第六章：等式约束的优化问题

（1）牛顿法

曹语

课程主页：https://yucaoyc.github.io/math3806

我们考虑如下的等式约束问题：

$$\begin{array}{ll} \text{minimize} & f(\boldsymbol{x}) \\ \text{subject to} & \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b} \end{array}$$

假设 f 是凸函数，则该问题是凸优化问题；由________________可知，强对偶成立。

目标： 学习算法来求解该问题

1. 转换成无约束问题
2. 基于强对偶，转换成对偶问题
3. 基于 KKT 条件，直接构造牛顿法

我们考虑如下的等式约束问题:

$$\begin{array}{ll} \text{minimize} & f(\boldsymbol{x}) \\ \text{subject to} & \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b} \end{array}$$

假设:

- f 是二次连续可微凸函数;
- $\boldsymbol{A} \in \mathbb{R}^{p \times n}$, rank $\boldsymbol{A} = p < n$;
- 该优化问题的可行集非空。

回顾练习: 请写出对该问题的 KKT 条件。

优化问题

$$\begin{array}{ll} \text{minimize} & f(\boldsymbol{x}) \\ \text{subject to} & \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b} \end{array}$$

的最优解满足如下的 KKT 条件:

$$\boxed{\begin{array}{ll} \text{(原可行方程)} & \boldsymbol{A}\boldsymbol{x}^{\star} = \boldsymbol{b}, \\ \text{(对偶可行方程)} & \nabla f\left(\boldsymbol{x}^{\star}\right) + \boldsymbol{A}^{\top}\boldsymbol{\nu}^{\star} = \mathbf{0}. \end{array}} \tag{1}$$

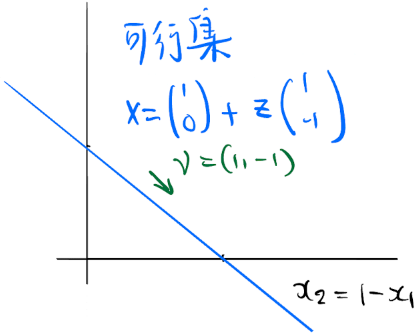
目录

例子

$$\begin{array}{ll} \text{minimize} & f(\boldsymbol{x}) \\ \text{subject to} & \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b} \end{array}$$

考虑具体的问题 $\boldsymbol{x} \in \mathbb{R}^2$： $x_1 + x_2 = 1$。我们可以把可行集参数化成

$$\begin{aligned} & \{\boldsymbol{x} : \ \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}\} \\ =& \{\boldsymbol{x} : \ \boldsymbol{x} = \boldsymbol{x}_0 + z\boldsymbol{v} \equiv (1,0) + z\big(1,-1\big)\} \end{aligned}$$



故而该优化问题变成了无约束的优化问题

$$\text{minimize} \quad f(\boldsymbol{x}_0 + z\boldsymbol{v})$$

因而我们可以使用梯度法、牛顿法等。

一般情景

- 由于矩阵 $\boldsymbol{A} \in \mathbb{R}^{p \times n}$ 具有 $\mathrm{rank}(\boldsymbol{A}) = p$（否则我们可以简化该等式限制），因此 $\dim\big(\mathrm{Range}(\boldsymbol{A}^\top)\big) = p$。因而，子空间 $\mathrm{Null}(\boldsymbol{A}) = \mathrm{Range}(\boldsymbol{A}^\top)^\perp$ 具有维数 $n-p$。
- 所以我们可以找到 $n-p$ 个独立的向量 $\boldsymbol{v}_i$ 使得

$$\mathrm{Null}(\boldsymbol{A}) = \left\{ \sum_{i=1}^{n-p} \boldsymbol{v}_i z_i = \boldsymbol{F}\boldsymbol{z} : \ \boldsymbol{z} \in \mathbb{R}^{n-p} \right\}, \qquad \boldsymbol{F} = \begin{bmatrix} \boldsymbol{v}_1 & \cdots \boldsymbol{v}_{n-p} \end{bmatrix} \in \mathbb{R}^{n \times (n-p)}$$

- 总结而言，我们可以找到 $\boldsymbol{F}$ 满足

$$\{\boldsymbol{x} : \ \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}\} = \{\boldsymbol{x}_0 + \boldsymbol{F}\boldsymbol{z} : \ \boldsymbol{z} \in \mathbb{R}^{n-p}\}$$

因此优化问题变成了

$$\text{minimize} \quad f\big(\boldsymbol{x}_0 + \boldsymbol{F}\boldsymbol{z}\big)$$

例子

例子： 对前面的 2D 问题 $\boldsymbol{x}_0 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$， $\boldsymbol{F} = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$， $z \in \mathbb{R}$，

$$\text{minimize} \quad \tilde{f}(z) = f(\boldsymbol{x}_0 + \boldsymbol{F}z)$$

请计算 $f'(z), f''(z)$。

答案：

$$\tilde{f}'(z) = \partial_{x_1} f(\boldsymbol{x}_0 + \boldsymbol{F}z)\boldsymbol{F}_{11} + \partial_{x_2} f(\boldsymbol{x}_0 + \boldsymbol{F}z)\boldsymbol{F}_{21} = \boldsymbol{F}^\top \nabla f(\boldsymbol{x}_0 + \boldsymbol{F}z)$$

$$\begin{aligned}
\tilde{f}''(z) &= \partial_{x_1,x_1} f(\boldsymbol{x}_0 + \boldsymbol{F}z)\big(\boldsymbol{F}_{11}\big)^2 + \partial_{x_1,x_2} f(\boldsymbol{x}_0 + \boldsymbol{F}z)\boldsymbol{F}_{11}\boldsymbol{F}_{21} \\
&\quad + \partial_{x_2,x_1} f(\boldsymbol{x}_0 + \boldsymbol{F}z)\boldsymbol{F}_{21}\boldsymbol{F}_{11} + \partial_{x_2,x_2} f(\boldsymbol{x}_0 + \boldsymbol{F}z)\big(\boldsymbol{F}_{21}\big)^2 \\
&= \boldsymbol{F}^\top \nabla^2 f(\boldsymbol{x}_0 + \boldsymbol{F}z)\boldsymbol{F}
\end{aligned}$$

一般情景

$$\text{minimize} \quad \tilde{f}(\boldsymbol{z}) = f(\boldsymbol{x}_0 + \boldsymbol{F}\boldsymbol{z})$$

通过直接计算可知（具体步骤见课内）:

$$\begin{aligned}
\nabla \tilde{f}(\boldsymbol{z}) &= \boldsymbol{F}^\top \nabla f(\boldsymbol{x}_0 + \boldsymbol{F}\boldsymbol{z}) \\
\nabla^2 \tilde{f}(\boldsymbol{z}) &= \boldsymbol{F}^\top \nabla^2 f(\boldsymbol{x}_0 + \boldsymbol{F}\boldsymbol{z})\boldsymbol{F}
\end{aligned}$$

因此，我们可以直接使用上面的公式来求解下降方向:

$$\begin{aligned}
&\text{梯度法} \qquad \Delta \boldsymbol{z} = -\nabla \tilde{f}(\boldsymbol{z}) \\
&\text{牛顿法} \qquad \Delta \boldsymbol{z} = -\left(\nabla^2 \tilde{f}(\boldsymbol{z})\right)^{-1} \nabla \tilde{f}(\boldsymbol{z})
\end{aligned}$$

好处: 思路简单

坏处: 需要 $\boldsymbol{F}$，构建矩阵 $\boldsymbol{F}$ 不一定容易

目录

对偶问题

$$
\begin{aligned}
g(\boldsymbol{\nu}) &= \inf_{\boldsymbol{x}} \; f(\boldsymbol{x}) + \boldsymbol{\nu}^\top (\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}) \\
&= -\boldsymbol{b}^\top \boldsymbol{\nu} + \inf_{\boldsymbol{x}} \left(f(\boldsymbol{x}) + \boldsymbol{\nu}^\top \boldsymbol{A}\boldsymbol{x} \right) \\
&= -\boldsymbol{b}^\top \boldsymbol{\nu} - f^* \left(-\boldsymbol{A}^\top \boldsymbol{\nu} \right)
\end{aligned}
$$

其中 f^* 是 f 的共轭函数

对偶问题是

$$
\begin{aligned}
&\text{maximize} \quad g(\boldsymbol{\nu}) \\
= \; &\text{maximize} \quad -\boldsymbol{b}^\top \boldsymbol{\nu} - f^* \left(-\boldsymbol{A}^\top \boldsymbol{\nu} \right)
\end{aligned}
$$

好处： 该对偶问题是无约束的优化问题

坏处： 依赖于 f^* 是否好算

目录

回顾

- 在无约束优化中，我们在 $\boldsymbol{x}$ 局部使用二次型函数来近似 f：

$$\hat{f}(\boldsymbol{y}) = f(\boldsymbol{x}) + \nabla f(\boldsymbol{x})^\top \boldsymbol{v} + \frac{1}{2}\boldsymbol{v}^\top \nabla^2 f(\boldsymbol{x}) \boldsymbol{v}$$

  其中，$\boldsymbol{v} = \boldsymbol{y} - \boldsymbol{x}$。通过最小化 $\hat{f}$，我们可得

$$\boldsymbol{v} = -\left(\nabla^2 f(\boldsymbol{x})\right)^{-1} \nabla f(\boldsymbol{x})$$

- 或者局部近似梯度

$$\nabla f(\boldsymbol{y}) \approx \nabla f(\boldsymbol{x}) + \nabla^2 f(\boldsymbol{x})(\boldsymbol{y} - \boldsymbol{x})$$

  通过求解 $\nabla f(\boldsymbol{y}) = \mathbf{0}$，我们依旧可以得到同样的步径 $\boldsymbol{v}$。

(1) 牛顿步径

通过二次型近似的思路，牛顿步径是如下问题中 $\boldsymbol{u}$ 的最优解

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

根据该二次型函数的 KKT 条件，最优解满足如下线性方程组：

$$\begin{bmatrix} \nabla^2 f(\boldsymbol{x}) & \boldsymbol{A}^\top \\ \boldsymbol{A} & 0 \end{bmatrix} \begin{bmatrix} \Delta \boldsymbol{x}_{\text{nt}} \\ \boldsymbol{\omega} \end{bmatrix} = \begin{bmatrix} -\nabla f(\boldsymbol{x}) \\ \boldsymbol{b} - \boldsymbol{A}\boldsymbol{x} \end{bmatrix} \tag{2}$$

如果当前位置 $\boldsymbol{x}$ 已经满足等式约束 $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$，则

$$\begin{bmatrix} \nabla^2 f(\boldsymbol{x}) & \boldsymbol{A}^\top \\ \boldsymbol{A} & 0 \end{bmatrix} \begin{bmatrix} \Delta \boldsymbol{x}_{\text{nt}} \\ \boldsymbol{\omega} \end{bmatrix} = \begin{bmatrix} -\nabla f(\boldsymbol{x}) \\ 0 \end{bmatrix} \tag{3}$$

通过局部近似梯度

对于公式(1)，我们可以求解在 $(\boldsymbol{x}, \boldsymbol{\nu})$ 的局部近似解

$$
\begin{aligned}
&\boldsymbol{A}\big(\boldsymbol{x} + \Delta\boldsymbol{x}_{\mathrm{nt}}\big) - \boldsymbol{b} = \mathbf{0} \\
&\nabla f(\boldsymbol{x}) + \nabla^2 f(\boldsymbol{x})\Delta\boldsymbol{x}_{\mathrm{nt}} + \boldsymbol{A}^\top(\boldsymbol{\nu} + \Delta\boldsymbol{\nu}) = \mathbf{0}
\end{aligned}
$$

引入符号 $\boldsymbol{\omega} = \boldsymbol{\nu} + \Delta\boldsymbol{\nu}$，该方程等价于

$$
\begin{bmatrix} \nabla^2 f(\boldsymbol{x}) & \boldsymbol{A}^\top \\ \boldsymbol{A} & 0 \end{bmatrix} \begin{bmatrix} \Delta\boldsymbol{x}_{\mathrm{nt}} \\ \boldsymbol{\omega} \end{bmatrix} = \begin{bmatrix} -\nabla f(\boldsymbol{x}) \\ \boldsymbol{b} - \boldsymbol{A}\boldsymbol{x} \end{bmatrix}
$$

由此可见，我们得到同样的方向 $\Delta\boldsymbol{x}_{\mathrm{nt}}$。

等式约束的牛顿法

Algorithm 1: 等式约束的牛顿法

Input: 满足等式约束的 $\boldsymbol{x}^{(0)}$，阈值 ϵ

```
while True do
    根据公式(3)计算牛顿步径 Δx^(k)
    停止准则：如果误差比较小，则退出
    通过回溯直线搜索选择步长 t^(k) > 0
    x^(k+1) = x^(k) + t^(k) Δx^(k)
end
```

（2）性质：算法能否满足等式约束？

- 如需找到某个 $\boldsymbol{x}^{(0)}$ 满足限制条件，我们仅需找到方程组 $\boldsymbol{A}\boldsymbol{x}^{(0)} = \boldsymbol{b}$ 的某个解。
- 如果 $\boldsymbol{x}^{(0)}$ 满足 $\boldsymbol{A}\boldsymbol{x}^{(0)} = \boldsymbol{b}$，则后续 $\boldsymbol{x}^{(j)}$（$j \geq 0$）永远满足等式约束，无论搜索步径是什么。
- 假使 $\boldsymbol{x}^{(0)}$ 不满足 $\boldsymbol{A}\boldsymbol{x}^{(0)} = \boldsymbol{b}$，取 $t^{(0)} = 1$，则 $\boldsymbol{x}^{(1)}$ 也可满足等式约束。

(2) 性质：算法能否满足函数值下降?

对于任意的定义域内的 $\boldsymbol{x}$，我们可以计算在牛顿方向，函数值变化如下：

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t} f(\boldsymbol{x} + t\Delta \boldsymbol{x}_{\mathrm{nt}})\Big|_{t=0} &= \nabla f(\boldsymbol{x})^\top \Delta \boldsymbol{x}_{\mathrm{nt}} \\ &= -\Delta \boldsymbol{x}_{\mathrm{nt}}^\top \left(\nabla^2 f(\boldsymbol{x}) \Delta \boldsymbol{x}_{\mathrm{nt}} + \boldsymbol{A}^\top \boldsymbol{\omega}\right) \\ &= -\Delta \boldsymbol{x}_{\mathrm{nt}}^\top \nabla^2 f(\boldsymbol{x}) \Delta \boldsymbol{x}_{\mathrm{nt}} + \boldsymbol{\omega}^\top (\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b})\end{aligned}$$

此处，我们使用了公式(2)来简化。

当 $\boldsymbol{x}$ 满足等式限制 $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$ 时，

$$\frac{\mathrm{d}}{\mathrm{d}t} f(\boldsymbol{x} + t\Delta \boldsymbol{x}_{\mathrm{nt}})\Big|_{t=0} = -\Delta \boldsymbol{x}_{\mathrm{nt}}^\top \nabla^2 f(\boldsymbol{x}) \Delta \boldsymbol{x}_{\mathrm{nt}} \leq 0$$

因此，该方向的确是下降方向。

(3) 和方法 1 的牛顿法的关联

对于思路 1：我们需要求解如下的问题

$$\text{minimize} \quad \tilde{f}(\boldsymbol{z}) = f(\boldsymbol{x}_0 + \boldsymbol{F}\boldsymbol{z})$$

其中牛顿步径为

$$\begin{aligned}
\Delta \boldsymbol{z} &= -\left(\nabla^2 \tilde{f}(\boldsymbol{z})\right)^{-1} \nabla \tilde{f}(\boldsymbol{z}) \\
&= -\left(\boldsymbol{F}^\top \nabla^2 f(\boldsymbol{x}_0 + \boldsymbol{F}\boldsymbol{z})\boldsymbol{F}\right)^{-1} \boldsymbol{F}^\top \nabla f(\boldsymbol{x}_0 + \boldsymbol{F}\boldsymbol{z})
\end{aligned}$$

因此，对于 $\boldsymbol{x} = \boldsymbol{x}_0 + \boldsymbol{F}\boldsymbol{z}$，$\boldsymbol{x}$ 方向的牛顿步径为

$$\Delta \boldsymbol{x} = \boldsymbol{F}\Delta \boldsymbol{z} = -\boldsymbol{F}\left(\boldsymbol{F}^\top \nabla^2 f(\boldsymbol{x})\boldsymbol{F}\right)^{-1} \boldsymbol{F}^\top \nabla f(\boldsymbol{x})$$

我们可以验证当 $\boldsymbol{x}$ 满足等式限制 $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$ 时，牛顿步径 $\Delta \boldsymbol{x} = -\boldsymbol{F}\left(\boldsymbol{F}^\top \nabla^2 f(\boldsymbol{x})\boldsymbol{F}\right)^{-1}\boldsymbol{F}^\top \nabla f(\boldsymbol{x})$ 是方程(2)的解，即

$$\begin{bmatrix} \nabla^2 f(\boldsymbol{x}) & \boldsymbol{A}^\top \\ \boldsymbol{A} & 0 \end{bmatrix} \begin{bmatrix} \Delta \boldsymbol{x} \\ \boldsymbol{\nu} \end{bmatrix} = \begin{bmatrix} -\nabla f(\boldsymbol{x}) \\ 0 \end{bmatrix}$$

因此方法 1 和方法 3 的牛顿法其实是等价的。但方法 3 不需要提前构造 $\boldsymbol{F}$。

由于无约束的牛顿法是二阶格式，因此该等式约束的牛顿法也是。

证明：由于 $\boldsymbol{A}\boldsymbol{F} = \mathbf{0}$，因此第 2 行得证。为证明第 1 行成立，我们仅需证明 $\nabla^2 f(\boldsymbol{x})\Delta \boldsymbol{x} + \nabla f(\boldsymbol{x}) = -\boldsymbol{A}^\top \boldsymbol{\nu}$ 有解；因而变成验证 $\nabla^2 f(\boldsymbol{x})\Delta \boldsymbol{x} + \nabla f(\boldsymbol{x}) \in \text{Range}(\boldsymbol{A}^\top)$。由于 $\text{Null}(\boldsymbol{A}) = \text{Range}(\boldsymbol{A}^\top)^\perp$，因此，我们仅需验证对于任意的 $\boldsymbol{z} \in \mathbb{R}^{n-p}$，

$$(\boldsymbol{F}\boldsymbol{z})^\top \left(\nabla^2 f(\boldsymbol{x})\Delta \boldsymbol{x} + \nabla f(\boldsymbol{x})\right) = \mathbf{0}$$

将 $\Delta \boldsymbol{x}$ 的表达式代入，我们可以看出该等式成立。

目录

问题来源： 等式约束的牛顿法本身很高效，一个小问题是：如果函数本身的定义域不是 $\mathbb{R}^n$，有时可能求解某个 $\boldsymbol{x}$ 来满足 $\boldsymbol{A}\boldsymbol{x}=\boldsymbol{b}$ 未必很容易；而使用牛顿法中 $t=1$ 来使得等式约束满足也未必可行。

问题：在不强制等式约束的情况下，能否依旧构造牛顿法?

考虑一般的 KKT 条件(1)，我们可以求解在 $(\boldsymbol{x},\boldsymbol{\nu})$ 的局部近似解

$$
\begin{aligned}
&\boldsymbol{A}\left(\boldsymbol{x}+\Delta\boldsymbol{x}_{\mathrm{nt}}\right)-\boldsymbol{b}=0 \\
&\nabla f(\boldsymbol{x})+\nabla^2 f(\boldsymbol{x})\Delta\boldsymbol{x}_{\mathrm{nt}}+\boldsymbol{A}^\top(\boldsymbol{\nu}+\Delta\boldsymbol{\nu}_{\mathrm{nt}})=\mathbf{0}
\end{aligned}
\tag{4}
$$

我们把 $(\boldsymbol{x},\boldsymbol{\nu})$ 看成是变量，而不是单纯仅更新 $\boldsymbol{x}$。

Input: 定义域内的 $\boldsymbol{x}^{(0)}$，$\boldsymbol{\nu}^{(0)}$

while *True* **do**

  根据公式(4)计算原对偶牛顿方向 $\Delta\boldsymbol{x}_{\text{nt}}^{(k)}, \Delta\boldsymbol{\nu}_{\text{nt}}^{(k)}$

  停止准则：如果误差比较小，则退出

  通过回溯直线搜索选择步长 $t^{(k)} > 0$

  $\boldsymbol{x}^{(k+1)} = \boldsymbol{x}^{(k)} + t^{(k)}\Delta\boldsymbol{x}_{\text{nt}}^{(k)}$

  $\boldsymbol{\nu}^{(k+1)} = \boldsymbol{\nu}^{(k)} + t^{(k)}\Delta\boldsymbol{\nu}_{\text{nt}}^{(k)}$

end

新问题： 因为涉及到了新变量 $\boldsymbol{\nu}$，

- 回溯直线搜索的标准是什么?
- 终止条件是什么?

可以考虑如下残差:

$$R(\boldsymbol{x}, \boldsymbol{\nu}) = \left\|\nabla f(\boldsymbol{x}) + \boldsymbol{A}^\top \boldsymbol{\nu}\right\|^2 + \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}\|^2 = r(\boldsymbol{x}, \boldsymbol{\nu})^2$$

则对于点 $(\boldsymbol{x}^{(k)}, \boldsymbol{\nu}^{(k)})$，通过方程(4)来计算牛顿方向 $(\Delta \boldsymbol{x}_{\mathrm{nt}}^{(k)}, \Delta \boldsymbol{\nu}_{\mathrm{nt}}^{(k)})$，可以验证

$$\frac{\mathrm{d}}{\mathrm{d}t} R\left(\boldsymbol{x}^{(k)} + t\Delta \boldsymbol{x}_{\mathrm{nt}}^{(k)}, \boldsymbol{\nu}^{(k)} + t\Delta \boldsymbol{\nu}_{\mathrm{nt}}^{(k)}\right) \Big|_{t=0} = -2R\left(\boldsymbol{x}^{(k)}, \boldsymbol{\nu}^{(k)}\right) \leq 0$$

【证明见该 PPT 的附录。】

因此，我们可以使用该函数做回溯直线搜索。

不可行初始点牛顿法

Algorithm 2: 不可行初始点牛顿法

Input: 定义域内的 $\boldsymbol{x}^{(0)}$，$\boldsymbol{\nu}^{(0)}$，阈值 ϵ, η

while *True* **do**

  根据公式(4)计算原对偶牛顿方向 $\Delta\boldsymbol{x}_{\text{nt}}^{(k)}$, $\Delta\boldsymbol{\nu}_{\text{nt}}^{(k)}$

  停止准则：如果误差 $r(\boldsymbol{x}^{(k)}, \boldsymbol{\nu}^{(k)}) \leq \epsilon$，且 $\left\|\boldsymbol{A}\boldsymbol{x}^{(k)} - \boldsymbol{b}\right\| \leq \eta$，则退出

  通过回溯直线搜索选择步长 $t^{(k)} > 0$

    令 $t = 1$

    如果 $r(\boldsymbol{x}^{(k)} + t\Delta\boldsymbol{x}_{\text{nt}}^{(k)}, \boldsymbol{\nu}^{(k)} + t\Delta\boldsymbol{\nu}_{\text{nt}}^{(k)}) > (1 - \alpha t) r(\boldsymbol{x}^{(k)}, \boldsymbol{\nu}^{(k)})$，则令 $t = \beta t$

  $\boldsymbol{x}^{(k+1)} = \boldsymbol{x}^{(k)} + t^{(k)}\Delta\boldsymbol{x}_{\text{nt}}^{(k)}$

  $\boldsymbol{\nu}^{(k+1)} = \boldsymbol{\nu}^{(k)} + t^{(k)}\Delta\boldsymbol{\nu}_{\text{nt}}^{(k)}$

end

性质

假使 $\boldsymbol{x}^{(0)}$ 不满足 $\boldsymbol{A}\boldsymbol{x}^{(0)}=\boldsymbol{b}$，并且我们一直取 $t^{(k)}\leq 1$，则

$$\boldsymbol{A}\boldsymbol{x}^{(k+1)}-\boldsymbol{b}=\boldsymbol{A}\boldsymbol{x}^{(k)}+t^{(k)}\mathbf{A}\Delta\mathbf{x}^{(k)}-\boldsymbol{b}=(1-t^{(k)})\big(\boldsymbol{A}\boldsymbol{x}^{(k)}-\boldsymbol{b}\big)$$

因此，残差 $\left\|\boldsymbol{A}\boldsymbol{x}^{(k+1)}-\boldsymbol{b}\right\|=\left(\prod_{j=0}^{k}(1-t^{(j)})\right)\left\|\boldsymbol{A}\boldsymbol{x}^{(0)}-\boldsymbol{b}\right\|$ 会快速衰减。

目录

总结

主要需要掌握的知识:

- 掌握三种不同的思路来解等式约束的优化问题
- 重点掌握公式(2), (3)，(4)，即牛顿步径的定义
- 能够复述写出等式约束的牛顿法
- 能够复述写出不可行初始点的牛顿法

阅读作业 & 参考资料:

- 课本第 10 章

附录

考虑求解 $\boldsymbol{g}(\boldsymbol{y})=\boldsymbol{0}$，其中对于任意的 $\boldsymbol{y}\in\mathbb{R}^m$，$\boldsymbol{g}(\boldsymbol{y})=\begin{bmatrix} g_1(\boldsymbol{y}) \\ g_2(\boldsymbol{y}) \\ \vdots \\ g_m(\boldsymbol{y}) \end{bmatrix}\in\mathbb{R}^m$ 依旧是一个向量。通过类似前面的局部近似法，更新的步径为

$$\boldsymbol{g}(\boldsymbol{y})+\nabla\boldsymbol{g}(\boldsymbol{y})\Delta\boldsymbol{y}=\boldsymbol{0}$$

其中矩阵 $\nabla\boldsymbol{g}$ 的 (i,j) 行的元素为 $\partial_{y_j}g_i(\boldsymbol{y})$。我们可以验证

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\|\boldsymbol{g}(\boldsymbol{y}+t\Delta\boldsymbol{y})\|^2\big|_{t=0} &= \frac{\mathrm{d}}{\mathrm{d}t}\left(\sum_{i=1}^{m}\left(g_i(\boldsymbol{y}+t\Delta\boldsymbol{y})\right)^2\right)\Big|_{t=0} \\ &= \sum_{i=1}^{m}2g_i(\boldsymbol{y})\left(\nabla g_i(\boldsymbol{y})\cdot\Delta\boldsymbol{y}\right)=\sum_{i=1}^{m}-2g_i(\boldsymbol{y})g_i(\boldsymbol{y})=-2\|\boldsymbol{g}(\boldsymbol{y})\|^2\le 0\end{aligned}$$

因此，该方向是残差函数 $\|\boldsymbol{g}(\boldsymbol{y})\|^2$ 的下降方向。

之前 KKT 条件可以变成该形式的特例：$m = n + p$，

$$\boldsymbol{y} = \begin{bmatrix} \boldsymbol{x} \\ \boldsymbol{\nu} \end{bmatrix} \in \mathbb{R}^{n+p}, \qquad \begin{bmatrix} g_1(\boldsymbol{y}) \\ \vdots \\ g_p(\boldsymbol{y}) \end{bmatrix} = \boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}, \qquad \begin{bmatrix} g_{p+1}(\boldsymbol{y}) \\ \vdots \\ g_{n+p}(\boldsymbol{y}) \end{bmatrix} = \nabla f(\boldsymbol{x}) + \boldsymbol{A}^\top \boldsymbol{\nu}$$

并且，残差 $R(\boldsymbol{x}, \boldsymbol{\nu}) = \|\boldsymbol{g}(\boldsymbol{y})\|_2^2$。