歩幅の天井を決めているのは勾配の大きさではなく、いちばん急な向きの曲率である

— 二次関数の条件数を $\kappa = 100$ とすると、歩幅の天井は $2/\kappa = 0.02000000$ で、その 1.01 倍の 0.02020000 では一歩 1.020000 倍と発散する — 分離子は、進みを止めているのが勾配の大きさか、曲率のばらつきか — [論文734]

山岸 勇輝 (Yuuki YAMAGISHI)
Yamagishi Complexity Torus Lab, Japan｜山岸勇輝ラボ
2026年9月

— From thinking to syncing. —

要旨

勾配降下で歩幅を上げれば速くなる。ただし天井が在り、その天井は問題の形が決めている。天井と、最適な歩幅と、そのときの歩数を、同じ一つの数から出す。新しい定理も法則も主張しない。 第一に、天井は $2/\kappa$ である——$\kappa = 100$ なら 0.02000000 で、勾配の大きさも出発点も入らない（第2節）。 第二に、天井を少し超えるだけで発散する——1.01 倍の 0.02020000 では一歩 1.020000 倍になる（第2節）。 第三に、これが本稿の芯である。最適な歩幅でも、一歩の縮みは 1 に近い——$2/(1+\kappa) = 0.01980198$ で一歩 0.980198 倍にとどまり、$(\kappa-1)/(\kappa+1)$ と差 0 で一致する（第3節）。 第四に、だから歩数が要る——100 分の 1 に縮めるのに 230.250834 歩、$\kappa = 1000$ なら 2302.584325 歩である（第3節）。 第五に、二つの曲率は違う働きをする——大きいほうが天井を決め、小さいほうは歩数にしか効かない（第3節）。 第六に、分離子は、進みを止めているのが勾配の大きさか、曲率のばらつきかである——止めているのは後者で、前者は歩幅に何の制限も与えない（第4節）。 キーワード：勾配降下、歩幅、条件数、収束率、曲率が決める天井、同じ語が別の量。

> 本稿の射程（射程注記）：新しい定理も法則も主張しない——二次関数に対する勾配降下の収束条件、最適な歩幅 $2/(L+\mu)$、収束率 $(\kappa-1)/(\kappa+1)$ は、いずれも標準的である（コーシー 1847、ネステロフ 2004）。二次関数に限る——非凸な面、鞍点、確率的勾配降下、運動量や適応的な手法は扱わない。実在の問題を主張しない——個別の学習器やデータは扱わず、条件数だけを動かす。前処理を設計しない——条件数を下げる手は存在を指すだけで、作らない。$\kappa = 1$ の場合に注意——そこで一歩で終わるのは二次関数だからであり、一般には成り立たない。既刊との関係：論文434 は、安くしているのが時間ではなく累積であることを示した——そこには「学習率」という語が 17 回出るが、それは学習曲線の勾配、すなわち累積生産が倍になると費用が何割下がるかという量である。本稿の「学習率」は最適化の歩幅で、次元も意味も違う。同じ語が別の量を指す例になる。加えたのは、天井を $2/\kappa$ と出して両側で確かめたこと、最適な歩幅の縮みが $(\kappa-1)/(\kappa+1)$ と一致することを示したこと、条件数ごとの歩数を並べたこと、二つの曲率の働きが違うことを分けたこと、分離子を「勾配か、曲率のばらつきか」に置いたことである。

1. 記法と問い

設定：二つの向きの曲率が 1 と κ の二次関数を考える。勾配降下は、各向きで一歩ごとに一定の倍率をかけることになる。

$$f(x) = \tfrac{1}{2}\left(x_1^2 + \kappa x_2^2\right), \qquad \text{縮み} = \max\left(|1-\eta|,\ |1-\eta\kappa|\right)$$

η は歩幅、κ は条件数である。縮みが 1 以上になれば発散する。

問い：歩幅はどこまで上げられるのか。

2. 天井は、いちばん急な向きが決めている

計算 2.1：$\kappa = 100$ として、歩幅ごとの一歩の縮みを出した。

歩幅	一歩の縮み	結果
0.00100000	0.999000	縮む
0.00500000	0.995000	縮む
0.01980000	0.980200	縮む
0.01980198	0.980198	縮む
0.02500000	1.500000	発散する

> **観察 2.2**
>
> 天井は $2/\kappa = 0.02000000$ である。その 1.01 倍の 0.02020000 では一歩 1.020000 倍、0.99 倍の 0.01980000 では 0.980200 倍になる——境目の両側で符号が変わることを確かめた。

> **観察 2.3**
>
> この天井に、勾配の大きさは入っていない。入っているのは、いちばん急な向きの曲率だけである——勾配がどれだけ大きくても、歩幅の上限は変わらない。

> **観察 2.4**
>
> 天井は曲率に反比例する。κ を 10 倍にすると天井は 10 分の 1 になり、0.02000000 から 0.00200000 へ下がる——急な向きが一つでも在れば、そこが全体の歩幅を縛る。

3. 最適な歩幅でも、縮みは一に近い

計算 3.1：条件数ごとに、最適な歩幅、一歩の縮み、100 分の 1 に縮めるのに要る歩数を出した。

κ	最適な歩幅	一歩の縮み	要る歩数
1	1.00000000	0.000000	1
10	0.18181818	0.818182	22.948892
100	0.01980198	0.980198	230.250834
1000	0.00199800	0.998002	2302.584325

> **観察 3.2**
>
> これが本稿の芯である。最適な歩幅でも、$\kappa = 100$ なら一歩 0.980198 倍にとどまる。これは $(\kappa - 1)/(\kappa + 1)$ と差 0 で一致する——歩幅を最善に選んでも、条件数が大きければ縮みは 1 に張り付く。

観察 3.3

だから歩数は条件数にほぼ比例する。$\kappa = 10$ で **22.948892** 歩、100 で **230.250834** 歩、1000 で **2302.584325** 歩である——10 倍ずつ増えていて、条件数を下げるほうが歩幅をいじるより効く。

観察 3.4

二つの曲率は、違う働きをしている。大きいほうは天井を決め、小さいほうは歩数を決める——「学習率をどうするか」という問いは、前者にしか触れていない。

観察 3.5

$\kappa = 1$ で一歩で終わるのは、二次関数だからである。曲率が一つなら歩幅 1 でちょうど底に着く——一般の関数では成り立たないので、この行は二次関数の中の話である。

4. 分離子

観察 4.1

分離子は、進みを止めているのが勾配の大きさか、曲率のばらつきかである。止めているのは後者で、前者は歩幅に何の制限も与えない——「急な坂だから慎重に」は、この式の中では理由になっていない。

観察 4.2

論文**434** と並べる。**434** は、安くしているのが時間ではなく累積であることを示した——そこにも「学習率」という語が出るが、それは累積生産に対する費用の下がり方で、本稿の歩幅とは次元も意味も違う。同じ語が別の量を指す例である。

> **観察 4.3**
>
> 言わないこと。非凸な面も鞍点も確率的な手法も前処理の設計も扱わず、二次関数に限った——示したのは、天井と歩数がそれぞれどの量から来ているかである。

5. 点呼と結び

1. 天井は $\mathbf{0.02000000}$（観察 2.2）：$2/\kappa$。
2. $\mathbf{0.02020000}$ で $\mathbf{1.020000}$ 倍（観察 2.2）：発散する。
3. $\mathbf{0.01980000}$ で $\mathbf{0.980200}$ 倍（計算 2.1）：境目の内側。
4. 0.02500000 では $\mathbf{1.500000}$ 倍（計算 2.1）：はっきり発散。
5. 最適は $\mathbf{0.01980198}$（計算 3.1）：$2/(1+\kappa)$。
6. 縮みは $\mathbf{0.980198}$ 倍（観察 3.2）：本稿の芯。
7. 歩数 $\mathbf{230.250834}$（計算 3.1）：$\kappa = 100$。
8. $\kappa = 10$ なら $\mathbf{22.948892}$（計算 3.1）：十分の一。
9. $\kappa = 1000$ なら $\mathbf{2302.584325}$（計算 3.1）：十倍。
10. 勾配か、曲率のばらつきか（観察 4.1）：分離子・論文434。

結び：歩幅の天井を決めているのは勾配の大きさではなく、いちばん急な向きの曲率である。二次関数の条件数を $\kappa = 100$ とすると天井は $2/\kappa = 0.02000000$ で、その 1.01 倍の 0.02020000 では一歩 1.020000 倍と発散し、0.99 倍の 0.01980000 なら 0.980200 倍で縮む——この天井に勾配の大きさも出発点も入っておらず、入っているのはいちばん急な向きの曲率だけである。最適な歩幅は $2/(1+\kappa) = 0.01980198$ だが、そのときの一歩の縮みは 0.980198 倍にとどまり、$(\kappa-1)/(\kappa+1)$ と差 0 で一致する——歩幅を最善に選んでも、条件数が大きければ縮みは 1 に張り付く。だから歩数が要る——100 分の 1 に縮めるのに $\kappa = 10$ で 22.948892 歩、100 で 230.250834 歩、1000 で 2302.584325 歩と、10 倍ずつ増える。二つの曲率は違う働きをしていて、大きいほうが天井を決め、小さいほうが歩数を決める——「学習率をどうするか」という問いは前者にしか触れていないので、条件数を下げるほうが歩幅をいじるより効く。分離子は、進みを止めているのが勾配の大きさか、曲率のばらつきかである。既刊との位置——論文434 にも「学習率」という語が出るが、そちらは累積生産に対する費用の下がり方で、次元も意味も違う。正直に言えば——二次関数に限り、非凸な面も確率的な手法も扱わず、$\kappa = 1$ で一歩で終わるのも二次関数だからである。

参考文献

1. Cauchy, A. (1847). Methode generale pour la resolution des systemes d’equations simultanees. *Comptes Rendus de l’Academie des Sciences*.
2. Nesterov, Y. (2004). *Introductory Lectures on Convex Optimization*. Springer.
3. 山岸勇輝 (2026). 『安くしているのは、時間ではなく累積である』[論文434]. Zenodo. 10.5281/zenodo.22708735

— From thinking to syncing. —

山岸 勇輝 (Yuuki YAMAGISHI)

© 2026 Yamagishi Complexity Torus Lab, Japan · Licensed under Creative Commons Attribution 4.0 International (CC BY 4.0)