凸优化 CH 1 凸集和凸函数

凸集和凸函数

- 凸集（凸组合连线）
  凸集的性质：交，加，仿射，水平集合是凸集
- 凸函数（凸组合的函数值）
  典型的凸函数：指数，负对数，最小二乘，仿射，p范数
- 凸函数的一阶判定
  - $\forall x, x_0, f(x) \geq f(x_0) + \bigtriangledown^T f(x_0)(x - x_0)$
  - 直观理解：切线总在下面
- 凸函数的二阶判定（二阶导/Hessian矩阵:对称半正定、对称正定）
  - $\forall x, \bigtriangledown^2 f(x)$对称半正定
    （正定：$\forall x, x^T A x \geq 0$，则称矩阵$A$是半正定的）
  - 直观理解：二次函数二次项系数大于0
  - 一个实例：标准型二次函数$f(x) = \frac{1}{2} x^T p x + q^T x + r$
    凹凸性：因为$\bigtriangledown^2 f(x) = p$，所以判断$p$的正定性即可
- 凸函数的性质（“xx也是凸函数”）：
  - 非负系数线性和
  - 仿射复合：$f(Ax + b)$
  - 凸集上的下确界（降维，注意形式）
  - 任意集合上的上确界（降维，注意形式）
  - 共轭必凸（造凸函数的方法）
  - 几个凸函数取max
  - Jensen：线性组合（系数和为1，保内部），引申到概率
  - 透视函数（本质是将向量进行伸缩，规范化）

优化问题的标准形式

- 优化问题=目标函数（最小化）+不等式限制（不大于0）+等式限制（等于0）
  $min\ f(x)$
  $g_i(x) \leq 0, m$个
  $h_j(x) = 0, n$个
  一定注意是$\leq$
- 有关概念
  - 可行解集X（由函数定义域和限制条件确定）
  - 局部极小解，全局极小解$\hat{x}$
  - 最优值p^*（下确界，不一定存在，一定唯一）
    最优解x^*（取最优值的解，不一定存在，不一定唯一）

凸优化问题的一般形式

针对上述的标准格式：

- $f(x)$是凸函数
- $h(x)$是仿射函数，等式条件可表述为$Ax = b$
- 可行解集是凸集（若可行解集只由不等条件确定，则可行解集是水平集合的交，必为凸集）

简单、常见的优化问题的解

- 线性规划：目标、限制为线性，有可行解，非分析，$O(n^2m)$
- 二次规划：目标为二次，限制为$Ax = b$和$x_i > 0$
- 二次规划之最小二乘（点积最小）：有可行解，分析，$O(n^2m)$
- 有限制的最小二乘问题

凸优化问题最优解的性质

- 凸优化问题：局部极小=全局极小（性质不错）
  凸优化问题利用这些性质判定解
- 无限制：（充要）梯度为0
- 等式限制：（KKT方程的解）$\bigtriangledown f_0(x^*) + A^T v^* = 0$

优化问题的等价转换

换一种表示方式->化简问题/有助于利用定式求解问题

- 上界，极小化，等效凸问题
- 减少约束个数：消除等式约束，消除线性约束
- 约束仿射化/松弛变量：简化思路，一个应用是SVM
- 变量独立化：使约束条件具有更好的结构
- 优化部分变量：可以利用确界降维优化函数，如约束条件维数低等情况
- 具体优化：最小化1范数、最小化无穷范数
- 优化表示：可行解问题、局部最优解问题

凸优化 CH 2 对偶与最优性条件

拉格朗日函数

- 拉格朗日函数：$L(x,\lambda)=f(x)+\lambda^T g(x)$
  带等式约束：$L(x,\lambda)=f(x)+\lambda^T g(x)+v^T h(x)$
- 拉格朗日对偶函数$L_D(\lambda)=\underset{x}{inf}\ L(x,\lambda)$
  带等式约束：$L_D(\lambda,v)=\underset{x}{inf}\ L(x,\lambda,v)$
- 性质
  - $L_D(x)$是凹函数
  - $\lambda\geq 0$时，对偶函数是原函数的下界
  - $L_D(\lambda)=-f^*(-\lambda)$
    （共轭函数：$f^*(\lambda)=\underset{x\in domf}{sup}(\lambda^T x-f(x))$，即f(x)与x的线性函数的最远距离）
    因为共轭函数必为凸函数，所以此性质也说明$L_D(x)$是凹函数

拉格朗日对偶问题

$$max\ L_D(\lambda,v)$$
$$\lambda\geq 0$$

弱对偶和强对偶

- 原函数：$L_P(x)=\underset{\lambda\geq 0}{sup}\ L(x,\lambda)$
  $min\ L_P(x)$和原问题等价（见纸质推导1）
- 弱对偶(极大极小不等式）：
  $$\begin{aligned}min\ L_P(x)&=\underset{x}{min}\ \underset{\lambda}{max}\ L(x,\lambda,v)\\&\geq\underset{\lambda}{max}\ \underset{x}{min}\ L(x,\lambda,v)\\&=max\ L_D\end{aligned}$$
  对偶问题的解是原问题解的下界
  启发：求解原问题时可以先求解对偶问题（凸优化问题），对偶问题的解必定不大于原问题的最优解（给出下界，保底）
- 强对偶：上述不等式等号严格成立
  一种很理想的情况，此时对偶问题的解就是原问题的解（纸质推导2）。
  解释：极大极小不等式的鞍点
  凸优化问题往往能满足强对偶条件。
- 强对偶的成立条件：Slater约束品性（内点）
  存在可行解集的内点严格满足不等式条件（严格 < ）
  解释：对应有严格凸子集的凸集

KKT方程

强对偶性成立时的充要条件：原问题的解是KKT方程的解
$g(x) \leq 0$,$h(x) = 0$（约束条件）
$\lambda \geq 0$（v没有要求）
$\lambda^T g(x) = 0$（对$g(x)$的松弛约束）
$\bigtriangledown f(x) + \lambda^T \bigtriangledown g(x) + v^T \bigtriangledown h(x) = 0$（主方程）

凸优化 CH 3 通用下降方法

准备工作

问题描述

目标函数是至少二阶连续可导的凸函数
性质：最优解充要于梯度为0，且水平集合为闭集
二阶泰勒近似：

$$f(x) = f(x_0) + \bigtriangledown^T f(x_0)(x - x_0) + \frac{1}{2}(x - x_0)^T \bigtriangledown^2 f(x)(x - x_0)$$

经常用这个式子直接相等时的情况做估计，以下的迭代求解都基于等式

强凸性

- 定义
  $\exists m > 0, s.t. \bigtriangledown^2 f(x) \geq mI, \forall x \in dom\ f$，即二阶梯度（Hessian矩阵）有正的下界
- 上部限界（经常成立）：

$$mI \leq \bigtriangledown^2 f(x) \leq MI$$

  估计m和M的好方法:矩阵的最大最小特征值
- 性质：
  - 结合二阶展开的等式使用

$$f(x) \geq f(x_0) + \bigtriangledown^T f(x_0)(x - x_0) + \frac{m}{2}||x - x_0||_2^2$$

  - 目标函数值与最优值（最小值的差）由梯度限定

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

    因此，通用下降方法可以判断梯度作为停止条件

无约束最小化方法：通用下降方法

求解无约束问题使用基于迭代的方法，取一点的序列以迫近梯度为0的点

```
find x # 取初值点
while |gradient x|>e # 有时也会采取不是梯度范数的其他停止条件
# 有时这里也用|gradient x|>2me，e对应原函数误差
   find d # 找下降方向d
   # 下降方向：梯度与方向内积<0，即局部一阶凸
   find t # 找下降步长t
   # 下降步长：只要使函数值下降即可，即f(x+td)<f(x)
   upgrade x: x=x+td
end
```

- 以下记原始误差$e_0 = e(x_0) = f(x_0) - p^*$。下文讨论收敛性时，实际讨论的$e$也是达到$f(x) - p^* < e$的情况,不是函数梯度模的$e$。算法中用梯度表示迭代停止的停止阈值对应这里的$2me$。
- 这里梯度没有要求具体是哪个范数，不过回溯下降和2范数联系紧密，这里最好采取2范数；最速下降最好采取和最速方向确定相同的范数。牛顿下降使用的是牛顿减小量。

线性下降方法介绍

确定下降方向的办法

- 梯度下降：取负梯度方向为下降方向
- 最速下降：基于方向导数，找下降速度最快的方向
  $d_{nsd} = argmin\{\bigtriangledown^T f(x) * d, s.t. ||d|| = 1\}$
  在下文最速下降部分，没有特殊声明的范数（没有下标的范数$||\cdot||$）就是梯度在此处的范数，下标为*的范数$||\cdot||_*$是其对偶范数

最速下降：一些补充

- 最速下降方向使方向导数最小（绝对值最大的负值）
- 最速下降方向是单位方向
- 非规范化最速下降：最速下降方向*梯度的对偶范数
  $d_{sd} = d_{nsd} * || \bigtriangledown f(x)||_*$
- 一些特殊情况
  - 2范数：非规范化最速下降方向就是负梯度方向
  - 1范数：等价于每步只改动梯度最大分量的维度，往下降方向（正或负）走$\frac{\partial f}{\partial xi}$（当前点的偏导数）
- 梯度和最速的重要区别：最小化的参考标准不同
  $\bigtriangledown^T f(x) * d = -||d||_2^2$
  $\bigtriangledown^T f(x) * d_{sd} = -||d_{sd}||_*^2$
  如果最速下降采用2范数，则确定的下降方向就是负梯度方向

确定下降步长的方法

- 精确直线搜索：针对步长t做单变量优化$\min_{t>0} f(x+td)$，找到精确的最小值
- 回溯直线搜索：先取步长$t=1$，然后每次乘一个公比$\beta<1$来缩
  回溯停止条件：$f(x+td)<f(x)+\alpha\bigtriangledown^T f(x)*td$，$0<\alpha<0.5$

回溯直线：一些补充

- $\bigtriangledown^T f(x)*td$对应目标函数在下降方向上的方向导数
  后文可以看到，有时方向导数的数值比较容易通过其他量计算得到，此时也会用跟其他量有关的表达式表示，但实质是一样的
- 参数α,β的作用
  - α的作用：$\bigtriangledown^T f(x)*td$是负的，而$f(x+td)\to f(x)+\bigtriangledown^T f(x)*td, t\to 0$，由此可知$\alpha$的作用是规定步长不能太小，同时也保证下降性
  - β的作用：β较大时内部迭代次数较多（用于确定步长的计算量较大），但估计得到的步长较好，外部迭代次数少；较小刚好相反。
- 实际应用时注意检查$x+td$是否在定义域内

线性下降方法的收敛性分析

通用收敛性证明

- 所谓线性下降，就是误差近似等比地减小
- 每次迭代，下降不小于$\mu||\bigtriangledown f(x)||_2^2$

$$f(x_k)-f(x_{k+1})\geq\mu||\bigtriangledown f(x_k)||_2^2$$

- 第k次迭代，误差$e_k\leq c^k e_0$

$$f(x_k)-p^*=e_k\leq c^k e_0=f(x_0)-p^*$$

  这里的c由估计值下界公式算出，$c=1-2m\mu$
- 收敛到误差e，所需步数$K\geq|\frac{log(e0/e)}{log(1/c)}|=-\frac{log(e0/e)}{log(c)}$

梯度下降+精确直线搜索

- $\mu=\frac{1}{2M}$，由强凸性易证
  $c=1-m/M$
- 收敛到误差e，所需步数$K\geq|\frac{log(e0/e)}{log(1/c)}|$
- 步长K估计值$\frac{log(e0/e)}{-log(1-m/M)}\to\frac{log(e0/e)}{m/M}$，故收敛步数大约正比于M/m，即收敛速度很强地线性依赖于M/m

梯度下降+回溯直线搜索

- $\forall t \leq \frac{1}{M}$满足回溯停止条件
  因此，每一步回溯直线搜索要么t=1，要么$t \geq \frac{\beta}{M}$
- $\mu = \alpha\ min(1, \frac{\beta}{M})$，由回溯终止条件直接指出的
  $c = 1 - 2m\alpha\ min(1, \frac{\beta}{M})$
- 步数大约正比于M/m(收敛速度很强地线性依赖于M/m)

最速下降+精确直线搜索

- 约束/参数：原范数，对偶范数至少是2范数的γ，$\tilde{\gamma}$倍
  $||\cdot|| \geq \gamma||\cdot||_2$,$||\cdot||_* \geq \tilde{\gamma}||\cdot||_2$
- $\mu = \frac{1}{2M}$，爆算+针对2范数合理放缩（见作业）
  计算结果和梯度+精确一样，过程却有所不同，很有意思
  $c = 1 - m/M$

最速下降+回溯直线搜索

- 参数：γ，$\tilde{\gamma}$
- 估计步长：$t \leq \frac{\gamma^2}{M}$时回溯停止条件满足，每步收敛步长$t \geq min(1, \beta\frac{\gamma^2}{M})$
- $\mu = \alpha\tilde{\gamma}^2 min(1, \frac{\beta\gamma^2}{M})$，步骤和梯度+回溯完全对应，针对2范数有一个系数放缩
- $c = 1 - 2m\alpha\tilde{\gamma}^2 min(1, \frac{\beta\gamma^2}{M})$

非线性下降方法（牛顿方法+回溯直线搜索）

- 下降方向
  $d_{nt} = -\bigtriangledown^2 f(x)^{-1} \bigtriangledown f(x)$
  意义：二阶泰勒展开最小值/Hessian矩阵范数
  下降性证明：因为二阶梯度是正定的，所以$\bigtriangledown^T f(x) d_{nt} = -\bigtriangledown^T f(x) \bigtriangledown^2 f(x)^{-1} \bigtriangledown f(x) < 0$
- 牛顿减小量
  $\lambda = (\bigtriangledown^T f(x) \bigtriangledown^2 f(x)^{-1} \bigtriangledown f(x))^{1/2}$
  $\lambda^2 = d_{nt}^T \bigtriangledown^2 f(x) d_{nt}$,$\lambda^2 = -\bigtriangledown^T f(x) d_{nt}$
  迭代停止：$\lambda^2 < e$
- 回溯直线搜索
  回溯停止条件：$f(x + td_{nt}) < f(x) - \alpha t\lambda^2$
  实际上$\lambda^2 = -\bigtriangledown^T f(x) d_{nt}$,停止条件和之前的回溯直线搜索是对应的。
- 收敛性分析（具体看作业）
  - 前提：F范数（方均根范数，Frobenius范数）Lipschitz连续

$$||\bigtriangledown^2 f(x) - \bigtriangledown^2 f(x_0)||_F < L||x - x_0||_2$$

    L可以解释为三阶导数的界，在L较小时牛顿法表现较好

- 阻尼牛顿：存在$0 < \eta \leq \frac{m^2}{L}$使得
  在$|| \bigtriangledown f(x)||_2 > \eta$时,每次至少下降$\gamma$
  证明思路：固定η并正常估计步长，计算每一步下降量来估计$\gamma = \alpha\beta\eta^2 \frac{m}{M^2}$
- 二次收敛：在$|| \bigtriangledown f(x)||_2 < \eta$,每一步都有$|| \bigtriangledown f(x_{k+1})||_2 < \frac{L}{2m^2}|| \bigtriangledown f(x_k)||_2^2$
  证明思路：证明步长为1（利用L，梯度差换变量形成二阶导，函数积分2次），爆算估计$\eta = min(1, 3(1-2\alpha))\frac{m^2}{L}$，顺便证明了二次收敛性

凸优化 CH 4 等式约束问题

问题描述

- $\underset{x}{min}\ f_0(x)$
  $s.t. Ax = b$
- $f_0(x)$是至少二阶连续可导的强凸函数
- 等式约束中A是p*n维矩阵（n是变量的维数）
  A是满秩的,$rank\ A = p$
- $h(x) = Ax - b$,$\bigtriangledown h(x) = A$

直接求解

- 理论上可以直接列解KKT方程来求解问题：存在最优解x^*等价于如下的方程组有解

$$\begin{cases} Ax^* = b \\ \bigtriangledown f(x^*) + {v^*}^T A = 0 \end{cases}$$

  不难发现$v^* = -(AA^T)^{-1}A \bigtriangledown f(x^*)$，这个结果后文有用
- 一个特殊情况
  对于二次函数求解，KKT方程可以写成矩阵形式：

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

  对后文的迭代求解有指导意义
- 大多数情况，KKT方程很难直接求解

破坏原问题结构求解

消除等式约束

- 对x进行仿射以降维，使p个等式约束被化解，形成的新变量为n-p维无约束变量

$$\{x : Ax = b\} = \{Fz + \hat{x} : z \in \mathbb{R}^{n-p}\}$$

  $\hat{x}$是特解，$A\hat{x} = b$
  F是A的零空间中的矩阵，n*n-p维，$AF = 0$
  注：变成无约束问题之后就没有KKT方程了，因为不再有对偶
- F的无关性
  任何一个零空间内的矩阵都可用作消除矩阵
  消除矩阵乘以一个满秩的矩阵还是消除矩阵

对偶方法求解

直接列出对偶问题并求解

$$\max_v \; -b^T v - f^*(-A^T v)$$

迭代求解

牛顿+回溯 从可行点开始

- 额外要求：从可行点$Ax_0 = b$开始；每一步下降都可行,即$Ad_{nt} = 0$
- 牛顿方向的选取：满足方程

$$\begin{bmatrix} \bigtriangledown^2 f(x) & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} d_x \\ w \end{bmatrix} = \begin{bmatrix} -\bigtriangledown f(x) \\ 0 \end{bmatrix}$$

- 讨论
  - 跟二次函数对应，可以还用之前二阶泰勒来解释
  - 可以解释成最优性条件的线性近似方程组的解
- 性质
  - 可以保住之前的一切关于收敛性的计算结果（减小量定义、阻尼、二次、各个数值）
  - 仿射不变性：$x = Ty$，则$d_{x,nt} = Td_{y,nt}$
    可以回代看看牛顿方向的定义方程是怎么变的
    这种换元可以让我们选取合适的T来简化计算
  - 从可行点开始牛顿，每步迭代过程和消除等式约束+无约束牛顿完全相同，即牛顿方向相同（乘个F），减小量相同，下一个序列点相同
    证明：根据仿射不变性，对应着算

牛顿+回溯 从不可行点开始

- 问题转化：最优化对偶残差

$$r\begin{pmatrix} x \\ v \end{pmatrix} = \begin{bmatrix} r_{dual}(y) & = & \bigtriangledown f(x) + A^T v \\ r_{pri}(y) & = & Ax - b \end{bmatrix}$$

，通过迭代使其下降到0（满足可行点的KKT）

- 牛顿方向的选取：尽快完成对偶残差的最优化，使得

$$r\begin{pmatrix} x + d_x \\ v + d_v \end{pmatrix}$$

下降最快
具体：牛顿方向应满足方程

$$\begin{bmatrix} \bigtriangledown^2 f(x) & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} d_x \\ w \end{bmatrix} = \begin{bmatrix} -\bigtriangledown f(x) \\ Ax - b \end{bmatrix}$$

理解：对应可行牛顿方向选取的线性近似理解，在迭代格式中$r_{pri}(y_k + d_{y,k})$的值应由上一次迭代的中心残差计算得到，所以约束条件的迭代格式写为$Ad_{x,k} = Ax_k - b$，其中$d_{x,k}$决定了$x_{k+1} - x_k$（的方向）

- 回溯停止条件：利用残差的2范数

$$||r(y + td_y)||_2 > (1 - \alpha t)||r||_2, y = \begin{pmatrix} x \\ v \end{pmatrix}$$

此处使用$||r||_2$是因为r在d_y方向上的方向导数为$-||r||_2$，即$t = 0$时$\frac{d}{dt}(||r(x + td_y)||_2) = -||r||_2$，使用这个量对应回溯搜索中要求的方向导数

- 可行性
    - 牛顿方向可以保证对偶残差的范数一直减小（在点处其2范数对步长t的导数是负的2范数），但是不能保证下降
      因此，残差以指数速度下降（线性收敛），迟早会降为0
      实际上残差方向一直不变，长度等比减小
    - 残差降为0的标志：步长取得1，再往下都是可行方向可行点了，已经达到误差允许的范围
      $Ax - b = 0$，与可行牛顿无异；此时应该修改牛顿方向的计算和回溯停止条件，回归正常的牛顿
      在下面的推导中没管这个误差，还是正常按设定误差算，经过两个阶段之后近似收敛到可行点，再按可行点牛顿下降来做
    - 有的时候就是不可行

- 收敛性分析
    - 要求和参数

> - r在初始点的水平集合S是凸集
> - KKT的逆矩阵$Dr(y)$有上界K，即$||Dr(y)^{-1}||_2 \le K$

> - Lipschitz连续：$||Dr(y)-Dr(y_0)||_2 \le L||y-y_0||_2$
> - 这些条件可以保证最优解和相应的对偶变量一定存在

- 取$t_{max}=inf\{t>0:r(y+td_y)=0\}$，有$||r(y+td_y)||_2 \le (1-t)||r(y)||_2+\frac{1}{2}K^2Lt^2||r(y)||_2^2, t \le t_{max}$
- 阻尼牛顿阶段

> - 回溯牛顿停止条件：$t_k \le \frac{1}{K^2L||r(yk)||_2^2}$
>   每一步的上界因梯度而异，但不妨碍导出结果
> - 下降量：$\gamma=\alpha\beta\frac{K^2}{L}$

- 二次收敛阶段：

> - 阶段的参数：$\eta=\frac{1}{K^2L}$
> - 步长：$t=1$
> - 二次收敛性：$||r_{k+1}||_2 \le \frac{K^2L}{2}||r_k||_2$

凸优化 CH 5 等式不等式约束问题

问题描述

- $\min\limits_x \ f_0(x)$
  $s.t. f_i(x) \le 0 (1 \le i \le m), Ax=b$
- $f_0(x)$是至少二阶连续可导的强凸函数
- 等式约束中A是p*n维矩阵（n是变量的维数）
  A是满秩的,$rank\ A=p$
- $h(x)=Ax-b$,$\bigtriangledown h(x)=A$

不等条件的消解

示性函数

示性函数$I_{-}(u)=\begin{cases}\infty, u>0\\0, u\le 0\end{cases}$
利用示性函数,问题转化为$\min\limits_x \ f_0(x)+\sum_{i=1}^m I_{-}(f_i(x)), s.t. Ax=b$，也是凸优化问题，也能给出同样的最优解
但我们希望目标函数性态能再好一点（可微）

对数障碍函数/近似示性函数

- 近似示性函数$\hat{I}_{-}(u) = -\frac{1}{t}log(-u)$对示性函数做了很好的模拟，t取越大越好
  使用对数障碍函数表示这个和，$\phi(x) = \sum_{i=1}^{m} -log(-f_i(x))$
  问题转化为$\min_{x}\ tf_0(x) + \phi(x), s.t. Ax = b$，也是凸优化问题，也能给出同样的最优解
- 对数障碍函数的实质是把不等约束放宽到了$f_i(x) \le 1/t$，下文的中心路径法对应KKT方程的连续变形，最终逼近最优解及其KKT
- 参数为t的近似问题，求解误差上界为m/t，即$p^*(t) - p^* \le m/t$
  （利用KKT和对偶函数性质证明）

障碍方法

- 直接求解
  因为t可以控制求解精度，所以可以直接设定t，求解对应的问题来获得对应精度的解
  问题：t较大时Hessian矩阵在可行集边界上误差较大
- 障碍方法求解
  针对一系列t解一系列问题，把每一个问题的结果作为下一个问题的初始点，直到t符合精度要求，求解结果序列称为中心路径
  t：从t_0开始，以μ为公比的等比数列
- 阶段2：从满足等式不等式条件的点出发，求解中心路径，迭代求问题的解

寻找可行点（阶段1）

障碍方法需要从一个满足等式不等式条件的点出发，需要提前找这个点
阶段1对于平凡的初始点x_0设计一个子问题（等式不等式优化问题），执行障碍方法求解以获得下一阶段的初始点；因此，问题设计的重点是初始点必须可行

满足等式约束条件的x_0

- 问题
  $\min_{s,x}\ s$
  $s.t. f_i(x) \le s, Ax = b$
- 此问题总严格可行，因为s也是优化变量，相当于不等式约束是$g_i(x, s) = f_i(x) - s \le 0$，初始值$s_0$取$max(f_i(x_0))$即可
- 迭代到$s < 0$说明有严格可行解，迭代到$s > 0$说明解不存在，但获得的解已经是这种情况下能用的最好的解了，所以也没事

不满足等式约束条件的x_0

- 对应不可行初始点的牛顿
- 先上松弛约束，把做不到的约束丢进目标函数，从而保证严格可行性，把不可行的部分变成优化目标；再通过迭代助教优化

$$\min_{z_i,s,x} t_0 f_0(x) - \sum_{i=1}^m log(s - f_i(x))$$
$$s.t. Ax = b, s = 0$$

不位于可行解集的x_0

- 对应不可行初始点的牛顿
- 思路还是为保证严格可行性而凑格式
  $$\min_{s,x} t_0 f_0(x + z_0) - \sum_{i=1}^m log(s - f_i(x + z_i))$$
  $$s.t. Ax = b, s = 0, z_i = 0 (0 \le i \le m)$$

基于原对偶残差的搜索

- 对标等式约束的原对偶残差方法
- 减小量：原对偶残差，多了一个中心残差

$$r_t(x, \lambda, v) = \begin{bmatrix} r_{dual} \\ r_{cent} \\ r_{pri} \end{bmatrix} = \begin{bmatrix} \bigtriangledown f_0(x) + Df(x)^T\lambda + A^T v \\ \lambda f(x) - 1/t \\ Ax - b \end{bmatrix}$$

- 下降方向（对应等式约束的原对偶下降方向）

$$\frac{\partial r_t(x, \lambda, v)}{\partial (x, \lambda, v)} \begin{bmatrix} d_x \\ d_\lambda \\ d_v \end{bmatrix} = -r_t(x, \lambda, v)$$

- 参数t：$t = \frac{\mu m}{\eta(x,\lambda)} = \frac{\mu m}{-Df(x)^T\lambda}$
- 各种停止的误差参数：r_{dual}和r_{pri}用一个e_{feas},$\eta(x, \lambda)$用一个e，不知道啥意思