2025 後期　均衡分析と数理最適化

最適化理論概論

経済研究所　大澤 実

osawa.minoru.4z@kyoto-u.ac.jp

本講の目的

変分不等式 (**variational inequalities**) を学ぶための前提知識として，最適化理論の基礎を速習する．

目標：

- 最適化問題とそれに関わる用語を知る
- 最適化問題の具体例をみる
- 最適化問題の主要な問題クラスを知る
- 特に重要な問題クラスとして凸最適化問題を知る
- 数値解法の基本的考え方を知る

あわせて，本講義で使う算数の基礎を復習する．

基本的な定義

最適化問題 (optimization problem)

最適化問題（数理計画問題）

- 与件：集合 $S \subseteq \mathbb{R}^n$，関数 $f : S \to \mathbb{R}$
- 問題：$x \in S$ を満たす x のうち，$f(x)$ を最小にするものを求める．

$$\text{minimize} \quad f(x) \quad \text{subject to} \quad x \in S.$$

- minimize を min.，subject to を s.t. などと略記する．
- f の最大化問題 $\Leftrightarrow$ $-f$ の最小化問題

- 最適化**問題**：問題インスタンス
- 最適化**理論**：問題を解く方法の研究
- 最適化**法**　：研究・応用の営み全般

基本用語

最適化問題 $\min_{x \in S} f(x)$ に対して：

- f：**目的関数** (objective function)
- S：**実行可能領域** or **許容領域** (feasible region)
- 条件「$x \in S$」：**制約** (constraint)
- 制約を満足する $x \in S$：**実行可能解** or **許容解** (feasible solution)
- f の最小値を達成する点 $\bar{x} \in S$：**最適解** (optimal solution)
- 最適解における目的関数値 $f(\bar{x})$：**最適値** (optimal value)

記号等

- n 次元数ベクトル $x \in \mathbb{R}^n$ を $(x_i)_{i=1}^n$，$(x_1, x_2, \ldots, x_n)$，(x_i) 等
- 適切な次元の行列 A との積を考えるときは全て列ベクトルとして扱う：$Ax, x^\top A$, etc.．ただし，$\top$ は転置．
- 同じサイズの数ベクトル x, y に対して $\langle x, y \rangle \equiv \sum_i x_i y_i$．
- 特に断らない限り，数ベクトル x に対して $\|x\| \equiv \left(\sum_i x_i^2\right)^{1/2}$．

最適解

$\boxed{\textbf{定義}}$ 実行可能解 $\bar{x} \in S$ に対して，ある $\epsilon > 0$ が存在して

$$f(x) \geq f(\bar{x}), \qquad \forall x \in S \text{ such that } \|x - \bar{x}\| < \epsilon$$

が成立するとき，$\bar{x}$ を**局所的最適解** (local optimal solution) と呼ぶ．

ある $\epsilon > 0$ に対して $0 < \|x - \bar{x}\| < \epsilon$ なる $\bar{x}$ 以外の局所的最適解が存在しないなら**孤立局所的最適解** (isolated local optimal solution) と呼ぶ．

$\boxed{\textbf{定義}}$ 最適化問題の実行可能解 $\bar{x} \in S$ に対して，

$$f(x) \geq f(\bar{x}), \qquad \forall x \in S$$

が成立するとき，$\bar{x}$ を**大域的最適解** (global optimal solution) と呼ぶ．

最適解の存在と実行可能性

最適化問題に対して，以下のいずれかが成立する．

1. **実行可能** (feasible) ：実行可能解が存在する（$S \neq \emptyset$）
   (a) 実行可能領域 S において目的関数 f の値が下に有界であり，かつ下限を達成する実行可能解（= 最適解）が存在する．
   (b) S において f の値が下に有界であるが，下限を達成する実行可能解が S に存在しない．
   (c) S において f の値が下に有界でない（**非有界**; unbounded）
2. **実行不能** (infeasible) ：実行可能解が存在しない（$S = \emptyset$）

Quiz $S \subseteq \mathbb{R}$ の場合で，1(b) となる状況の例を考えてみよ．

応用における最適化問題

実際の例では，S は具体的な**制約関数**たちによって表現される．

$$
\begin{array}{lll}
\text{minimize} & f(x) & \\
\text{subject to} & g_j(x) \le 0, & (j = 1, 2, \ldots, J) \\
& h_i(x) = 0, & (i = 1, 2, \ldots, I)
\end{array}
$$

- 不等号の向きは「小なり $\le$」が標準形．この向きに取ることで制約関数たちが S の（相対的な）「内側」を表現する．
- 一般性を失わない：$\tilde{g}(x) \ge 0 \Leftrightarrow g(x) \equiv -\tilde{g}(x) \le 0$

  例：非負制約 $x_i \ge 0$ は $g(x_i) = -x_i \le 0$

Quiz 2 変数問題について $S = \{x \in \mathbb{R}^2 : g(x) = x_1^2 + x_2^2 - a \le 0\}$ が $a \in \mathbb{R}$ の変化に対してどう振る舞うか考察せよ．

$\boxed{\text{例}}$ 効用最大化問題

n 種の財の消費量 $x = (x_i)_{i=1}^n$，価格 $p = (p_i)_{i=1}^n$，所得 $y > 0$.

効用関数が準凹関数 $U(x)$ で与えられるとき

$$
\begin{aligned}
&\text{maximize} && U(x) \\
&\text{subject to} && \sum_i p_i x_i \leq y,\ x \geq \mathbf{0}
\end{aligned}
$$

目的関数 $f(x) = -U(x)$，実行可能領域を $S \equiv \{\langle p, x \rangle \leq y, x \geq \mathbf{0}\}$ とすれば最初にみた一般形の最小化問題と一致.

U として扱いやすい関数を選べば解析的に解けるのだった.

例 施設配置問題

n 世帯が居住する村に集会所を設置しよう．村の土地を $S \subset \mathbb{R}^2$ とする．世帯 k の座標を $a^k = (a_1^k, a_2^k) \in S$ で，集会所の座標を $x \in S$ で表す．

1. 学校までの二乗距離を平均的に最小化： $\min_{x \in S} \frac{1}{n} \sum_{k=1}^n \|a^k - x\|^2$
2. 学校までの距離を平均的に最小化： $\min_{x \in S} \frac{1}{n} \sum_{k=1}^n \|a^k - x\|$
3. 最も遠い家から学校までの距離を最小化： $\min_{x \in S} \max_k \|a^k - x\|$

Quiz 1 の最適解を考え，S の形状との関係を考察せよ．

教訓 （当然のことながら）S の形状が解の性質に影響．

$\boxed{\text{例}}$ 線形回帰と最小二乗法

m 個のデータ点 (a_i, b_i) $(i = 1, 2, \ldots, m)$ が与えられている．
ただし，$a_i = (a_{ij})_{j=1}^n \in \mathbb{R}^n$ であり，$b_i \in \mathbb{R}$．

共通の $x \in \mathbb{R}^n$ で各点を $b_i = \langle a_i, x \rangle$ の形の線形式で近似したい．このような近似のよさを $A = (a_{ij})$ として**二乗誤差**で測ってみてもよい：

$$f(x) \equiv \|Ax - b\|^2 = \sum_{i=1}^{m} (\langle a_i, x \rangle - b_i)^2$$

この f を x について最小化する基準を考えられる：$\min_{x \in \mathbb{R}^n} f(x)$．その解 $\bar{x}$ による近似 $\langle a_i, \bar{x} \rangle$ は，<u>線形関数</u>のうちで<u>二乗誤差の意味で</u>データを最もよく近似する．もちろん他の関数型・誤差基準も想定できる．

制約無し最適化問題：x は全空間 $\mathbb{R}^n$ を動いてよい（$S = \mathbb{R}^n$）．

$\boxed{例}$ 栄養問題 (diet problem)

野菜 $j = 1, 2, \ldots, n$，栄養素 $i = 1, 2, \ldots, m$ がある．

必要な最低限度の栄養素の量 $b = (b_i)_{i=1}^m$ を満足しつつ，
原料費の合計を最小にする野菜購入量 $x = (x_j)_{j=1}^n$ を決めたい．

野菜 j について，単位数量あたりの価格は c_j，栄養素 i の含有量は a_{ij}．

Quiz 最適化問題として表現してみよ．

$\boxed{\text{例}}$ 輸送計画問題 (transportation problem)

m 箇所の工場から n 箇所の顧客へ納入したい．ただし，

- 工場 i の最大限の生産量は a_i
- 顧客 j の最低限の需要量は b_j
- 工場 i から顧客 j への単位あたり輸送費用は c_{ij}

このとき総輸送費用を最小化する輸送量のパターン $x = (x_{ij})$ は？
ただし x_{ij} は工場 i から顧客 j への輸送量を意味する．

$\boxed{\textbf{Quiz}}$ 最適化問題として表現してみよ．

線形計画問題 (linear programming; LP)

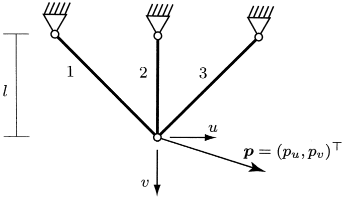
栄養計画問題および輸送計画問題は**線形計画問題**，すなわち目的関数・制約のすべてが線形関数で与えられる最適化問題の例である．

$$
\begin{aligned}
\text{maximize} \quad & \sum_{j=1}^{n} c_j x_j, & \\
\text{subject to} \quad & \sum_{j=1}^{n} a_{ij} x_j \le b_i & (i = 1, ..., m), \\
& x_j \ge 0 & (j = 1, ..., n).
\end{aligned}
$$

$\boxed{\text{例}}$ 構造最適化問題 (structural optimization)

下図のようなトラスの部材断面積 (a_1, a_2, a_3) を決定したい $(a_3 = a_1)$

ここで，p はトラスに加えられている荷重，u, v はそれによる頂点の変位を表す．全ての部材の Young 率は E．変位は $(\hat{u}, \hat{v})$ を，各部材に働く応力は $\hat{\sigma}$ を超過してはならない．また，a_i は $\hat{a}$ を下回ってはならない．



（出典：東京大学工学教程）

例 構造最適化問題

よい構造を選ぶために，次のような最適化問題を考えることができる：

$$
\begin{array}{ll}
\text{minimize} & 2\sqrt{2}la_1 + la_2 \\
\text{subject to} & \frac{1}{\sqrt{2}}\left(\frac{p_u}{a_1} + \frac{p_v}{a_1+\sqrt{2}a_2}\right) \le \hat{\sigma},\ \frac{\sqrt{2}p_v}{a_1+\sqrt{2}a_2} \le \hat{\sigma}, \\
& \frac{1}{\sqrt{2}}\left(-\frac{p_u}{a_1} + \frac{p_v}{a_1+\sqrt{2}a_2}\right) \ge -\hat{\sigma}, \\
& \frac{\sqrt{2}lp_u}{Ea_1} \le \hat{u},\ \frac{\sqrt{2}lp_v}{E(a_1+\sqrt{2}a_2)} \le \hat{v}, \\
& a_1 \ge \hat{a}, a_2 \ge \hat{a}.
\end{array}
$$

この問題は制約が非線形な関数で与えらる**非線形計画問題**である．

最適化問題の分類

- 制約付最適化 (constraind) vs. 無制約最適化 (unconstrained)
- 線形 (linear) vs. 非線形 (nonlinear)
- 凸 (convex) vs. 非凸 (nonconvex)
- 連続 (continuous) vs. 離散 (discrete)
- 静学的 (static) vs. 動学的 (dynamic)

本講義では，基本的に静学的な連続最適化を取り扱う．

凸最適化問題

in fact, the great watershed in optimization isn't between linearity and nonlinearity, but convexity and nonconvexity.

– R. Tyrrell Rockafellar, in SIAM Review, 1993

復習 集合と関数の凸性

定義 **凸集合** (convex set) S は次の条件を満足する集合である：

$$\theta x + (1-\theta)y \in S, \qquad \forall x, y \in S, \theta \in [0,1].$$

定義 **凸関数** (convex function) f とは次の条件を満足する関数である：

$$\theta f(x) + (1-\theta) f(y) \geq f\left(\theta x + (1-\theta)y\right), \qquad \forall x, y, \forall \theta \in [0,1].$$

不等号が狭義で成立するならば**狭義凸関数**と呼ぶ．

f の定義域を適当な集合 S へ制限した場合についても同様に定義される．

凸最適化問題

f が S 上で凸関数かつ S が凸集合である最適化問題を特に**凸最適化問題** (convex optimization problem) と呼ぶ．

$$\text{minimize} \quad f(x) \quad \text{subject to} \quad x \in S \quad (S\text{: convex}, f\text{: convex})$$

凸最適化問題には様々な著しくよい性質がある．例えば

- 局所的最適解が存在するならば大域的最適解となる．
- 最適解の集合が凸集合となる．
- 目的関数が狭義凸関数なら，最適解が存在すれば唯一である．
- **双対性** (duality) により問題の等価変換・計算を効率化可能．

制約関数による凸最適化問題の特徴づけ

応用上，S は具体的な制約関数たちによって表現された．

$$
\begin{array}{lll}
\text{minimize} & f(x) & \\
\text{subject to} & g_j(x) \leq 0, & (j = 1, 2, \ldots, J) \\
& h_i(x) = 0, & (i = 1, 2, \ldots, I)
\end{array}
$$

凸最適化問題になるのは以下が成立するときである：

- f が凸関数
- $\{g_j\}$ が全て凸関数
- $\{h_i\}$ が全て affine 関数（$h_i(x) = \langle a_i, x \rangle + b_i$）

Quiz このとき確かに上の問題が凸最適化問題になることを確かめよ．

余談 解きやすさと“面白さ”

- 実空間を対象とした均衡分析は**計算機利用が前提**．数千以上の変数．解析解はない．都市・地域の分析では計算機が大いに活用されている．
- 問題の数理的性質の特徴づけは，意味解釈および数値計算の効率化の双方で有用．大抵 f や S のようすは自明ではない．
- 計算機で解き易い問題に落とし込むことはモデリングの鍵．凸最適化問題は解き易いクラスの問題の代表例．
- しかし，理論モデルとしては解き易さ $=$ 興味深さとは限らない．
  - 例：問題の凸性（解き易さ）は均衡の一意性に対応．複数均衡を許す理論モデルの面白さ・表現力を取り除く．
- 解きやすくかつ面白いモデルを作るのはとても難しい．

数値解法

本節の目的

よく調べられた性質のよい問題に対してはソルバがある．経済学の研究では自分でプログラムを組む必要がある場合も多い．

- 問題が大規模になった場合，素朴な定式化ではメモリ不足になったりするため，自分でアルゴリズムを工夫する必要がある．
- 経済学分野で出てくる非線形計画問題に対しては安定したソルバが存在しない場合がある．

以上のような背景を念頭に，数値計算法の一般的な考え方を学ぶ：

- 数値計算法における反復法 (iterative method) の考え方を知る
- 無制約最適化問題とそれに対する反復法を知る
- 制約つき最適化問題に対する反復法を知る

反復法 (iterative method)

反復法：初期解からスタートして**反復計算**により真の解に近づけることで数値解を求めるためのアルゴリズム．

- **数値解**：所望の条件を何らかの意味で近似的に満足する変数値
  - 例：方程式 $F(x) = \mathbf{0}$ の数値解：$\|F(x)\| < \delta \ll 1$ なる x
- **反復計算**：初期解 x^0 から x^k を反復的に更新 $x^k \mapsto x^{k+1}$ して，**解候補**の列 $\{x^k\}_{k=0}^{\infty}$ を得ること
- 例：非線形方程式の解法としての **Newton 法**

対義語：**直接法** (direct method)

初期解からスタートして有限回の手続きで数値解を定めるアルゴリズム．

- 例：線形方程式の解法としての **Gauss の消去法**（掃き出し法）

最適化問題に対する反復法

最適化問題 $\min_{x \in S} f(x)$ に対しては，初期実行可能解 (initial feasible solution) $x^0 \in S$ からスタートして解を更新していく $x^k \mapsto x^{k+1} \in S$ ことで**局所的最適解**を発見しようとする．

- 反復法により**大域的最適解**を発見することは一般の問題に対しては困難．大域的最適化 (global optimization) に対しては**メタヒューリスティクス** (metaheuristics) と総称される様々なアルゴリズムがあるが，本講義では取り扱わない．
  - 例：焼きなまし (simulated annealing) 法，遺伝的アルゴリズム
- 凸最適化問題では全ての局所的最適解が大域的最適解となる．

どのような解更新を行うかによって，それぞれ別の解法となる．

復習 スカラー値関数の勾配と Hesse 行列

$f : \mathbb{R}^n \to \mathbb{R}$ に対して，その点 x における**勾配** (gradient) $\nabla f(x)$ は

$$\nabla f(x) \equiv \left(\frac{\partial f}{\partial x_1}(x), \frac{\partial f}{\partial x_2}(x), \ldots, \frac{\partial f}{\partial x_n}(x) \right)^\top$$

で定義され，また **Hesse 行列** (Hessian matrix)$\nabla^2 f(x)$ は

$$\left[\nabla^2 f(x)\right]_{ij} = \frac{\partial^2 f}{\partial x_i \partial x_j}(x), \qquad i, j = 1, \ldots, n$$

次のページで復習する微分可能性が満足されるとき意味を持つ．

復習 スカラー値関数の微分可能性と近似

$\|\delta\| \to \mathbf{0}$ のとき，$f : \mathbb{R}^n \to \mathbb{R}$ が

$$f(x+\delta) = f(x) + \nabla f(x)^\top \delta + o(\|\delta\|)$$
$$f(x+\delta) = f(x) + \nabla f(x)^\top \delta + \frac{1}{2}\delta^\top \nabla^2 f(x) \delta + o(\|\delta\|^2)$$

を満足するかどうかを考える．$o(h(\delta))$ は $\delta \to \mathbf{0}$ において $h(\delta)$より速く小さくなる項．

- **1 回微分可能**であるとは全 x において $\nabla f(x)$ が存在すること．
- **2 回微分可能**であるとは全 x において $\nabla^2 f(x)$ が存在すること．

（このような関係を満足する量を微分と呼ぶ．）

復習 スカラー値関数の勾配の性質

スカラー値関数 f の等高線 (contour) と勾配 ∇f とは直交する．

略証：$c \in \mathbb{R}$ に対して関数 f がその値をとる点の集合を

$$X_c = \{x \in \mathbb{R}^n \mid f(x) = c\}$$

とし，$t \in \mathbb{R}$ に対して X_c の一点を対応させる写像を $x(t)$ で表現する．

このとき，X_c 上で $f(x(t)) = c$ なる関係を両辺微分して $t = 0$ とすると

$$\frac{\mathrm{d}f(x(0))}{\mathrm{d}t} = \nabla f(x(0))^\top \frac{\mathrm{d}x(0)}{\mathrm{d}t} = 0$$

$x(0)$ と c の選び方は任意だから所望の結果を得る．

復習 スカラー値関数の勾配の性質

スカラー値関数 f の勾配 ∇f は f が最も増加する方向を示す．

f が 1 回微分可能ならば $f(x+\delta) \approx f(x) + \nabla f(x)^\top \delta$ が小さい δ に対して成立する．いま，δ だけ動いた際の f の増加量を

$$\Delta f(\delta; x) \equiv \langle \nabla f(x), \delta \rangle \qquad (\approx f(x+\delta) - f(x))$$

としよう．$\mathbb{R}^n$ における通常の内積は，$\nabla f(x)$ と δ のなす角を θ とすれば

$$\langle \nabla f(x), \delta \rangle = \|\nabla f(x)\| \|\delta\| \cos(\theta)$$

であったから，$\|\delta\|$ が固定のとき，$\Delta f(\delta; x)$ は $\theta = 0$ で最大値をとる．

反復法の基本

最適化問題に対する反復法の基本形は次の通り．

$$x^{k+1} := x^k + \alpha^k d^k$$

ここで $x^k \in S$ は k 回反復時点における解候補であり，

- $d^k \in \mathbb{R}^n$ を**探索方向** (search direction)，
- $\alpha^k \in \mathbb{R}$ を**ステップ幅** (step size)，と呼ぶ．

反復法の**収束判定条件**には様々な考え方があるが，局所最適解であるための**最適性条件** (optimality conditions) を確認することが基本．

最適性条件については後に非線形最適化問題を題材に系統的に復習する．

降下方向

素朴には，$x^{k+1} := x^k + \alpha^k d^k$ のような反復で局所的最適解が求まるためには，探索方向 d^k が目的関数を減少させる方向であることが必要であろうと考えられる．つまり d^k は

$$\Delta f(d^k; x) = \langle \nabla f(x), d^k \rangle < 0$$

を満足していてほしいということを意味する．

この条件を満足する d^k を**降下方向** (descent direction) と呼ぶ．

※一部の反復法（**加速勾配法**等）ではこの条件は必ずしも満足されない．

反復法の基本

- **探索方向** d^k およびステップ幅 α^k をどう設定するかについて，様々な方針がある．
  - 微分情報を使わない方法 (derivertive-free method)
  - 勾配の情報を使う方法 (first-order method/gradient method)
  - Hesse 行列の情報を使う方法 (second-order method)
- 例：**最急降下法** (method of steepest descent)：$d^k = -\nabla f(x^k)$．
- 理論上は目的関数の情報をより多く使えた方が収束のスピードが速い (解に収束するまでに必要な反復回数が少なくなる)．
- 問題の次元が高くなると，勾配や Hesse 行列の計算機による評価自体のコストが高くなる $\Rightarrow$ 目的関数について高次の微分情報を要求するアルゴリズムが実用的に効率的かどうかは問題による．

反復法と進化ダイナミクス

反復法は**進化ゲーム理論**とも関係がある．講義後半では進化ゲーム理論との関係についても触れる．

進化ゲームでは，エージェントが戦略を改訂する状況で戦略分布 x の時間変化を**進化ダイナミクス** (evolutionary dynamics) でモデル化する．特に離散時間のモデルはある種の反復法と直接見なすことができる．連続時間の進化ダイナミクスでも時間離散化により反復法が導かれる．

例 BR(x) を戦略分布 x に対する混合最適応答とすれば，**最適応答動学** (best-response dynamic) は

$$x^{k+1} \in \mathrm{BR}(x^k)$$

で与えられ，その Nash 均衡への収束性が問題になる．

無制約最適化問題に対する反復法

無制約最適化問題

制約付き最適化問題においては，d^k をどのように決めるかも非自明．

ここでは単純な反復法の例を導入するため無制約最適化問題を考える：

$$\text{minimize} \quad f(x) \quad \text{subject to} \quad x \in \mathbb{R}^n$$

まず**最適性条件**を知り，次の 2 つのアルゴリズムを学ぶ：

- 最急降下法
- Newton 法

無制約最適化問題の最適性条件

定理 **1 次の必要条件** (1st-order necessary condition)：f が 1 回微分可能とする．もし $\bar{x} \in \mathbb{R}^n$ が無制約最適化問題の局所最適解ならば，次の条件を満足しなければならない．

$$\nabla f(\bar{x}) = \mathbf{0}.$$

このような $\bar{x}$ を**停留点** (stationary point) と呼ぶ．

- 必要条件：$\nabla f(\bar{x}) = \mathbf{0} \Longleftarrow \bar{x}$ が局所最適解．

Quiz 1 変数の場合に，十分性が成り立たない場合（停留点であるが局所最適解ではない場合）を図示せよ．

無制約最適化問題の最適性条件

$\boxed{\textbf{定理}}$ **2 次の必要条件** (2nd-order necessary condition)．f が 2 回微分可能とする．もし $\bar{x} \in \mathbb{R}^n$ が無制約最適化問題の局所最適解ならば，$\nabla^2 f(\bar{x})$ は半正定値である．

- 行列 A が半正定値：$\delta^\top A\delta \geq 0\ (\forall \delta \neq 0)$
- 必要条件：$\nabla^2 f(\bar{x})$ が**半正定値** $\Longleftarrow$ $\bar{x}$ が局所最適解．

$\boxed{\textbf{Quiz}}$ 1 変数の場合に，十分性が成り立たない場合を図示せよ．

無制約最適化問題の最適性条件

定理 **2 次の十分条件** (2nd-order sufficient condition)．f が 2 回微分可能とする．もし $\bar{x} \in \mathbb{R}^n$ において $\nabla f(\bar{x}) = \mathbf{0}$ かつ $\nabla^2 f(\bar{x})$ が**正定値**ならば，無制約最適化問題の局所最適解である．

- 行列 A が正定値：$\delta^\top A\delta > 0$ $(\forall \delta \neq \mathbf{0})$
- 十分条件：$\Longrightarrow$ $\bar{x}$ が局所最適解．

Quiz 1 変数の場合で必要性が成立しない場合を図示せよ．

以上の最適性条件は，後に講義する一般の非線形最適化問題に対する最適性条件に特殊ケースとして含まれる．

最急降下法 (steepest descent)

$f^k \equiv f(x^k)$, $\nabla f^k \equiv \nabla f(x^k)$, etc.

Step 0. 初期化：$x^0 \in \mathbb{R}^n$，$k \leftarrow 0$

Step 1. 停止条件を満足するならば，x^k を数値解として出力.

Step 2. $d^k \leftarrow -\nabla f^k$

Step 3. $\alpha^k \leftarrow \arg\min_{\alpha \geq 0} \phi(\alpha) \equiv f(x^k + \alpha d^k)$

Step 4. $x^{k+1} \leftarrow x^k + \alpha^k d^k$

Step 5. $k \leftarrow k + 1$ として Step 1 へ.

- 収束条件としては小さい δ に対して $|f^{k+1} - f^k| < \delta$ などを考える (≈ 一次の必要条件 $\nabla f(x^k) = \mathbf{0}$).
- 実際的には，Step 3 における**直線探索** (line search) は厳密に行わず，**バックトラッキング** (backtracking; Armijo's rule) により満足できる程度に関数値を下げる α^k を選ぶ.

理論分析可能な例：凸二次計画問題　（出典：金森ら）

次の凸二次関数を $\mathbb{R}^n$ 上で最小化する無制約最小化問題を考える．

$$f(x) = \frac{1}{2}(x - \bar{x})^\top Q(x - \bar{x}) \qquad Q：n \times n \text{ 正定値対称行列}$$

Quiz $x^k \neq \bar{x}$ を考える．

1. 点 x^k における最急降下方向 d^k を求めよ．Q が対称なことに注意．
2. 点 x^k における最適ステップ幅 α^k を求めよ．
3. $f(x^{k+1})$ を，$f(x^k)$ および Q の固有値を用いて評価せよ．
4. この問題では最急降下法の収束速度は行列 Q の**条件数** $\mathrm{cond}(Q) \equiv \lambda_{\mathsf{max}}/\lambda_{\mathsf{min}} \geq 1$ に依存する．その図的意味を考えよ．なお λ_{max} と λ_{min} はそれぞれ Q の最大・最小固有値．（※ 行列の条件数は一般には特異値を用いて定義される．）

補足 条件数 (condition number)

条件数：数値解析で問題の感度を表す指標．条件数が大きいほど，問題は数値的に取り扱いづらく，入力データの摂動が解に大きく影響を与える．

例 線形方程式 $Ax = b$ の場合，非特異な正規行列 A の条件数 $\kappa(A)$ は $\kappa(A) = \max_i |\lambda_i| / \min_i |\lambda_i|$（$\{\lambda_i\}$：$A$ の固有値たち）．

条件数が小（良条件; well-conditioned）なら数値計算は安定しやすい．条件数が大（悪条件; ill-conditioned）なら不安定になりやすい．特に，A が特異なら $\min_i |\lambda_i| = 0$ で条件数 $=$ 無限大．

例 線形方程式系 $Ax = b$ について A の条件数は，データ b が変化すると解がどの程度変化するかを表現．A が特異なとき，$Ax = b$ の解は一意ではなかった．条件数 A がどれだけ "特異に近い" かを示す．

See: シャトラン『行列の固有値』Ch.4 誤差解析
Judd "Numerical Methods in Economics" Sec 3.5

Newton 法

最急降下法は目的関数を x^k において線形近似して降下方向を決定した．
Newton 法では，二次近似によって降下方向を決定する．

f の x^k における二次近似は

$$f(x^k+d) \simeq f_k^{\mathrm{Q}}(d) \equiv f(x^k) + \langle \nabla f^k, d\rangle + \frac{1}{2} d^\top \nabla^2 f^k d.$$

f_k^{Q} の停留条件は $\nabla f_k^{\mathrm{Q}}(d) = \mathbf{0}$ であるから，f^{Q} **が凸であれば**

$$\nabla f_k^{\mathrm{Q}}(d) = \nabla f^k + \nabla^2 f^k d = \mathbf{0}$$

の解を $d = d^k$ とすれば $x^{k+1} = x^k + d^k$ で f_k^{Q} が最小．

Newton 法

Step 0. 初期化：$x^0 \in \mathbb{R}^n$，$k \leftarrow 0$

Step 1. 停止条件を満足するならば，x^k を数値解として出力．

Step 2. $d^k \leftarrow -\left(\nabla^2 f^k\right)^{-1} \nabla f^k$

Step 3. $\alpha^k \leftarrow 1$

Step 4. $x^{k+1} \leftarrow x^k + \alpha^k d^k$

Step 5. $k \leftarrow k+1$ として Step 1 へ．

Quiz 探索方向 d^k が降下方向となるための十分条件を述べよ．なお，降下方向とは $\langle \nabla f^k, d \rangle < 0$ を満足する d だった．

Quiz 目的関数の形状によっては目的関数値が解更新で改悪する場合，最終的に解が振動し収束しない場合がある．1 変数の場合で図示せよ．

Newton 法の性質

- 局所最適解の近傍，すなわち関数がある意味「二次関数っぽい」範囲では非常に収束が速い．(e.g., 目的関数が真に二次なら？)
- 二次近似のよさに決定的に依存．**局所的収束性** (local convergence) は保証されるものの，**大域的収束性** (global convergence) はない．
  - 大域的収束性：任意の初期点から局所最適解へ収束する．
    最急降下法は大域収束性が保証されている．
  - 局所的収束性：局所最適解 $\bar{x}$ に十分近い初期点から $\bar{x}$ へ収束する．

  $\rightarrow$ 最急降下法と Newton 法系の手法を組み合わせることがある．
- 次元が大きな問題では，Step 2 の Hesse 行列の評価と線形方程式の求解のコストが高い．Hesse 行列の計算を回避し，効率的に最適化を行う手法として**準 Newton 法** (quasi-Newton method) がある．

制約つき最適化問題に対する反復法

許容方向降下法 (feasible direction method)

制約付き最適化問題では，d^k を自由に選べるとは限らない．

例えば，最急降下方向 $d^k = -\nabla f^k$ が実行可能領域 S の外側に向かってしまう場合，その方向に降下することはできず，最急降下法は使えない．

各点で，許容領域内であり，かつ目的関数値を減少させる方向 d を選択する方針が考えられる．このような反復法を**許容方向降下法**と呼ぶ．

Frank–Wolfe 法

凸計画問題に対する古典的な許容方向降下法．

x^k において最適化問題を解き d^k を決定する．

$$\begin{aligned} y^k &\leftarrow \operatorname*{arg\,min}_{y \in S} \langle \nabla f(x^k), y \rangle \\ d^k &\leftarrow y^k - x^k \end{aligned}$$

- 各点において目的関数の線形近似を最小化する点の方向へ移動する．
- S を表現する制約関数が全て線形であれば，y^k を決定する問題は線形最適化問題になる．Frank–Wolfe 法は，この線形最適化問題が高速に解けるならば有効．

Quiz このようにして得られる d^k が降下方向であることを確認せよ．

- 制約付凸最適化問題の非常に重要な問題クラスである**線形最適化問題** (linear optimization problem) を学ぶ．