拡張 Lagrange 法

中村 祐介

(2015 年 10 月 19 日)

1　制約条件なし極小化問題

以下の様な極小化問題を考える。

$E[f]$ を極小化する f を求める。
ただし $f = (f_0, f_1, \cdots, f_{M-1})$。

極小値を与える f を f^o とすると、以下の式が成立していることが必要である:

$$\frac{\partial E[f^o]}{\partial f_m} = 0 \tag{1}$$

これを「 1 次の必要条件」という。1 次の必要条件を満たすだけでは、極小とは限らない。しかし以下の条件が成立する場合、極小であると十分いえる:

$$\frac{\partial^2 E[f^o]}{\partial f_m \partial f_n} \quad \text{が正定値} \tag{2}$$

これを「 2 次の十分条件」という。

極小値問題は数値計算では「 降下法」を用いると効率的に解ける。降下法ではある適当な点 f^1 を仮定して、山を下る要領で $E[f^1] > E[f^2] > E[f^3] > \cdots$ を満たす f^i を次々に求めていく方法である。その中でも一番な「 最急降下法」について述べる。最急降下法では現在地点を f^i として、$E[f^i] > E[f^{i+1}]$ となる f^{i+1} を

$$f^{i+1} = f^i - \alpha \frac{\partial E[f^i]}{\partial f} \tag{3}$$

と選ぶ。ただし α は十分に小さい正の数である。つまり現在地点の勾配を見て、最も下る方向（ いわゆる最急降下方向）にステップ幅 α で移動する。α が十分小さければ、$E[f^{i+1}] < E[f^i]$ が保証されるので、近くに極小値があればそこに収束するはずである。つまり f^o において 2 次の十分条件が満たされているのならば、その近傍で最急降下法は収束性を持っている。

1.1　α の決め方

α が十分に小さければ収束性が良くなるが、小さい α では収束までにたくさんのステップ数が必要となる。計算の高速化のためには、収束を保証しつつも、出来るだけ α の値を大きく選びた

い。この観点では α は $E[f^{i+1}]$ が出来るだけ小さくなるように、選ぶのが良い。つまり

$$\frac{\partial}{\partial \alpha} E\left[f^i - \alpha \frac{\partial E[f^i]}{\partial f}\right] = 0 \tag{4}$$

と選ぶ。この選び方を α についての「正確な直線探索」という。上式は合成関数の微分より

$$\sum_m \frac{\partial E[f_i]}{\partial f_m} \frac{\partial E[f_i - \frac{\partial E[f^i]}{\partial f}]}{\partial f_m} = 0 \tag{5}$$

と書くことも出来る。つまり、f^{i+1} における勾配と f^i における勾配が直交するような α を選ぶことになる。これが最適であることが直感的にもすぐ分かる。

実際には正確な直線探索は計算コストが膨大なので、ある程度手を抜いて求めるのが良い。

2 制約条件付き極小化問題

解きたい問題以下の通り

<table><tr><td>制約条件 $G_r[f]=0$ の下で、$E[f]$ を極小化する f を求める。
ただし $f=(f_0, f_1, \cdots, f_{M-1})$、$r=0,1,\cdots,R-1$。</td></tr></table>

この時の 1 次の必要条件は

$$^{\exists}v_r^o \quad \text{s.t.} \quad \frac{\partial E[f^o]}{\partial f_m} + \sum_r v_r^o \frac{\partial G_r[f^o]}{\partial f_m} = 0\,, \tag{6}$$

$$G_r[f^o] = 0 \tag{7}$$

である。

2.1 Lagrange 未定乗数法

f, v を独立変数に持つ以下の関数を考える:

$$E'[f, v] = E[f] + \sum_r v_r G_r[f] \tag{8}$$

すると 1 次の必要条件 (6) は

$$\frac{\partial E'[f^o, v^o]}{\partial f_m} = 0\,, \qquad \frac{\partial E'[f^o, v^o]}{\partial v_r} = 0 \tag{9}$$

と書き換えることが出来る。これは $(M+R)$ 次元の制約条件なし極小化問題の 1 次の必要条件と同じ構造をしている。このようにして次元を上げる代わりに、制約条件を無くす方法を Lagrange 未定乗数法という。v は Lagrange 未定乗数と呼ばれる。

注意しなくてはいけないのは、E' は 2 次の十分条件を満たさないということである。「満たすとは限らない」ではなく、$G \equiv 0$ の様な自明な例を除けば「必ず満たさない」のである。つまり E の制約条件付きの極小点 (f^o, v^o) は E' の極小点になることはなく、必ず鞍点になる。(f^o, v^o) は 2 次の十分条件を満たさないので、最急降下法は収束しない。

降下法は使えないが、例えば $M + R$ 元の非線形連立方程式 (6) を解くことで、停留点 (f^o, v^o) を見つけることも出来る。しかしそれは一般には非常に難しい問題である。

2.2 拡張 Lagrange 法

拡張 Lagrange 法では Lagrange 未定乗数を導入しつつも、問題が停留値問題ではなく、極小値問題になるように工夫している。具体的には Lagrange 関数と呼ばれる以下の関数を考える:

$$L[f, v] = E[f] + \sum_r \left(v_r G_r[f] + \frac{\rho}{2} G_r^2[f] \right) \tag{10}$$

ここで ρ は手で与える十分に大きいパラメータである。

2.2.1 Lagrange 関数に対する極小値問題

本来未知の v^o を既に知っているとし、$L[f, v^o]$ を f のみの関数だとして、制約なし極小値問題を考えよう。この時の 1 次の必要条件は

$$\frac{\partial L[\bar{f}, v^o]}{\partial f_m} = \frac{\partial E[\bar{f}]}{\partial f_m} + \sum_r \Big(v_r^o + \rho G_r[\bar{f}] \Big) \frac{\partial G_r[\bar{f}]}{\partial f_m} = 0 \tag{11}$$

となる。ここで $\bar{f}$ は $L[f, v^o]$ の停留点である。もともとの 1 次の必要条件 (6),(7) を満たす (f^0, v^o) ならば、L の 1 次の必要条件を見たすことが分かる。逆に $G_r[\bar{f}] = 0$ を満たすのならば、$\bar{f} = f^o$ であることもすぐ分かる。

つづいて L の 2 次の十分条件を調べると

$$\frac{\partial^2 L[\bar{f}, v^o]}{\partial f_m \partial f_n} = \frac{\partial^2 E[\bar{f}, v^o]}{\partial f_m \partial f_n} + \sum_r \Big(v_r^o + \rho G_r[\bar{f}] \Big) \frac{\partial^2 G_r[\bar{f}]}{\partial f_m \partial f_n} + \rho \sum_r \frac{\partial G_r[\bar{f}]}{\partial f_m} \frac{\partial G_r[\bar{f}]}{\partial f_n} \tag{12}$$

となっている。特に $\bar{f} = f^o$ の場合を考えると

$$\frac{\partial^2 L[f^o, v^o]}{\partial f_m \partial f_n} = \frac{\partial^2 E[f^o, v^o]}{\partial f_m \partial f_n} + \sum_r \left(v_r^o \frac{\partial^2 G_r[f^o]}{\partial f_m \partial f_n} + \rho \frac{\partial G_r[f^o]}{\partial f_m} \frac{\partial G_r[f^o]}{\partial f_n} \right) \tag{13}$$

となっている。ここで $\frac{\partial G_r[\bar{f}]}{\partial f_m} \frac{\partial G_r[\bar{f}]}{\partial f_n}$ が半正定値行列であることに注目する。もし ρ が十分に大きければ、この項が支配的になるので、何となく $\frac{\partial^2 L[f^o, v^o]}{\partial f_m \partial f_n}$ が正定値行列っぽくなる[*1]。すると 2 次の十分条件を満たすので、降下法が使えるようになる。また停留点 $\bar{f}$ は極小点となる。

[*1] もっと数学的にぎりぎりを攻めれば、証明できるらしい。

まとめると以下の通り。ρ が十分大きくかつ v^o が既知であれば、$L[f, v^o]$ の f についての条件なし極小値問題が降下法で解ける。さらにこの時の極小点 $\bar{f}$ が $G_r[\bar{f}] = 0$ を満たすのであれば、それが求めたかった f^o である。

2.2.2　v^o の求め方

実際は v^o が既知であることはありえない。つまり v^o を求める手順が必要である。そこである適当な v を選んで $L[f, v]$ を f について極小化してみる。1 次の必要条件は

$$\frac{\partial L[\bar{f}, v]}{\partial f_m} = \frac{\partial E[\bar{f}, v]}{\partial f_m} + \sum_r \Big(v_r + \rho G_r[\bar{f}]\Big) \frac{\partial G_r[\bar{f}]}{\partial f_m} = 0 \tag{14}$$

である。ここでもし $G_r[\bar{f}] = 0$ ならば

$$\frac{\partial E[\bar{f}, v]}{\partial f_m} + \sum_r v_r \frac{\partial G_r[\bar{f}]}{\partial f_m} = 0 \tag{15}$$

となる。これはもともとの 1 次の必要条件 (6) そのものであるので、問題はすでに解けており、$v = v^o$, $\bar{f} = f^o$ であったことになる。逆に $G_r[\bar{f}] \neq 0$ となってしまったのならば、それは $v \neq v^o$ であったためである。その場合は、v を選び直さなくてはいけない。式 (14) と式 (15) を見比べると

$$v_r + \rho G_r[\bar{f}] \longrightarrow v_r \tag{16}$$

と取り直すのが、良さそうだと予想できる[*2]。

まとめると以下の通り。ある v を与えて $L[f, v]$ の f に関する制約条件なし極小値問題を解く。その後、式 (16) で v を定義し直すと、v は前よりも v^o に近づいている。

3　拡張 Lagrange 法のアルゴリズム

3.1　プロトタイプ

以下に示すのが拡張 Lagrange 法のアルゴリズムである。

1. v^o に対する推定値 v^1 を与える。$k = 1$ とする。
2. v^k を固定して $L[f, v^k]$ の制約条件なし極小値問題を解き、極小値を与える f を f^k とする。
3. もし $G[f^k, v^k] = 0$ ならば、$(f^o, v^o) = (f^k, v^k)$ であるので、計算を終了する。
4. 一方 $G[f^k, v^k] \neq 0$ ならば、それは $v^k \neq v^o$ であったためである。そこで v 以下の公式で修正する:
$$v^{k+1} = v^k + \rho G[f^k] \tag{17}$$

[*2] 実際に良いことが証明できるが、難しいので略。

5. $k = k + 1$として、2. に戻る。

手順 2. で条件なしの極小値問題を解かなくてはいけないが、その手法は任意である。

3.2 改良版その1

ここから先は、中村が考えたアルゴリズムである。まだ改善途中なので、ほかのアイデアも試して欲しい。

プロトタイプのアルゴリズムは数学的に見通しが良いが、数値計算上では無駄が多い。手順 2. で正確に極小値問題を解かず、途中で打ち切ってしまうのが良い。中でも最急降下法で 1 ステップだけ下るのが効率が良い。

一般により高級な降下法（例えばニュートン法など）で下るほうが1 ステップで大きく下れるが、そもそも v^k が v^o に近い保証はないので、大きく下るのは危険である。小さなステップを数多く重ねるのなら、計算量の観点から最急降下法が優れているのである。

従ってアルゴリズムは以下の様になる：

1. v^o, f^o に対する推定値 v^1, f^1 を与える。$k = 1$とする。
2. v^k を固定して $L[f, v^k]$ に対して f の最急降下法で 1 ステップ下る。具体的には $f^{k+1} = f^k - \alpha \frac{\partial L}{\partial f}$ とする。ここで α は適切なステップ幅。
3. もし $G[f^k, v^k] = 0$ かつ $\frac{\partial L}{\partial f} = 0$ ならば、$(f^o, v^o) = (f^k, v^k)$として計算を終了する。
4. そうでなければ以下の公式で v を修正する：
$$v^{k+1} = v^k + \rho G[f^k] \tag{18}$$
5. $k = k + 1$として、2. に戻る。

つまり f の降下と v の修正を交互にしていく。手順 2. の α は大きすぎると破綻するが、小さすぎると計算が遅い。毎回 α を直線探索（もしくは Armijo の方法、Wolfe の方法など）で決める方が、収束までのループ回数が減る。しかし、逆に直線探索に時間が掛かるので、1 ループあたりに掛かる時間が増える。

3.3 改良版その2

計算が速くなるように、できだけ大きな α を使いたいが、α の直線探索で決めようとすると、それに時間が掛かる。そこでループごとに少しずつ改善する方法を採用しよう。つまり最初はある程度小さな α_{ini} で計算を始めて、L の変化を見ながら次のループの α を決定していく。以下の方法で上手くいった：

- もし $L[f^{k+1}, v^{k+1}] > L[f^k, v^k]$ ならば $\alpha \leftarrow \max[\alpha_{\mathrm{ini}}, \alpha\gamma_1]$。

- もし $L[f^{k+1}, v^{k+1}] < L[f^k, v^k]$ ならば $\alpha \leftarrow \alpha\gamma_2$。

つまり L の極小化が上手く働いたら次のループでは α は大胆にとり、失敗したら控え目にする。ここで $\alpha_{\mathrm{ini}} = 0.01, \gamma_1 = 0.5, \gamma_2 = 1.1$ くらいが丁度良い。このアルゴリズムをさらに数値計算向けに書き換えたものを「ALSolver3 の使い方」に載せてある。

4 さらに改良版

ここまでの拡張 Lagrange 法に以下の 2 つの改良を加える。

4.1 制約条件の改良

例えば、$G_1 \equiv 1 - \sum \cdots = 0$, $\quad G_2 \equiv N - \sum \cdots = 0$ の様な規格化と粒子数に関する制約条件を課すことがある。しかし N が大きい場合、両者の強さに違いがありすぎるので、計算がかなり不安定になってしまう。これは $G_r = 0$ という制約条件では $G \to cG$ の様なスケール変換を許してしまっていることに起因する。最適な ρ の選び方がほぼ不可能になっている。(「G_1 に対しては大きすぎ、G_2 に対しては小さすぎる」というジレンマに陥る可能性がある？)

つまり、制約条件の強さを揃える形式すべきである。そこで制約条件は

$$G_r \equiv 1 - C_r = 0 \tag{19}$$

という形に制限することにする。これで、r 毎の制約条件の強さのばらつきが抑えられるはずである。先ほどの例で云うならば、

$$G_2 = 1 - \frac{1}{N} \sum \cdots \tag{20}$$

のように予め N で割っておくことを強制する。

4.2 条件数の改善

L の f に関する最急降下法を使うわけだが、最急降下法は L の条件数（Hesse 行列の最大固有値と最小固有値の比）が大きいと極端に収束性が悪くなってしまう。そこで、出来るだけ低コストで条件数を下げたい。

ここで行うような条件数を下げる処理を「前処理」という。今回はその中でも最も単純で低コスト（で当然低性能）な Jacobi 前処理を応用した前処理を施すことにする。

4.2.1 Jacobi 前処理

以下の様に正定値行列 A を使って定義される $L[\boldsymbol{x}]$ の最小値問題を考える：

$$L = \boldsymbol{x}^\dagger A \boldsymbol{x} \tag{21}$$

さて、ある正則行列 P を用いて

$$L = \boldsymbol{y}^\dagger B \boldsymbol{y}\,, \quad \boldsymbol{y} = P^{-1}\boldsymbol{x}\,, \qquad B = P^\dagger A P \tag{22}$$

と書き換えても数学的に当然等価である。そして、B の条件数を A の条件数よりも小さくするような P を得ることが出来たとする。すると B に対して最急降下法を適用すれば、A に対して適用するより、速く収束してくれる。極端な例として、$P = A^{-1/2}$ と選ぶと、B は単位行列となるので、最急降下法は任意の初期推定値に対して、反復回数 1 で真の解に収束する。とはいえ $A^{-1/2}$ の計算は L の最小化というもとの問題よりもはるかに難しいので、実用的ではない。重要なのは、いかに少ない計算量で $A^{-1/2}$ に近い行列を見つけてくるかということである。

Jacobi 前処理では P を

$$P_{nm} = \delta_{nm}\frac{1}{\sqrt{A_{nn}}} \tag{23}$$

と選ぶ。A の対角優位性が強い場合、この選び方は $A^{-1/2}$ の良い近似になっている。

4.2.2 Jacobi 前処理を適用する

もともとの Lagrange 関数は

$$L[f,v] = E[f] + \sum_r \left(v_r G_r[f] + \frac{\rho}{2} G_r^2[f] \right) \tag{24}$$

である。ここで

$$g_m = \sqrt{p_m}\, f_m \tag{25}$$

という変数変換をする。p_m の具体形は後で考える。L を g の関数だと思って、最急降下法を適用しよう。L の g に関する Hesse 行列は

$$\frac{\partial^2 L}{\partial g_m \partial g_n} = \frac{\partial^2 E}{\partial g_m \partial g_n} + \sum_r \left[(v_r + \rho G_r) \frac{\partial^2 G}{\partial g_m \partial g_n} + \rho \frac{\partial G_r}{\partial g_m} \frac{\partial G_r}{\partial g_n} \right] \tag{26}$$

$$= \frac{1}{\sqrt{p_m p_n}} \left\{ \frac{\partial^2 E}{\partial f_m \partial f_n} + \sum_r \left[(v_r + \rho G_r) \frac{\partial^2 G}{\partial f_m \partial f_n} + \rho \frac{\partial G_r}{\partial f_m} \frac{\partial G_r}{\partial f_n} \right] \right\} \tag{27}$$

Jacobi 前処理の精神に則れば

$$p_m = \frac{\partial^2 E}{\partial f_m \partial f_m} + \sum_r \left[(v_r + \rho G_r) \frac{\partial^2 G}{\partial f_m \partial f_m} + \rho \left(\frac{\partial G_r}{\partial f_m} \right)^2 \right] \tag{28}$$

と選ぶのが良さそうである。しかし式 (25) から分かる通り、$p_m > 0$ でないといけない。数学的には p_m が非常に小さな正数でも OK であるが、数値計算上は危険なので、ある適当な $p_{\min}$ を設定して、

$$p = \max\left\{ \frac{\partial^2 E}{\partial f_m \partial f_m} + \sum_r \left[(v_r + \rho G_r) \frac{\partial^2 G_r}{\partial f_m \partial f_m} + \rho \left(\frac{\partial G_r}{\partial f_m} \right)^2 \right],\ p_{\min} \right\} \tag{29}$$

と選ぶことにする。

最急降下法の 1 ステップは

$$g_m^{k+1} = g_m^k - \alpha \frac{\partial L}{\partial g_m} \tag{30}$$

である。式 (25) を用いて、f の言葉に戻すと

$$f_m^{k+1} = f_m^k - \frac{\alpha}{p_m} \frac{\partial L}{\partial f_m} \tag{31}$$

となる。手順 2 の降下法の部分を上式に置き換えればよい。

以上2 つの改良を施したソルバーの使い方を「 ALSolver4 の使い方」に載せてある。