运筹学

（非线性规划）

王焕钢

清华大学自动化系

基础知识（凸函数、可行下降迭代算法）

一维搜索（直线搜索）

无约束优化（下降方向）

约束优化（KKT条件）

算法（简约梯度、罚函数、障碍函数）

拉格朗日对偶

运筹学

（非线性规划基础）

王焕钢

清华大学自动化系

要点：局部最优解与全局最优解

非线性规划的一般形式

$$\begin{aligned}\min\ & f(X)\\ \text{s.t.}\ & h_i(X)=0,\ i=1,2,\cdots,m\\ & g_j(X)\le 0,\ j=1,2,\cdots,l\end{aligned}$$

其中 $X=(x_1,x_2,\cdots,x_n)^T\in R^n$

定以可行集为

$$\Omega=\left\{X\in R^n \,\middle|\, h_i(X)=0,\ 1\le i\le m, g_j(X)\le 0, 1\le j\le l\right\}$$

上述一般形式可简写成 $\min\limits_{X\in\Omega} f(X)$

局部最优解和全局最优解

范数： $\|X\|=\left(\sum_{i=1}^{n} x_i^{\,2}\right)^{\frac{1}{2}}, \forall X \in R^n$

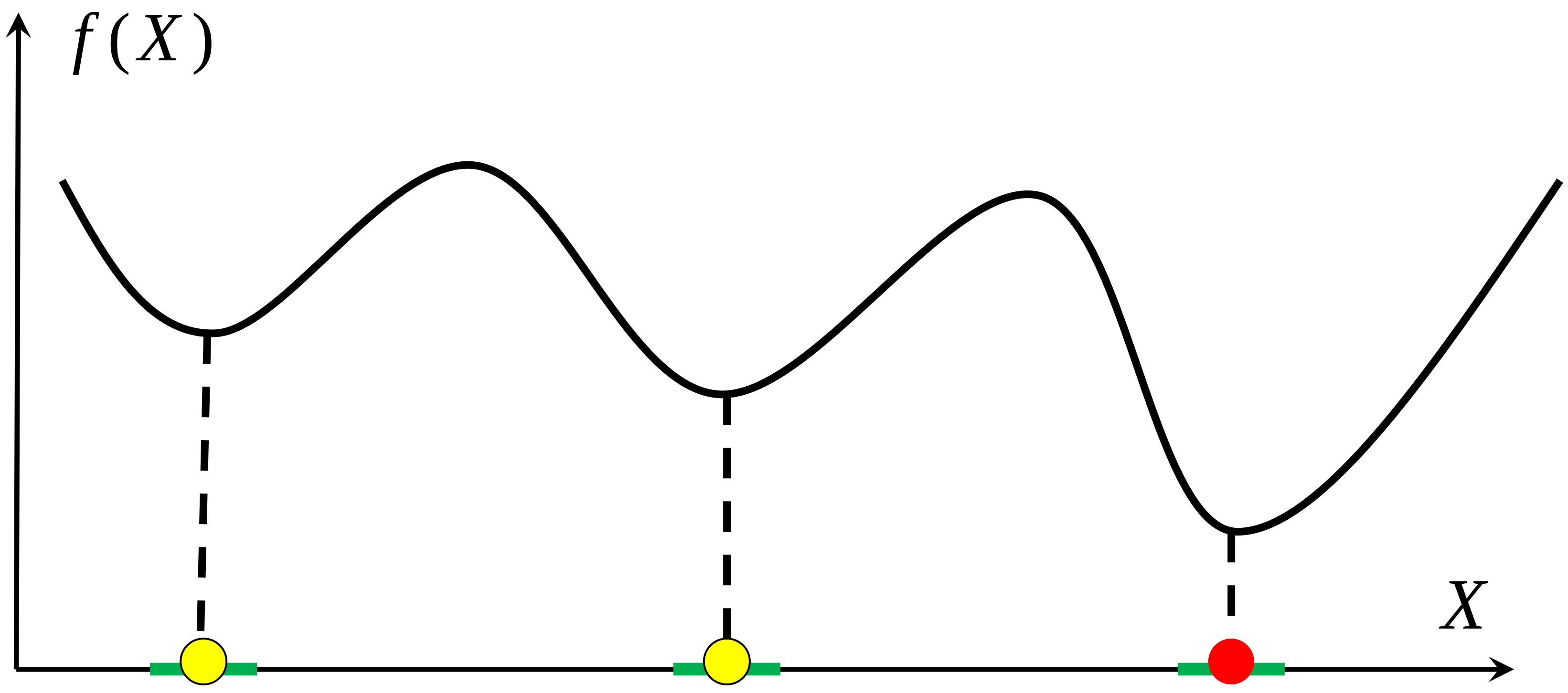
$\hat{X}$ 的 ε 邻域： $B\left(\hat{X},\varepsilon\right)=\left\{X \in R^n \,\middle|\, \left\|X-\hat{X}\right\|<\varepsilon\right\}$

局部最优解 $\hat{X}$ ： $\hat{X} \in \Omega$ ，且存在 $\varepsilon > 0$ 满足

$$f\left(\hat{X}\right)\le f\left(X\right),\ \forall X \in B\left(\hat{X},\varepsilon\right)\cap\Omega$$

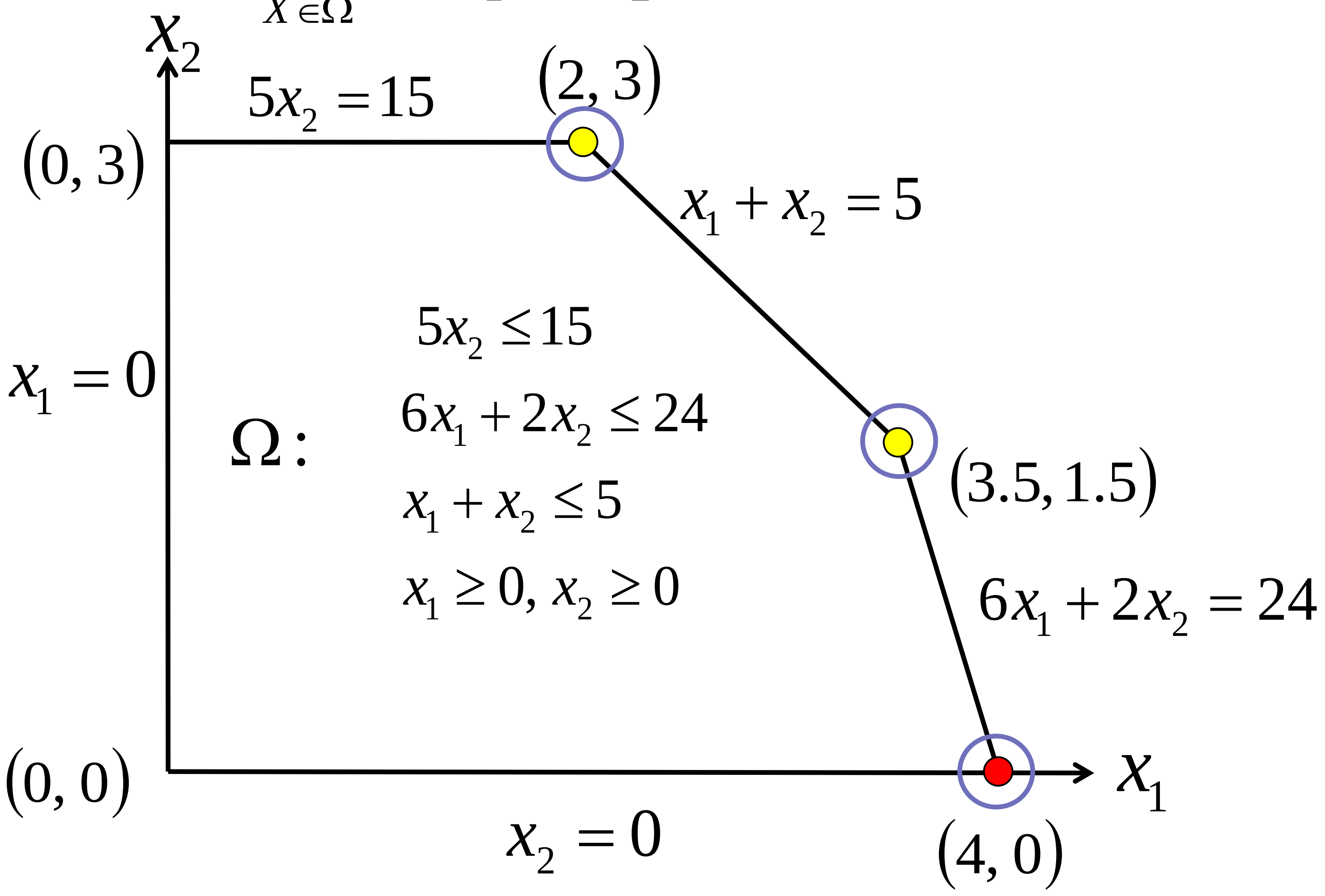
全局最优解 $\hat{X}$ ： $\hat{X} \in \Omega$ ， $f\left(\hat{X}\right)\le f\left(X\right),\ \forall X \in \Omega$

如果在上面的定义中满足 $f\left(\hat{X}\right)< f\left(X\right)$ ，则称为严格局部最优解和严格全局最优解



邻域　局部最优解　全局最优解

$$\min_{X \in \Omega} \; -x_1^{\,2} - x_2^{\,2}$$



邻域　局部最优解　全局最优解

要点：梯度与海赛（Hesse）矩阵

标量函数求偏导数（梯度）

$$\nabla f(X)=\frac{\partial f(X)}{\partial X}=\begin{pmatrix}\dfrac{\partial f(X)}{\partial x_1}\\ \dfrac{\partial f(X)}{\partial x_2}\\ \vdots\\ \dfrac{\partial f(X)}{\partial x_n}\end{pmatrix}$$

$$\nabla^T f(X)=\frac{\partial f(X)}{\partial X^T}=\left(\frac{\partial f(X)}{\partial x_1},\frac{\partial f(X)}{\partial x_2},\cdots,\frac{\partial f(X)}{\partial x_n}\right)$$

向量函数求偏导数

$$F(X)=\left(f_1(X),\, f_2(X),\cdots,\, f_m(X)\right)^T$$

$$\frac{\partial F^T(X)}{\partial X}=\left(\frac{\partial f_1(X)}{\partial X},\frac{\partial f_2(X)}{\partial X},\cdots,\frac{\partial f_m(X)}{\partial X}\right)_{n\times m}$$

$$=\left(\nabla f_1(X),\nabla f_2(X),\cdots,\nabla f_m(X)\right)_{n\times m}$$

$$\frac{\partial F(X)}{\partial X^T}=\begin{pmatrix}\frac{\partial f_1(X)}{\partial X^T}\\ \frac{\partial f_2(X)}{\partial X^T}\\ \vdots\\ \frac{\partial f_m(X)}{\partial X^T}\end{pmatrix}_{m\times n}=\begin{pmatrix}\nabla^T f_1(X)\\ \nabla^T f_2(X)\\ \vdots\\ \nabla^T f_m(X)\end{pmatrix}_{m\times n}$$

海赛（Hesse）矩阵

$$\nabla^2 f\left(X\right)=\frac{\partial\nabla f(X)}{\partial X^T}=\begin{bmatrix} \frac{\partial^2 f(X)}{\partial x_1 x_1} & \frac{\partial^2 f(X)}{\partial x_1 x_2} & \cdots & \frac{\partial^2 f(X)}{\partial x_1 x_n} \\ \frac{\partial^2 f(X)}{\partial x_2 x_1} & \frac{\partial^2 f(X)}{\partial x_2 x_2} & \cdots & \frac{\partial^2 f(X)}{\partial x_2 x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f(X)}{\partial x_n x_1} & \frac{\partial^2 f(X)}{\partial x_n x_2} & \cdots & \frac{\partial^2 f(X)}{\partial x_n x_n} \end{bmatrix}$$

对向量函数的点积求偏导数

$$F(X)=\left(f_1(X), f_2(X),\cdots, f_m(X)\right)^T$$

$$G(X)=\left(g_1(X), g_2(X),\cdots, g_m(X)\right)^T$$

$$\frac{\partial\left(F^T(X)G(X)\right)}{\partial X}=\frac{\partial F^T(X)}{\partial X}G(X)+\frac{\partial G^T(X)}{\partial X}F(X)$$

$$\frac{\partial\left(F^T(X)G(X)\right)}{\partial X^T}=F^T(X)\frac{\partial G(X)}{\partial X^T}+G^T(X)\frac{\partial F(X)}{\partial X^T}$$

对常数矩阵和向量函数的乘积求偏导数

$$A \in R^{m\times m} \quad F(X) = \left(f_1(X), f_2(X), \cdots, f_m(X)\right)^T$$

$$\frac{\partial\left(AF(X)\right)^T}{\partial X} = \frac{\partial\left(F^T(X)A^T\right)}{\partial X} = \frac{\partial F^T(X)}{\partial X} A^T$$

$$\frac{\partial AF(X)}{\partial X^T} = A\frac{\partial F(X)}{\partial X^T}$$

对二次函数求偏导数（ $A^T = A \in R^{m\times m}$ ）

$$\frac{\partial\left(X^T AX\right)}{\partial X} = \frac{\partial X^T}{\partial X} AX + \frac{\partial\left(AX\right)^T}{\partial X} X = \left(A + A^T\right)X = 2AX$$

$$\frac{\partial\left(X^T AX\right)}{\partial X^T} = X^T \frac{\partial\left(AX\right)}{\partial X^T} + \left(AX\right)^T \frac{\partial X}{\partial X^T} = X^T\left(A + A^T\right) = 2X^T A$$

要点：泰勒（Taylor）展开

一元函数在原点的二阶泰勒（Taylor）展开

设一元函数 $g(t)$ 在包含原点的某个开区间内有连续的二阶导数，则对该区间的任意的 $t \geq 0$，存在和 t 有关的 $0 \leq \xi \leq t$ 满足

$$g(t) = g(0) + c_1 t + \frac{1}{2} c_2(\xi) t^2$$

其中

$$c_1 = \left.\frac{dg(t)}{dt}\right|_{t=0} \qquad c_2(\xi) = \left.\frac{d^2 g(t)}{dt^2}\right|_{t=\xi}$$

多元函数在给定点沿给定方向的二阶泰勒展开

设多元函数 $f(X)$ 在包含 $\hat{X}$ 的某个邻域内有连续的二阶偏导数，则对任意的常数向量 $D \in R^n$ ，存在包含原点的某个开区间，对该区间任意的 $t \geq 0$ 存在和 t 与 D 有关的 $0 \leq \xi \leq t$ 满足

$$f(\hat{X}+tD)=f(\hat{X})+c_1 t+\frac{1}{2}c_2(\xi)t^2$$

其中

$$c_1=\left.\frac{df(\hat{X}+tD)}{dt}\right|_{t=0} \qquad c_2(\xi)=\left.\frac{d^2 f(\hat{X}+tD)}{dt^2}\right|_{t=\xi}$$

二阶泰勒展开

$$\frac{df(\hat{X}+tD)}{dt}=\frac{\partial f(\hat{X}+tD)}{\partial X^T}\frac{d(\hat{X}+tD)}{dt}=\nabla^T f(\hat{X}+tD)D$$

$$\frac{d^2 f(\hat{X}+tD)}{dt^2}=\frac{d(\hat{X}+tD)^T}{dt}\nabla^2 f\left(\hat{X}+tD\right)D=D^T\nabla^2 f\left(\hat{X}+tD\right)D$$

其中 $\nabla^2 f\left(X\right)=\dfrac{\partial\nabla^T f(X)}{\partial X}$ 为海赛（Hesse）矩阵

所以 $c_1=\nabla^T f(\hat{X})D,\ c_2(\xi)=D^T\nabla^2 f\left(\hat{X}+\xi D\right)D$

$$f(\hat{X}+tD)=f(\hat{X})+c_1 t+\frac{1}{2}c_2(\xi)t^2$$

$$=f(\hat{X})+\nabla^T f(\hat{X})Dt+\frac{1}{2}D^T\nabla^2 f(\hat{X}+\xi D)Dt^2$$

要点：凸函数和凹函数

凸集上的凸函数和凹函数

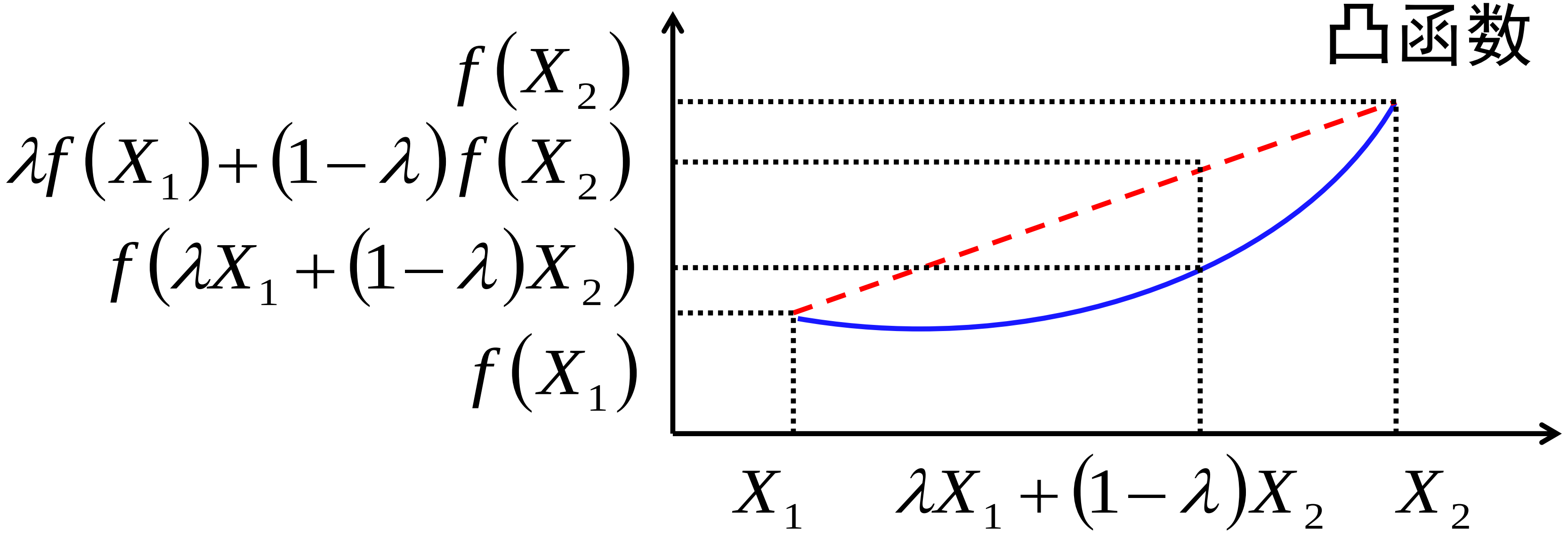
设 $f(X)$ 是定义在集合 $\Omega \subset R^n$ 上的函数，如果 Ω 是凸集，并且对 Ω 中任意两点 X_1, X_2 以及闭区间 $[0,\ 1]$ 中任意一点 λ 都满足

$$f\left(\lambda X_1 + (1-\lambda) X_2\right) \leq \lambda f\left(X_1\right) + (1-\lambda) f\left(X_2\right)$$

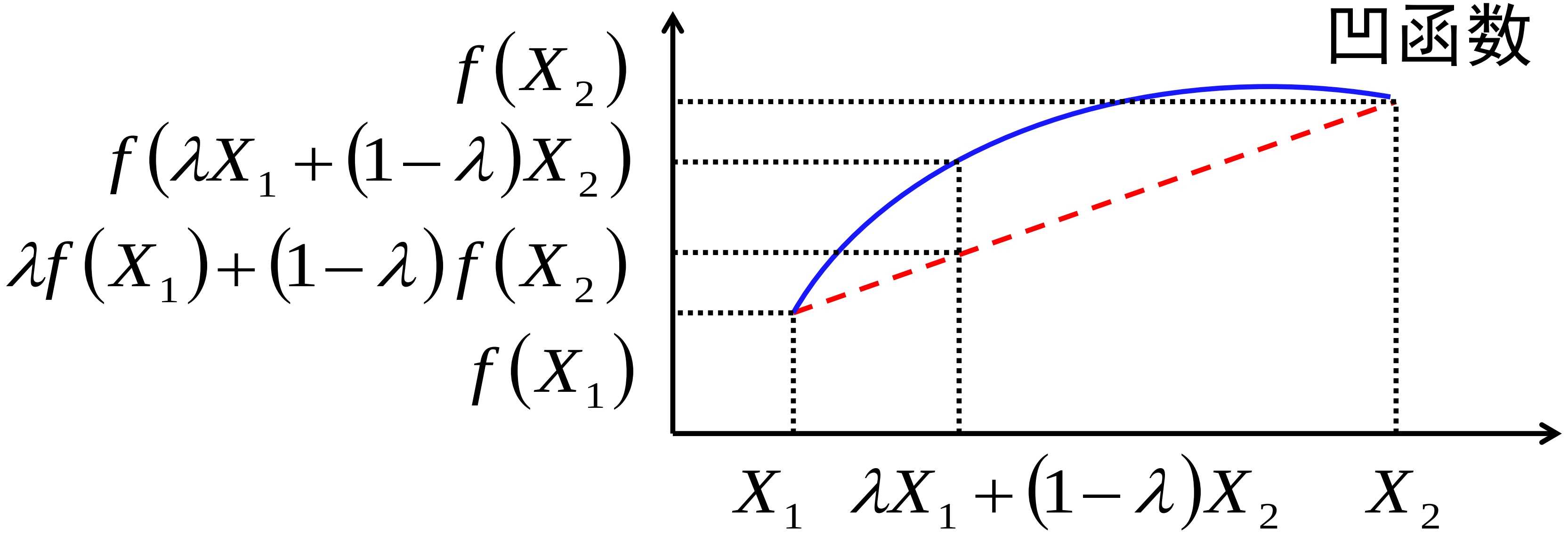
则称 $f(X)$ 是（凸集 Ω上的）凸函数，如果 $-f(X)$ 是（凸集 Ω上的）凸函数，则称 $f(X)$ 是（凸集 Ω上的）凹函数，此时在上面的条件下应满足

$$f\left(\lambda X_1 + (1-\lambda) X_2\right) \geq \lambda f\left(X_1\right) + (1-\lambda) f\left(X_2\right)$$

一元凸（凹）函数的图象



凸函数
$f(X_2)$
$\lambda f(X_1)+(1-\lambda)f(X_2)$
$f(\lambda X_1+(1-\lambda)X_2)$
$f(X_1)$
X_1
$\lambda X_1+(1-\lambda)X_2$
X_2



凹函数
$f(X_2)$
$f(\lambda X_1+(1-\lambda)X_2)$
$\lambda f(X_1)+(1-\lambda)f(X_2)$
$f(X_1)$
X_1
$\lambda X_1+(1-\lambda)X_2$
X_2

要点：多元凸函数的判别

多元可导凸（凹）函数的一阶充要条件

$$\nabla^T f(X_1)(X_2-X_1)\le(\ge)\ f(X_2)-f(X_1),\ \ \forall X_1,X_2\in\Omega$$

必要性：

记 $D=X_2-X_1,\quad X_\lambda=X_1+\lambda D=(1-\lambda)X_1+\lambda X_2,$

利用二阶泰勒展开可得

$$f(X_\lambda)=f(X_1)+\nabla^T f(X_1)D\lambda+\frac{1}{2}D^T\nabla^2 f(X_1+\xi D)D\lambda^2$$

$$f(X_\lambda)-\left((1-\lambda)f(X_1)+\lambda f(X_2)\right)$$

$$\Rightarrow =\lambda\left(f(X_1)-f(X_2)+\nabla^T f(X_1)D+\frac{1}{2}D^T\nabla^2 f(X_1+\xi D)D\lambda\right)$$

令 λ 充分小，由凸（凹）性可得上面的不等式

充分性：

记 $X_\lambda=(1-\lambda)X_1+\lambda X_2$，则 $X_\lambda\in\Omega$，利用给定条件

$$\nabla^T f(X_1)(X_2-X_1)\le\ (\ge)\ f(X_2)-f(X_1),\ \forall X_1,X_2\in\Omega$$

可得

$$\nabla^T f(X_\lambda)(X_1-X_\lambda)\le\ (\ge)\ f(X_1)-f(X_\lambda)$$
$$\nabla^T f(X_\lambda)(X_2-X_\lambda)\le\ (\ge)\ f(X_2)-f(X_\lambda)$$

用 λ 和 $1-\lambda$ 分别乘以上两式再相加，再利用

$$\lambda(X_1-X_\lambda)+(1-\lambda)(X_2-X_\lambda)=0$$

可得

$$f(X_\lambda)\le(\ge)\lambda f(X_1)+(1-\lambda)f(X_2)$$

要点：多元凸函数与一元凸函数关系

多元凸函数和一元凸函数的关系

对定义在凸集 $\Omega \subset R^n$ 上的多元函数 $f(\cdot)$，任取 $X \in \Omega$ 和 $D \in R^n$ 定义一元函数 $\varphi(\cdot \mid X, D)$ 如下

$$\varphi(t \mid X, D) = f(X + tD),\ \forall X + tD \in \Omega$$

则存在以下关系

$$f(\cdot)\ \text{是凸函数} \Leftrightarrow \varphi(\cdot \mid X, D), \forall X \in \Omega, D \in R^n\ \text{是凸函数}$$

证明：正向利用

$$\varphi\left((1-\lambda)t_1 + \lambda t_2 \mid X, D\right) = f\left((1-\lambda)(X + t_1 D) + \lambda(X + t_2 D)\right)$$

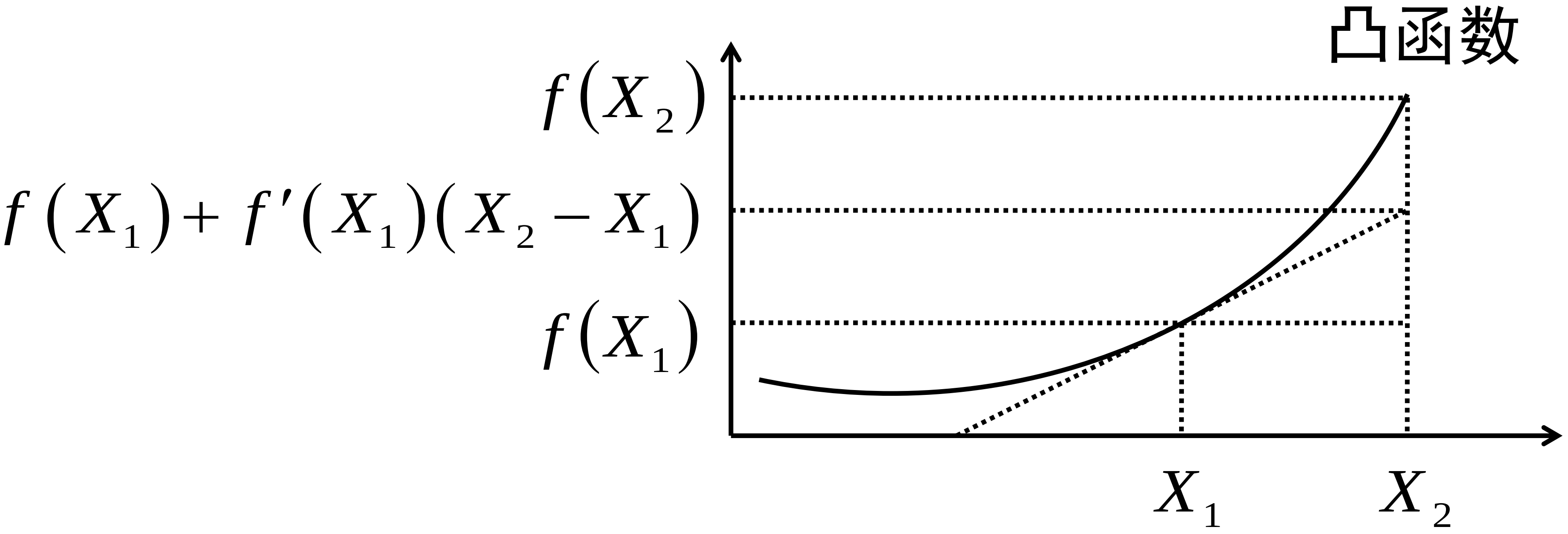
反向利用 $f\left((1-t)X_1 + tX_2\right) = \varphi\left((1-t)\times 0 + t\times 1 \mid X_1, X_2 - X_1\right)$

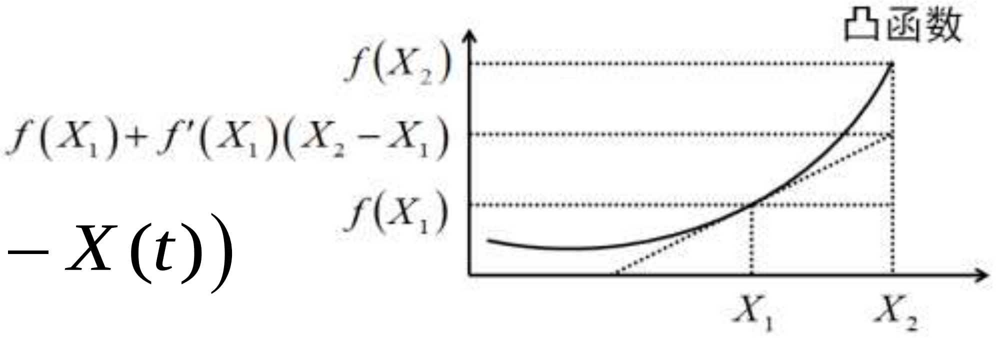
要点：凸函数的一阶条件

一元可导凸（凹）函数的充要条件

凸函数 $f(X_1)+\nabla f(X_1)(X_2-X_1)\le f(X_2)$

凹函数 $f(X_1)+\nabla f(X_1)(X_2-X_1)\ge f(X_2)$





充分性： $X(t)=(1-t)X_1+tX_2$

$$f(X_1)-f(X(t))\geq f'(X(t))(X_1-X(t))$$

$$f(X_2)-f(X(t))\geq f'(X(t))(X_2-X(t))$$

$$\Rightarrow\quad (1-t)\big(f(X_1)-f(X(t))\big)+t\big(f(X_2)-f(X(t))\big)\geq 0$$

$$\Leftrightarrow\quad (1-t)f(X_1)+tf(X_2)\geq f\big((1-t)X_1+tX_2\big)$$

必要性： $f\big((1-t)X_1+tX_2\big)\leq(1-t)f(X_1)+tf(X_2)$

$$\Leftrightarrow\quad \frac{f\big(X_1+t(X_2-X_1)\big)-f(X_1)}{t}\leq f(X_2)-f(X_1)$$

（只需要一阶导数存在）

多元可导凸函数的一阶充要条件

$$\nabla^T f(X_1)(X_2 - X_1) \le f(X_2) - f(X_1),\ \forall X_1, X_2 \in \Omega$$

充分性： $g(t_i) = f(X + t_i D) = f(X_i),\ i = 1, 2$

$$\Rightarrow \quad g'(t_1) = \nabla f(X + t_1 D)^T D$$

$$\Rightarrow \quad g(t_2) - g(t_1) = f(X_2) - f(X_1) \ge \nabla f(X_1)^T (X_2 - X_1)$$

$$= \nabla f(X + t_1 D)^T D(t_2 - t_1) = g'(t_1)(t_2 - t_1)$$

必要性： $g'(t) = \nabla f(X_1 + t(X_2 - X_1))^T (X_2 - X_1)$

$$g'(0) = \nabla f(X_1)^T (X_2 - X_1)$$

$$g'(0)(1-0) \le g(1) - g(0),\ \forall X_1, X_2 \in \Omega$$

要点：凸函数的二阶条件

一元可导凸函数的二阶充分条件

$$f''(X) \geq 0,\ \forall X \in \Omega$$

证明

已知条件 $\Rightarrow$ $f'(X_2) \geq f'(X_1),\ \forall X_2 \geq X_1$

情况1 $X_2 \geq X_1\ (\xi \geq X_1)$

$$f(X_2) - f(X_1) = f'(\xi)(X_2 - X_1) \geq f'(X_1)(X_2 - X_1)$$

情况2 $X_2 \leq X_1\ (\xi \leq X_1)$

$$f(X_2) - f(X_1) = f'(\xi)(X_2 - X_1) \geq f'(X_1)(X_2 - X_1)$$

由一阶充要条件知结论成立（只需要二阶导数存在）

一元可导凸函数的二阶必要条件

$$f''(X) \geq 0,\ \forall X \in \Omega \quad \text{其中 } \Omega \text{ 不是单点集}$$

证明　必存在 $|\Delta X| \in B(0,\varepsilon) \quad \Rightarrow \quad X+\Delta X \in \Omega$

$$\Rightarrow \quad f'(X)(\Delta X) \leq f(X+\Delta X) - f(X)$$

$$f'(X+\Delta X)(-\Delta X) \leq f\left(X+\Delta X+(-\Delta X)\right) - f(X+\Delta X)$$

$$\Rightarrow \quad \left(f'(X+\Delta X) - f'(X)\right)\Delta X \geq 0$$

$$\Rightarrow \quad \frac{\left(f'(X+\Delta X) - f'(X)\right)(\Delta X)^2}{\Delta X} \geq 0 \quad \Rightarrow \quad f''(X) \geq 0$$

多元可导凸函数的二阶充分条件

$$\nabla^2 f(X) \geq 0,\ \forall X \in \Omega$$

证明途径

$$g(t) = f\left(X_1 + t(X_2 - X_1)\right)$$

$$g'(t) = \nabla f\left(X_1 + t(X_2 - X_1)\right)^T (X_2 - X_1)$$

$$g''(t) = (X_2 - X_1)^T \nabla^2 f\left(X_1 + t(X_2 - X_1)\right)(X_2 - X_1)$$

$$\nabla^2 f(X) \geq 0,\ \forall X \in \Omega \quad \Rightarrow \quad g''(t) \geq 0$$

$\Rightarrow \quad g(t)$ 是一元凸函数

多元可导凸函数的二阶必要条件

$$\nabla^2 f(X) \geq 0,\ \forall X \in \Omega$$ 其中 Ω 是开集

证明途径

对任意的 $X \in \Omega$ 和模充分小的 $\Delta X \in R^n$

$g(t) = f(X + t\Delta X)$ 是含零开区间上的凸函数

$$g'(t) = \nabla f(X_1 + t\Delta X)^T \Delta X$$

$$g''(t) = \Delta X^T \nabla^2 f(X + t\Delta X)\Delta X$$

$$\Delta X^T \nabla^2 f(X)\Delta X = g''(0) \geq 0 \quad \Rightarrow \quad \nabla^2 f(X) \geq 0$$

要点：凸规划问题

凸性对优化问题的基本作用

如果 Ω 是凸集，$f(X)$ 是其上的连续凸函数，称

$$\min_{X \in \Omega} f(X)$$

是凸规划问题

如果 $X^* \in \Omega$ 是凸规划问题的任意一个局部最优解，那么它也是该问题的全局最优解

证明：如果存在 $\hat{X} \in \Omega$ 满足 $f(\hat{X}) < f(X^*)$

$$\Rightarrow \lambda f(\hat{X}) < \lambda f(X^*), \quad \forall \lambda > 0$$

又因为

$$f\left(\lambda \hat{X} + (1-\lambda) X^*\right) \le \lambda f(\hat{X}) + (1-\lambda) f(X^*), \ \forall 0 \le \lambda \le 1$$

$$\Rightarrow \ f\left(\lambda \hat{X} + (1-\lambda) X^*\right) < f(X^*), \ \forall 0 < \lambda \le 1$$

因为对充分小的 $\lambda > 0$，$\lambda \hat{X} + (1-\lambda) X^*$ 能够充分接近 X^*，说明 X^* 不是局部最优解，矛盾！

要点：求解非线性规划的基本方法

函数求极值问题 $\min f(X),\ X \in R^n \quad \Rightarrow \quad \nabla f(X^*) = 0$

无约束最优化是非线性规划（一般函数）的典型问题

求解非线性规划问题的基本途径

“迭代算法”

$$X_{k+1} = X_k + \lambda_k D_k$$

长记性的做法

$\lambda_k \in R^1$ 一维搜索步长、 $D_k \in R^n$ 寻优方向

可行下降方向

对于优化问题 $\min\limits_{X\in\Omega} f(X)$，给定可行解 $\hat{X}\in\Omega$ 以及向量 $D\in R^n$，如果存在 $\bar{t}>0$ 满足

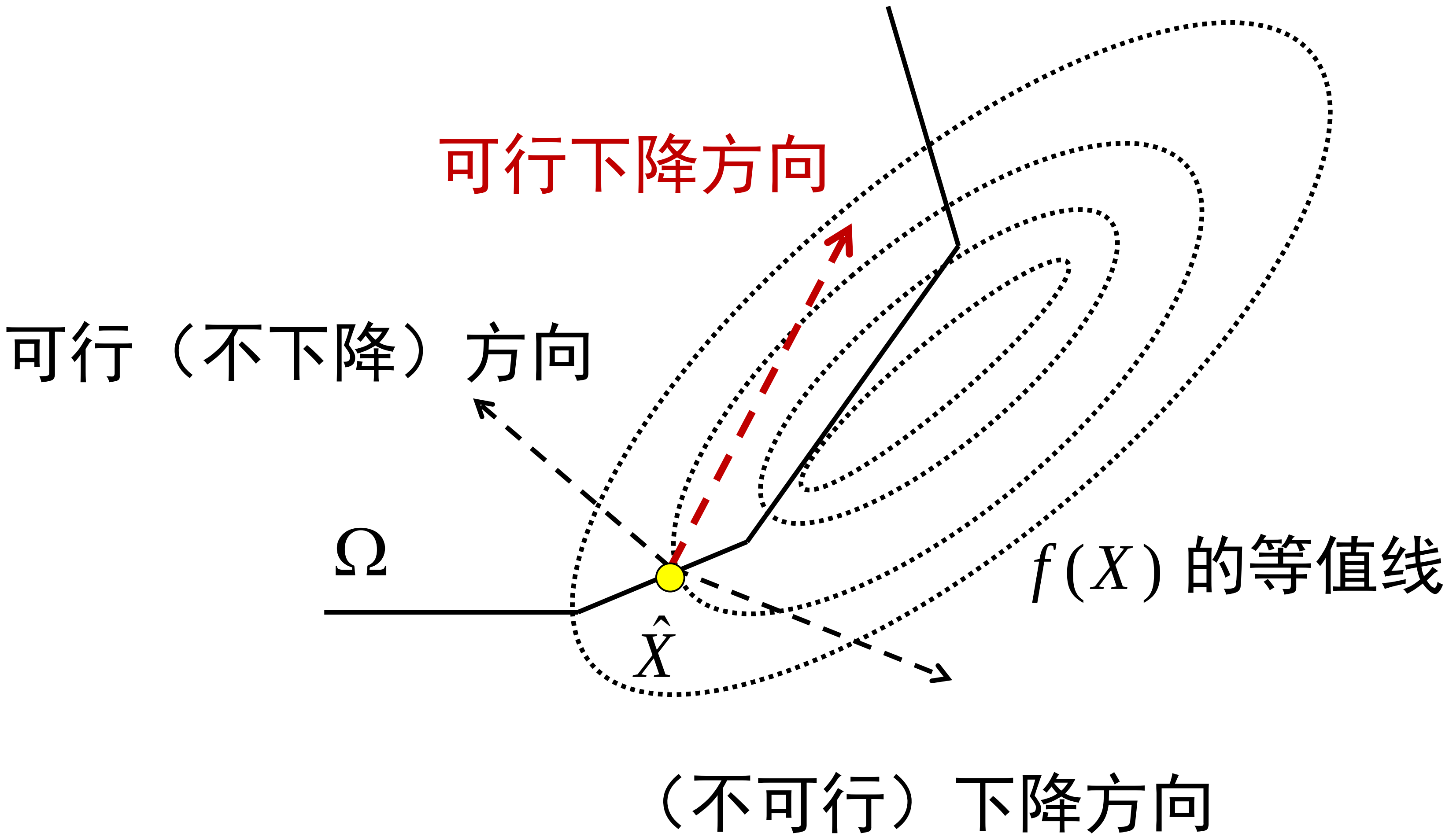
$$\hat{X}+tD\in\Omega,\quad \forall 0<t\le\bar{t}$$

称 D 是 $\hat{X}$ 处的可行方向，如果存在 $\bar{t}>0$ 满足

$$f\left(\hat{X}+tD\right)<f(\hat{X}),\quad \forall 0<t\le\bar{t}$$

称 D 是 $\hat{X}$ 处的下降方向，既可行又下降的方向

称为可行下降方向


可行下降方向
可行（不下降）方向
Ω
f(X) 的等值线
X̂
（不可行）下降方向

$\min\limits_{X\in\Omega} f(X)$ 寻优算法的基本思路：可行下降迭代算法

确定初始可行解 $X_0\in\Omega$ $\rightarrow$ 确定可行下降方向 D_0

$\Rightarrow$ 一维搜索确定 λ_0 满足 $X_1=X_0+\lambda_0 D_0\in\Omega$ 以及

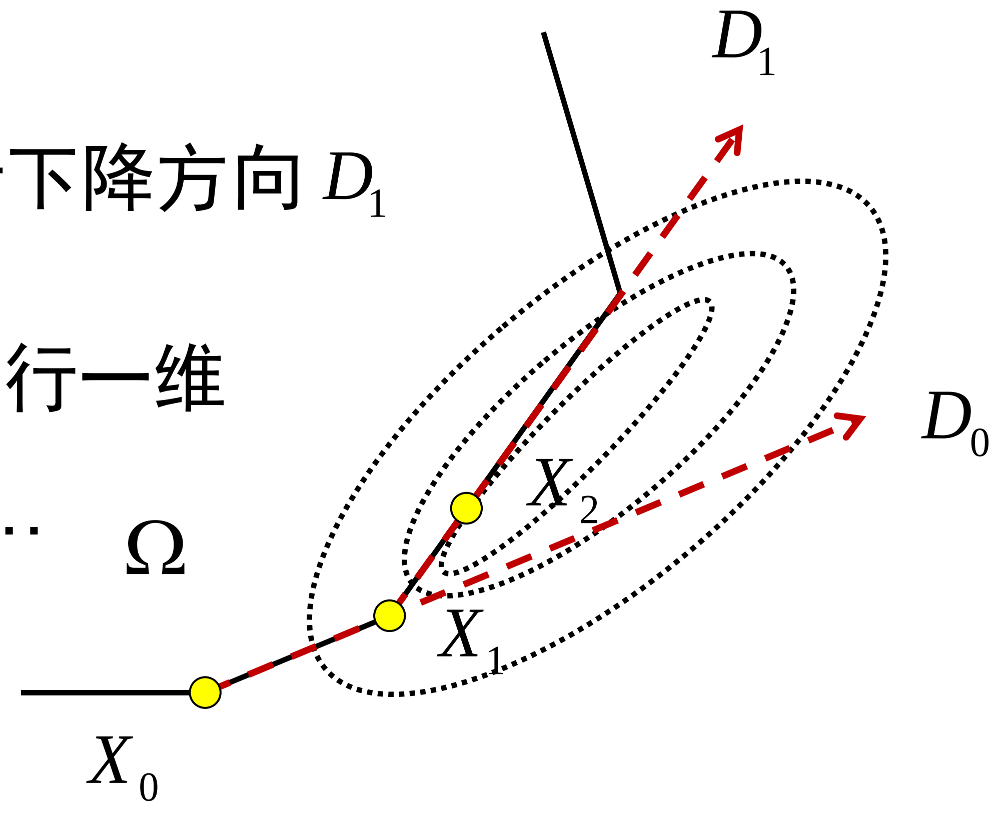
$f(X_1)<f(X_0)$

$\Rightarrow$ 确定 X_1 处的可行下降方向 D_1

$\Rightarrow$ 在 X_1 处沿 D_1 进行一维

搜索确定 X_2 ……

$$X_k=X_{k-1}+\lambda_{k-1}D_{k-1}$$



凸规划最优解与可行下降方向的关系

对于凸规划 $\min\limits_{X\in\Omega} f(X)$ （Ω 凸集，$f(X)$ 凸函数）

$X^* \in \Omega$ 是最优解的充要条件是在该点不存在可行下降方向

必要性显然。充分性反证：如果有另一点更好，连接两点可得可行下降方向。

要点：一维精确搜索的性质

单变量非线性函数寻优问题

$$\min \varphi(t)$$
$$\text{s.t.} \quad t \in [a,b]$$

一维搜索：求解单变量优化问题

$$\begin{aligned} \min \quad & f(X_0 + tD) \\ \text{s.t.} \quad & X_0 + tD \in \Omega \\ & t \geq 0 \end{aligned} \quad \Rightarrow \quad \min_{0 \leq t \leq t_{\max}} \varphi(t)$$

其中 $\varphi(t) = f(X_0 + tD)$，

$$X_0 + tD \in \Omega, \forall 0 \leq t \leq t_{\max}$$

条件：已知 $X_0 \in \Omega$ 和可行下降方向 $D \in R^n$

精确搜索：求解单变量优化问题

$$\begin{aligned} &\min\ f(X_0 + tD) \\ &\text{s.t. } X_0 + tD \in \Omega \\ &\qquad t \geq 0 \end{aligned} \quad \Rightarrow \quad \min_{0 \leq t \leq t_{\max}} \varphi(t)$$

其中 $\varphi(t) = f(X_0 + tD)$ ， $X_0 + tD \in \Omega, \forall 0 \leq t \leq t_{\max}$

非精确搜索：找到一个 $\hat{t} > 0$ 满足

$$X_0 + \hat{t}D \in \Omega,\ f(X_0 + \hat{t}D) < f(X_0 + tD)$$ 且 $\hat{t}$ 不能太小

由于 D 是下降方向，所以 $\varphi'(0) < 0$

一维精确搜索的特性

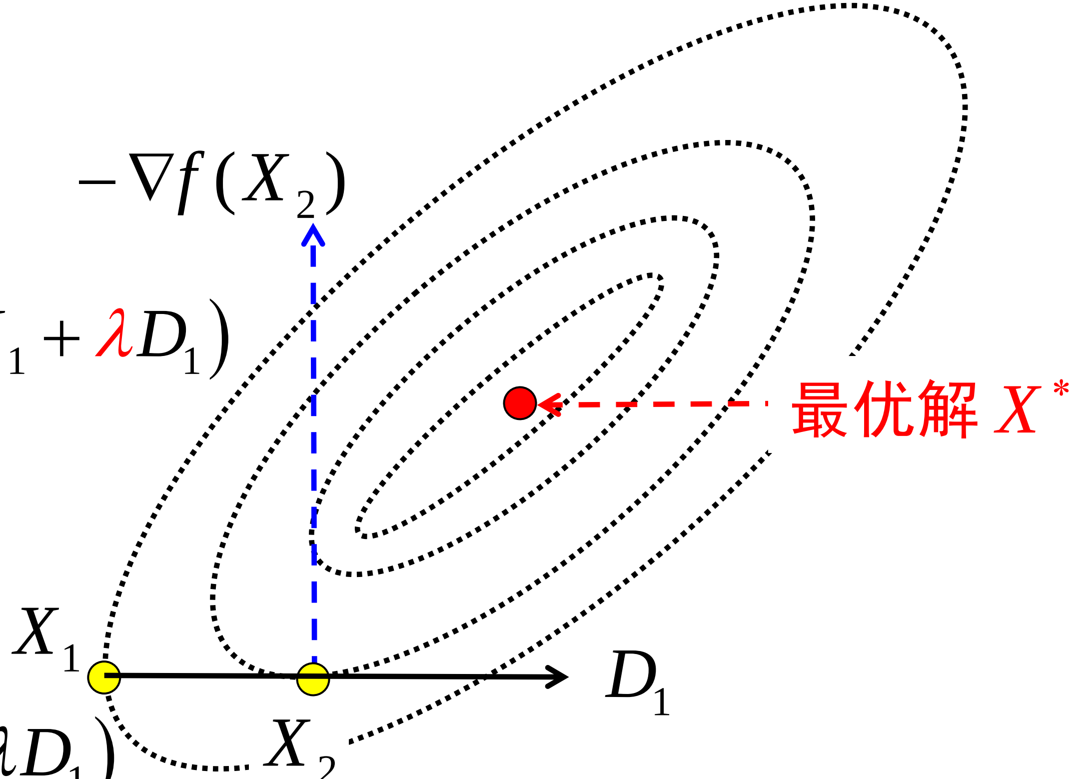
下降方向 D_1

$$X_2 = X_1 + \lambda_1^* D_1$$

λ_1^* 是优化问题 $\min\limits_{\lambda>0} f\left(X_1 + \lambda D_1\right)$ 的最优解

$$\frac{\mathrm{d} f\left(X_1 + \lambda D_1\right)}{\mathrm{d}\lambda} = \frac{\mathrm{d} f\left(X_1 + \lambda D_1\right)}{\mathrm{d} X^T} \frac{\mathrm{d}\left(X_1 + \lambda D_1\right)}{\mathrm{d}\lambda}$$

$$\Rightarrow \quad \nabla^T f\left(X_1 + \lambda_1^* D_1\right) D_1 = 0 \quad \Rightarrow \quad \nabla^T f\left(X_2\right) D_1 = 0$$



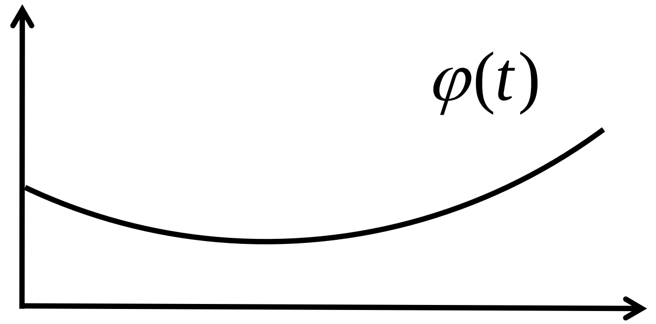
精确搜索得到新点的梯度方向与搜索方向正交

要点：精确搜索的基本途经

精确搜索的基本途经

1）确定初始区间

用试探法确定一个单谷区间可用步长加倍或减倍的方法

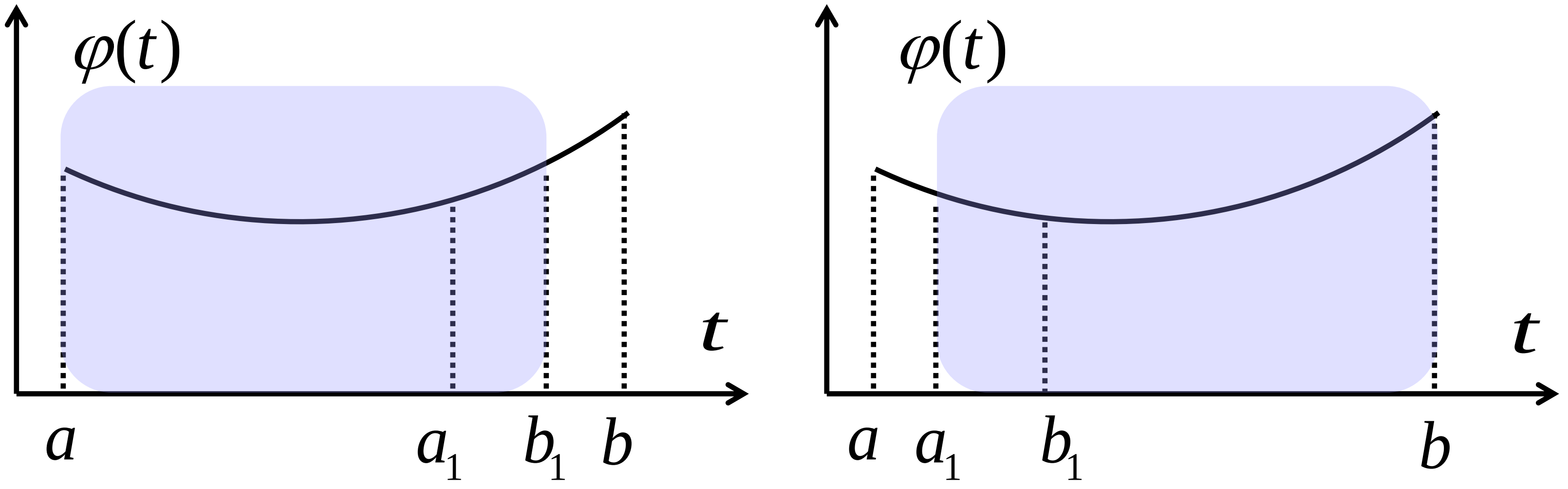


2）逐步压缩区间

按照一定规则在上述区间内选点，通过计算并比较这些点上的函数值，逐步缩小包含局部最优解的区间直至区间长度小于给定阈值

区间压缩原理

已知闭区间 $[a,b]$ 是单谷区间，在其内部任取两点 $a_1 < b_1$ ，计算 $\varphi(a_1),\varphi(b_1)$，如果 $\varphi(a_1) < \varphi(b_1)$ ，局部最优解在区间 $[a,b_1]$，否则局部最优解在区间 $[a_1,b]$，两种情况均能压缩区间

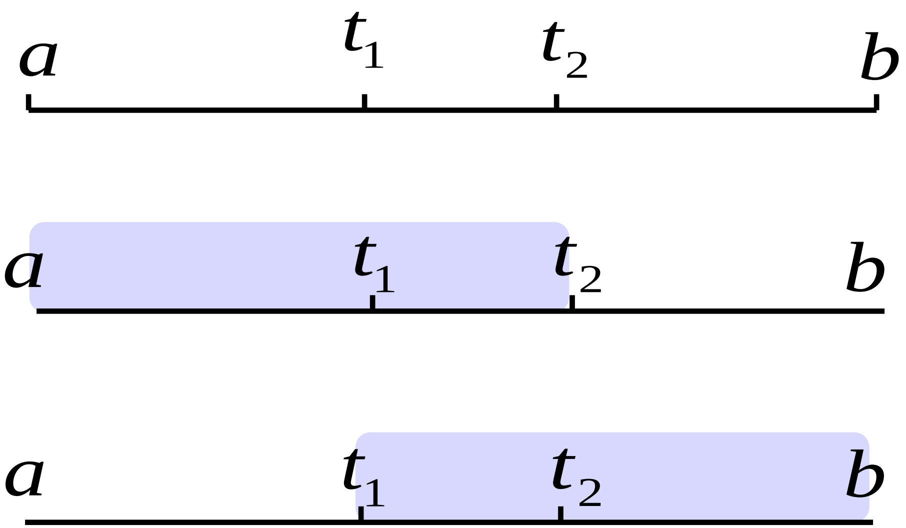


第一种情况　　第二种情况

要点：0.618 法

精确搜索的0.618 法的原理

基本想法：每计算一个函数值能将区间压缩一个固定比值 c



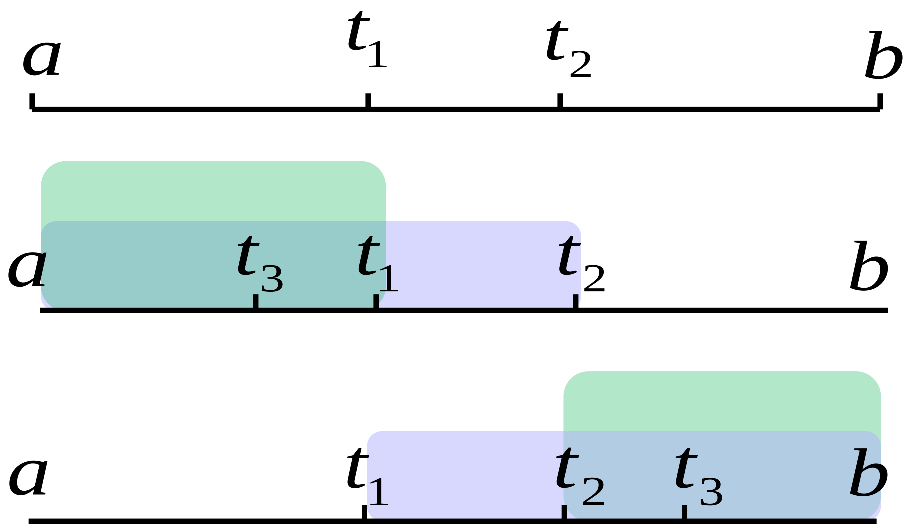
$\varphi(t_1) < \varphi(t_2)$

$\varphi(t_1) > \varphi(t_2)$

则有： $$\frac{t_2 - a}{b - a} = \frac{b - t_1}{b - a} = c$$

精确搜索的0.618 法的原理

基本想法：每计算一个函数值能将区间压缩一个固定比值 c



$\varphi(t_1) < \varphi(t_2)$

$\varphi(t_1) > \varphi(t_2)$

为实现上述想法，必须：

$$\frac{t_2 - a}{b - a} = \frac{b - t_1}{b - a} = c \qquad \frac{t_1 - a}{t_2 - a} = \frac{b - t_2}{b - t_1} = c$$

$$\frac{t_2-a}{b-a}=\frac{b-t_1}{b-a}=c$$

$$\frac{t_1-a}{t_2-a}=\frac{b-t_2}{b-t_1}=c$$

$$\Rightarrow \quad \begin{aligned} t_1&=b-c(b-a)\\ t_2&=a+c(b-a)\\ t_1&=a+c(t_2-a)\\ t_2&=b-c(b-t_1) \end{aligned}$$

由上面右边前三个方程可以解得

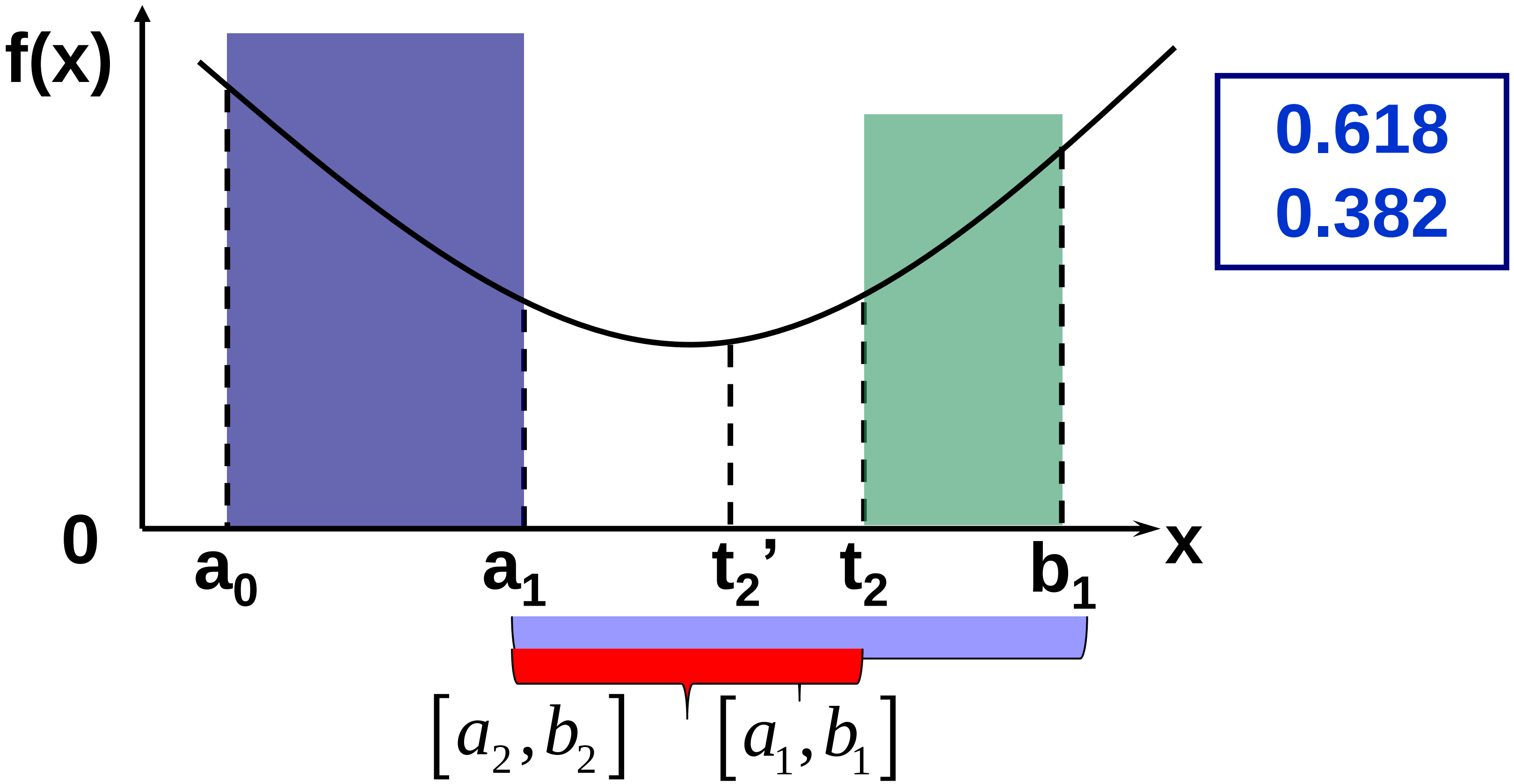
$$c^2=1-c$$

将前两个方程和新得到的方程代入第四个方程可使等式成立，所以满足上式的 c 即所求常数

解二次方程得正数解 $c=\frac{1}{2}\left(\sqrt{5}-1\right)\approx 0.618$

在单谷区间 $[a,b]$ 搜索局部最优解的0.618法

f(x)是区间 $[a,b]$ 上的单峰函数（有唯一极小点）



在单谷区间 $[a,b]$ 搜索局部最优解的0.618法

1）确定误差阈值 δ 及满足 $0.618^{n-1}(b-a)\leq\delta$ 的 n

2）令 $a_0=a, b_0=b$

3）对于 $k=1,2,\cdots,n$ 依次完成以下运算

a）令 $t_k=a_{k-1}+0.618(b_{k-1}-a_{k-1})$

$t'_k=b_{k-1}+0.618(a_{k-1}-b_{k-1})$

b）计算 $\varphi(t_k)$ 和 $\varphi(t'_k)$ 中未知的数值

c）比较 $\varphi(t_k)$ 和 $\varphi(t'_k)$ 的大小确定 $[a_k,b_k]$

4）取所求局部最优解为 $0.5(a_n+b_n)$

要点：Fibonacci 法

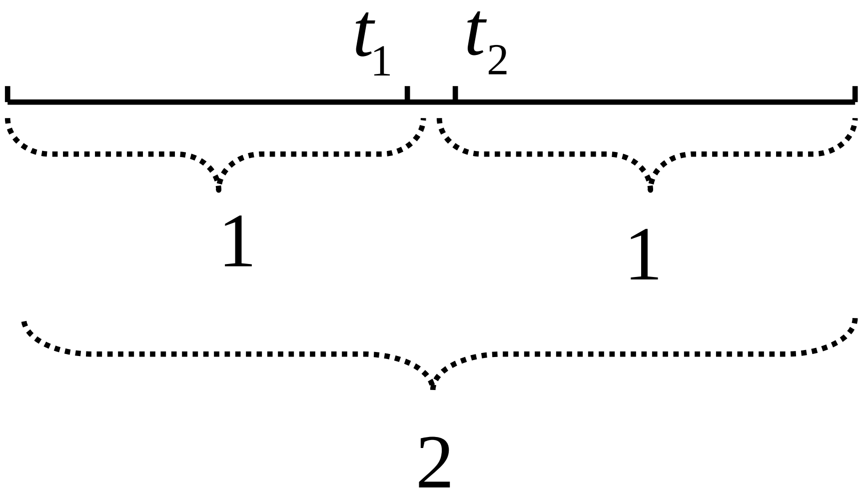
斐波那契（Fibonacci）法（最优选点方法）

用 F_n 表示某个区间长度，通过计算 n 点函数值能将该区间压缩为一个单位长度，但任何大于 F_n 的区间长度都不能保证做到这一点

由于不计算函数值和只计算一点的函数值都不能压缩区间，所以

$$F_0 = F_1 = 1$$

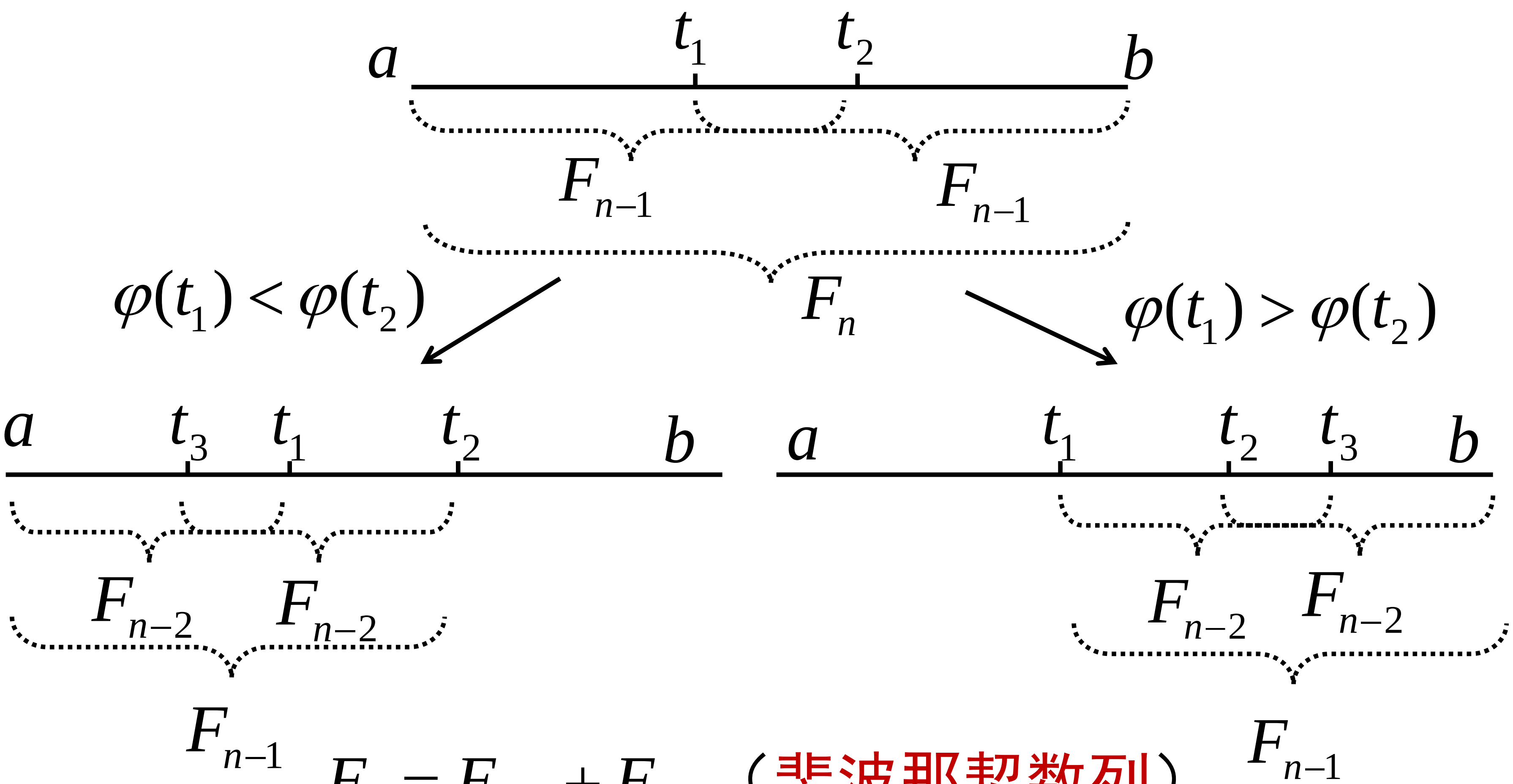
当 $n=2$ 时，由于两个分点可以任意接近中点，如下面的 t_1 和 t_2 所示，所以能将两个单位的区间压缩为一个单位的区间，即 $F_2=2$



前三个数的关系

$$F_0=1,\ F_1=1,\ F_2=F_1+F_0$$

一般情况下，F_n, F_{n-1}, F_{n-2} 必须满足下图所示关系



$F_n = F_{n-1} + F_{n-2}$（斐波那契数列）

结论：

$$t_1 = b + \frac{F_{n-1}}{F_n}(a-b),\ t_2 = a + \frac{F_{n-1}}{F_n}(b-a)$$

如何确定分点数 n ？

假设初始区间是 $[a,b]$，容许误差是 $\delta>0$

将 δ 视为一个单位的长度，则在 $[a,b]$ 中一共有

$\frac{b-a}{\delta}$ 个单位长度，只要取 n 为满足

$$F_n \geq \frac{b-a}{\delta}$$

的最小整数，通过计算 n 个分点的函数值，一定能将包含最优解的区间压缩为不大于一个单位的长度，即小于或等于 δ

在单谷区间 $[a,b]$ 搜索局部最优解的斐波那契法

1）确定误差阈值 δ 以及满足 $F_n \geq \delta^{-1}(b-a)$ 的 n

2）令 $a_0 = a, b_0 = b$

3）对于 $k = 1,2,\cdots,n-1$ 依次完成以下运算

a）令 $t_k = a_{k-1} + (F_{n-k}/F_{n-k+1})(b_{k-1} - a_{k-1})$

$$t_k' = b_{k-1} + (F_{n-k}/F_{n-k+1})(a_{k-1} - b_{k-1})$$

b）计算 $\varphi(t_k)$ 和 $\varphi(t_k')$ 中未知的数值

c）比较 $\varphi(t_k)$ 和 $\varphi(t_k')$ 的大小确定 $[a_k, b_k]$

4）取所求局部最优解为？

0.618法和斐波那契法的关系

由 $F_n = F_{n-1} + F_{n-2}$ 可得 $\dfrac{F_n}{F_{n-1}} = 1 + \dfrac{F_{n-2}}{F_{n-1}}$

在上面右式令 $n \to \infty$ 并定义 $F_* = \lim\limits_{n\to\infty} \dfrac{F_{n-1}}{F_n}$

可得 $\dfrac{1}{F_*} = 1 + F_*$ 等价于 $(F_*)^2 + F_* - 1 = 0$

解二次方程得正数解 $F_* = \dfrac{1}{2}\left(\sqrt{5} - 1\right) \approx 0.618$

所以0.618法实际上就是用斐波那契法中分数数列的极限代替每个分数值所得到的方法

斐波那契法小结

用 F_n 表示某个区间长度，通过计算 n 点函数值能将该区间压缩为一个单位长度，但任何大于 F_n 的区间长度都不能保证做到这一点

斐波那契数列：$F_n = F_{n-1} + F_{n-2}$

$$F_0 = 1$$
$$F_1 = 1$$
$$F_2 = 2$$
$$F_3 = 3$$
$$F_4 = 5$$
$$\vdots$$

0.618 法和斐波那契法的关系

$$\frac{F_n}{F_{n-1}} = 1 + \frac{F_{n-2}}{F_{n-1}} \qquad F_* = \lim_{n\to\infty} \frac{F_{n-1}}{F_n}$$

$$F_* = \frac{1}{2}\left(\sqrt{5}-1\right) \approx 0.618$$

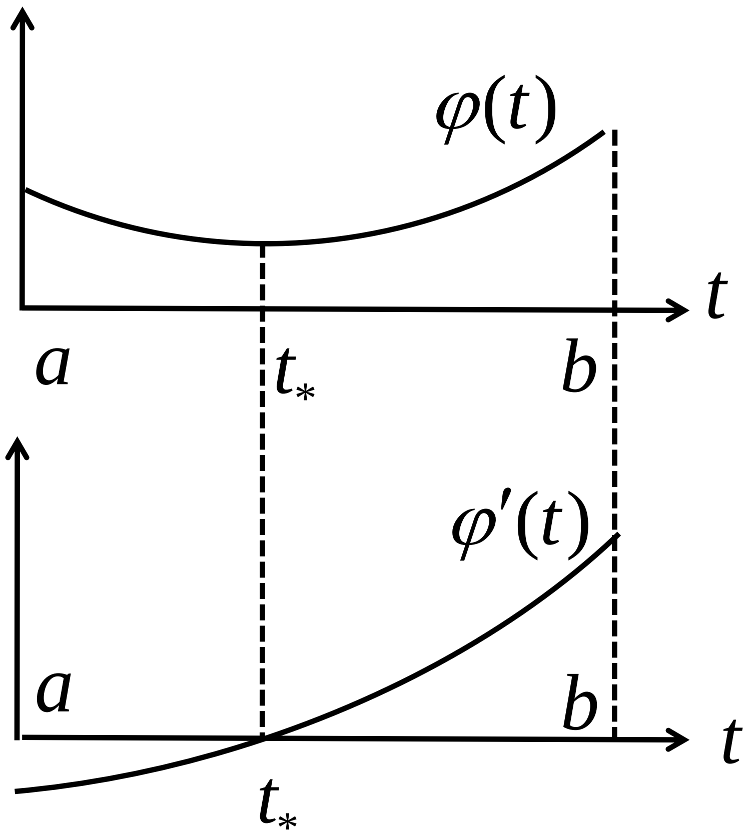
要点：利用导数的精确搜索法

利用导数的精确搜索算法（区间对分法）

单谷区间 $[a,b]$ 一定满足

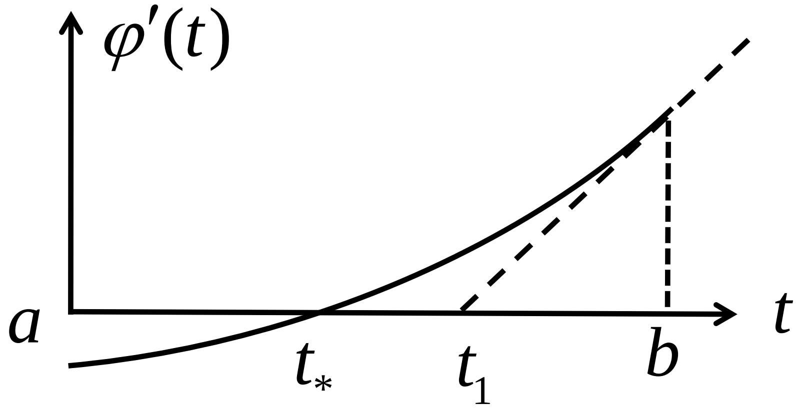
$\varphi'(a)<0, \varphi'(b)>0$

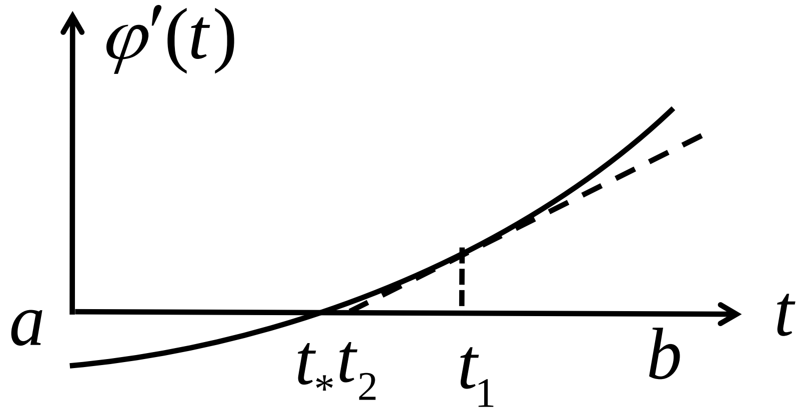
取 $t=0.5(a+b)$，若 $\varphi'(t)>0$，将区间压缩为 $[a,t]$，否则将区间压缩为 $[t,b]$



这种方法区间压缩比等于 0.5，比仅计算函数值的最好方法斐波那契法好，实际效果取决于导数计算量

利用二阶导数的精确搜索算法（Newton法）





如左图，在 b 点用切线近似 $\varphi'(t)$，求该切线的零点

切线方程：$g(t) = \varphi'(b) + \varphi''(b)(t - b)$

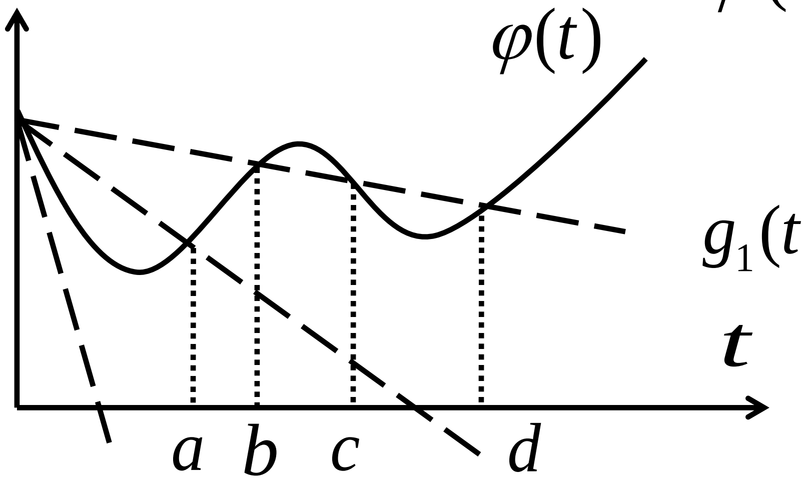
$$g(t_1) = 0 \quad \Rightarrow \quad t_1 = b - \frac{\varphi'(b)}{\varphi''(b)}$$

如上所示
单谷区间
一定收敛

再在 t_1 点重复上述过程 $\Rightarrow \quad t_2 = t_1 - \dfrac{\varphi'(t_1)}{\varphi''(t_1)}$

要点：非精确搜索

非精确搜索的Goldstein法原理



$\varphi'(0) < 0,\ 0 < m_1 < m_2 < 1$

$g_1(t) = \varphi(0) + m_1\varphi'(0)t$

$g_0(t) = \varphi(0) + \varphi'(0)t$　　$g_2(t) = \varphi(0) + m_2\varphi'(0)t$

基本想法：$\hat{t}$ 不能太小　　$\Rightarrow \quad \varphi(\hat{t}) \geq g_2(\hat{t})$

函数减量不能太小　$\Rightarrow \quad \varphi(\hat{t}) \leq g_1(\hat{t})$

对于上图，合格的 $\hat{t}$ 属于区间 $[a,b]$ 或 $[c,d]$

非精确搜索的Goldstein法

1）选择 $\alpha > 1$，$0 < m_1 < m_2 < 1$，$0 = a_0 < t_0 < b_0 = \infty$，$k = 0$

2）计算 $\varphi(t_k)$，若 $\varphi(t_k) \le g_1(t_k)$，到 3，否则，令

$a_{k+1} = a_k$， $b_{k+1} = t_k$ ，然后到 5 （$\varphi(b_k) > g_1(b_k)$）

3）若 $\varphi(t_k) \ge g_2(t_k)$，停止迭代，输出 t_k，否则，令

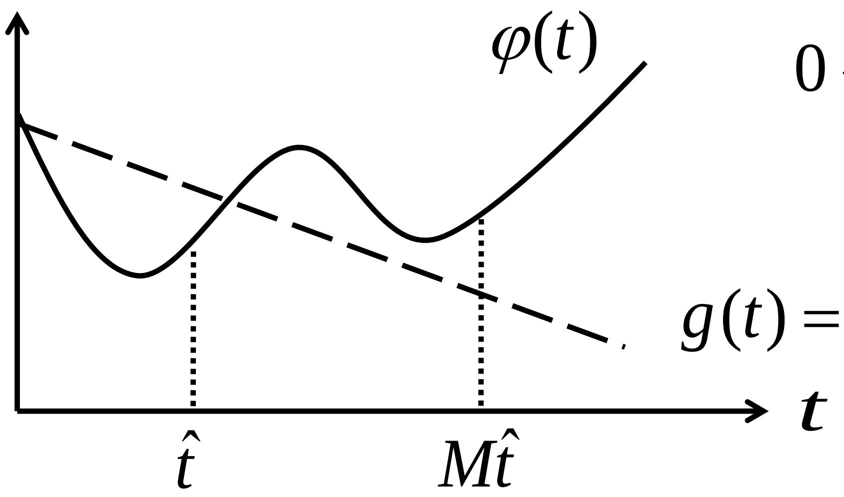
$a_{k+1} = t_k$ ， $b_{k+1} = b_k$ ，然后到 4 （$\varphi(a_k) < g_2(a_k)$）

4）若 $b_{k+1} < \infty$，到 5，否则，令 $t_{k+1} = \alpha t_k$ ，然后到 6

5）令 $t_{k+1} = 0.5(a_{k+1} + b_{k+1})$，然后到 6

6）用 $k+1$ 替换 k，然后回到 2

Armijo法



$0<m<1<M$

减量不太小 $\Rightarrow \varphi(\hat{t})\le g(\hat{t})$　　　$\hat{t}$ 不太小 $\Rightarrow \varphi(M\hat{t})>g(M\hat{t})$

一种实现方法：选择足够大的 $\bar{t}>0$ ，取

$$\hat{t}=\max\left\{\beta(s)=\frac{\bar{t}}{2^{s-1}},\ s=1,2,\ldots \,\middle|\, \text{s.t. } \varphi\big(\beta(s)\big)\le\varphi\big(0\big)+\frac{1}{2}\beta(s)\varphi'\big(0\big)\right\}$$

这样的 $\hat{t}$ 对 $m=\dfrac{1}{2}$ 和 $M=2$ 满足上述条件