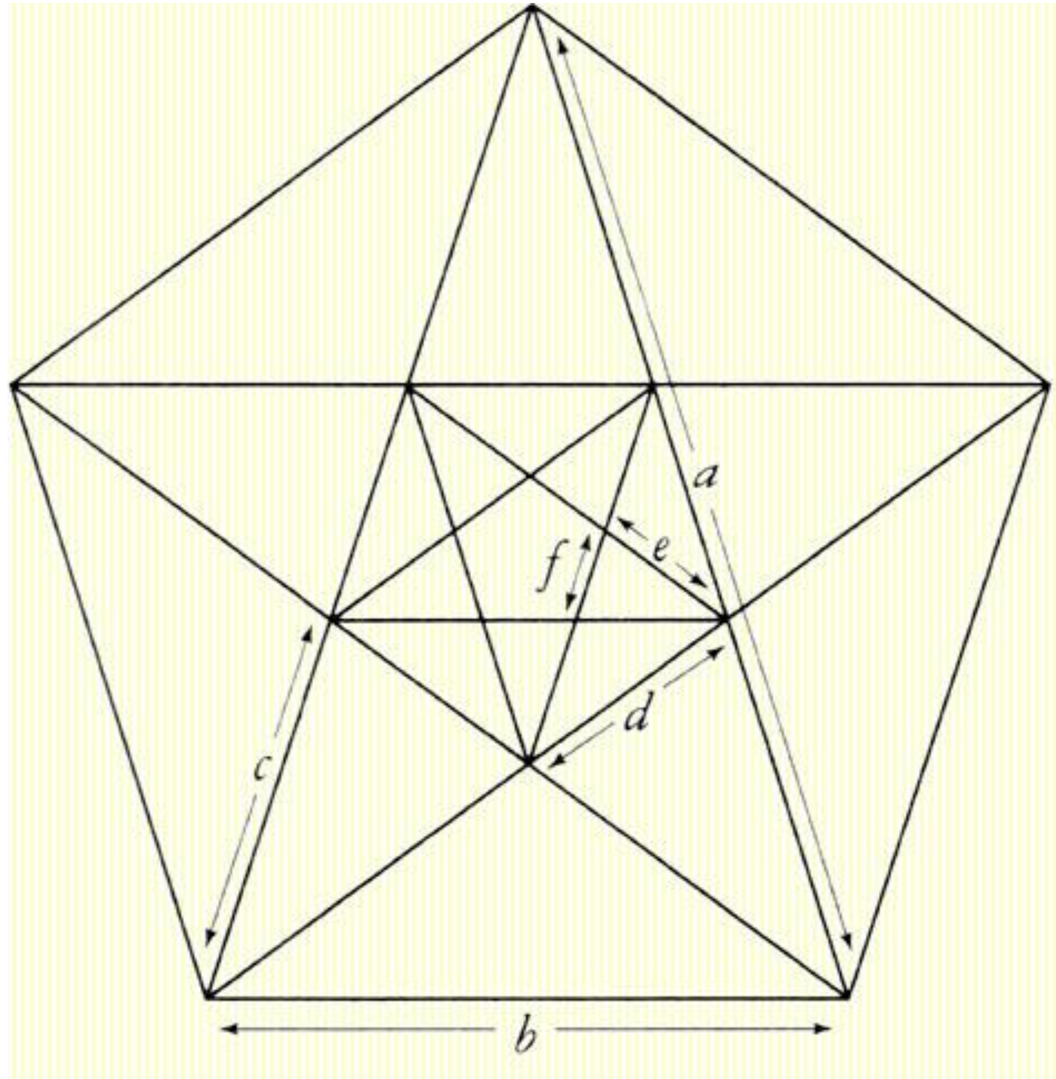
"Da Vinci" redirects here. For other uses, see Da Vinci (disambiguation).

For the 17th century Italian composer, see Leonardo Vinci

Leonardo di ser Piero da Vinci (pronunciation (help·info), April 15, 1452 – May 2, 1519) was an Italian polymath, scientist, mathematician, engineer, inventor, anatomist, painter, sculptor, architect, botanist, musician and writer. Leonardo has often been described as the archetype of the renaissance man, a man whose unquenchable curiosity was equaled only by his powers of invention.[1] He is widely considered to be one of the greatest painters of all time and perhaps the most diversely talented person ever to have lived.[2] Helen Gardner says "The scope and depth of his interests were without precedent...His mind and personality seem to us superhuman, the man himself mysterious and remote".[1]

Euclid

(约325BC—265BC)



extreme and mean ratio

最速下降法

$$\min_{x\in\Re^n} f(x)$$

$$d_k = -\nabla f(x_k)$$

$$x_{k+1} = x_k + \alpha_k d_k$$

α_k 使 f(x + α d) 达到最小

（精确搜索）

A. Cauchy, Comptes Rendus de L’Acadmia des Sciences 25(1847) 536-538

Cauchy （1789 – 1859）

最速下降法收敛速度

假定 f(x) 是二次凸函数

$$f(x) = g^T x + x^T H x$$

收敛速度:

$$\frac{f(x_{k+1}) - f(x^*)}{f(x_k) - f(x^*)} \leq \left[\frac{\sigma_1(H) - \sigma_n(H)}{\sigma_1(H) + \sigma_n(H)}\right]^2$$

最好 + 最好 = 最好 ???

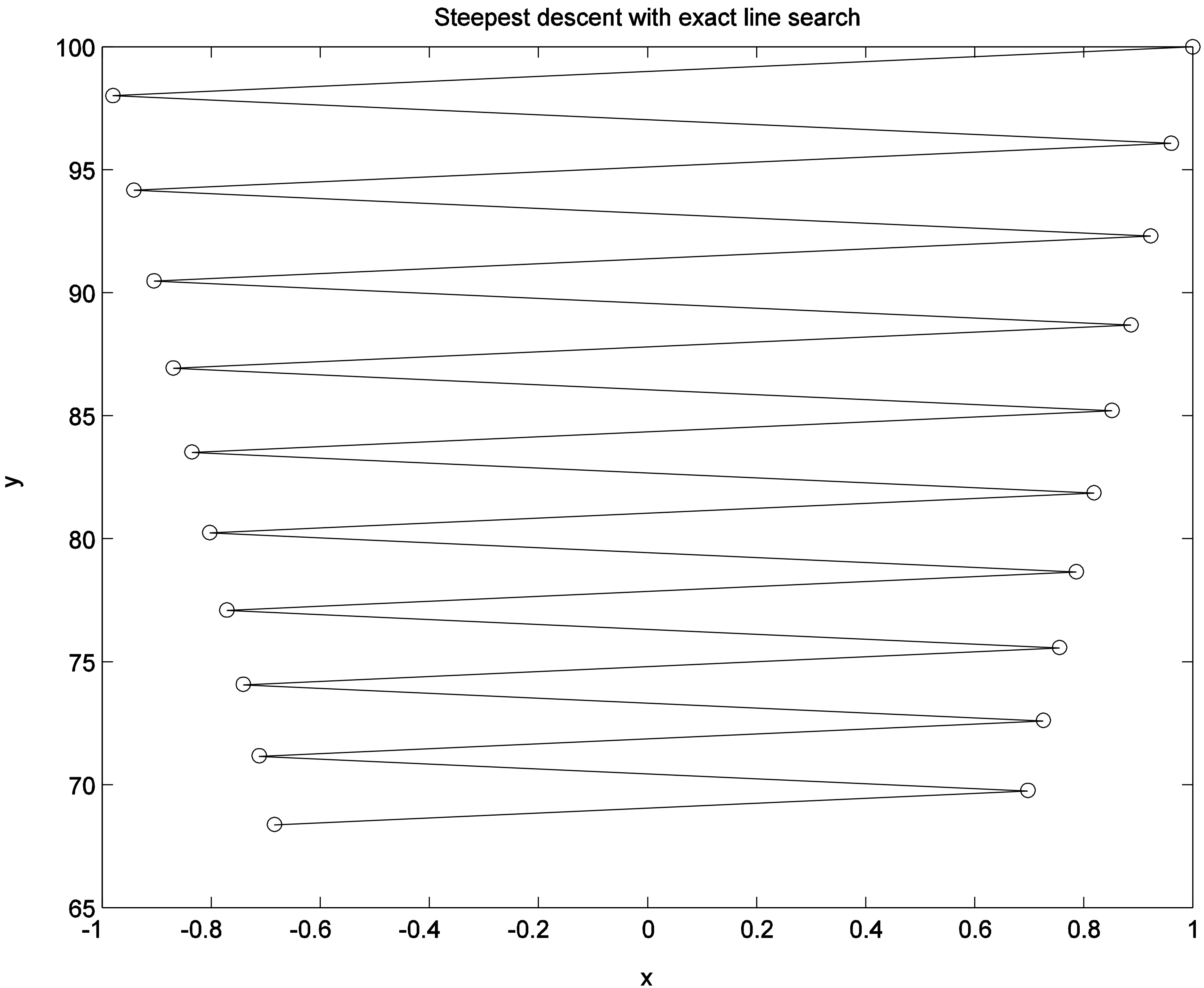
- 方向 (最速下降) （best d_k)

$$d^T \nabla f(x_k) \geq d_k^T \nabla f(x_k), \quad \forall d, ||d|| = ||d_k||$$

- 步长 (精确搜索) (best α_k)

$$f(x_k + \alpha d_k) \geq f(x_k + \alpha_k d_k), \quad \forall \alpha$$

- $x_{k+1} = x_k + \alpha_k d_k$ 是否最好 ????



最速下降法应用于 $f(x,y)=100x^2+y^2$

最速下降法的启示

最好方向 + 最好步长 ≠ 最好的方法

推论：

班上最好的女生不应该嫁给
班上最好的男生！

Barzilai & Borwein Method

方向

(最速下降 – 最好方向)

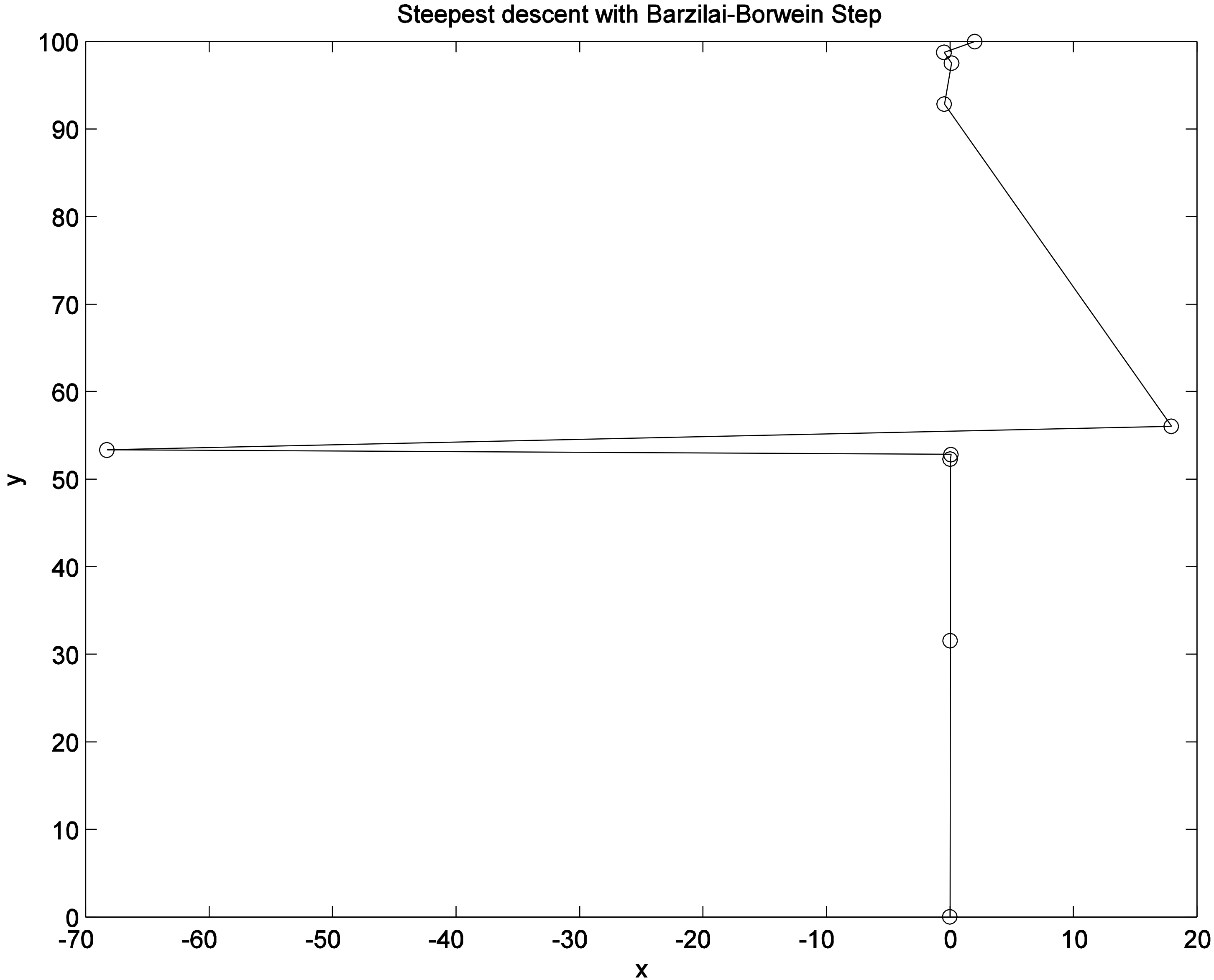
步长

(上一次的精确搜索步长)

最好的d + 上一步最好的 α

→ 最好

J.M. Borwein(1951-



BB 方法 应用于 $f(x,y)=100x^2+ y^2$

BB方法的启示

最好方向 + 上一次最好步长 = 最好

推论:

班上最好的女生应该嫁给
高年级最好的男生!

共轭梯度法/Conjugate gradient method/ Метод сопряжённых градиентов

- 共轭梯度法是介于最速下降法与牛顿法之间的一个方法，它仅需利用一阶导数信息，但克服了最速下降法收敛慢的缺点，又避免了牛顿法需要存储和计算Hessen矩阵并求逆的缺点。
- 优点：存储量小，具有n步收敛性，稳定性高，而且不需要任何外来参数。

$$\min_{x \in \Re^n} f(x)$$

$$x_{k+1} = x_k + \alpha_k d_k$$

- α_k 使 $f(x + \alpha d)$ 达到最小（精确搜索）
- 如何选取 d_k?

共轭梯度法

$$d_k = -\nabla f(x_k) + \beta_k d_{k-1}$$

前后两次的方向 d 相互共轭

d_k and d_{k-1} are conjugate!

$$d_k^T \nabla^2 f(x) d_{k-1} = 0$$

- **收敛速度：** N次迭代内收敛，其中N是变量的维度。
- **但是：在 ill-posed problems 和 机器学习中，它不一定好！**

共轭梯度法基本思想

考虑 $f(x) = (1/2)\ x^T A x - b^T x$

共轭方向 $d_1, d_2, \cdots, d_n$

$$x = \sum_{i=1}^{n} \alpha_i d_i$$

$$f(x) = \sum_{i=1}^{n} [\frac{1}{2}\alpha_i^2 d_i^T A d_i - \alpha_i b^T d_i]$$

一个 n 维问题 转化为 n 个 一维问题

共轭梯度法的实现

$$d_{k+1} = -g_{k+1} + \beta_k d_k$$

著名的β选取:

$$\beta_k^{HS} = \frac{(g_{k+1} - g_{k+1})^T g_{k+1}}{d_k^T (g_{k+1} - g_k)}$$

$$\beta_k^{FR} = \frac{||g_{k+1}||_2^2}{||g_k||_2^2}$$

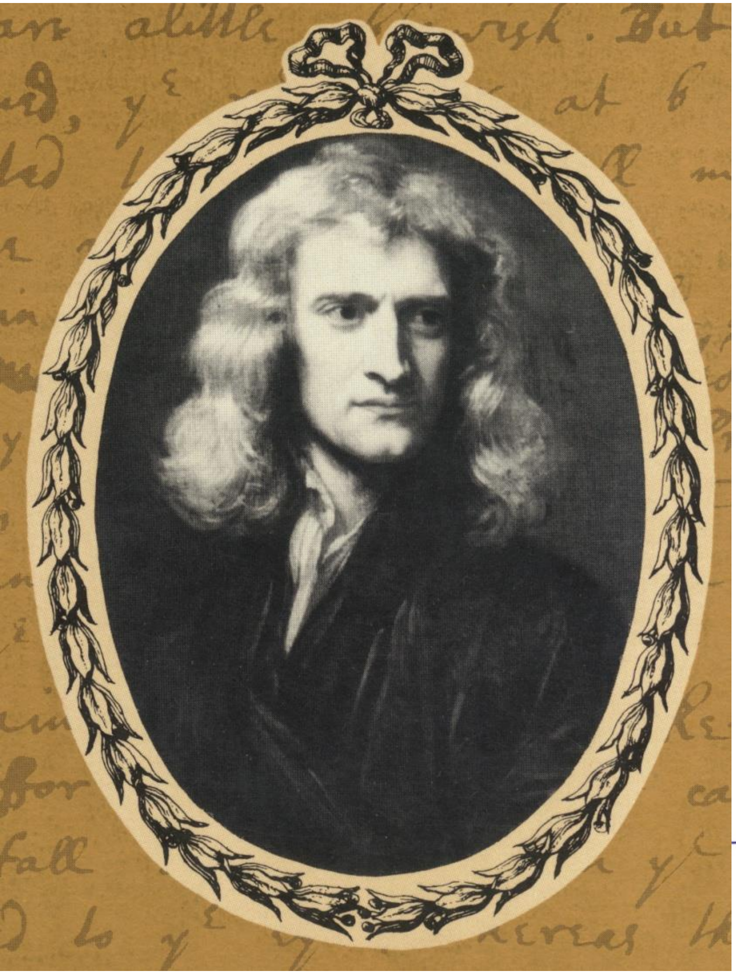
$$\beta_k^{PRP} = \frac{(g_{k+1} - g_k)^T g_{k+1}}{||g_k||_2^2}$$

Hestenes-Stiefel(1952), Fletcher-Reeves(1964), Polak-Ribiére(1969), Polyak(1969)

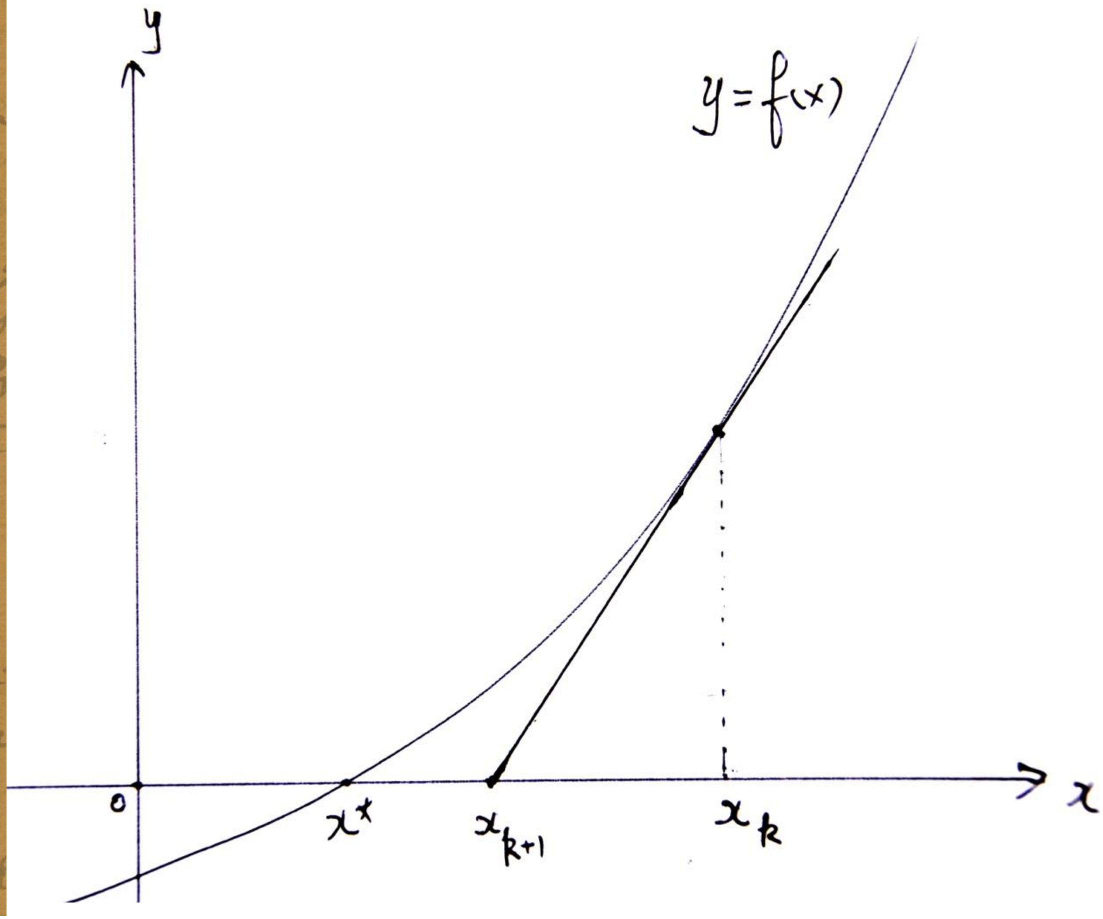
Hestenes
(1906-1991)

Stiefel
(1909-1978)

牛顿法：切线代替曲线



Newton(1643-1727)



牛顿法求f(x)=0 的根

牛顿法性质

迭代公式：

$$x_{k+1} = x_k - (\nabla^2 f(x_k))^{-1} \nabla f(x_k)$$

优缺点：

1) 优点：速度快（二次收敛）

$$||x_{k+1} - x^*|| = O(||x_k - x^*||^2)$$

2) 缺点：计算量大（需要计算二阶导数）

拟牛顿法

牛顿:

$$x_{k+1} = x_k - (\nabla^2 f(x_k))^{-1} \nabla f(x_k)$$

拟牛顿:

Quasi-Newton: $x_{k+1} = x_k - (B_k)^{-1} \nabla f(x_k)$

如何选取 B?

如何“拟”牛顿？

拟牛顿方程：

$$B_{k+1}s_k = y_k$$

$$\begin{aligned} s_k &= \alpha_k d_k \\ y_k &= g_{k+1} - g_k. \end{aligned}$$

Davidon(1959), Fletcher and Powell (1963):

$$B_{k+1} = B_k - \frac{B_k s_k y_k^T + y_k s_k^T B_k}{s_k^T y_k} + \left(1 + \frac{s_k^T B_k s_k}{s_k^T y_k}\right)\frac{y_k y_k^T}{s_k^T y_k},$$

F(x)=0

非线性最小二乘问题

$$\min_{x\in\Re^n} ||F(x)||_2^2$$

$$F(x) = (f_1(x), \ldots. f_m(x))^T$$

最小二乘问题

- 超定方程组求解

$$Ax = b, \quad A \in \Re^{m\times n}, b \in \Re^{m}, x \in \Re^{n}. \qquad m >> n.$$

- 数值模拟，曲线拟合

$$y_i \approx f(x_i, \beta), i = 1, ..., m.$$

- 反问题

$$\begin{aligned} h(x,y) &= \int\int k(x-\xi, y-\eta) f(\xi,\eta) d\xi d\eta + e(x,y) \\ &= (k \star f)(x,y) + e(x,y). \end{aligned}$$

高斯－牛顿法

$x_{k+1} = x_k + d_k$，如何求 d_k？

$$\min \|F(x_k + d)\|_2^2 \quad \Longrightarrow \quad \min \|F(x_k) + A(x_k)^T d\|_2^2$$

A(x) 是 F(x) 的 JACOBI 阵

$$d_k = -(A(x_k)^T)^+ F(x_k)$$

J. C. F. Gauss

(1777-1855)

I. Newton

(1642-1727)

Levenberg-Marquardt 方法

当 A 坏条件时 高斯－牛顿步 很可能不好。

Levenberg (1944)，Marquardt(1963) 提出：

$$d_k = -(A(x_k)A(x_k)^T + \lambda_k I)^{-1} A(x_k) F(x_k)$$

L-M步的最优性

设 d_k 是 Levenberg-Marquardt 步:

$$d_k = -(A(x_k)A(x_k)^T + \lambda_k I)^{-1} A(x_k) F(x_k)$$

则它也是如下问题的解

$$\min_{d \in \Re^n} ||F(x_k) + A(x_k)^T d||_2^2$$

$$s.\ t. \qquad ||d||_2 \leq \|d_k\|_2$$

信赖域方法

信赖域方法基本思想

1) 局部区域
2) 逼近模型
3) 调节模型和区域

孙悟空的信赖域

- **袁亚湘 院士 团队**

交替方向方法/ADMM

Alternating Direction Method of Multipliers

$$\min f(x)$$

$$x=(x_1,\ldots, x_n)$$

固定 x_j $(j \neq i)$，对 x_i 进行优化

例子：

Gauss-Seidel 方法 求解 $Ax = b$

- **何炳生教授**
- **http://maths.nju.edu.cn/~hebma （很多资料）**

交替方向方法 -- L1 优化

$$\min \quad ||x||_1$$

$$\text{s.t.} \quad Ax = b$$

罚函数方法：

$$\min \sigma ||x||_1 + ||Ax - b||_2^2$$

交替方向方法 -- L1 优化

$$\min \sigma ||x||_1 + ||Ay - b||_2^2$$

$$\text{s. t.} \quad y = x$$

对于: min $||Ay-b||^2$

可以 用梯度法 下降;

对于 min $||x||_1$

$|x_{i+1}| <= |x_i| - \delta$ (shrinkage operator)

祝福大家

优化自己的一生！

表 11-1　　MATLAB 优化工具箱函数列表

类　别	适用问题	公式描述	可用函数
极小值优化	标量最小值优化问题	$\min_x f(x)$ $l < x < u$ (x 是标量)	fminbnd
	无约束最小值优化问题	$\min_x f(x)$	fminunc fminsearch
	线性规划	$\min_x f^T x$ $A \bullet x \le b$ $Aeq \bullet x = beq$ $l \le x \le u$	linprog
	二次规划	$\min_x \frac{1}{2} x^T Hx + f^T x$ $A \bullet x \le b$ $Aeq \bullet x = beq$ $l \le x \le u$	quadprog
	约束最小值优化问题	$\min_x f(x)$ $c(x) \le 0$ $ceq(x) = 0$ $A \bullet x \le b$ $Aeq \bullet x = beq$ $l \le x \le u$	fmincon
	半无限问题	$\min_x f(x)$ $K(x, w) \le 0$ (对于所有的 w) $c(x) \le 0$ $ceq(x) = 0$ $A \bullet x \le b$ $Aeq \bullet x = beq$ $l \le x \le u$	fseminf
	0-1 规划	$\min_x f^T x$ $A \bullet x \le b$ $Aeq \bullet x = beq$ x 为二进制	bintprog

Пример 1

Find values of x that minimize f(x) = –x1*x2*x3, starting at the point x = [10;10;10], subject to the constraints: 0 ≤ x1 + 2x2 + 2x3 ≤ 72.

Write a file that returns a scalar value f of the objective function evaluated at x:

```
function f = myfun(x)
f = -x(1) * x(2) * x(3);
```

Rewrite the constraints as both less than or equal to a constant,

–x1–2x2–2x3 ≤ 0

x1 + 2x2 + 2x3≤ 72

Since both constraints are linear, formulate them as the matrix inequality A·x ≤ b, where

```
A = [-1 -2 -2; ...
     1  2  2];
b = [0;72];
```

Supply a starting point and invoke an optimization routine:

```
x0 = [10;10;10];    % Starting guess at the solution
[x,fval] = fmincon(@myfun,x0,A,b);
```

Пример 2

find values of x that minimize $f(x) = \frac{1}{2}x_1^2 + x_2^2 - x_1x_2 - 2x_1 - 6x_2$, subject to

x1 + x2 ≤ 2

–x1 + 2x2 ≤ 2

2x1 + x2 ≤ 3

0 ≤ x1, 0 ≤ x2.

In matrix notation this is $f(x) = \frac{1}{2}x^T Hx + f^T x,\quad H = \begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix},\ f = \begin{bmatrix} -2 \\ -6 \end{bmatrix},\ x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}.$

- **Enter the coefficient matrices:**

```
H = [1 -1; -1 2];
f = [-2; -6];
A = [1 1; -1 2; 2 1];
b = [2; 2; 3];
lb = zeros(2,1);
```

- [x,fval] = quadprog(H,f,A,b,[],[],lb,[],[])

课后作业

作业：给定 k(x,t) 和 g(t), 用最小二乘法求解如下两个问题：

问题1：

$$f(x)+\int_0^1 k(x,t)\,f(x)\,dx=g(t)$$

问题2：

$$\int_0^1 k(x,t)\,f(x)\,dx=g(t)$$