制約つき問題の解法

山下信雄

講義内容

1. 非線形方程式，最小二乗問題の解法

2. 制約つき最小化問題の解法
    - 射影勾配法
    - 有効制約法
    - 内点法
    - 逐次２次計画法
    - 乗数法

非線形方程式と最小二乗問題

非線形方程式

$$F(x) = 0$$

ただし, $F : R^n \to R^m$

最小二乗問題

$$\min \ \ \phi(x)$$

$$\text{s.t.} \ \ x \in R^n$$

ただし, $\phi(x) = \dfrac{1}{2}\|F(x)\|^2$

非線形方程式 $F(x)=0$ の解法

ニュートン法:

次のニュートン方程式(線形方程式)の解を d^k とする.

$$F'(x^k)d+F(x^k)=0$$

$$x^{k+1}=x^k+d^k$$ とする.

- ヤコビ行列 $F'(x^k)$ が正則でないと使えない.
- 大域的収束しない.
- 収束するときは速い.
- 制約なし最小化問題のニュートン法は
  非線形方程式 $\nabla f(x)=0$ に対するニュートン法

ガウスニュートン法とLevenberg-Marqurdt法

ガウスニュートン法：

次の線形方程式の解を探索方向 d^k とする．

$$(F'(x^k))^\top F'(x^k)d + F'(x^k)^\top F(x^k) = 0$$

$$\min \frac{1}{2}\left\| F'(x^k)d + F(x^k) \right\|^2$$

Levenberg-Marquadt法：

次の線形方程式の解を探索方向 d^k とする．

$$\left\{ \underbrace{(F'(x^k))^\top F'(x^k) + \delta I}_{\text{正定値行列}} \right\} d + F'(x^k)^\top F(x^k) = 0$$

最小二乗問題とLebenberg-Marquadt法

最小二乗問題：

$$\min \ f(x) := \frac{1}{2}\|F(x)\|^2$$

ここで，$\nabla f(x) = (F'(x))^\top F(x)$ である．

LM法の探索方向 d^k は，

$$\nabla f(x^k)^\top d^k = -\nabla f(x^k)^\top \left((F'(x^k))^\top F'(x^k) + \delta I\right)^{-1} \nabla f(x^k) < 0$$

となり，最小二乗問題の降下方向となる．

⇒　大域的収束するアルゴリズム(直線探索法)が構築できる

講義内容

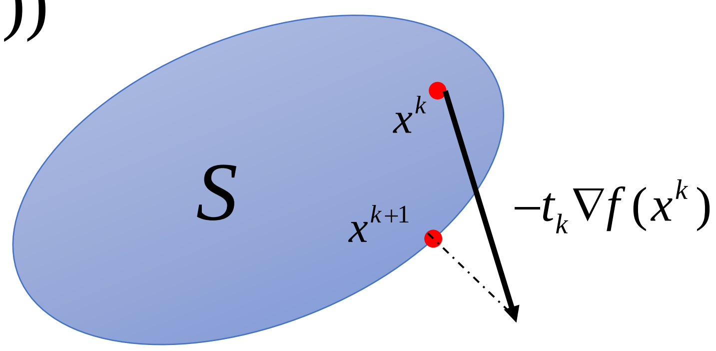
1. 非線形方程式，最小二乗問題の解法
2. 制約つき最小化問題の解法
   - 射影勾配法
   - 有効制約法
   - 内点法
   - 逐次２次計画法
   - 乗数法

射影勾配法(最急降下法の拡張)

$$\boxed{\begin{aligned}\min \quad & f(x) \\ \text{s. t.} \quad & x \in S\end{aligned}}$$

最急降下法を実施してから，実行可能領域に射影する．

$$x^{k+1} = P_S(x^k - t_k \nabla f(x^k))$$



- 大規模な問題で，S が単純な構造のときに使われる．
- 収束は遅い．S が複雑なときは射影に時間がかかる．

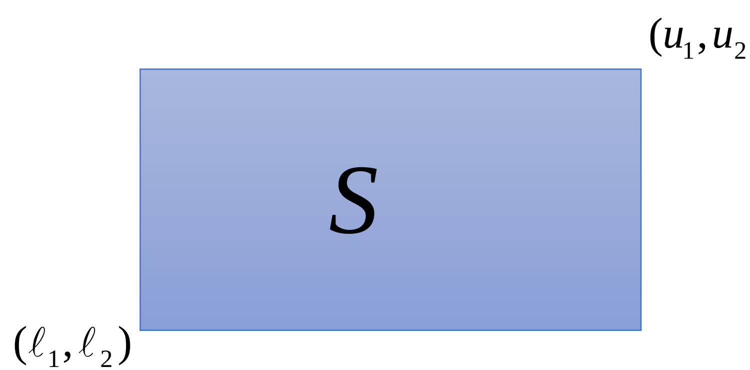
射影について

射影 $P_S(y)$　は次の最適化問題の最適解

$$
\begin{aligned}
&\min \ \frac{1}{2}\|x-y\|^2 \\
&\text{s.t.} \quad x \in S
\end{aligned}
$$

Box制約：$S=\left\{x \in R^n \mid \ell_i \le x_i \le u_i \ (i=1,\ldots,n)\right\}$

$$
\left(P_S(x)\right)_i = \begin{cases} \ell_i & \text{if } x_i \le \ell_i \\ x_i & \text{if } \ell_i < x_i < u_i \\ u_i & \text{if } u_i \le x_i \end{cases}
$$



線形等式制約 $S=\{x \mid Ax=b\}$ への射影

$$\begin{aligned}\min &\quad \frac{1}{2}\|x-y\|^2 \\ \text{s.t.} &\quad x \in S\end{aligned}$$

射影の最適化問題のKKT条件：

$$x^* - y + A^\top \mu^* = 0$$

$$Ax^* = b$$

$x^* = y - A^\top \mu^*$ より，　$Ay - AA^\top \mu^* = b$

A をフルランク行列とすると，$\mu^* = (AA^\top)^{-1}Ay - (AA^\top)^{-1}b$

よって，　$$\begin{aligned}P_S(y) = x^* &= y - A^\top (AA^\top)^{-1}Ay + A^\top (AA^\top)^{-1}b \\ &= \left(I - A^\top (AA^\top)^{-1}A\right)y + A^\top (AA^\top)^{-1}b\end{aligned}$$

補足：　$\left(I - A^\top (AA^\top)^{-1}A\right)$ を射影行列という．

射影勾配法の別バージョン

探索方向を決めてから，ステップ幅を決める．

$$d^k = P_S(x^k - \nabla f(x^k)) - x^k$$

$$x^{k+1} = x^k + t_k d^k$$

S が凸集合のとき，最適解でなければ $\nabla f(x^k)^\top d^k < 0$

- ステップ幅を求めやすい．
- 点列が実行可能領域の境界になりにくい．

講義内容

1. 非線形方程式，最小二乗問題の解法
2. 制約つき最小化問題の解法
   - 射影勾配法
   - 有効制約法
   - 内点法
   - 逐次２次計画法
   - 乗数法

制約つき問題の最適性の条件

考える問題：

$$\begin{aligned} &\min \quad f(x) \\ &\text{s.t.} \quad h_i(x) = 0,\ i = 1, \ldots, m \\ &\qquad\quad g_j(x) \le 0,\ \ j = 1, \ldots, r \end{aligned}$$

有効集合：

$$A(x) = \{ j \mid g_j(x) = 0 \}$$

KKT条件に基づく解法

$$\nabla f(x^*) + \sum_{i=1}^{m} \lambda_i^* \nabla h_i(x^*) + \sum_{j=1}^{r} \mu_j^* \nabla g_j(x^*) = 0 \quad (1)$$

$$h_i(x^*) = 0,\ i = 1, \ldots, m \quad (2)$$

$$g_j(x^*) \le 0,\ \mu_j^* \ge 0,\ \mu_j^* g_j(x^*) = 0,\ j = 1, \ldots, r \quad (3)$$

不等式制約があると，非線形方程式に対する
ニュートン法が適用できない．

近似的に等式制約のみの問題として扱う

有効制約法

最適解における有効集合$A(x^*) = \{ j \mid g_j(x^*) = 0\}$が既知のとき
元の問題は等式制約のみの問題

$$\begin{aligned} &\min\ f(x) \\ &\text{s. t.}\ \ h_i(x) = 0,\ i = 1, \ldots, m \\ &\qquad\ \ g_j(x) = 0,\ j \in A(x^*) \end{aligned}$$

と等価.
この問題のKKT条件は非線形方程式とみなせるので,
ニュートン型の手法で解ける.

有効制約法

各反復で，有効集合を推定.
ニュートン型の手法で次の反復点を求める手法.

良いところ：
- 解くべきニュートン方程式が小さくなる.
- 初期点を解のそばに選べば，すぐに解が求まる.

悪いところ：
- 有効集合の推定に時間がかかることも

代表的手法： 単体法，凸2次計画問題に対する双対法

内点法

以下では不等式制約は非負制約とする.

$$\begin{aligned} &\min\ f(x) \\ &\text{s.t.}\ \ h_i(x) = 0,\ i = 1, \ldots, m \\ &\qquad\ \ x \geq 0 \end{aligned}$$

非負制約をペナルティに持つ近似問題を考え,
等式制約のみの問題として解く. （ $\rho > 0$ ）

$$\begin{aligned} \mathrm{P}(\rho):\ &\min\ f(x) - \rho \sum_{i=1}^{n} \ln x_i \\ &\text{s.t.}\ \ h_i(x) = 0,\ i = 1, \ldots, m \end{aligned}$$

注意： $x_i > 0$ の領域しか定義できない.

対数障壁関数について

$-\rho_k \sum \ln x_i$ を対数障壁関数という.

部分問題のKKT条件

$$\mathrm{P}(\rho):\ \min\ f(x)-\rho\sum_{i=1}^{n}\ln x_i$$
$$\text{s.t.}\ \ h_i(x)=0,\ i=1,\ldots,m$$

$$\nabla f(x)+\sum_{i=1}^{m}\lambda_i\nabla h_i(x)+\rho\begin{pmatrix}-\dfrac{1}{x_1}\\ \vdots \\ -\dfrac{1}{x_n}\end{pmatrix}=0$$
$$h_i(x)=0,\ i=1,\ldots,m$$

ここで，$\mu_i=\rho\dfrac{1}{x_i}$ とおくと，

$$\nabla f(x)+\sum_{i=1}^{m}\lambda_i\nabla h_i(x)-\mu=0$$
$$h_i(x)=0,\ i=1,\ldots,m$$
$$x_j>0,\ \mu_j>0,\ x_j\mu_j=\rho$$

元の問題のKKT条件

$$\nabla f(x)+\sum_{i=1}^{m}\lambda_i\nabla h_i(x)-\mu=0$$
$$h_i(x)=0,\ i=1,\ldots,m$$
$$x_j\geq 0,\ \mu_j\geq 0,\ x_j\mu_j=0$$

内点法

- 部分問題 $\mathrm{P}(\rho_k)$ のKKT条件(非線形方程式）：

$$\nabla f(x^k) + \sum_{i=1}^{m} \lambda_i \nabla h_i(x) - \mu = 0$$
$$h_i(x) = 0, \quad i = 1, \ldots, m$$
$$x_j \mu_j = \rho_k, \quad j = 1, \ldots, n$$

に対して，ニュートン型の手法で近似解を求める．

- 反復点の情報を用いて，ρ_k を $\rho_k \to 0$ となるように更新．
- 各反復では，$x^k > 0, \mu^k > 0$ となるように点列を生成する．

性質：

- 通常，数十回の反復で解ける．
- 初期点を任意に選べない
- 解のそばで数値的に不安定になる

講義内容

1. 非線形方程式，最小二乗問題の解法
2. 制約つき最小化問題の解法
   - 射影勾配法
   - 有効制約法
   - 内点法
   - 逐次２次計画法
   - 乗数法

逐次２次計画法

制約なし最小化問題の（準）ニュートン法の拡張

部分問題

$$
\begin{aligned}
\min \quad & f(x^k) + \nabla f(x^k)^\top d + \frac{1}{2} d^\top B_k d \\
\text{s. t.} \quad & h_i(x^k) + \nabla h_i(x^k)^\top d = 0, \ \ i = 1, \ldots, m \\
& g_j(x^k) + \nabla g_j(x^k)^\top d \le 0, \ \ j = 1, \ldots, r
\end{aligned}
$$

- 目的関数は２次近似
- 制約条件は１次近似

逐次２次計画法

L1ペナルティ関数：

$$p_c(x) = f(x) + c\left\{\sum_{i=1}^{m}|h_i(x)| + \sum_{j=1}^{r}\max\{0, g_j(x)\}\right\}$$

$p_c(x^k + t_k d^k) < p_c(x^k)$ となるようにステップ幅　t_k を決め,

$$x^{k+1} = x^k + t_k d^k$$

とする.

良い点：問題固有の性質を利用した拡張がしやすい

悪い点：毎回，２次計画問題を解かなければならない

ペナルティ法

制約条件を満たさないとき大きい値を持つ関数を目的関数に加え，制約なし問題として解く方法

L1ペナルティ関数（ ⇒ 逐次２次計画法）

$$p_c(x) = f(x) + c\left\{\sum_{i=1}^{m}\left|h_i(x)\right| + \sum_{j=1}^{r}\max\left\{0, g_j(x)\right\}\right\}$$

(混合)対数障壁関数　（ ⇒ 内点法）

$$r_{c,\rho}(x) = f(x) + c\sum_{i=1}^{m}|h_i(x)| - \rho\sum_{i=1}^{n}\ln(x_i)$$

拡張ラグランジュ関数　（ ⇒ 乗数法）

乗数法(拡張ラグランジュ関数法)

拡張ラグランジュ関数(等式制約のみのとき)：

$$L_c(x,\lambda) = f(x) + \sum_{i=1}^{m} \lambda_i h_i(x) + c\sum_{i=1}^{m} h_i(x)^2 = L(x,\lambda) + c\sum_{i=1}^{m} h_i(x)^2$$

制約なし問題(λ 固定)：

$$\begin{aligned} &\min\ L_c(x,\lambda) \\ &\text{s. t.}\ \ x \in R^n \end{aligned}$$

＊ c を大きくすれば，最適解は元の問題のよい近似解.

＊ λ がKKT点のラグランジュ乗数に近いとき,
最適解は元の問題のよい近似解.

乗数法

$$L_c(x,\lambda) = f(x) + \sum_{i=1}^{m} \lambda_i h_i(x) + c\sum_{i=1}^{m} h_i(x)^2$$

1．部分問題：

$$\min L_{c_k}(x,\lambda^k)$$

$$\text{s.t.}\ \ x \in R^n$$

$$0 \approx \nabla_x L_{c_k}(x^{k+1},\lambda^k)$$

$$= \nabla f(x^{k+1}) + \sum_{i=1}^{m} \lambda_i^k \nabla h_i(x^k) + 2c_k \sum_{i=1}^{m} h_i(x^k)\nabla h_i(x^k)$$

$$= \nabla f(x^{k+1}) + \sum_{i=1}^{m} (\lambda_i^k + 2c_k h_i(x^k))\nabla h_i(x^k)$$

の近似解を x^{k+1} とする.

2．ラグランジュ乗数 λ^k を更新する

$$\lambda_i^{k+1} = \lambda_i^k + 2c_k h_i(x^{k+1}),\ \ i = 1,\ldots,m$$

3．ペナルティパラメータ c_k を更新する.

まとめ

- 大規模 $(n \geq 10000)$，制約が簡単なとき

  ⇒　射影勾配法
- 大規模，制約が複雑なとき

  ⇒　乗数法
- １回問題を解くとき

  ⇒　内点法など
- 何回も繰り返し同じような問題を解くとき

  ⇒　有効制約法，逐次２次計画法