システム制御最適化特論

前期後半 月 5, 6限 14:00-16:10 5号館 第16講義室

担当:平田 健太郎

8/5 第8回 非線形最適化

講義日程（予定）

6/17	第1回	最適化問題と線形計画法（LP）
6/24	第2回	内点法
7/1	第3回	最短経路問題と動的計画法（DP）
7/8	第4回	最適制御
7/18*	第5回	二次計画法(QP)とモデル予測制御(MPC)
7/22	第6回	凸解析と線形行列不等式
7/29	第7回	線形行列不等式(LMI)による制御系解析・設計
8/5	第8回	非線形最適化

* irregular

局所最適性 vs. 大域的最適性

最適化問題： min $f(x)$ s.t. $x \in S$ において，点 $x^* \in S$ とその近傍 $N(x^*) = \{x \in S \mid \|x - x^*\| < \delta\}$ に対して

$$f(x^*) \leq f(x), \ \forall x \in N(x^*)$$

が成り立つとき x^* を局所最適解という．

$$f(x^*) \leq f(x), \ \forall x \in S$$

が成り立つとき x^* を大域的最適解という．

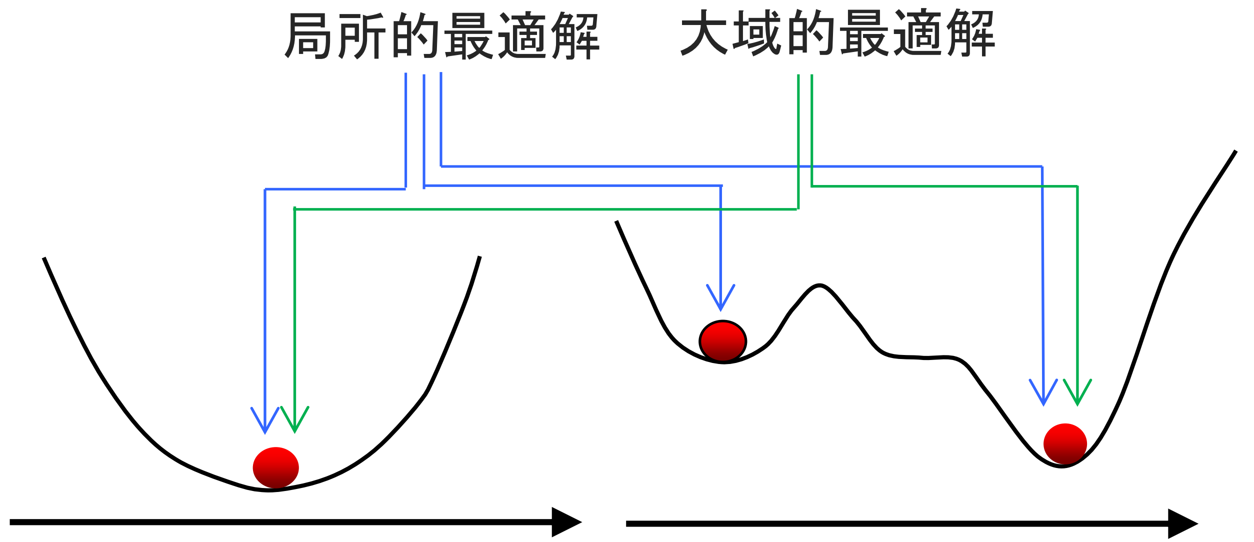
目的関数が凸であれば，局所最適→大域的最適

2次の評価関数 $J(x)$ を用いるメリットは何か？
(ex. 最小二乗法や最適制御)

- $J'(x) = 0$ が線形方程式になる.
- 凸なので, その解が大域的最適解

評価関数が高次の(非線形)関数であるとき, どうすべきか？

- とにかく局所最適解を見つける. (「bestな解」であるかどうかは運任せ, あるいは探索の仕方に工夫をこらす.)

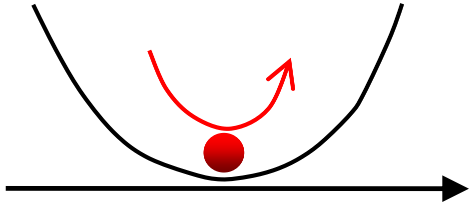


$f(x)$: 2次 or convex　　$f(x)$: highly nonlinear

最適性条件が解析的に解けない/分からない場合の局所最適化を考える.

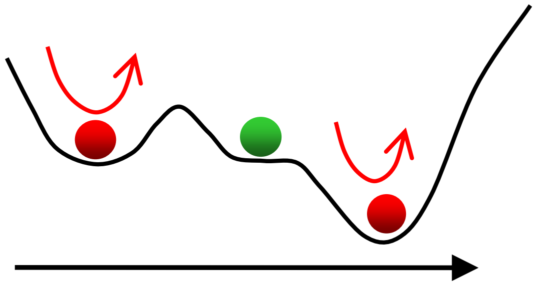
目的関数は微分可能とし, 順次, 高次の微分情報を活用する.

局所的最適性条件　(スカラ関数の場合)



停留点であるための条件

$$\left.\frac{\partial f(x)}{\partial x}\right|_{x=x^*} = 0$$



極小点であるための条件（上に加えて）

$$\left.\frac{\partial^2 f(x)}{\partial x^2}\right|_{x=x^*} > 0$$

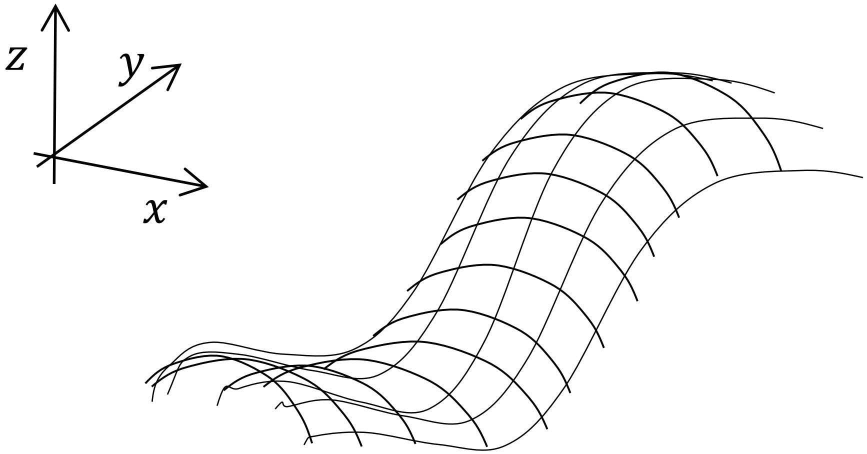
ベクトル値関数へ拡張するために

勾配 (gradient) ベクトル

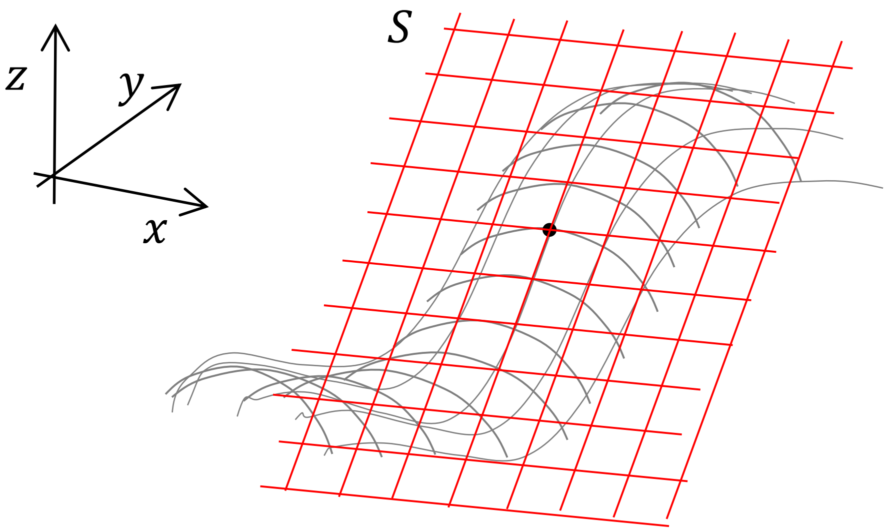
$$\nabla f(x^*) := \left.\frac{\partial f(x)}{\partial x}\right|_{x=x^*} := \begin{bmatrix} \frac{\partial f(x)}{\partial x_1} \\ \vdots \\ \frac{\partial f(x)}{\partial x_n} \end{bmatrix}_{x=x^*}$$

ヘッセ行列 (Hessian)

$$\nabla^2 f(x^*) := \left.\frac{\partial^2 f(x)}{\partial x^2}\right|_{x=x^*} = \begin{bmatrix} \frac{\partial^2 f(x)}{\partial x_1^2} & \frac{\partial^2 f(x)}{\partial x_1 \partial x_2} & \cdots \\ \frac{\partial^2 f(x)}{\partial x_2 \partial x_1} & \ddots & \\ \vdots & & \end{bmatrix}_{x=x^*}$$



曲面: $z = f(x, y)$

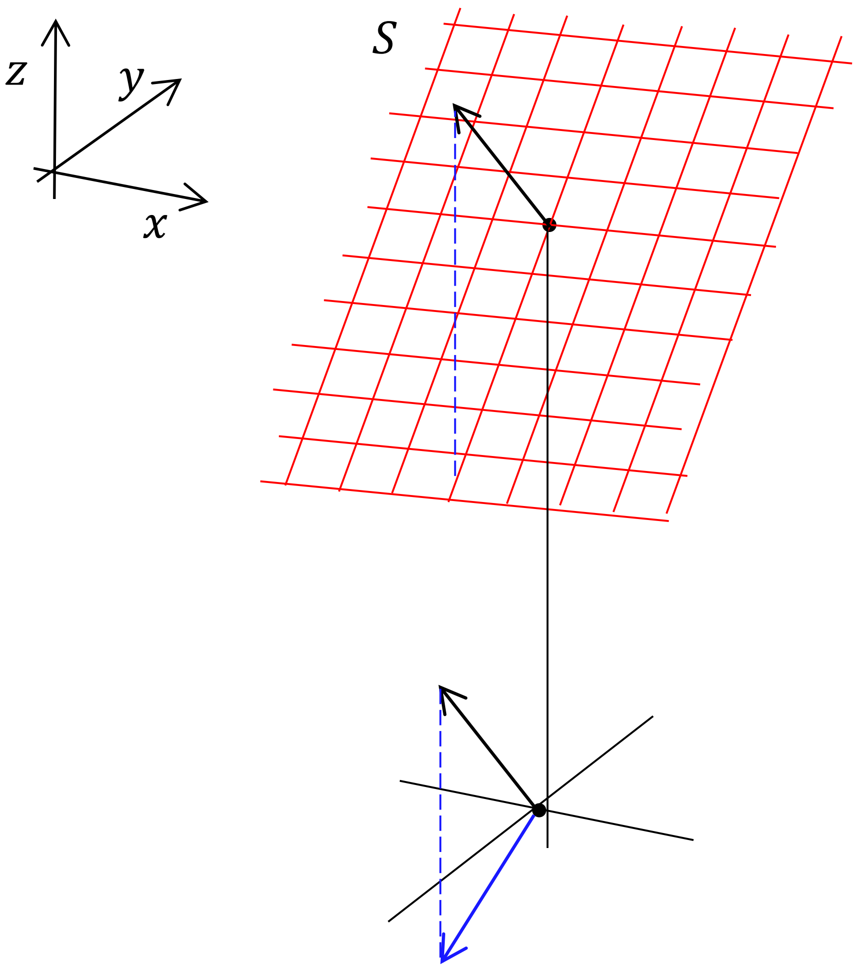


曲面上のある点(x_0, y_0)
における接平面S:

$$z - z_0 = a(x - x_0) + b(y - y_0)$$

$$a = \frac{\partial f}{\partial x}(x_0, y_0),$$

$$b = \frac{\partial f}{\partial y}(x_0, y_0)$$



平面の式を標準の形にすると

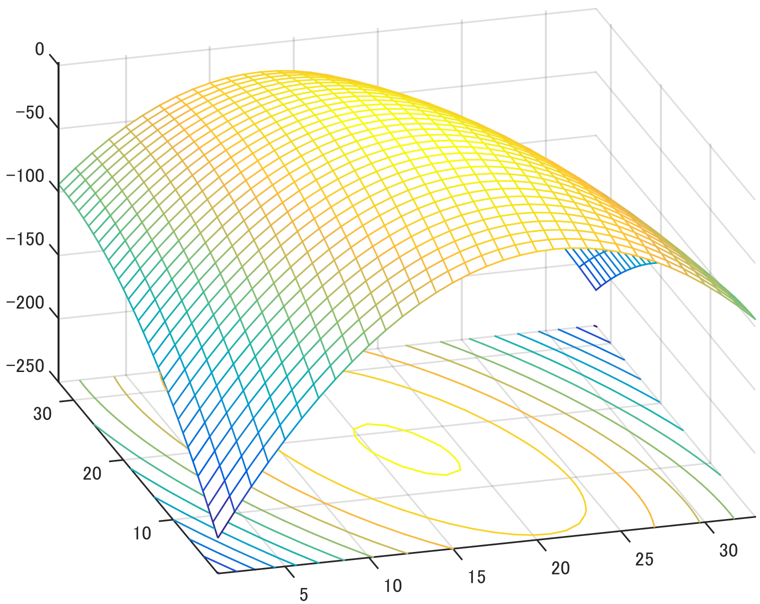
$$(z - z_0) - a(x - x_0) - b(y - y_0) = 0$$

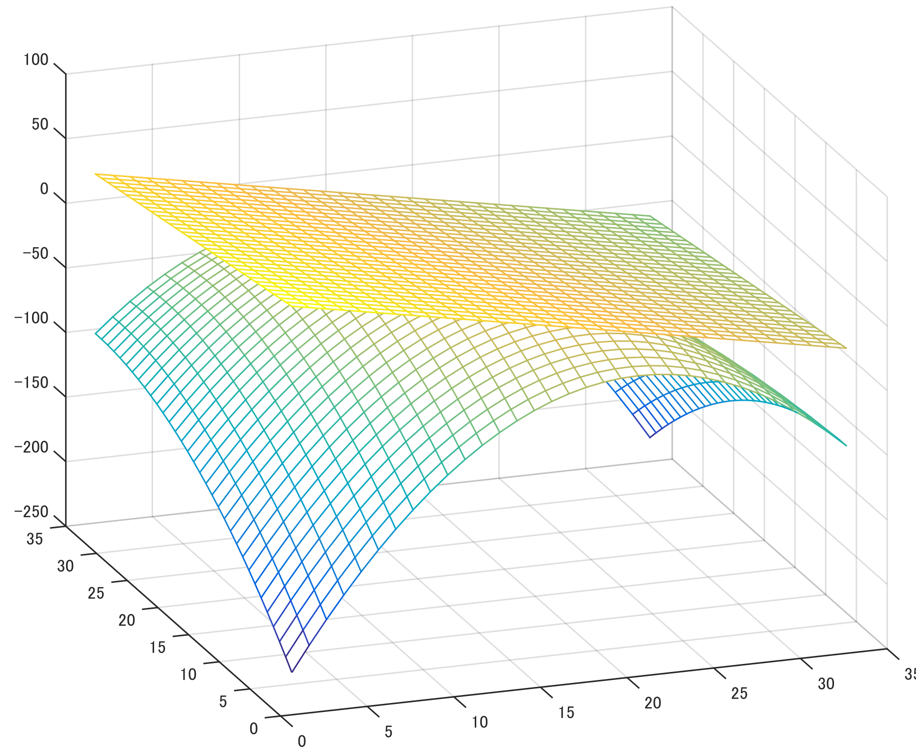
法線ベクトル　$\boldsymbol{a} = (-a, -b, 1)$

法線ベクトルの(x, y)平面への射影
$(-a, -b)$は, (x, y)平面において
平面 S のz座標値が最も減少する
方向を示している.

$$(a, b) = \left(\frac{\partial f}{\partial x}(x_0, y_0), \frac{\partial f}{\partial y}(x_0, y_0) \right)$$

は, (x, y)平面において, $z = f(x, y)$
の値が, 局所的に最も増加する方向
(勾配方向)を示している.

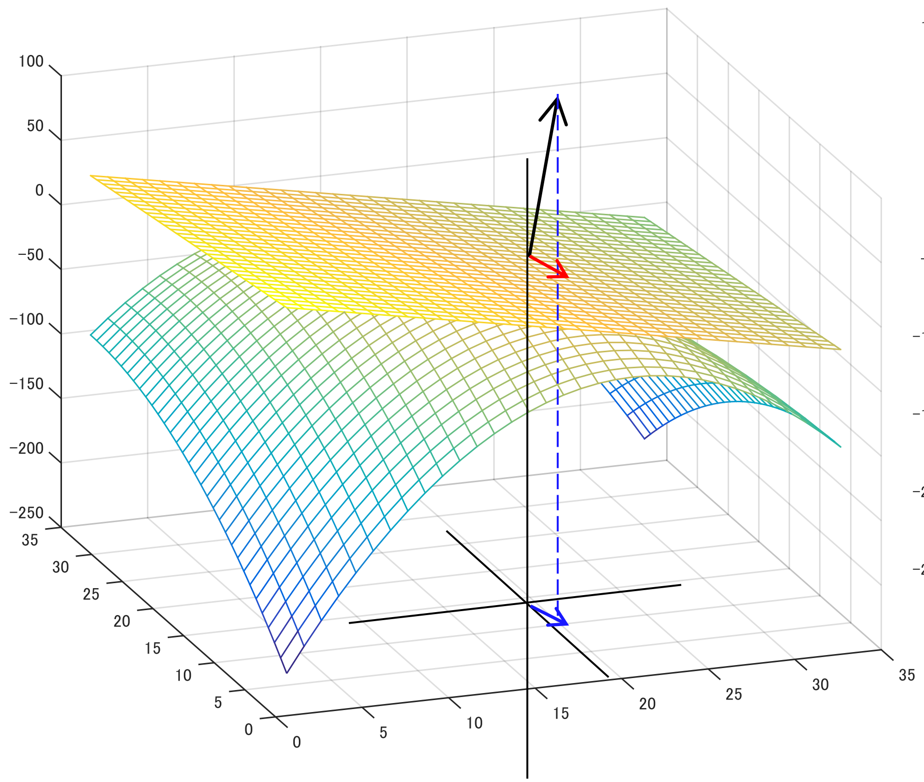


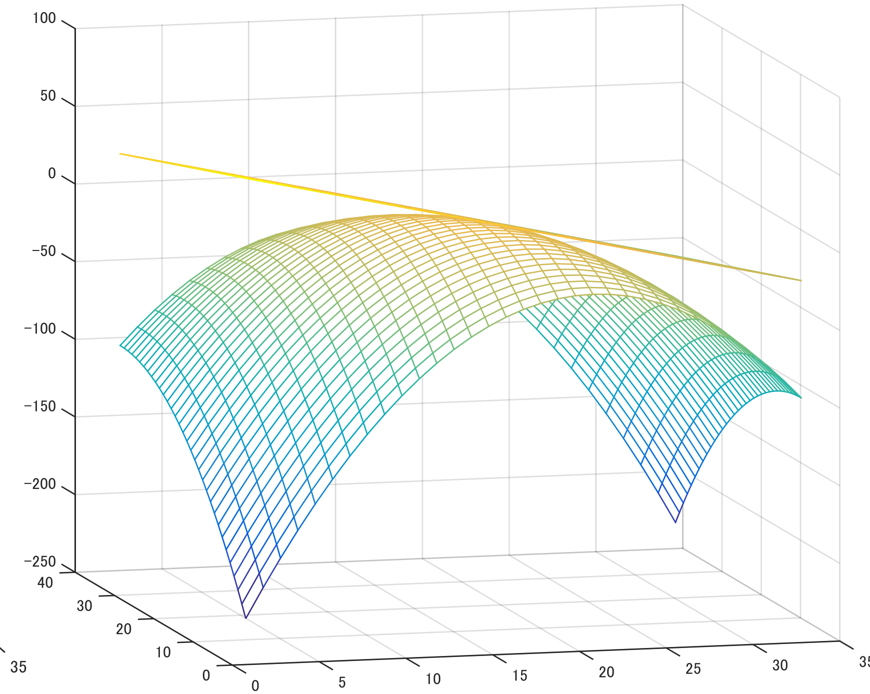


$$z = f(x, y) = 2 - (2x^2 + xy + \frac{1}{2}y^2)$$

$(x, y) = (1,1)$ における接平面

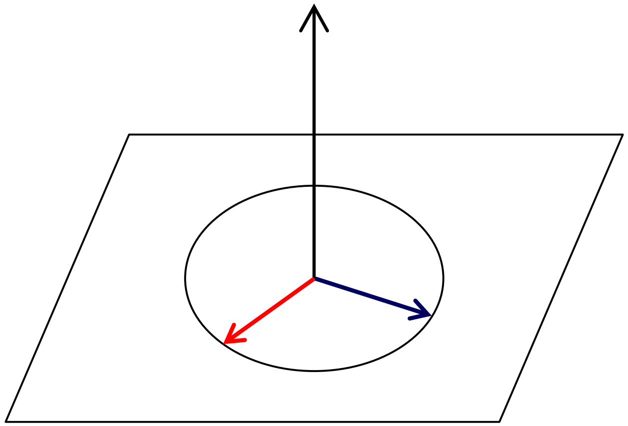
$$z = -5x - 2y + \frac{11}{2}$$

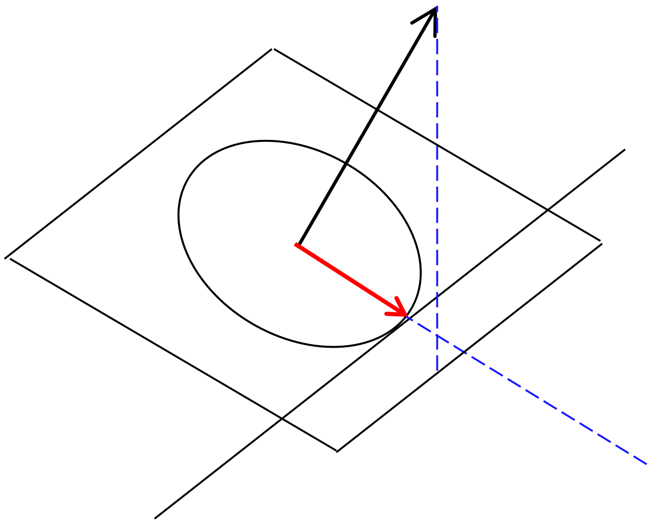




$(x, y) = (1,1)$ における接平面

$$z = -5x - 2y + \frac{11}{2}$$





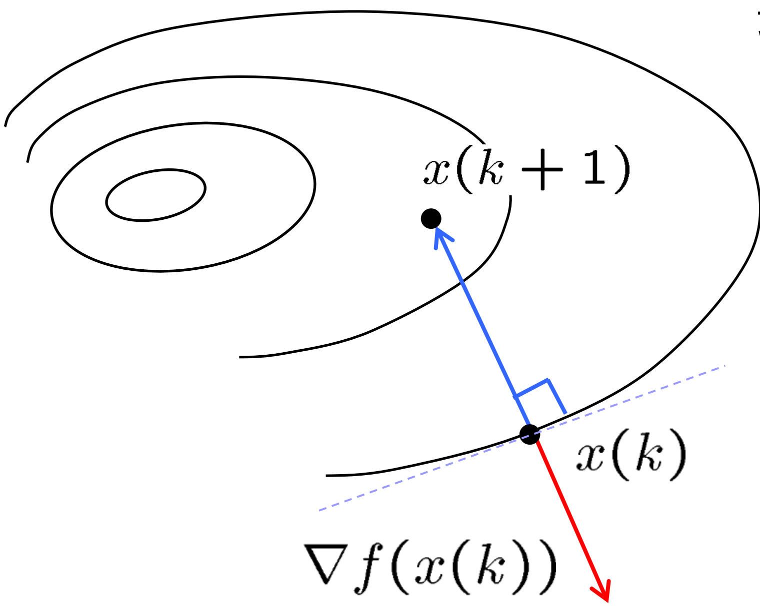
法線ベクトルが z 軸方向に
向いていれば, 単位円周上の
どちらに動いても, z 座標値は
変化しない

法線ベクトルが z 軸方向から
ずれているとき, 単位円周上の
z 座標が最も低い点が決まる.
その方向は法線ベクトルを
平面へ射影した方向である.

■ 最急降下法 (steepest descent method)

勾配が既知のとき　現在の場所から, 目的関数値が最も減少する方向を選んで前進することを繰り返せばよい

各点での勾配 $\nabla f(x)$ は目的関数の値が最も大きく増加する方向

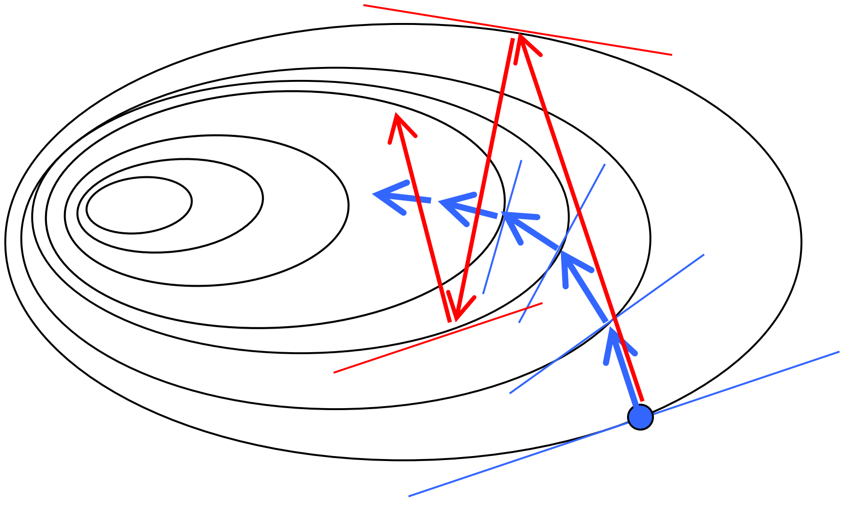


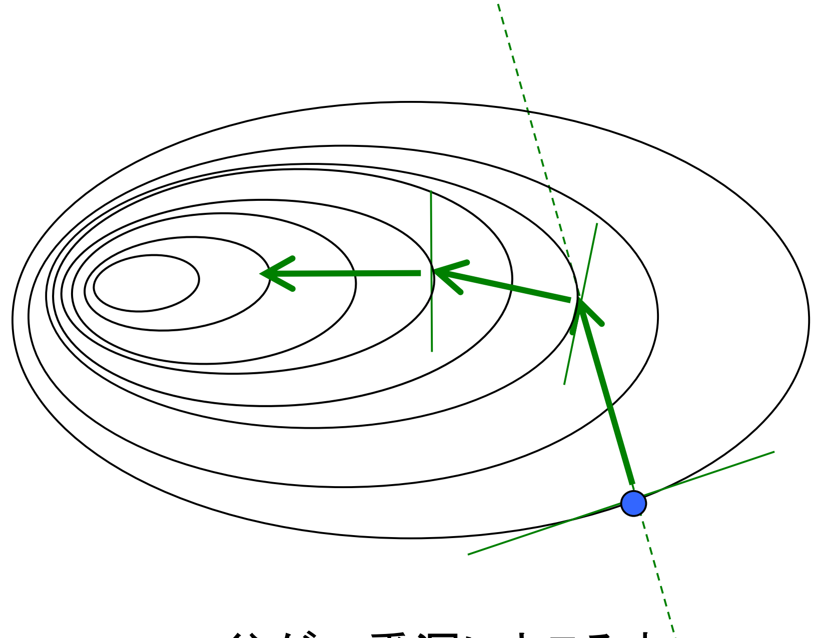
最も減少するのは, その反対方向

$$d_k = -\nabla f(x(k)),$$
$$x(k+1) = x(k) + \alpha_k d_k$$

勾配ベクトルと等高線は直交

◆ どれだけ進むか？





ステップ幅が小さいと繰り返し回数が多くなる

ステップ幅が大きいと不正確になる

谷が一番深いところまで進むのがよい

ステップ幅 α_k は

$$f(x(k)+\alpha_k d_k) \simeq \min_{\alpha \geq 0} f(x(k)+\alpha d_k)$$

となるよう決定　(直線探索)

停止条件:

$$\|\nabla f(x(k))\| < \epsilon$$

最急降下法のalgorithm

1. Choose $x(0)$. Set $k := 0$.
2. Compute the value of $\nabla f(x)$ at $x(k)$.
   Let $d_k := -\nabla f(x(k))$. Stop if $\|d_k\| < \epsilon$.
3. Choose step size α_k via a line search, i.e.,

$$\alpha_k = \arg\min_{\alpha \geq 0} f(x(k) + \alpha d_k)$$

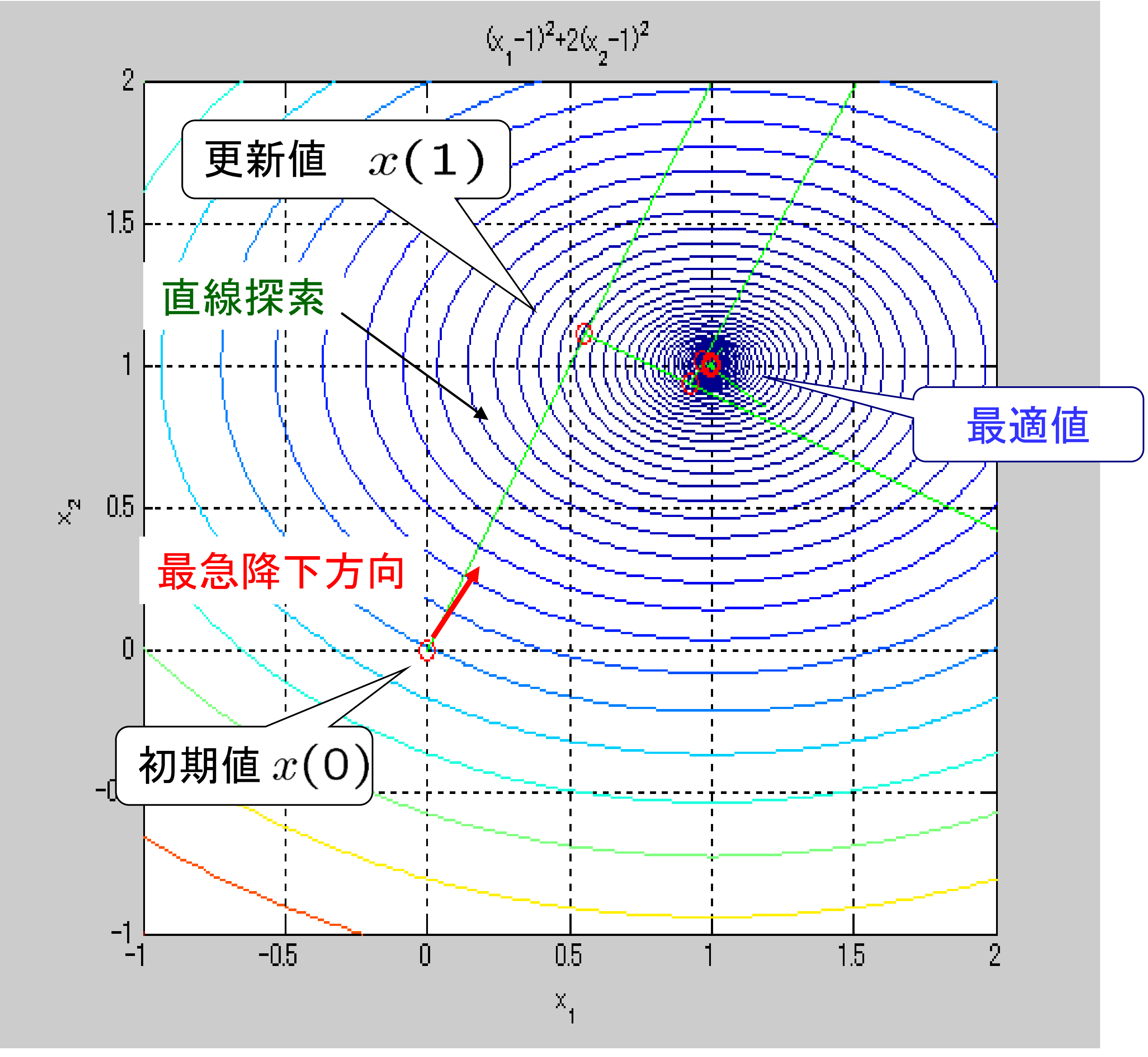
4. Update x by $x(k+1) = x(k) + \alpha_k d_k$.
5. Set $k := k + 1$. Go back to 2.

■ 例題 1　　$\min f(x)$

$$x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \quad f(x) = (x_1-1)^2+2(x_2-1)^2, \quad x^* = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

k	x_1	x_2	$f(x(k))$	$\|\nabla f(x(k))\|$
1	0.00000	0.00000	3.00000E+00	4.47214E+00
2	0.55284	1.10567	2.22289E−01	9.89187E−01
3	0.92351	0.93048	1.55172E−02	3.17387E−01
4	0.96820	1.01171	1.28569E−03	7.89917E−02
5	0.99183	0.99431	1.31580E−04	2.80303E−02
6	0.99673	1.00114	1.32717E−05	7.96437E−03
7	0.99923	0.99940	1.31604E−06	2.85385E−03
8	0.99967	1.00009	1.23634E−07	7.51949E−04

停止条件:
$\|\nabla f(x(k))\| < 10^{-3}$

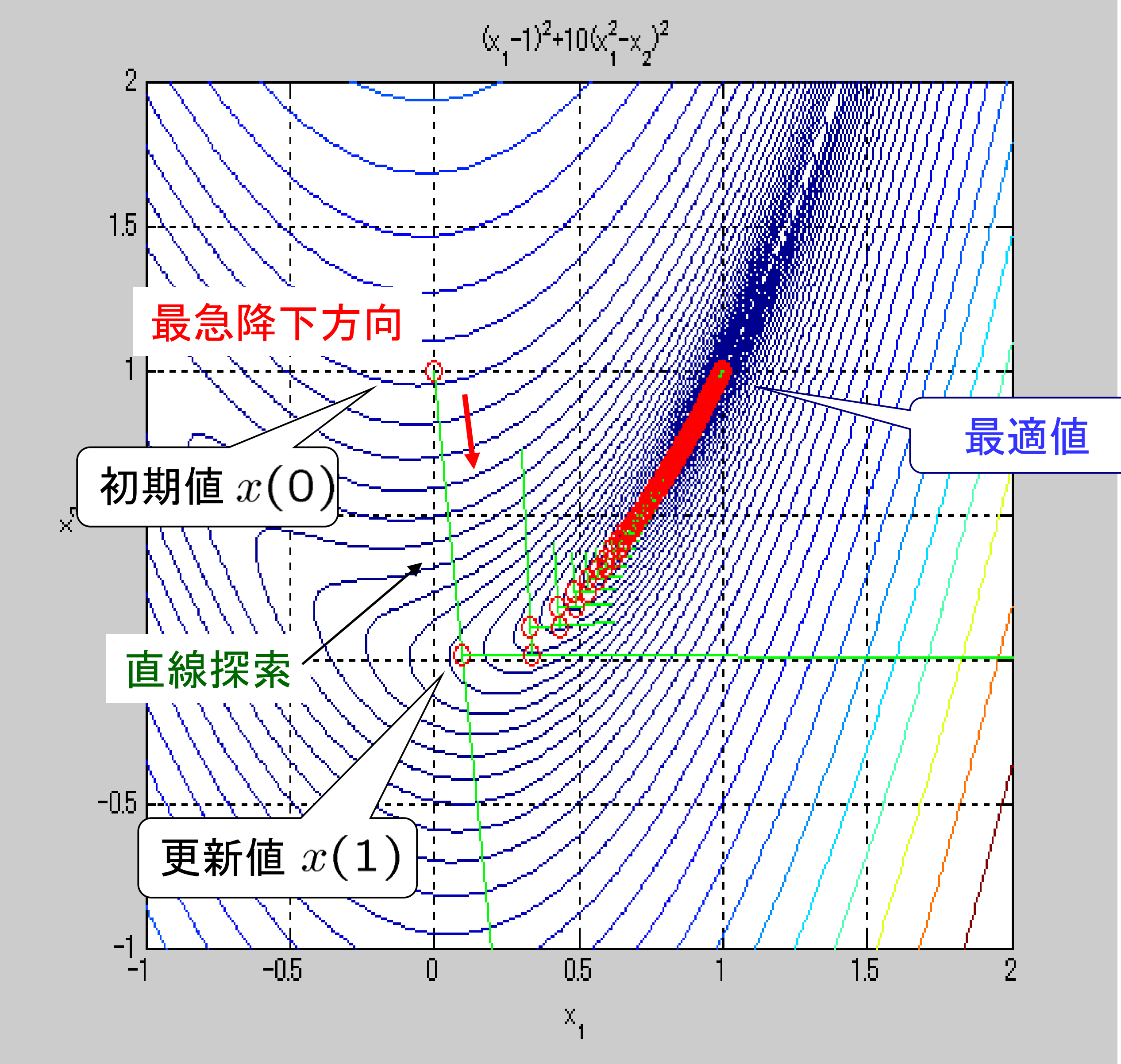

$(x_1-1)^2+2(x_2-1)^2$
更新値 $x(1)$
直線探索
最適値
最急降下方向
初期値 $x(0)$
x_1
x_2

■ 例題２ $\min f(x)$ $\quad f(x) = (x_1 - 1)^2 + 10(x_1^2 - x_2)^2, x^* = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$

k	x_1	x_2	$f(x(k))$	$\|\nabla f(x(k))\|$
1	0.00000E+00	1.00000E+00	1.10000E+01	2.00998E+01
2	9.90037E−02	9.96300E−03	8.11795E−01	1.80263E+00
3	3.38997E−01	9.53360E−03	5.47986E−01	2.11043E+00
4	3.34044E−01	1.07094E−01	4.43699E−01	1.27507E+00
5	4.30998E−01	1.13941E−01	3.75341E−01	1.43984E+00
10	5.28588E−01	2.76497E−01	2.22314E−01	8.83249E−01
20	6.57193E−01	4.30128E−01	1.17548E−01	6.39942E−01
30	7.28946E−01	5.30180E−01	7.34845E−02	5.08218E−01
40	7.77117E−01	6.02786E−01	4.96895E−02	4.11426E−01
50	8.13594E−01	6.61159E−01	3.47532E−02	3.47871E−01
100	9.08144E−01	8.24018E−01	8.44254E−03	1.58631E−01
200	9.77606E−01	9.55603E−01	5.01633E−04	4.05543E−02
300	9.92288E−01	9.84603E−01	5.94799E−05	1.41382E−02
400	9.97204E−01	9.94381E−01	7.83029E−06	4.27191E−03
500	9.98832E−01	9.97650E−01	1.36711E−06	1.79332E−03
$\boxed{529}$	9.99101E−01	9.98191E−01	8.09872E−07	$\boxed{9.94169E−04}$

停止条件:

$\|\nabla f(x(k))\| < 10^{-3}$



$(x_1-1)^2+10(x_1^2-x_2)^2$
最急降下方向
初期値 $x(0)$
最適値
直線探索
更新値 $x(1)$
x_1
x_2
2
1.5
1
-0.5
-1
-1
-0.5
0
0.5
1
1.5
2

■ 最急降下法の特徴

出発点をどこに選んでも，生成される点列が有界ならば，その点列の集積点は関数 f の停留点となる．

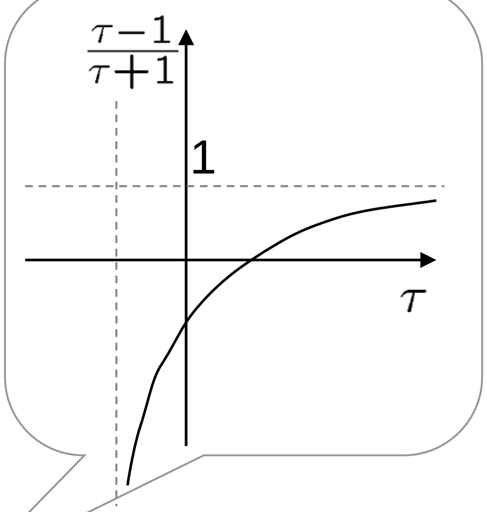
⇒ 大域的収束性をもつ

$G := \nabla^2 f(x^*) > 0$(最適性の十分条件)であるとき，ノルム$\|\cdot\|_G$を以下のように定める．

$$\|x\|_G = \sqrt{x^T G x}.$$

このとき，$\forall \epsilon > 0$，$\exists k_0 \in \mathbf{N}$ s.t.

$$\|x(k+1) - x^*\|_G \leq \left(\frac{\tau - 1}{\tau + 1} + \epsilon\right) \|x(k) - x^*\|_G, \ \forall k > k_0.$$

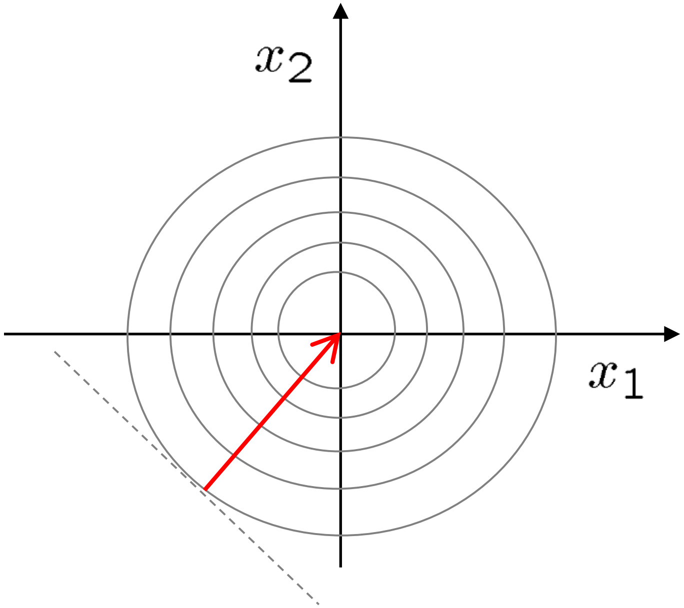


τ: G の条件数(最大固有値と最小固有値の比)

$\epsilon \simeq 0$ とすれば，局所最適解に一定の比率 $(\tau - 1)/(\tau + 1)$ で収束（1 次収束）

⇒ $\tau >> 1$ のとき収束は非常に遅い

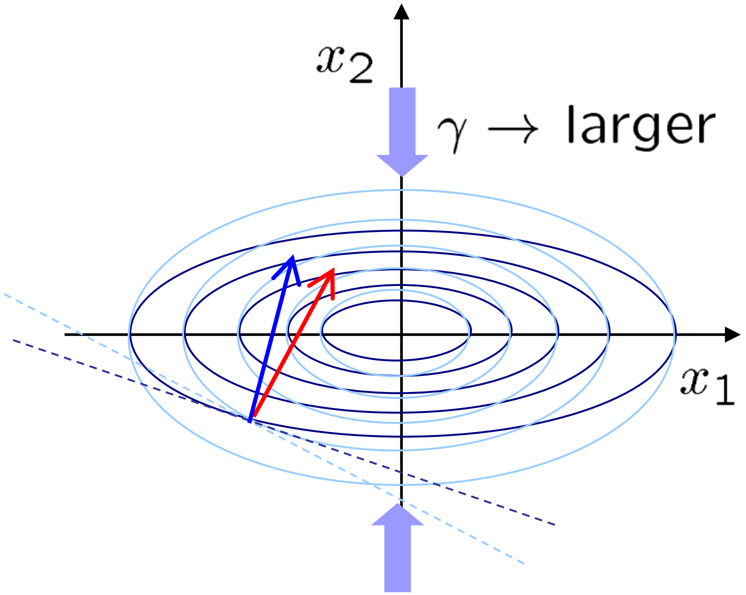
Today's minutes paper



$$f(x) = x_1^2 + x_2^2, \quad x^* = (0,0)$$

$$\nabla^2 f(x^*) = \begin{bmatrix} \boxed{} \end{bmatrix} =: G$$

$$\tau = \lambda_{\max}(G)/\lambda_{\min}(G) = \boxed{}$$



$$f(x) = x_1^2 + \gamma x_2^2, \quad x^* = (0,0)$$

$$\nabla^2 f(x^*) = \begin{bmatrix} \boxed{} \end{bmatrix} =: G$$

$$\tau = \lambda_{\max}(G)/\lambda_{\min}(G) = \boxed{}$$

最急降下法の特徴

-  大域的収束性をもつ
（収束性が初期値の選び方に依存しない）
-  一次収束
（収束が非常に遅くなる場合がある）

停止条件: $\|\nabla f(x(k))\| < 10^{-3}$ のとき

	目的関数	初期値	最適値	繰り返し回数
例題1	$(x_1 - 1)^2 + 2(x_2 - 1)^2$	(0,0)	(1,1)	8
例題2	$(x_1 - 1)^2 + 10(x_1^2 - x_2)^2$	(0,1)	(1,1)	529

■ より高次の微分情報が得られる場合

(非線形）目的関数のテーラー展開を考えると…

ここまでの情報で局所最適化をおこなう ➡ 最急降下法

$$f(x+\delta x) = f(x)+\nabla f(x)^T\delta x+\frac{1}{2}\delta x^T\nabla^2 f(x)\delta x+\cdots$$

2階微分係数の情報まで利用して，さらに効率よい数値最適化をおこなう

➡ ニュートン法

以前の例では $f(x) = x^2 - 2 = 0$ の解を求めるためにニュートン法を用いたが, ここでは $f(x)$ が極値を取る条件 $f'(x) = 0$ を満たす点を求めるために使う.

また以前の例は1変数関数を対象としていたが,
今回はベクトル変数に対する関数を対象としている.

ニュートン法

非線形目的関数を2次関数で近似し, 平方完成によって最小化

$$
\begin{aligned}
f(x+\delta x) &\simeq f(x) + \nabla f(x)^T \delta x + \frac{1}{2}\delta x^T \nabla^2 f(x) \delta x \\
&=: c + b^T y + \frac{1}{2} y^T A y \qquad (A = A^T) \\
&= \boxed{\qquad\qquad\qquad\qquad\qquad\qquad}
\end{aligned}
$$

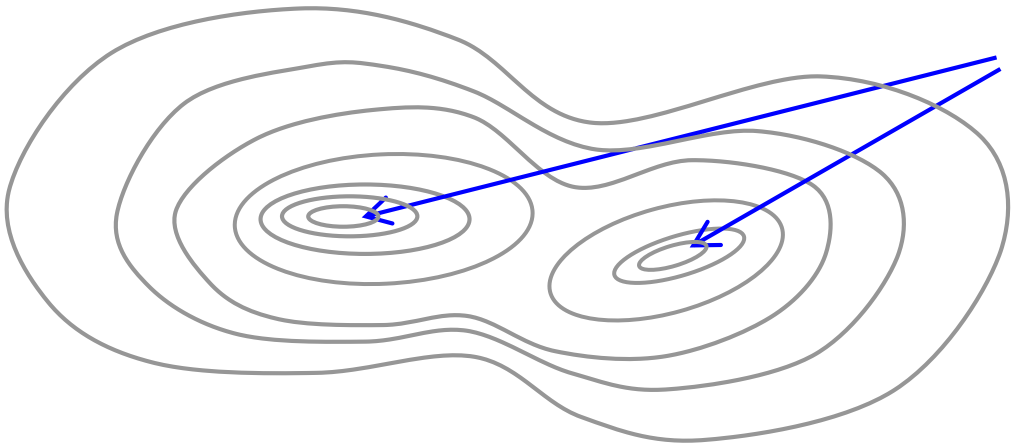
$y = \boxed{\qquad}$ のとき最小 $\Rightarrow \delta x = \boxed{\qquad\qquad\qquad}$

更新則

$$x_{\text{new}} = \boxed{\qquad\qquad\qquad\qquad}$$

暗黙裡に $A > 0$ を仮定

Q1．仮定 $A > 0(\nabla^2 f(x) \geq 0)$ は妥当か？



A1. 局所最適値の近傍では常に成立

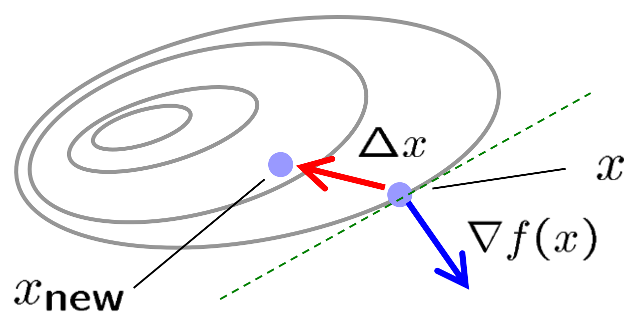
Q2．元々２次関数ではないものに対して，平方完成に基づく更新則を用いてよいか？

$$x_{\text{new}} = x - \nabla^2 f(x)^{-1} \nabla f(x) =: x + \Delta x$$

$\nabla^2 f(x) > 0$ のとき

$$\langle \nabla f(x), \Delta x \rangle \ = \ -\nabla f(x)^T \nabla^2 f(x)^{-1} \nabla f(x) < 0$$

A2. Yes. 勾配と更新方向の交角が90度より大 ⇒ 関数値は必ず減少



$\nabla^2 f(x) > 0$が成り立つような範囲に初期値をとる必要あり

⇒ 局所的収束性

収束性に関しては，ある定数$\beta > 0$が存在して

$$\|x(k+1) - x^*\| \leq \beta \|x(k) - x^*\|^2$$

⇒ 2次収束

ニュートン法の特徴

[○]　収束が非常に速い　(2次収束)
[×]　初期値を最適解近傍にとる必要あり
(局所的収束性しかもたない)

■ ニュートン法のalgorithm

1. Choose x_0. Set $k := 0$.

2. Compute the values of $\nabla f(x)$ and $\nabla^2 f(x)$ at $x(k)$.
   Stop if $\|\nabla f(x(k))\| < \epsilon$.

3. Update x by

$$x(k+1) = x(k) - \nabla^2 f(x(k))^{-1} \nabla f(x(k))$$

4. Set $k := k + 1$. Go back to 2.

例題 [最急降下法の例題2]

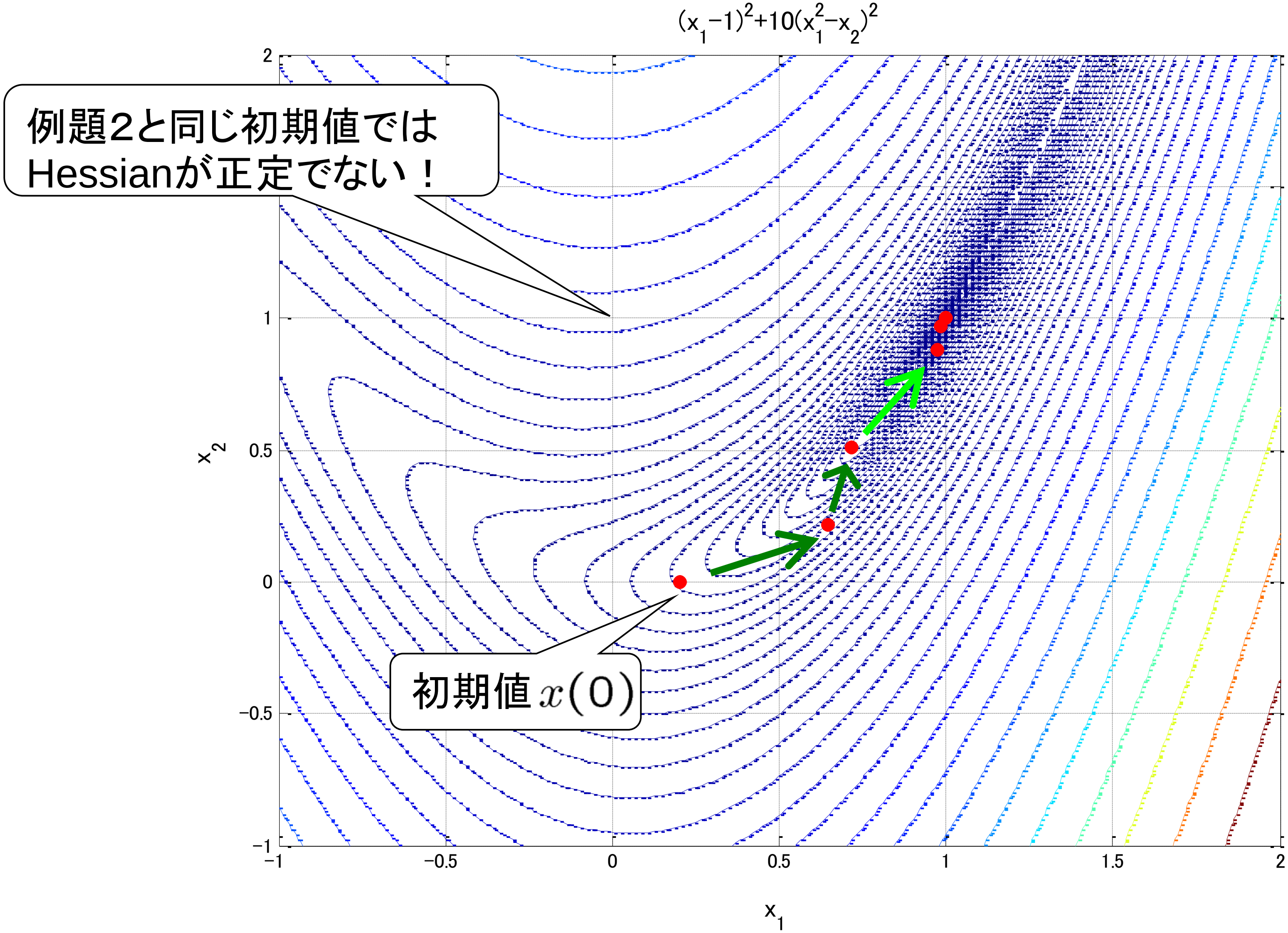
$$f(x) = (x_1 - 1)^2 + 10(x_1^2 - x_2)^2, \qquad x^* = \begin{bmatrix} 1 & 1 \end{bmatrix}^T$$

初期値: $x(0) = \begin{bmatrix} 0.2 & 0 \end{bmatrix}^T$

停止条件: $\|\nabla f(x(k))\| < 10^{-3}$ のとき

k	x_1	x_2	$f(x(k))$	$\|\nabla f(x(k))\|$
1	2.000E-01	0.000E+00	6.560E-01	1.509E+00
2	6.444E-01	2.178E-01	5.166E-01	5.899E+00
3	7.163E-01	5.079E-01	8.077E-02	4.322E-01
4	9.735E-01	8.815E-01	4.447E-02	2.849E+00
5	9.849E-01	9.699E-01	2.285E-04	2.522E-02
6	1.000E+00	9.997E-01	5.177E-07	1.009E-02
7	1.000E+00	1.000E+00	3.167E-14	2.961E-07

最急降下法
では529回



$(x_1-1)^2+10(x_1^2-x_2)^2$
例題2と同じ初期値では
Hessianが正定でない！
初期値 $x(0)$
x_1
x_2

最急降下法

[○]　大域的収束性をもつ
　　勾配のみ必要
[×]　1次収束

ニュートン法

[○]　2次収束
[×]　局所的収束性
　　ヘッセ行列の計算必要

ニュートン法の欠点を解消しつつ, 速い収束性を保持

準ニュートン法

$\nabla^2 f(x)$ を直接用いる代わりに，行列 H_k を逐次計算しつつ，探索方向を決定．（$\nabla^2 f(x)$ の近似に相当）

準ニュートン法のalgorithm

1. Choose x_0. Set $H_0 = I$, $k := 0$.
2. Let $d_k := -H_k \nabla f(x_k)$. Stop if $\|d_k\| < \epsilon$.
3. Choose step size α_k via line search for

$$x_{k+1} = x_k + \alpha_k d_k.$$

4. Update H_k.
5. Set $k := k + 1$. Go back to 2.

At $k = 0$, d_k is the steepest descent direction.

注意: これまではベクトルxのk番目の要素x_kと区別してk回目のxを$x(k)$と表記してきたが，以下では混乱は生じないのでx_kと表記.

■ 準ニュートン法の性質

$H_k > 0$が維持される限り，繰り返しごとに目的関数値は単調減少する．

$d_k = -H_k \nabla f(x_k)$より

$$\begin{aligned} \langle d_k, \nabla f(x_k) \rangle &= \nabla f(x_k)^T d_k \\ &= -\nabla f(x_k)^T H_k \nabla f(x_k) < 0. \end{aligned}$$

（ニュートン法と同様）

Hessianの近似（スカラ変数の場合）

$$\nabla^2 f(x) = \frac{\partial}{\partial x}\nabla f(x) = \lim_{\Delta x \to 0} \frac{\nabla f(x+\Delta x) - \nabla f(x)}{\Delta x}$$

よって Δx が小さいとき

$$\nabla^2 f(x) \simeq \frac{\nabla f(x+\Delta x) - \nabla f(x)}{\Delta x}$$

$$\Rightarrow \quad \nabla^2 f(x_{k+1}) \simeq \frac{\nabla f(x_{k+1}) - \nabla f(x_k)}{x_{k+1} - x_k}$$

$$s_k := x_{k+1} - x_k, \quad y_k := \nabla f(x_{k+1}) - \nabla f(x_k)$$

ベクトル変数の場合

$H_{k+1} s_k = y_k$ なる正定行列 H_k をいかに定めるか？

H_k の更新則

$$s_k := x_{k+1} - x_k,\ \ y_k := \nabla f(x_{k+1}) - \nabla f(x_k)$$

- DFP公式

$$H_{k+1} = H_k + \frac{s_k s_k^T}{s_k^T y_k} - \frac{H_k y_k y_k^T H_k}{y_k^T H_k y_k}$$

- BFGS法

$$H_{k+1} = H_k + \frac{y_k y_k^T}{s_k^T y_k} - \frac{H_k s_k s_k^T H_k}{s_k^T H_k s_k}$$

David Shanno: KarmakerのKORBX (890万ドル）
に対抗するOB1 (5万ドル）の開発メンバのひとり

数学的帰納法による$H_k > 0$の証明（DFP）

$H_0 = I$より$k = 0$のとき成り立つ．

$H_i > 0$を仮定したとき，$H_{i+1} > 0$を示す．

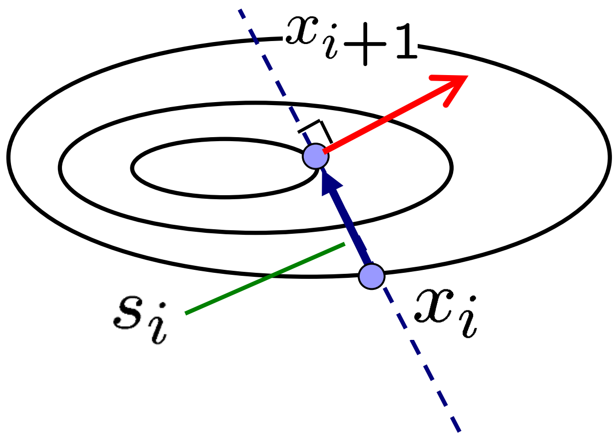
現在のx_iは局所最適解ではないとする．すなわち$\nabla f(x_i) \neq 0$.

いま$x \neq 0$なるベクトルを考え，更新則の前後からx^T，xをかけると

$$x^T H_{i+1} x = x^T H_i x + \frac{x^T s_i s_i^T x}{s_i^T y_i} - \frac{x^T H_i y_i y_i^T H_i x}{y_i^T H_i y_i}. \qquad (1)$$

式(1)右辺第２項の分子は$|s_i^T x|^2 \geq 0$．分母は

$$s_i^T y_i = s_i^T\{\nabla f(x_{i+1}) - \nabla f(x_i)\}. \quad (2)$$



s_iはx_iからx_{i+1}への増分であることに注意．精密な直線探索の結果，s_iを含む直線は目的関数の等高線に対する接線になっている．すなわち

$$s_i^T \nabla f(x_{i+1}) = 0. \quad (3)$$

またs_iは$\alpha_i > 0$，$d_i = -H_i \nabla f(x_i)$ により$s_i = \alpha_i d_i$とも書ける．よって

$$-s_i^T \nabla f(x_i) = \alpha_i \nabla f(x_i)^T H_i \nabla f(x_i) > 0, \quad (4)$$

($H_i > 0$なので)．よって(2), (3), (4)から$s_i^T y_i > 0$となり，式(1)右辺第２項は非負．

一方，式(1)右辺第１および第３項をまとめると

$$\frac{(x^T H_i x)(y_i^T H_i y_i) - (x^T H_i y_i)^2}{y_i^T H_i y_i}. \qquad (5)$$

Schwarzの不等式

$$(u^T Q v)^2 \leq (u^T Q u)(v^T Q v), \ \forall Q > 0$$

(等号成立は$v = \lambda u$, $\lambda \in \mathbf{R}$のときのみ) より，式(5)は非負.

cf. $\langle v, u \rangle \leq |u|^2 |v|^2$

式(5)が0，すなわち等号成立のとき，$x = \lambda y_i$ となっているが，このとき第２項は

$$\frac{x^T s_i s_i^T x}{s_i^T y_i} = \lambda^2 y_i^T s_i.$$

先に示したように $y_i^T s_i = s_i^T y_i > 0$ であり，$x \neq 0$ より $\lambda \neq 0$．したがって第１，３項の和が０であっても，そのとき第２項は正．したがって $\forall x \neq 0$ に対して $x^T H_{i+1} x > 0$．すなわち $H_{i+1} > 0$．

Q.E.D.

準ニュートン法の特徴

目的関数 f が凸関数ならば，任意の出発点に対して生成される点列 $\{x_k\}$ の集積点は最適解

⇒ 大域的収束性

収束性に関しては，最適解 x^* において $\nabla^2 f(x^*) > 0$ ならば

$$\lim_{k \to \infty} \frac{\|x_{k+1} - x^*\|}{\|x_k - x^*\|} = 0.$$

⇒ 超1次収束

1次収束 (最急降下法)

$$\|x_{k+1} - x^*\|_G \leq \bar{\beta} \|x_k - x^*\|_G$$

2次収束 (ニュートン法)

$$\|x_{k+1} - x^*\| \leq \beta \|x_k - x^*\|^2$$

勾配法

正定行列の列 H_k によって定まる方向 $d_k = -H_k \nabla f(x_k)$ に関する精密な直線探索によって解を更新する方法を一般に勾配法という.

- ✓ $H_k > 0$ となるような更新則を与えている準ニュートン法は勾配法である.
- ✓ 最急降下法は，常に $H_k = I$ としている勾配法の特殊ケースである.
- ✓ ヘッセ行列は正定とは限らず，直線探索もおこなわないためニュートン法は勾配法ではない.

期末レポート

例題2を最急降下法, ニュートン法,
準ニュートン法で解くプログラムを
Octave, Matlab等で作成,
プログラムリスト, 計算結果をまとめて
8/19までに提出. (PDFにしてメール添付)