最优化方法

东南大学

计算机&人工智能学院

宋沫飞

songmf@seu.edu.cn

无约束优化

- 术语和假设
- 梯度下降法
- 最速下降法
- **Newton**法
- 自和谐
- 实现

无约束优化

无约束优化

$$\text{minimize} \quad f(x)$$

无约束优化

$$\text{minimize} \quad f(x)$$

- 函数为凸函数，二次连续可微

无约束优化

$$\text{minimize} \quad f(x)$$

- 函数为凸函数，二次连续可微
- 假设最优值 $p^\star = \inf_x f(x)$存在且为有限值

无约束优化

$$\text{minimize} \quad f(x)$$

- 函数为凸函数，二次连续可微
- 假设最优值 $p^\star = \inf_x f(x)$存在且为有限值
- 无约束优化方法

无约束优化

$$\text{minimize} \quad f(x)$$

- 函数为凸函数，二次连续可微
- 假设最优值 $p^\star = \inf_x f(x)$存在且为有限值
- 无约束优化方法
  - 产生点序列 $x^{(k)} \in \mathbf{dom}\, f,\ k = 0, 1, \ldots$ 其中

无约束优化

$$\text{minimize} \quad f(x)$$

- 函数为凸函数，二次连续可微
- 假设最优值 $p^\star = \inf_x f(x)$存在且为有限值
- 无约束优化方法
  - 产生点序列 $x^{(k)} \in \mathbf{dom}\, f,\ k = 0, 1, \ldots$ 其中 $f(x^{(k)}) \to p^\star$

无约束优化

$$\text{minimize} \quad f(x)$$

- 函数为凸函数，二次连续可微
- 假设最优值 $p^\star = \inf_x f(x)$存在且为有限值
- 无约束优化方法
  - 产生点序列 $x^{(k)} \in \mathbf{dom}\, f,\ k = 0, 1, \ldots$ 其中
    $$f(x^{(k)}) \to p^\star$$
- 通过迭代算法求解满足最优化条件的方程

无约束优化

$$\text{minimize} \quad f(x)$$

- 函数为凸函数，二次连续可微
- 假设最优值 $p^\star = \inf_x f(x)$存在且为有限值
- 无约束优化方法
  - 产生点序列 $x^{(k)} \in \mathbf{dom}\, f,\ k = 0, 1, \ldots$ 其中

$$f(x^{(k)}) \to p^\star$$

- 通过迭代算法求解满足最优化条件的方程

$$\nabla f(x^\star) = 0$$

初始点和下水平集

初始点和下水平集

- 本章算法需要一个初始点 $x^{(0)}$ 满足

初始点和下水平集

- 本章算法需要一个初始点 $x^{(0)}$ 满足

$$x^{(0)} \in \mathbf{dom}\, f$$

初始点和下水平集

- 本章算法需要一个初始点 $x^{(0)}$ 满足

$$x^{(0)} \in \mathbf{dom}\, f$$

- 下水平集 $S = \{x \mid f(x) \leq f(x^{(0)})\}$ 为闭集

初始点和下水平集

- 本章算法需要一个初始点 $x^{(0)}$ 满足

$$x^{(0)} \in \mathbf{dom}\, f$$

- 下水平集 $S = \{x \mid f(x) \leq f(x^{(0)})\}$ 为闭集
- 条件**2**较难验证，除非所有的下水平集是闭集：

初始点和下水平集

- 本章算法需要一个初始点 $x^{(0)}$ 满足

$$x^{(0)} \in \mathbf{dom}\, f$$

- 下水平集 $S = \{x \mid f(x) \leq f(x^{(0)})\}$ 为闭集
- 条件**2**较难验证，除非所有的下水平集是闭集：
  - 等价于 $\mathbf{epi}\, f$ 是闭的

初始点和下水平集

- 本章算法需要一个初始点 $x^{(0)}$ 满足

$$x^{(0)} \in \mathbf{dom}\, f$$

- 下水平集 $S = \{x \mid f(x) \leq f(x^{(0)})\}$ 为闭集
- 条件**2**较难验证，除非所有的下水平集是闭集：
  - 等价于 $\mathbf{epi}\, f$ 是闭的
  - 若 $\mathbf{dom}\, f = \mathbf{R}^n$，则为真

初始点和下水平集

- 本章算法需要一个初始点 $x^{(0)}$ 满足

  $$x^{(0)} \in \mathbf{dom}\, f$$

- 下水平集 $S = \{x \mid f(x) \leq f(x^{(0)})\}$ 为闭集
- 条件**2**较难验证，除非所有的下水平集是闭集：
  - 等价于 $\mathbf{epi}\, f$ 是闭的
  - 若 $\mathbf{dom}\, f = \mathbf{R}^n$，则为真
  - 若 $f(x) \to \infty$ as $x \to \mathbf{bd}\,\mathbf{dom}\, f$，则为真

初始点和下水平集

- 本章算法需要一个初始点 $x^{(0)}$ 满足
$$x^{(0)} \in \mathbf{dom}\, f$$
- 下水平集 $S = \{x \mid f(x) \le f(x^{(0)})\}$ 为闭集
- 条件**2**较难验证，除非所有的下水平集是闭集：
  - 等价于 $\mathbf{epi}\, f$ 是闭的
  - 若 $\mathbf{dom}\, f = \mathbf{R}^n$，则为真
  - 若 $f(x) \to \infty$ as $x \to \mathbf{bd}\,\mathbf{dom}\, f$，则为真
- 不同带有闭的下水平集的可微函数

初始点和下水平集

- 本章算法需要一个初始点 $x^{(0)}$ 满足
$$x^{(0)} \in \mathbf{dom}\, f$$
- 下水平集 $S = \{x \mid f(x) \le f(x^{(0)})\}$ 为闭集
- 条件**2**较难验证，除非所有的下水平集是闭集：
  - 等价于 $\mathbf{epi}\, f$ 是闭的
  - 若 $\mathbf{dom}\, f = \mathbf{R}^n$，则为真
  - 若 $f(x) \to \infty$ as $x \to \mathbf{bd}\,\mathbf{dom}\, f$，则为真
- 不同带有闭的下水平集的可微函数
$$f(x) = \log(\sum_{i=1}^{m} \exp(a_i^T x + b_i)), \qquad f(x) = -\sum_{i=1}^{m} \log(b_i - a_i^T x)$$

强凸性及其含义

强凸性及其含义

- 目标函数在S上强凸的，指存在 $m > 0$，满足

强凸性及其含义

- 目标函数在S上强凸的，指存在 $m > 0$，满足

$$\nabla^2 f(x) \succeq mI \qquad \text{for all } x \in S$$

强凸性及其含义

- 目标函数在S上强凸的，指存在 $m > 0$，满足

$$\nabla^2 f(x) \succeq mI \qquad \text{for all } x \in S$$

- 对 $x, y \in S$,

强凸性及其含义

- 目标函数在S上强凸的，指存在 $m > 0$，满足

$$\nabla^2 f(x) \succeq mI \qquad \text{for all } x \in S$$

- 对 $x, y \in S$,

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

强凸性及其含义

- 目标函数在S上强凸的，指存在 $m > 0$，满足

$$\nabla^2 f(x) \succeq mI \qquad \text{for all } x \in S$$

- 对 $x, y \in S$,

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

- $p^\star > -\infty$，且对 $x \in S$,

强凸性及其含义

- 目标函数在S上强凸的，指存在 $m > 0$，满足

$$\nabla^2 f(x) \succeq mI \qquad \text{for all } x \in S$$

- 对 $x, y \in S$,

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

- $p^\star > -\infty$，且对 $x \in S$,

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

强凸性及其含义

- 目标函数在S上强凸的，指存在 $m > 0$，满足

$$\nabla^2 f(x) \succeq mI \qquad \text{for all } x \in S$$

- 对 $x, y \in S$,

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

- $p^\star > -\infty$，且对 $x \in S$,

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

- 作为停止条件十分有用（若m已知）

下降方法

下降方法

$$x^{(k+1)} = x^{(k)} + t^{(k)}\Delta x^{(k)} \quad \text{with } f(x^{(k+1)}) < f(x^{(k)})$$

下降方法

$$x^{(k+1)} = x^{(k)} + t^{(k)}\Delta x^{(k)} \quad \text{with } f(x^{(k+1)}) < f(x^{(k)})$$

- 其他表示方法：$x^+ = x + t\Delta x,\ x := x + t\Delta x$

下降方法

$$x^{(k+1)} = x^{(k)} + t^{(k)}\Delta x^{(k)} \quad \text{with } f(x^{(k+1)}) < f(x^{(k)})$$

- 其他表示方法：$x^+ = x + t\Delta x,\ x := x + t\Delta x$
- Δx为步径或搜索方向；t为步长

下降方法

$$x^{(k+1)} = x^{(k)} + t^{(k)}\Delta x^{(k)} \quad \text{with } f(x^{(k+1)}) < f(x^{(k)})$$

- 其他表示方法：$x^+ = x + t\Delta x,\ x := x + t\Delta x$
- Δx为步径或搜索方向；t为步长
- 根据凸性，$f(x^+) < f(x)$则 $\nabla f(x)^T \Delta x < 0$ ， Δx 为下降方向

下降方法

$$x^{(k+1)} = x^{(k)} + t^{(k)}\Delta x^{(k)} \quad \text{with } f(x^{(k+1)}) < f(x^{(k)})$$

- 其他表示方法：$x^+ = x + t\Delta x,\ x := x + t\Delta x$
- Δx为步径或搜索方向；t为步长
- 根据凸性，$f(x^+) < f(x)$则 $\nabla f(x)^T \Delta x < 0$ ， Δx 为下降方向

算法 9.1　通用下降方法。

给定 初始点 $x \in \mathbf{dom}\, f$。

重复进行

1. **确定下降方向** Δx。
2. **直线搜索。** 选择步长 $t > 0$。
3. **修改。** $x := x + t\Delta x$。

直到 满足停止准则。

直线搜索

直线搜索

- 精确搜索：$t = \text{argmin}_{t>0} f(x + t\Delta x)$

直线搜索

- 精确搜索：$t = \operatorname{argmin}_{t>0} f(x + t\Delta x)$
- 回溯直线搜索（参数为 $\alpha \in (0, 1/2)$, $\beta \in (0, 1)$）

直线搜索

- 精确搜索：$t = \operatorname{argmin}_{t>0} f(x + t\Delta x)$
- 回溯直线搜索（参数为 $\alpha \in (0, 1/2)$, $\beta \in (0, 1)$）
- 从$t = 1$开始，反复执行 $t := \beta t$ 直到

直线搜索

- 精确搜索：$t = \operatorname{argmin}_{t>0} f(x + t\Delta x)$
- 回溯直线搜索（参数为 $\alpha \in (0, 1/2)$, $\beta \in (0, 1)$）
- 从$t = 1$开始，反复执行 $t := \beta t$ 直到

$$f(x + t\Delta x) < f(x) + \alpha t \nabla f(x)^T \Delta x$$

直线搜索

- 精确搜索：$t = \operatorname{argmin}_{t>0} f(x + t\Delta x)$
- 回溯直线搜索（参数为 $\alpha \in (0, 1/2)$, $\beta \in (0, 1)$）
- 从$t = 1$开始，反复执行 $t := \beta t$ 直到

$$f(x + t\Delta x) < f(x) + \alpha t \nabla f(x)^T \Delta x$$

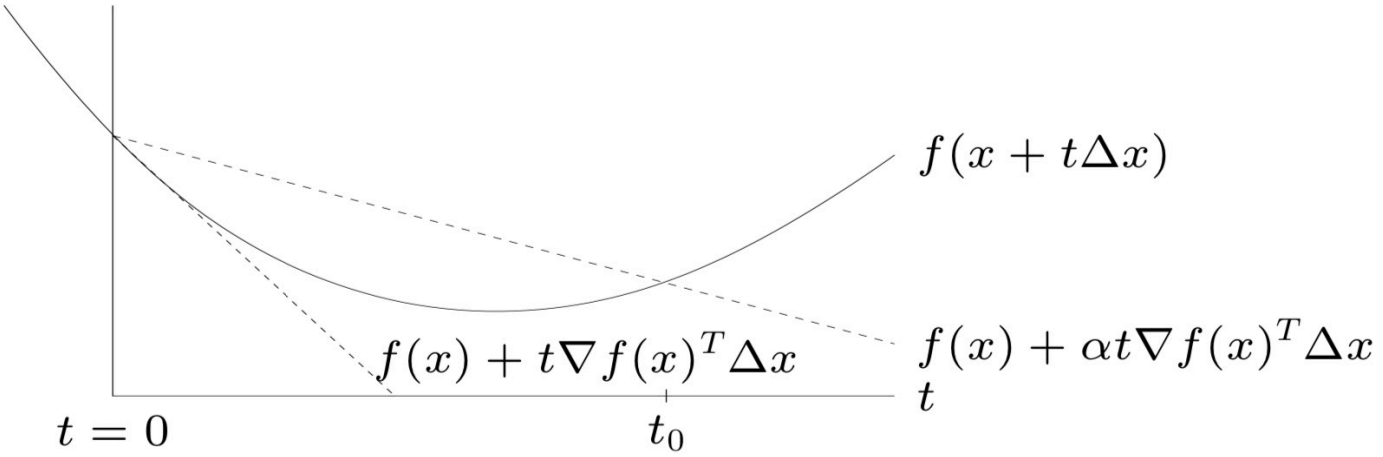
- 图例说明：回溯到 $t \leq t_0$

直线搜索

- 精确搜索：$t = \operatorname{argmin}_{t>0} f(x + t\Delta x)$
- 回溯直线搜索（参数为 $\alpha \in (0, 1/2)$, $\beta \in (0, 1)$）
- 从$t = 1$开始，反复执行 $t := \beta t$ 直到

$$f(x + t\Delta x) < f(x) + \alpha t \nabla f(x)^T \Delta x$$

- 图例说明：回溯到 $t \leq t_0$



梯度下降法

梯度下降法

- 梯度下降法，其中 $\Delta x = -\nabla f(x)$

梯度下降法

- 梯度下降法，其中 $\Delta x = -\nabla f(x)$

算法 9.3　梯度下降方法。

给定 初始点 $x \in \mathbf{dom}\, f$。

重复进行

1. $\Delta x := -\nabla f(x)$。
2. **直线搜索。** 通过精确或回溯直线搜索方法确定步长 t。
3. **修改。** $x := x + t\Delta x$。

直到 满足停止准则。

梯度下降法

- 梯度下降法，其中 $\Delta x = -\nabla f(x)$

算法 9.3　梯度下降方法。

给定 初始点 $x \in \mathbf{dom}\, f$。

重复进行

1. $\Delta x := -\nabla f(x)$。
2. **直线搜索。** 通过精确或回溯直线搜索方法确定步长 t。
3. **修改。** $x := x + t\Delta x$。

直到 满足停止准则。

- 停止条件：通常形如 $\|\nabla f(x)\|_2 \leq \epsilon$

梯度下降法

- 梯度下降法，其中 $\Delta x = -\nabla f(x)$

算法 9.3　梯度下降方法。

给定 初始点 $x \in \mathbf{dom}\, f$。

重复进行

1. $\Delta x := -\nabla f(x)$。
2. **直线搜索。** 通过精确或回溯直线搜索方法确定步长 t。
3. **修改。** $x := x + t\Delta x$。

直到 满足停止准则。

- 停止条件：通常形如 $\|\nabla f(x)\|_2 \leq \epsilon$
- 收敛结果：对强凸函数$f, f(x^{(k)}) - p^\star \leq c^k(f(x^{(0)}) - p^\star)$

梯度下降法

- 梯度下降法，其中 $\Delta x = -\nabla f(x)$

算法 9.3　梯度下降方法。

给定 初始点 $x \in \mathbf{dom}\, f$。

重复进行

1. $\Delta x := -\nabla f(x)$。
2. **直线搜索。** 通过精确或回溯直线搜索方法确定步长 t。
3. **修改。** $x := x + t\Delta x$。

直到 满足停止准则。

- 停止条件：通常形如 $\|\nabla f(x)\|_2 \leq \epsilon$
- 收敛结果：对强凸函数$f, f(x^{(k)}) - p^\star \leq c^k(f(x^{(0)}) - p^\star)$
- $c \in (0,1)$ 取决于 m, $x^{(0)}$, 直线搜索类型

梯度下降法

- 梯度下降法，其中 $\Delta x = -\nabla f(x)$

算法 9.3　梯度下降方法。

给定 初始点 $x \in \mathbf{dom}\, f$。

重复进行

1. $\Delta x := -\nabla f(x)$。
2. **直线搜索。** 通过精确或回溯直线搜索方法确定步长 t。
3. **修改。** $x := x + t\Delta x$。

直到 满足停止准则。

- 停止条件：通常形如 $\|\nabla f(x)\|_2 \leq \epsilon$
- 收敛结果：对强凸函数f, $f(x^{(k)}) - p^\star \leq c^k(f(x^{(0)}) - p^\star)$
- $c \in (0,1)$ 取决于 m, $x^{(0)}$, 直线搜索类型
- 非常简单，但通常较慢；实际中很少使用

R^2空间中的二次问题

R^2空间中的二次问题

$$f(x) = (1/2)(x_1^2 + \gamma x_2^2) \qquad (\gamma > 0)$$

R^2空间中的二次问题

$$f(x) = (1/2)(x_1^2 + \gamma x_2^2) \qquad (\gamma > 0)$$

- 进行精确直线搜索，起始点为$x^{(0)} = (\gamma, 1)$:

R^2空间中的二次问题

$$f(x) = (1/2)(x_1^2 + \gamma x_2^2) \qquad (\gamma > 0)$$

- 进行精确直线搜索，起始点为$x^{(0)} = (\gamma, 1)$:

$$x_1^{(k)} = \gamma \left(\frac{\gamma - 1}{\gamma + 1}\right)^k, \qquad x_2^{(k)} = \left(-\frac{\gamma - 1}{\gamma + 1}\right)^k$$

R^2空间中的二次问题

$$f(x) = (1/2)(x_1^2 + \gamma x_2^2) \qquad (\gamma > 0)$$

- 进行精确直线搜索，起始点为$x^{(0)} = (\gamma, 1)$:

$$x_1^{(k)} = \gamma \left(\frac{\gamma - 1}{\gamma + 1}\right)^k, \qquad x_2^{(k)} = \left(-\frac{\gamma - 1}{\gamma + 1}\right)^k$$

- 若 $\gamma \gg 1$ or $\gamma \ll 1$ ，速度非常慢

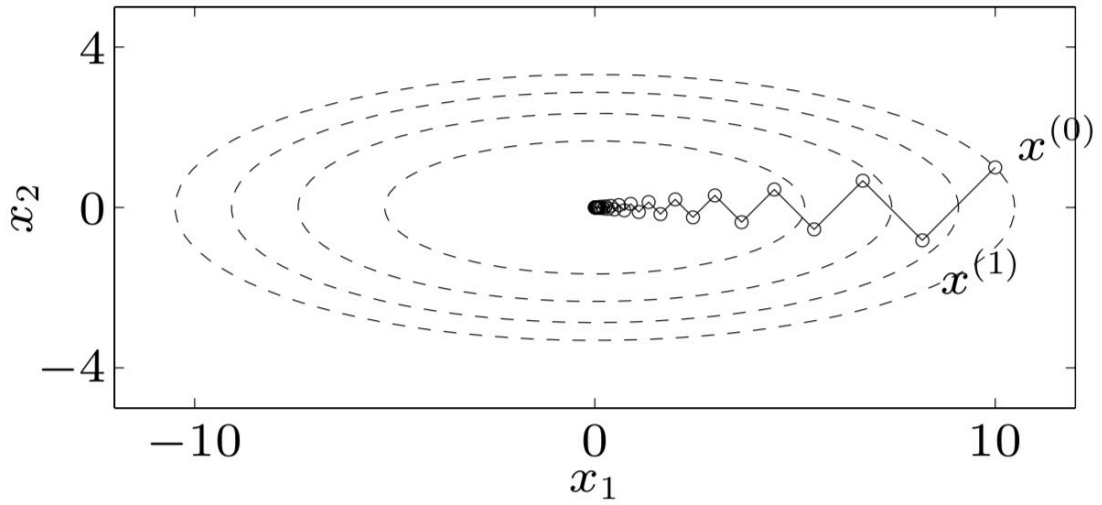
R^2空间中的二次问题

$$f(x) = (1/2)(x_1^2 + \gamma x_2^2) \qquad (\gamma > 0)$$

- 进行精确直线搜索，起始点为$x^{(0)} = (\gamma, 1)$:

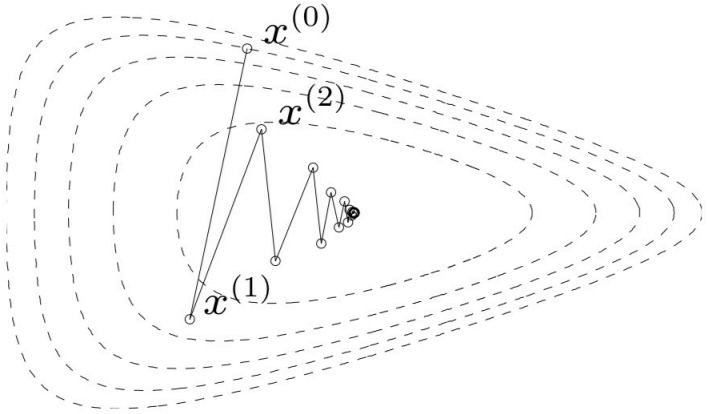
$$x_1^{(k)} = \gamma \left(\frac{\gamma - 1}{\gamma + 1}\right)^k, \qquad x_2^{(k)} = \left(-\frac{\gamma - 1}{\gamma + 1}\right)^k$$

- 若 $\gamma \gg 1$ or $\gamma \ll 1$ ，速度非常慢
- $\gamma = 10$

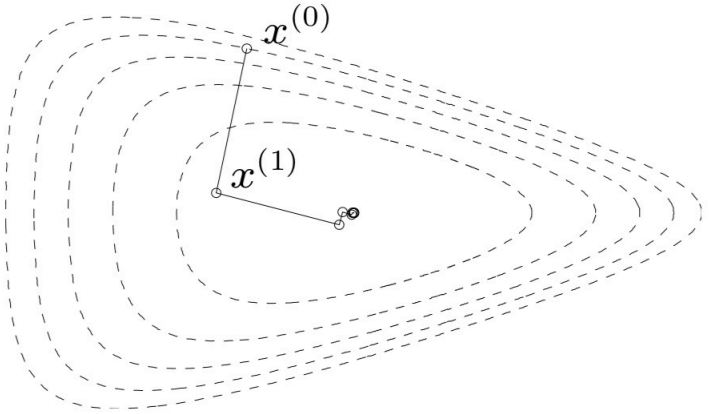


非二次型问题

$$f(x_1, x_2) = e^{x_1+3x_2-0.1} + e^{x_1-3x_2-0.1} + e^{-x_1-0.1}$$



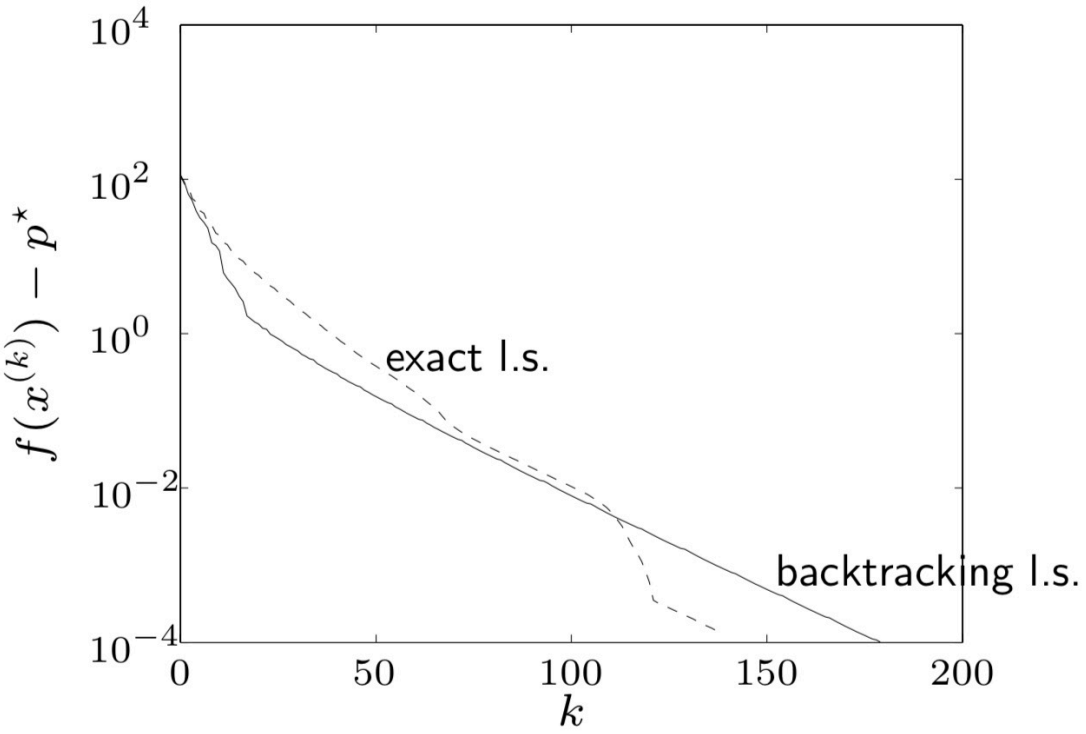
回溯直线搜索



精确直线搜索

R^{100}空间中的问题

$$f(x) = c^T x - \sum_{i=1}^{500} \log(b_i - a_i^T x)$$



线性收敛

最速下降法

最速下降法

- 规范化的最速下降方向（相对于范数$\|\cdot\|$）：

最速下降法

- 规范化的最速下降方向（相对于范数$\|\cdot\|$）：

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

最速下降法

- 规范化的最速下降方向（相对于范数$\|\cdot\|$）：

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

- 解释：对较小的v，有 $f(x+v) \approx f(x) + \nabla f(x)^T v$

最速下降法

- 规范化的最速下降方向（相对于范数$\|\cdot\|$）：

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

- 解释：对较小的v，有 $f(x+v) \approx f(x) + \nabla f(x)^T v$
- 搜索方向Δx_{nsd}是具有最大线性下降的单位范数的步径

最速下降法

- 规范化的最速下降方向（相对于范数$\|\cdot\|$）：

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

- 解释：对较小的v，有 $f(x+v) \approx f(x) + \nabla f(x)^T v$
- 搜索方向Δx_{nsd}是具有最大线性下降的单位范数的步径
- 非规范化的最速下降方向

最速下降法

- 规范化的最速下降方向（相对于范数$\|\cdot\|$）：

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

- 解释：对较小的v，有 $f(x+v) \approx f(x) + \nabla f(x)^T v$
- 搜索方向Δx_{nsd}是具有最大线性下降的单位范数的步径
- 非规范化的最速下降方向

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

最速下降法

- 规范化的最速下降方向（相对于范数$\|\cdot\|$）：

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

- 解释：对较小的v，有 $f(x+v) \approx f(x) + \nabla f(x)^T v$
- 搜索方向Δx_{nsd}是具有最大线性下降的单位范数的步径
- 非规范化的最速下降方向

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

  - 满足 $\nabla f(x)^T \Delta x_{\text{sd}} = -\|\nabla f(x)\|_*^2$

最速下降法

- 规范化的最速下降方向（相对于范数$\|\cdot\|$）：

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

- 解释：对较小的v，有 $f(x+v) \approx f(x) + \nabla f(x)^T v$
- 搜索方向Δx_{nsd}是具有最大线性下降的单位范数的步径
- 非规范化的最速下降方向

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

  - 满足 $\nabla f(x)^T \Delta x_{\text{sd}} = -\|\nabla f(x)\|_*^2$
- 最速下降法

最速下降法

- 规范化的最速下降方向（相对于范数$\|\cdot\|$）：

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

- 解释：对较小的v，有 $f(x+v) \approx f(x) + \nabla f(x)^T v$
- 搜索方向Δx_{nsd}是具有最大线性下降的单位范数的步径
- 非规范化的最速下降方向

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

  - 满足 $\nabla f(x)^T \Delta x_{\text{sd}} = -\|\nabla f(x)\|_*^2$
- 最速下降法
  - 通用下降法 $\Delta x = \Delta x_{\text{sd}}$

最速下降法

- 规范化的最速下降方向（相对于范数$\|\cdot\|$）：

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

- 解释：对较小的v，有 $f(x+v) \approx f(x) + \nabla f(x)^T v$
- 搜索方向Δx_{nsd}是具有最大线性下降的单位范数的步径
- 非规范化的最速下降方向

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

  - 满足 $\nabla f(x)^T \Delta x_{\text{sd}} = -\|\nabla f(x)\|_*^2$
- 最速下降法
  - 通用下降法 $\Delta x = \Delta x_{\text{sd}}$
  - 收敛特性和梯度下降一致

例

例

- 欧式范数：$\Delta x_{\mathrm{sd}} = -\nabla f(x)$

例

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

例

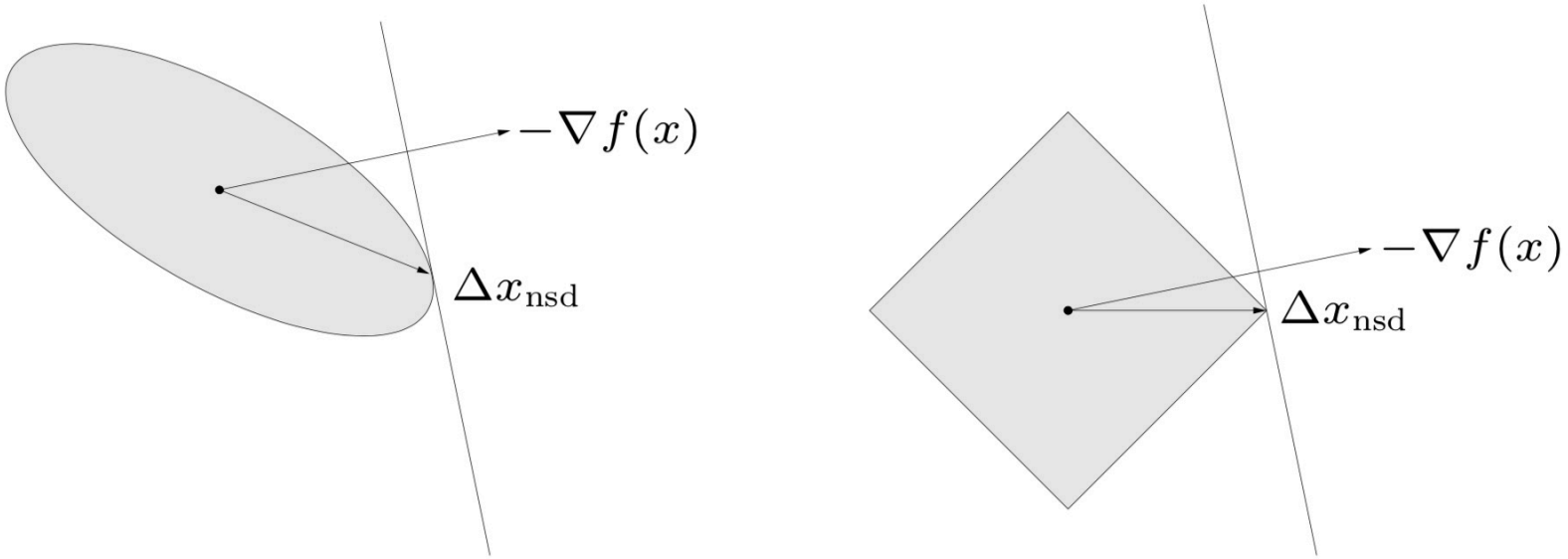
- 欧式范数：$\Delta x_{\mathrm{sd}} = -\nabla f(x)$
- 二次范数 $\|x\|_P = (x^T P x)^{1/2}$ $(P \in \mathbf{S}^n_{++})$: $\Delta x_{\mathrm{sd}} = -P^{-1}\nabla f(x)$
- ℓ_1- 范数 $\Delta x_{\mathrm{sd}} = -(\partial f(x)/\partial x_i)e_i$, where $|\partial f(x)/\partial x_i| = \|\nabla f(x)\|_\infty$

例

- 欧式范数：$\Delta x_{\mathrm{sd}} = -\nabla f(x)$
- 二次范数 $\|x\|_P = (x^T P x)^{1/2}$ $(P \in \mathbf{S}^n_{++})$: $\Delta x_{\mathrm{sd}} = -P^{-1}\nabla f(x)$
- ℓ_1- 范数 $\Delta x_{\mathrm{sd}} = -(\partial f(x)/\partial x_i)e_i$, where $|\partial f(x)/\partial x_i| = \|\nabla f(x)\|_\infty$
- 关于二次型和ℓ_1-范数的单位球和规范化最速下降方向

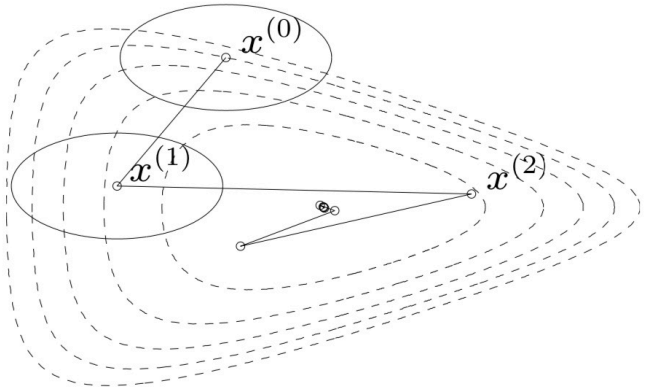
例

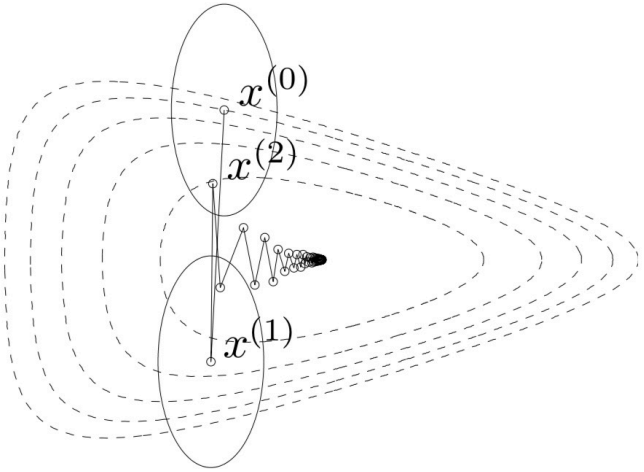
- 欧式范数：$\Delta x_{\mathrm{sd}} = -\nabla f(x)$
- 二次范数 $\|x\|_P = (x^T P x)^{1/2}$ $(P \in \mathbf{S}^n_{++})$: $\Delta x_{\mathrm{sd}} = -P^{-1}\nabla f(x)$
- ℓ_1- 范数 $\Delta x_{\mathrm{sd}} = -(\partial f(x)/\partial x_i)e_i$, where $|\partial f(x)/\partial x_i| = \|\nabla f(x)\|_\infty$
- 关于二次型和ℓ_1-范数的单位球和规范化最速下降方向



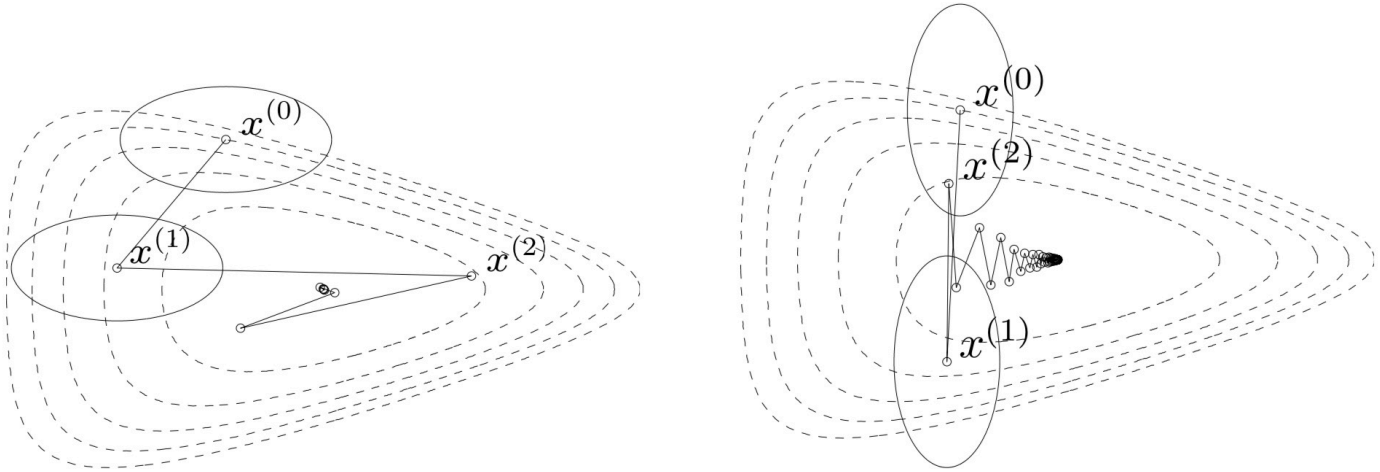
最速下降的范数选择

最速下降的范数选择



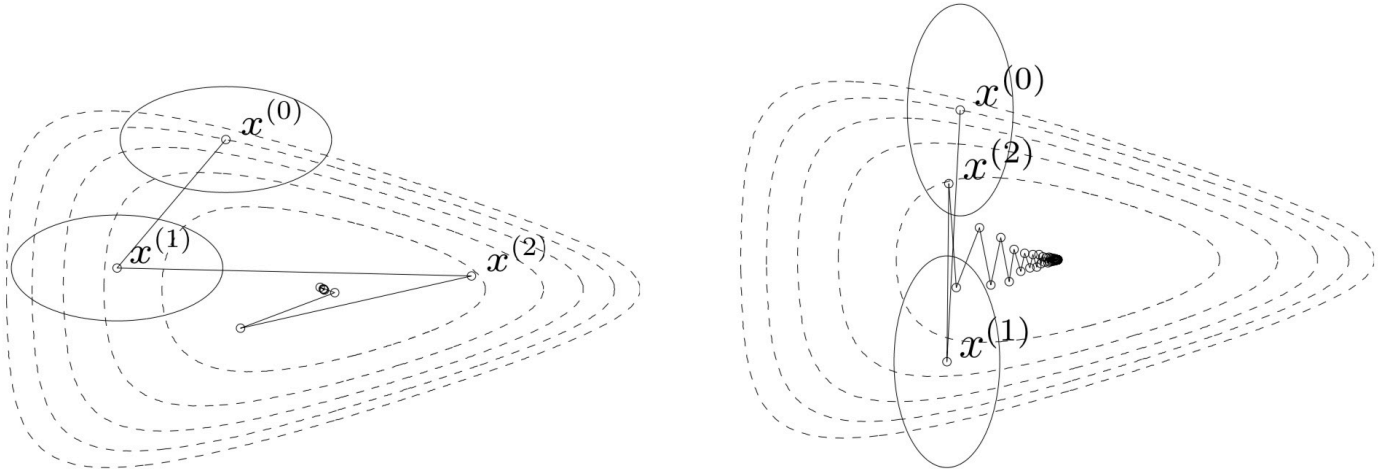


最速下降的范数选择



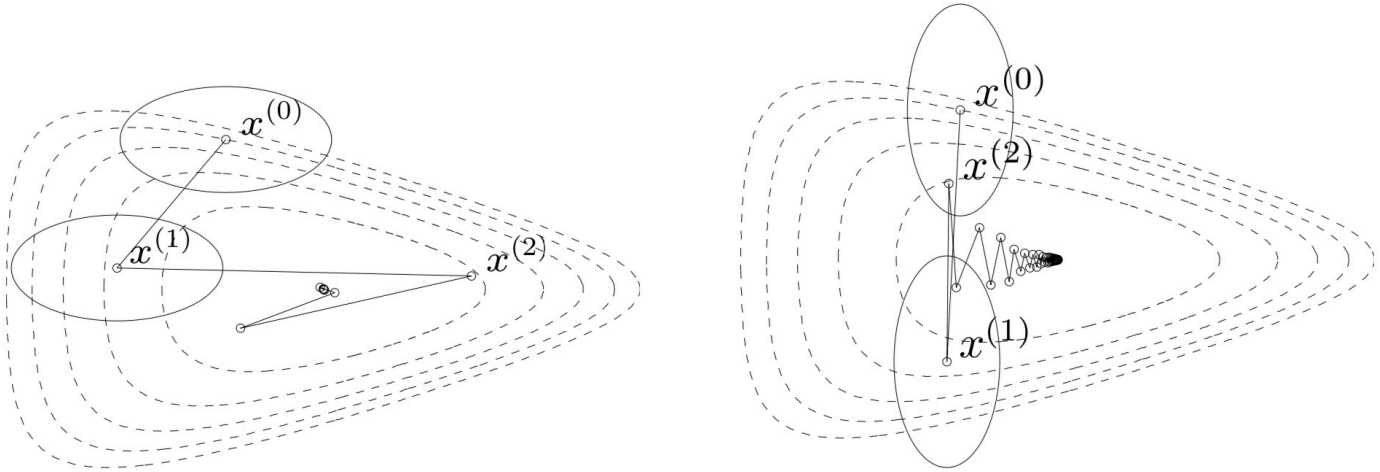
- 分别使用两个二次范数和回溯直线搜索的最速下降法

最速下降的范数选择



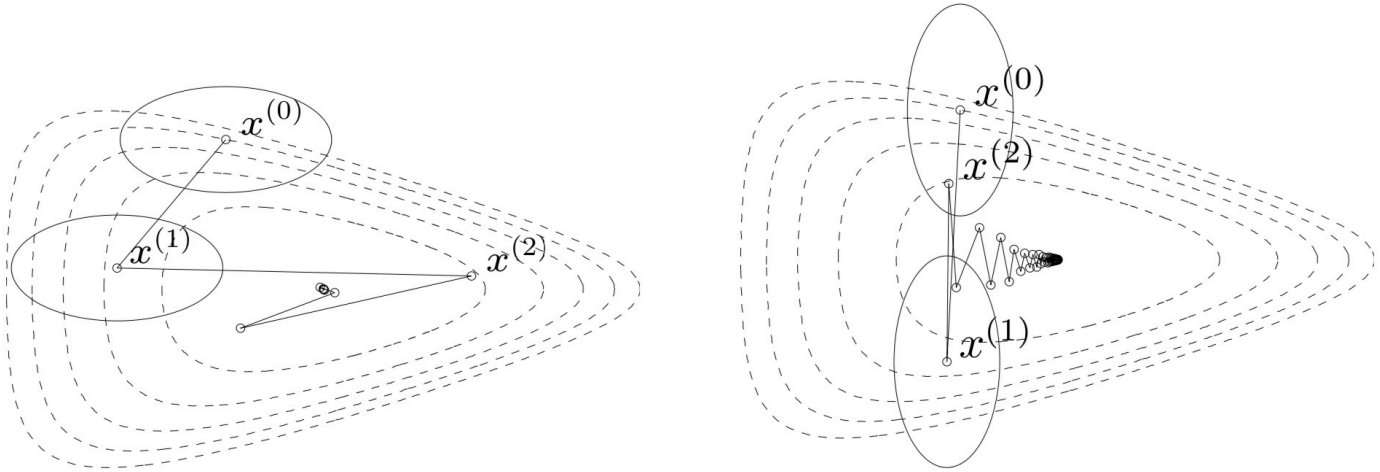
- 分别使用两个二次范数和回溯直线搜索的最速下降法
- 椭圆 $\{x \mid \|x - x^{(k)}\|_P = 1\}$

最速下降的范数选择



- 分别使用两个二次范数和回溯直线搜索的最速下降法
- 椭圆 $\{x \mid \|x - x^{(k)}\|_P = 1\}$
- 二次型最速下降的等价解释：修改优化变量 $\bar{x} = P^{1/2}x$ 的梯度下降

最速下降的范数选择



- 分别使用两个二次范数和回溯直线搜索的最速下降法
- 椭圆 $\{x \mid \|x - x^{(k)}\|_P = 1\}$
- 二次型最速下降的等价解释：修改优化变量 $\bar{x} = P^{1/2}x$ 的梯度下降
- 范数选择对收敛速度有很大影响

Newton步径

Newton步径

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

Newton步径

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

- $x + \Delta x_{\mathrm{nt}}$最小化二阶近似

Newton步径

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

- $x + \Delta x_{\mathrm{nt}}$最小化二阶近似

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

Newton步径

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

- $x + \Delta x_{\mathrm{nt}}$最小化二阶近似

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

- $x + \Delta x_{\mathrm{nt}}$ 求解线性最优条件

Newton步径

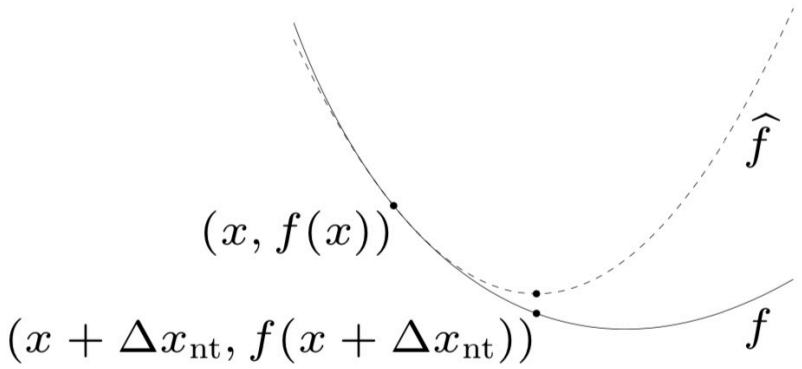
$$\Delta x_{\mathrm{nt}} = -\nabla^2 f(x)^{-1} \nabla f(x)$$

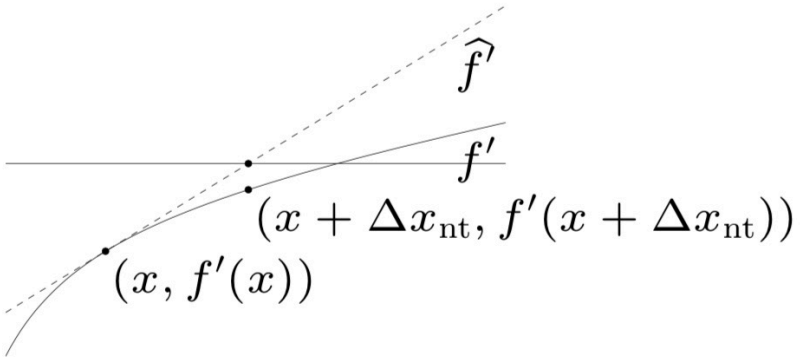
- $x + \Delta x_{\mathrm{nt}}$ 最小化二阶近似

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

- $x + \Delta x_{\mathrm{nt}}$ 求解线性最优条件

$$\nabla f(x+v) \approx \nabla \widehat{f}(x+v) = \nabla f(x) + \nabla^2 f(x) v = 0$$





Newton步径

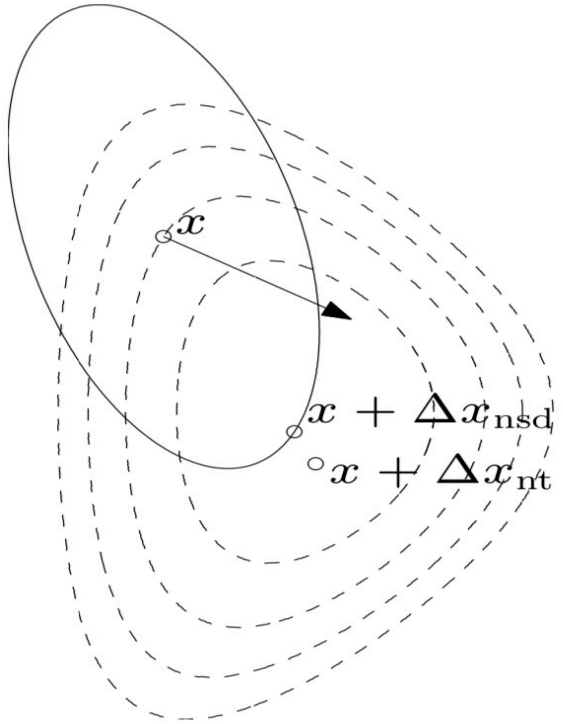
Newton步径

- Δx_{nt} 为局部**Hessian**范数的最速梯度下降方向

Newton步径

- Δx_{nt} 为局部**Hessian**范数的最速梯度下降方向

$$\|u\|_{\nabla^2 f(x)} = \left(u^T \nabla^2 f(x) u\right)^{1/2}$$

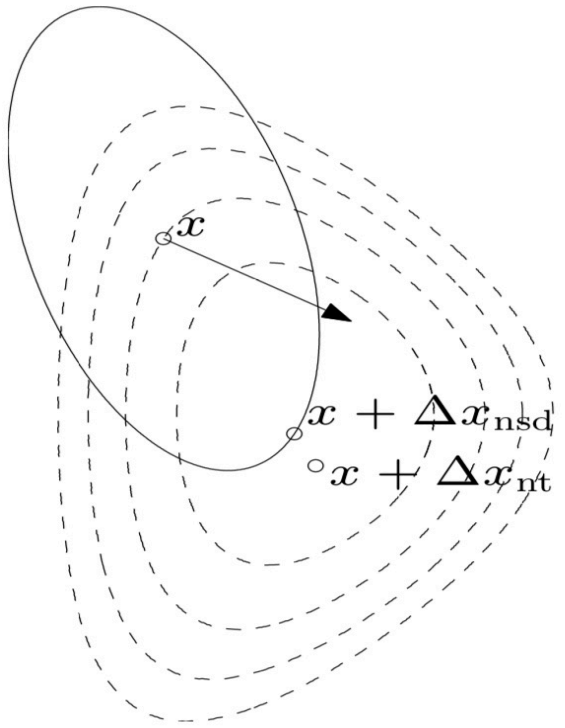


Newton步径

- Δx_{nt} 为局部**Hessian**范数的最速梯度下降方向

$$\|u\|_{\nabla^2 f(x)} = \left(u^T \nabla^2 f(x) u\right)^{1/2}$$

- 虚线为函数的等高线

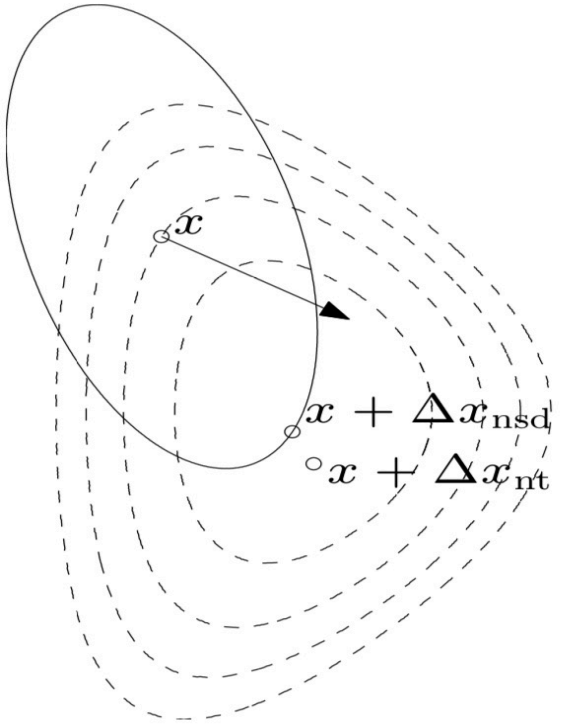


Newton步径

- Δx_{nt} 为局部**Hessian**范数的最速梯度下降方向

$$\|u\|_{\nabla^2 f(x)} = \left(u^T \nabla^2 f(x) u\right)^{1/2}$$

- 虚线为函数的等高线
- 椭圆为 $\{x + v \mid v^T \nabla^2 f(x) v = 1\}$

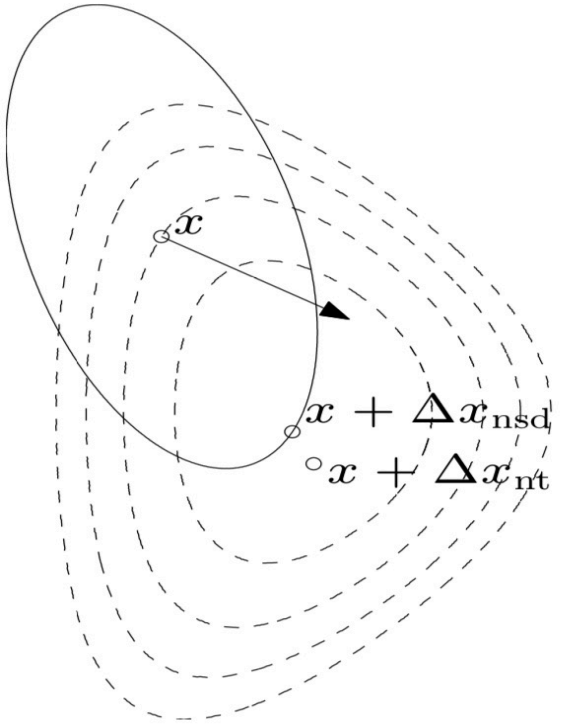


Newton步径

- Δx_{nt} 为局部**Hessian**范数的最速梯度下降方向

$$\|u\|_{\nabla^2 f(x)} = \left(u^T \nabla^2 f(x) u\right)^{1/2}$$

- 虚线为函数的等高线
- 椭圆为 $\{x + v \mid v^T \nabla^2 f(x) v = 1\}$
- 箭头表示 $-\nabla f(x)$



Newton减量

Newton减量

$$\lambda(x) = \left(\nabla f(x)^T \nabla^2 f(x)^{-1} \nabla f(x)\right)^{1/2}$$

Newton减量

$$\lambda(x) = \left(\nabla f(x)^T \nabla^2 f(x)^{-1} \nabla f(x)\right)^{1/2}$$

- 对x和x^*接近程度的度量

Newton减量

$$\lambda(x) = \left(\nabla f(x)^T \nabla^2 f(x)^{-1} \nabla f(x)\right)^{1/2}$$

- 对x和x^*接近程度的度量
- 性质

Newton减量

$$\lambda(x) = \left(\nabla f(x)^T \nabla^2 f(x)^{-1} \nabla f(x)\right)^{1/2}$$

- 对x和x^*接近程度的度量
- 性质
  - 给出了 $f(x) - p^\star$ 的一个估计，使用二次近似：

Newton减量

$$\lambda(x) = \left(\nabla f(x)^T \nabla^2 f(x)^{-1} \nabla f(x)\right)^{1/2}$$

- 对x和x^*接近程度的度量
- 性质
  - 给出了 $f(x) - p^\star$ 的一个估计，使用二次近似：

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

Newton减量

$$\lambda(x) = \left(\nabla f(x)^T \nabla^2 f(x)^{-1} \nabla f(x)\right)^{1/2}$$

- 对*x*和*x**接近程度的度量
- 性质
  - 给出了 $f(x) - p^\star$ 的一个估计，使用二次近似：

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

- 为**Newton**步径的二次范数，具有二次**Hessian**形式

Newton减量

$$\lambda(x) = \left(\nabla f(x)^T \nabla^2 f(x)^{-1} \nabla f(x)\right)^{1/2}$$

- 对x和x^*接近程度的度量
- 性质
  - 给出了 $f(x) - p^\star$ 的一个估计，使用二次近似：

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

- 为**Newton**步径的二次范数，具有二次**Hessian**形式

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

Newton减量

$$\lambda(x) = \left(\nabla f(x)^T \nabla^2 f(x)^{-1} \nabla f(x)\right)^{1/2}$$

- 对x和x^*接近程度的度量
- 性质
  - 给出了 $f(x) - p^\star$ 的一个估计，使用二次近似：

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

- 为**Newton**步径的二次范数，具有二次**Hessian**形式

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

  - 仿射不变性（和 $\|\nabla f(x)\|_2$ 不一样）

Newton方法

Newton方法

算法 9.5　Newton 方法。

给定 初始点 $x \in \mathbf{dom}\, f$，误差阈值 $\epsilon > 0$。

重复进行

1. **计算 Newton 步径和减量。**
   $$\Delta x_{\mathrm{nt}} := -\nabla^2 f(x)^{-1}\nabla f(x)\ ; \quad \lambda^2 := \nabla f(x)^T \nabla^2 f(x)^{-1}\nabla f(x)。$$
2. **停止准则。** 如果 $\lambda^2/2 \leqslant \epsilon$，**退出。**
3. **直线搜索。** 通过回溯直线搜索确定步长 t。
4. **改进。** $x := x + t\Delta x_{\mathrm{nt}}$。

Newton方法

算法 9.5　Newton 方法。

给定 初始点 $x \in \mathbf{dom}\, f$，误差阈值 $\epsilon > 0$。

重复进行

1. **计算 Newton 步径和减量。**
   $\Delta x_{\mathrm{nt}} := -\nabla^2 f(x)^{-1}\nabla f(x)$；　$\lambda^2 := \nabla f(x)^T \nabla^2 f(x)^{-1}\nabla f(x)$。
2. **停止准则。** 如果 $\lambda^2/2 \leqslant \epsilon$，**退出。**
3. **直线搜索。** 通过回溯直线搜索确定步长 t。
4. **改进。** $x := x + t\Delta x_{\mathrm{nt}}$。

- 仿射不变性，表示和坐标的线性变换无关：

Newton方法

算法 9.5　Newton 方法。

给定 初始点 $x \in \mathbf{dom}\, f$，误差阈值 $\epsilon > 0$。

重复进行

1. **计算 Newton 步径和减量。**
   $\Delta x_{\mathrm{nt}} := -\nabla^2 f(x)^{-1}\nabla f(x)$；　$\lambda^2 := \nabla f(x)^T \nabla^2 f(x)^{-1}\nabla f(x)$。
2. **停止准则。** 如果 $\lambda^2/2 \leqslant \epsilon$，**退出。**
3. **直线搜索。** 通过回溯直线搜索确定步长 t。
4. **改进。** $x := x + t\Delta x_{\mathrm{nt}}$。

- 仿射不变性，表示和坐标的线性变换无关：
- **Newton**迭代进行 $\tilde{f}(y) = f(Ty)$

Newton方法

算法 9.5　Newton 方法。

给定 初始点 $x \in \mathbf{dom}\, f$，误差阈值 $\epsilon > 0$。

重复进行

1. **计算 Newton 步径和减量。**
   $\Delta x_{\mathrm{nt}} := -\nabla^2 f(x)^{-1}\nabla f(x)$；　$\lambda^2 := \nabla f(x)^T \nabla^2 f(x)^{-1}\nabla f(x)$。
2. **停止准则。** 如果 $\lambda^2/2 \leqslant \epsilon$，**退出。**
3. **直线搜索。** 通过回溯直线搜索确定步长 t。
4. **改进。** $x := x + t\Delta x_{\mathrm{nt}}$。

- 仿射不变性，表示和坐标的线性变换无关：
- **Newton**迭代进行 $\tilde{f}(y) = f(Ty)$
  - 起始点为 $y^{(0)} = T^{-1}x^{(0)}$

Newton方法

算法 9.5 Newton 方法。

给定 初始点 $x \in \mathbf{dom}\, f$，误差阈值 $\epsilon > 0$。

重复进行

1. **计算 Newton 步径和减量。**
   $\Delta x_{\mathrm{nt}} := -\nabla^2 f(x)^{-1}\nabla f(x)$；　$\lambda^2 := \nabla f(x)^T \nabla^2 f(x)^{-1}\nabla f(x)$。
2. **停止准则。** 如果 $\lambda^2/2 \leqslant \epsilon$，**退出。**
3. **直线搜索。** 通过回溯直线搜索确定步长 t。
4. **改进。** $x := x + t\Delta x_{\mathrm{nt}}$。

- 仿射不变性，表示和坐标的线性变换无关：
- **Newton**迭代进行 $\tilde{f}(y) = f(Ty)$
  - 起始点为 $y^{(0)} = T^{-1}x^{(0)}$

$$y^{(k)} = T^{-1}x^{(k)}$$

经典收敛分析

经典收敛分析

- 假设f在S上具有常数为m的强凸性

经典收敛分析

- 假设f在S上具有常数为m的强凸性
  - $\nabla^2 f$在S上具有以L为常数的**Lipschitz**连续性：

经典收敛分析

- 假设f在S上具有常数为m的强凸性
  - $\nabla^2 f$在S上具有以L为常数的**Lipschitz**连续性：

$$\|\nabla^2 f(x) - \nabla^2 f(y)\|_2 \leq L\|x - y\|_2$$

经典收敛分析

- 假设f在S上具有常数为m的强凸性
  - $\nabla^2 f$在S上具有以L为常数的**Lipschitz**连续性：

$$\|\nabla^2 f(x) - \nabla^2 f(y)\|_2 \leq L\|x - y\|_2$$

- L测量了使用二次函数拟合函数f的近似程度

经典收敛分析

- 假设f在S上具有常数为m的强凸性
  - $\nabla^2 f$在S上具有以L为常数的**Lipschitz**连续性：

$$\|\nabla^2 f(x) - \nabla^2 f(y)\|_2 \leq L\|x - y\|_2$$

- L测量了使用二次函数拟合函数f的近似程度
- 框架：存在常数$\eta \in (0, m^2/L)$, $\gamma > 0$满足

经典收敛分析

- 假设f在S上具有常数为m的强凸性
  - $\nabla^2 f$在S上具有以L为常数的**Lipschitz**连续性：

$$\|\nabla^2 f(x) - \nabla^2 f(y)\|_2 \leq L\|x - y\|_2$$

- L测量了使用二次函数拟合函数f的近似程度
- 框架：存在常数$\eta \in (0, m^2/L)$, $\gamma > 0$满足

if $\|\nabla f(x)\|_2 \geq \eta$, then $f(x^{(k+1)}) - f(x^{(k)}) \leq -\gamma$

if $\|\nabla f(x)\|_2 < \eta$, then

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

阻尼Newton阶段

阻尼Newton阶段

$$(\|\nabla f(x)\|_2 \geq \eta)$$

阻尼Newton阶段

$(\|\nabla f(x)\|_2 \geq \eta)$

- 大多数迭代过程需要回溯步骤

阻尼Newton阶段

$$(\|\nabla f(x)\|_2 \geq \eta)$$

- 大多数迭代过程需要回溯步骤
- 目标函数减少量至少为γ

阻尼Newton阶段

$(\|\nabla f(x)\|_2 \geq \eta)$

- 大多数迭代过程需要回溯步骤
- 目标函数减少量至少为γ
- 若$p^\star > -\infty$，该阶段最多需要的迭代次数是

阻尼Newton阶段

$(\|\nabla f(x)\|_2 \geq \eta)$

- 大多数迭代过程需要回溯步骤
- 目标函数减少量至少为γ
- 若$p^\star > -\infty$，该阶段最多需要的迭代次数是

$$(f(x^{(0)}) - p^\star)/\gamma$$

阻尼Newton阶段

$(\|\nabla f(x)\|_2 \geq \eta)$

- 大多数迭代过程需要回溯步骤
- 目标函数减少量至少为γ
- 若$p^\star > -\infty$，该阶段最多需要的迭代次数是

$$(f(x^{(0)}) - p^\star)/\gamma$$

- 二次收敛阶段$(\|\nabla f(x)\|_2 < \eta)$

阻尼Newton阶段

$(\|\nabla f(x)\|_2 \geq \eta)$

- 大多数迭代过程需要回溯步骤
- 目标函数减少量至少为 γ
- 若 $p^\star > -\infty$，该阶段最多需要的迭代次数是

$$(f(x^{(0)}) - p^\star)/\gamma$$

- 二次收敛阶段 $(\|\nabla f(x)\|_2 < \eta)$
  - 所有迭代使用相同步长 $t = 1$

阻尼Newton阶段

$(\|\nabla f(x)\|_2 \geq \eta)$

- 大多数迭代过程需要回溯步骤
- 目标函数减少量至少为γ
- 若$p^\star > -\infty$，该阶段最多需要的迭代次数是

$$(f(x^{(0)}) - p^\star)/\gamma$$

- 二次收敛阶段$(\|\nabla f(x)\|_2 < \eta)$
  - 所有迭代使用相同步长$t = 1$
- $\|\nabla f(x)\|_2$二次收敛到**0**：若$\|\nabla f(x^{(k)})\|_2 < \eta$

阻尼Newton阶段

$(\|\nabla f(x)\|_2 \geq \eta)$

- 大多数迭代过程需要回溯步骤
- 目标函数减少量至少为 γ
- 若 $p^\star > -\infty$，该阶段最多需要的迭代次数是

$$(f(x^{(0)}) - p^\star)/\gamma$$

- 二次收敛阶段 $(\|\nabla f(x)\|_2 < \eta)$
  - 所有迭代使用相同步长 $t = 1$
- $\|\nabla f(x)\|_2$ 二次收敛到**0**：若 $\|\nabla f(x^{(k)})\|_2 < \eta$

$$\frac{L}{2m^2}\|\nabla f(x^l)\|_2 \leq \left(\frac{L}{2m^2}\|\nabla f(x^k)\|_2\right)^{2^{l-k}} \leq \left(\frac{1}{2}\right)^{2^{l-k}}, \qquad l \geq k$$

结论

结论

- 达到 $f(x) - p^\star \leq \epsilon$ 的迭代次数上界为

结论

- 达到 $f(x) - p^\star \leq \epsilon$ 的迭代次数上界为

$$\frac{f(x^{(0)}) - p^\star}{\gamma} + \log_2 \log_2(\epsilon_0/\epsilon)$$

结论

- 达到 $f(x) - p^\star \leq \epsilon$ 的迭代次数上界为

$$\frac{f(x^{(0)}) - p^\star}{\gamma} + \log_2 \log_2(\epsilon_0/\epsilon)$$

- γ, ϵ_0 为常数，取决于 m, L, $x^{(0)}$

结论

- 达到 $f(x) - p^\star \leq \epsilon$ 的迭代次数上界为

$$\frac{f(x^{(0)}) - p^\star}{\gamma} + \log_2 \log_2(\epsilon_0/\epsilon)$$

- γ，ϵ_0 为常数，取决于 m，L，$x^{(0)}$
- 第二项较小，实际中基本为常数

结论

- 达到 $f(x) - p^\star \leq \epsilon$ 的迭代次数上界为

$$\frac{f(x^{(0)}) - p^\star}{\gamma} + \log_2 \log_2(\epsilon_0/\epsilon)$$

- γ，ϵ_0 为常数，取决于 m，L，$x^{(0)}$
- 第二项较小，实际中基本为常数
- 实际上，常数 m，L 通常未知（因此，γ，ϵ_0 也未知）

结论

- 达到 $f(x) - p^\star \leq \epsilon$ 的迭代次数上界为

$$\frac{f(x^{(0)}) - p^\star}{\gamma} + \log_2 \log_2(\epsilon_0/\epsilon)$$

- γ，ϵ_0 为常数，取决于 m，L，$x^{(0)}$
- 第二项较小，实际中基本为常数
- 实际上，常数 m，L 通常未知（因此，γ，ϵ_0 也未知）
- 为收敛性提供了定性理解（指解释了两个阶段）

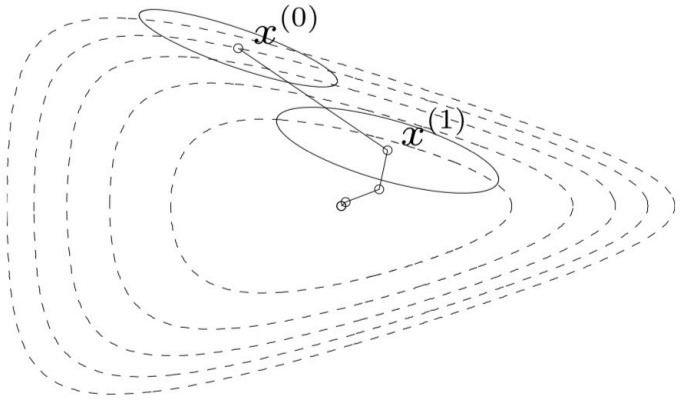
例

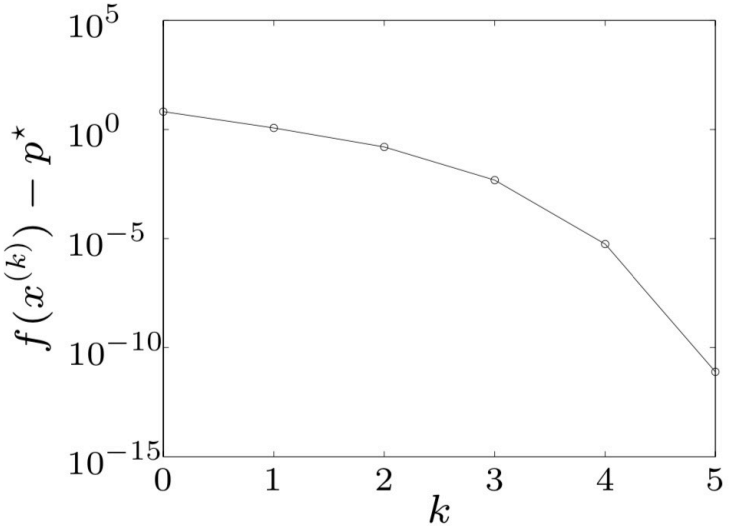
例

- R^2空间中的例子

例

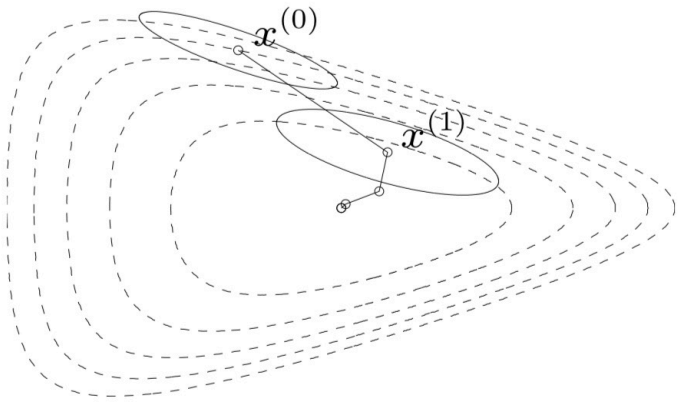
- R^2空间中的例子

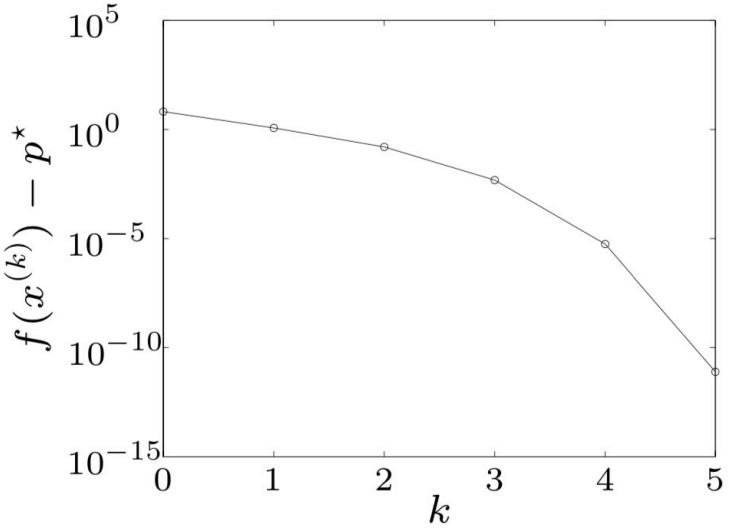




例

- R^2空间中的例子

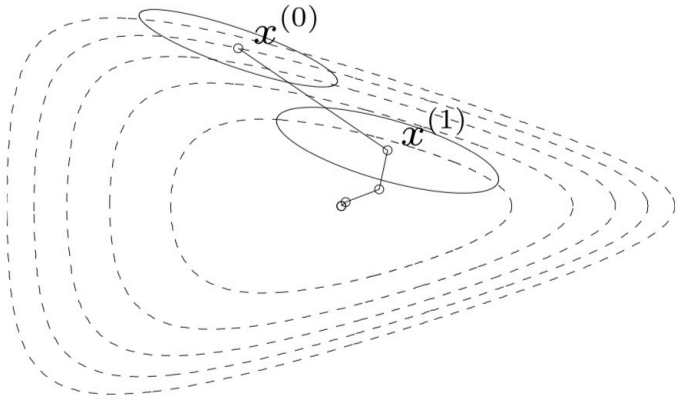


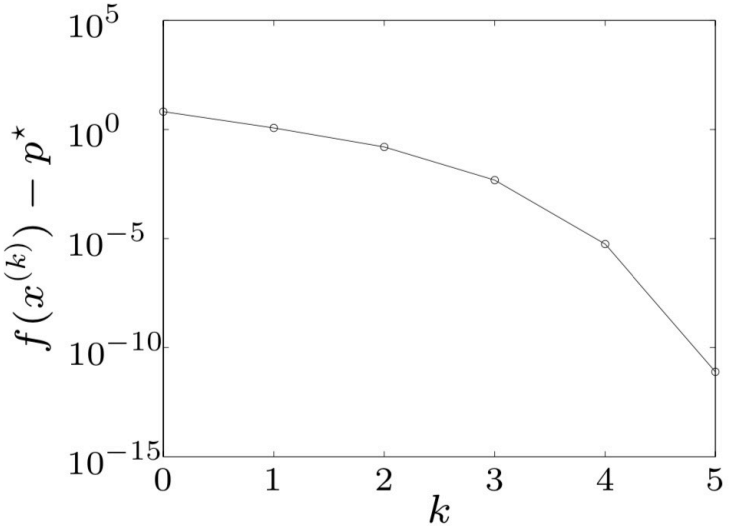


- 回溯参数 $\alpha = 0.1,\ \beta = 0.7$

例

- R^2空间中的例子

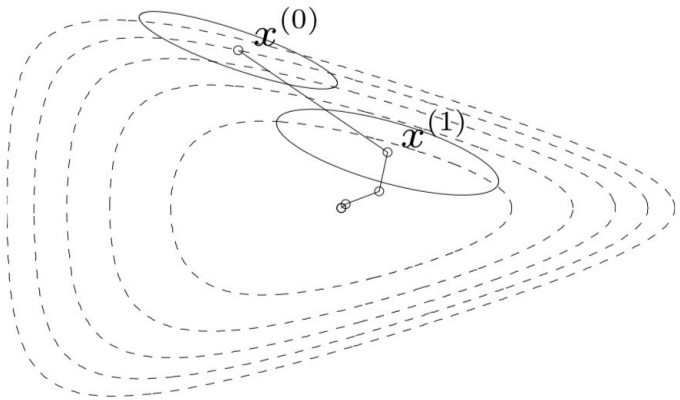


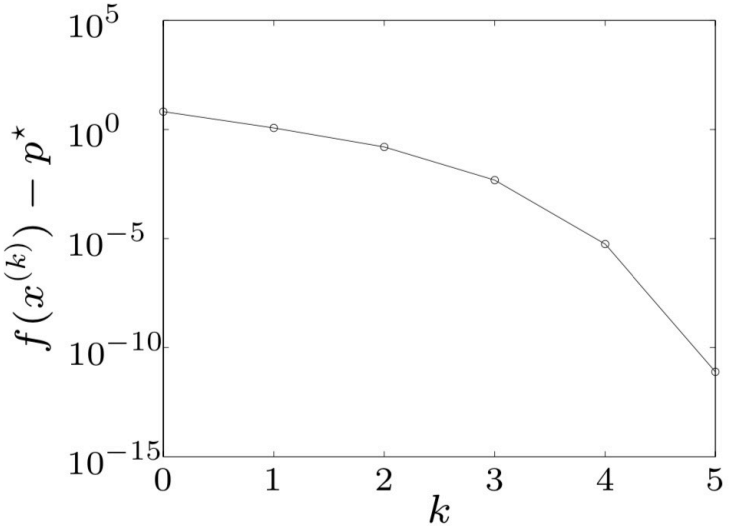


- 回溯参数 $\alpha = 0.1,\ \beta = 0.7$
- 使用**5**步即收敛

例

- R^2空间中的例子





- 回溯参数 $\alpha = 0.1,\ \beta = 0.7$
- 使用**5**步即收敛
- 二次局部收敛

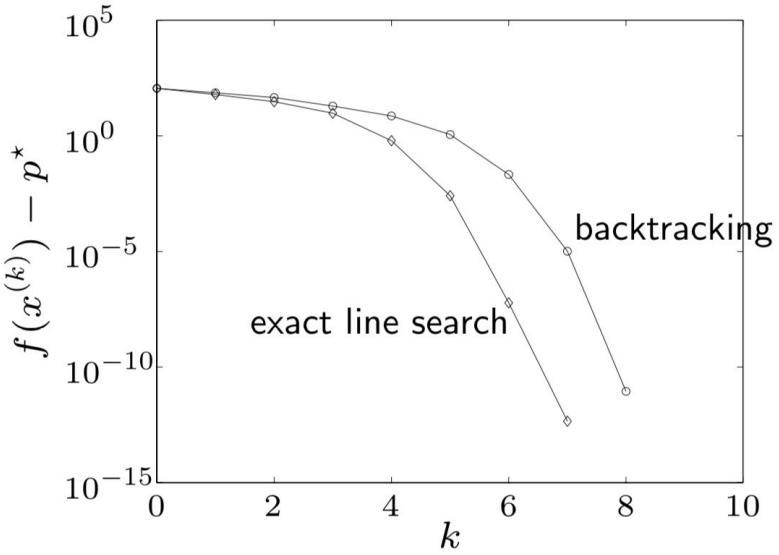
例

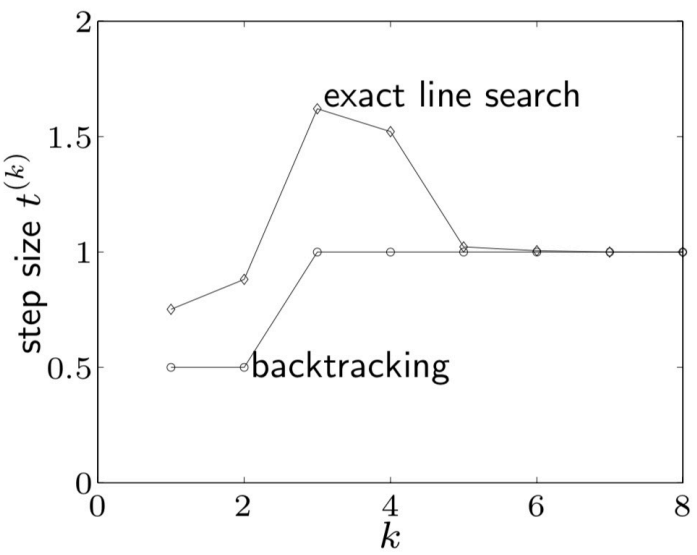
例

- R^{100}空间中的例子

例

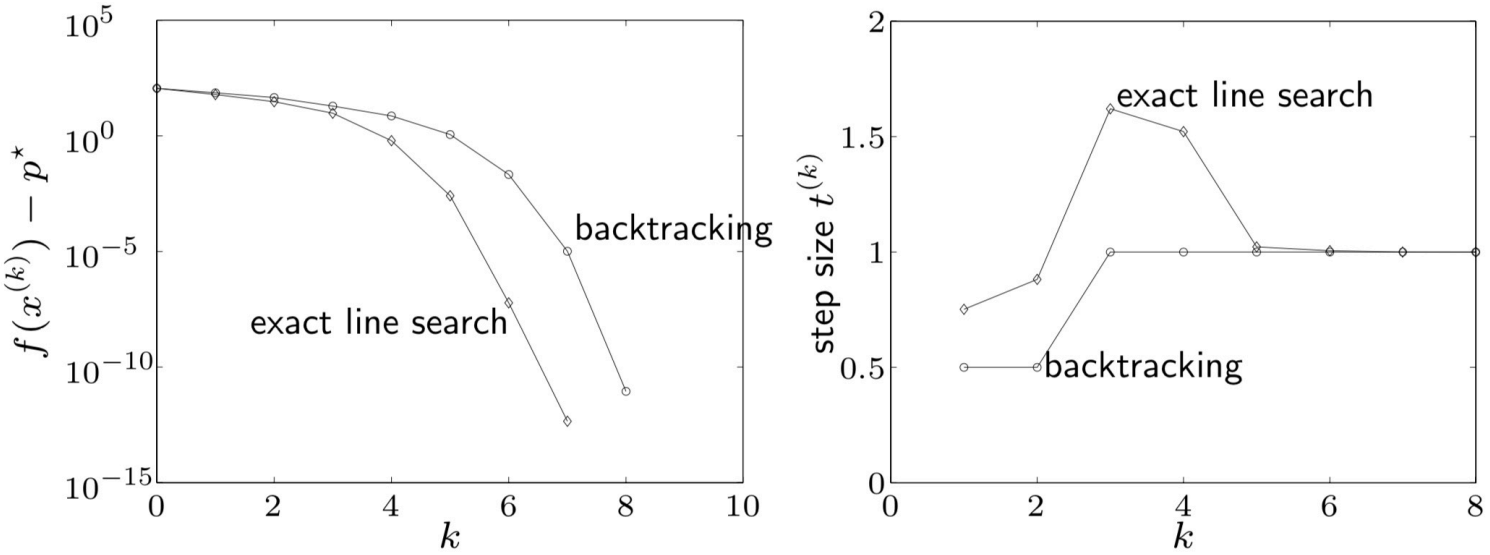
- R^{100}空间中的例子





例

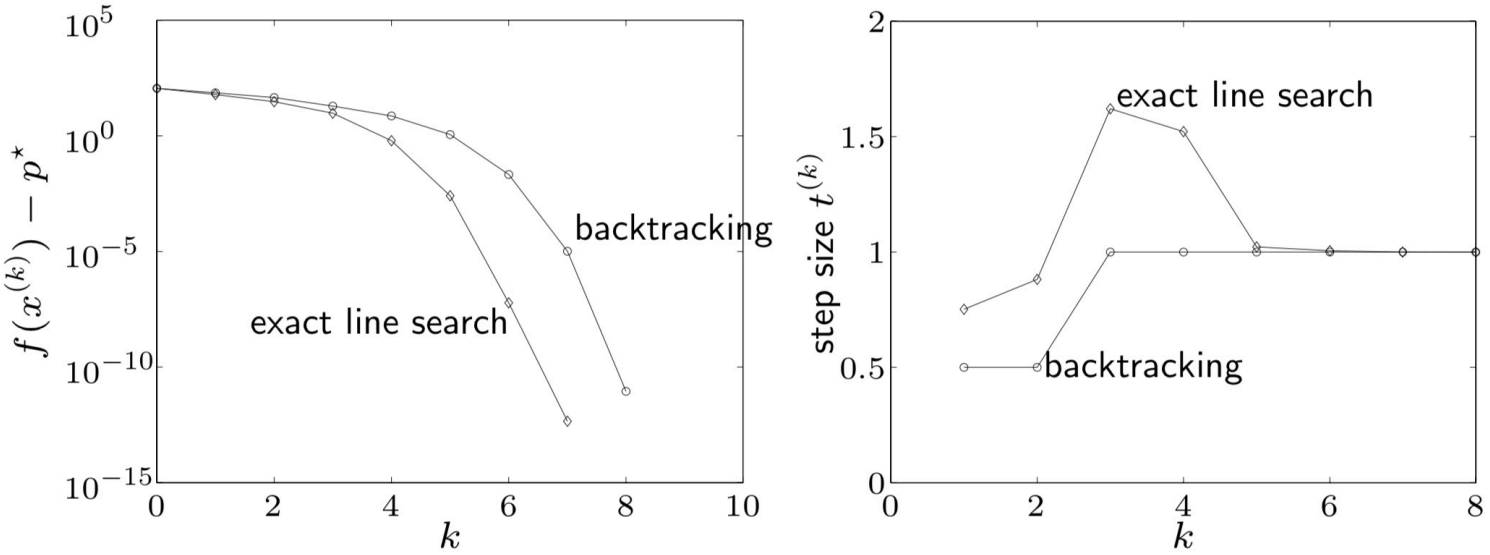
- R^{100}空间中的例子



- 回溯参数 $\alpha = 0.01$, $\beta = 0.5$

例

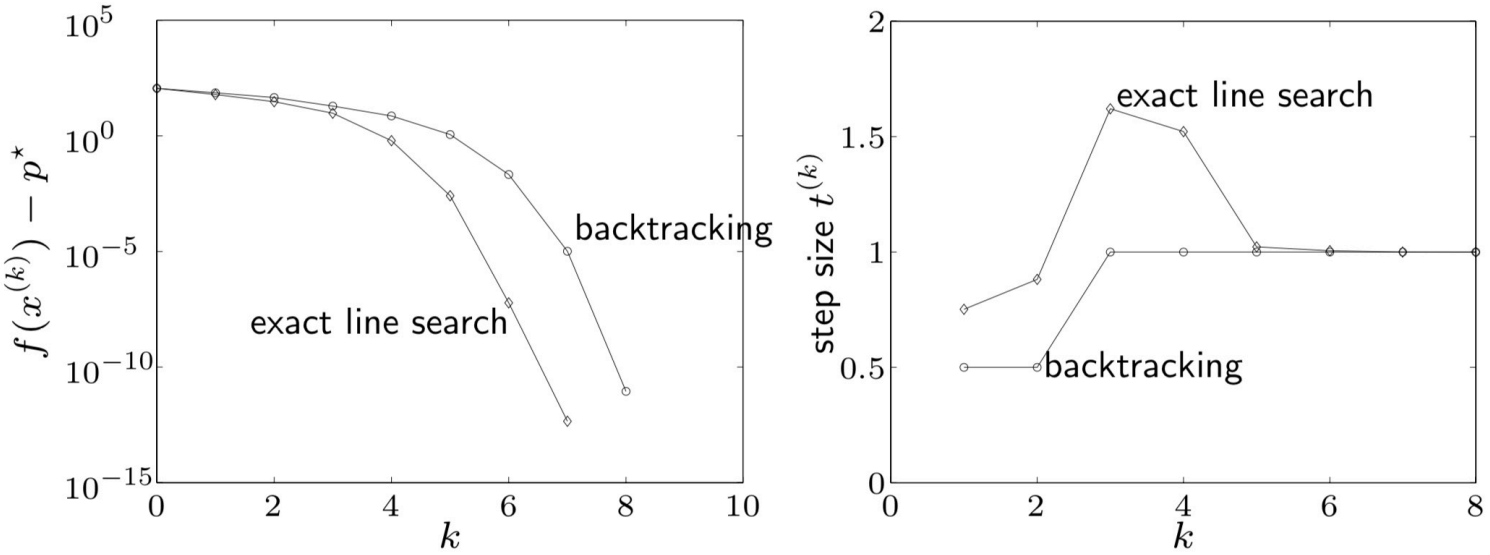
- R^{100}空间中的例子



- 回溯参数 $\alpha = 0.01, \beta = 0.5$
- 回溯直线搜索和精确搜索基本一样快（且更为简单）

例

- R^{100}空间中的例子



- 回溯参数 $\alpha = 0.01$, $\beta = 0.5$
- 回溯直线搜索和精确搜索基本一样快（且更为简单）
- 清楚的显示了算法的两个阶段

例

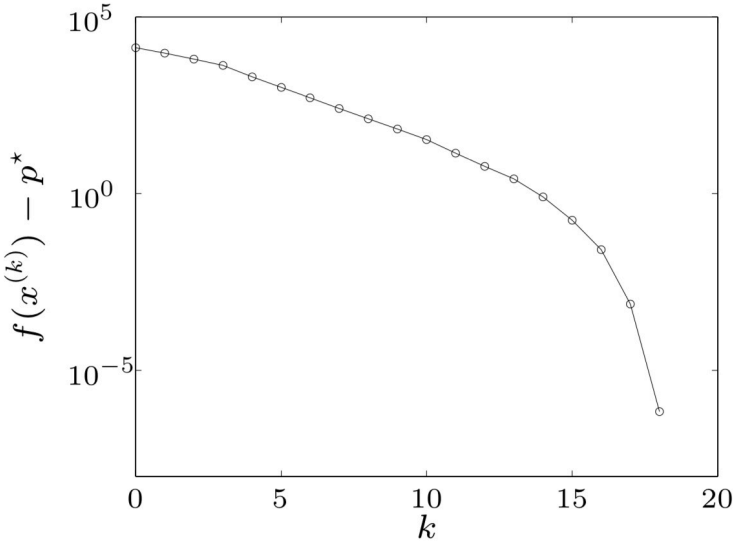
例

- R^{10000}空间的例子（带有稀疏的a_i）

例

- R^{10000}空间的例子（带有稀疏的a_i）

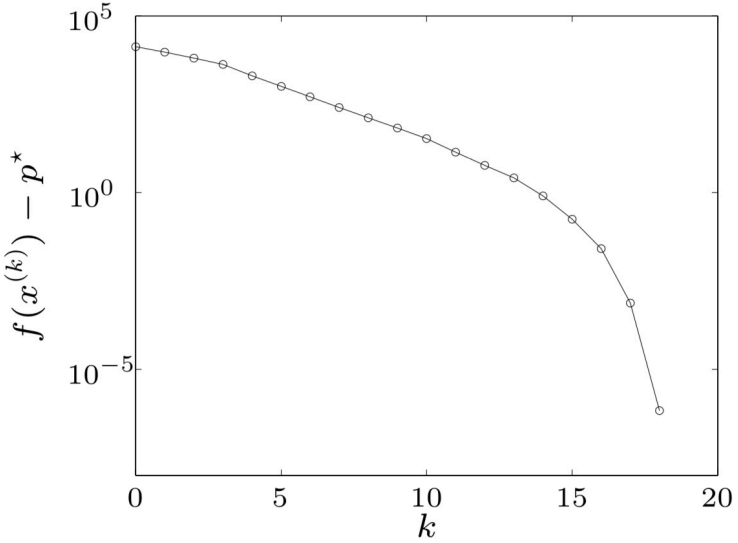
$$f(x) = -\sum_{i=1}^{10000} \log(1 - x_i^2) - \sum_{i=1}^{100000} \log(b_i - a_i^T x)$$



例

- R^{10000}空间的例子（带有稀疏的a_i）

$$f(x) = -\sum_{i=1}^{10000} \log(1 - x_i^2) - \sum_{i=1}^{100000} \log(b_i - a_i^T x)$$

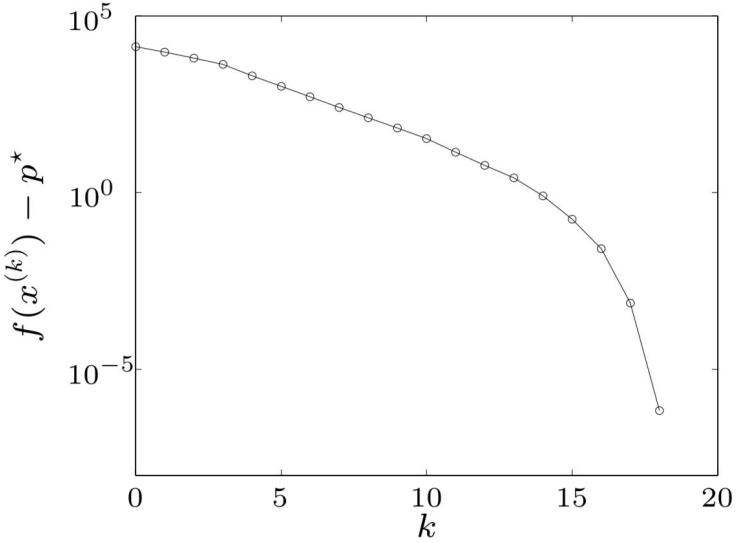


- 回溯参数 $\alpha = 0.01$, $\beta = 0.5$

例

- R^{10000}空间的例子（带有稀疏的a_i）

$$f(x) = -\sum_{i=1}^{10000} \log(1 - x_i^2) - \sum_{i=1}^{100000} \log(b_i - a_i^T x)$$



- 回溯参数 $\alpha = 0.01$, $\beta = 0.5$
- 和较小的例子具有同样的性能

自和谐

自和谐

- 经典收敛性分析的不足

自和谐

- 经典收敛性分析的不足
  - 取决于未知常量（m，L，…）

自和谐

- 经典收敛性分析的不足
  - 取决于未知常量（m，L，…）
  - 界限不具仿射不变性

自和谐

- 经典收敛性分析的不足
  - 取决于未知常量（m，L，…）
  - 界限不具仿射不变性
- 使用自和谐的收敛性分析（**Nesterov**和**Nemirovski**）

自和谐

- 经典收敛性分析的不足
  - 取决于未知常量（m，L，…）
  - 界限不具仿射不变性
- 使用自和谐的收敛性分析（**Nesterov**和**Nemirovski**）
  - 不受任何未知常量影响

自和谐

- 经典收敛性分析的不足
  - 取决于未知常量（m，L，…）
  - 界限不具仿射不变性
- 使用自和谐的收敛性分析（**Nesterov**和**Nemirovski**）
  - 不受任何未知常量影响
  - 给出了具有仿射不变性的边界

自和谐

- 经典收敛性分析的不足
  - 取决于未知常量（m，L，…）
  - 界限不具仿射不变性
- 使用自和谐的收敛性分析（**Nesterov**和**Nemirovski**）
  - 不受任何未知常量影响
  - 给出了具有仿射不变性的边界
  - 可用于特殊的凸函数（自和谐函数）

自和谐

- 经典收敛性分析的不足
  - 取决于未知常量（m，L，…）
  - 界限不具仿射不变性
- 使用自和谐的收敛性分析（**Nesterov**和**Nemirovski**）
  - 不受任何未知常量影响
  - 给出了具有仿射不变性的边界
  - 可用于特殊的凸函数（自和谐函数）
  - 可用来分析凸优化的多项式时间内点法

自和谐方程

自和谐方程

- 定义

自和谐方程

- 定义
- 凸函数 $f : \mathbf{R} \to \mathbf{R}$为自和谐函数

自和谐方程

- 定义
- 凸函数 $f : \mathbf{R} \to \mathbf{R}$ 为自和谐函数
  - 对所有 $x \in \mathbf{dom}\, f$ ，满足

自和谐方程

- 定义
- 凸函数 $f : \mathbf{R} \to \mathbf{R}$ 为自和谐函数
  - 对所有 $x \in \mathbf{dom}\, f$ ，满足

$$|f'''(x)| \leq 2f''(x)^{3/2}$$

自和谐方程

- 定义
- 凸函数 $f:\mathbf{R}\to\mathbf{R}$为自和谐函数
  - 对所有$x\in\mathbf{dom}\,f$，满足

$$|f'''(x)|\le 2f''(x)^{3/2}$$

- 函数 $f:\mathbf{R}^n\to\mathbf{R}$ 为自和谐函数

自和谐方程

- 定义
- 凸函数 $f:\mathbf{R}\to\mathbf{R}$为自和谐函数
  - 对所有 $x\in\mathbf{dom}\,f$ ，满足

$$|f'''(x)|\le 2f''(x)^{3/2}$$

- 函数 $f:\mathbf{R}^n\to\mathbf{R}$ 为自和谐函数
  - 对所有 $x\in\mathbf{dom}\,f$, $v\in\mathbf{R}^n$ 满足 $g(t)=f(x+tv)$ 为自和谐函数

自和谐方程

自和谐方程

- R空间

自和谐方程

- *R*空间
  - 线性和二次函数

自和谐方程

- R空间
  - 线性和二次函数
  - 负对数函数 $f(x) = -\log x$

自和谐方程

- R空间
  - 线性和二次函数
  - 负对数函数 $f(x) = -\log x$
  - 负熵加负对数 $f(x) = x\log x - \log x$

自和谐方程

- R空间
  - 线性和二次函数
  - 负对数函数 $f(x) = -\log x$
  - 负熵加负对数 $f(x) = x\log x - \log x$
- 仿射不变性：

自和谐方程

- R空间
  - 线性和二次函数
  - 负对数函数 $f(x) = -\log x$
  - 负熵加负对数 $f(x) = x\log x - \log x$
- 仿射不变性：
  - 若 $f : \mathbf{R} \to \mathbf{R}$ 为自和谐，则 $\tilde{f}(y) = f(ay + b)$ 为自和谐

自和谐方程

- *R*空间
  - 线性和二次函数
  - 负对数函数 $f(x) = -\log x$
  - 负熵加负对数 $f(x) = x\log x - \log x$
- 仿射不变性：
  - 若 $f : \mathbf{R} \to \mathbf{R}$ 为自和谐，则 $\tilde{f}(y) = f(ay + b)$ 为自和谐

$$\tilde{f}'''(y) = a^3 f'''(ay + b), \qquad \tilde{f}''(y) = a^2 f''(ay + b)$$

基于自和谐函数的收敛分析

基于自和谐函数的收敛分析

- 存在常数 $\eta \in (0, 1/4], \gamma > 0$ 满足:

基于自和谐函数的收敛分析

- 存在常数 $\eta \in (0, 1/4], \gamma > 0$ 满足：
  - 若 $\lambda(x) > \eta$，有 $f(x^{(k+1)}) - f(x^{(k)}) \leq -\gamma$

基于自和谐函数的收敛分析

- 存在常数 $\eta \in (0, 1/4], \gamma > 0$ 满足：
  - 若 $\lambda(x) > \eta$，有 $f(x^{(k+1)}) - f(x^{(k)}) \leq -\gamma$
  - 若 $\lambda(x) \leq \eta$，有 $2\lambda(x^{(k+1)}) \leq \left(2\lambda(x^{(k)})\right)^2$

基于自和谐函数的收敛分析

- 存在常数 $\eta \in (0, 1/4], \gamma > 0$ 满足：
  - 若 $\lambda(x) > \eta$，有 $f(x^{(k+1)}) - f(x^{(k)}) \leq -\gamma$
  - 若 $\lambda(x) \leq \eta$，有 $2\lambda(x^{(k+1)}) \leq \left(2\lambda(x^{(k)})\right)^2$
- （η 和γ 只取决于回溯参数 α, β ）

基于自和谐函数的收敛分析

- 存在常数 $\eta \in (0, 1/4], \gamma > 0$ 满足：
  - 若 $\lambda(x) > \eta$，有 $f(x^{(k+1)}) - f(x^{(k)}) \leq -\gamma$
  - 若 $\lambda(x) \leq \eta$，有 $2\lambda(x^{(k+1)}) \leq \left(2\lambda(x^{(k)})\right)^2$
- （η 和γ 只取决于回溯参数 α, β ）
- 复杂性边界：**Newton**收敛次数上界为

基于自和谐函数的收敛分析

- 存在常数 $\eta \in (0, 1/4], \gamma > 0$ 满足：
  - 若 $\lambda(x) > \eta$，有 $f(x^{(k+1)}) - f(x^{(k)}) \leq -\gamma$
  - 若 $\lambda(x) \leq \eta$，有 $2\lambda(x^{(k+1)}) \leq \left(2\lambda(x^{(k)})\right)^2$
- （η 和γ 只取决于回溯参数 α, β ）
- 复杂性边界：**Newton**收敛次数上界为

$$\frac{f(x^{(0)}) - p^\star}{\gamma} + \log_2 \log_2(1/\epsilon)$$

基于自和谐函数的收敛分析

- 存在常数 $\eta \in (0, 1/4], \gamma > 0$ 满足：
  - 若 $\lambda(x) > \eta$，有 $f(x^{(k+1)}) - f(x^{(k)}) \leq -\gamma$
  - 若 $\lambda(x) \leq \eta$，有 $2\lambda(x^{(k+1)}) \leq \left(2\lambda(x^{(k)})\right)^2$
- （η 和γ 只取决于回溯参数 α, β ）
- 复杂性边界：**Newton**收敛次数上界为

$$\frac{f(x^{(0)}) - p^\star}{\gamma} + \log_2 \log_2(1/\epsilon)$$

- 若 $\alpha = 0.1, \beta = 0.8, \epsilon = 10^{-10}$，上界估计为

基于自和谐函数的收敛分析

- 存在常数 $\eta \in (0, 1/4], \gamma > 0$ 满足：
  - 若 $\lambda(x) > \eta$，有 $f(x^{(k+1)}) - f(x^{(k)}) \leq -\gamma$
  - 若 $\lambda(x) \leq \eta$，有 $2\lambda(x^{(k+1)}) \leq \left(2\lambda(x^{(k)})\right)^2$
- （η 和γ 只取决于回溯参数 α, β ）
- 复杂性边界：**Newton**收敛次数上界为

$$\frac{f(x^{(0)}) - p^\star}{\gamma} + \log_2 \log_2(1/\epsilon)$$

- 若 $\alpha = 0.1, \beta = 0.8, \epsilon = 10^{-10}$，上界估计为

$$375(f(x^{(0)}) - p^\star) + 6$$

数值例子

数值例子

- **150**个随机生成的实例

数值例子

- **150**个随机生成的实例

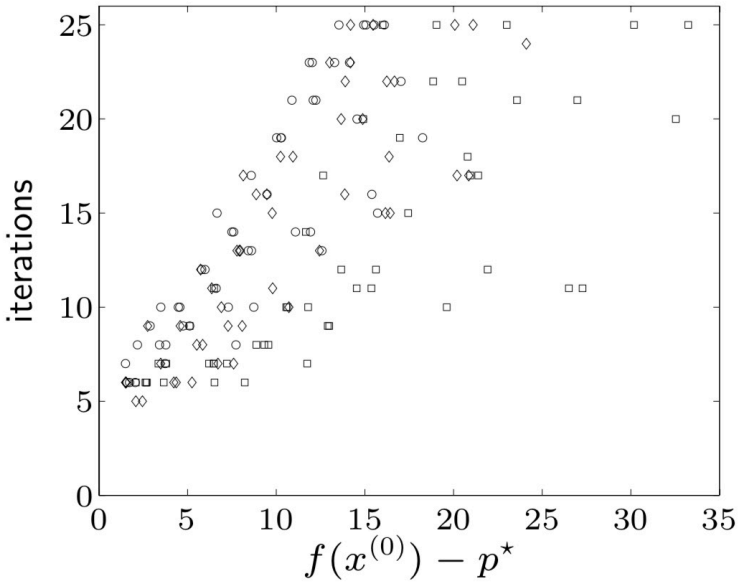
$$\text{minimize} \quad f(x) = -\sum_{i=1}^{m} \log(b_i - a_i^T x)$$

数值例子

- **150**个随机生成的实例

$$\text{minimize} \quad f(x) = -\sum_{i=1}^{m} \log(b_i - a_i^T x)$$

○: $m = 100,\ n = 50$
□: $m = 1000,\ n = 500$
◊: $m = 1000,\ n = 50$

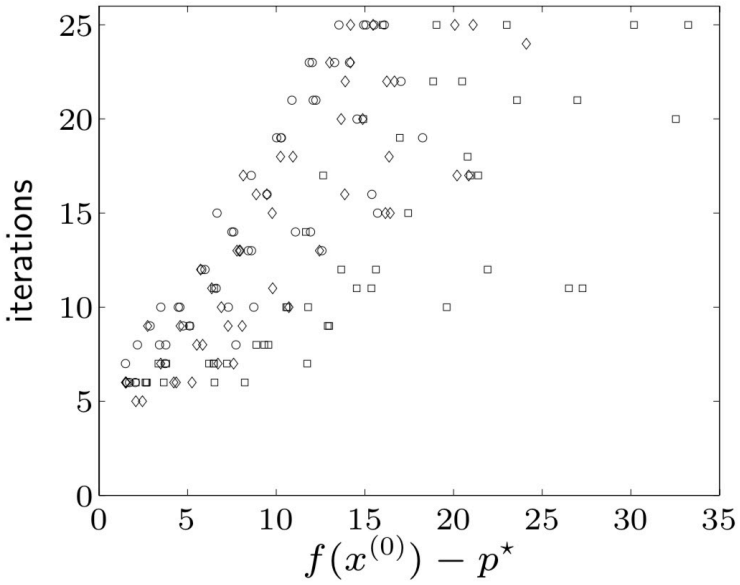


数值例子

- **150**个随机生成的实例

$$\text{minimize} \quad f(x) = -\sum_{i=1}^{m} \log(b_i - a_i^T x)$$

○: $m = 100,\ n = 50$
□: $m = 1000,\ n = 500$
◇: $m = 1000,\ n = 50$



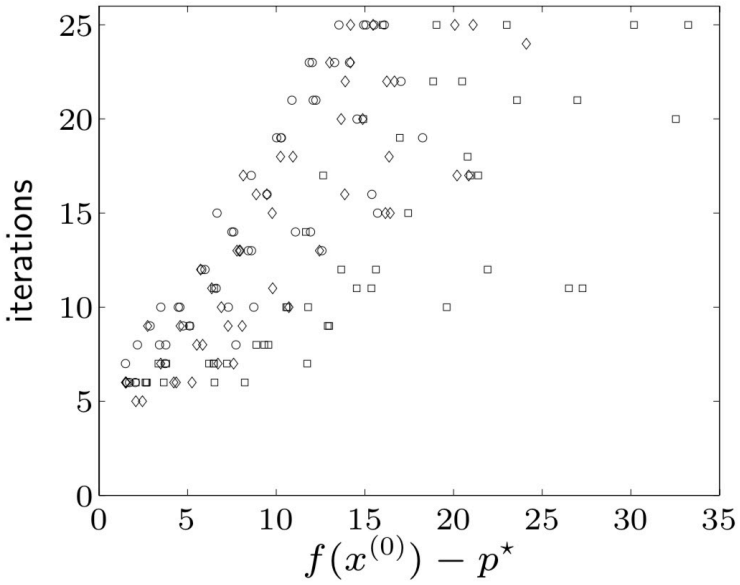
- 收敛次数大大少于 $375(f(x^{(0)}) - p^\star) + 6$

数值例子

- **150**个随机生成的实例

$$\text{minimize} \quad f(x) = -\sum_{i=1}^{m} \log(b_i - a_i^T x)$$

○: $m = 100,\ n = 50$
□: $m = 1000,\ n = 500$
◇: $m = 1000,\ n = 50$



- 收敛次数大大少于 $375(f(x^{(0)}) - p^\star) + 6$
- $c(f(x^{(0)}) - p^\star) + 6$是所需迭代次数的一个不坏的估计（$c$较小）

实现

实现

- 每次迭代的主要代价：估计导数并求解**Newton**系统 $H\Delta x = -g$

实现

- 每次迭代的主要代价：估计导数并求解**Newton**系统 $H\Delta x = -g$
  - 其中，$H = \nabla^2 f(x)$，$g = \nabla f(x)$

实现

- 每次迭代的主要代价：估计导数并求解**Newton**系统 $H\Delta x = -g$
  - 其中，$H = \nabla^2 f(x)$，$g = \nabla f(x)$
- 通过**Cholesky**因式分解

实现

- 每次迭代的主要代价：估计导数并求解**Newton**系统 $H\Delta x = -g$
  - 其中，$H = \nabla^2 f(x)$, $g = \nabla f(x)$
- 通过**Cholesky**因式分解

$$H = LL^T, \qquad \Delta x_{\mathrm{nt}} = -L^{-T}L^{-1}g, \qquad \lambda(x) = \|L^{-1}g\|_2$$

实现

- 每次迭代的主要代价：估计导数并求解**Newton**系统 $H\Delta x = -g$
  - 其中，$H = \nabla^2 f(x)$，$g = \nabla f(x)$
- 通过**Cholesky**因式分解

$$H = LL^T, \qquad \Delta x_{\mathrm{nt}} = -L^{-T}L^{-1}g, \qquad \lambda(x) = \|L^{-1}g\|_2$$

- 对非结构系统需要$(1/3)n^3$次浮点运算

实现

- 每次迭代的主要代价：估计导数并求解**Newton**系统 $H\Delta x = -g$
  - 其中，$H = \nabla^2 f(x)$, $g = \nabla f(x)$
- 通过**Cholesky**因式分解

$$H = LL^T, \qquad \Delta x_{\mathrm{nt}} = -L^{-T}L^{-1}g, \qquad \lambda(x) = \|L^{-1}g\|_2$$

- 对非结构系统需要$(1/3)n^3$次浮点运算
- 若H为稀疏或带状的，代价 $\ll (1/3)n^3$