最急降下法の加速法

☆この資料は，期末試験の範囲外☆

---

1. 反復法の収束率

反復法が生成する点列 $\{\boldsymbol{x}_k\}$ がある点 $\bar{\boldsymbol{x}}$ に収束するとき，その収束の速さを収束率とよび，次のように定義する．

ある定数 $c \in (0,1)$ と整数 k' が存在して条件

$$\frac{\|\boldsymbol{x}_{k+1} - \bar{\boldsymbol{x}}\|}{\|\boldsymbol{x}_k - \bar{\boldsymbol{x}}\|} \leq c, \quad \forall k \geq k'$$

が成り立つとき，その反復法の収束率は 1 次である（あるいは，その反復法は 1 次収束する）という．また，条件

$$\lim_{k \to +\infty} \frac{\|\boldsymbol{x}_{k+1} - \bar{\boldsymbol{x}}\|}{\|\boldsymbol{x}_k - \bar{\boldsymbol{x}}\|} = 0$$

が成り立つとき，収束率は超 1 次であるという．さらに，ある定数 $c > 0$ と整数 k' が存在して条件

$$\|\boldsymbol{x}_k - \bar{\boldsymbol{x}}\|^2 \quad \leq c, \quad \forall \quad \geq \quad '$$

が成り立つとき，収束率は 2 次であるという．

超 1 次収束や 2 次収束は，速い収束である．これに比べて，1 次収束は遅い（つまり，収束率が 1 次の反復法は，終了までに多くの反復回数を必要とする）ため，実用的とは言えない．

無制約最適化の解法としては，適当な仮定の下で，ニュートン法の収束率は 2 次であることを示すことができる．また，準ニュートン法の収束率は超 1 次である．これに対して，最急降下法の収束率は 1 次であり，解を得るまでに多くの反復を必要とするため，実用的とは言えない．そこで，以下で述べるような改良が考えられており，大規模な問題（つまり，決定変数の数が多い問題）に対して特に有用とされている．

2. 共役勾配法

正定値対称行列 $A \in \mathbb{R}^{n \times n}$ を係数にもつ連立 1 次方程式

$$A\boldsymbol{x} = \boldsymbol{b}$$

を解く手法に，共役勾配法とよばれる方法がある．一方，この連立 1 次方程式を解くことと，最適化問題

$$\text{Minimize} \quad \frac{1}{2}\boldsymbol{x}^\top A\boldsymbol{x} - \boldsymbol{b}^\top \boldsymbol{x}$$

を解くこととは等価である．この対応に基づいて，共役勾配法は一般の無制約最適化問題

$$\text{Minimize} \quad f(\boldsymbol{x})$$

に対しても拡張されている．この最適化手法は，元の連立 1 次方程式の解法と区別するために，非線形共役勾配法とよばれることもある．

共役勾配法の概要を，アルゴリズム 1 に示す．最急降下法との違いは，アルゴリズム 1 の 2 行目で，前の反復で用いた探索方向 $\boldsymbol{d}_{k-1}$ の情報も使って現在の探索方向 $\boldsymbol{d}_k$ を定めていることにある．係数 β_{k+1} の決め方には，ここに示した以外に，さまざまな提案がある（文献 [1] や文献 [5, § 4.6] を参照されたい）．

アルゴリズム 1 （共役勾配法）

Require: $\boldsymbol{x}_0 \in \mathbb{R}^n$, $\boldsymbol{d}_0 = \boldsymbol{0}$, $\beta_0 = 0$.

1: **for** $k = 0, 1, 2, \ldots$ **do**
2: $\quad \boldsymbol{d}_k \leftarrow -\nabla f(\boldsymbol{x}_k) + \beta_k \boldsymbol{d}_{k-1}$.
3: $\quad$ 直線探索によりステップ幅 $\alpha_k > 0$ を定める．
4: $\quad \boldsymbol{x}_{k+1} \leftarrow \boldsymbol{x}_k + \alpha_k \boldsymbol{d}_k$.
5: $\quad \beta_{k+1} \leftarrow \dfrac{\|\nabla f(\boldsymbol{x}_{k+1})\|^2}{\|\nabla f(\boldsymbol{x}_k)\|^2}$.
6: **end for**

---

3. 加速法

アルゴリズム 1（共役勾配法）において $\boldsymbol{x}_k = \boldsymbol{x}_{k-1} + \alpha_{k-1}\boldsymbol{d}_{k-1}$ という関係を用いると，2 行目は

$$\boldsymbol{d}_k \leftarrow -\nabla f(\boldsymbol{x}_k) + \frac{\beta_k}{\alpha_{k-1}}(\boldsymbol{x}_k - \boldsymbol{x}_{k-1})$$

と書き直せる．つまり，共役勾配法は一つ前の反復の点 $\boldsymbol{x}_{k-1}$ を利用しているとみることもできる．これと同じ様に，過去の点列の情報を利用して勾配法の収束を速める技法として，ネステロフ (Nesterov) の加速法 *1 とよばれるものがある．

アルゴリズム 2 は，ネステロフの加速法を最急降下法に適用したものである．ここで，7 行目に注目すると，補助的な点列 $\{\boldsymbol{y}_k\}$ を用いることで点列 $\{\boldsymbol{x}_k\}$ の過去の情報を取り込んでいることがわかる．目的関数の凸性や微分可能性などを仮定すると，アルゴリズム 2 は 2 次収束することが知られている [4].

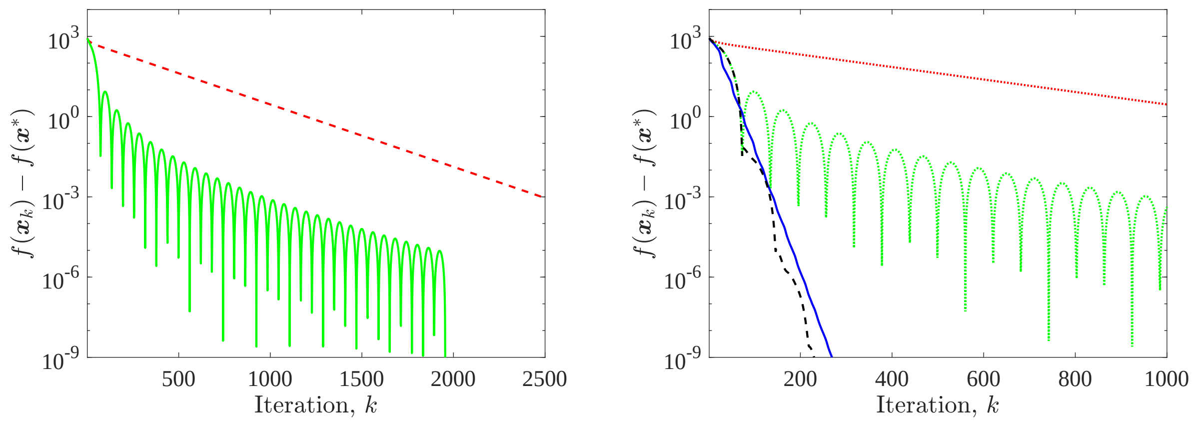
アルゴリズム 2 （ネステロフの加速付き最急降下法）

Require: $\boldsymbol{x}_0 \in \mathbb{R}^n$, $\tau_0 = 0$.

1: $\boldsymbol{y}_0 \leftarrow \boldsymbol{x}_0$.
2: **for** $k = 0, 1, 2, \ldots$ **do**
3: $\quad \boldsymbol{d}_k \leftarrow -\nabla f(\boldsymbol{y}_k)$.
4: $\quad$ ステップ幅 $\alpha_k > 0$ を定める．
5: $\quad \boldsymbol{x}_{k+1} \leftarrow \boldsymbol{y}_k + \alpha_k \boldsymbol{d}_k$.
6: $\quad \tau_{k+1} \leftarrow \dfrac{1}{2}\left(1 + \sqrt{1 + 4\tau_k^2}\right)$.
7: $\quad \boldsymbol{y}_{k+1} \leftarrow \boldsymbol{x}_{k+1} + \dfrac{\tau_k - 1}{\tau_{k+1}}(\boldsymbol{x}_{k+1} - \boldsymbol{x}_k)$.
8: **end for**

---

*1 Y. Nesterov が文献 [2] で初めて提案した手法であるので，こうよばれている．



(a) (b)

図 1: 最急降下法とその加速．(a) 破線は最急降下法，実線は加速付き最急降下法，(b) 破線は再スタート付きの加速付き最急降下法，実線は共役勾配法

ネステロフの加速法は，元のアルゴリズムにわずかな修正を加えるだけで，解の近傍での収束率を大きく改善するものである．しかし，最急降下法では目的関数値が反復ごとに単調に減少するのに対して，ネステロフの加速法を用いるとそのような単調性は一般に失われてしまう．この弱点を克服する方法として，加速の再スタート法が提案されている [3]．これは，目的関数値が増加し始めた時点で，アルゴリズム 2 の $\boldsymbol{y}_{k+1}$ と τ_{k+1} を初期化する（つまり，$\boldsymbol{y}_{k+1} \leftarrow \mathbf{0}$, $\tau_{k+1} \leftarrow 0$ とする）というものである．

図 1 は，最急降下法，共役勾配法，加速付き最急降下法の収束の例を示している．ここで，横軸は反復回数を表し，縦軸は目的関数値と最適値の差を表している．図 1(a) において，最急降下法（破線）は 2500 回の反復の後も収束していない．これに対して，加速付き最急降下法（実線）はこの例では 2000 回程度で収束しており，解の近傍では収束が速いことも観察できる．しかし，目的関数値が増加した反復も多く存在する．これに再スタート法を組み込むと，図 1(b) の破線のように，この例では 200 回程度で収束する．また，共役勾配法も図 1(b) の実線のように，少ない反復回数で収束している．

参考文献

[1] 成島 康史：無制約最適化問題に対するアルゴリズムの最前線—非線形共役勾配法を中心に—．オペレーションズ・リサーチ，**59**, pp. 131–137 (2014).

[2] Y. Nesterov: A method of solving a convex programming problem with convergence rate $O(1/k^2)$. *Soviet Mathematics Doklady*, **27**, 372–376 (1983).

[3] B. O'Donoghue, E. Candès: Adaptive restart for accelerated gradient schemes. *Foundations of Computational Mathematics*, **15**, 715–732 (2015).

[4] 鈴木 大慈：『確率的最適化』．講談社 (2015).

[5] 矢部 博：『工学基礎・最適化とその応用』，数理工学社 (2006).

（以上）