

曲がった空間での最適化

佐藤 寛之

2 次関数の最小化問題という慣れ親しんだ最適化問題を通して、最急降下法と呼ばれるアルゴリズムを紹介する。これはいわゆるまっすぐな空間での最適化である。続いて、各変数の 2 乗和が 1 という条件が付いた場合は、最適化問題が単位円や単位球面上で考えられることを説明し、曲がった空間での最急降下法を紹介する。さらに、このような曲がった空間での最適化、すなわちリーマン多様体上の最適化という比較的新しい研究分野の概要をなるべく簡明に紹介する。最後に、諸問題への応用例についても触れる。

キーワード：最適化、ユークリッド空間、超球面、リーマン多様体、最急降下法、2 次関数

1. はじめに

皆さんは、この記事のタイトルを見てどのような感想を抱くでしょうか。「最適化」は、この特集の他の記事でも多く取り扱われていますね。この記事では、最適化問題と言えば「与えられた関数（目的関数）を制約条件の下で最小化する問題」だとします。

では、「曲がった空間」とは何でしょうか。「曲がった空間」について考えるためには、まず「まっすぐな空間」から考える必要があるでしょう。それでは「まっすぐな空間」を記述する数学的道具とは何でしょうか？

高校で習う平面ベクトルや空間ベクトルは、それぞれ 2 次元の世界、3 次元の世界を記述する道具です。平面ベクトルのことを 2 次元ベクトル、空間ベクトルのことを 3 次元ベクトルとも言います。4 次元や 5 次元、あるいは一般に n 次元の世界がどんなものなのかを想像するのは難しいですが、2 次元ベクトルが成分を二つもち、3 次元ベクトルが成分を三つもちことから、 n 次元ベクトルというものがあるとすれば、成分を n 個もつのが自然でしょう。また、2 次元ベクトルと 3 次元ベクトルは成分の個数こそ異なるものの、共通する性質がたくさんあります。たとえば、「二つのベクトル \vec{x}, \vec{y} に対して $\vec{x} + \vec{y} = \vec{y} + \vec{x}$ が成り立つ」という性質（交換法則）は n 次元ベクトルに関しても成り立つべきでしょう。このように、 n 個の成分をもち、かつ、2 次元ベクトルや 3 次元ベクトルが満たす交換法則などの性質を満たすもの全体の集合として n 次元ベクトル空間（これを \mathbb{R}^n と表します）を定義することができ、 \mathbb{R}^n の要素として n 次元ベクトル $\vec{x} = (x_1, x_2, \dots, x_n)$

を定義することができます。したがって、平面は 2 次元ベクトル空間 \mathbb{R}^2 であり、（高校で学習する）空間は 3 次元ベクトル空間 \mathbb{R}^3 です。ベクトル空間の正確な定義についてはたとえば [1] を参照してください。さらに、高校で学習するのと同様に \mathbb{R}^n でも内積を定義することができます。このような内積が定義されている \mathbb{R}^n を（標準的）ユークリッド空間と言います。要するに、「まっすぐな空間」がユークリッド空間です。

次に「曲がった空間」についてですが、簡単な例としては、 \mathbb{R}^2 の中での単位円や \mathbb{R}^3 の中での単位球面が挙げられます。さらに、これらの概念を拡張したものを考えたいのですが、この先はこれ以降の節を通して少しずつ紹介していきます。

なお、ベクトルを「 \vec{x} 」ではなく「 \mathbf{x} 」のように表す記法もよく使われます。この記事では、これ以降はそれにならってベクトルを太字で表すことにします。また、 $y = f(x)$ という関係があるとき、 $f(x)$ を関数と言いますが、 x に y を対応させる規則そのもの、つまり f のことを関数と言うこともあります。そのため、この記事では「関数 $f(x)$ 」や「関数 f 」のように引数のある表記とない表記が混在しており、微妙にニュアンスが異なるのですが、さほど気にする必要はありません。

2. 「まっすぐな空間」での最適化

2.1 具体的な問題

ここで、少し唐突ですが次の問題を考えてみましょう。

問題 1. $2x^2 + 4xy + 5y^2$ が最小となる実数の組 (x, y) を求めよ。

高校の数学でもよく見かける問題ですが、これも最適化問題です。さて、この問題は少し手計算をする

と簡単に解けます. すなわち, $2x^2 + 4xy + 5y^2 = 2(x+y)^2 + 3y^2$ より, $x+y=y=0$, つまり $x=y=0$ のとき最小となり, 最小値は 0 となります. この実数の組 $(x, y) = (0, 0)$ のように目的関数を最小とするものを, この問題に対する最適解と言います.

ところで, (x, y) は実数の組ですが, 2次元ベクトルとみなすこともできます. これを \mathbf{x} とおきたいのですが, x と \mathbf{x} が混在するとややこしいので, 今後は (x, y) の代わりに (x_1, x_2) と書くことにしましょう. これらは単に記法が異なるだけです, 本質的には全く同じことです. さて, $\mathbf{x} = (x_1, x_2)$ とおきましょう. また, 先ほどの x と y の 2 次式も x_1 と x_2 で書き直し, さらにこれを 2 変数 x_1 と x_2 の関数とみなして $f(x_1, x_2)$ と書くことにします. すなわち,

$$f(x_1, x_2) = 2x_1^2 + 4x_1x_2 + 5x_2^2 \quad (1)$$

とおきます. すると, 先ほどの問題は 2 変数関数 $f(x_1, x_2)$ が最小となる (x_1, x_2) を求める問題ですが, $\mathbf{x} = (x_1, x_2)$ とおくと, ベクトルを変数とする関数 $f(\mathbf{x})$ が最小となる \mathbf{x} を求める問題と考えることもできます. この問題自体は手計算で簡単に解けましたが, もっと一般的な問題, すなわち n 変数関数 $F(x_1, x_2, \dots, x_n)$, あるいは同じことですが n 次元ベクトル \mathbf{x} を変数とする関数 $F(\mathbf{x})$ が最小となる \mathbf{x} を求める問題は, 手計算ではそう簡単に解けそうにないことが想像できると思います. そこで, コンピュータを利用した解法が重要となってきます. しかし, コンピュータに問題を解かせるためには, その具体的な計算手順 (アルゴリズム) を人間が指示する必要があります. ここからは最適化のアルゴリズムを考えていきますが, 一般論を扱うには高校の学習内容を越えた予備知識が必要となってしまいますので, まずは解のわかっている問題 1 を詳しく考えていきましょう.

ところで, \mathbf{x} を, 原点を基点とした位置ベクトルとみなせば, \mathbf{x} を点と考えることもできます. すると, 関数 f は平面上の点 \mathbf{x} に対して値 $f(\mathbf{x})$ を割り当てるものであり, この値を「高さ」とみなせば, $f(\mathbf{x})$ が最小となる \mathbf{x} は, この関数によって定義される「地形」の「谷底」になります. $f(\mathbf{x})$ の値が等しくなるような点 \mathbf{x} を曲線で結ぶと, これはちょうど地図上の等高線に対応するものになります. 最適解 $(0, 0)$ のことを \mathbf{x}^* と書くことにして, 実際に \mathbf{x}^* が「谷底」になっていることを見るために, 図 1 において 4 本の楕円で示した f の「等高線」を観察してみましょう.

点 \mathbf{x} が与えられたとき, どの方向に進めば最も関数

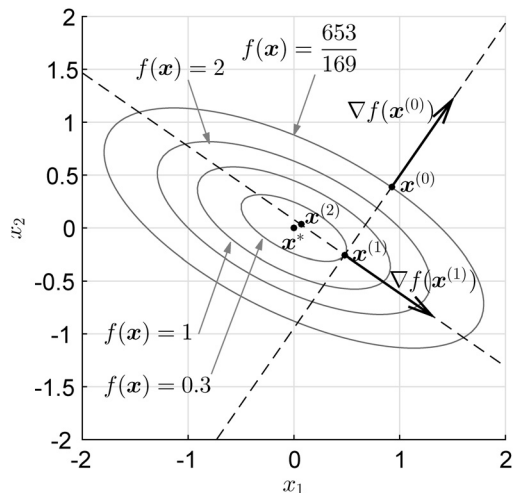


図 1 関数 f の等高線および \mathbb{R}^2 上の最急降下法により生成される点列

f の値が減少するでしょうか. 実際に自分が谷を下っていると考えると, とりあえず一番傾斜が急な下り坂の方向に進むのがよさそうだとと言えるでしょう. このことをもう少し数学的に考えてみましょう.

方向だけを考えたいので, ベクトルの長さは 1 に固定して考えます. 高校で習うベクトルの長さのことを標準ノルム (この記事ではこれを単にノルムと言います) と言い, \mathbf{a} のノルムを $\|\mathbf{a}\|$ と表します. つまり, $\mathbf{a} = (a_1, a_2)$ に対して $\|\mathbf{a}\| = \sqrt{a_1^2 + a_2^2}$ です.

さて, 点 \mathbf{x} からノルム 1 の任意のベクトル $\mathbf{d} = (d_1, d_2)$ の方向に進んだときの f の「方向微分係数」を計算するために, $g_d(t) = f(\mathbf{x} + t\mathbf{d})$ において $g'_d(0)$ を求めましょう. つまり, 点 \mathbf{x} から \mathbf{d} の方向に微小に進んだときの f の変化率を計算しようというわけです. \mathbf{d} はいったん固定しておいて, $g_d(t)$ は単なる t の 1 変数関数だとみなします. すると, $g_d(t) = f(x_1 + td_1, x_2 + td_2)$ より, 式 (1) に注意して計算を進めると,

$$g'_d(t) = (4(x_1 + td_1) + 4(x_2 + td_2))d_1 + (4(x_1 + td_1) + 10(x_2 + td_2))d_2 \quad (2)$$

となり, $g'_d(0) = (4x_1 + 4x_2)d_1 + (4x_1 + 10x_2)d_2$ とわかります. ここで, 2 変数関数 $f(x_1, x_2)$ において x_2 を定数とみなし, x_1 だけを変数だと考えたうえで f を x_1 について微分したものを f の x_1 に関する偏導関数といい, $\frac{\partial f}{\partial x_1}$ のように表します. この例では

$$\frac{\partial f}{\partial x_1}(\mathbf{x}) = 4x_1 + 4x_2, \quad \frac{\partial f}{\partial x_2}(\mathbf{x}) = 4x_1 + 10x_2 \quad (3)$$

ですから, 高校での記法と同様に内積の記号として「 \cdot 」

を用いると、実は

$$\begin{aligned} g'_a(0) &= \frac{\partial f}{\partial x_1}(\mathbf{x}) \times d_1 + \frac{\partial f}{\partial x_2}(\mathbf{x}) \times d_2 \\ &= \left(\frac{\partial f}{\partial x_1}(\mathbf{x}), \frac{\partial f}{\partial x_2}(\mathbf{x}) \right) \cdot (d_1, d_2) \\ &= \nabla f(\mathbf{x}) \cdot \mathbf{d} \end{aligned} \quad (4)$$

となります。ここで、 $\nabla f(\mathbf{x}) = \left(\frac{\partial f}{\partial x_1}(\mathbf{x}), \frac{\partial f}{\partial x_2}(\mathbf{x}) \right)$ とおきました。これを \mathbb{R}^2 での点 \mathbf{x} における f の勾配ベクトルと呼び、 ∇ は「ナブラ」と読みます。

式(4)より、ノルム1のベクトル \mathbf{d} が $\nabla f(\mathbf{x})$ とちょうど反対の方向を向いているとき、すなわち $\mathbf{d} = -\nabla f(\mathbf{x})/\|\nabla f(\mathbf{x})\|$ のとき、 $g'_a(0)$ が最小となり、この方向に進むと目的関数を最も小さくすることができるとわかりました。この意味で、 $-\nabla f(\mathbf{x})$ のことを f の \mathbf{x} における最急降下方向とも言います。したがって、最急降下方向 $-\nabla f(\mathbf{x})$ に進むことは、まさに、最も傾斜が急な下り坂の方向に進むことに他ならないのです。そこで、 \mathbf{x} から $\mathbf{d} = -\nabla f(\mathbf{x})/\|\nabla f(\mathbf{x})\|$ の方向に進んで、目的関数 f の値が $f(\mathbf{x})$ より小さくなる点を探そうにしましょう。このように、点 \mathbf{x} から次の点を探するための方向を \mathbf{x} における探索方向と言います。

しかし、 $g'_a(0)$ は微分係数ですから、点 \mathbf{x} のごく近くでは関数 f の値が最も小さくなる方向が $-\nabla f(\mathbf{x})$ だということに過ぎません。したがって、点 \mathbf{x} から $-\nabla f(\mathbf{x})$ の方向に進めば進むほど f がどんどん小さくなる、というわけではないことに注意しましょう。また、 \mathbf{x} から $-\nabla f(\mathbf{x})$ の方向に適切な距離だけ進んだとしても、一般には最適解が見つかるとは限りません。

それでは、実際にはどのくらい進めばよいのでしょうか？ 理想的には、1変数 t の関数 $g_a(t) = f(\mathbf{x} + t\mathbf{d})$ が最小となるような正の実数 t が見つければよいと言えるでしょう。このような t を探すプロセスを（正確な）直線探索と言います。ただし、現実的には $g_a(t)$ が「ある程度」小さくなるような $t > 0$ を見つければ上手くいく場合も多く、直線探索を必ずしも正確に行う必要はありません。直線探索の詳細について興味がある読者は[2]を参照してください。

いずれにせよ、このようにして決まる、点 \mathbf{x} から探索方向 \mathbf{d} の方向にどれくらい進むかを表す t のことをステップ幅と言ひ、次の点を $\mathbf{x} + t\mathbf{d}$ として計算することになります。ところで、探索方向を表すベクトルのノルムが変わっても、どれくらい進むかを表すステップ幅をそれに応じて変えれば同じことです。たとえば、 $-\nabla f(\mathbf{x})/\|\nabla f(\mathbf{x})\|$ の方向にステップ幅1だけ進むの

表1 問題1に対する \mathbb{R}^2 上の最急降下法の計算結果

k	$\ \mathbf{x}^{(k)}\ $	$f(\mathbf{x}^{(k)})$	$\ \nabla f(\mathbf{x}^{(k)})\ $
0	1.000	3.864	9.175
1	5.430×10^{-1}	2.957×10^{-1}	1.097
2	7.652×10^{-2}	2.262×10^{-2}	7.021×10^{-1}
5	3.179×10^{-3}	1.014×10^{-5}	6.422×10^{-3}
10	2.623×10^{-6}	2.659×10^{-11}	2.407×10^{-5}
15	8.339×10^{-9}	6.974×10^{-17}	1.685×10^{-8}

と、 $-\nabla f(\mathbf{x})$ の方向にステップ幅 $1/\|\nabla f(\mathbf{x})\|$ だけ進むのは同じです。つまり、探索方向を $-\nabla f(\mathbf{x})/\|\nabla f(\mathbf{x})\|$ としても $-\nabla f(\mathbf{x})$ としても考え方は同じですから、これ以降では探索方向は簡単に $\mathbf{d} = -\nabla f(\mathbf{x})$ とします。そして、ここまで述べてきたように、現在の点が \mathbf{x} のとき、探索方向を $\mathbf{d} = -\nabla f(\mathbf{x})$ として適切にステップ幅 t を決めて次の点 $\mathbf{x} + t\mathbf{d}$ を計算する、という操作を次々に反復するアルゴリズムのことを、最急降下法と言います。 k を0以上の整数として、第 k 反復目の点を $\mathbf{x}^{(k)}$ 、探索方向を $\mathbf{d}^{(k)}$ 、ステップ幅を $t^{(k)}$ と書くことにすると、最急降下法での点の更新式は

$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + t^{(k)}\mathbf{d}^{(k)} = \mathbf{x}^{(k)} - t^{(k)}\nabla f(\mathbf{x}^{(k)}) \quad (5)$$

となります。つまり、現在 $\mathbf{x}^{(k)}$ という点にいるとすると、ここから最も急な方向に下るという役割を $-\nabla f(\mathbf{x}^{(k)})$ が果たしており、この方向にどれくらい進むかを決めるのが $t^{(k)}$ です。最初は最も急な方向に下り始めたとしても、同じ方向に進み続けられればいずれは上り坂になるかもしれません。上り坂になるまで進んでは困りますので、点 $\mathbf{x}^{(k)}$ から方向 $-\nabla f(\mathbf{x}^{(k)})$ に進んで下り続けて、平坦になったところで止まるように $t^{(k)}$ を決めるのが正確な直線探索というわけです。

例として、初期点を $\mathbf{x}^{(0)} = (12/13, 5/13)$ とおき、コンピュータを用いて数値を計算した結果を表1に載せます。ステップ幅は、正確な直線探索によるものを用いました。ちなみに実際に手計算をすると、正確な直線探索では $t^{(k)} = -(\nabla f(\mathbf{x}^{(k)}) \cdot \mathbf{d}^{(k)})/(2f(\mathbf{d}^{(k)}))$ となることが確かめられます。是非確認してみてください。表1を見ると、最急降下法の反復を繰り返すと $\|\mathbf{x}^{(k)}\|$ が0に近づいているのがわかります。つまり、 $\mathbf{x}^{(k)}$ が最適解 $(0, 0)$ に近づいているということです。また、 $f(\mathbf{x}^{(k)})$ はもちろん最小値0に近づいています。さらに、 $\|\nabla f(\mathbf{x}^{(k)})\|$ も0に近づいていることがわかりますね。高校で学習するように、1変数関数が極小となる点においてはその微分係数が0となりますが、同様に、(制約条件なしの下で)多変数関数が極小となる点においてはその勾配が零ベクトルとなります。した

がって、ここの結果でもし $\|\nabla f(\mathbf{x}^{(k)})\|$ が 0 に近づいていなかったならば、プログラミングのミスなどにより最適化の計算が適切にできていないと考えられます。

また、実際にどのように $\mathbf{x}^{(k)}$ が計算されるのかを、 $k = 0, 1, 2$ まで図 1 に示しました。図では、勾配ベクトル $\nabla f(\mathbf{x}^{(k)})$ を矢印で表していますが、実際にはその反対方向である最急降下方向 $-\nabla f(\mathbf{x}^{(k)})$ に直線探索をすることに注意してください。また、図中の 653/169 という値は初期点における目的関数値 $f(\mathbf{x}^{(0)})$ です。

2.2 一般化

問題 1 を一般化すると、次のような一般の多変数の 2 次関数の最小化問題が得られます。

問題 2. $n^2 + n + 1$ 個の実数の定数 $a_{1,1}, a_{1,2}, \dots, a_{1,n}, a_{2,1}, a_{2,2}, \dots, a_{2,n}, \dots, a_{n,1}, a_{n,2}, \dots, a_{n,n}, b_1, b_2, \dots, b_n, c$ に対して、 n 変数関数 $F(x_1, x_2, \dots, x_n) = \sum_{k=1}^n \sum_{l=1}^n a_{k,l} x_k x_l + \sum_{m=1}^n b_m x_m + c$ が最小となる実数の組 (x_1, x_2, \dots, x_n) を求めよ。

もちろんこの問題も、上に述べた最急降下法で解くことができます。しかし、最急降下法は考え方が自然でわかりやすいものの、最適解に近づくまでに反復回数を比較的多く要し、あまり実用的ではありません。そこで、最急降下法を改良したアルゴリズムに共役勾配法というものがあり、こちらは最適解への収束速度が速く、現在でも盛んに研究が行われています。そして、2 次項の係数 $a_{k,l}$ がある条件を満たすとき、問題 2 の最適解はある n 元連立 1 次方程式の解と一致することが知られています [2]。 n 元連立 1 次方程式はありとあらゆる分野で現れる重要な問題ですが、 n が大きくなるとそう簡単には解けません。そこで、実用的な解法として、問題 2 に共役勾配法を適用するというものがあり、実際によく用いられています。2 次関数の最小化より連立 1 次方程式を解くほうが簡単に思えるかもしれませんが、コンピュータにとっては必ずしもそうではないというのが面白いですね。

3. 「曲がった空間」での最適化

それではいよいよ「曲がった空間」での最適化に進みましょう。先ほどと同じく、まずは簡単な問題から考えることにします。

3.1 具体的な問題

問題 3. $x^2 + y^2 = 1$ のとき、 $2x^2 + 4xy + 5y^2$ が最小となる実数の組 (x, y) を求めよ。

この問題も高校の数学の知識で解けますね。実際、条

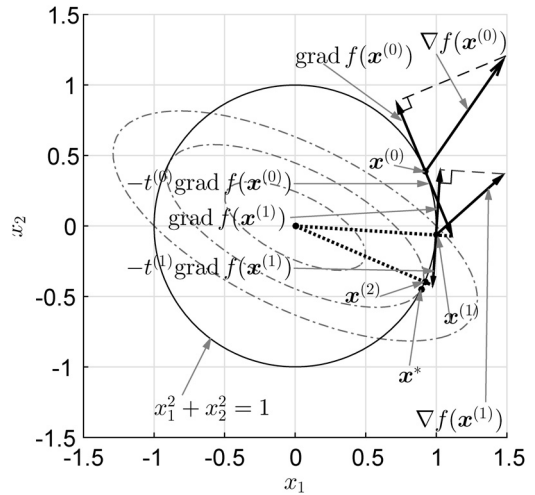


図 2 関数 f の単位円 S^1 における勾配および S^1 上の最急降下法により生成される点列

件 $x^2 + y^2 = 1$ から、 (x, y) は単位円上の点と考えられるので、 $x = \cos \theta$, $y = \sin \theta$ ($0 \leq \theta < 2\pi$) とおくことができます。すると、 $2x^2 + 4xy + 5y^2$ は 1 変数 θ の関数となるので、これを $h(\theta)$ とおくと、

$$\begin{aligned} h(\theta) &= 2 \cos^2 \theta + 4 \cos \theta \sin \theta + 5 \sin^2 \theta \\ &= 2 \sin 2\theta - \frac{3}{2} \cos 2\theta + \frac{7}{2} \\ &= \frac{5}{2} \sin(2\theta + \alpha) + \frac{7}{2} \end{aligned} \quad (6)$$

となり、ここで、 α は $-\pi/2 < \alpha < 0$ かつ $\tan \alpha = -3/4$ を満たす角です。 $-1 \leq \sin(2\theta + \alpha) \leq 1$ より、 $\sin(2\theta + \alpha) = -1$ のとき $h(\theta)$ は最小値 1 を取ります。このとき $\cos \theta$ と $\sin \theta$ を計算することで、元の問題の解は、 $(x, y) = (2/\sqrt{5}, -1/\sqrt{5}), (-2/\sqrt{5}, 1/\sqrt{5})$ であると求まります。

このような問題は制約条件つき最適化問題と呼ばれ、さまざまな最適化アルゴリズムが研究されています [3]。この記事では、通常の「まっすぐな空間」での制約条件付きの問題を、「曲がった空間」での制約条件なしの最適化問題とみなす比較的新しい考え方を紹介します。それはリーマン多様体上の最適化と呼ばれる分野で、特に 2000 年代に入ってから盛んに研究されています。

問題 3 も解が手計算で求められましたが、より一般的な問題をコンピュータに解かせるためのアルゴリズムを導出するために、この問題を詳しく見ていきましょう。やはり (x, y) の代わりに $\mathbf{x} = (x_1, x_2)$ という記号を用いることにし、関数 f は先ほどと同じものとします。問題 1 で見たように、 $\nabla f(\mathbf{x}) = (4x_1 + 4x_2, 4x_1 + 10x_2)$

となりますが、先ほどと同様に $-\nabla f(\mathbf{x})$ の方向に直線探索を行うと、明らかに制約条件が満たされなくなります。実は、上述の解答のように、 \mathbf{x} が単位円という「曲がった空間」に存在するという考え方が、これから述べるアルゴリズムの鍵となります。この場合は「曲がった空間」というより「曲がった線（曲線）」と言うほうがしっくりくるかもしれません。

問題 1 では平面 \mathbb{R}^2 全体で解を探しましたが、ここでは $x_1^2 + x_2^2 = 1$ を満たす点全体、すなわち単位円 $S^1 = \{\mathbf{x} \in \mathbb{R}^2 \mid \|\mathbf{x}\| = 1\}$ 上で解を探すのだと考えましょう。そこで、まず $\nabla f(\mathbf{x})$ は一般に単位円からはみ出た方向を向いていますから、せめて接線方向を向くように射影しましょう。点 \mathbf{x} を始点とするベクトル $\nabla f(\mathbf{x})$ を \mathbf{x} における S^1 の接線に直交射影（正射影）すると、 $\nabla f(\mathbf{x}) - (\mathbf{x} \cdot \nabla f(\mathbf{x}))\mathbf{x}$ となります。これを、 S^1 上の点 \mathbf{x} における f の勾配と言い、 $\text{grad } f(\mathbf{x})$ と書くことにします（図 2）。勾配という名称は ∇f と共通しますが、平面 \mathbb{R}^2 上で考えるか単位円 S^1 上で考えるかというところが異なります。なお、 grad は勾配を表す英単語 *gradient* の略称です。このように点 \mathbf{x} における S^1 の接線方向を向いたベクトルを、 S^1 の \mathbf{x} における接ベクトルと呼びます。

さて、問題 1 と同じように $-\text{grad } f(\mathbf{x})$ を探索方向 \mathbf{d} とします。この探索方向は \mathbf{x} において S^1 の接線方向を向いていますから、 $-\nabla f(\mathbf{x})$ に比べると「単位円に沿って探そう」という気持ちが強く表れていると言えるでしょう。しかし、それでもなお、適切なステップ幅 t が見つかった場合に次の点を $\mathbf{x} + t\mathbf{d}$ とすることはできません。これもまた単位円 S^1 からはみ出てしまうためです。そこで、さらにこのはみ出た点を S^1 に引き戻す必要があります。そのための方法は色々ありますが、ここでは最も簡単なものを紹介しましょう。すなわち、はみ出た点 $\mathbf{x} + t\mathbf{d}$ をそれ自身のノルムで割って $(\mathbf{x} + t\mathbf{d})/\|\mathbf{x} + t\mathbf{d}\|$ とすることで、ノルムを 1 にして S^1 に乗せてしまうというものです。幾何学的に言えば、はみ出た点 $\mathbf{x} + t\mathbf{d}$ から原点 $(0, 0)$ に線分を引いて、単位円 S^1 との交点を次の点にすることです。この様子も図 2 に示しています。

以上のことから、 S^1 上でのステップ幅 $t > 0$ を決めるには、直線 $\mathbf{x} + t\mathbf{d}$ の代わりに単位円 S^1 上の曲線 $(\mathbf{x} + t\mathbf{d})/\|\mathbf{x} + t\mathbf{d}\|$ を用いて、いわば、直線探索ならぬ「曲線探索」を行うこととなります。

これで準備は整いました。まとめる前に記号を導入しておきます。 \mathbb{R}^2 上の任意のベクトル \mathbf{b} の、 S^1 上の点 \mathbf{x} における接線への直交射影を $P_{\mathbf{x}}(\mathbf{b})$ とし、点 \mathbf{x} に

表 2 問題 3 に対する S^1 上の最急降下法の計算結果

k	$\ \mathbf{x}^{(k)} - \mathbf{x}^*\ $	$f(\mathbf{x}^{(k)}) - 1$	$\ \text{grad } f(\mathbf{x}^{(k)})\ $
0	8.323×10^{-1}	2.864	4.947
1	3.964×10^{-1}	7.549×10^{-1}	3.580
2	5.525×10^{-2}	1.525×10^{-2}	5.514×10^{-1}
3	1.682×10^{-4}	1.415×10^{-7}	1.682×10^{-3}
4	4.759×10^{-12}	0	4.758×10^{-11}

における S^1 の接ベクトル \mathbf{c} に対して $\mathbf{x} + \mathbf{c}$ を S^1 に引き戻した点を $R_{\mathbf{x}}(\mathbf{c})$ と書くことにしましょう。つまり、

$$P_{\mathbf{x}}(\mathbf{b}) = \mathbf{b} - (\mathbf{x} \cdot \mathbf{b})\mathbf{x}, \quad R_{\mathbf{x}}(\mathbf{c}) = \frac{\mathbf{x} + \mathbf{c}}{\|\mathbf{x} + \mathbf{c}\|} \quad (7)$$

です。すると、 S^1 上の f の勾配 $\text{grad } f(\mathbf{x})$ は

$$\text{grad } f(\mathbf{x}) = P_{\mathbf{x}}(\nabla f(\mathbf{x})) \quad (8)$$

と書けること、および、「曲線探索」の際には $f(R_{\mathbf{x}}(t\mathbf{d}))$ の値を小さくするようなステップ幅 $t > 0$ を探すことになるということに注意してください。さらに、 S^1 上での関数 f が極小となる \mathbf{x} では $\text{grad } f(\mathbf{x})$ が 0 になることが知られています。そこで、 $\text{grad } f(\mathbf{x})$ が 0 になった時点で計算を停止すればよいわけです。ただし、コンピュータによる数値計算には誤差がありますから、実際には $\|\text{grad } f(\mathbf{x})\|$ が十分小さくなったら（たとえば 10^{-6} 未満）計算を停止します。こうして単位円 S^1 上の最急降下法を記述することができます。

S^1 上の目的関数 f に対する最急降下法

手順 0：初期点 $\mathbf{x}^{(0)} \in S^1$ を選び、 $k = 0$ とする。

手順 1： $\|\text{grad } f(\mathbf{x}^{(k)})\|$ が十分小さければ計算を終了し、そうでなければ $\mathbf{d}^{(k)} = -\text{grad } f(\mathbf{x}^{(k)})$ とする。

手順 2： $t^{(k)} > 0$ を $f(R_{\mathbf{x}^{(k)}}(t\mathbf{d}^{(k)}))$ が（近似的に）最小となるような t として計算し、次の点 $\mathbf{x}^{(k+1)} = R_{\mathbf{x}^{(k)}}(t^{(k)}\mathbf{d}^{(k)})$ を計算する。 k の値を 1 増やして手順 1 へ戻る。

最適解の一方 $(2/\sqrt{5}, -1/\sqrt{5})$ を \mathbf{x}^* とおきます。 S^1 上の最急降下法を用いて実際に数値計算をした結果が表 2 です。ただし、話の都合上、ステップ幅は常に $t^{(k)} = 0.1$ と固定しました。後述するように S^1 は 1 次元なので、正確な探索を行うといきなり最適解が得られてしまうためです。ここまで考え方をわかりやすく説明するために単位円 S^1 上で議論しましたが、本来このような方法はより高次元の問題に対して用いるほうが真価を発揮します。ともかく、表 2 を見ると、 $\mathbf{x}^{(k)}$ がどんどん最適解 \mathbf{x}^* に近づき、勾配のノルムも 0 に近づいているため、最適化が上手く行えていることがわかります。また、目的関数の値は最小値 $f(\mathbf{x}^*) = 1$

に近づいていることがわかります。なお、 $f(\mathbf{x}^{(4)}) - 1$ が 0 となっていますが、これはあくまでもコンピュータの精度では 0 と区別できないほど 0 に近いということです。

3.2 一般化

問題 4. $\sum_{i=1}^n x_i^2 = 1$ のとき、 n^2 個の実数の定数 $a_{1,1}, a_{1,2}, \dots, a_{1,n}, \dots, a_{n,1}, a_{n,2}, \dots, a_{n,n}$ に対して n 変数関数 $G(x_1, x_2, \dots, x_n) = \sum_{k=1}^n \sum_{l=1}^n a_{k,l} x_k x_l$ が最小となる実数の組 (x_1, x_2, \dots, x_n) を求めよ。

問題 3 で単位円 $S^1 = \{(x_1, x_2) \in \mathbb{R}^2 \mid x_1^2 + x_2^2 = 1\}$ を考えたのと同様に、集合 $S^{n-1} = \{(x_1, x_2, \dots, x_n) \in \mathbb{R}^n \mid x_1^2 + x_2^2 + \dots + x_n^2 = 1\}$ を考えると、問題 4 に対しても同様にアルゴリズムを導出することができます。そして、この集合 S^{n-1} のことを、 $(n-1)$ 次元超球面（あるいは単に球面）と言います。 \mathbb{R}^n の 1 点 (x_1, x_2, \dots, x_n) を定めるには n 個の成分 x_1, x_2, \dots, x_n の値を定める必要がありますが、 S^{n-1} 上の点 (x_1, x_2, \dots, x_n) については、 $n-1$ 個の成分を定めると関係式 $x_1^2 + x_2^2 + \dots + x_n^2 = 1$ から残りの成分は（符号を除いて）自動的に定まります。これが S^{n-1} の次元が $n-1$ であることの説明です。あるいは、 S^{n-1} 上の点 (x_1, x_2, \dots, x_n) の成分は $n-1$ 個が自由に動けるので、自由度が $n-1$ であるとも言えます。 \mathbb{R}^2 内の円を S^1 と書くことを不思議に感じた読者もいるかもしれませんが、このような事情があったわけです。 S^1 上の点 (x_1, x_2) は一方の成分 x_1 を定めると関係式 $x_1^2 + x_2^2 = 1$ から他方の成分 x_2 が $x_2 = \pm\sqrt{1-x_1^2}$ と定まります。したがって、 S^1 は 1 次元、あるいは自由度が 1 であると言えます。これが、 $(x_1, x_2) \in S^1$ を $x_1 = \cos \alpha, x_2 = \sin \alpha$ と一つの変数 α で表せることにも関係しているのです。同様に、 \mathbb{R}^3 の中のいわゆる普通の球面 S^2 は 2 次元球面です。

ここまで見てきた（超）球面のように、「曲がった空間」を抽象化した概念を多様体と呼びます。また、多様体 M 上の各点 x に対して、 x における接ベクトル全体の集合を x における接空間と呼び、各接空間に内積が適切に定められているとき、 M をリーマン多様体と呼びます。これらの用語の正確な定義はここでは述べませんが、要するに、リーマン多様体とは曲面を一般化した概念です。この記事のタイトルにある「曲がった空間」はリーマン多様体のことを指していたのです。そこで、より一般的な次の問題が考えられます。

問題 5. リーマン多様体 M 上の関数 f を最小化せよ。

この問題では、 M は球面に限りませんし、目的関数

f は問題 1 や 3 における 2 次関数を指すわけではなく、より一般の関数です。もちろん M や f が具体的に与えられないと解きようがありませんが、解法アルゴリズムは S^1 の場合と同様に考えることができます。

S^1 のときと同様、ユークリッド空間の中で適切に定義されたリーマン多様体 M では、その上の関数に対してユークリッド空間における勾配を接空間に射影することで、 M 上の勾配が得られます。また、式 (7) のように、 M からはみ出た点を M 上に引き戻す働きをする R をレトラクション (retraction) と呼びます。ちなみに、retract にはまさに「引き戻す」という意味があります。こうした道具を使えば、問題 5 に対する一般の最急降下法も同様に導出できます。さらに、最近ではリーマン多様体上の共役勾配法も研究が進んでいます。

4. 終わりに

「まっすぐな空間（ユークリッド空間）での最適化」から始めて、その考えを基礎にして「曲がった空間（リーマン多様体）での最適化」を紹介してきました。ユークリッド空間での話はリーマン多様体での話の特殊な場合ですが、ユークリッド空間でのさまざまな研究が積み重ねられてきたからこそ、リーマン多様体での議論への拡張が可能になったということは重要です。

最後に、「曲がった空間での最適化」、すなわちリーマン多様体上の最適化の応用例を紹介します。まず、問題 4 は行列の固有値問題という重要な問題と密接な関係があります。たとえば Google で語句を検索すると、重要と判断されたページから順に並んだ結果が出てきますが、ここでも固有値問題が応用されています。また、統計学において膨大なデータから重要な情報を取り出したり、制御工学において制御モデルの次元（制御対象の状態変数の個数）を小さくするためにもさまざまな最適化問題を解く必要があるのですが、これらの中にも問題 5 の形で定式化できるものが多くあり、研究が盛んに行われています。

謝辞 本稿の執筆の機会を与えてくださり、原稿に対しても多数の有益なコメントをくださった宮代隆平先生に、厚く御礼申し上げます。

参考文献

- [1] 佐武一郎、『線形代数』、共立出版、1997.
- [2] 矢部博、『工学基礎 最適化とその応用』、数理工学社、2006.
- [3] 福島雅夫、『新版 数理計画入門』、朝倉書店、2011.