凸优化理论与应用

第一章
凸集

信息与通信工程学院 庄伯金
bjzhuang@bupt.edu.cn

仿射集(Affine sets)

- 直线的表示：

$$y=\theta x_1+(1-\theta)x_2,\ \theta\in\mathcal{R}.$$

- 线段的表示：

$$y=\theta x_1+(1-\theta)x_2,\ \theta\in[0,1].$$

- 仿射集的定义：过集合C内任意两点的直线均在集合C内，则称集合C为仿射集。
- 仿射集的例：直线、平面、超平面

$$Ax=b$$

仿射集

- 仿射包：包含集合C的最小的仿射集。

$$\operatorname{aff} C = \{\sum \theta_i x_i \mid x_i \in C, \sum \theta_i = 1\}$$

- 仿射维数：仿射包的维数。

- 相对内点（relative interior）：

$$\operatorname{relint} C = \{x \mid B(x, r) \cap \operatorname{aff} C \subseteq C, r > 0\}$$

凸集(Convex Sets)

- **凸集的定义：集合C内任意两点间的线段均在集合C内，则称集合C为凸集。**

$$\forall x_1, x_2 \in C, \theta \in [0,1], \text{则} \, \theta x_1 + (1-\theta) x_2 \in C$$

$$\forall x_1, \ldots, x_k \in C, \theta_i \in [0,1] \text{且} \sum_{i=1}^{k} \theta_i = 1,$$

$$\text{则} \sum_{i=1}^{k} \theta_i x_i \in C$$

信息与通信工程学院 庄伯金
bjzhuang@bupt.edu.cn

凸集

- 凸包的定义：包含集合C的最小的凸集。

$$\operatorname{conv} C = \{\sum_{i=1}^{k} \theta_i x_i \mid x_i \in C, \theta_i \geq 0, \sum_{i=1}^{k} \theta_i = 1\}$$

锥(Cones)

- 锥的定义：

$$\forall x \in C, \theta \geq 0, \text{则有}\ \theta x \in C.$$

- 凸锥的定义：集合C既是凸集又是锥。

$$\forall x_1, x_2 \in C, \theta_1, \theta_2 \geq 0, \text{则有}\ \theta_1 x_1 + \theta_2 x_2 \in C.$$

- 锥包的定义：集合C内点的所有锥组合。

$$\{\sum_{i=1}^{k} \theta_i x_i \mid x_i \in C, \theta_i \geq 0\}$$

信息与通信工程学院 庄伯金
bjzhuang@bupt.edu.cn

超平面和半空间

- **超平面(hyperplane)：** $\{x \mid a^T x = b\}$

- **半空间(Halfspace):** $\{x \mid a^T x \le b\}$ $\{x \mid a^T x \ge b\}$

欧氏球和椭球

- **欧氏球(euclidean ball):**

$$B(x_c, r) = \{x \mid \|x - x_c\|_2 \le r\}$$
$$= \{x \mid (x - x_c)^T (x - x_c) \le r^2\}$$

- **椭球(ellipsoid):**

$$E = \{x \mid (x - x_c)^T P^{-1} (x - x_c) \le r^2\},\ P\text{为对称正定矩阵}$$

范数球和范数锥

- **范数(norm)**： $\|x\| \ge 0, \|x\| = 0$当且仅当$x = 0$;

  $\|tx\| = |t|\,\|x\|, t \in \mathcal{R}$;

  $\|x + y\| \le \|x\| + \|y\|$

- **范数球(norm ball)**：

$$B(x_c, r) = \{x \mid \|x - x_c\| \le r\}$$

- **范数锥(norm cone)**：

$$\{(x, t) \mid \|x\| \le t\}$$

信息与通信工程学院 庄伯金
bjzhuang@bupt.edu.cn

多面体(Polyhedra)

- 多面体： $P=\{x \mid a_j^T x \le b_j, c_i^T x = d_i\}$
- 单纯形(simplex)：

$$\{\sum_{i=0}^{k}\theta_i v_i \mid \theta_i \ge 0, \sum_{i=0}^{k}\theta_i = 1, v_1 - v_0, \ldots, v_k - v_0 \text{线性无关}\}$$

半正定锥(Positive semidefinite cone)

- **n阶对称矩阵集：**

$$S^n = \{X \in \mathcal{R}^{n\times n} \mid X = X^T\}$$

- **n阶半正定矩阵集：**

$$S_+^n = \{X \in S^n \mid X \succ= 0\}$$

- **n阶正定矩阵集：**

$$S_{++}^n = \{X$$

n阶半正定矩阵集为凸锥！

信息与通信工程学院 庄伯金
bjzhuang@bupt.edu.cn

保持凸性的运算

- 集合交运算
- 仿射变换
- 透视函数(perspective function)

$$P(z,t) = z/t, z \in \mathcal{R}^n, t \in \mathcal{R}_{++}$$

- 线性分式函数(linear-fractional function)

$$f(x) = (Ax+b)/(c^T x + d)$$

$$A \in \mathcal{R}^{m\times n}, b \in \mathcal{R}^m, c \in \mathcal{R}^n, d \in \mathcal{R}, c^T x + d > 0$$

真锥(proper cone)

- **真锥的定义：锥 $K \subset R^n$ 满足如下条件**

  1.K为凸集；

  2. K为闭集；

  3. K非中空；

  4.K有端点。

K具有内点

K内不含直线

广义不等式

- 真锥 K下的偏序关系：

$$x \preceq_K y \Leftrightarrow y - x \in K$$

广义不等式

$$x \prec_K y \Leftrightarrow y - x \in \operatorname{int} K$$

严格广义不等式

- 例：
  - 逐项不等式
  - 矩阵不等式

广义不等式的性质

$1.\ x \preceq_K x;$

$2.\ x \preceq_K y, y \preceq_K x \Rightarrow x = y;$

$3.\ x \preceq_K y, y \preceq_K z \Rightarrow x \preceq_K z;$

$4.\ x \preceq_K y, u \preceq_K v \Rightarrow x + u \preceq_K y + v;$

$5.\ x \preceq_K y, \alpha > 0 \Rightarrow \alpha x \preceq_K \alpha y;$

$6.\ x_i \preceq_K y_i, \lim x_i = x, \lim y_i = y \Rightarrow x \preceq_K y.$

严格广义不等式的性质

$1. x \prec_K y \Rightarrow x \preceq_K y;$

$2. x \nprec_K x;$

$3. x \prec_K y, u \preceq_K v \Rightarrow x + u \prec_K y + v;$

$4. x \prec_K y, \alpha > 0 \Rightarrow \alpha x \prec_K \alpha y$

$5. x \prec_K y, u$足够小$\Rightarrow x + u \prec_K y.$

信息与通信工程学院 庄伯金
bjzhuang@bupt.edu.cn

最值和极值

- 最小元的定义：设 $x \in S$ ，对$\forall y \in S$，都有 $x \prec=_K y$ 成立，则称 x 为S 的最小元。

- 极小元的定义：设 $x \in S$，对于$y \in S$ ，若 $y \prec=_K x$，则 $y = x$ 成立，则称 x 为 S 的极小元。

信息与通信工程学院 庄伯金
bjzhuang@bupt.edu.cn

分割超平面(separating hyperplane)

- 定理：设C 和 D 为两不相交凸集，则存在超平面将C和D分离。即：

$$\forall x \in C, a^T x \le b \text{且} \forall x \in D, a^T x \ge b.$$

支撑超平面(supporting hyperplane)

- 定义：设集合 C，x_0 为 C 边界上的点。若存在 $a \neq 0$，满足对任意 $x \in C$，都有 $a^T x \leq a^T x_0$ 成立，则称超平面 $\{x \mid a^T x = a^T x_0\}$ 为集合 C 在点 x_0 处的支撑超平面。

- 定理：凸集边界上任意一点均存在支撑超平面。
- 定理：若一个闭的非中空集合，在边界上的任意一点存在支撑超平面，则该集合为凸集。

对偶锥(dual cone)

- 对偶锥的定义：设 K为锥，则集合

$$K^* = \{y \mid x^T y \ge 0, \forall x \in K\}$$

称为对偶锥。

- 对偶锥的性质：

真锥的对偶锥仍然是真锥！

1.K^*是闭凸集；

2.若K非中空，则K^*有端点；

3.若K的闭包有端点，则K^*非中空；

4. K^{**}是K的闭凸包；

信息与通信工程学院 庄伯金
bjzhuang@bupt.edu.cn

对偶广义不等式

- **广义不等式与对偶等价性质**

$$x \preceq_K y \Leftrightarrow \lambda^T x \le \lambda^T y, \text{for all } \lambda \succeq_{K^*} 0;$$

$$x \prec_K y \Leftrightarrow \lambda^T x \le \lambda^T y, \text{for all } \lambda \succeq_{K^*} 0, \lambda \ne 0.$$

- **最小元的对偶特性：**

x为集合S中关于K偏序的最小元

$\Leftrightarrow$ 对所有$\lambda \succ_{K^*} 0$, x为使$\lambda^T z, z \in S$最小的值.

对偶广义不等式

- **极小元的对偶特性**

$\lambda \succ_{K^*} 0$, x为使$\lambda^T z, z \in S$最小的值 $\Rightarrow$ x为极小元.

反过来不一定成立！

信息与通信工程学院 庄伯金
bjzhuang@bupt.edu.cn

作业(1)

- **P60 2.8**
- **P60 2.10**
- **P60 2.14**

作业(2)

- **P62 2.16**
- **P62 2.18**
- **P64 2.30**
- **P64 2.31**
- **P64 2.32**

凸优化理论与应用

第二章　凸函数

信息与通信工程学院 庄伯金
bjzhuang@bupt.edu.cn

凸函数的定义

- 凸函数的定义：函数 $f:\mathcal{R}^n \to \mathcal{R}$，满足

  1.定义域 dom f为凸集；

  2.$\forall x, y \in \mathrm{dom}\, f, 0 \le \theta \le 1$，有

$$f(\theta x + (1-\theta) y) \le \theta f(x) + (1-\theta) f(y).$$

- 凸函数的扩展定义：若 f 为凸函数，
展函数 $\tilde{f}:\mathcal{R}^n \to \mathcal{R} \cup \{\infty\}$ 为

$$\tilde{f}(x) = \begin{cases} f(x) & x \in \mathrm{dom} f \\ \infty & x \notin \mathrm{dom} f \end{cases}$$

凸函数的扩展函数也是凸函数！

信息与通信工程学院 庄伯金
bjzhuang@bupt.edu.cn

凸函数的一阶微分条件

- 若函数 f 的定义域 dom$f$为开集，且函数 f 一阶可微，则函数 f为凸函数当且仅当 dom$f$ 为凸集，且对

  $\forall x, y \in \mathrm{dom} f$

$$f(y) \geq f(x) + \nabla f(x)^T (y - x)$$

凸函数的二阶微分条件

- 若函数 f 的定义域 $\mathrm{dom}f$ 为开集，且函数 f 二阶可微，则函数 f 为凸函数当且仅当 $\mathrm{dom}f$ 为凸集，且对 $\forall x \in \mathrm{dom}f$ ，其**Hessian**矩阵

$$\nabla^2 f(x) \succ= 0.$$

凸函数的例

- 指数函数 e^{ax}
- 幂函数 $x^a, x \in \mathcal{R}_{++}, a \geq 1 \text{ or } a \leq 0.$
- 负对数函数 $-\log x$
- 负熵函数 $x \log x$
- 范数函数 $\|x\|_p$

凸函数的例

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

$$f(x) = x^2 / y$$

$$f(x) = \log(e^{x_1} + \ldots + e^{x_n})$$

$$f(x) = (\prod_{i=1}^{n} x_i)^{1/n}, \mathrm{dom} f = \mathcal{R}_{++}^n$$

$$f(X) = \log(\det X), \mathrm{dom} f = S_{++}^n$$

下水平集(sublevel set)

- 定义：集合

$$C_\alpha = \{x \in \mathrm{dom} f \mid f(x) \le \alpha\}$$

称为 f 的 α 下水平集。

- 定理：凸函数的任一下水平集均为凸集。

- 任一下水平集均为凸集的函数不一定为凸函数。

函数上半图(epigraph)

- 定义：集合

$$\mathrm{epi}f = \{(x,t) \mid x \in \mathrm{dom}f,\ f(x) \le t\}$$

称为函数 f 的上半图。

- 定理：函数 f 为凸函数当且仅当 f 的上半图为凸集。

信息与通信工程学院 庄伯金
bjzhuang@bupt.edu.cn

Jensen不等式

- f 为凸函数，则有：

$$f(\theta_1 x_1 + \ldots + \theta_n x_n) \le \theta_1 f(x_1) + \ldots + \theta_n f(x_n)$$

其中$0 \le \theta_i \le 1, \theta_1 + \ldots + \theta_n = 1.$

- **Jensen不等式的另外形式：**

$$f(\int_S p(x) x dx) \le \int_S p(x) f(x) dx.$$

保持函数凸性的算子

- 凸函数的非负加权和

$$f(x) = \omega_1 f_1(x) + \ldots + \omega_n f_n(x)$$

- 凸函数与仿射变换的复合

$$g(x) = f(Ax + b)$$

- 凸函数的逐点最大值

$$f(x) = \max(f_1(x), \ldots, f_n(x))$$

$$f(x) = \sup_{y \in \mathcal{A}} g(x, y)$$

保持函数凸性的算子

- 复合运算

$$g:\mathcal{R}^n \to \mathcal{R}, h:\mathcal{R} \to \mathcal{R}$$

$$f(x) = h(g(x))$$

- 最小值算子

$$g(x) = \inf_{y\in C} f(x, y)$$

- 凸函数的透视算子

$$g(x, t) = tf(x/t)$$

共轭函数(conjugate function)

- 定义：设函数 $f:\mathcal{R}^n \to \mathcal{R}$ ，其共轭函数 $f^*:\mathcal{R}^n \to \mathcal{R}$ ，定义为

$$f^*(y) = \sup_{x \in \mathrm{dom} f} (y^T x - f(x)).$$

共轭函数具有凸性！

- 共轭函数的例

共轭函数的性质

- **Fenchel's inequality**

$$f(x)+f^*(y)\geq y^Tx.$$

- 性质：若 $f(x)$为凸函数，且 $f(x)$的上半图是闭集，则有

$$f^{**}=f.$$

- 性质：设 $f(x)$ 为凸函数，且可微，对于 $z\in R^n$ ，若

$$y=\nabla f(z)$$

则
$$f^*(y)=z^T\nabla f(z)-f(z)$$

准凸函数(quasiconvex function)

- 定义：设函数 $f:\mathcal{R}^n \to \mathcal{R}$ ，若函数的定义域和任意下水平集

$$S_\alpha = \{x \mid f(x) \le \alpha, x \in \mathrm{dom} f\}$$

则称函数 $f(x)$ 为准凸函数。

- 准凸函数的例

准凸函数的判定定理

- 定理：函数 $f(x)$为准凸函数，当且仅当 $\mathrm{dom}f$ 为凸集，且对 $\forall x, y \in \mathrm{dom}f, 0 \le \theta \le 1$，有
$$f(\theta x + (1-\theta)y) \le \max\{f(x), f(y)\}$$
- 定理：若函数 $f(x)$ 一阶可微，则 $f(x)$ 为准凸函数，当且仅当 $\mathrm{dom}f$ 为凸集，且对 $\forall x, y \in \mathrm{dom}f$ ，有
$$f(y) \le f(x) \Rightarrow \nabla^T f(x)(y-x) \le 0$$
- 定理：若函数 $f(x)$ 二阶可微，且满足对
$\forall x \in \mathrm{dom}f, y \in \mathcal{R}^n, y \ne 0$ ，有
$$y^T \nabla f(x) = 0 \Rightarrow y^T \nabla^2 f(x) y > 0$$
则函数 $f(x)$准凸函数。

保持准凸性的算子

- 非负权值函数的最大值函数

- 复合函数

- 最小值函数

准凸函数的凸函数族表示

- 若 $f(x)$ 为准凸函数，根据 $f(x)$ 的任意 t 下水平集，我们可以构造一个凸函数族 $\phi_t(x)$，使得

$$f(x) \le t \Leftrightarrow \phi_t(x) \le 0$$

- 性质：若 $\phi_t(x)$ 为准凸函数 $f(x)$ 的凸函数族表示，对每一个 $x \in \mathrm{dom} f$，若 $s \ge t$，则有

$$\phi_s(x) \le \phi_t(x).$$

对数凸函数

- 定义：函数$f(x)$称为对数凸函数，若函数 $f(x)$满足：

  1.dom$f$ 为凸集

  2.$f(x)>0$

  3.$\log f(x)$ 为凸函数。

- 定理：函数$f(x)$的定义域为凸集，且$f(x)>0$，则$f(x)$为对数凸函数，当且仅当对$\forall x, y \in \mathrm{dom} f, 0\le\theta\le 1$有

$$f(\theta x+(1-\theta)y)\le f(x)^{\theta} f(y)^{1-\theta}$$

- 对数凸函数的例

对数凸函数和凹函数的性质

- 定理：函数 $f(x)$ 二阶可微，则 $f(x)$ 为对数凸函数当且仅当

$$f(x)\nabla^2 f(x) \succ= \nabla f(x)\nabla f(x)^T$$

- 性质：对数凸性与凹性对函数乘积和正数数乘运算均保持封闭。
- 性质：对数凸性对函数加运算保持封闭。但对数凹性对函数加运算不封闭。
- 推论：函数 $f(x,y)$ 对每一个 $y \in C$ 在 x 上对数凸，则函数 $g(x)=\int_C f(x,y)dy$ 也是对数凸函数。

对数凸函数和凹函数的性质

- 定理：函数 $f(x,y):\mathcal{R}^n\times\mathcal{R}^m\to\mathcal{R}$ 为对数凹函数，则函数 $g(x)=\int f(x,y)dy$是对数凹函数。

广义不等式下的凸性

- 广义单调性的定义：设$K \subseteq \mathcal{R}^n$为真锥，函数$f:\mathcal{R}^n \to \mathcal{R}$称为$K-$单调增，若函数$f(x)$满足：

$$x \preceq_K y \Rightarrow f(x) \le f(y)$$

- 广义凸函数的定义：设$K \subseteq \mathcal{R}^m$为真锥，函数$f:\mathcal{R}^n \to \mathcal{R}^m$称为$K-$凸，若函数$f(x)$满足对$\forall x, y \in \mathrm{dom} f, 0 \le \theta \le 1$均有 $f(\theta x + (1-\theta)y) \preceq_K \theta f(x) + (1-\theta) f(y)$.
- 定理(对偶等价)：函数$f(x)$为$K-$凸函数，当且仅当对所有$w \succeq_{K^*} 0$，$w^T f(x)$为凸函数。

信息与通信工程学院 庄伯金
bjzhuang@bupt.edu.cn

练习

- **P116 3.16**
- **P120 3.53**
- **P120 3.41**
- **P120 3.49 (1) (2)**

凸优化理论与应用

第三章　凸优化

信息与通信工程学院 庄伯金
bjzhuang@bupt.edu.cn

优化问题的基本形式

- 优化问题的基本描述：

$$\begin{aligned} &\text{minimize} \quad f_0(x), \quad x \in \mathcal{R}^n \\ &\text{subject to} \quad f_i(x) \le 0, \quad i = 1, \dots, m \\ &\qquad\qquad\quad\; h_i(x) = 0, \quad j = 1, \dots, p \end{aligned}$$

优化变量 $x \in \mathcal{R}^n$

不等式约束 $f_i(x) \le 0$

等式约束 $h_i(x) = 0$

无约束优化 $m = p = 0$

优化问题的基本形式

优化问题的域 $D = \bigcap_{i=0}^{m} \mathrm{dom} f_i \cap \bigcap_{i=1}^{p} \mathrm{dom} h_i$

可行点(解) (feasible) $x \in D$ 满足约束条件

可行域(可解集) 所有可行点的集合

最优化值

$$p^* = \inf\{ f_0(x) \mid f_i(x) \le 0, i = 1, \dots, m, h_i(x) = 0, i = 1, \dots, p\}$$

最优化解 $p^* = f_0(x^*)$

局部最优问题

- **局部最优问题**

$$
\begin{aligned}
\text{minimize} \quad & f_0(x), \quad x \in \mathcal{R}^n \\
\text{subject to} \quad & f_i(x) \le 0, \quad i = 1, \ldots, m \\
& h_i(x) = 0, \quad j = 1, \ldots, p \\
& \|x - z\|_2 \le R, \; R > 0
\end{aligned}
$$

优化问题的等价形式(1)

- 定理：若 $\alpha_i > 0, i = 0, \ldots, m, \beta_i \neq 0, i = 1, \ldots, p$
则原优化问题与以下优化问题等价

$$
\begin{array}{ll}
\text{minimize} & \tilde{f}_0(x) = \alpha_0 f_0(x), \quad x \in \mathcal{R}^n \\
\text{subject to} & \tilde{f}_i(x) = \alpha_i f_i(x) \leq 0, \quad i = 1, \ldots, m \\
& \tilde{h}_i(x) = h_i(x) = 0, \quad i = 1, \ldots, p
\end{array}
$$

信息与通信工程学院 庄伯金
bjzhuang@bupt.edu.cn

优化问题的等价形式(2)

- **定理：设 $\phi:\mathcal{R}^n \to \mathcal{R}^n$ 为一一对应，且 $D \subseteq \phi(\mathrm{dom}\,\phi)$ 则原优化问题与以下优化问题等价**

$$
\begin{array}{ll}
\text{minimize} & \tilde{f}_0(z) = f_0(\phi(z)),\;\; z \in \mathcal{R}^n \\
\text{subject to} & \tilde{f}_i(z) = f_i(\phi(z)) \le 0,\;\; i = 1,\dots,m \\
& \tilde{h}_i(z) = h_i(\phi(z)) = 0,\;\; i = 1,\dots,p
\end{array}
$$

信息与通信工程学院 庄伯金
bjzhuang@bupt.edu.cn

优化问题的等价形式(3)

- **定理：设$\psi_0:\mathcal{R}\to\mathcal{R}$ 为严格单调增函数；$\psi_1,\ldots,\psi_m$ 满足$\psi_i(u)\le 0$ 当且仅当$u\le 0$；$\omega_1,\ldots,\omega_p$ 满足$\omega_i(u)=0$ 当且仅当$u=0$。则原优化问题与以下优化问题等价**

$$
\begin{aligned}
\text{minimize}\quad & \tilde{f}_0(z)=\psi_0(f_0(z)),\quad z\in\mathcal{R}^n\\
\text{subject to}\quad & \tilde{f}_i(z)=\psi_i(f_i(z))\le 0,\quad i=1,\ldots,m\\
& \tilde{h}_i(z)=\omega_i(h_i(z))=0,\quad \mathrm{i}=1,\ldots,p
\end{aligned}
$$

优化问题的等价形式(4)

- 定理：原优化问题与以下优化问题等价

$$
\begin{aligned}
\text{minimize} \quad & f_0(x), \quad x \in \mathcal{R}^n \\
\text{subject to} \quad & f_i(x) + s_i = 0, \quad i = 1, \ldots, m \\
& s_i \geq 0 \\
& h_i(x) = 0, \quad j = 1, \ldots, p
\end{aligned}
$$

- s 称为松弛变量

优化问题的等价形式(5)

- 定理：设 $\phi:\mathcal{R}^k \to \mathcal{R}^n$ 满足等式 $h_i(x)=0,\ \ j=1,\dots,p$ 成立，当且仅当 $x=\phi(z)$ 。则原优化问题与以下优化问题等价

$$
\begin{array}{ll}
\text{minimize} & f_0(\phi(z)),\ \ x\in\boldsymbol{\mathcal{R}}^n \\
\text{subject to} & f_i(\phi(z))\le 0,\ \ i=1,\dots,m
\end{array}
$$

可分离变量优化问题

- 性质：$$\inf_{x,y} f(x,y) = \inf_x \tilde{f}(x)$$

  其中 $$\tilde{f}(x) = \inf_y f(x,y)$$

- 定理：优化问题

$$\begin{array}{ll} \text{minimize} & f_0(x_1, x_2), \quad x \in \mathcal{R}^n \\ \text{subject to} & f_i(x_1) \le 0, \quad i = 1, \dots, m_1 \\ & \tilde{f}_i(x_2) \le 0, \quad i = 1, \dots, m_2 \end{array}$$

可以分离变量 x_1, x_2

信息与通信工程学院 庄伯金
bjzhuang@bupt.edu.cn

优化问题的上半图形式

$$
\begin{array}{ll}
\text{minimize} & t \\
\text{subject to} & f_0(x) - t \le 0, \\
& f_i(x) \le 0, \quad i = 1, \ldots, m \\
& h_i(x) = 0, \quad j = 1, \ldots, p
\end{array}
$$

信息与通信工程学院 庄伯金
bjzhuang@bupt.edu.cn

凸优化问题的基本形式

- 凸优化问题的基本描述：

$$\begin{aligned} \text{minimize} \quad & f_0(x), \quad x \in \mathcal{R}^n \\ \text{subject to} \quad & f_i(x) \le 0, \quad i = 1, \dots, m \\ & h_i(x) = 0, \quad j = 1, \dots, p \end{aligned}$$

$f_i(x)$为凸函数

$h_i(x)$为仿射函数

若 $f_0(x)$为准凸函数，则优化问题称为准凸优化问题。

性质：凸优化问题的可行域是凸集。

凸优化问题的例

- 例：

$$\begin{aligned} \text{minimize} \quad & f_0(x) = x_1^2 + x_2^2 \\ \text{subject to} \quad & f_1(x) = x_1/(1+x_2^2) \le 0 \\ & h_1(x) = (x_1 + x_2)^2 = 0 \end{aligned}$$

等价于凸优化问题

$$\begin{aligned} \text{minimize} \quad & f_0(x) = x_1^2 + x_2^2 \\ \text{subject to} \quad & \tilde{f}_1(x) = x_1 \le 0 \\ & \tilde{h}_1(x) = x_1 + x_2 = 0 \end{aligned}$$

凸优化问题的局部最优解

- 定理：凸优化问题的局部最优解均是全局最优解。

凸优化问题最优解的微分条件

- 定理：设 X 为凸优化问题的可行域，$f_0(x)$可微。则 x 为最优解当且仅当$\nabla f_0(x)^T(y-x)\geq 0, y\in X$成立。

- 定理：非约束凸优化问题中，若$f_0(x)$可微。则 x 为最优解当且仅当 $\nabla f_0(x)=0$成立。

信息与通信工程学院 庄伯金
bjzhuang@bupt.edu.cn

凸优化问题的等价形式

- 定理：设凸优化问题仅有等式约束

$$\begin{aligned} &\text{minimize} \quad f_0(x), \quad x \in \mathcal{R}^n \\ &\text{subject to} \quad Ax = b \end{aligned}$$

则 x 为最优解当且仅当 $x \in X$ ，且存在向量 v 满足

$$\nabla f_0(x) + A^T v = 0$$

信息与通信工程学院 庄伯金
bjzhuang@bupt.edu.cn

凸优化问题的等价形式

- **定理：设凸优化问题**

$$\text{minimize} \quad f_0(x), \quad x \in \mathcal{R}^n$$

$$\text{subject to} \quad x \succ= 0$$

则 x **为最优解当且仅当** $x \in X$ **，且**

$$\nabla f_0(x)^T x = 0$$

凸优化问题的等价形式

- **定理：设凸优化问题**

$$\begin{aligned} \text{minimize} \quad & f_0(x), \quad x \in \mathcal{R}^n \\ \text{subject to} \quad & f_i(x) \le 0, \quad i = 1, \ldots, m \\ & A^T x = b \end{aligned}$$

等价于

$$\begin{aligned} \text{minimize} \quad & f_0(Fz + x_0), \quad x \in \mathcal{R}^n \\ \text{subject to} \quad & f_i(Fz + x_0) \le 0, \quad i = 1, \ldots, m \end{aligned}$$

其中

$$A^T x = b \Leftrightarrow x = Fz + x_0$$

信息与通信工程学院 庄伯金
bjzhuang@bupt.edu.cn

凸优化问题的等价形式

- **定理：设凸优化问题**

$$\begin{aligned} \text{minimize} \quad & f_0(x), \quad x \in \mathcal{R}^n \\ \text{subject to} \quad & a_i^T x \le b_i, \quad i = 1, \ldots, m \end{aligned}$$

等价于

$$\begin{aligned} \text{minimize} \quad & f_0(x), \quad x \in \mathcal{R}^n \\ \text{subject to} \quad & a_i^T x + s_i = b_i, \quad i = 1, \ldots, m \\ & s_i \ge 0, \quad i = 1, \ldots, m \end{aligned}$$

准凸优化问题

- **准凸优化问题的基本描述**

$$\begin{aligned} \text{minimize} \quad & f_0(x), \quad x \in \mathcal{R}^n \\ \text{subject to} \quad & f_i(x) \le 0, \quad i = 1, \ldots, m \\ & h_i(x) = 0, \quad j = 1, \ldots, p \end{aligned}$$

$f_0(x)$为准凸函数，$f_1(x), \ldots, f_m(x)$为凸函数。

- 注：准凸优化问题的局部最优解不一定是全局最优解。

准凸优化问题的上半图形式

- 设 $\phi_t(x)$为准凸函数 $f_0(x)$ 的凸函数族表示，即

$$f_0(x) \le t \Leftrightarrow \phi_t(x) \le 0$$

则准凸优化问题的可行解问题为

$$\begin{array}{ll} \text{find} & x \\ \text{subject to} & \phi_t(x) \le 0, \\ & f_i(x) \le 0, \quad i = 1, \ldots, m \\ & Ax = b \end{array}$$

- 设 p^* 为准凸优化问题的最优解，若上述问题可解，则 $p^* \le t$ 。否则 $p^* \ge t$ 。

信息与通信工程学院 庄伯金
bjzhuang@bupt.edu.cn

准凸优化问题二分法求解

- 给定一个足够小的 l 和足够大的 u，使得区间 $[l,u]$ 能包含最优解 p^*。给定 $\varepsilon > 0$
- **LOOP**：
  - 令 $t = (l+u)/2$
  - 求解可行解问题；
  - 若可解，则令 $u = t$，否则令 $l = t$
  - 若 $|u - l| < \varepsilon$，则结束，否则**goto LOOP**。

信息与通信工程学院 庄伯金
bjzhuang@bupt.edu.cn

线性规划(linear program,LP)

- **LP问题的一般描述**

$$
\begin{aligned}
\text{minimize} \quad & c^T x + d \\
\text{subject to} \quad & Gx \preceq h \\
& Ax = b
\end{aligned}
$$

$$G \in \mathcal{R}^{m\times n}, A \in \mathcal{R}^{p\times n}$$

LP问题的几种类型

- **标准LP问题**

$$
\begin{aligned}
\text{minimize} \quad & c^T x \\
\text{subject to} \quad & Ax = b \\
& x \succ= 0
\end{aligned}
$$

- **不等式形式LP问题**

$$
\begin{aligned}
\text{minimize} \quad & c^T x + d \\
\text{subject to} \quad & Ax \prec= b
\end{aligned}
$$

信息与通信工程学院 庄伯金
bjzhuang@bupt.edu.cn

标准LP问题的转换

$$\text{minimize} \quad c^T x^+ - c^T x^- + b$$

$$\text{subject to} \quad Gx^+ - Gx^- + s = h$$

$$Ax^+ - Ax^- = b$$

$$x^+ \succ= 0, x^- \succ= 0, s \succ= 0$$

信息与通信工程学院 庄伯金
bjzhuang@bupt.edu.cn

LP问题的例

- Chebyshev center of a polyhedron

- Piecewise-linear minimization

- Linear-fractional programming

Chebyshev center of a polyhedron

- 多面体 $P=\{x\in R^n \mid Ax \prec= b\}$
- Chebyshev center：到多面体边界距离最大的内点（最深的点）
- 问题描述

$$\begin{array}{ll} \text{maximize} & r \\ \text{subject to} & B(x_c,r)\subseteq P \end{array}$$

$$B(x_c,r)=\{x_c+u \mid \|u\|_2\le r\}$$

- LP形式

$$\begin{array}{ll} \text{minimize} & 1/r \\ \text{subject to} & a_i^T x_c + r\|a_i\|_2 \le b_i, i=1,\dots,m \end{array}$$

Piecewise-linear minimization

- 问题描述

$$\text{minimize} \quad f(x) = \max_{i=1,\dots,m} (a_i^T x + b_i)$$

- 上半图形式

$$\begin{aligned} &\text{minimize} && t \\ &\text{subject to} && \max_{i=1,\dots,m} (a_i^T x + b_i) \le t \end{aligned}$$

- LP形式

$$\begin{aligned} &\text{minimize} && t \\ &\text{subject to} && a_i^T x + b_i \le t, i = 1, \dots, m \end{aligned}$$

信息与通信工程学院 庄伯金
bjzhuang@bupt.edu.cn

Linear-fractional programming

- 问题描述

$$\text{minimize} \quad f_0(x) = \frac{c^T x + d}{e^T x + f}, \text{dom} f_0 = \{x \mid e^T x + f > 0\}$$

$$\text{subject to} \quad Gx \preceq h$$

$$Ax = b$$

- LP形式

$$\text{minimize} \quad c^T y + dz$$

$$\text{subject to} \quad Gy - hz \preceq 0$$

$$Ay - bz = 0$$

$$e^T y + fz = 1$$

$$z \geq 0$$

$$y = \frac{x}{e^T x + f}$$

$$z = \frac{1}{e^T x + f}$$

信息与通信工程学院 庄伯金
bjzhuang@bupt.edu.cn

二次规划(quadratic program,QP)

- **QP问题的基本描述**

$$
\begin{aligned}
\text{minimize} \quad & (1/2)x^T P x + q^T x + r \\
\text{subject to} \quad & Gx \prec= h \\
& Ax = b \\
& P \in S_+^n, G \in R^{m \times n}, A \in R^{p \times n}
\end{aligned}
$$

二次约束二次规划

- quadratically constrained quadratic program (QCQP)

$$\begin{array}{ll} \text{minimize} & (1/2)x^T P_0 x + q_0^T x + r_0 \\ \text{subject to} & (1/2)x^T P_i x + q_i^T x + r_i \le 0 \\ & Ax = b \\ & P_i \in S_+^n, A \in R^{p\times n} \end{array}$$

QP问题的例

- Least-squares and regression

- Distance between polyhedra

Least-squares and regression

- 问题描述

$$\text{minimize} \quad \|Ax-b\|_2^2$$

$$\|Ax-b\|_2^2 = x^T A^T A x - 2b^T A x + b^T b$$

Distance between polyhedra

- 问题描述

$$\mathrm{dist}(P_1, P_2) = \inf\{\|x_1 - x_2\| \mid x_1 \in P_1, x_2 \in P_2\}$$

$$P_1 = \{x \mid A_1 x \preceq b_1\} \quad P_2 = \{x \mid A_2 x \preceq b_2\}$$

- QP形式

$$\begin{array}{ll} \text{minimize} & \|x_1 - x_2\|_2^2 \\ \text{subject to} & A_1 x \preceq b_1, A_2 x \preceq b_2 \end{array}$$

Second-order cone program, SOCP

- SOCP问题的基本描述

$$\begin{aligned} \text{minimize} \quad & f^T x \\ \text{subject to} \quad & \|A_i x + b\|_2 \le c_i^T x + d_i \\ & Fx = g \end{aligned}$$

- 二次锥约束条件

$$\|Ax + b\|_2 \le c^T x + d$$

SOCP问题的例－Robust linear programming

- 问题描述

$$\begin{aligned} \text{minimize} \quad & c^T x \\ \text{subject to} \quad & a_i^T x \le b_i \end{aligned}$$

其中 c, a_i, b_i 不是完全确定的常数。

- 例：c, b_i 为确定的常数，a_i为变量，其范围满足

$$a_i \in \varepsilon_i = \{\overline{a}_i + P_i u \mid \|u\|_2 \le 1\}$$

SOCP形式

$$\begin{aligned} \text{minimize} \quad & c^T x \\ \text{subject to} \quad & \overline{a}_i^T x + \left\|P_i^T x\right\|_2 \le b_i \end{aligned}$$

几何规划(Geometric programming)

- 几何规划的基本描述

$$\begin{aligned} \text{minimize} \quad & f_0(x) \\ \text{subject to} \quad & f_i(x) \le 1, i = 1, \ldots, m \\ & h_i(x) = 1, i = 1, \ldots, p \end{aligned}$$

f_i为多项式，h_i为单项式, $D = R_{++}^n$

- 单项式与多项式

$$f(x) = c x_1^{a_1} \cdots x_n^{a_n}$$

$$f(x) = \sum_{k=1}^{K} c_k x_1^{a_{1k}} \cdots x_n^{a_{nk}}$$

几何规划的凸形式转换

- 令

$$y_i = \log x_i$$

- 几何规划的凸形式

$$\begin{aligned}
\text{minimize} \quad & \tilde{f}_0(y) = \log\left(\sum_{k=1}^{K_0} e^{a_{0k}^T y + b_{0k}}\right) \\
\text{subject to} \quad & \tilde{f}_i(y) = \log\left(\sum_{k=1}^{K_i} e^{a_{ik}^T y + b_{ik}}\right) \le 0, i = 1, \ldots, m \\
& \tilde{h}_i(y) = g_i^T y + h_i = 0, i = 1, \ldots, p
\end{aligned}$$

广义不等式约束

- 广义不等式约束的优化问题

$$\begin{aligned} \text{minimize} \quad & f_0(x) \\ \text{subject to} \quad & f_i(x) \preceq_{K_i} 0, i = 1, \ldots, m \\ & Ax = b \end{aligned}$$

- SOCP的描述

$$\begin{aligned} \text{minimize} \quad & c^T x \\ \text{subject to} \quad & -(A_i x + b_i, c_i^T x + d_i) \preceq_{K_i} 0, i = 1, \ldots, m \\ & Fx = g \\ & K_i = \{(y, t) \in R^{k_i+1} \mid \|y\|_2 \le t\} \end{aligned}$$

凸优化理论与应用

第四章　对偶问题

信息与通信工程学院 庄伯金
bjzhuang@bupt.edu.cn

优化问题的拉格朗日函数

- 设优化问题：

$$
\begin{aligned}
\text{minimize} \quad & f_0(x), \quad x \in \mathcal{R}^n \\
\text{subject to} \quad & f_i(x) \le 0, \quad i = 1, \ldots, m \\
& h_i(x) = 0, \quad j = 1, \ldots, p
\end{aligned}
$$

- 拉格朗日(**Lagrangian**)函数：

$$L(x, \lambda, \nu) = f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{p} \nu_i h_i(x)$$

- 对固定的 x，拉格朗日函数 $L(x, \lambda, \nu)$ 为关于 λ 和 ν 的仿射函数。

拉格朗日对偶函数

- **拉格朗日对偶函数(lagrange dual function)：**

$$g(\lambda,\nu)=\inf_{x\in D} L(x,\lambda,\nu)=\inf_{x\in D}\left(f_0(x)+\sum_{i=1}^{m}\lambda_i f_i(x)+\sum_{i=1}^{p}\nu_i h_i(x)\right)$$

若拉格朗日函数没有下界，则令

$$g(\lambda,\nu)=-\infty$$

- 拉格朗日对偶函数为凹函数。
- 对$\forall\lambda\succ=0$和 $\forall\nu$，若原最优化问题有最优值p^*，则

$$g(\lambda,\nu)\le p^*$$

信息与通信工程学院 庄伯金
bjzhuang@bupt.edu.cn

对偶函数的例

- **Least-squares solution of linear equations**

- **Standard form LP**

- **Two-way partitioning problem**

Least-squares solution of linear equations

- 原问题：

$$\begin{aligned} &\text{minimize} \quad x^T x, \quad x \in \mathcal{R}^n \\ &\text{subject to} \quad Ax = b \end{aligned}$$

- 拉格朗日函数：

$$L(x,v) = x^T x + v^T(Ax - b)$$

- 拉格朗日对偶函数：

$$g(v) = -\frac{1}{4} v^T A A^T v - b^T v$$

信息与通信工程学院 庄伯金
bjzhuang@bupt.edu.cn

Standard form LP

- 原问题：

$$\begin{aligned} \text{minimize} \quad & c^T x \\ \text{subject to} \quad & Ax = b \\ & x \succ= 0 \end{aligned}$$

- 拉格朗日函数：

$$\begin{aligned} L(x,\lambda,\nu) &= c^T x - \lambda^T x + \nu^T (Ax - b) \\ &= -\nu^T b + (c - \lambda + A^T \nu)^T x \end{aligned}$$

- 拉格朗日对偶函数：

$$g(\lambda,\nu) = \begin{cases} -b^T \nu & A^T \nu - \lambda + c = 0 \\ -\infty & otherwise \end{cases}$$

Two-way partitioning problem

- 原问题：

$$\begin{aligned} &\text{minimize} \quad x^T W x, W \in S^n \\ &\text{subject to} \quad x_i^2 = 1, i = 1, \dots, n \end{aligned}$$

- 拉格朗日函数：

$$\begin{aligned} L(x,\nu) &= x^T W x + \sum_{i=1}^{n} \nu_i (x_i^2 - 1) \\ &= x^T (W + \mathrm{diag}(\nu)) x - 1^T \nu \end{aligned}$$

- 拉格朗日对偶函数：

$$g(\nu) = \begin{cases} -1^T \nu & W + \mathrm{diag}(\nu) \succ= 0 \\ -\infty & otherwise \end{cases}$$

对偶函数与共轭函数

- 共轭函数 $f^*(y) = \sup_{x \in \mathrm{dom} f} (y^T x - f(x))$
- 共轭函数与对偶函数存在密切联系
- 具有线性不等式约束和线性等式约束的优化问题：

$$\begin{aligned} \text{minimize} \quad & f_0(x) \\ \text{subject to} \quad & Ax \preceq b \\ & Cx = d \end{aligned}$$

对偶函数：

$$g(\lambda, \nu) = -b^T \lambda - d^T \nu - f_0^*(-A^T \lambda - C^T \nu)$$

信息与通信工程学院 庄伯金
bjzhuang@bupt.edu.cn

Equality constrained norm minimization

- 问题描述：
$$\begin{array}{ll} \text{minimize} & \|x\| \\ \text{subject to} & Ax=b \end{array}$$

- 共轭函数：
$$f_0^*(y)=\begin{cases} 0 & \|y\|_* \le 1 \\ \infty & otherwise \end{cases}$$

- 对偶函数：
$$g(\nu)=-b^T\nu - f_0^*(-A^T\nu)=\begin{cases} -b^T\nu & \left\|-A^T\nu\right\|_* \le 1 \\ -\infty & otherwise \end{cases}$$

Entropy maximization

- 原始问题：

$$\begin{aligned} \text{minimize} \quad & \sum_{i=1}^{n} x_i \log x_i, \qquad D = R_{+}^{n} \\ \text{subject to} \quad & Ax \preceq b \\ & 1^T x = 1 \end{aligned}$$

- 共轭函数：

$$f_0^*(y) = \sum_{i=1}^{m} e^{y_i - 1}$$

- 对偶函数：

$$\begin{aligned} g(\lambda, \nu) &= -b^T \lambda - \nu - f_0^*(-A^T \lambda - 1\nu) \\ &= -b^T \lambda - \nu - \sum_{i=1}^{m} e^{-a_i^T \lambda - \nu - 1} \\ &= -b^T \lambda - \nu - e^{-\nu - 1} \sum_{i=1}^{m} e^{-a_i^T \lambda} \end{aligned}$$

信息与通信工程学院 庄伯金
bjzhuang@bupt.edu.cn

Minimum volume covering ellipsoid

- 原始问题：

$$\begin{array}{ll} \text{minimize} & \log\det X^{-1}, \quad D = S_{++}^{n} \\ \text{subject to} & a_i^T X a_i \le 1, i = 1, \ldots, m \end{array}$$

- 共轭函数：

$$f_0^*(Y) = \log\det(-Y)^{-1} - n$$

- 对偶函数：

$$g(\lambda) = \begin{cases} \log\det(\sum_{i=1}^{m} \lambda_i a_i a_i^T) - 1^T \lambda + n & \sum_{i=1}^{m} \lambda_i a_i a_i^T \succ 0 \\ -\infty & otherwise \end{cases}$$

信息与通信工程学院 庄伯金
bjzhuang@bupt.edu.cn

拉格朗日对偶问题

- **拉格朗日对偶问题的描述：**

$$\begin{aligned} &\text{maximize} \quad g(\lambda, \nu) \\ &\text{subject to} \quad \lambda \succ= 0 \end{aligned}$$

- **对偶可行域** $\lambda \succ= 0$

$$g(\lambda, \nu) > -\infty$$

- **最优值** d^*

- **最优解** (λ^*, ν^*)

信息与通信工程学院 庄伯金
bjzhuang@bupt.edu.cn

LP问题的对偶问题

- **标准LP问题**
$$\begin{aligned} \text{minimize} \quad & c^T x \\ \text{subject to} \quad & Ax = b \\ & x \succ= 0 \end{aligned}$$

- **对偶函数**
$$g(\lambda, \nu) = \begin{cases} -b^T \nu & A^T \nu - \lambda + c = 0 \\ -\infty & otherwise \end{cases}$$

- **对偶问题**
$$\begin{aligned} \text{minimize} \quad & g(\lambda, \nu) \\ \text{subject to} \quad & \lambda \succ= 0 \end{aligned}$$

- **等价描述**
$$\begin{aligned} \text{minimize} \quad & g(\lambda, \nu) \\ \text{subject to} \quad & A^T \nu - \lambda + c = 0 \\ & \lambda \succ= 0 \end{aligned}$$

信息与通信工程学院 庄伯金
bjzhuang@bupt.edu.cn

弱对偶性

- 定理（弱对偶性）：设原始问题的最优值为 p^*，对偶问题的最优值为d^*，则 $d^* \le p^*$ 成立。
- **optimal duality gap**

$$p^* - d^*$$

- 可以利用对偶问题找到原始问题最优解的下界。

强对偶性

- 定义（强对偶性）：设原始问题的最优值为 p^*，对偶问题的最优值为 d^*。若 $d^* = p^*$ 成立，则称原始问题和对偶问题之间具有强对偶性。
- 强对偶性并不是总是成立的。
- 凸优化问题通常（但并不总是）具有强对偶性。
- **Slater**定理：若凸优化问题存在严格可行解，即存在 $x \in \operatorname{relint} D$，满足

$$f_i(x) < 0, i = 1, \ldots, m,$$

$$Ax = b$$

则优化问题具有强对偶性。该条件称为**Slater**条件。

信息与通信工程学院 庄伯金
bjzhuang@bupt.edu.cn

强对偶性

- 弱化的**Slater**条件：若不等式约束条件的前 k 个为线性不等式约束条件，则**Slater**条件可以弱化为：

  存在 $x \in \operatorname{rel}\operatorname{int} D$，满足

$$
\begin{aligned}
&f_i(x) \le 0, i = 1, \ldots, k, \\
&f_i(x) < 0, i = k+1, \ldots, m, \\
&Ax = b
\end{aligned}
$$

Least-squares solution of linear equations

- 原问题：

$$\begin{aligned} &\text{minimize} \quad x^T x, \quad x \in \mathcal{R}^n \\ &\text{subject to} \quad Ax = b \end{aligned}$$

- 对偶问题：

$$\text{maximize} \quad g(v) = -\frac{1}{4} v^T A A^T v - b^T v$$

- 具有强对偶性

Lagrange dual of QCQP

- **QCQP：**
$$\begin{array}{ll}\text{minimize} & (1/2)x^T P_0 x + q_0^T x + r_0 \\ \text{subject to} & (1/2)x^T P_i x + q_i^T x + r_i \le 0\end{array}$$

- **拉格朗日函数：**

$$L(x,\lambda) = \frac{1}{2}x^T P(\lambda)x + q(\lambda)^T x + r(\lambda)$$

$$P(\lambda) = P_0 + \sum_{i=1}^{m}\lambda_i P_i \quad q(\lambda) = q_0 + \sum_{i=1}^{m}\lambda_i q_i \quad r(\lambda) = r_0 + \sum_{i=1}^{m}\lambda_i r_i$$

- **对偶函数：**

$$g(\lambda) = \inf_x L(x,\lambda) = -\frac{1}{2}q(\lambda)^T P(\lambda)^{-1} q(\lambda) + r(\lambda)$$

信息与通信工程学院 庄伯金
bjzhuang@bupt.edu.cn

Lagrange dual of QCQP

- 对偶问题：

$$\begin{aligned} &\text{maximize} \quad -\frac{1}{2}q(\lambda)^T P(\lambda)^{-1} q(\lambda) + r(\lambda) \\ &\text{subject to} \quad \lambda \succ= 0 \end{aligned}$$

- **Slater条件**：存在 x，满足

$$(1/2)x^T P_i x + q_i^T x + r_i < 0, i = 1, \ldots, m$$

Entropy maximization

- 原始问题：

$$\begin{aligned} &\text{minimize} \quad \sum_{i=1}^{n} x_i \log x_i, \qquad D = R_+^n \\ &\text{subject to} \quad Ax \preceq b \\ &\qquad\qquad\quad\; 1^T x = 1 \end{aligned}$$

- 对偶函数：

$$g(\lambda, \nu) = -b^T\lambda - \nu - e^{-\nu-1}\sum_{i=1}^{n} e^{-a_i^T\lambda}$$

- 对偶问题：

$$\begin{aligned} &\text{maximize} \quad -b^T\lambda - \nu - e^{-\nu-1}\sum_{i=1}^{n} e^{-a_i^T\lambda} \\ &\text{subject to} \quad \lambda \succeq 0 \end{aligned}$$

Entropy maximization

- 弱化的**Slater**条件：存在$x \succ 0$ ，满足

$$Ax \prec= b$$

$$1^T x = 1$$

Minimum volume covering ellipsoid

- 原始问题：

$$\begin{array}{ll} \text{minimize} & \log\det X^{-1}, \quad D = S_{++}^{n} \\ \text{subject to} & a_i^T X a_i \le 1, i = 1, \ldots, m \end{array}$$

- 对偶函数：

$$g(\lambda) = \begin{cases} \log\det(\sum_{i=1}^{m} \lambda_i a_i a_i^T) - 1^T \lambda + n & \sum_{i=1}^{m} \lambda_i a_i a_i^T \succ 0 \\ -\infty & otherwise \end{cases}$$

- 对偶问题：

$$\begin{array}{ll} \text{maximize} & \log\det(\sum_{i=1}^{m} \lambda_i a_i a_i^T) - 1^T \lambda + n \\ \text{subject to} & \lambda \succ= 0 \end{array}$$

Minimum volume covering ellipsoid

- 弱化的**Slater**条件：存在 $X \in S_{++}^{n}$，满足

$$a_i^T X a_i \le 1, i = 1, \ldots, m$$

- 弱化的**Slater**条件总成立，因此该优化问题具有强对偶性。

对偶可行解的不等式

- 对于一优化问题，若 x 为一可行解，(λ,ν) 为对偶问题可行解，则有如下不等式：

$$f_0(x)-p^* \le f_0(x)-g(\lambda,\nu)$$

  x 为$\varepsilon-$次优解，其中

$$\varepsilon = f_0(x)-g(\lambda,\nu)$$

- 不等式可以用于对次优解的精度估计

互补松弛条件

- 设 x^* 为原始优化问题的最优解，(λ^*, ν^*) 为对偶问题的最优解，若两者具有强对偶性，则

$$
\begin{aligned}
f_0(x^*) &= g(\lambda^*, \nu^*) \\
&= \inf_x \left(f_0(x) + \sum_i \lambda_i^* f_i(x) + \sum_i \nu_i^* h_i(x) \right) \\
&\le f_0(x^*) + \sum_i \lambda_i^* f_i(x^*) + \sum_i \nu_i^* h_i(x^*) \\
&\le f_0(x^*)
\end{aligned}
$$

所以 $$\sum_i \lambda_i^* f_i(x^*) = 0$$

即 $$\lambda_i^* f_i(x^*) = 0, i = 1, \dots, m$$

KKT优化条件

- 设优化问题中，函数 $f_0(x),\ldots,f_m(x),h_0(x),\ldots,h_p(x)$ 可微。设x^*为原始优化问题的最优解，(λ^*,ν^*) 为对偶问题的最优解，且两者具有强对偶性，则 (x^*,λ^*,ν^*) 满足如下条件：

1. $f_i(x^*)\le 0, i=1,\ldots,m$
2. $h_i(x^*)=0, i=1,\ldots,p$
3. $\lambda_i^*\ge 0, i=1,\ldots,m$
4. $\lambda_i^* f_i(x^*)=0, i=1,\ldots,m$
5. $\nabla f_0(x^*)+\sum_{i=1}^{m}\lambda_i^*\nabla f_i(x^*)+\sum_{i=1}^{p}\nu_i^*\nabla h_i(x^*)=0$

KKT条件为必要条件！

信息与通信工程学院 庄伯金
bjzhuang@bupt.edu.cn

凸优化问题的KKT条件

- 设原始问题为凸优化问题中，函数

$$f_0(x),\ldots,f_m(x),h_0(x),\ldots,h_p(x)$$

可微。设$(\tilde{x},\tilde{\lambda},\tilde{v})$ 满足**KKT**条件，则 $\tilde{x}$ 为原始问题的最优解，而 $(\tilde{\lambda},\tilde{v})$ 为对偶问题的最优解，且两者具有强对偶性。

例

- 原始凸优化问题 minimize $\quad -\sum_i \log(\alpha_i + x_i)$

  subject to $\quad x \succ= 0$

  $1^T x = 1$

KKT条件 $\quad x^* \succ= 0, 1^T x^* = 1, \lambda^* \succ= 0,$

$\lambda_i^* x_i^* = 0, i = 1, \ldots, n,$

$-1/(\alpha_i + x_i^*) - \lambda_i^* + \nu^* = 0, i = 1, \ldots, n$

例

解得

$$x_i^* = \begin{cases} 1/v^* - \alpha_i & v^* < 1/\alpha_i \\ 0 & v^* \geq 1/\alpha_i \end{cases}$$

其中

$$\sum_{i=1}^{n} \max\{0, 1/v^* - \alpha_i\} = 1$$

凸优化问题的对偶求解

- 设原始优化问题与对偶问题具有强对偶性，且(λ^*, ν^*) 为对偶问题的最优解。$L(x, \lambda^*, \nu^*)$ 存在唯一的最小解，即

$$\text{minimize} \quad f_0(x) + \sum_{i=1}^{m} \lambda_i^* f_i(x) + \sum_{i=1}^{p} \nu_i^* h_i(x)$$

存在唯一解 x^* 。若 x^* 为原始问题的可行解，则x^* 即为原始问题的最优解；若 x^* 不是原始问题的可行解，则原始问题不存在最优解。

扰动问题

- 扰动问题：

$$\begin{aligned} \text{minimize} \quad & f_0(x) \\ \text{subject to} \quad & f_i(x) \le u_i, \quad i = 1, \ldots, m \\ & h_j(x) = v_j, \quad j = 1, \ldots, p \end{aligned}$$

- 当 $u = 0, v = 0$ 时即为原始问题。
- 若 u_i 为正，则第 i 个不等式约束被放宽；若 u_i 为负，则第 i 个不等式约束被收紧。
- 记 $p^*(u,v)$ 为扰动问题的最优解。若扰动问题无最优解，则记

$$p^*(u,v) = \infty$$

灵敏度分析

- 设对偶问题存在最优解，且与原始问题具有强对偶性，若非干扰问题的最优对偶解为(λ^*, v^*)，则有

$$p^*(u,v) \geq p^*(0,0) - \lambda^{*T}u - v^{*T}v$$

- 若 $p^*(u,v)$ 在$u=0, v=0$ 处可微，则

$$\lambda_i^* = -\frac{\partial p^*(0,0)}{\partial u_i} \qquad v_i^* = -\frac{\partial p^*(0,0)}{\partial v_i}$$

选择定理

- 定义（弱选择性）：若两个不等式（等式）系统，至多有一个可解，则称这两个系统具有弱选择性。
- 设原始问题的约束条件：

$$f_i(x) \le 0, i = 1, \ldots, m, h_j(x) = 0, j = 1, \ldots, p$$

- 对偶问题

$$g(\lambda, \nu) = \inf\left(\sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{p} \nu_i h_i(x)\right)$$

- 对偶不等式组

$$\lambda \succ= 0, g(\lambda, \nu) > 0$$

- 原始问题的约束条件与对偶不等式组具有弱选择性。

信息与通信工程学院 庄伯金
bjzhuang@bupt.edu.cn

选择定理

- 设原始问题的严格不等式约束条件：

$$f_i(x) < 0, i = 1, \ldots, m, h_j(x) = 0, j = 1, \ldots, p$$

- 对偶不等式组

$$\lambda \succ= 0, \lambda \neq 0, g(\lambda, v) \geq 0$$

- 原始问题的严格不等式约束条件与对偶不等式组具有弱选择性。

选择定理

- 定义（强选择性）：若两个不等式（等式）系统，恰有一个可解，则称这两个系统具有强选择性。

- 设原始问题为凸优化问题，其严格不等式约束条件为：

$$f_i(x) < 0, i = 1, \ldots, m, Ax = b$$

- 对偶不等式组 $\lambda \succ= 0, \lambda \neq 0, g(\lambda, v) > 0$

- 若存在 $x \in \mathrm{renlint}D$ ，满足 $Ax = b$，则上述两不等式约束系统具有强选择性。

选择定理

- 设原始问题为凸优化问题，其不等式约束条件为：

$$f_i(x) \le 0, i = 1, \ldots, m, Ax = b$$

- 对偶不等式组

$$\lambda \succ= 0, g(\lambda, \nu) > 0$$

- 若存在 $x \in \mathrm{renlint} D$ ，满足$Ax = b$ ，且下述优化问题存在最优解

$$\begin{aligned} &\text{minimize} \quad s \\ &\text{subject to} \quad f_i(x) - s \le 0, \quad i = 1, \ldots, m \\ &\qquad\qquad\qquad Ax = b \end{aligned}$$

则原始问题的不等式约束条件与对偶不等式组具有强选择性。

信息与通信工程学院 庄伯金
bjzhuang@bupt.edu.cn

凸优化理论与应用

第五章　逼近与拟合

信息与通信工程学院 庄伯金
bjzhuang@bupt.edu.cn

范数逼近问题

- 问题描述：

$$\text{minimize} \quad \|Ax-b\|$$

$$A \in R^{m\times n}, m \geq n,$$

- 残差向量 $r = Ax - b$

- 范数逼近问题的原型
  - 几何原型：在线性变换象集到某一点的最小距离。
  - 优化设计原型：一个线性系统中与目标最接近的输入变量。

例

- 最小平方逼近：范数 $\|\bullet\|_2$

  最优解满足： $A^T Ax = A^T b$

- 切比雪夫逼近：范数 $\|\bullet\|_\infty$，原问题转换为**LP**问题：

$$\begin{array}{ll} \text{minimize} & t \\ \text{subject to} & -t1 \le Ax - b \le t1 \end{array}$$

- 残差绝对值和逼近：范数 $\|\bullet\|_1$ ，原问题转换为**LP**问题：

$$\begin{array}{ll} \text{minimize} & 1^T y \\ \text{subject to} & -y \prec= Ax - b \prec= y \end{array}$$

罚函数逼近

- 问题描述：

$$\text{minimize} \quad \sum_{i=1}^{n} \phi(r_i)$$

$$\text{subject to} \quad r = Ax - b$$

- 罚函数 $\phi(r)$ 表示逼近问题误差的代价，一般为对称、非负且 $\phi(0) = 0$
- 若罚函数为凸函数，则罚函数逼近问题为凸优化问题。

罚函数的例

- $l_p, p \geq 1$ 范数：

$$\phi(r) = |r|^p$$

- 死区线性罚函数：

$$\phi(r) = \begin{cases} 0 & |r| \leq a \\ |r| - a & |r| > a \end{cases}$$

- 对数门限罚函数

$$\phi(r) = \begin{cases} -a^2 \log(1-(r/a)^2) & |r| < a \\ \infty & |r| \geq a \end{cases}$$

鲁棒的罚函数

- 若 $|r|$ 大到一定程度时，罚函数为$|r|$ 的线性函数，则称该罚函数为鲁棒的罚函数。
- **Huber罚函数**

$$\phi(r)=\begin{cases} r^2 & |r|\le M \\ M(2|r|-M) & |r|>M \end{cases}$$

最小范数问题

- 问题描述：$\text{minimize} \quad \|x\|$

$$\text{subject to} \quad Ax = b, A \in R^{m\times n}, m < n$$

- 可以消去等式约束将其转换为范数逼近问题：

$$\text{minimize} \quad \|x_0 + zu\|$$

其中 $x_0 + zu$ 为方程组 $Ax = b$ 的解。

最小范数问题

- 最小平方范数问题：范数 $\|\bullet\|_2$ ，最优解满足：

$$2x^* + A^T v^* = 0, Ax^* = b$$

- 最小罚问题：

$$\begin{aligned} &\text{minimize} \quad \sum_{i=1}^{n} \phi(x_i) \\ &\text{subject to} \quad Ax = b \end{aligned}$$

- 绝对值和最小问题：范数 $\|\bullet\|_1$ ，原问题可转换为**LP**问题：

$$\begin{aligned} &\text{minimize} \quad 1^T y \\ &\text{subject to} \quad Ax = b, -y \prec= x \prec= y \end{aligned}$$

正则逼近

- 二元矢量优化问题描述：

$$\text{minimize}(\text{w.r.t. } R_+^2) \quad (\|Ax-b\|, \|x\|)$$

- 正则化问题：

$$\text{minimize} \quad \|Ax-b\| + \gamma\|x\|, \gamma > 0$$

最优解描述了两分量的一条折中曲线。

正则逼近

- **Tikhonov正则化问题：**

$$\text{minimize} \quad \|Ax-b\|^2+\delta\|x\|^2, \delta>0$$

为二次优化问题：

$$\text{minimize} \quad x^T(A^TA+\delta I)x-2b^TAx+b^Tb$$

最优解的形式：

$$x=(A^TA+\delta I)^{-1}A^Tb$$

正则逼近

- **Tikhonov光滑正则化问题：**

$$\text{minimize} \quad \|Ax-b\|^2+\delta\|\Delta x\|^2, \delta>0$$

Δx 为二阶差分算子：

$$\Delta x = n^2\begin{bmatrix} 1 & -2 & 1 & 0 & \dots & 0 & 0 & 0 \\ 0 & 1 & -2 & 1 & \dots & 0 & 0 & 0 \\ 0 & 0 & 1 & -2 & \dots & 0 & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & 0 & \dots & -2 & 1 & 0 \\ 0 & 0 & 0 & 0 & \dots & 1 & -2 & 1 \end{bmatrix} x$$

信息与通信工程学院 庄伯金
bjzhuang@bupt.edu.cn

信号复原

- 已知加噪信号：

$$x_{cor} = x + v$$

信号复原问题的描述：

$$\text{minimize(w.r.t. } R_+^2) \quad (\|\hat{x} - x_{cor}\|_2, \phi(\hat{x}))$$

函数 $\phi(\hat{x}): R^n \to R$ 为正则函数或光滑函数。

$$\phi_{quad}(\hat{x}) = \sum_{i=1}^{n-1} (\hat{x}_{i+1} - \hat{x}_i)^2 \qquad \phi_{tv}(\hat{x}) = \sum_{i=1}^{n-1} |\hat{x}_{i+1} - \hat{x}_i|$$

信号复原

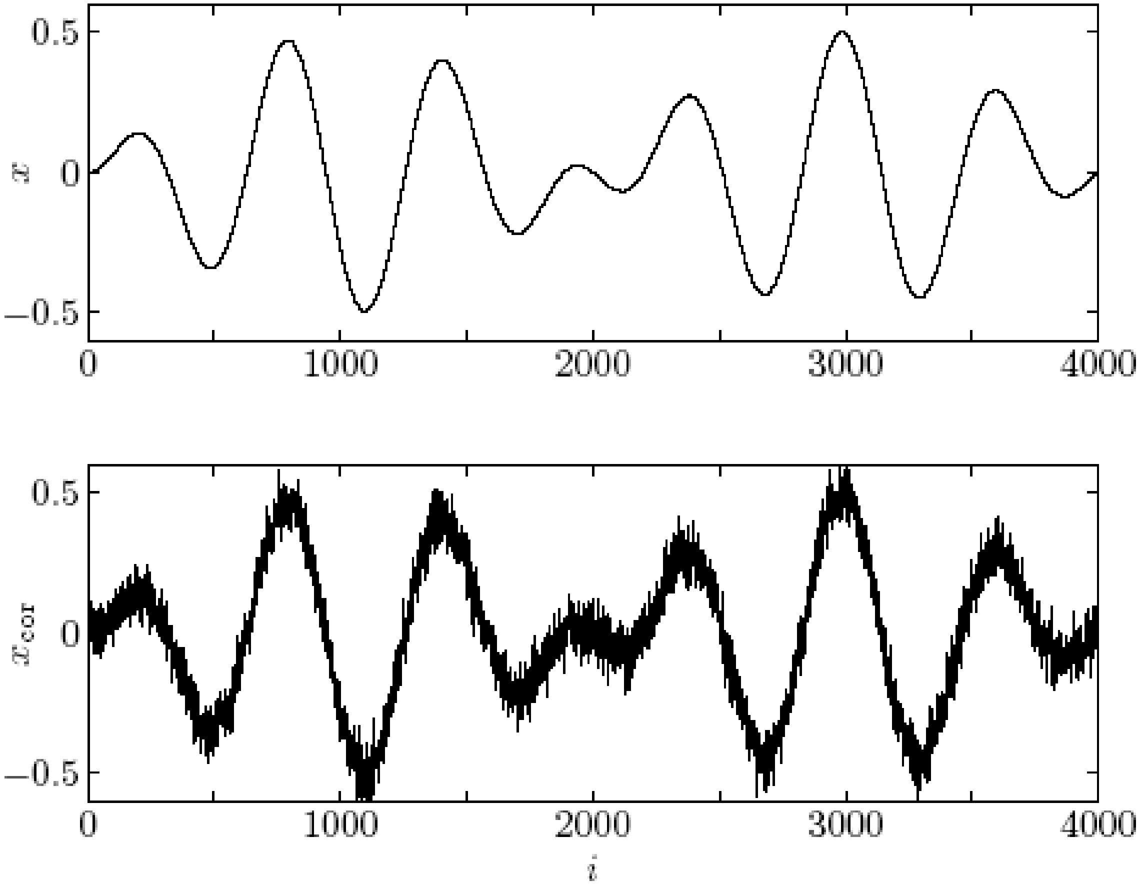


Figure 6.8 *Top:* the original signal $x \in \mathbf{R}^{4000}$. *Bottom:* the corrupted signal x_{cor}.

信息与通信工程学院 庄伯金
bjzhuang@bupt.edu.cn

信号复原

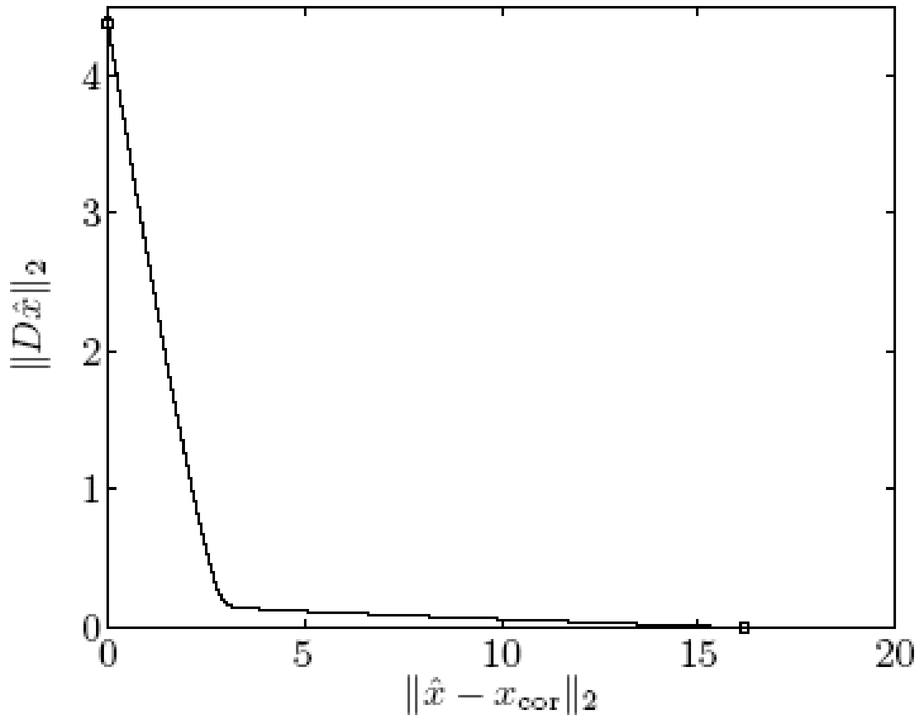


Figure 6.9 Optimal trade-off curve between $\|D\hat{x}\|_2$ and $\|\hat{x} - x_{\rm cor}\|_2$. The curve has a clear knee near $\|\hat{x} - x_{\rm cor}\| \approx 3$.

信号复原

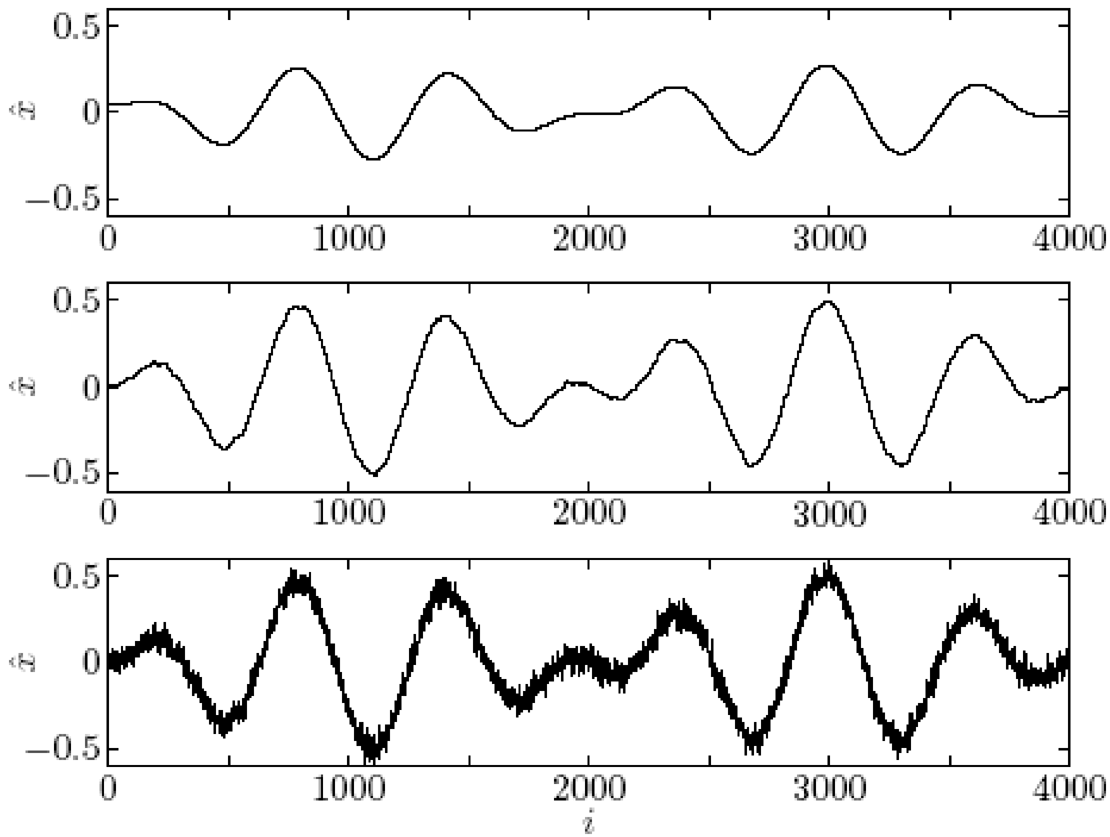


Figure 6.10 Three smoothed or reconstructed signals $\hat{x}$. The top one corresponds to $\|\hat{x} - x_{\mathrm{cor}}\|_2 = 8$, the middle one to $\|\hat{x} - x_{\mathrm{cor}}\|_2 = 3$, and the bottom one to $\|\hat{x} - x_{\mathrm{cor}}\|_2 = 1$.

鲁棒逼近

- 问题描述：minimize $\quad \|Ax-b\|, A$不确定
- 随机鲁棒逼近：A为随机变量，逼近问题转换为最小化期望

$$\text{minimize} \quad \mathrm{E}(\|Ax-b\|)$$

- 例：

$$P(A=A_i)=p_i$$

随机鲁棒逼近为：

$$\text{minimize} \quad \sum_{i=1}^{n} p_i \|A_i x-b\|$$

转换为：

$$\begin{array}{ll} \text{minimize} & p^T t \\ \text{subject to} & \|A_i x-b\| \le t_i, i=1,\dots,n \end{array}$$

随机鲁棒逼近

- A为随机变量： $A = \bar{A} + U$

最小平方随机鲁棒逼近为：

$$\text{minimize} \quad \mathrm{E}\left\| Ax - b \right\|_2^2$$

转换为：

$$\text{minimize} \quad \left\| \bar{A}x - b \right\|_2^2 + \left\| P^{1/2} x \right\|_2^2$$

其中 $P = \mathrm{E} U^T U$

最坏情况鲁棒逼近

- 考虑 $A \in I_A$，最坏情况鲁棒逼近为：

$$\text{minimize} \quad \sup_{A \in I_A} (\|Ax - b\|)$$

- 例：

随机鲁棒逼近为：

$$\text{minimize} \quad \sum_{i=1}^{n} p_i \|A_i x - b\|$$

转换为：

$$\begin{array}{ll} \text{minimize} & p^T t \\ \text{subject to} & \|A_i x - b\| \le t_i, i = 1, \ldots, n \end{array}$$

信息与通信工程学院 庄伯金
bjzhuang@bupt.edu.cn

凸优化理论与应用

第6章　统计估计

信息与通信工程学院 庄伯金
bjzhuang@bupt.edu.cn

概率分布的参数估计

- 随机变量的概率密度为 $p_x(\cdot)$，其中x为概率分布的参数，且参数未知。参数估计的目标就是通过一些已知样本估计获得参数的最优近似值。
- 问题描述

$$\text{maximize}(x \in C) \quad \log p_x(y)$$

- y 为样本观测值；
- $l(x) = \log p_x(y)$为对数似然函数；
- 若似然函数为凹函数，则优化问题为凸优化问题。

信息与通信工程学院 庄伯金
bjzhuang@bupt.edu.cn

具有独立同分布噪声的线性测量

- 线性测量模型： $y_i = a_i^T x + v_i$
- y_i为观测值或测量值；
- $x \in R^n$ 为未知参数向量；
- v_i 独立同分布噪声，其概率密度为 p。
- 似然函数为

$$p_x(y) = \prod_{i=1}^{n} p(y_i - a_i^T x)$$

- 最大似然估计问题为：

$$\text{maximize} \quad \sum_{i=1}^{n} \log p(y_i - a_i^T x)$$

信息与通信工程学院 庄伯金
bjzhuang@bupt.edu.cn

例

- 高斯白噪声

$$p(r) = N(0,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{r^2}{2\sigma^2}}$$

对数似然函数：

$$l(x) = -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(a_i^T x - y_i)^2$$

- 区间 $[-a, a]$ 上均匀分布的噪声：

对数似然函数：

$$l(x) = \begin{cases} -n\log(2a) & \left|a_i^T x - y_i \le a\right|, i = 1,\ldots,n \\ -\infty & otherwise \end{cases}$$

逻辑回归

- 随机变量 $y\in\{0,1\}$ 的概率分布为：

$$p=\mathrm{prob}(y=1)=\frac{\exp(a^T u+b)}{1+\exp(a^T u+b)}$$

- $a\in R^n, b\in R$ 为参数；
- $u\in R^n$ 为可观测的解释变量；(u_i, y_i)为观察值。

对数似然函数：$(y_1=\ldots=y_k=1, y_{k+1}=\ldots=y_n=0)$

$$l(a,b)=\log\left(\prod_{i=1}^{k}\frac{\exp(a^T u_i+b)}{1+\exp(a^T u_i+b)}\prod_{i=k+1}^{n}\frac{1}{1+\exp(a^T u_i+b)}\right)$$

$$=\sum_{i=1}^{k}(a^T u_i+b)-\sum_{i=1}^{n}(1+\exp(a^T u_i+b))$$

信息与通信工程学院 庄伯金
bjzhuang@bupt.edu.cn

假定测验

- 随机变量 $X \in \{1,2,\ldots,n\}$，有m种可能（假定）的分布；
- 假定 j：X 的概率分布为 $p_j = \{p_{1j}, p_{2j}, \ldots, p_{nj}\}$
- 假定测验的目标：由观察值猜测随机变量最有可能服从哪种假定的分布。
- 当 $m=2$ 时，称为二元假定测验。
- 随机检测子：非负元素矩阵 $T \in R^{2\times n}$

$$t_{ik} = prob(\widehat{\theta} = i \mid x = k)$$

假定测验

- 检测概率矩阵

$$D=[T_1\ \ T_2]=\begin{bmatrix}1-P_{fp} & P_{fn}\\ P_{fp} & 1-P_{fn}\end{bmatrix}$$

- P_{fp} 为当 X 实际服从第**1**种假定分布而猜测为第**2**种假定分布的概率；
- P_{fn} 为当 X 实际服从第**2**种假定分布而猜测为第**1**种假定分布的概率；
- 多目标优化形式：

$$\begin{aligned}&\text{minimize(w.r.t. } R_+^2) && (P_{fp}, P_{fn})\\ &\text{subject to} && t_{1k}+t_{2k}=1, k=1,\ldots,n\\ & && t_{1k}\geq 0, t_{2k}\geq 0, k=1,\ldots,n\end{aligned}$$

假定测验

- 尺度优化形式：

$$\begin{aligned} \text{minimize} \quad & P_{fp} + \lambda P_{fn} \\ \text{subject to} \quad & t_{1k} + t_{2k} = 1, k = 1, \ldots, n \\ & t_{1k} \geq 0, t_{2k} \geq 0, k = 1, \ldots, n \end{aligned}$$

- 最小最大值形式

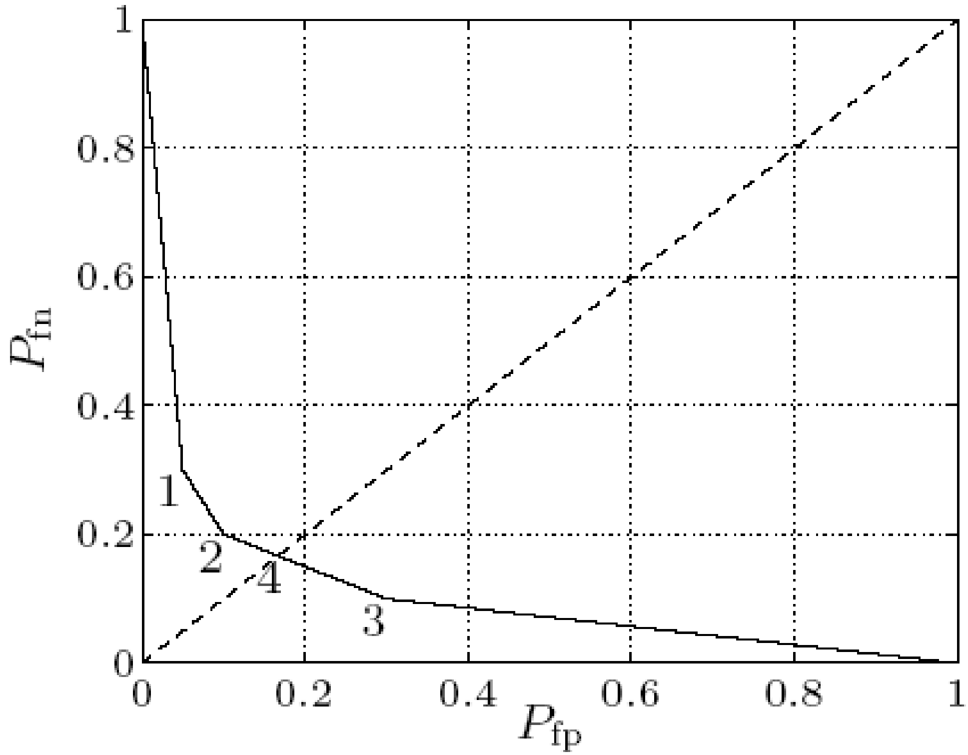
$$\begin{aligned} \text{minimize} \quad & \max(P_{fp}, P_{fn}) \\ \text{subject to} \quad & t_{1k} + t_{2k} = 1, k = 1, \ldots, n \\ & t_{1k} \geq 0, t_{2k} \geq 0, k = 1, \ldots, n \end{aligned}$$

例

- X 在两种假设下的概率分布为：

$$P=\begin{bmatrix} 0.7 & 0.1 \\ 0.2 & 0.1 \\ 0.05 & 0.7 \\ 0.05 & 0.1 \end{bmatrix}$$

例



solutions 1, 2, 3 (and endpoints) are deterministic; 4 is minimax detector

实验设计

- 线性测量问题 $y_i = a_i^T x + w_i$
- w_i独立同分布高斯白噪声，服从分布 $N(0,1)$。
- 最大似然估计值：

$$\widehat{x} = (\sum_{i=1}^{n} a_i a_i^T)^{-1} \sum_{i=1}^{n} y_i a_i$$

- 估计误差 $e = \widehat{x} - x$ 均值为**0**，方差为

$$E = Eee^T = (\sum_{i=1}^{n} a_i a_i^T)^{-1}$$

信赖椭圆为 $\{x \mid (x-\widehat{x})^T E^{-1} (x-\widehat{x}) \le \beta\}$

实验设计

- 实验设计的目标：寻找 $a_i \in \{v_1,\dots,v_p\}$，使得误差的方差矩阵最小。
- 向量优化形式：

$$
\begin{aligned}
&\text{minimize(w.r.t. } S_+^n) && E = \left(\sum_{k=1}^{p} m_k v_k v_k^T\right) \\
&\text{subject to} && m_k \geq 0, m_1 + \dots + m_p = n \\
& && m_k \in Z
\end{aligned}
$$

- 为整数问题，求解较困难。

实验设计

- 当 $n \gg p$ 时，令 $\lambda_p = m_p / n$ 近似为一连续实数，原问题可松弛转换为连续实数优化：

$$\text{minimize(w.r.t. } S_+^n) \quad E = \frac{1}{n}\left(\sum_{k=1}^{p} \lambda_k v_k v_k^T\right)$$

$$\text{subject to} \quad \lambda \succ= 0, 1^T \lambda = 1$$

凸优化理论与应用

第7章　几何问题

信息与通信工程学院 庄伯金
bjzhuang@bupt.edu.cn

体积问题

- 已知集合 C ， E 为包含 C 的椭球，满足：

$$E = \{v \mid \|Av+b\| \le 1\}$$

- 求包含 C 的体积最小的椭球问题：

$$\begin{array}{ll} \text{minimize} & \log\det A^{-1} \\ \text{subject to} & \sup_{v\in C}\|Av+b\| \le 1 \end{array}$$

- 若 C 为有限集，则问题变为：

$$\begin{array}{ll} \text{minimize} & \log\det A^{-1} \\ \text{subject to} & \sup\|Av_i+b\| \le 1, i=1,\dots,m \end{array}$$

体积问题

- 已知凸集 C ， E 为包含在C 内的椭球，满足：

$$E=\{Bu+b \mid \|u\|_2 \le 1\}$$

- 求包含在 C 内的体积最大的椭球问题：

$$\begin{array}{ll} \text{maximize} & \log\det B \\ \text{subject to} & \sup\limits_{\|u\|_2 \le 1} I_C(Bu+d) \le 0 \end{array}$$

- 若C 为多面体，则问题变为：

$$\begin{array}{ll} \text{maximize} & \log\det B \\ \text{subject to} & \|Ba_i\|_2 + a_i^T d \le b_i, i=1,\ldots,m \end{array}$$

中心问题

- 已知凸集 C，包含在C 内的最大体积球的球心，称为**Chebyshev**中心。

- 已知凸集C，包含在 C 内的最大体积椭球的球心，称为**MVE**中心。

线性判别

- 两个可分离点集 $\{x_1, \ldots, x_N\}$和$\{y_1, \ldots, y_M\}$

分割超平面满足：

$$a^T x_i + b > 0, i = 1, \ldots, N, \text{且} a^T y_i + b < 0, i = 1, \ldots, M$$

可对其归一化：

$$a^T x_i + b \geq 1, i = 1, \ldots, N, \text{且} a^T y_i + b \leq -1, i = 1, \ldots, M$$

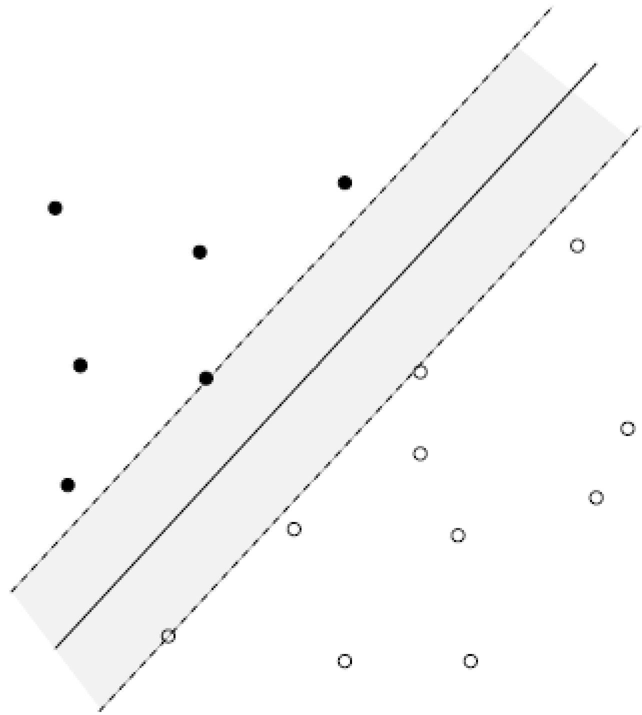
线性判别

- 支撑超平面 $H1: a^T x_s + b = 1$

  $H2: a^T y_t + b = -1$

两超平面之间的距离：

$$d(H1, H2) = 2 / \|a\|_2$$



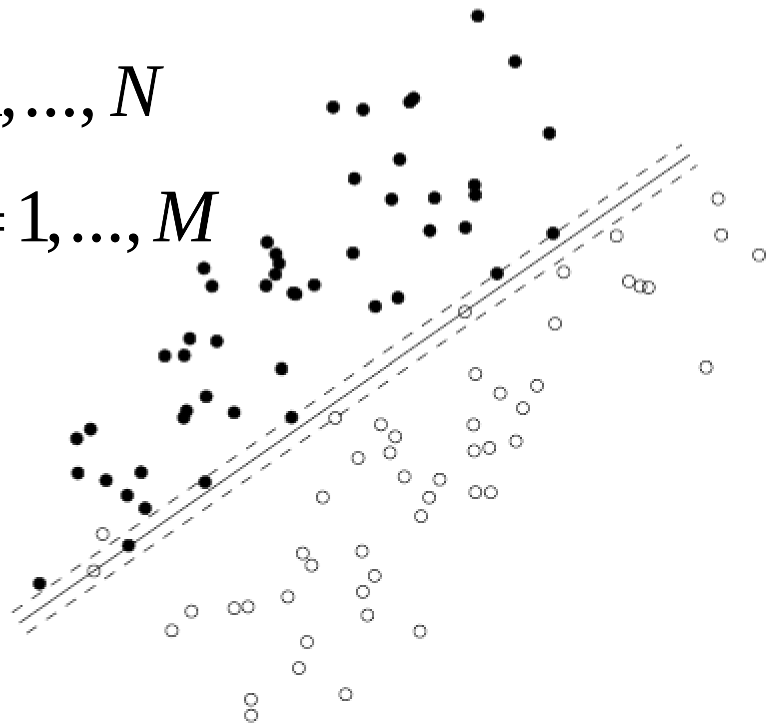
最优线性分割问题：

$$\begin{aligned} &\text{minimize} \quad \|a\|_2 / 2 \\ &\text{subject to} \quad a^T x_i + b \geq 1, i = 1, \ldots, N \\ &\qquad\qquad\quad a^T y_i + b \leq -1, i = 1, \ldots, M \end{aligned}$$

线性判别

- 若两点集 $\{x_1,\dots,x_N\}$ 和 $\{y_1,\dots,y_M\}$ 不完全可分，近似分割超平面问题：

$$
\begin{aligned}
\text{minimize} \quad & 1^T u + 1^T v \\
\text{subject to} \quad & a^T x_i + b \geq 1 - u_i, i = 1,\dots,N \\
& a^T y_i + b \leq -1 + v_i, i = 1,\dots,M \\
& u_i \geq 0, v_i \geq 0
\end{aligned}
$$



线性判别

- 支持向量分类器问题：

$$\begin{aligned} &\text{minimize} \quad && \|a\|_2 + \lambda(1^T u + 1^T v) \\ &\text{subject to} \quad && a^T x_i + b \geq 1 - u_i, i = 1, \ldots, N \\ & && a^T y_i + b \leq -1 + v_i, i = 1, \ldots, M \\ & && u_i \geq 0, v_i \geq 0 \end{aligned}$$

凸优化理论与应用

第7章　无约束优化

信息与通信工程学院 庄伯金
bjzhuang@bupt.edu.cn

无约束优化问题

- 问题描述： $\text{minimize} \quad f(x)$
- $f(x)$ 为凸函数，且二次可微。
- 无约束问题求解的两种方法：
  - 求解梯度方程：
    $$\nabla f(x^*) = 0$$
  - 迭代逼近：
    $$f(x^{(k)}) \to p^*$$

例

- 二次优化：

$$\text{minimize} \quad \frac{1}{2}x^T P x + q^T x + r, P \in S_+^n$$

梯度方程

$$Px^* + q = 0$$

信息与通信工程学院 庄伯金
bjzhuang@bupt.edu.cn

迭代起始点

- 起始点 $x^{(0)}$ 满足：

$1.x^{(0)} \in \mathrm{dom}f$；

$2.S=\{x \in \mathrm{dom}f \mid f(x) \le f(x^{(0)})\}$为闭集。

- 满足条件**2**的几种函数：
  - 函数 $f(x)$任意下水平集都是闭集；
  - 函数的定义域为 R^n
  - 当 $x \to \mathrm{bd}\ \mathrm{dom}f$ 时，$f(x) \to \infty$

强凸性

- 定义：函数 $f(x)$ 在 S 上具有强凸性，若$f(x)$ 满足

$$\nabla^2 f(x) \succ= mI, m > 0$$

- 若函数 $f(x)$具有强凸性，则有

$$f(y) \geq f(x) + \nabla f(x)^T (y-x) + \frac{m}{2}\|y-x\|_2^2$$

$$\geq f(x) - \frac{1}{2m}\|\nabla f(x)\|_2^2$$

- p^* 为最优值，则

$$f(x) - p^* \leq \frac{1}{2m}\|\nabla f(x)\|_2^2$$

信息与通信工程学院 庄伯金
bjzhuang@bupt.edu.cn

强凸性

- 若函数 $f(x)$ 在 S 上具有强凸性，则可以证明存在$M>0$，满足
$$\nabla^2 f(x) \preceq MI$$
则有
$$f(y) \le f(x) + \nabla f(x)^T (y-x) + \frac{M}{2}\|y-x\|_2^2$$

- p^* 为最优值，则
$$p^* \le f(x) - \frac{1}{2M}\|\nabla f(x)\|_2^2$$

强凸性

- 对于 $x \in S$ ，矩阵 $\nabla^2 f(x) \in S_{++}^n$ 的特征值从大到小依次为 $\{\lambda_1, \ldots, \lambda_n\}$ 。则有：

$$mI \prec= \lambda_n I \prec= \nabla^2 f(x) \prec= \lambda_1 I \prec= MI$$

- 定义：矩阵$\nabla^2 f(x) \in S_{++}^n$ 的条件数为最大特征值与最小特征值之比，即 $r = \lambda_1 / \lambda_n$ 。
- 条件数的上界：

$$r \leq M / m$$

下降法

- 下降法的基本原理：

  迭代 $x^{(k+1)} = x^{(k)} + t\Delta x$ ，满足 $f(x^{(k+1)}) < f(x^{(k)})$

  Δx 为下降方向，t 为步长因子。

- 对于凸函数 $f(x)$ ，当Δx 满足 $\nabla f(x)^T \Delta x < 0$ 时，存在某个 t，使得 $f(x^{(k+1)}) < f(x^{(k)})$ 。

信息与通信工程学院 庄伯金
bjzhuang@bupt.edu.cn

下降法

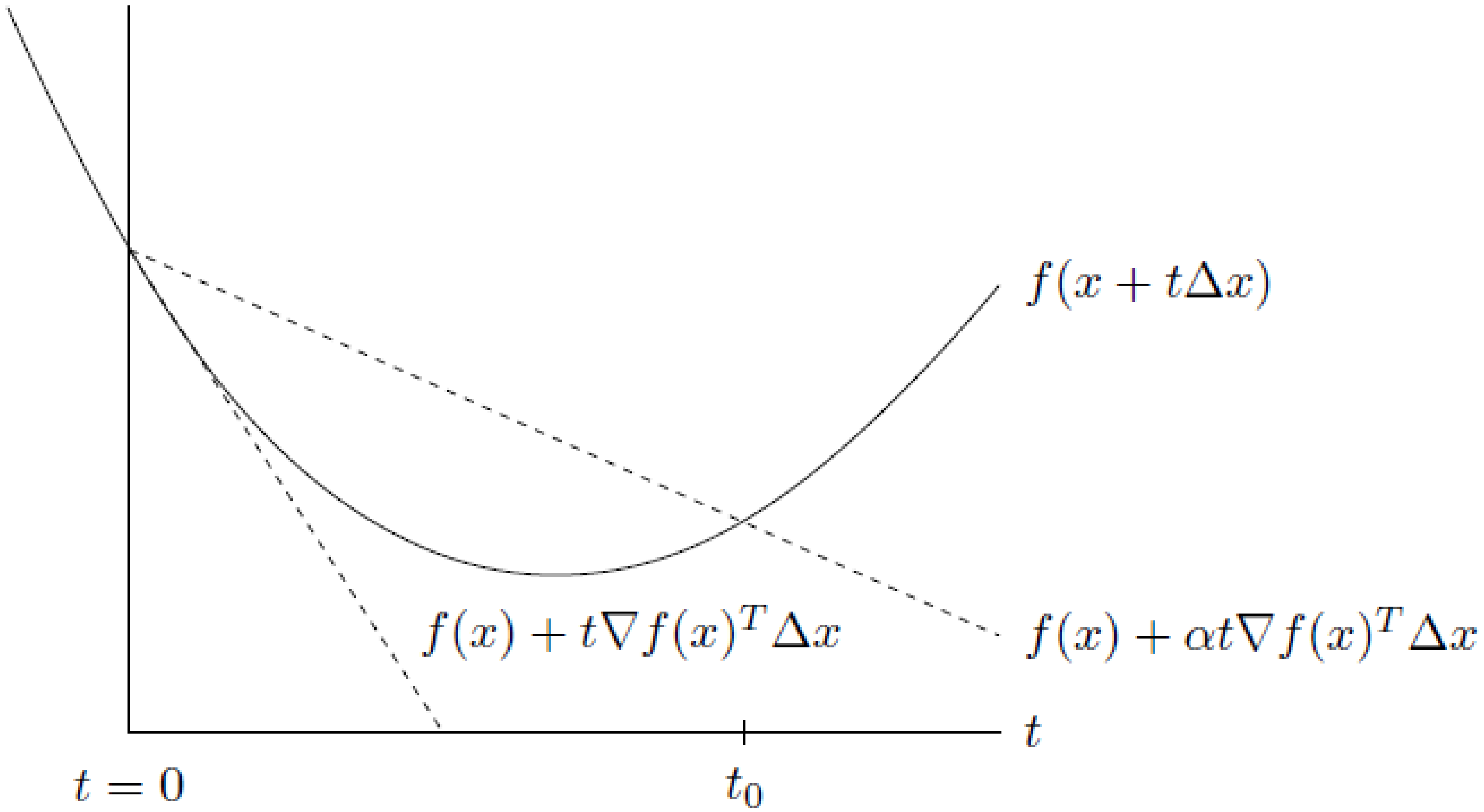
- 下降法的一般步骤：
  - 给出初始点 $x \in \operatorname{dom} f$ ；
  - 循环迭代
    - 计算下降方向 Δx；
    - 搜索步长因子 t ；
    - 迭代：$x = x + t\Delta x$

步长因子搜索

- 精确一维搜索：$t = \arg\min_{t>0} f(x + t\Delta x)$
- 回溯一维搜索：给定参数 $\alpha \in (0, 0.5), \beta \in (0,1)$
  - 初始化：令 $t = 1$；
  - 循环迭代
    - 若 $f(x + t\Delta x) < f(x) + \alpha t \nabla f(x)^T \Delta x$，则终止退出；
    - 否则令 $t = \beta t$

信息与通信工程学院 庄伯金
bjzhuang@bupt.edu.cn

步长因子搜索



$f(x + t\Delta x)$
$f(x) + t\nabla f(x)^T \Delta x$
$f(x) + \alpha t\nabla f(x)^T \Delta x$
t
$t = 0$
t_0

梯度下降法

- 下降方向：$\Delta x = -\nabla f(x)$
- 终止条件：$\|\nabla f(x)\|_2 \le \varepsilon$
- 收敛性：

$$f(x^{(k)}) - p^* \le c^k (f(x^{(0)}) - p^*)$$

其中 $c \in (0,1)$ 。

- 算法简单，但收敛速度较慢。

收敛性分析

- 设函数 $f(x)$ 具有强凸性，则存在$m>0$ 和 $M>0$，满足：

$$mI \preceq= \nabla^2 f(x) \preceq= MI$$

则有：

$$f(x-t\Delta x) \le f(x) - t\|\nabla f(x)\|_2^2 + \frac{Mt^2}{2}\|\nabla f(x)\|_2^2$$

- 若 t 采用精确一维搜索，则$t=1/M$，收敛速度因子：

$$c = 1 - m/M$$

- 若 t 采用回溯一维搜索，收敛速度因子：

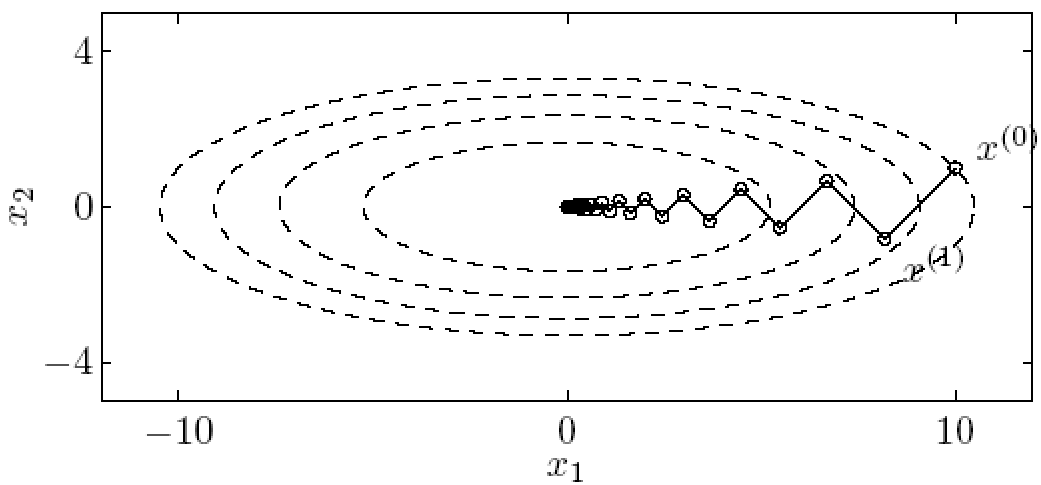
$$c = 1 - \min\{2m\alpha, 2\beta\alpha m/M\}$$

- 条件数越大，收敛速度越小。

例

$$\text{minimize} \quad \frac{1}{2}(x_1^2+\gamma x_2^2), \gamma>0$$

- 初始解为 $(\gamma,1)$，采用精确一维搜索；
- 迭代： $x_1^{(k)}=\gamma\left(\frac{\gamma-1}{\gamma+1}\right)^k \qquad x_2^{(k)}=\left(-\frac{\gamma-1}{\gamma+1}\right)^k$
- 当 $\gamma \ll 1 \; or \; \gamma \gg 1$ 时，算法收敛速度很慢。

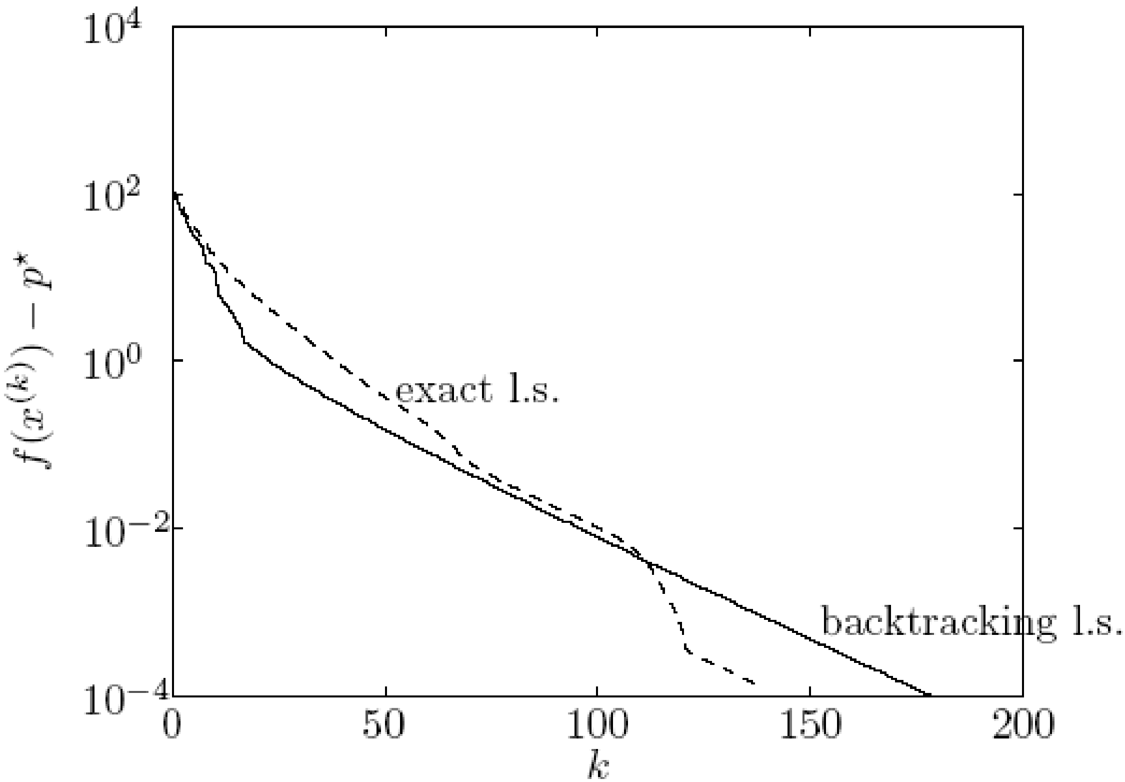


信息与通信工程学院 庄伯金
bjzhuang@bupt.edu.cn

例

$$\text{minimize} \quad c^T x - \sum_{i=1}^{m} \log(b_i - a_i^T x), m = 500, n = 100$$

- 步长因子采用回溯一维搜索。



最速下降法

- 归一化最速下降方向：

$$\Delta x_{nsd} = \arg\min_{v}\{\nabla f(x)^T v \mid \|v\| = 1\}$$

- 非归一化最速下降方向

$$\Delta x_{sd} = \|\nabla f(x)\|_* \Delta x_{nsd}$$

- 欧式范数： $\Delta x_{sd} = -\nabla f(x)$
- 二次范数 $\|x\|_P = (x^T P x)^{1/2}$ ：

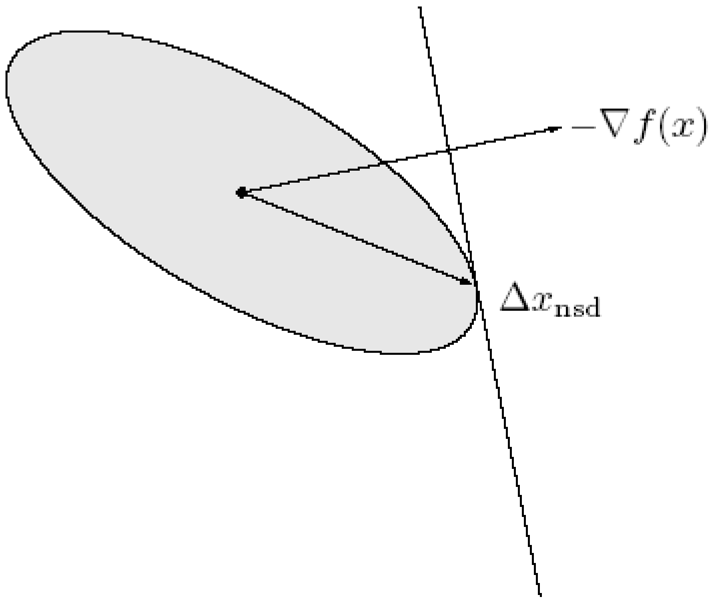
$$\Delta x_{sd} = -P^{-1}\nabla f(x)$$

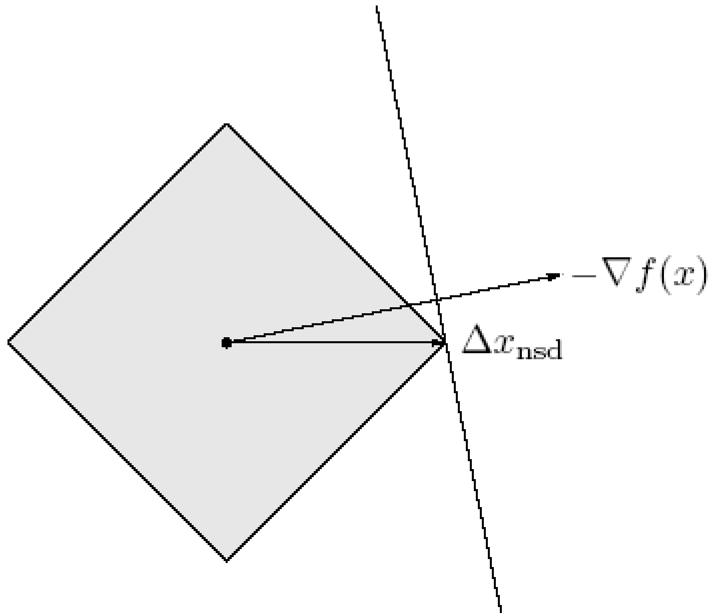
- $l_1 -$范数：

$$\Delta x_{sd} = -\frac{\partial f(x)}{\partial x_i} e_i$$

信息与通信工程学院 庄伯金
bjzhuang@bupt.edu.cn

最速下降法





收敛性分析

- 范数界：

$$\|x\|_* \geq \gamma \|x\|_2, \gamma \in (0,1]$$

- 收敛速度因子：

$$c = 1 - 2m\alpha\gamma^2 \min\{1, \beta\gamma^2 / M\}$$

牛顿法

- 设函数 $f(x)$ 二阶可微，则在 x 附近, $f(x)$ 的泰勒展式为：

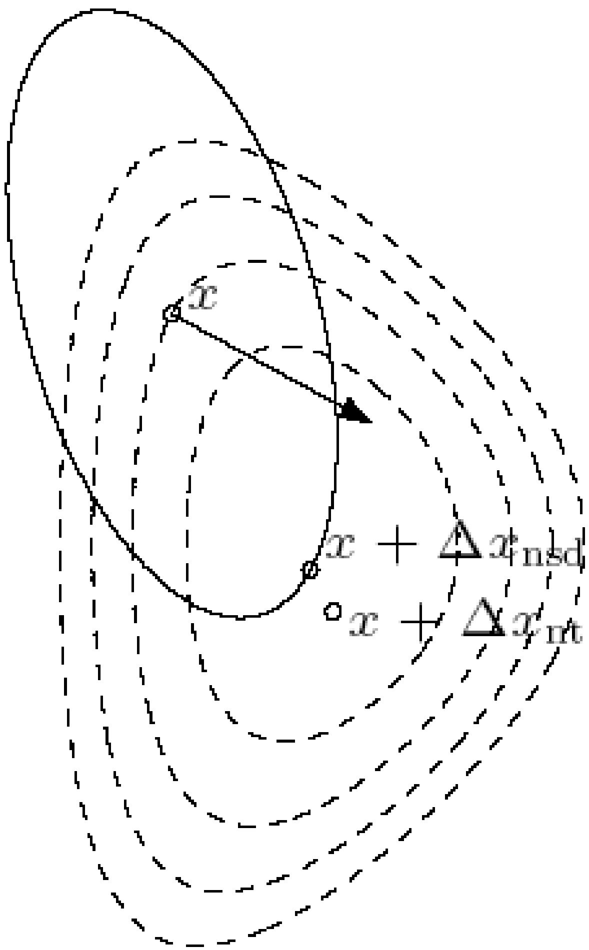
$$\widehat{f}(x+\Delta x)=f(x)+\nabla f(x)^T\Delta x+\frac{1}{2}\Delta x^T\nabla^2 f(x)\Delta x$$

- 泰勒展开可作为 $f(x)$ 在 x 附近的近似；
- 下降方向：

$$\Delta x_{nt}=-\nabla^2 f(x)^{-1}\nabla f(x)$$

- 为二次范数 $\|x\|_{\nabla^2 f(x)}=(x^T\nabla^2 f(x)x)^{1/2}$ 上的最速下降方向。

牛顿法



牛顿减量

- 令 $$\lambda(x) = (\nabla f(x)^T \nabla^2 f(x)^{-1} \nabla f(x))^{1/2}$$
- $\lambda(x)$ 为 $f(x)$ 在 x 处的牛顿减量。
- 牛顿减量的性质**1**：

$$f(x) - \inf_y \widehat{f}(y) = f(x) - \widehat{f}(x + \Delta x_{nt}) = \frac{1}{2}\lambda(x)^2$$

牛顿减量可作为迭代求解的误差估计。

- 性质**2**： $$\nabla f(x)^T \Delta x_{nt} = -\lambda(x)^2$$
- 性质**3**：牛顿减量具有仿射不变性。

牛顿方法

- 初始化：给定初始解 $x \in \mathrm{dom}f$ 以及 $\varepsilon > 0$
- **LOOP：**
  - 计算： $\Delta x_{nt} = -\nabla^2 f(x)^{-1} \nabla f(x)$

    $\lambda^2 = \nabla f(x)^T \nabla^2 f(x)^{-1} \nabla f(x)$
  - 若 $\lambda^2 / 2 < \varepsilon$ 则终止退出；
  - 一维线性搜索：计算步长因子 t ；
  - 迭代： $x = x + t\Delta x_{nt}$

收敛性分析

- 定理：假设 $f(x)$ 二阶连续可微，具有强凸性，即存在 $M > m > 0$ ，满足：

$$mI \preceq \nabla^2 f(x) \preceq MI$$

且海森矩阵满足**Lipschitz**条件，即存在 $L > 0$ ，满足：

$$\left\|\nabla^2 f(x) - \nabla^2 f(y)\right\|_2 \le L\left\|x - y\right\|_2$$

则存在 $0 < \eta < m^2 / L, \gamma > 0$ ，

若$\left\|\nabla f(x^{(k)})\right\|_2 \ge \eta$，则 $f(x^{(k+1)}) - f(x^{(k)}) < -\gamma$ ；

若$\left\|\nabla f(x^{(k)})\right\|_2 < \eta$，则$t^{(k)} = 1$ ，且

$$\frac{L}{2m^2}\left\|\nabla f(x^{(k+1)})\right\|_2 \le \left(\frac{L}{2m^2}\left\|\nabla f(x^{(k)})\right\|_2\right)^2$$

收敛性分析

- 定理：假设 $f(x)$ 二阶连续可微，具有强凸性，即存在 $M>m>0$ ，满足：

$$mI \preceq \nabla^2 f(x) \preceq MI$$

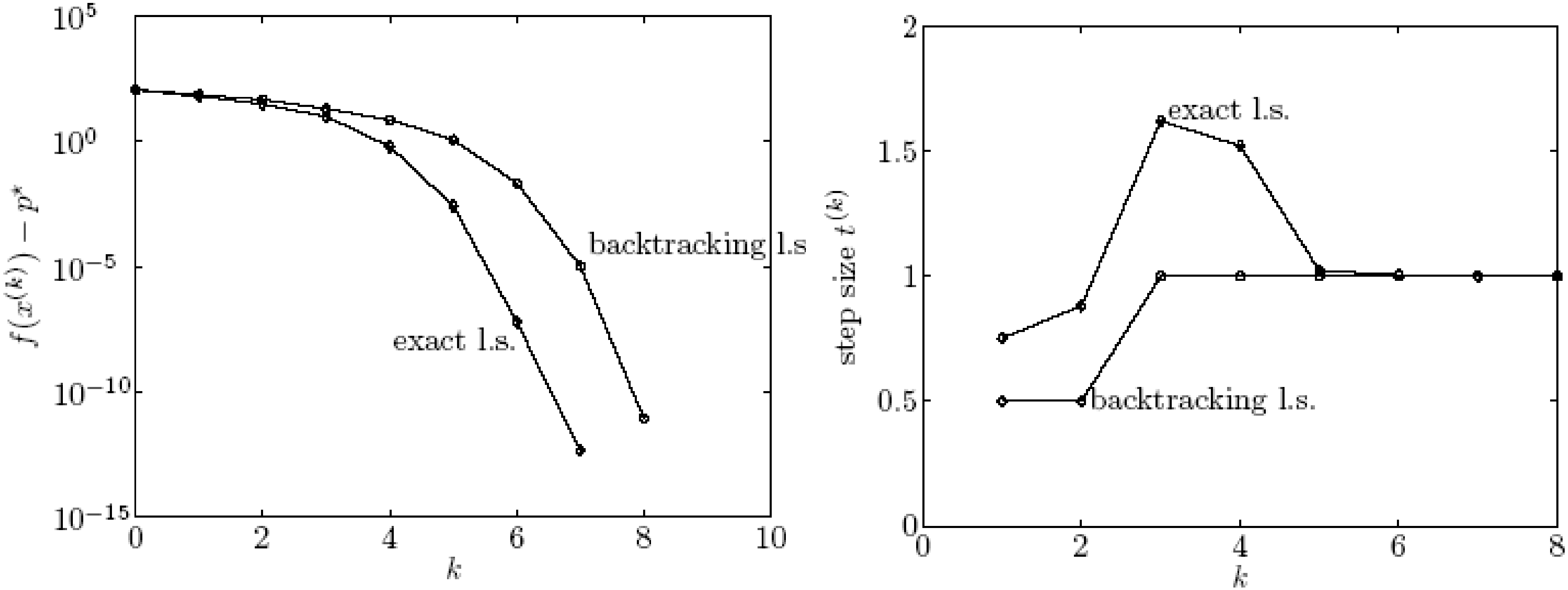
且海森矩阵满足**Lipschitz**条件，即存在 $L>0$ ，满足：

$$\left\|\nabla^2 f(x)-\nabla^2 f(y)\right\|_2 \le L\|x-y\|_2$$

则存在 $k>0$，对于 $l>k$，有

$$f(x^{(l)})-p^* \le \frac{1}{2m}\left\|\nabla f(x^{(l)})\right\|_2^2 \le \frac{2m^3}{L^2}\left(\frac{1}{2}\right)^{2^{l-k+1}}$$

例



信息与通信工程学院 庄伯金
bjzhuang@bupt.edu.cn

凸优化理论与应用

第8章　等式约束优化

信息与通信工程学院 庄伯金
bjzhuang@bupt.edu.cn

等式约束优化问题

- 问题描述：

$$\begin{aligned} &\text{minimize} \quad f(x) \\ &\text{subject to} \quad Ax = b \end{aligned}$$

- $f(x)$ 为凸函数，且二次连续可微，且

$$A \in R^{p\times n}, p < n, \operatorname{rank} A = p$$

- 假设最优值 p^* 存在，则 x^* 为最优解当且仅当存在 v^*，满足（**KKT**条件）：

$$\nabla f(x^*) + A^T v^* = 0, Ax^* = b$$

例

- 二次优化： minimize $\frac{1}{2}x^T Px + q^T x + r, P \in S_+^n$

  subject to $Ax = b$

  KKT系统：

$$\begin{bmatrix} P & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} x^* \\ v^* \end{bmatrix} = \begin{bmatrix} -q \\ b \end{bmatrix}$$

- **KKT系统可解，则二次优化问题存在最优解。**
- 系数矩阵称为**KKT**矩阵。**KKT**矩阵非奇异当且仅当：

$$Ax = 0, x \neq 0 \Rightarrow x^T Px > 0$$

消去等式约束

- 方程组 $Ax=b$ 的解集：

$$\{x \mid Ax=b\}=\{Fz+\tilde{x} \mid z \in R^{n-p}\}$$

- $\tilde{x}$ 为方程组的一个特解， F 为 A 的零空间范围。
- 无约束优化形式：

$$\text{minimize} \quad f(Fz+\tilde{x}), z \in R^{n-p}$$

- 若 z^* 为最优解，则有

$$x^*=Fz^*+\tilde{x} \qquad \nu^*=-(A^TA)^{-1}A\nabla f(x^*)$$

信息与通信工程学院 庄伯金
bjzhuang@bupt.edu.cn

对偶问题

- 对偶形式：

$$\text{maximize} \quad -b^T \nu - f^*(A^T \nu)$$

牛顿法

- x 为等式约束优化的可行解，则在 x 附近原问题的二次近似为：

$$\begin{array}{ll} \text{minimize} & \widehat{f}(x+v) = f(x) + \nabla f(x)^T v + \frac{1}{2} v^T \nabla^2 f(x) v \\ \text{subject to} & A(x+v) = b \end{array}$$

- 设 Δx_{nt} 和 ω 分别为该问题和对偶问题的最优解，则满足：

$$\begin{bmatrix} \nabla^2 f(x) & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} \Delta x_{nt} \\ \omega \end{bmatrix} = \begin{bmatrix} -\nabla f(x) \\ 0 \end{bmatrix}$$

- 牛顿减量

$$\lambda(x) = (\Delta x_{nt}^T \nabla^2 f(x)^{-1} \Delta x_{nt})^{1/2}$$

牛顿减量

- 牛顿减量 $\lambda(x)=(\Delta x_{nt}^{T}\nabla^2 f(x)^{-1}\Delta x_{nt})^{1/2}$
- 牛顿减量的性质：

$$f(x)-\inf\{\widehat{f}(x+v)\,|\,A(x+v)=b\}=\frac{1}{2}\lambda(x)^2$$

- 性质2：牛顿减量具有仿射不变性。

信息与通信工程学院 庄伯金
bjzhuang@bupt.edu.cn

可行下降方向

- 可行下降方向：设 x 满足方程组 $Ax=b$ 。若 v 满足方程组 $Av=0$ ，则 $A(x+tv)=b$ 。v 称为可行方向。若对于较小的 $t>0$，有 $f(x+tv)<f(x)$ ，则 v 为可行下降方向。

等式约束的牛顿方法

- 初始化：给定初始解 $x \in \mathrm{dom} f$ 满足 $Ax = b$，以及 $\varepsilon > 0$
- **LOOP**：
  - 计算 Δx_{nt} 及 λ^2 ；
  - 若 $\lambda^2 / 2 < \varepsilon$ 则终止退出；
  - 一维线性搜索：计算步长因子 t ；
  - 迭代：$x = x + t\Delta x_{nt}$

消去等式约束的牛顿方法

$$\text{minimize} \quad \tilde{f}(z) = f(Fz + \tilde{x}), z \in R^{n-p}$$

- 初始值 $z^{(0)}$，第k 次迭代值$z^{(k)}$；
- 转换为等式约束下的牛顿方法：
  - 初始值：$x^{(0)} = Fz^{(0)} + \tilde{x}$
  - 迭代值：$x^{(k)} = Fz^{(k)} + \tilde{x}$

非可行解为初始点的牛顿法

- x 为等式约束优化的非可行解，则增量 Δx 应尽可能使 $x+\Delta x$ 满足**KKT**条件，即：

$$A(x+\Delta x)=b \qquad \nabla f(x+\Delta x)+A^T\omega=0$$

- 函数 $f(x)$ 二阶连续可微，因此有

$$\nabla f(x+\Delta x)\approx\nabla f(x)+\nabla^2 f(x)\Delta x$$

- 设 Δx_{nt}和 ω 为**KKT**条件的解，即有：

$$\begin{bmatrix}\nabla^2 f(x) & A^T\\ A & 0\end{bmatrix}\begin{bmatrix}\Delta x_{nt}\\ \omega\end{bmatrix}=-\begin{bmatrix}\nabla f(x)\\ Ax-b\end{bmatrix}$$

非可行解为初始点的牛顿法

- 令 $y=(x,v)$ $\quad r(y)=(\nabla f(x)+A^T v, Ax-b)$
- 则**KKT**条件可表示为： $r(y)=0$
- 设 y 为不满足**KKT**条件，则其迭代量需满足：

$$r(y+\Delta y)\approx r(y)+Dr(y)\Delta y=0$$

即

$$\begin{bmatrix} \nabla^2 f(x) & A^T \\ A & 0 \end{bmatrix}\begin{bmatrix} \Delta x_{nt} \\ \Delta v_{nt} \end{bmatrix}=-\begin{bmatrix} \nabla f(x)+A^T v \\ Ax-b \end{bmatrix}$$

非可行解为初始点的牛顿法

- 初始化：给定初始解 $x \in \mathrm{dom} f$ 及 v，以及 $\varepsilon > 0$
- **LOOP**：
  - 计算 Δx_{nt} 和 Δv_{nt}；
  - 回溯一维线性搜索：
    - 令 $t = 1$；
    - **While** $\|r(x + t\Delta x_{nt}, v + t\Delta v_{nt})\|_2 \geq (1 - \alpha t)\|r(x, v)\|_2$

$$t = \beta t$$

  - 迭代：$x = x + t\Delta x_{nt} \quad v = v + t\Delta v_{nt}$
  - 当 $Ax = b$ 且 $\|r(y)\|_2 < \varepsilon$ 时，终止迭代。

KKT系统的求解

- **KKT系统：** $\begin{bmatrix} H & A^T \\ A & 0 \end{bmatrix}\begin{bmatrix} v \\ w \end{bmatrix} = -\begin{bmatrix} g \\ h \end{bmatrix}$
- **1.** LDL^T **分解；**
- **2.若** H **非奇异，则可消元：**

$$AH^{-1}A^T w = h - AH^{-1}g, Hv = -(g + A^T w)$$

- **3.若** H **奇异，则KKT系统可改写为：**

$$\begin{bmatrix} H + A^T QA & A^T \\ A & 0 \end{bmatrix}\begin{bmatrix} v \\ w \end{bmatrix} = -\begin{bmatrix} g + A^T Qh \\ h \end{bmatrix}$$

其中 $Q \succ= 0$ **，且满足** $H + A^T QA \succ 0$ **。**

凸优化理论与应用

第9章　内点法

信息与通信工程学院 庄伯金
bjzhuang@bupt.edu.cn

不等式约束优化问题

- 问题描述：
$$\begin{aligned} \text{minimize} \quad & f_0(x) \\ \text{subject to} \quad & f_i(x) \le 0, i = 1, \ldots, m \\ & Ax = b \end{aligned}$$
- $f_i(x)$ 为凸函数，且二次连续可微，且
$$A \in R^{p \times n}, p < n, \mathrm{rank}A = p$$
- 假设最优值 p^* 存在；
- 假设存在 $\tilde{x} \in \mathrm{dom} f$，满足严格不等式条件 $f_i(x) < 0$
- 则优化问题具有强对偶性，其对偶问题亦可解。

不等式约束的消去

- 示性函数消去不等式约束：

$$\begin{aligned}&\text{minimize} \quad f_0(x)+\sum_{i=1}^{m} I\_(f_i(x))\\&\text{subject to} \quad Ax=b\end{aligned}$$

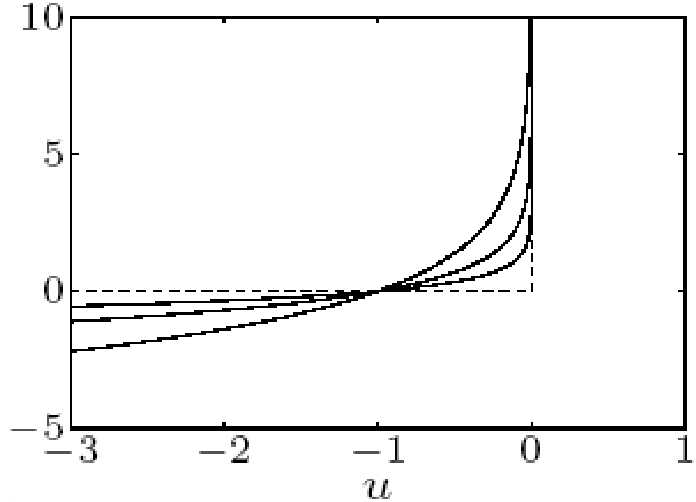
$$I\_(u)=\begin{cases}0 & u\le 0\\ \infty & u>0\end{cases}$$

- $I_(u)$不具备良好的连续可微性，考虑用对数阀函数来近似替代。

对数阀函数

- 对于 $t>0$，$-1/t\log(-u)$ 是 $I_{\_}(u)$ 的光滑逼近。且当 $t\to\infty$ 时，有

$$-1/t\log(-u)\to I_{\_}(u)$$



- 令 $$\phi(x)=-\sum_{i=1}^{m}\log(-f_i(x))$$

- 带示性函数的优化问题可近似为：

$$\begin{array}{ll}\text{minimize} & f_0(x)+\dfrac{1}{t}\phi(x), t>0\\ \text{subject to} & Ax=b\end{array}$$

对数阀函数

- 对数阀函数 $\phi(x)$是凸函数
- 对数阀函数二阶连续可微，导数为：

$$\nabla\phi(x)=\sum_{i=1}^{m}\frac{1}{-f_i(x)}\nabla f_i(x)$$

$$\nabla^2\phi(x)=\sum_{i=1}^{m}\frac{1}{f_i(x)^2}\nabla f_i(x)\nabla f_i(x)^T+\sum_{i=1}^{m}\frac{1}{-f_i(x)}\nabla^2 f_i(x)$$

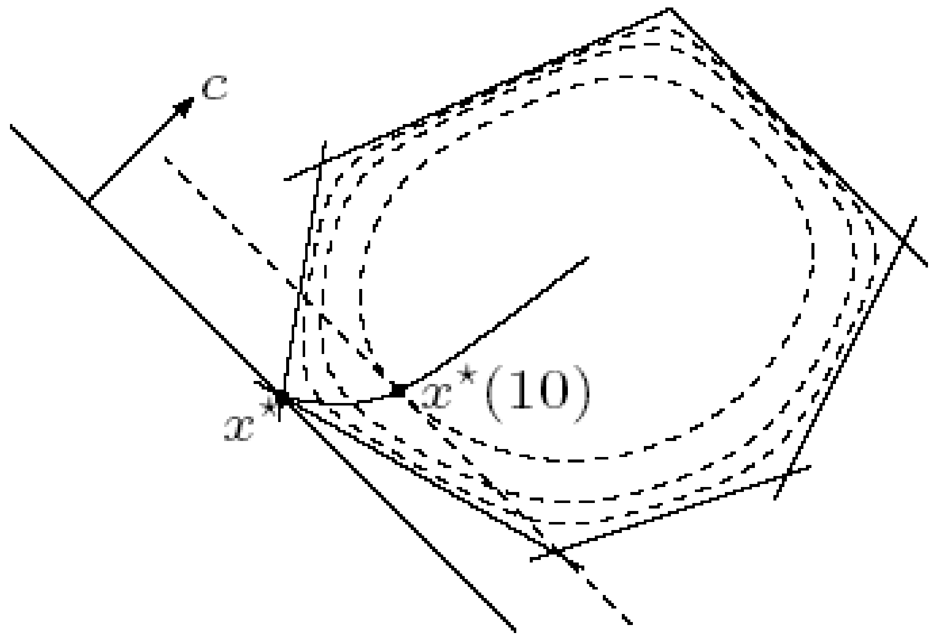
信息与通信工程学院 庄伯金
bjzhuang@bupt.edu.cn

中心线

- 对数阀近似问题的等价问题：

$$\begin{aligned} &\text{minimize} \quad tf_0(x)+\phi(x), t>0 \\ &\text{subject to} \quad Ax=b \end{aligned}$$

- 最优解为 $x^*(t)$，则最优解集 $\{x^*(t) \mid t>0\}$ 称为优化问题的中心线。



中心线的对偶点

- 设 $x = x^*(t)$，则存在w满足**KKT**条件：

$$t\nabla f_0(x) + \sum_{i=1}^{m} \frac{1}{-f_i(x)} \nabla f_i(x) + A^T w = 0, Ax = b$$

- 令 $\lambda_i^*(t) = -\dfrac{1}{tf_i(x^*(t))}, v^*(t) = w/t$

则 $x = x^*(t)$ 是拉格朗日函数$L(x, \lambda^*(t), v^*(t))$的最小值解。

$$L(x, \lambda^*(t), v^*(t)) = f_0(x) + \sum_{i=1}^{m} \lambda_i^*(t) f_i(x) + v^*(t)(Ax - b)$$

- $(\lambda^*(t), v^*(t))$为对偶问题的可行解。

中心线的对偶点

- 设 p^* 为原始问题的最优值，则有：

$$\begin{aligned} p^* &\geq g(\lambda^*(t), v^*(t)) \\ &= L(x^*(t), \lambda^*(t), v^*(t)) \\ &= f_0(x^*(t)) - m/t \end{aligned}$$

- 因此，当 $t \to \infty$ 时，有 $f_0(x^*(t)) \to p^*$。$x^*(t)$ 为原始问题的 $m/t-$次优解。

阀方法

- 初始化：给定严格可行解 x , $t>0$, $\mu>1$ ， 及 $\varepsilon>0$
- **LOOP**：
  - 中心步骤：以 x 为初始点求解优化问题 $x^*(t)$,

$$\begin{array}{ll} \text{minimize} & tf_0(x)+\phi(x) \\ \text{subject to} & Ax=b \end{array}$$

  - 迭代： $x=x^*(t)$
  - 终止条件：若 $m/t<\varepsilon$ ， 则终止退出。
  - 更新 t : $t=\mu t$

收敛性分析

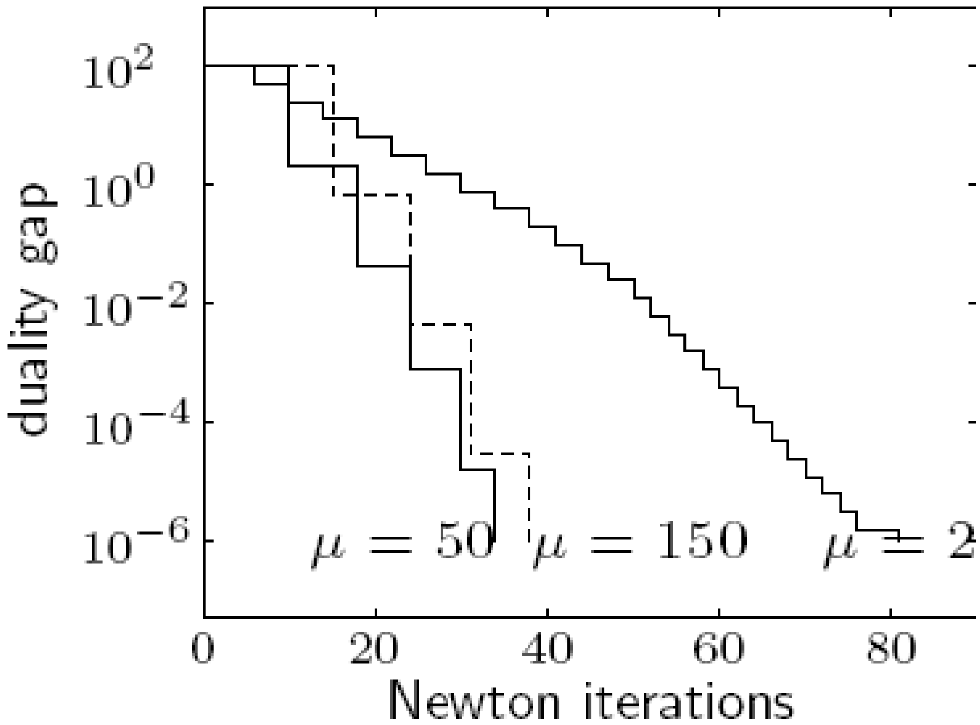
- 外层循环迭代次数：

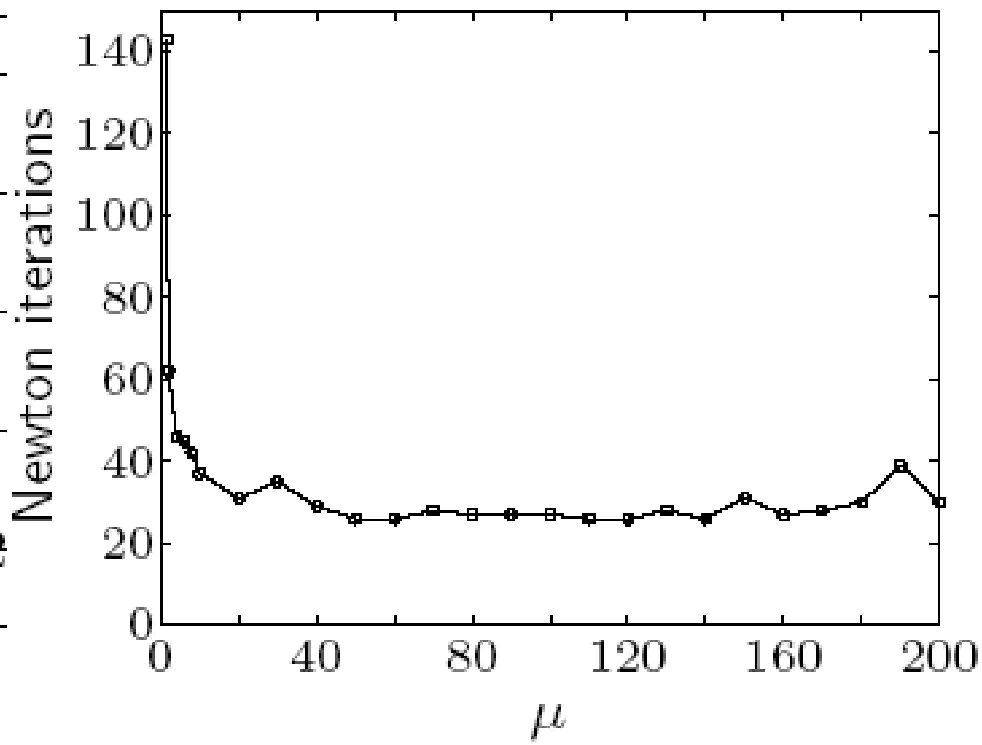
$$\left\lceil \frac{\log(m/(\varepsilon t^{(0)}))}{\log \mu} \right\rceil$$

- 中心步骤实质为一个无约束或等式约束优化问题，其收敛性分析与相应优化问题的收敛性分析结果一致。

例：

- **LP问题**：$m = 100, n = 50$
- 初始值：$t^{(0)} = 1, \varepsilon = 10^{-6}$





信息与通信工程学院 庄伯金
bjzhuang@bupt.edu.cn

第一阶段方法

- 对于不等式约束的优化问题，如何寻找严格可行解或验证不可解？
- 求解优化问题：

$$\begin{aligned} &\text{minimize} && s \\ &\text{subject to} && f_i(x) \le s, i=1,\dots,m \\ &&& Ax=b \end{aligned}$$

- 该问题最优解存在，假设最优值为 $\overline{p}^*$
  - 当 $\overline{p}^* < 0$ 时，存在严格可行解；
  - 当 $\overline{p}^* > 0$ 时，原始问题不可解；
  - 当 $\overline{p}^* = 0$ 时，无法准确确定。

第一阶段方法

- 优化目标为逐项之和：

$$
\begin{aligned}
\text{minimize} \quad & 1^T s \\
\text{subject to} \quad & f_i(x) \le s_i, i = 1, \ldots, m \\
& Ax = b
\end{aligned}
$$

- 对于固定的 x ，$s_i = \max\{f_i(x), 0\}$

寻找严格可行解的方法

- 牛顿法求解优化问题：

$$\begin{aligned} &\text{minimize} \quad s \\ &\text{subject to} \quad f_i(x) \le s, i = 1, \ldots, m \\ &\qquad\qquad\qquad Ax = b \end{aligned}$$

- 迭代终止条件：当前解 $s^{(k)} < 0$ ，即终止迭代，严格可行解为 $x^{(k)}$。