劣勾配法の話

松井 知己

1 劣勾配

本稿では、微分不可能な連続凸関数を閉凸領域で最小化する際に用いる劣勾配法についてまとめてみる。劣勾配法は、区分線形な連続凸関数の最小化等に良く使われる。また特に離散最適化問題のラグランジュ双対問題を解く際に、実際に使われている。劣勾配法の収束について書かれた本は非常に少なく、結果のみが載っている場合が多い。本稿では、収束性についてきちんと証明をしてみた。

まず最初に勾配の概念を拡張した劣勾配を定義する。集合 $\Omega \subseteq R^k$ 上で定義された関数 $g : \Omega \to R$ に対し、ベクトル $\boldsymbol{d} \in R^k$ が点 $\boldsymbol{x}_0 \in \Omega$ での g に対する劣勾配であるとは、任意の $\boldsymbol{x} \in \Omega$ に対し $g(\boldsymbol{x}_0) + \boldsymbol{d}^t(\boldsymbol{x} - \boldsymbol{x}_0) \leq g(\boldsymbol{x})$ が成り立つことである。また、 $\boldsymbol{x}_0$ における g の劣勾配の集合を $\partial g(\boldsymbol{x}_0)$ と書き、劣微分と呼ぶ。集合 $\Omega' \subseteq \Omega$ 中の任意の点 $\boldsymbol{x} \in \Omega'$ において、 $\partial g(\boldsymbol{x}) \neq \emptyset$ のとき、 g は Ω' 上で劣微分可能と言う。

容易に分かるように、関数 g が連続凸関数であれば、任意の点において劣勾配は 1 つ以上存在する。劣勾配方向は、その方向と逆に進むことによって関数値が下がることが期待される方向であるが、下がることが保証されている訳ではない。

2 劣勾配法

以下では、劣勾配を用いて一般的な最適化問題

$$\inf\{g(\boldsymbol{y}) \mid \boldsymbol{y} \in \Omega\}$$

を解く算法について議論する。ただし、集合 $\Omega \subseteq R^k$ は非空な閉凸集合とし、関数 $g : \Omega \to R$ は Ω 上で劣微分可能と仮定する。上記の最適化問題において $\inf$ という記号を用いているのは、許容解集合 Ω の有界性を仮定していないために、最適解が存在していない可能性があるためである。

任意のベクトル $\boldsymbol{y} \in R^k$ に対し、 Ω 中の点で $\boldsymbol{y}$ からの最も距離の近い点を $\mathrm{P}_\Omega(\boldsymbol{y})$ で表す。集合 $\Omega \subseteq R^k$ は閉凸集合であることから、点 $\mathrm{P}_\Omega(\boldsymbol{y})$ は必ず丁度一つ存在する。

劣勾配法

Step 0: 初期ベクトル $\boldsymbol{y}_1 \in \Omega$ を適当なものに固定する。 $i = 1$ とする。

Step 1: 劣勾配 $\boldsymbol{d}_i \in \partial g(\boldsymbol{y}_i)$ を求め、 $\hat{\boldsymbol{y}}_i := \mathrm{P}_\Omega(\boldsymbol{y}_i - s_i \boldsymbol{d}_i)$ を計算する。

Step 2: $\boldsymbol{y}_i = \hat{\boldsymbol{y}}_i$ ならば、 $\boldsymbol{y}_i$ を出力して終了する。

Step 3: $\boldsymbol{y}_i \neq \hat{\boldsymbol{y}}_i$ ならば、 $\boldsymbol{y}_{i+1} := \hat{\boldsymbol{y}}_i$ 、 $i := i + 1$ と更新し、 Step 1 に戻る。

上記のアルゴリズムは第 i 反復でのステップ長を表す s_i の設定によって、その動きは異なったものとなる。上記の劣勾配法は、必ずしも有限回の反復で終了するとは限らないため、許容解の無限列を生成することもあるが、現実にはある程度の回数で打ち切ることによって有限時間で終了させる事が多い。

数列 $\{s_i\}$ の設定について提案されているものはいくつかあるが、 代表的なものとしては以下が挙げられる。

1. 等比数列 $s_{i+1} = r s_i$ 、ただし $0 < r < 1,\ s_1 > 0$ である。

2. 性質 $s_i > 0$ と $\lim_{i\to\infty} s_i = 0$ と $\lim_{i\to\infty} \sum_{j=1}^{i} s_j = +\infty$ を満たすもの。

3. 数列 $s_i = (g(\boldsymbol{y}_i) - \underline{z})/||\boldsymbol{d}_i||^2$ 、 ただし $\underline{z}$ は最適値の下界である。

4. 数列 $s_i = 2^{2-i}(g(\boldsymbol{y}_i) - \underline{z})/||\boldsymbol{d}_i||^2$ 、 ただし $\underline{z}$ は最適値の下界である。

等比数列の収束性については、理論的に保証される性質はあまりない。また、最適解がある問題において初期点が最適解より非常に遠いと、最適解に到達することができずに収束してしまうことがある。しかしながら、実用上は非常に速く収束することが多い。2 つ目の数列の具体例としては、 $s_i = 1/i$ という数列がある。2 つ目の性質を満たす数列については、(最適解が存在するならば) 最適解に収束することが知られているが、収束が非常に遅いことが経験的に分かっており、あまり使われないようである。3 つ目の数列は、良い下界が求まっているならば、 その挙動は Newton 法に似ており、 最適解付近へ到達するまでの収束が速い。しかしながら、最適解付近での挙動が不安定となることがある。4 つ目の数列は、1 つ目と 3 つ目の数列の組合せとも言うべきものであり、 実用上最もよく使われるものであるが、 理論的に良い性質はあまりない。

3 最適性と収束性

以下では、理論的な結果についてまとめる。最初に後の証明で用いる補題を示す。

補題 **3.1**:

$$\forall \boldsymbol{y} \in \Omega,\ \forall \boldsymbol{x} \in R^k,\ \ (\boldsymbol{x} - \mathrm{P}_\Omega(\boldsymbol{x}))^t(\boldsymbol{y} - \mathrm{P}_\Omega(\boldsymbol{x})) \le 0$$

証明 ベクトル $\boldsymbol{y}(\varepsilon) = \mathrm{P}_\Omega(\boldsymbol{x}) + \varepsilon(\boldsymbol{y} - \mathrm{P}_\Omega(\boldsymbol{x}))$ とすると、 Ω の凸性より、任意の $1 \ge \varepsilon \ge 0$ に対し $\boldsymbol{y}(\varepsilon) \in \Omega$ である。 $\mathrm{P}_\Omega(\boldsymbol{x})$ の定義より、 $||\mathrm{P}_\Omega(\boldsymbol{x}) - \boldsymbol{x}||^2 \le ||\boldsymbol{y}(\varepsilon) - \boldsymbol{x}||^2$ が成り立つ。これより、

$$\begin{aligned} ||\mathrm{P}_\Omega(\boldsymbol{x}) - \boldsymbol{x}||^2 &\le ||\boldsymbol{y}(\varepsilon) - \boldsymbol{x}||^2 = ||\mathrm{P}_\Omega(\boldsymbol{x}) + \varepsilon(\boldsymbol{y} - \mathrm{P}_\Omega(\boldsymbol{x})) - \boldsymbol{x}||^2 \\ &= ||\mathrm{P}_\Omega(\boldsymbol{x}) - \boldsymbol{x}||^2 + 2\varepsilon(\mathrm{P}_\Omega(\boldsymbol{x}) - \boldsymbol{x})^t(\boldsymbol{y} - \mathrm{P}_\Omega(\boldsymbol{x})) + \varepsilon^2||\boldsymbol{y} - \mathrm{P}_\Omega(\boldsymbol{x})||^2 \end{aligned}$$

である。上式を変形して $\varepsilon^2||\boldsymbol{y} - \mathrm{P}_\Omega(\boldsymbol{x})||^2 + 2\varepsilon(\mathrm{P}_\Omega(\boldsymbol{x}) - \boldsymbol{x})^t(\boldsymbol{y} - \mathrm{P}_\Omega(\boldsymbol{x})) \ge 0$ が得られるが、ε が十分小さな正の数のときもこれが成り立つには、 $(\mathrm{P}_\Omega(\boldsymbol{x}) - \boldsymbol{x})^t(\boldsymbol{y} - \mathrm{P}_\Omega(\boldsymbol{x})) \ge 0$ でなければならない。ゆえに補題の性質が満たされる。 //

劣勾配法が (有限回の反復で) 終了したならば、以下の性質が成り立つ。

定理 **3.2**: 関数 g が Ω 上で凸であり、ステップ長列 $\{s_i\}$ が $s_i > 0$ を満たすならば、劣勾配法が有限回の反復で終了した際に求められた解は最適解である。

証明 劣勾配法が有限回の反復で終了し、 $\boldsymbol{y}^*$ を出力したとする。劣勾配法の終了判定に用いられた $\boldsymbol{y}^*$ での劣勾配を $\boldsymbol{d}$ とし、終了時点での反復でのステップ長を $s > 0$ とする。

$\bar{\boldsymbol{y}}$ を Ω 中の点とすると、上記補題より

$$((\boldsymbol{y}^* - s\boldsymbol{d}) - \mathrm{P}_\Omega(\boldsymbol{y}^* - s\boldsymbol{d}))^t(\bar{\boldsymbol{y}} - \mathrm{P}_\Omega(\boldsymbol{y}^* - s\boldsymbol{d})) \le 0$$

が成り立つ。劣勾配法の終了条件より $\mathrm{P}_\Omega(\boldsymbol{y}^* - s\boldsymbol{d}) = \boldsymbol{y}^*$ であるから、 $\forall \bar{\boldsymbol{y}} \in \Omega$ について

$$\begin{aligned} (\boldsymbol{y}^* - s\boldsymbol{d} - \boldsymbol{y}^*)^t(\bar{\boldsymbol{y}} - \boldsymbol{y}^*) &\le 0 \\ \boldsymbol{d}^t(\bar{\boldsymbol{y}} - \boldsymbol{y}^*) &\ge 0 \end{aligned}$$

となる。ベクトル $\boldsymbol{d}$ が $\boldsymbol{y}^*$ における劣勾配であることから、

$$\forall \bar{\boldsymbol{y}} \in \Omega,\ \ g(\bar{\boldsymbol{y}}) \geq g(\boldsymbol{y}^*) + \boldsymbol{d}^t(\bar{\boldsymbol{y}} - \boldsymbol{y}^*) \geq g(\boldsymbol{y}^*)$$

が成り立ち、 $\boldsymbol{y}^*$ が最適解であることが示された。 //

最適解での劣勾配が唯一でない時は、劣勾配法のある反復で最適解が見つかっていながらも、最適解であることを認識できずに、劣勾配法が停止しない可能性が存在する。

2 つ目の数列については、以下の性質が成り立つ。

定理 **3.3**: 関数 $g : \Omega \to R$ は凸関数で、 Ω 上で劣微分可能とする。最適化問題 $\inf\{g(\boldsymbol{x}) \mid \boldsymbol{x} \in \Omega\}$ に劣勾配法を適用した際、第 i 反復で得られた許容解と劣勾配を $\boldsymbol{y}_i$, $\boldsymbol{d}_i$ と書く。ある定数 C が存在して、 $\forall i > 0,\ ||\boldsymbol{d}_i||^2 < C$ と仮定する。劣勾配法において使用する数列 $\{s_i\}$ が $s_i > 0$ と $\lim_{i\to\infty} s_i = 0$ と $\lim_{i\to\infty} \sum_{j=1}^{i} s_j = +\infty$ を満たすならば、

$$\forall z > \inf\{g(\boldsymbol{y}) \mid \boldsymbol{y} \in \Omega\},\ \ \exists i > 0,\ \ g(\boldsymbol{y}_i) < z$$

が成り立つ。

証明　劣勾配法が有限回の反復で終了したならば、上記の定理から得られた解は最適解であり、定理は成り立つ。ゆえに、劣勾配法が有限回の反復で終了しなかったと仮定して証明を行なう。

ある $z_0 > \inf\{g(\boldsymbol{y}) \mid \boldsymbol{y} \in \Omega\}$ が存在して、 $\forall i > 0,\ g(\boldsymbol{y}^i) \geq z_0$ であったとして矛盾を導く。ここで $z_0 > \inf\{g(\boldsymbol{y}) \mid \boldsymbol{y} \in \Omega\}$ より、 $\{\boldsymbol{y} \in \Omega \mid g(\boldsymbol{y}) < z_0\} \neq \emptyset$ となりベクトル $\bar{\boldsymbol{y}} \in \{\boldsymbol{y} \in \Omega \mid g(\boldsymbol{y}) < z_0\}$ を適当に選ぶことができる。

任意のベクトル $\boldsymbol{x} \in R^k$ に対し、 $\bar{\boldsymbol{y}} \in \Omega$ と上記補題より、

$$\begin{aligned} ||\boldsymbol{x} - \bar{\boldsymbol{y}}||^2 &= ||\boldsymbol{x} - \mathrm{P}_\Omega(\boldsymbol{x}) + \mathrm{P}_\Omega(\boldsymbol{x}) - \bar{\boldsymbol{y}}||^2 \\ &= ||\boldsymbol{x} - \mathrm{P}_\Omega(\boldsymbol{x})||^2 + 2(\boldsymbol{x} - \mathrm{P}_\Omega(\boldsymbol{x}))(\mathrm{P}_\Omega(\boldsymbol{x}) - \bar{\boldsymbol{y}}) + ||\mathrm{P}_\Omega(\boldsymbol{x}) - \bar{\boldsymbol{y}}||^2 \\ &\geq ||\boldsymbol{x} - \mathrm{P}_\Omega(\boldsymbol{x})||^2 + ||\mathrm{P}_\Omega(\boldsymbol{x}) - \bar{\boldsymbol{y}}||^2 \geq ||\mathrm{P}_\Omega(\boldsymbol{x}) - \bar{\boldsymbol{y}}||^2 \end{aligned}$$

が成り立つ。これより、

$$\begin{aligned} ||\boldsymbol{y}_{i+1} - \bar{\boldsymbol{y}}||^2 &= ||\mathrm{P}_\Omega(\boldsymbol{y}_i - s_i\boldsymbol{d}_i) - \bar{\boldsymbol{y}}||^2 \leq ||\boldsymbol{y}_i - s_i\boldsymbol{d}_i - \bar{\boldsymbol{y}}||^2 \\ &= ||\boldsymbol{y}_i - \bar{\boldsymbol{y}}||^2 - 2s_i(\boldsymbol{d}_i)^t(\boldsymbol{y}_i - \bar{\boldsymbol{y}}) + (s_i)^2||\boldsymbol{d}_i||^2 \\ &\leq ||\boldsymbol{y}_i - \bar{\boldsymbol{y}}||^2 - 2s_i(g(\boldsymbol{y}_i) - g(\bar{\boldsymbol{y}})) + (s_i)^2C \\ &\leq ||\boldsymbol{y}_i - \bar{\boldsymbol{y}}||^2 - 2s_i(z_0 - g(\bar{\boldsymbol{y}})) + (s_i)^2C \end{aligned}$$

が成り立つ。数列 $\{s_i\}$ が $s_i > 0$ と $\lim_{i\to\infty} s_i = 0$ を満たすことから、(十分大きな) 数字 i_0 が存在して、 $\forall i \geq i_0$ に対し $-s_i(z_0 - g(\bar{\boldsymbol{y}})) + (s_i)^2C \leq 0$ が成り立ち、ゆえに

$$\forall i \geq i_0,\ \ ||\boldsymbol{y}_{i+1} - \bar{\boldsymbol{y}}||^2 \leq ||\boldsymbol{y}_i - \bar{\boldsymbol{y}}||^2 - s_i(z_0 - g(\bar{\boldsymbol{y}}))$$

となる。数列 $\{s_i\}$ が $\lim_{i\to\infty} \sum_{j=1}^{i} s_j = +\infty$ を満たすことと、 $z_0 - g(\bar{\boldsymbol{y}}) > 0$ から、(十分大きな) 数字 $i_1 \geq i_0$ が存在して、

$$||\boldsymbol{y}_{i_0} - \bar{\boldsymbol{y}}||^2 < \sum_{i=i_0}^{i_1} s_i(z_0 - g(\bar{\boldsymbol{y}}))$$

が成り立つ。以上より、

$$||\boldsymbol{y}_{i_1+1} - \bar{\boldsymbol{y}}||^2 \leq ||\boldsymbol{y}_{i_0} - \bar{\boldsymbol{y}}||^2 - \sum_{i=i_0}^{i_1} s_i(z_0 - g(\bar{\boldsymbol{y}})) < 0$$

となり矛盾が導かれる。 //

上記の定理での仮定 $\lim_{i\to\infty} s_i = 0$ は、許容解列 $\{\boldsymbol{y}_i\}$ が巡回しないためには当然必要を思われる仮定である。また $\lim_{i\to\infty} \sum_{j=1}^{i} s_j = +\infty$ という仮定は、最適解が存在した時にどんな遠い初期点からもそこにたどり着くには、必要な仮定である。

3 つ目の数列については、以下の性質が成り立つ。

定理 **3.4**: 関数 $g : \Omega \to R$ は凸関数で、 Ω 上で劣微分可能とする。最適化問題 $\inf\{g(\boldsymbol{x}) \mid \boldsymbol{x} \in \Omega\}$ に劣勾配法を適用した際、第 i 反復で得られた許容解と劣勾配を $\boldsymbol{y}_i$, $\boldsymbol{d}_i$ と書く。ある定数 C が存在して、 $\forall i > 0,\ ||\boldsymbol{d}_i||^2 < C$ と仮定する。劣勾配法において使用する数列 $\{s_i\}$ が $s_i = (g(\boldsymbol{y}_i) - \underline{z})/||\boldsymbol{d}_i||^2$ を満たすならば、

$$\forall z > z^* + (z^* - \underline{z}),\ \exists i > 0,\ g(\boldsymbol{y}_i) < z$$

が成り立つ。ただし $z^* = \inf\{g(\boldsymbol{y}) \mid \boldsymbol{y} \in \Omega\}$ である。

証明　上記の定理と同様の理由で、、劣勾配法が有限回の反復で終了しなかったときのみ示せば良い。

ある $\varepsilon > 0$ が存在して、 $\forall i,\ g(\boldsymbol{y}_i) \geq 2z^* - \underline{z} + \varepsilon$ と仮定して矛盾を導く。ベクトル $\bar{\boldsymbol{y}} \in \{\boldsymbol{y} \in \Omega | g(\boldsymbol{y}) < z^* + (1/3)\varepsilon\}$ とする。上記の定理の証明と同様に、

$$||\boldsymbol{y}_{i+1} - \bar{\boldsymbol{y}}||^2 \leq ||\boldsymbol{y}_i - \bar{\boldsymbol{y}}||^2 - 2s_i(g(\boldsymbol{y}_i) - g(\bar{\boldsymbol{y}})) + (s_i)^2||\boldsymbol{d}_i||^2$$

が成り立つ。 $\{s_i\}$ の定義と $s_i > 0,\ \ g(\bar{\boldsymbol{y}}) < z^* + (1/3)\varepsilon,\ \ 2z^* - \underline{z} - g(\boldsymbol{y}_i) \leq -\varepsilon$ より、

$$\begin{aligned}
||\boldsymbol{y}_{i+1} - \bar{\boldsymbol{y}}||^2 &\leq ||\boldsymbol{y}_i - \bar{\boldsymbol{y}}||^2 + s_i(-2(g(\boldsymbol{y}_i) - g(\bar{\boldsymbol{y}})) + s_i||\boldsymbol{d}_i||^2) \\
&= ||\boldsymbol{y}_i - \bar{\boldsymbol{y}}||^2 + s_i(-2(g(\boldsymbol{y}_i) - g(\bar{\boldsymbol{y}})) + g(\boldsymbol{y}_i) - \underline{z}) \\
&= ||\boldsymbol{y}_i - \bar{\boldsymbol{y}}||^2 + s_i(2g(\bar{\boldsymbol{y}}) - g(\boldsymbol{y}_i) - \underline{z}) \\
&\leq ||\boldsymbol{y}_i - \bar{\boldsymbol{y}}||^2 + s_i(2z^* + (2/3)\varepsilon - g(\boldsymbol{y}_i) - \underline{z}) \\
&\leq ||\boldsymbol{y}_i - \bar{\boldsymbol{y}}||^2 + s_i(-\varepsilon + (2/3)\varepsilon) \leq ||\boldsymbol{y}_i - \bar{\boldsymbol{y}}||^2 - (1/3)s_i\varepsilon \\
&= ||\boldsymbol{y}_i - \bar{\boldsymbol{y}}||^2 - (1/3)\varepsilon(g(\boldsymbol{y}_i) - \underline{z})/||\boldsymbol{d}_i||^2 \leq ||\boldsymbol{y}_i - \bar{\boldsymbol{y}}||^2 - (1/3)(\varepsilon^2/C)
\end{aligned}$$

が成り立つ。上記の式より、(十分大きな) i_0 が存在して、

$$||\boldsymbol{y}_{i_0} - \bar{\boldsymbol{y}}||^2 \leq ||\boldsymbol{y}_1 - \bar{\boldsymbol{y}}||^2 - (i_0 - 1)(\varepsilon^2/3C) < 0$$

となり矛盾。 //

3 つ目の数列を用いた場合、用いた下界が最適値と離れていたならば、最適解付近ではステップ長が長過ぎるため、最適解を通り越してしまう。特に最適解付近での劣勾配が短かった時は、次のステップで最適解から非常に遠いところへ移動してしまうことがあるため、挙動が安定しない。