最適化のための基礎

最適化に基づく画像処理の必要性

理由

ビジョン問題では，観測時の雑音やオクルージョンに起因する不確かさが存在し，完全に正確な解を得ることは困難．

⇒ 多くのビジョン問題が最適化問題として定式化される

画像処理の最適化と統計的手法

画像処理における最適化とは、ある基準に基づいて最も良く目的を達成することを目指すもの。

統計的画像処理とは、画像の表現に統計モデルを導入し、そのモデルパラメータを推定することで、結果として目的とする画像を得る手法。

統計的画像処理手法では、MRFやGRFによるモデル化がしばしば利用される。

最適化手法における三要素

１．問題の表現

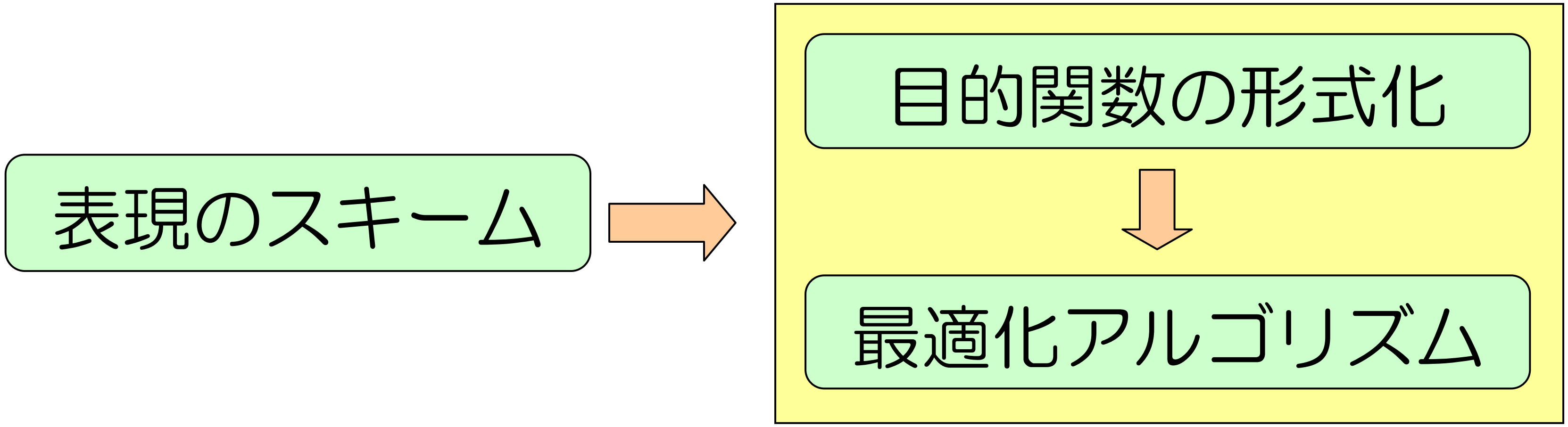
どんな特徴で解を形式化するか

２．目的関数

最適化をどのように基準化するか

３．最適化アルゴリズム

どのようにして最適値を探索するか



- 目的関数とその最適化法とは密接に関係.
- 目的関数としては，エネルギー関数や（対数）尤度関数，事後確率など.
- 最適化手法には様々な方法があって，それぞれ特徴あり.

最適化基準

データ分布のみがあり、推定量の事前情報がない

⇒ 最大尤度 (maximum likelihood)推定

逆の場合

⇒ 最大エントロピー (maximum entropy)推定

推定量の事前情報とデータ分布がともに既知

⇒ Bayes基準

最大事後確率(maximum a posterior;MAP)推定

最大事後確率平均(maximum a posterior mean;)推定

エネルギー関数

エネルギー関数の役割

１．解の大局性を定量的に表現.

２．最小解に対する探索の指針を提供.

エネルギー関数の定式化

基本的方法 $\begin{cases} \text{パラメトリックな表現} \\ \text{ノンパラメトリックな表現} \end{cases}$

最小解 f^*: 関数形式 E とパラメータ θ に依存

$$f^* = \arg\min_f E(f \mid \boldsymbol{d}, \theta)$$

$\boldsymbol{d}$ は観測データ

事象・集合・確率

・ある操作を行うこと：　試行
・試行によって起こる事がら：　事象
・基本的な事象の１つ１つ：　根元事象
・根元事象全体の集合：　$\boldsymbol{\Omega}=\{\Omega_1, \Omega_2, \cdots \Omega_n\}$
・試行の結果起こる事象A：　$A \subseteq \boldsymbol{\Omega}$
・２つの事象AとBが同時には起こらないとき、$A \cap B = \phi$ と書き、２つの事象は互いに排反であるという。
・根元事象は互いに排反である。
・事象Aの確率$P(A)=\dfrac{n(A)}{n(\boldsymbol{\Omega})}$;　$n(A)$:事象Aの要素の数

$$P(\boldsymbol{\Omega})=1$$
$$P(\phi)=0$$

Bayesの定理

$H_n(n=1,\cdots,n):$ 排反、網羅的な事象系列

$E:$ 任意の事象

$$P(H_n \mid E)P(E) = P(E \mid H_n)P(H_n) = P(H_n, E)$$

従って、$P(E) \neq 0$が与えられると、

$$P(H_n \mid E) = \frac{P(H_n)P(E \mid H_n)}{P(E)}$$

$$= \frac{P(E \mid H_n)P(H_n)}{\sum_m P(E \mid H_m)P(H_m)}$$

$$\propto P(E \mid H_n)P(H_n)$$

確率変数と確率分布

『例』サイコロ

サイコロを振る操作：試行

いずれかの目が出る：根元事象

根元事象に割り当てられた数値：実現値{1,2,3,4,5,6}

試行にともなって出る目を表す変数 X：確率変数

確率変数Xの実現値が離散的：離散的確率変数

連続的：連続的確率変数

離散的確率分布

離散的確率変数の各実現値の得られる確率が既知のとき

$$P(\{\Omega_i; X(\Omega_i) = x_i\}) = P(X = x_i) = P_i,$$

$$\sum_{i=1}^{n} P_i = 1, \quad P_i$$：確率関数

と表し、確率変数 X に確率分布が与えられているという。

連続的確率分布

確率変数Xの実現値xが連続な場合、実現値の微小区間を考え、Xの実現値がこの間に入る確率を次のように表す。

$$P(x \le X \le x + \delta x) = p(x)\delta x$$

$p(x)$を確率密度関数といい、$\int_{-\infty}^{+\infty} p(x)dx = 1$

（累積）分布関数 $F(x)$

$$F(x) = \sum_{x_i \le x} P_i$$ 離散的確率分布

$$F(x) = \int_{-\infty}^{x} p(x)dx$$ 連続的確率分布

正規分布

ある連続的確率変数Xが平均μ, 分散σ^2の正規分布に従うとき、

$$P(X=x)=p(x)=\frac{1}{\sqrt{2\pi\sigma^2}}e^{-(x-\mu)^2/2\sigma^2}$$

であり、$x \sim N(\mu,\sigma^2)$と表す。

中心極限定理

同一の分布に従う、互いに独立なN個の離散的確率変数$X_1, X_2, \cdots, X_N$の和で表される確率変数の分布は、適当な変数変換と$N\to\infty$で、正規分布$N(0,1)$となる。

標本抽出と母集団

- 全体から選び出されたもの：標本
- 標本の背後に存在する全体：母集団
- 母集団から標本を抽出する操作：標本抽出

大きな母集団から$x_1, x_2, \cdots, x_n$という値の標本をサンプリング：

⇒ 母集団と同じ確率的構造をもつ，互いに独立なn個の確率変数$X_1, X_2, \cdots, X_n$が，それぞれ$x_1, x_2, \cdots, x_n$の実現値を持つことと同等。

互いに独立で同一な確率分布に従うn個の確率変数:

⇒ i.i.d.(independent identically distributed)な確率変数

乱数発生法

- 一様乱数:: メルセンヌ・ツイスタ法

  M.Matsumoto and T.Nishimura, Mersenne twister: A 623 dimensionally equidistributed uniform pseudorandom number generator, ACM Trans. on Modeling and Computer Simulation,1998

- 正規分布乱数:: ボックス・ミューラ法

  ２個の一様乱数r_1, r_2から，正規分布$N(0,1)$に従う２個の正規乱数z_1, z_2を以下の式で発生.

  $$z_1 = (-2\log r_1)^{1/2}\cos 2\pi r_2$$

  $$z_2 = (-2\log r_1)^{1/2}\sin 2\pi r_2$$

  $N(\mu, \sigma^2)$に従う乱数は、変換 $\varsigma_i = \mu + \sigma^2 z_i$で発生可能.

- その他の分布の乱数::

  棄却サンプリング、SIR、MCMC等の方法で発生

非線形最適化問題

問題の分類

- 非線形最適化問題
  - 制約なし最適化
    - 単峰性関数最適化
    - 多峰性関数最適化
  - 制約つき最適化
    - 等式制約下での最適化
    - 不等式制約下での最適化

手法の分類

- 非線形最適化手法
  - 勾配法
    - 最急降下法
    - Newton法
    - 共役勾配法
  - 直接探索法
    - シンプレックス法
    - 実数値GA

制約なし非線形最適化問題

〔問題1〕

$$\min_{\boldsymbol{x} \in R^n} : f(\boldsymbol{x})$$

ここで、 f は R^n 上で定義される非線形関数.

$$f(x^*) \le f(x) \quad for\ \forall x \in R^n$$

が成り立つとき、 x^* を〔問題１〕の大域的最適解（global optimal solution）という。

$$f(x^*) \le f(x)\ for\ \forall x \in N(x^*, \delta) = \left\{ x \,\middle|\, \left\| x^* - x \right\| \le \delta \right\}$$

が成り立つとき、 x^* を〔問題１〕の局所的最適解（local optimal solution）という。

最適性条件

（仮定）$f : R^n \to R$は2階連続微分可能とする．
$\nabla f(\boldsymbol{x}) = \boldsymbol{0}$を満たす点$\boldsymbol{x}^*$: f の停留点(stationary point)

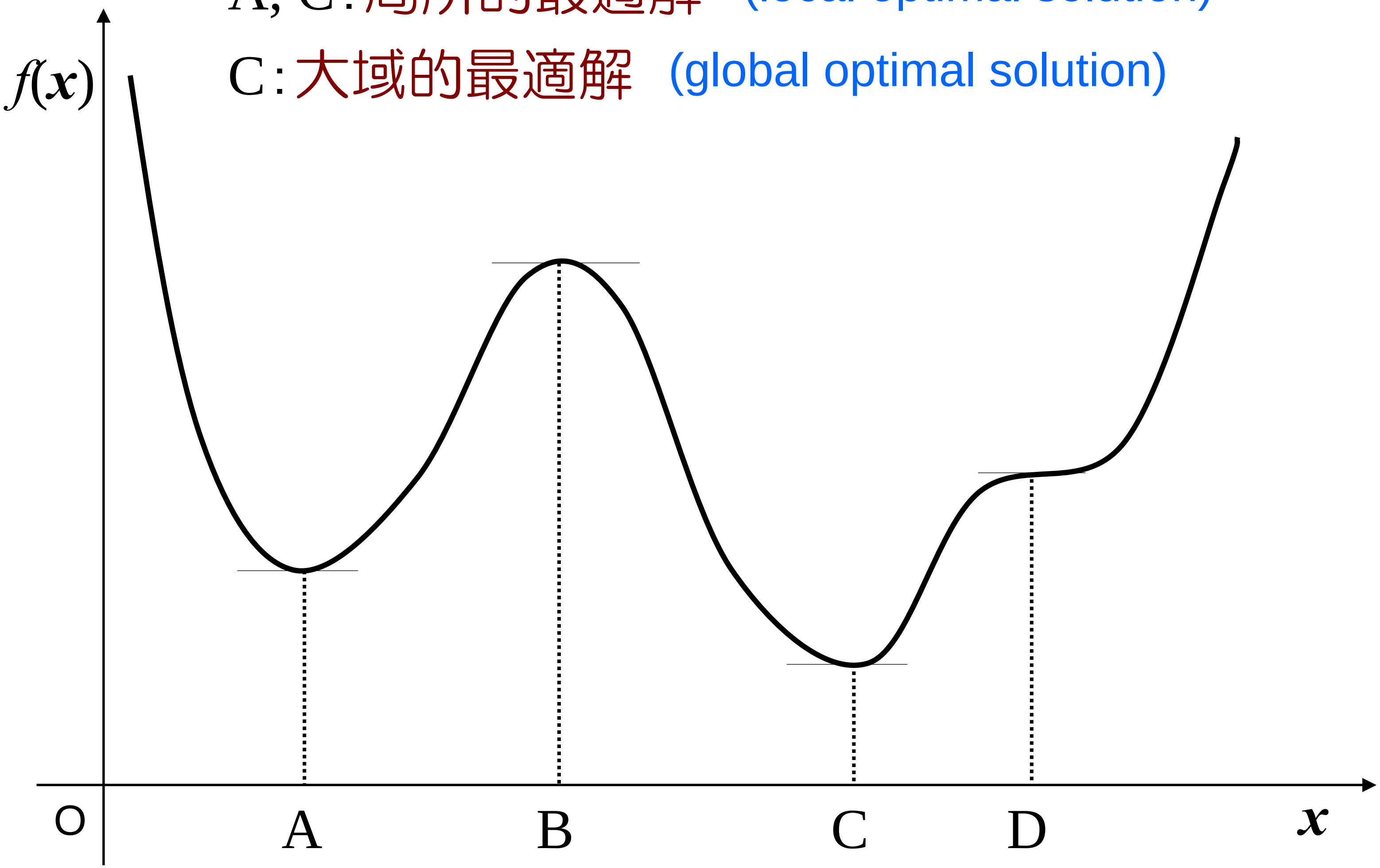
点$\boldsymbol{x}^*$が[問題 1]の局所的最適解ならば，$\nabla f(\boldsymbol{x}^*) = \boldsymbol{0}$.

f が凸関数であるとき，点$\boldsymbol{x}^*$が[問題 1]の大域的最適解であるための必要条件は，$\nabla f(\boldsymbol{x}^*) = \boldsymbol{0}$.

A, B, C, D : 停留点 (stationary point)

A, C : 局所的最適解 (local optimal solution)

C : 大域的最適解 (global optimal solution)



降下方向

$\nabla f(\boldsymbol{x})^T \boldsymbol{d} < 0$なる$\boldsymbol{d}$を、
点 $\boldsymbol{x}$ における f の降下方向と言う.

(n,n)の正定値対称行列 $\boldsymbol{B}$に対して

$$\boldsymbol{d} = -\boldsymbol{B}\nabla f(\boldsymbol{x})$$

とすれば、

$$\nabla f(\boldsymbol{x})^T \boldsymbol{d} = -\nabla f(\boldsymbol{x})^T \boldsymbol{B}\nabla f(\boldsymbol{x}) < 0$$ となる.

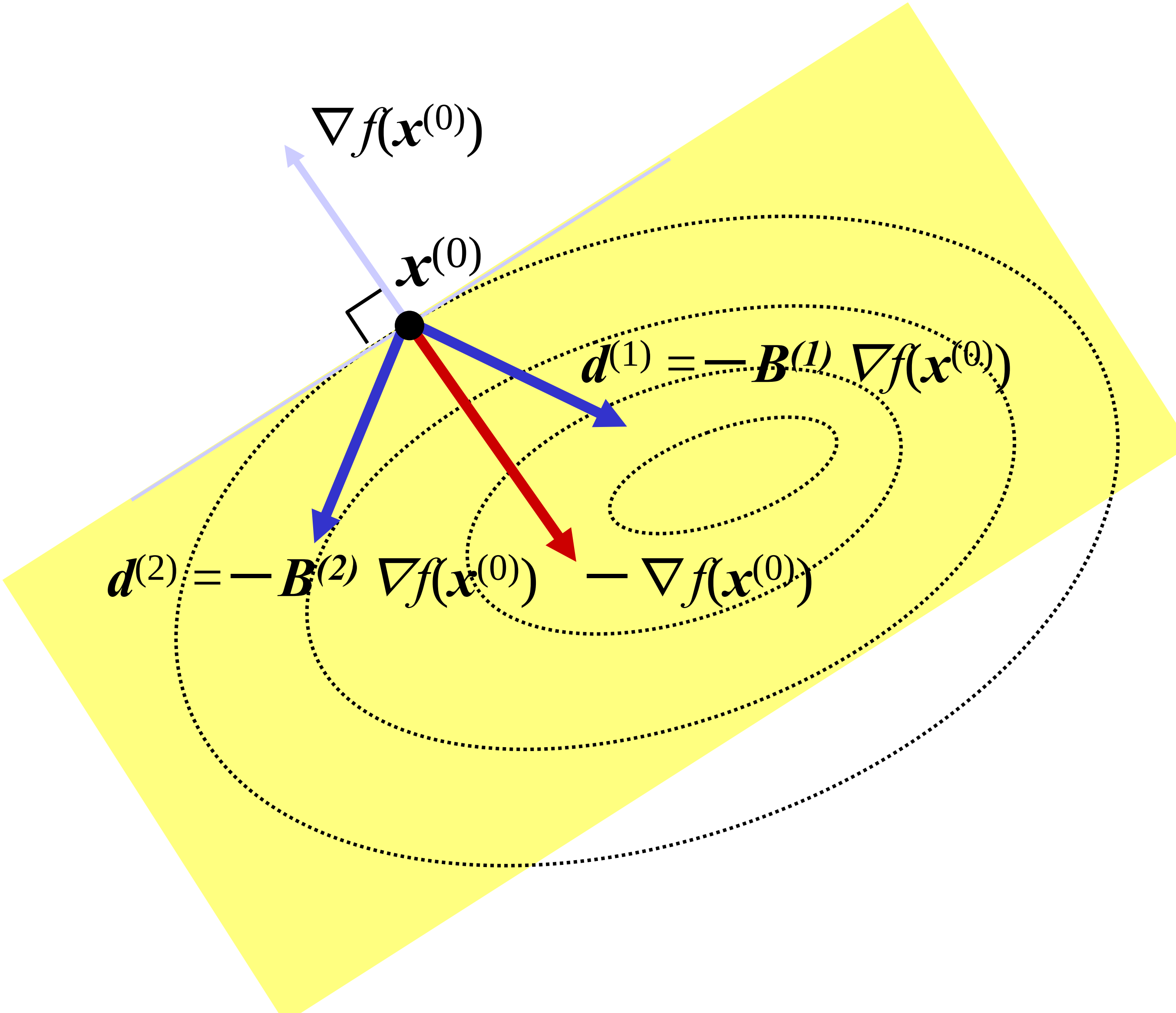
従って、 $\boldsymbol{d} = -\boldsymbol{B}\nabla f(\boldsymbol{x})$ は降下方向条件を満た す.

特に、 Bが単位行列のとき、

$$\boldsymbol{d} = -\nabla f(\boldsymbol{x})$$

を最急降下方向 (steepest descent direction) と言う.

降下方向の例


$\nabla f(\boldsymbol{x}^{(0)})$
$\boldsymbol{x}^{(0)}$
$\boldsymbol{d}^{(1)} = -\boldsymbol{B}^{(1)} \nabla f(\boldsymbol{x}^{(0)})$
$\boldsymbol{d}^{(2)} = -\boldsymbol{B}^{(2)} \nabla f(\boldsymbol{x}^{(0)})$
$-\nabla f(\boldsymbol{x}^{(0)})$

勾配法の概要

Step1: 初期探索点 $\boldsymbol{x}^{(0)}$を選択. $k=0$

Step2 : 適当な正定値対称行列$\boldsymbol{B}^{(k)}$で降下方向$\boldsymbol{d}^{(k)}$を決定.

$$\boldsymbol{d}^{(k)} = -\boldsymbol{B}\nabla f(\boldsymbol{x}^{(k)})$$

Step3 : ステップ幅$a^{(k)}$を決定. ⇒ 直線探索

$$a^{(k)} = \arg\min_{a} f(\boldsymbol{x}^{(k)} + a\boldsymbol{d}^{(k)})$$

を解き、$x^{(k)}$を更新.

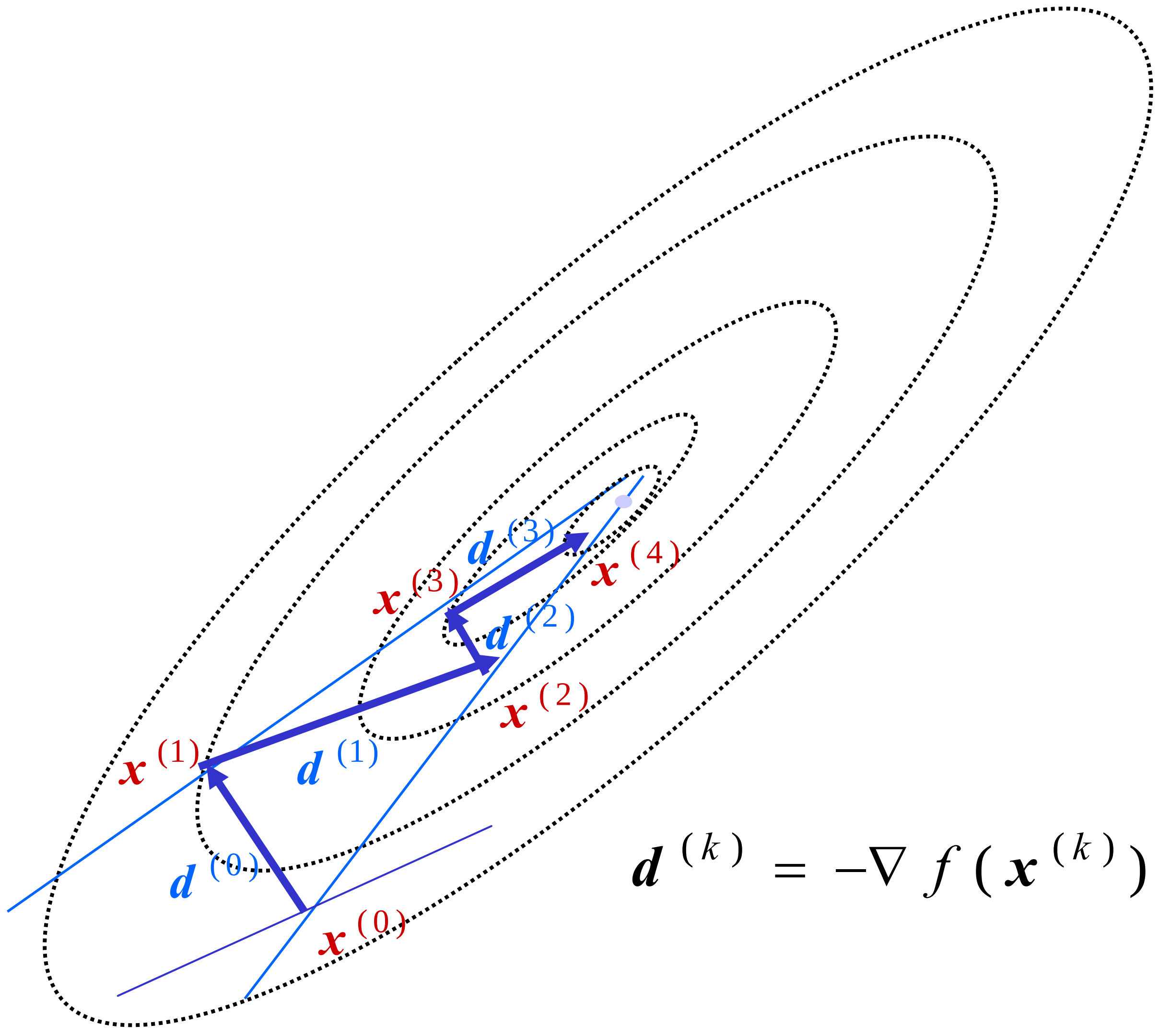
$$\boldsymbol{x}^{(k+1)} = \boldsymbol{x}^{(k)} + a^{(k)}\boldsymbol{d}^{(k)}$$

Step4 : $\nabla f(\boldsymbol{x}^{(k+1)}) = 0$であれば探索終了.

そうでなければ $k = k+1$としてStep2へ.

$\boldsymbol{B}^{(k)}$が単位行列のとき, 最急降下法（steepest descent method）

最急降下法の収束性



$$\boldsymbol{d}^{(k)} = -\nabla f(\boldsymbol{x}^{(k)})$$

ニュートン法の概要

Step1 初期探索点$x^{(0)}$を選択．$k=0$

Step2 正定値行列$\boldsymbol{B}^{(k)}$を，$\boldsymbol{B}^{(k)}=(\nabla^2 f(\boldsymbol{x}^{(k)}))^{-1}$とし，

$$\boldsymbol{d}^{(k)}=-(\nabla^2 f(\boldsymbol{x}^{(k)}))^{-1}\cdot\nabla f(\boldsymbol{x}^{(k)})$$ とする．

Step3 ステップ幅αを，$\alpha^{(k)}=1$ として，

$$\boldsymbol{x}^{(k+1)}=\boldsymbol{x}^{(k)}+\alpha^{(k)}\boldsymbol{d}^{(k)}$$

Step4 $\nabla f(x^{(k+1)})=0$であれば探索終了．
そうでなければ，$k=k+1$ としてStep2へ

Step3で，ステップ幅を$\alpha^{(k)}=\arg\min_{\alpha} f(\boldsymbol{x}^{(k)}+\alpha\boldsymbol{d}^{(k)})$

と直線探索：→ 直線探索付きニュートン法

修正ニュートン法

最適解近傍は凸２次関数で近似可能であり、探索が進めばニュートン法は効率的．しかし,$(\nabla^2 f(x))^{-1}$の存在が保証されない．

$\nabla^2 f(x)$の対角要素に適当な値を加え、正定値化した行列を作成．

⇒ 修正ニュートン法（Levenberg－Marquart法）．

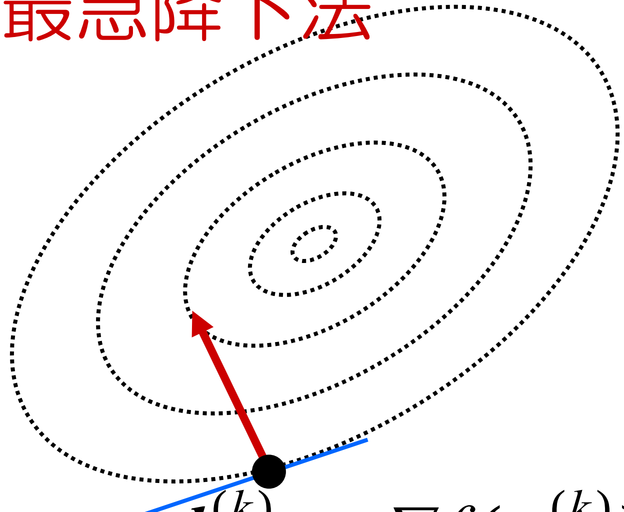
$$\boldsymbol{d}^{(k)} = -\left\{(1-\alpha)\nabla^2 f(\boldsymbol{x}^{(k)}) + \alpha \boldsymbol{I}\right\}^{-1} \nabla f(\boldsymbol{x}^{(k)})$$

ここで,　$0 \le \alpha \le 1$

ニュートン法と最急降下法の混合．

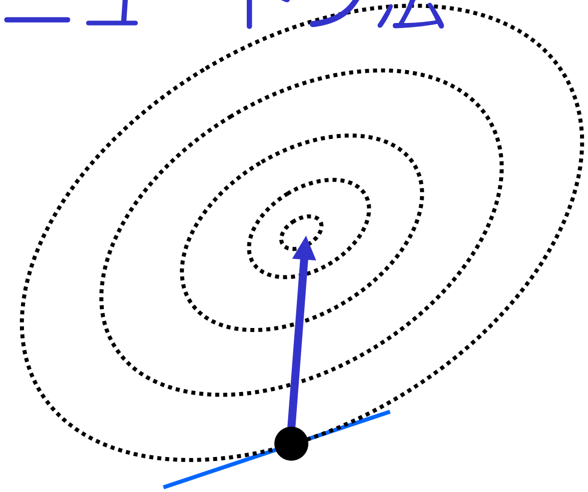
ヘッセ行列の計算が重い．

最急降下法



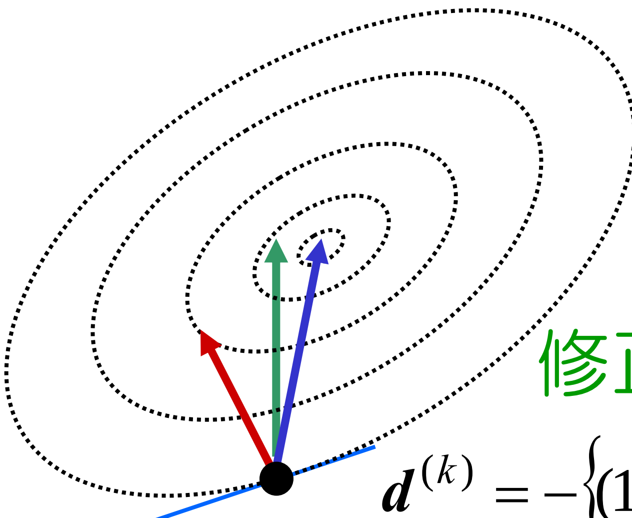
$$\boldsymbol{d}^{(k)} = -\nabla f(\boldsymbol{x}^{(k)})$$

ニュートン法



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

修正ニュートン法



$$\boldsymbol{d}^{(k)} = -\left\{(1-\alpha)\nabla^2 f(\boldsymbol{x}^{(k)}) + \alpha \boldsymbol{I}\right\}^{-1}\nabla f(\boldsymbol{x}^{(k)})$$

Fletcher-Reeves法（共役勾配法）

Step1 初期探索点$\boldsymbol{x}^{(0)}$を選択.

$$\boldsymbol{d}^{(0)} = -\nabla f(\boldsymbol{x}^{(0)}), \quad k = 0$$

Step2 $\nabla f(\boldsymbol{x}^{(k)}) = \boldsymbol{0}$ならば終了.

$$\boldsymbol{d}^{(k)} = -\nabla f(\boldsymbol{x}^{(k)}) + \frac{\nabla f(\boldsymbol{x}^{(k)})^T \nabla f(\boldsymbol{x}^{(k)})}{\nabla f(\boldsymbol{x}^{(k-1)})^T \nabla f(\boldsymbol{x}^{(k-1)})} \boldsymbol{d}^{(k-1)}, k \geq 1$$

Step3 ステップサイズの決定

$\alpha^{(k)} = \arg\min_{\alpha>0} f(\boldsymbol{x}^{(k)} + \alpha \boldsymbol{d}^{(k)})$ を解き,

$\boldsymbol{x}^{(k+1)} = \boldsymbol{x}^{(k)} + \alpha^{(k)} \boldsymbol{d}^{(k)}, k = k+1$ とする.

Step2へ戻る.

等式制約の下での最適化

〔問題４：等式制約〕

$$\min_{\boldsymbol{x}\in R^n} f(\boldsymbol{x}) \quad \text{subj.to} \quad g_i(\boldsymbol{x}) = 0 \quad (i = 1, \cdots, m)$$

ラグランジュ関数

〔問題４：等式制約〕 のラグランジュ関数は

$$L(\boldsymbol{x}, \boldsymbol{\lambda}) = f(\boldsymbol{x}) + \sum_{i=1}^{m} \lambda_i g_i(\boldsymbol{x}),$$

$\lambda_i (i = 1, \cdots, m)$:ラグランジュ乗数

ラグランジュの未定乗数法

〔問題４〕のラグランジュ関数が$\boldsymbol{x}^*$と$\boldsymbol{\lambda}^*$で局所的最適解を持つための必要条件は,

$$\nabla_{\boldsymbol{x}} L(\boldsymbol{x},\boldsymbol{\lambda})\Big|_{\boldsymbol{x}=\boldsymbol{x}^*,\boldsymbol{\lambda}=\boldsymbol{\lambda}^*}=\boldsymbol{0},\quad \nabla_{\boldsymbol{\lambda}} L(\boldsymbol{x},\boldsymbol{\lambda})\Big|_{\boldsymbol{x}=\boldsymbol{x}^*,\boldsymbol{\lambda}=\boldsymbol{\lambda}^*}=\boldsymbol{0}$$

$$\nabla_{\boldsymbol{x}} L(\boldsymbol{x},\boldsymbol{\lambda})\Big|_{\boldsymbol{x}=\boldsymbol{x}^*,\boldsymbol{\lambda}=\boldsymbol{\lambda}^*}=\nabla f(\boldsymbol{x}^*)+\sum_{i=1}^{m}\lambda_i^*\nabla g_i(\boldsymbol{x}^*)=\boldsymbol{0}$$

$$\nabla_{\boldsymbol{\lambda}} L(\boldsymbol{x},\boldsymbol{\lambda})\Big|_{\boldsymbol{x}=\boldsymbol{x}^*,\boldsymbol{\lambda}=\boldsymbol{\lambda}^*}=(g_i(\boldsymbol{x}^*),i=1,\cdots,m)^T=\boldsymbol{0}$$

⇨ 〔問題4〕のKKT条件と等しい.

ラグランジュ未定乗数法（最小化問題）の例

$$\min_{x_1, x_2, x_3} f = x_1^2 + x_2^2 + x_3^2 \quad \text{subj.to} \quad 6x_1 + 2x_2 + 4x_3 = 42$$

ラグランジュ関数は,

$$L(x_1, x_2, x_3, \lambda) = (x_1^2 + x_2^2 + x_3^2) + \lambda(6x_1 + 2x_2 + 4x_3 - 42)$$

最適性条件より

$$\nabla_{\boldsymbol{x}} L = \begin{pmatrix} 2x_1 \\ 2x_2 \\ 2x_3 \end{pmatrix} + \lambda \begin{pmatrix} 6 \\ 2 \\ 4 \end{pmatrix} = \boldsymbol{0}, \quad \nabla_{\lambda} L = 6x_1 + 2x_2 + 4x_3 - 42 = 0.$$

が得られる．各変数をλで表し，制約式に代入すれば，

$\lambda^* = -3/2$．従って，最適解は$(9/2, 3/2, 3)$，最適値は$63/2$．

EMアルゴリズム

EMアルゴリズムとは，不完全データから最尤推定値を求める理論的な枠組み。(1977, Dempster, Laird, Rubin)

不完全データとは，欠損値を含むデータだけではなく，本来観測できない変数（隠れ変数や潜在変数）も含めたデータの総称.

観測データ集合 $\boldsymbol{D}$，隠れ変数集合 $\boldsymbol{Z}$，モデルパラメータ $\boldsymbol{\theta}$ とすると，対数尤度関数は

$$L(\boldsymbol{D},\boldsymbol{\theta}) = \log p(\boldsymbol{D} \mid \boldsymbol{\theta}) = \log \sum_{Z} p(\boldsymbol{D},\boldsymbol{Z} \mid \boldsymbol{\theta})$$

EM法は，対数尤度関数の代わりに，次に示す条件付期待値を逐次的に最大化する.

$$Q(\boldsymbol{\theta} \mid \boldsymbol{\theta}^{(t)}) = E_Z\{\log p(\boldsymbol{D}, \boldsymbol{Z} \mid \boldsymbol{\theta}) \mid \boldsymbol{D}, \boldsymbol{\theta}^{(t)}\}$$
$$= \sum_{\boldsymbol{Z}} P(\boldsymbol{Z} \mid \boldsymbol{D}, \boldsymbol{\theta}^{(t)}) \log p(\boldsymbol{D}, \boldsymbol{Z} \mid \boldsymbol{\theta})$$

θに関して以下のように逐次的に最大化を行う.

EMアルゴリズム

Step1. 初期値$\theta^{(0)}$を設定し, $t \leftarrow 0$とする.

Step2. 収束するまで以下の処理を繰り返す.

E-step : $Q(\theta \mid \theta^{(t)})$を計算.

M-step : $\theta^{(t+1)} = \arg\max_{\theta} Q(\theta \mid \theta^{(t)})$とし, $t \leftarrow t+1$とする.

Q関数の最大化 ⇨ 対数尤度の最大化（最尤法）

正規混合分布の最尤推定

正規混合分布：m個の正規分布の混合からなる分布

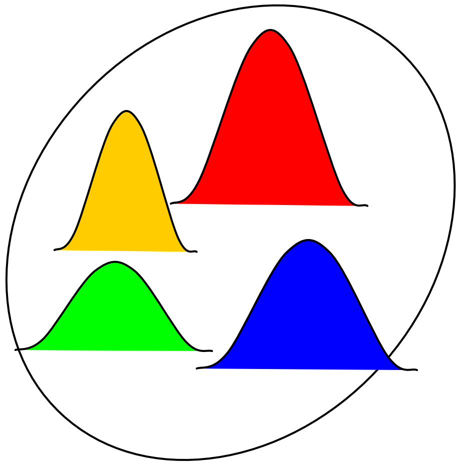
$$p(\boldsymbol{x}\,|\,\boldsymbol{\theta}) = \sum_{i=1}^{m} \alpha_i N(\boldsymbol{x}; \boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i), \qquad \text{ただし}, \sum_{i=1}^{m} \alpha_i = 1$$

$N(\cdot\,; \boldsymbol{\mu}, \boldsymbol{\Sigma})$は，平均ベクトル$\boldsymbol{\mu} \in R^d$，共分散行列$\boldsymbol{\Sigma} \in R^{d \times d}$を持つ多次元正規分布．

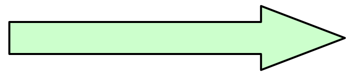
$$N(\boldsymbol{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = (2\pi)^{-d/2} \, |\boldsymbol{\Sigma}|^{-1/2} \exp\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\}$$

観測データ集合 $\boldsymbol{D} = \{\boldsymbol{x}_j\}_{j=1}^{N}$から，最尤法によって未知パラメータ $\boldsymbol{\theta} = \{\alpha_i, \boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i\}_{i=1}^{m}$を推定する問題．

画像の統計モデル

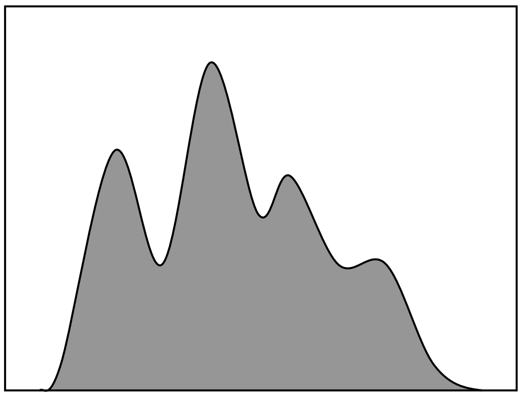


m個の正規分布の混合分布
として画像が存在（仮定）



多変量混合正規分布モデル

$$p(\boldsymbol{x} \mid \boldsymbol{\theta}) = \sum_{i=1}^{m} \alpha_i N(\boldsymbol{x}; \boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$$



EMアルゴリズムによる推定

隠れ変数$z_i \in \boldsymbol{Z} = \{z_j\}_{j=1}^{N}$は，観測データ$\boldsymbol{x}_i$がどの混合成分から発生したのかを示す指標で，$\{1,\cdots,m\}$のいずれかの値をとる．

$$P(z_n = i \mid \boldsymbol{x}_n, \boldsymbol{\theta}^{(t)}) = \frac{p(\boldsymbol{x}_n, z_n = i \mid \boldsymbol{\theta}^{(t)})}{p(\boldsymbol{x}_n \mid \boldsymbol{\theta}^{(t)})} = \frac{p(\boldsymbol{x}_n, z_n = i \mid \boldsymbol{\theta}^{(t)})}{\sum_{i=1}^{m} p(\boldsymbol{x}_n, z_n = i \mid \boldsymbol{\theta}^{(t)})}$$

$$p(\boldsymbol{x}_n, z_n = i \mid \boldsymbol{\theta}) = \alpha_i N(\boldsymbol{x}_n; \boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$$

より，

$$Q(\boldsymbol{\theta} \mid \boldsymbol{\theta}^{(t)}) = \sum_{i=1}^{m} \sum_{n=1}^{N} P(z_n = i \mid \boldsymbol{x}_n, \boldsymbol{\theta}^{(t)}) \log p(\boldsymbol{x}_n, z_n = i \mid \boldsymbol{\theta})$$

$$= \sum_{i=1}^{m} \sum_{n=1}^{N} \frac{\alpha_i^{(t)} N(\boldsymbol{x}_n; \boldsymbol{\mu}_i^{(t)}, \boldsymbol{\Sigma}_i^{(t)})}{\sum_{j=1}^{m} \alpha_j^{(t)} N(\boldsymbol{x}_n; \boldsymbol{\mu}_j^{(t)}, \boldsymbol{\Sigma}_j^{(t)})} \log\{\alpha_i N(\boldsymbol{x}_n; \boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)\}$$

M-step：等式拘束条件$(\sum_{i=1}^{m} \alpha_i = 1)$つき$Q(\theta \mid \theta^{(t)})$の最大化.

⟹ ラグランジェの未定乗数法による解法

$G(\boldsymbol{\theta}, \lambda) = Q(\boldsymbol{\theta} \mid \boldsymbol{\theta}^{(t)}) + \lambda(\sum_{i=1}^{m} \alpha_i - 1)$として（ラグランジェ関数）

$$\nabla_{\boldsymbol{\theta}} G(\boldsymbol{\theta}, \lambda) \mid_{\boldsymbol{\theta}=\boldsymbol{\theta}^*, \lambda=\lambda^*} = \boldsymbol{0}, \quad \nabla_{\lambda} G(\boldsymbol{\theta}, \lambda) \mid_{\boldsymbol{\theta}=\boldsymbol{\theta}^*, \lambda=\lambda^*} = \boldsymbol{0}$$

を満たす(θ^*, λ^*)を求める.

これを各成分毎に表すと,

$$\frac{\partial Q(\theta \mid \theta^{(t)})}{\partial \mu_i} = 0, \frac{\partial Q(\theta \mid \theta^{(t)})}{\partial \Sigma_i^{-1}} = 0, \frac{\partial G(\theta, \lambda)}{\partial \alpha_i} = 0$$

が条件となる.

以上のことを踏まえると，以下の更新式が得られる。

$$\boldsymbol{\mu}_i^{(t+1)} = \frac{1}{N_i^{(t)}} \sum_{n=1}^{N} P(z_n = i \mid \boldsymbol{x}_n, \boldsymbol{\theta}^{(t)}) \boldsymbol{x}_n,$$

$$\boldsymbol{\Sigma}_i^{(t+1)} = \frac{1}{N_i^{(t)}} \sum_{n=1}^{N} P(z_n = i \mid \boldsymbol{x}_n, \boldsymbol{\theta}^{(t)}) \boldsymbol{V}_{ni}^{(t+1)},$$

$$\alpha_i^{(t+1)} = \frac{N_i^{(t)}}{N}$$

ただし，
$$\boldsymbol{V}_{ni}^{(t+1)} = (\boldsymbol{x}_n - \boldsymbol{\mu}_i^{(t+1)})(\boldsymbol{x}_n - \boldsymbol{\mu}_i^{(t+1)})^T,$$

$$N_i^{(t)} = \sum_{n=1}^{N} P(z_n = i \mid \boldsymbol{x}_n, \boldsymbol{\theta}^{(t)})$$

$\alpha_i^{(0)}, \mu_i^{(0)}, \Sigma_i^{(0)}$に適当な初期値を与え，上記反復処理を実行.