A6 Iterative Algorithms

樋口さぶろお

龍谷大学大学院理工学研究科数理情報学専攻

理論物理学特論 L14(2020-12-23 Wed)

最終更新: Time-stamp: "2021-01-06 Wed 07:45 JST hig"

今日の目標

- 確率的/決定的反復法のそれぞれの特徴が説明できる

ここまで来たよ

14 A6 Iterative Algorithms

- 反復法による最小化
- 勾配法
- Simulated Annealing
- パラメタ分割法
- 補助関数法

(多次元連続) 最小化 (最適化) 問題

$\theta_{\mathsf{min}} = \mathrm{argmin}_{\theta}\phi(\theta) = ?$

解法の分類　　数理計画法

- 厳密解 θ_{min} が計算可能な式で書ける
  - explicit solution, closed expression
- 近似解
  - 決定的反復法 iterative methods. Loop をまわると, だんだん θ_{min} に近づいていく KA MDA A.6.1
    - ★ パラメタ分割法 parameter partition algorithms ♠
    - ★ 補助関数法 auxiliary function algorithms
      EM=Expectation-Maximization Algorithm, 途中で MCMC も ♡
    - ★ 勾配法 gradient algorithms KA MDA A.6.2, A.6.3　　最急降下法 gradient descent
  - 近似的確率的反復法
    - ★ 焼き鈍し法 Simulated Annealing ← Markov Chain Monte Carlo ♡ $\rightsquigarrow$ Gibbs sampler ♠

implicit solution 解き方はわからないけど, 厳密な代数/超越方程式の解として表現できる

二分法やニュートン法じゃだめ?

$\nabla\phi(\theta_{\mathrm{min}}) = \mathbf{0}$ を二分法やニュートン法で解けばいい?
1 次元の場合の困難

- 二分法やニュートン法では, 初期条件次第で 零点のうちの 1 個がみつかるのだった

2 次元以上の場合の困難

- 二分法 (そのままの形の) 中間値の定理はない
- ニュートン法 $\rightsquigarrow$ 勾配法

高次元の二分法と似た?アプローチ 'grid search'? 計算量 $(\frac{L}{\Delta x})^d$.
安易に頼ってはいけない.

決定的反復法の一般論

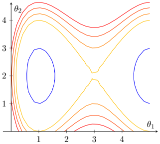
Step1 $\theta_{[0]} :=$ 適当な初期値

Step2 $\theta_{[t]}$ を小さく変更して $\theta_{[t+1]} = \cdots$ に '改善' する local search

- '改善': $\phi(\theta_{[t+1]}) \leq \phi(\theta_{[t]})$ となるように $\rightsquigarrow$ 単調減少列

Step3 実質的に '収束' したなら $\theta_{[t+1]}$ を出力して終了, そうでなければ, $t := t + 1$ して Step2 にもどる

- '収束' 判定はさまざまなものがありうる.
  $\phi(\theta_{[t]}) - \phi(\theta_{[t+1]}) \geq 0$ が小さい, 何らかの距離
  $d(\theta_{[t]}, \theta_{[t+1]})$ が小さい, など.



ここまで来たよ

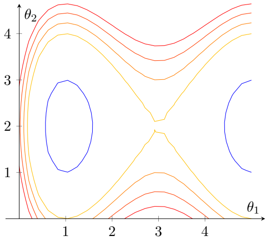
14 A6 Iterative Algorithms

- 反復法による最小化
- 勾配法
- Simulated Annealing
- パラメタ分割法
- 補助関数法

勾配法 Gradient Algorithms

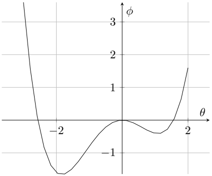
'改善'(Update): $\theta_{[t+1]} = \theta_{[t]} - s\nabla\phi(\theta_{[t]})$ KA MDA A.6.7 ベクトル解析
s の決定は自明でない. ϕ:m, θ:g, s:g^2/m のように次元が異なる.

Step2.1 $s := 1$
Step2.2 $\theta_{[t+1]} = \theta_{[t]} - s\nabla\phi(\theta_{[t]})$
Step2.3 もし $\phi(\theta_{[t+1]}) \leq \phi(\theta_{[t]})$ なら Step2 終了, そうでなければ, $s := s/\sqrt{2}$ して Step2.2 にもどる



勾配法の弱点

- Local minimum にとらわれて global minimum を見つけられないことがある
- s の設定に考察が必要



ここまで来たよ

14 A6 Iterative Algorithms
- 反復法による最小化
- 勾配法
- Simulated Annealing
- パラメタ分割法
- 補助関数法

Simulated Annealing

エネルギー ϕ, 外部パラメタ (逆温度) $\beta : 0 \to +\infty$.

Step1 $\theta_{[0]} :=$ 適当な初期値

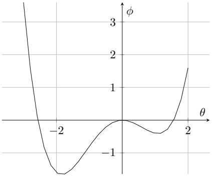
Step2.2 $\theta_{[t+1]} = \theta_{[t]} +$ ランダムな小さいベクトル

Step2.3 もし $\phi(\theta_{[t+1]}) \leq \phi(\theta_{[t]})$ なら Step2 終了, そうでなければ,

- **Step2.3.1** $R \sim U(0,1)$ ($[0,1)$ 連続型一様乱数)
- **Step2.3.1**

$$R < \mathrm{e}^{-\beta\cdot(\phi(\theta_{[t+1]})-\phi(\theta_{[t]}))}$$

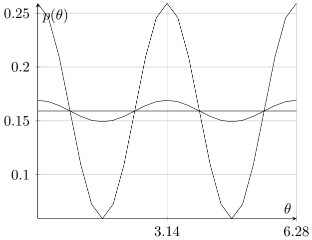
なら $\theta_{[t+1]}$ を返す. そうでなければ元の $\theta_{[t+1]} := \theta_{[t]}$ を返す.



Step2.4 β を少し増やす.

Simulated Annealing の統計力学的意味

マルコフ連鎖 は極限分布=定常分布を持つことがある.
極限分布=定常分布がボルツマン分布 $\frac{1}{Z}\mathrm{e}^{-\beta\phi(\theta)}$ に従うように設計.
一様分布 $\rightarrow$ 最低エネルギー状態に集中した分布, as $\beta : 0 \rightarrow +\infty$.



Simulated Annealing の弱点

- 収束の判定が難しい
- β の増加スケジュールに考察が必要
- local minimum につかまらないくらいゆっくり β を増加させると, 計算時間がかかる (たりないと local minimum につかまる)

ここまで来たよ

14 A6 Iterative Algorithms
- 反復法による最小化
- 勾配法
- Simulated Annealing
- パラメタ分割法
- 補助関数法

パラメタ分割法 Parameter partition algorithms

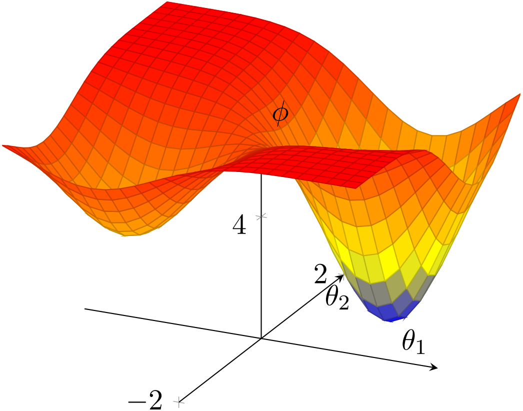
$\theta_{[t]} = [\theta_1^{[t]}, \theta_2^{[t]}]$.

Step 2.1. $\theta_2^{[t]}$ を固定して, $\theta_1^{[t+1]} = \mathrm{argmin}_{\theta_1} \phi(\theta_1, \theta_2^{[t]})$.
Step 2.2. $\theta_1^{[t+1]}$ を固定して, $\theta_2^{[t+1]} = \mathrm{argmin}_{\theta_2} \phi(\theta_1^{[t+1]}, \theta_2)$.

分割されたパラメタについて, 最小化までできなくても, 改善するだけでもいい.
MCMC, Simulated Annealing の文脈では ‘Gibbs sampling’.

パラメタ分割法の弱点

- Local minimum にとらわれて global minimum を見つけられないことがある



$$\mathsf{minarg}_{\theta_1,\theta_2}\phi(\theta_1,\theta_2) = \mathsf{minarg}_{\theta_1}\mathsf{minarg}_{\theta_2}\phi(\theta_1,\theta_2)$$

じゃないの? $\rightsquigarrow$ 自明じゃないし, この書き方じゃ両辺の型がそろってない…

正しくは,

$$\begin{aligned}\mathsf{min}_{\theta_1,\theta_2}\phi(\theta_1,\theta_2)& \\ &= \mathsf{min}_{\theta_1}(\mathsf{min}_{\theta_2}\phi(\theta_1,\theta_2)) \\ &= \mathsf{min}_{\theta_1}\phi(\theta_1,\theta_2^{\mathsf{min}}(\theta_1))\end{aligned}$$

パラメタ分割法 では $\theta_2(\theta_1)$ を考えてない.

ここまで来たよ

14 A6 Iterative Algorithms

- 反復法による最小化
- 勾配法
- Simulated Annealing
- パラメタ分割法
- 補助関数法

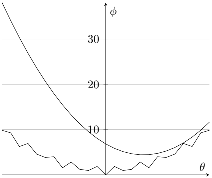
補助関数法 Auxiliary function algorithms

$\phi(\theta)$ と別の補助関数 $\eta(\theta)$ をもってくる．ただし，

$$\phi(\theta_{[t]}) = \eta(\theta_{[t]}),$$
$$\phi(\theta) \leq \eta(\theta)$$

Step2 (ϕ を気にせず)η を減少させる $\theta_{[t+1]}$ を作る．

$$\phi(\theta_{[t]}) = \eta(\theta_{[t]}) \geq \eta(\theta_{[t+1]}) \geq \phi(\theta_{[t+1]})$$



補助関数法の別の記述方法

$\phi(\theta)$ と別の補助関数 $\eta(\theta,\xi)$ をもってくる. ただし,

$$\phi(\theta) = \min_{\xi} \eta(\theta,\xi).$$

すると,

Step2.1 $\xi_{[t+1]} = \mathrm{argmin}_{\xi}(\theta_{[t]},\xi)$
Step2.2 $\theta_{[t+1]} = \mathrm{argmin}_{\theta}(\theta,\xi_{[t+1]})$

で, $\phi(\theta_{[t]}) \geq \phi(\theta_{[t]})$ となる.

Hunter, D. R., Lange, K. A tutorial on MM algorithms. The American Statistician, 58(1), 30-37, 2004.

補助関数法の例: EM Algorithm

EM=Expectation-Maximization KA MDA Note12.1
log likelihood: $\phi(\theta) = \ell(\theta, X_0) = \log P(X_0, X_1|\theta)$
X_0: 未測定 (因子分析における latent factor), X_1: 測定済

$$\begin{aligned} \text{E-Step} \quad & Q(\theta) = \mathrm{E}_{X_0}[\ell(\theta_{[t]}, X_0)] \\ \text{M-Step} \quad & \theta_{[t+1]} = \mathrm{argmax}_\theta(Q(\theta)) \end{aligned}$$

Q, ℓ は, 変数 $p(X_0)$ を追加した補助関数と思える.

補助関数法の弱点

- よい補助関数を発見し, アルゴリズムの振舞いを証明するのは容易でない.