短講

最急降下法を使ってみよう！

星 貴之

平成 18 年 11 月 28 日

1. はじめに

設計の最適化やフィッティング, ソース推定など, 最小化問題に出会うことは多い. 最小化関数が複雑で解析的に解けない（解きたくない）場合には, 反復計算により数値的に解くことになる. 最急降下法は, 最小化問題を解く最も直感的な方法である. 関数が下に凸でない場合に局所的最小解へ落ち込む可能性もあるが, 関数の形が不明なときには, とりあえず試してみるのも手かと思う.

本稿では最急降下法について簡単に説明し, サンプルプログラムを示す.

2. 最急降下法

2·1 定式化

次のような制約なし最小化問題を考える[1]. ここで $\boldsymbol{x} \equiv [x_1, x_2, \cdots, x_n]^T$ は n 次元の変数ベクトルである.

$$f(\boldsymbol{x}) \to \min. \tag{1}$$

最急降下法では, ある初期値から関数 f を最もよく減少させる方向（最急降下方向）へ進んでいくことで最小化問題の解を得ようとする. ちょうど山の斜面にボールを転がすイメージである.

最急降下方向ベクトル $\boldsymbol{d}$ は, 関数 f のグラディエントをとって式 (2) のように書かれる. ここで $f_x \equiv \frac{\partial f}{\partial x}$.

$$\boldsymbol{d} = -\nabla f(\boldsymbol{x}) \equiv - \begin{bmatrix} f_{x_1}(\boldsymbol{x}) \\ f_{x_2}(\boldsymbol{x}) \\ \vdots \\ f_{x_n}(\boldsymbol{x}) \end{bmatrix} \tag{2}$$

これを用いて式 (3) の漸化式に従って変数を更新していけば（局所的）最小解 $\boldsymbol{x}^*$ に到達できる. ここで i は反復回数, α^i は重みを表す.

$$\boldsymbol{x}^{i+1} = \boldsymbol{x}^i + \alpha^i \boldsymbol{d}^i \tag{3}$$

$\boldsymbol{d}^i \approx \boldsymbol{0}$ となった時点で更新を終了する.

2·2 使用上の注意

1. 得られる $\boldsymbol{x}^*$ は局所的最小解ではあるが, 大域的最小解である保証はない. もしも f が下に凸ならば大域的最小解であることが保証される. そうでない場合は, 複数の初期値から試してその中で f が最小となる解を最小解とみなす, $f(\boldsymbol{x}^*)$ が満足できるほど小さいので納得する, などの判断を行う.
2. 初期値 $\boldsymbol{x}^0$ の選び方は収束の速さに利いてくる. とりあえず $\boldsymbol{0}$ とおく手もあるが, 先見情報が使えるなら解に近い値を選ぶとよい. 局所的最小解に落ちることを予防する効果も期待できる.
3. 重み α^i の選び方も収束の速さや解の精度に利いてくる. 最急降下法では経験的に決めた正の定数を用いる. 値が大きすぎると解の周辺で振動する.
4. 反復計算の終了判定も収束の速さや解の精度に利いてくる. $\boldsymbol{d}^i \approx \boldsymbol{0}$ を判定するスレッショルドを決めるには多少の試行錯誤を必要とする.
5. プログラミングの際には, $\nabla f(\boldsymbol{x})$ をあらかじめ手計算しておく必要がある. もしくは $f(\boldsymbol{x})$ の各変数についての差分を近似式として使う. 計算ミスやタイプミスが致命傷になるので気をつける.

[1] 等式制約がある場合には “ラグランジュの未定乗数法”, 不等式制約がある場合には “クーン・タッカーの条件” を用いて解くらしい.

3. 最小化問題の例

3·1 最小二乗法（サンプルプログラム）※

平面上の二つの三角形 A$\{\boldsymbol{a}_1, \boldsymbol{a}_2, \boldsymbol{a}_3\}$, B$\{\boldsymbol{b}_1, \boldsymbol{b}_2, \boldsymbol{b}_3\}$ を比較したい. B を $\boldsymbol{r} = [r_x, r_y]^T$ だけ平行移動させ, θ [rad] だけ回転させた三角形を B' とする. A と B' の各頂点位置の差の二乗和を最小化することで “なるべくぴったり” 重ね合わせる.

$\boldsymbol{b}_k = [b_{kx}, b_{ky}]^T$ $(k = 1, 2, 3)$ とすると, 前述の操作は次のように書ける.

$$\begin{bmatrix} b'_{kx} \\ b'_{ky} \end{bmatrix} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} b_{kx} + r_x \\ b_{ky} + r_y \end{bmatrix} \tag{4}$$

A と B' の差の二乗和 $f(\boldsymbol{r}, \theta)$ は次のようになる.

$$f(\boldsymbol{r}, \theta) = \sum_{k=1}^{3} \left\{ (b'_{kx} - a_{kx})^2 + (b'_{ky} - a_{ky})^2 \right\} \tag{5}$$

あとは $f(\boldsymbol{r}, \theta)$ を最小化するだけである. $(f_{\boldsymbol{r}}, f_\theta)$ を書き下すか差分近似して, 最急降下法を適用する.

※ http://star.web.nitech.ac.jp/pdf/061128short.zip

3·2 非線形方程式

一般の非線形方程式 $g(\boldsymbol{x}) = h(\boldsymbol{x})$ も, $f(\boldsymbol{x}) \equiv \{g(\boldsymbol{x}) - h(\boldsymbol{x})\}^2$ とおくことで最小化問題に帰着できる. 最急降下法などで解 $\boldsymbol{x}^*$ が得られたとき, $f(\boldsymbol{x}^*) = 0$ が成り立っていれば, それが最初の非線形方程式の解である.

4. おわりに

最急降下法は直感的ではあるが, 収束が遅いのが欠点である（一次収束）. それを改善するため, 効率のよい α^i を決定する直線探索やニュートン・ラプソン法などが考案されてきた. また $\boldsymbol{d}^i$ の決め方についても, Powell の方法や共役勾配法などがある.