

導入と数学的準備

中野 慎也

「データ科学: 理論から実用へ」

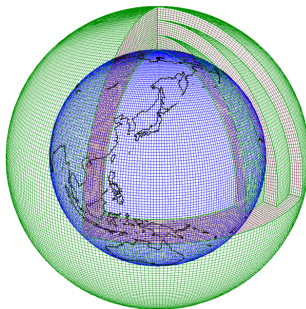
7 September 2023

本講義の概要

- 1 導入と数学的準備
- 2 最小二乗法とベイズ推定
- 3 カルマンフィルタ
- 4 アンサンブルカルマンフィルタ
- 5 アンサンブル変換カルマンフィルタ
- 6 4次元変分法の基礎
- 7 粒子フィルタ, 局所化, その他の話題

データ同化とは

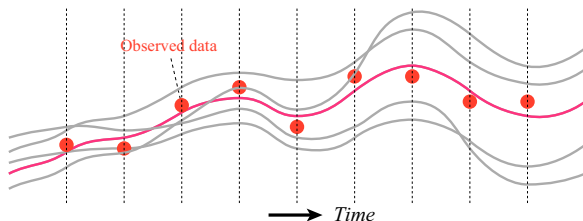
- システムの時間発展を記述するシミュレーションモデルと、観測とを組み合わせ、システム全体の時間発展を推定しようという考え方.
- もともとは、気象学の分野で発展.
 - 気象学では、大気の物理過程を記述した数値シミュレーションモデルが発達.
 - しかし、シミュレーションの初期条件をどう与えるかは問題. システム全体の物理量をすべて観測して、初期値とするのは不可能.



気象庁の Web ページより.

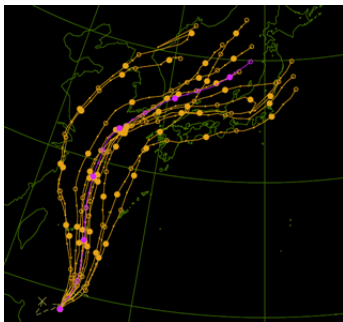
データ同化とは

- 数値シミュレーションを行う際には、初期条件、境界条件、パラメータなどを適切に与える必要があるが、その正確な値はわからない場合が多い。
- 初期条件や境界条件、モデルの不確かさなどを考慮すると、様々なシナリオが考えられる。
- データ同化は、実際の観測データをなるべくうまく説明する尤もらしいシナリオを提示する。



データ同化とは

- 気象シミュレーションによる予測は、初期条件によって大きく変わる。
- しかし、シミュレーションでの予測に必要な細かさで気圧や気温の空間分布を観測するのは不可能。

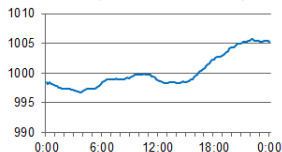


気象庁の Web ページより.

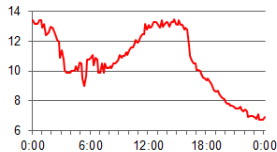
データ同化とは

- 特に気象の場合は、ある場所の予測に、別の場所の情報が有用である。
 - 台風は南方の海で発生して日本列島まで移動してくる。
 - 気温には周辺の気圧の空間分布が影響する。
- 別の場所の情報は、物理法則にしたがって伝搬する。
- 物理法則に基づいて作られたシミュレーションモデル自体が、別の場所の情報を活用するのに役立つ。

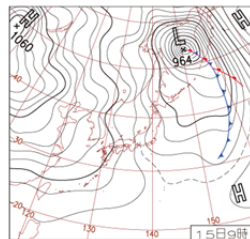
2010年12月15日の東京の気圧



2010年12月15日の東京の気温



2010年12月15日の天気図



データ同化の目的

- 実測データを用いて数値シミュレーションモデルの精度，性能を改善する。
 - 適切な初期条件の設定．（現在からスタートさせる場合，過去からスタートさせる場合を含む．）
 - 適切な境界条件，その他のパラメータの設定．
- 物理法則など，システムの時間発展に関する我々の知見が埋め込まれたシミュレーションモデルを用いることで，観測の不足を補ったり，観測誤差を修正したりする。
 - 観測の得られない時間・場所における物理量の推定．
 - そもそも観測できない物理量の推定．

具体的なデータ同化の用途

■ 気象予報・予測.

- 気象システムは、初期値鋭敏性を持つが決定論的なので、現在だけでなく、過去のデータも活用してよい初期値を作れば、将来予測の性能も上がることが期待できる.

■ 再解析データの生成.

- 過去の気象現象について調べる際に、観測データだけでは情報が不十分なことが多い. 過去の観測をデータ同化した出力結果 (再解析データと呼ばれる) は、その時の気象システムの全体像に関して、信頼できる推定値となる.

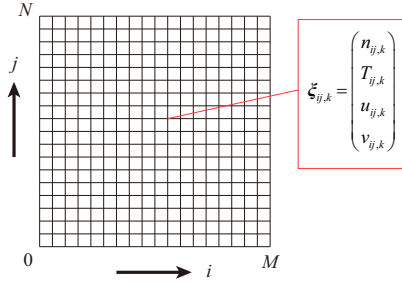
表記法

- a : ベクトル (確率変数)
- A : 行列
- a^\top, A^\top : ベクトル, 行列の転置
- x : 状態変数
- y : 観測値
- k : 時間ステップ
- t_k : k 番目の時間ステップに対応する時刻
 - y_1, \dots, y_k をまとめて, $y_{1:k}$ などと表記することもある.
 - $p(x_k | y_{1:k-1}) = p(x_k | y_1, \dots, y_{k-1})$ など.
- $\mathcal{N}(\mu, \Sigma)$: 平均 μ , 分散共分散行列 Σ のガウス分布 (正規分布)

定式化

まず、シミュレーションの中で扱っている全ての変数の時刻 t_k における値を一つのベクトルにまとめて x_k とおく。

各グリッドに種々の変数が割り当てられている場合、ある時刻 t_k 、あるグリッド (i, j) での全変数の値をまとめたベクトルを $\xi_{ij,k}$ とすると、

$$x_k = \begin{pmatrix} \xi_{00,k} \\ \vdots \\ \xi_{M0,k} \\ \xi_{01,k} \\ \vdots \\ \xi_{M1,k} \\ \vdots \\ \xi_{0N,k} \\ \vdots \\ \xi_{MN,k} \end{pmatrix}.$$


The diagram shows a grid with horizontal axis i (from 0 to M) and vertical axis j (from 0 to N). A red arrow points from a specific grid cell (i, j) to a box containing the vector $\xi_{ij,k} = \begin{pmatrix} n_{ij,k} \\ T_{ij,k} \\ u_{ij,k} \\ v_{ij,k} \end{pmatrix}$.

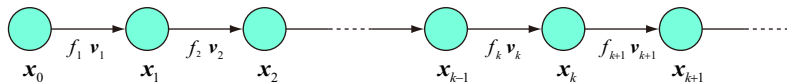
システムモデル

- 時刻 t_{k-1} の状態 x_{k-1} からシミュレーションをスタートさせて、時刻 t_k まで走らせたことによる x の時間発展を

$$x_k = f_k(x_{k-1}) + v_k \quad (1)$$

のように f_k で表現する。

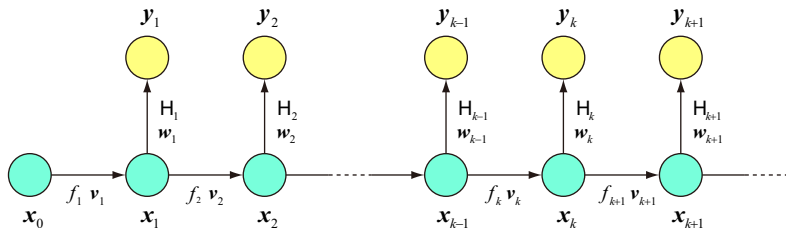
- ここで、 $x_k = f_k(x_{k-1})$ としないのは、シミュレーションがシステムの時間発展を精確に記述できない可能性があるため。
- v_k をシステムノイズと呼ぶ。



観測モデル

- 用いる観測データを一つのベクトルにまとめて、 y_k とする。
- シミュレーション変数 x_k と観測との関係を H_k で表し、以下のように書く：

$$y_k = H_k x_k + w_k. \quad (2)$$



観測モデル

- 観測モデル $y_k = H_k x_k + w_k$ の行列 H_k は、例えば、シミュレーション変数 x_k の中の x_1, x_2, x_3 のみが直接観測できた場合、

$$H_k = \begin{pmatrix} 1 & 0 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 1 & 0 & \cdots & 0 \end{pmatrix} \quad (3)$$

のような行列を定義すればよい。(周囲のグリッドで重み付き平均を取って観測と比較する場合もある。)

- 積分量のデータなど、もっと複雑な形になる場合もある。

観測モデル

- シミュレーション変数と観測データとの関係を

$$y_k = H_k x_k + w_k.$$

と書いたが、 $y_k = H_k x_k$ としないのは、

- データには観測誤差が含まれている。
- いくらパラメータや初期条件を改善しても、数値シミュレーションで現実の世界を完璧に表現できるわけではない。(シミュレーションモデルには、時間・空間分解能に限界があり、それ以外にも様々な近似が使われている。)

といったことを考慮するため。

- w_k を観測ノイズと呼ぶ。

定式化

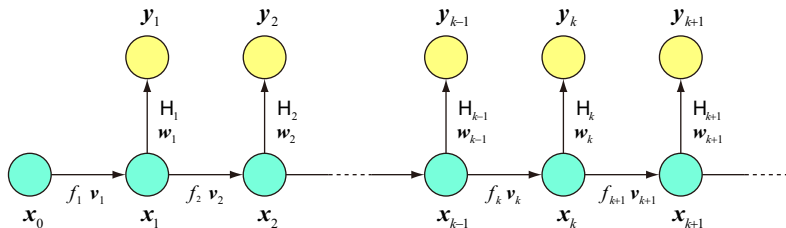
- 結局，ここで取り扱うべき変数同士の関係は，

$$\mathbf{x}_k = \mathbf{f}_k(\mathbf{x}_{k-1}) + \mathbf{v}_k, \quad (4a)$$

$$\mathbf{y}_k = \mathbf{H}_k \mathbf{x}_k + \mathbf{w}_k \quad (4b)$$

の 2 種類の式で表現できる．

- この 2 つをまとめて (非線型) 状態空間モデルと呼ぶ．
- データ同化は，上のような関係が与えられ，更に各時刻の観測データ \mathbf{y}_k が与えられた下で，各時刻の \mathbf{x}_k を推定する問題と捉えられる．



やや具体的な例

常微分方程式 $\frac{dx}{dt} = g(x)$ で「近似的に」表されるようなシステムを考える。
このシステムにおいて、時刻 t_{k-1} から t_k までの x の時間発展は、

$$x(t_k) \approx x(t_{k-1}) + \int_{t_{k-1}}^{t_k} g(x(\tau)) d\tau. \quad (5)$$

$x_k = x(t_k)$, $x_{k-1} = x(t_{k-1})$ とし、さらに、関数 f_k を

$$f_k(x_{k-1}) = x(t_{k-1}) + \int_{t_{k-1}}^{t_k} g(x(\tau)) d\tau \quad (6)$$

と定義すると、 $x_k \approx f_k(x_{k-1})$ となる。

やや具体的な例

$$\boldsymbol{x}_k \approx \boldsymbol{f}_k(\boldsymbol{x}_{k-1})$$

の $\boldsymbol{f}_k(\boldsymbol{x}_{k-1})$ は、近似式から得られた \boldsymbol{x}_k の予測であることを考慮し、(未知の) 予測誤差 \boldsymbol{v}_k を導入して、

$$\boldsymbol{x}_k = \boldsymbol{f}_k(\boldsymbol{x}_{k-1}) + \boldsymbol{v}_k.$$

観測については、 \boldsymbol{x}_k の各成分がノイズ \boldsymbol{w}_k つきで観測されると仮定すると、

$$\boldsymbol{y}_k = \boldsymbol{x}_k + \boldsymbol{w}_k.$$

これは \mathbf{H}_k を単位行列 \mathbf{I} とすれば、

$$\boldsymbol{y}_k = \mathbf{H}_k \boldsymbol{x}_k + \boldsymbol{w}_k.$$

データ同化問題の特徴

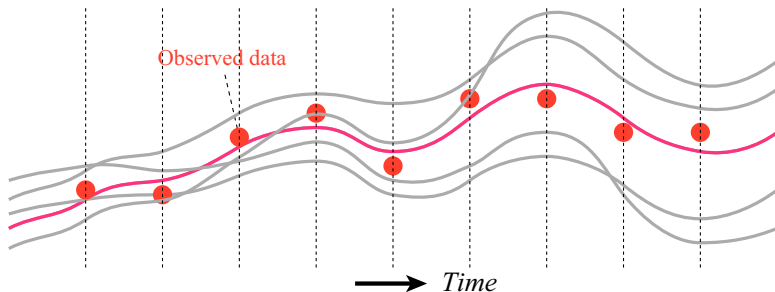
- 高次元ベクトル，あるいは空間構造を推定する．
- 大量かつ多様なデータを用いる．
- 高次元ベクトルの推定だけでなく，時間変動も推定する．
- システムの非線型性も考慮する必要がある．
- 時間変動の計算にコストが掛かる．

主なデータ同化手法

- 静的データ同化手法
 - 最適内挿法
 - 3次元変分法
- 逐次データ同化手法
 - (拡張) カルマンフィルタ
 - アンサンブルカルマンフィルタ
 - アンサンブル変換カルマンフィルタ
- 4次元変分法
 - アジョイント法
 - アンサンブル 4次元変分法

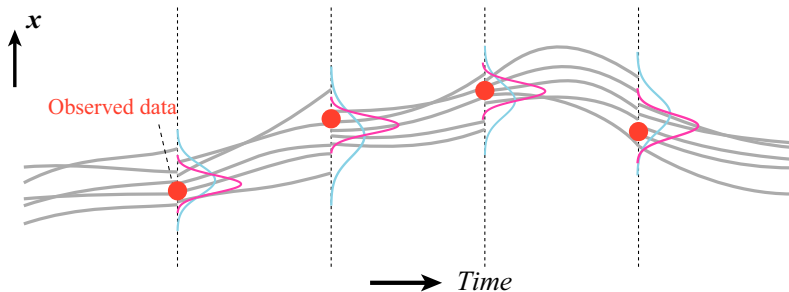
4 次元変分法

- 初期値やパラメータを調整して、ある一定の期間に得られたデータすべてに適合するようなシナリオを求める。
- 基本的にシミュレーションを信用し、 $x_k = f_k(x_{k-1})$ という拘束条件の下で $\sum_{i=1}^N \|y_k - H_k x_k\|^2$ を最小化する x_k を求めるという手続きになる。



逐次データ同化手法

- 時間ステップごとに推定を行う。
- まず前の時間ステップからの予測を行い、データの情報を取り入れて、その予測を修正する。
- 得られた推定結果を、さらにその次の時間ステップの予測に用いる。



平均と分散共分散行列

ここで、データ同化でよく用いられる確率分布、特に正規分布の性質について概観しておく。

m 次元確率変数 x の平均 μ ，分散共分散行列 P は以下のように定義される：

$$\mu = E[x], \quad (7)$$

$$P = E[(x - \mu)(x - \mu)^T]. \quad (8)$$

但し、 E は期待値を表し、上付きの T は転置行列を表す。

x の確率密度関数 $p(x)$ が与えられているとき、 μ 、 P はそれぞれ、

$$\mu = \int x p(x) dx, \quad (9)$$

$$P = \int (x - \mu)(x - \mu)^T p(x) dx \quad (10)$$

のように得られる。但し、積分領域は x の定義域全体とする。

分散共分散行列について

$$\mathbf{P} = E[(\mathbf{x} - \boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})^T] \quad (11)$$

で定義される分散共分散行列 \mathbf{P} は、 m 次の正方行列で半正定値対称行列となる。
すなわち、任意の m 次元ベクトル \mathbf{z} に対して $\mathbf{z}^T \mathbf{P} \mathbf{z} \geq 0$ が成り立ち、かつ
 $\mathbf{P}^T = \mathbf{P}$ となる。

(正定値とは限らないことに注意。)

対称行列になることは定義より明らかだが、半正定値性についても

$$\begin{aligned} \mathbf{z}^T \mathbf{P} \mathbf{z} &= \mathbf{z}^T E[(\mathbf{x} - \boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})^T] \mathbf{z} \\ &= E[\mathbf{z}^T (\mathbf{x} - \boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})^T \mathbf{z}] \\ &= E[(\mathbf{z}^T (\mathbf{x} - \boldsymbol{\mu}))^2] \geq 0 \end{aligned}$$

のように直ちに確認できる。

分散共分散行列について

実対称行列は直交行列で対角化でき、得られる対角行列の対角要素 (固有値) はすべて実数となる。さらに、半正定値行列ならば、固有値はすべて非負。

このことから、分散共分散行列 P も直交行列で対角化でき、固有値はすべて非負となる。

さらに、固有値がすべて正の (0 でない) 実数の場合 (正定値の場合)、 P は正則行列となる (逆行列 P^{-1} を持つ)。このとき、 P を

$$P = U\Lambda U^T \quad (12)$$

のように固有値分解すると、 P^{-1} は

$$P^{-1} = U\Lambda^{-1}U^T \quad (13)$$

のように得られる。

なお、 $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_m)$ とおいたとき、 $\Lambda^{-1} = \text{diag}(1/\lambda_1, \dots, 1/\lambda_m)$ であり、各対角要素 $1/\lambda_i$ は正の実数となるので、 P^{-1} も正定値。

同時分布と周辺分布

2つの確率ベクトル x_a, x_b を考えたとき, x_a, x_b をまとめた

$$\boldsymbol{x}' = \begin{pmatrix} \boldsymbol{x}_a \\ \boldsymbol{x}_b \end{pmatrix} \quad (14)$$

の確率分布を同時確率分布あるいは同時分布と言う.

同時分布の確率密度関数 (同時確率密度関数) $p(\boldsymbol{x}_a, \boldsymbol{x}_b)$ が与えられているとき,

$$\iint p(\boldsymbol{x}_a, \boldsymbol{x}_b) d\boldsymbol{x}_a d\boldsymbol{x}_b = 1, \quad (15)$$

$$\int p(\boldsymbol{x}_a, \boldsymbol{x}_b) d\boldsymbol{x}_a = p(\boldsymbol{x}_b), \quad \int p(\boldsymbol{x}_a, \boldsymbol{x}_b) d\boldsymbol{x}_b = p(\boldsymbol{x}_a) \quad (16)$$

が成り立つ. 同時分布から, 式 (16) のような計算によって得られる確率密度関数 $p(\boldsymbol{x}_a), p(\boldsymbol{x}_b)$ を, それぞれ x_a, x_b の周辺分布と呼び, 周辺分布を求める式 (16) の操作を周辺化と呼ぶ.

条件付き分布

x_a, x_b のうち, x_a の値が与えられたときの x_b の分布を, x_a が与えられたもとの x_b の条件付き分布と呼び, その確率密度関数は $p(x_b|x_a)$ のように書く.

同時確率密度関数と条件付き分布の確率密度関数との間には,

$$p(x_a, x_b) = p(x_b|x_a)p(x_a) = p(x_a|x_b)p(x_b) \quad (17)$$

が成り立つ. 式 (16) の周辺化の操作は, 式 (17) を適用した

$$\int p(x_b|x_a)p(x_a) dx_a = p(x_b), \quad \int p(x_a|x_b)p(x_b) dx_b = p(x_a) \quad (18)$$

の形で書かれることも多い.

独立な確率変数の性質

同時確率密度関数 $p(x_a, x_b)$ が任意の x_a, x_b に対して,

$$p(x_a, x_b) = p(x_a)p(x_b) \quad (19)$$

を満たすとき, x_a と x_b が独立であると言う. x_a と x_b が独立のとき, 式 (17) より,

$$p(x_b|x_a) = p(x_b), \quad p(x_a|x_b) = p(x_a) \quad (20)$$

が成り立つ.

確率密度関数の変数変換

ある単射で微分可能なベクトル値関数 g を用いて, m 次元の確率変数 x と m 次元確率変数 u が

$$x = g(u)$$

のように対応付けられているとする. x の確率密度関数 $p(x)$ が

$$p(x) = f(x) \tag{21}$$

の形で与えられたとき, u の確率密度関数は,

$$p(u) = f(g(u)) \left| \frac{\partial g}{\partial u^\top} \right| \tag{22}$$

のように書くことができる. 但し, $\left| \frac{\partial g}{\partial u^\top} \right|$ は関数 g のヤコビ行列の行列式の絶対値を表し, 式 (22) が成り立つには, u が定義される任意の点で $\left| \frac{\partial g}{\partial u^\top} \right| \neq 0$ が成り立つ必要がある.

確率密度関数の変数変換

$p(x)$ から $p(u)$ を得るときにヤコビ行列式を掛けるのは,

$$\begin{aligned} P(x \in A) &= \int_A f(x) dx = \int_{g(A)} f(g(u)) \left| \frac{\partial g}{\partial u^\top} \right| du \\ &= \int_{g(A)} p(u) du = P(u \in g(A)) \end{aligned} \quad (23)$$

のように、 x が領域 A に含まれる確率を領域 A の g による像 $g(A)$ に u が含まれる確率と対応づけるためである。

(確率密度関数は、積分して確率を与える関数なので、ヤコビ行列式を掛けて辻褄を合わせる必要がある。)

式 (22) のように x の確率密度関数 $p(x)$ を u の確率密度関数 $p(u)$ に変換する操作は、 x から u への変数変換と呼ばれる。

正規分布

1 次元 (1 変量) の確率変数 z の確率密度関数が

$$p(z) = \frac{1}{\sqrt{2\pi\sigma_z^2}} \exp \left[-\frac{(z - \mu_z)^2}{2\sigma_z^2} \right] \quad (24)$$

のように書けるとき、 z が平均 μ_z 、分散 σ_z^2 の正規分布にしたがうと言う。

平均が μ_z 、分散が σ_z^2 なので、

$$E[z] = \int z p(z) dz = \mu_z, \quad (25)$$

$$E[(z - \mu_z)^2] = \int (z - \mu_z)^2 p(z) dz = \sigma_z^2 \quad (26)$$

が成り立つ。(計算過程は省略。)

標準正規分布

正規分布のうち、平均が $\mu_z = 0$ 、分散が $\sigma_z^2 = 1$ となるものを特に (1 次元) 標準正規分布と呼ぶ。

さらに、 m 個の確率変数 z_1, \dots, z_m がそれぞれ独立に標準正規分布にしたがうとき、 z_1, \dots, z_m をまとめた m 次元ベクトル

$$z = \begin{pmatrix} z_1 \\ \vdots \\ z_m \end{pmatrix} \quad (27)$$

のしたがう分布を m 次元標準正規分布、あるいは単に標準正規分布と呼ぶ。

標準正規分布の密度関数

m 次元標準正規分布 $p(\mathbf{z})$ の確率密度関数は、 z_1, \dots, z_m がそれぞれ式 (24) の 1 次元標準正規分布 (平均 0, 分散 1) に独立にしたがうことから、

$$\begin{aligned} p(\mathbf{z}) &= \prod_{i=1}^m p(z_i) = \prod_{i=1}^m \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{z_i^2}{2}\right] \\ &= \frac{1}{\sqrt{(2\pi)^m}} \exp\left[-\frac{1}{2} \sum_{i=1}^m z_i^2\right] = \frac{1}{\sqrt{(2\pi)^m}} \exp\left[-\frac{1}{2} \mathbf{z}^\top \mathbf{z}\right] \end{aligned} \quad (28)$$

となる。 z_1, \dots, z_m のそれぞれの平均が 0 なので、

$$E[\mathbf{z}] = \begin{pmatrix} E[z_1] \\ \vdots \\ E[z_m] \end{pmatrix} = \mathbf{0}. \quad (29)$$

但し、 $\mathbf{0}$ は零ベクトルを表す。

標準正規分布の分散共分散行列

また, z_i, z_j ($i \neq j$) に対して

$$E[z_i z_j] = \iint z_i z_j p(z_i) p(z_j) dz_i dz_j = \int z_i p(z_i) dz_i \int z_j p(z_j) dz_j = 0 \quad (30)$$

が成り立つので, z_i と z_j の共分散は $i \neq j$ のとき 0 となる. z_1, \dots, z_m の分散はそれぞれ 1 なので, m 次元標準正規分布 $p(\mathbf{z})$ にしたがう \mathbf{z} の分散共分散行列は

$$E[(\mathbf{z} - \mathbf{0})(\mathbf{z} - \mathbf{0})^T] = E[\mathbf{z}\mathbf{z}^T] = \mathbf{I}_m \quad (31)$$

となる. 但し, \mathbf{I}_m は m 次の単位行列を表す.

多次元の正規分布

一般の m 次元正規分布は、 n 次元標準正規分布から得られる．すなわち、 m 次元正規分布とは、 n 次元標準正規分布にしたがう確率変数 z を

$$x = \mu + \mathbf{T}z \quad (32)$$

のように変換した m 次元確率変数 x がしたがう確率分布のことを言う．但し、 μ は m 次元の定数ベクトル、 \mathbf{T} は $m \times n$ 行列である．

式 (32) は、 z に \mathbf{T} による線型変換と μ による平行移動を適用すると x が得られることを示している．(線型変換と平行移動を組み合わせた変換をアフィン変換と言う．)

以下、単に正規分布と言った場合には、 m 次元などの多次元正規分布も含むものとする．

多次元の正規分布

x が $x = \mu + \mathbf{T}z$ のように表せるとき x の平均は,

$$\begin{aligned} E[x] &= E[\mu + \mathbf{T}z] = \int (\mu + \mathbf{T}z) p(z) dz = \mu + \mathbf{T} \int zp(z) dz \\ &= \mu + \mathbf{T}E[z] = \mu, \end{aligned} \quad (33)$$

分散共分散行列は,

$$\begin{aligned} E[(x - \mu)(x - \mu)^{\top}] &= E[\mathbf{T}zz^{\top}\mathbf{T}^{\top}] = \int \mathbf{T}zz^{\top}\mathbf{T}^{\top} p(z) dz \\ &= \mathbf{T} \left(\int zz^{\top} p(z) dz \right) \mathbf{T}^{\top} = \mathbf{T}E[zz^{\top}] \mathbf{T}^{\top} = \mathbf{T}\mathbf{T}^{\top} \end{aligned} \quad (34)$$

となる.

多次元の正規分布

一方、任意の m 次元の平均ベクトル μ と m 次の分散共分散行列 P を与えた場合に、その平均、分散共分散行列をもつ正規分布を得ることもできる．そのためには、

$$P = TT^T, \quad (35)$$

を満たす行列 T を求め、標準正規分布を $x = \mu + Tz$ でアフィン変換する．

式 (35) を満たす行列 T は、例えば、固有値分解 $P = U\Lambda U^T$ を利用すると求まる．このとき、対角行列 Λ はすべての対角要素が非負なので、 Λ の各対角要素の平方根を取った対角行列 $\Lambda^{\frac{1}{2}} = \text{diag}(\sqrt{\lambda_1}, \dots, \sqrt{\lambda_m})$ を考えると、

$$T = U\Lambda^{\frac{1}{2}}, \quad (36)$$

で定義される行列 T は式 (35) を満たす．

多次元の正規分布

なお、 $\mathbf{T} = \mathbf{U}\mathbf{\Lambda}^{\frac{1}{2}}$ ならば、

$$\mathbf{P} = \mathbf{T}\mathbf{T}^{\top}, \quad (35)$$

が満たされるが、 \mathbf{W} を任意の m 次の直交行列として、

$$\mathbf{T} = \mathbf{U}\mathbf{\Lambda}^{\frac{1}{2}}\mathbf{W}^{\top} \quad (37)$$

とおいても、式 (35) は成り立つ。

ただ、 \mathbf{T} が異なっても、平均 μ と分散共分散行列 $\mathbf{P} = \mathbf{T}\mathbf{T}^{\top}$ が同じであれば、分布の形状は同じになるので、 \mathbf{T} よりも μ と \mathbf{P} が本質的である。

そのため、通常、正規分布の形状は μ と \mathbf{P} で指定する。以下では、平均 μ 、分散共分散行列 \mathbf{P} の正規分布を $\mathcal{N}(\mu, \mathbf{P})$ と表記する。

正規分布のアフィン変換

今、標準正規分布にしたがう z のアフィン変換から正規分布を得たが、一般の n 次元正規分布 $\mathcal{N}(\mu_a, \mathbf{P}_a)$ にしたがう確率変数 x_a のアフィン変換

$$x_b = b + \mathbf{A}x_a \quad (38)$$

も正規分布 $\mathcal{N}(b + \mathbf{A}\mu_a, \mathbf{A}\mathbf{P}_a\mathbf{A}^\top)$ にしたがう。

このことを示すために、先程と同様にして、 $\mathbf{P}_a = \mathbf{T}_a\mathbf{T}_a^\top$ を満たす行列 \mathbf{T}_a を用意する。 x_a は、 n 次元標準正規分布にしたがう確率変数 z と行列 \mathbf{T}_a を用いて、

$$x_a = \mu_a + \mathbf{T}_a z \quad (39)$$

と表せる。これを式 (38) に代入すると、

$$x_b = b + \mathbf{A}(\mu_a + \mathbf{T}_a z) = b + \mathbf{A}\mu_a + \mathbf{A}\mathbf{T}_a z \quad (40)$$

と書けるので、 x_b は z のアフィン変換であることが分かる。

このとき、 x_b の平均は $b + \mathbf{A}\mu_a$ 、分散共分散行列は $\mathbf{A}\mathbf{T}_a\mathbf{T}_a^\top\mathbf{A}^\top = \mathbf{A}\mathbf{P}_a\mathbf{A}^\top$ となる。したがって、 x_b は正規分布 $\mathcal{N}(b + \mathbf{A}\mu_a, \mathbf{A}\mathbf{P}_a\mathbf{A}^\top)$ にしたがう。

正規分布にしたがう確率変数の和

2つの m 次元の確率変数 x_a, x_b が互いに独立でそれぞれ正規分布 $\mathcal{N}(\mu_a, P_a)$, $\mathcal{N}(\mu_b, P_b)$ にしたがうとき、その和 $x_a + x_b$ は正規分布 $\mathcal{N}(\mu_a + \mu_b, P_a + P_b)$ にしたがう性質がある。

これについても、アフィン変換を利用して確認することができる。まず $P_a = T_a T_a^\top$, $P_b = T_b T_b^\top$ を満たすように行列 T_a, T_b を選ぶ。このとき、確率変数 x_a, x_b は、独立に標準正規分布にしたがう確率変数 z_a, z_b を用いて、

$$x_a = \mu_a + T_a z_a, \quad x_b = \mu_b + T_b z_b \quad (41)$$

のように書ける。これを1つにまとめると、

$$\begin{pmatrix} x_a \\ x_b \end{pmatrix} = \begin{pmatrix} \mu_a \\ \mu_b \end{pmatrix} + \begin{pmatrix} T_a & \mathbf{O} \\ \mathbf{O} & T_b \end{pmatrix} \begin{pmatrix} z_a \\ z_b \end{pmatrix}. \quad (42)$$

正規分布にしたがう確率変数の和

z_a, z_b が独立に m 次元標準正規分布にしたがうので、 $\begin{pmatrix} z_a \\ z_b \end{pmatrix}$ は $2m$ 次元標準正規分布にしたがう。したがって、

$$\begin{pmatrix} x_a \\ x_b \end{pmatrix} = \begin{pmatrix} \mu_a \\ \mu_b \end{pmatrix} + \begin{pmatrix} \mathbf{T}_a & \mathbf{O} \\ \mathbf{O} & \mathbf{T}_b \end{pmatrix} \begin{pmatrix} z_a \\ z_b \end{pmatrix}$$

は、 $2m$ 次元正規分布にしたがい、平均、分散共分散行列はそれぞれ

$$\begin{pmatrix} \mu_a \\ \mu_b \end{pmatrix}, \begin{pmatrix} \mathbf{T}_a & \mathbf{O} \\ \mathbf{O} & \mathbf{T}_b \end{pmatrix} \begin{pmatrix} \mathbf{T}_a & \mathbf{O} \\ \mathbf{O} & \mathbf{T}_b \end{pmatrix}^T = \begin{pmatrix} \mathbf{T}_a \mathbf{T}_a^T & \mathbf{O} \\ \mathbf{O} & \mathbf{T}_b \mathbf{T}_b^T \end{pmatrix} = \begin{pmatrix} \mathbf{P}_a & \mathbf{O} \\ \mathbf{O} & \mathbf{P}_b \end{pmatrix} \quad (43)$$

となる。

正規分布にしたがう確率変数の和

和 $\boldsymbol{x}_a + \boldsymbol{x}_b$ は,

$$\boldsymbol{x}_a + \boldsymbol{x}_b = \begin{pmatrix} \mathbf{I}_m & \mathbf{I}_m \end{pmatrix} \begin{pmatrix} \boldsymbol{x}_a \\ \boldsymbol{x}_b \end{pmatrix} \quad (44)$$

から得られるので、先程の結果から、正規分布にしたがう。また、平均ベクトル $\boldsymbol{\mu}_{\text{sum}}$ ，分散共分散行列 \mathbf{P}_{sum} はそれぞれ

$$\boldsymbol{\mu}_{\text{sum}} = \begin{pmatrix} \mathbf{I}_m & \mathbf{I}_m \end{pmatrix} \begin{pmatrix} \boldsymbol{\mu}_a \\ \boldsymbol{\mu}_b \end{pmatrix} = \boldsymbol{\mu}_a + \boldsymbol{\mu}_b \quad (45)$$

$$\mathbf{P}_{\text{sum}} = \begin{pmatrix} \mathbf{I}_m & \mathbf{I}_m \end{pmatrix} \begin{pmatrix} \mathbf{P}_a & \mathbf{O} \\ \mathbf{O} & \mathbf{P}_b \end{pmatrix} \begin{pmatrix} \mathbf{I}_m & \mathbf{I}_m \end{pmatrix}^T = \mathbf{P}_a + \mathbf{P}_b \quad (46)$$

となる。したがって、 $\boldsymbol{x}_a + \boldsymbol{x}_b$ の従う分布は正規分布 $\mathcal{N}(\boldsymbol{\mu}_a + \boldsymbol{\mu}_b, \mathbf{P}_a + \mathbf{P}_b)$ となることが言える。

多次元正規分布の確率密度関数

行列 \mathbf{T} が正則行列である場合、つまり逆行列をもつ場合には、 x の確率密度関数を得ることができる。 \mathbf{T} が正則のとき、元のアフィン変換の式 $x = \mu + \mathbf{T}z$ から

$$z = \mathbf{T}^{-1}(x - \mu) \quad (47)$$

のように書き直せる。 z および x の次元を m とし、式 (47) を用いて、標準正規分布の確率密度関数を変数変換すると、

$$\begin{aligned} p(x) &= p(z) \left| \frac{dz}{dx^\top} \right| \\ &= \frac{1}{\sqrt{(2\pi)^m}} \exp \left[-\frac{1}{2} z^\top z \right] \left| \frac{dz}{dx^\top} \right| \\ &= \frac{|\mathbf{T}^{-1}|}{\sqrt{(2\pi)^m}} \exp \left[-\frac{1}{2} (x - \mu)^\top \mathbf{T}^{-1\top} \mathbf{T}^{-1} (x - \mu) \right]. \end{aligned}$$

多次元正規分布の確率密度関数

したがって,

$$\begin{aligned} p(\boldsymbol{x}) &= \frac{|\mathbf{T}^{-1}|}{\sqrt{(2\pi)^m}} \exp \left[-\frac{1}{2} (\boldsymbol{x} - \boldsymbol{\mu})^\top \mathbf{T}^{-1} \mathbf{T}^{-1} (\boldsymbol{x} - \boldsymbol{\mu}) \right] \\ &= \frac{1}{\sqrt{(2\pi)^m |\mathbf{T}|}} \exp \left[-\frac{1}{2} (\boldsymbol{x} - \boldsymbol{\mu})^\top (\mathbf{T} \mathbf{T}^\top)^{-1} (\boldsymbol{x} - \boldsymbol{\mu}) \right]. \end{aligned} \quad (48)$$

分散共分散行列を $\mathbf{P} = \mathbf{T} \mathbf{T}^\top$ とおき,

$$|\mathbf{T}| = \sqrt{|\mathbf{T}|^2} = \sqrt{|\mathbf{P}|}$$

が成り立つことを用いると,

$$p(\boldsymbol{x}) = \frac{1}{\sqrt{(2\pi)^m |\mathbf{P}|}} \exp \left[-\frac{1}{2} (\boldsymbol{x} - \boldsymbol{\mu})^\top \mathbf{P}^{-1} (\boldsymbol{x} - \boldsymbol{\mu}) \right]. \quad (49)$$

多次元正規分布の確率密度関数

今、 T が正則であることを仮定して式 (49) を導いたが、分散共分散行列 P が正則行列であれば、 $P = TT^T$ を満たす正則行列 T を得ることは可能である。

例えば、 P が正則行列のとき、固有値分解 $P = U\Lambda U^T$ を行くと、 Λ はすべての対角要素が 0 でない正の値を持つ対角行列なので、逆行列を持つ。したがって、

$$T = U\Lambda^{\frac{1}{2}}, \quad (36)$$

のように T をおくと、 $T^{-1} = (\Lambda^{\frac{1}{2}})^{-1}U^T$ のように T の逆行列が存在し、 T は正則行列となる。

一般に、平均 μ 、分散共分散行列 P の m 次元正規分布は、その P が正則行列であれば、確率密度関数が式 (49) のように書ける。

($* T$ が非正則であっても、 $P = TT^T$ が正則行列になる場合がある。それは、 x の次元 m より z の次元 n が大きく、行列 T の階数 r が m に等しい $m = r < n$ の場合である。ただ、この場合も確率密度関数は同じものになる。)

多次元正規分布の確率密度関数

一方、 P が正則ではない場合、 P が逆行列を持たないので、確率密度関数を式 (49) の形で書くことはできない。

この場合、 x はある部分空間 (アフィン部分空間) 上に分布しており、 x が分布する部分空間上で確率密度を考えることはできるが、 m 次元空間上での確率密度 $p(x)$ は定義できない。

ただ、そのような場合でも、 $p(x)$ を部分空間上の確率密度と読み替えて平均、分散共分散行列を計算することにすれば事足りるので、以下では便宜上、 P が正則ではなくても、確率密度関数 $p(x)$ で x の分布を表すことにする。

正規乱数の生成

正規分布に従う確率変数のアフィン変換は、多変量正規分布に従う乱数の生成にも応用される。

標準正規分布 $\mathcal{N}(0, \mathbf{I})$ に従う乱数はを得るには、一様分布にしたがう 2 つの乱数 $u_1, u_2 \sim U[0, 1)$ から

$$z_1 = \sqrt{-2 \log u_1} \cos(2\pi u_2), \quad (50a)$$

$$z_2 = \sqrt{-2 \log u_1} \sin(2\pi u_2) \quad (50b)$$

を求める．すると、 z_1, z_2 が標準正規分布に従う乱数となる．この方法は Box-Muller 法と呼ばれる． m 次元標準正規分布に従う乱数 z は、1 次元標準正規分布に従う乱数を m 個組み合わせれば得られる．

多変量正規分布 $\mathcal{N}(\mu, \mathbf{P})$ からの乱数を生成したい場合には、 $\mathbf{P} = \mathbf{T}\mathbf{T}^\top$ のように分解して、

$$x = \mu + \mathbf{T}z$$

と変換すればよい．

特異値分解

任意の $m \times n$ 行列 \mathbf{A} (今回, 要素はすべて実数とする) は以下のように分解できる:

$$\mathbf{A} = \mathbf{U} \mathbf{\Lambda} \mathbf{V}^{\top}. \quad (51)$$

ここで \mathbf{U} は $m \times m$ の, \mathbf{V} は $n \times n$ の直交行列であり, また,

$$\mathbf{\Lambda} = \begin{pmatrix} \lambda_1 & & & \mathbf{O} \\ & \ddots & & \\ & & \lambda_r & \\ \mathbf{O} & & & \mathbf{O}_{r \times (n-r)} \\ & \mathbf{O}_{(m-r) \times r} & & \mathbf{O}_{(m-r) \times (n-r)} \end{pmatrix}. \quad (52)$$

$\mathbf{U} = [\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_m]$, $\mathbf{V} = [\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n]$ と縦ベクトルに分解すると, $\{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_m\}$, $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$ は, それぞれ m 次元, n 次元ベクトル空間の正規直交基底をなす.

- $\{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_m\}$, $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$ を特異ベクトルと呼ぶ. また, $\{\lambda_1, \dots, \lambda_r\}$ を特異値と呼ぶ.

特異値分解

A をさらに変形すると以下ようになる。

$$\begin{aligned}
 A &= U \Lambda V^T \\
 &= (\mathbf{u}_1 \quad \cdots \quad \mathbf{u}_m) \begin{pmatrix} \lambda_1 & & \mathbf{O} & \\ & \ddots & & \mathbf{O}_{r \times (n-r)} \\ \mathbf{O} & & \lambda_r & \\ & \mathbf{O}_{(m-r) \times r} & & \mathbf{O}_{(m-r) \times (n-r)} \end{pmatrix} \begin{pmatrix} \mathbf{v}_1^T \\ \vdots \\ \mathbf{v}_n^T \end{pmatrix} \quad (53) \\
 &= \sum_{i=1}^r \lambda_i \mathbf{u}_i \mathbf{v}_i^T.
 \end{aligned}$$

- 通常, $\{\lambda_1, \dots, \lambda_r\}$ はすべて正になるように取る.
- $\{\lambda_1, \dots, \lambda_r\}$ の中に同じ値を持つものが含まれなければ, 特異値を $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_r > 0$ のように順番づけすると, 上記の変換は (正負の符号等の任意性を除いて) 一意に定まる.

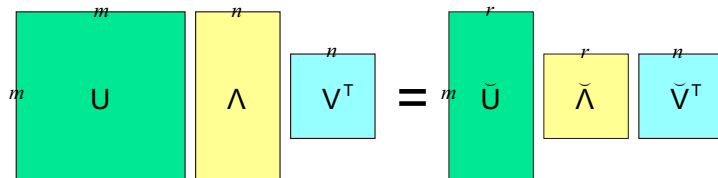
特異値分解

後で便利のように、以下の行列も定義しておく。これは、特異値 λ_i が 0 となる成分を無視しただけで、先程の特異値分解と等価である。

$$\mathbf{A} = \check{\mathbf{U}} \check{\mathbf{\Lambda}} \check{\mathbf{V}}^{\top}. \quad (54)$$

ただし、 $\check{\mathbf{U}} = [\mathbf{u}_1, \dots, \mathbf{u}_r]$, $\check{\mathbf{V}} = [\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_r]$. また,

$$\check{\mathbf{\Lambda}} = \begin{pmatrix} \lambda_1 & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \lambda_r \end{pmatrix}.$$



特異値分解定理の証明

$n \times m$ 行列 A から m 次の正方行列 $A^T A$ を作ると、 $A^T A$ は明らかに対称行列となる。また任意の m 次元ベクトル x に対して

$$x^T A^T A x = \|Ax\|^2 \geq 0$$

が成り立つので、 $A^T A$ は半正定値行列である。したがって、

$$A^T A = V D V^T \tag{55}$$

と固有値分解すると、 V は直交行列となり、対角行列 $D = \text{diag}(d_1, \dots, d_m)$ の各対角要素 d_i はすべて非負の実数となる。

ここで、行列 A の階数を r とおくと、 $A^T A$ の階数も r となるので、 D の 0 でない対角要素の数も r となる。

特異値分解定理の証明

$d_1 \geq d_2 \geq \cdots \geq d_r > 0, d_i = 0 \ (i > r)$ となるように d_1, \dots, d_m をとり, r 番目までの要素から作られる r 次の対角行列

$$\check{\mathbf{D}} = \text{diag}(d_1, \dots, d_r) = \begin{pmatrix} d_1 & & \mathbf{O} \\ & \ddots & \\ \mathbf{O} & & d_r \end{pmatrix} \quad (56)$$

を定義する. また, 行列 \mathbf{V} を $\mathbf{V} = (\mathbf{v}_1 \ \cdots \ \mathbf{v}_m)$ のように分解し, その r 番目までの縦ベクトルまでを取った $m \times r$ 行列

$$\check{\mathbf{V}} = (\mathbf{v}_1 \ \cdots \ \mathbf{v}_r) \quad (57)$$

を定義する. このとき, $\mathbf{A}^\top \mathbf{A}$ の固有値分解は以下のように書き直せる:

$$\mathbf{A}^\top \mathbf{A} = \check{\mathbf{V}} \check{\mathbf{D}} \check{\mathbf{V}}^\top. \quad (58)$$

特異値分解定理の証明

各 v_i は行列 $A^\top A$ の固有値 d_i 固有ベクトルとなるので,

$$A^\top A v_i = d_i v_i \quad (59)$$

を満たす. この両辺に A を掛けると

$$A A^\top A v_i = d_i A v_i \quad (60)$$

となり, ベクトル $A v_i$ が行列 $A A^\top$ の固有ベクトルになっていることが分かる. そこで, $1 \leq i \leq r$ に対して

$$u_i = d_i^{-\frac{1}{2}} A v_i \quad (61)$$

とおき, u_1, \dots, u_r を並べた行列

$$\check{U} = (u_1 \ u_2 \ \cdots \ u_r) = A \check{V} \check{D}^{-\frac{1}{2}} \quad (62)$$

を定義する. 但し, $\check{D}^{-\frac{1}{2}} = \text{diag}(d_1^{-1/2}, \dots, d_r^{-1/2})$ である.

特異値分解定理の証明

$\check{\mathbf{U}} = \mathbf{A}\check{\mathbf{V}}\check{\mathbf{D}}^{-\frac{1}{2}}$ としたとき,

$$\begin{aligned}\check{\mathbf{U}}^T\check{\mathbf{U}} &= \check{\mathbf{D}}^{-\frac{1}{2}}\check{\mathbf{V}}^T\mathbf{A}^T\mathbf{A}\check{\mathbf{V}}\check{\mathbf{D}}^{-\frac{1}{2}} = \check{\mathbf{D}}^{-\frac{1}{2}}\check{\mathbf{V}}^T\check{\mathbf{V}}\check{\mathbf{D}}\check{\mathbf{V}}^T\check{\mathbf{V}}\check{\mathbf{D}}^{-\frac{1}{2}} \\ &= \check{\mathbf{D}}^{-\frac{1}{2}}\check{\mathbf{D}}\check{\mathbf{D}}^{-\frac{1}{2}} = \mathbf{I}_r\end{aligned}\tag{63}$$

が成り立つ。但し, \mathbf{I}_r は r 次の単位行列である。

式 (63) は, $\mathbf{u}_i^T\mathbf{u}_i = 1$, $i \neq j$ のとき $\mathbf{u}_i^T\mathbf{u}_j = 0$ となることを示しているので, 適当な $\mathbf{u}_{r+1}, \dots, \mathbf{u}_n$ を選ぶことにより, $\{\mathbf{u}_1, \dots, \mathbf{u}_n\}$ を n 次元ベクトル空間の正規直交基底にすることができる。

そこで $\{\mathbf{u}_1, \dots, \mathbf{u}_n\}$ をまとめた n 次の正方行列

$$\mathbf{U} = (\mathbf{u}_1 \ \mathbf{u}_2 \ \cdots \ \mathbf{u}_m)$$

を定義する。

特異値分解定理の証明

また, $n \times m$ 行列 Λ を

$$\Lambda = \begin{pmatrix} \sqrt{d_1} & & \mathbf{O} & \\ & \ddots & & \\ \mathbf{O} & & \sqrt{d_r} & \mathbf{O}_{r \times (m-r)} \\ & \mathbf{O}_{(n-r) \times r} & & \mathbf{O}_{(n-r) \times (m-r)} \end{pmatrix}$$

と定義する. u_i の定義などより,

$$\mathbf{A}v_i = \sqrt{d_i}u_i \quad (1 \leq i \leq r), \quad \mathbf{A}v_i = \mathbf{0} \quad (r < i \leq n)$$

が成り立つので, まとめると,

$$\mathbf{A}\mathbf{V} = \mathbf{U}\Lambda. \quad (64)$$

\mathbf{V} は直交行列なので, \mathbf{V}^\top を右から掛けて以下の特異値分解の式を得る:

$$\mathbf{A} = \mathbf{U}\Lambda\mathbf{V}^\top. \quad (65)$$

正方行列の場合

- 行列 A が正則な場合、これを

$$A = U\Lambda V^T$$

のように特異値分解すると、 Λ は対角行列ですべての i に対して $\lambda_i > 0$.
(全対角成分が 0 でない)

- このとき、 A の逆行列は以下のように得られる:

$$A^{-1} = V\Lambda^{-1}U^T. \quad (66)$$

- A が (半) 正定値実対称行列 (固有値が全部非負) なら、 A は

$$A = U\Lambda U^T \quad (67)$$

のように固有値分解でき、その結果は特異値分解と同じ.

一般の行列の場合

- 一般の行列 A についても特異値分解

$$A = \check{U} \check{\Lambda} \check{V}^T$$

を利用して,

$$A^- = \check{V} \check{\Lambda}^{-1} \check{U}^T \quad (68)$$

のような、逆行列みたいなものが得られる.

- これは、ムーア・ペンローズの一般逆行列と呼ばれるもの.
- $y = Ax$ という方程式があり、 $\dim y > \dim x$ のとき、この一般逆行列 A^- を使うと、 y と Ax との差のユークリッドノルムを最小とする x が

$$x_{\text{est}} = A^- y \quad (69)$$

の形で得られる (後で少し触れる).

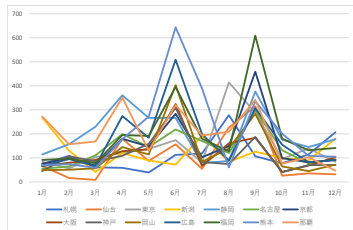
特異値分解等の計算について

- 自分でプログラムを書くよりは、LAPACK というライブラリを使うのが一般的と思われる。
- 大学・研究所等のスパコンで提供されている数値計算ライブラリの中に LAPACK も含まれていることが多い。
- Python を使うなら、numpy でできる。
- R にも関数が用意されている。

特異値分解の応用例

以下のデータに基づく行列を特異値分解する。

	Jan.	...	Dec.
札幌の降水量	$x_{1,1}$...	$x_{12,1}$
仙台の降水量	$x_{1,2}$...	$x_{12,2}$
東京の降水量	$x_{1,3}$...	$x_{12,3}$
新潟の降水量	$x_{1,4}$...	$x_{12,4}$
静岡の降水量	$x_{1,5}$...	$x_{12,5}$
名古屋の降水量	$x_{1,6}$...	$x_{12,6}$
京都の降水量	$x_{1,7}$...	$x_{12,7}$
大阪の降水量	$x_{1,8}$...	$x_{12,8}$
神戸の降水量	$x_{1,9}$...	$x_{12,9}$
岡山の降水量	$x_{1,10}$...	$x_{12,10}$
広島	$x_{1,11}$...	$x_{12,11}$
福岡の降水量	$x_{1,12}$...	$x_{12,12}$
熊本の降水量	$x_{1,13}$...	$x_{12,13}$
那覇の降水量	$x_{1,14}$...	$x_{12,14}$



$$\mathbf{x}_i = \begin{pmatrix} x_{i,1} \\ \vdots \\ x_{i,m} \end{pmatrix} \quad (i = 1, \dots, n),$$

$$\boldsymbol{\mu} = \sum_i \mathbf{x}_i / n \text{ として,}$$

$$\mathbf{X} = \begin{pmatrix} \mathbf{x}_1 - \boldsymbol{\mu} & \cdots & \mathbf{x}_n - \boldsymbol{\mu} \end{pmatrix}.$$

データ行列の例 (2016 年の各月の降水量)

特異値分解の応用例

データ行列 \mathbf{X} の特異値分解

$$\mathbf{X} = \check{\mathbf{U}} \check{\mathbf{\Lambda}} \check{\mathbf{V}}^T$$

を考える。 $\check{\mathbf{\Lambda}} \check{\mathbf{V}}^T$ を成分で書くと、

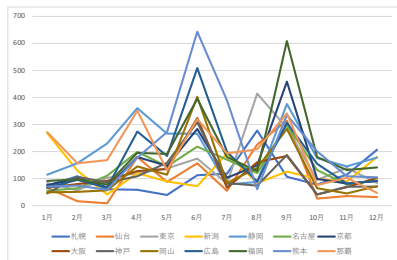
$$\check{\mathbf{\Lambda}} \check{\mathbf{V}}^T = \begin{pmatrix} \lambda_1 v_{1,1} & \cdots & \lambda_1 v_{1,n} \\ \vdots & \ddots & \vdots \\ \lambda_m v_{m,1} & \cdots & \lambda_m v_{m,n} \end{pmatrix}. \quad (70)$$

これを用いると、

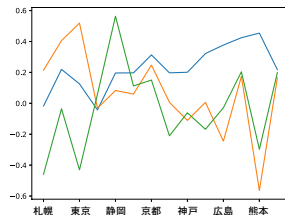
$$(\mathbf{x}_1 \quad \cdots \quad \mathbf{x}_n) = \left(\sum_{j=1}^r \lambda_j v_{j,1} \mathbf{u}_j \quad \cdots \quad \sum_{j=1}^r \lambda_j v_{j,n} \mathbf{u}_j \right). \quad (71)$$

つまり、特異値分解により、各 \mathbf{x}_i が共通の基底 $\{\mathbf{u}_j\}$ による直交基底展開が得られる。

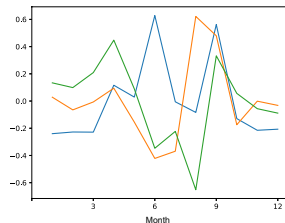
特異値分解の応用例



各月の降水量 (2016)



主成分ベクトル u_1, u_2, u_3



主成分の変動 v_1, v_2, v_3

特異値分解の応用例

$$(\mathbf{x}_1 \quad \cdots \quad \mathbf{x}_n) = \left(\sum_{j=1}^r \lambda_j v_{j,1} \mathbf{u}_j \quad \cdots \quad \sum_{j=1}^r \lambda_j v_{j,n} \mathbf{u}_j \right)$$

において、 $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_r \geq 0$ なので、ある閾値 ε を与えたとき、 $\lambda_L > \varepsilon \geq \lambda_{L+1} (\geq \lambda_r)$ なる L をとることができる。このとき

$$(\tilde{\mathbf{x}}_1 \quad \cdots \quad \tilde{\mathbf{x}}_n) = \left(\sum_{j=1}^L \lambda_j v_{j,1} \mathbf{u}_j \quad \cdots \quad \sum_{j=1}^L \lambda_j v_{j,n} \mathbf{u}_j \right) \quad (72)$$

のように $L+1$ 番目以降の成分を落とすと、 $\tilde{\mathbf{x}}_1, \dots, \tilde{\mathbf{x}}_n$ は L 個の基底ベクトルによる $\mathbf{x}_1, \dots, \mathbf{x}_n$ の近似となる。実際、

$$\|\tilde{\mathbf{x}}_i - \mathbf{x}_i\|^2 = \left\| \sum_{j=L+1}^r \lambda_j v_{j,i} \mathbf{u}_j \right\|^2 = \sum_{j=L+1}^r (\lambda_j v_{j,i})^2 < \sum_{j=L+1}^r \varepsilon^2 v_{j,i}^2 < \varepsilon^2 \sum_{j=1}^n v_{j,i}^2 = \varepsilon^2$$

が言える。(最後の等号は V が直交行列であることから言える。)

この操作は主成分分析と呼ばれる手法に対応している。

主成分分析について

主成分分析は、多次元データ $x^{(1)}, \dots, x^{(n)}$ が与えられたとき、長さ 1 のベクトル a ($\|a\|^2 = 1$) を、各 $x^{(i)}$ との (平均 μ を引いた上での) 内積

$$z^{(i)} = a^T (x^{(i)} - \mu) \quad (i = 1, \dots, n) \quad (73)$$

で得られる $\{z^{(i)}\}$ の分散が最大になるように選ぶというのが本来の定義である。

$$d = \begin{pmatrix} z^{(1)} \\ \vdots \\ z^{(n)} \end{pmatrix} \quad (74)$$

とくと、その $\{z^{(i)}\}$ の分散は以下のように書ける。

$$P_z = \frac{1}{n-1} d^T d. \quad (75)$$

これを最大化する a を探せばよい。

主成分分析について

平均を引いたデータ行列

$$\mathbf{X} = \begin{pmatrix} \mathbf{x}^{(1)} - \boldsymbol{\mu} & \cdots & \mathbf{x}^{(n)} - \boldsymbol{\mu} \end{pmatrix} \quad (76)$$

を考えると, $\mathbf{d}^\top = \mathbf{a}^\top \mathbf{X}$ となる.

\mathbf{X} の特異値分解

$$\mathbf{X} = \check{\mathbf{U}} \check{\boldsymbol{\Lambda}} \check{\mathbf{V}}^\top$$

を考えると,

$$\mathbf{d}^\top = \mathbf{a}^\top \mathbf{X} = \mathbf{a}^\top \check{\mathbf{U}} \check{\boldsymbol{\Lambda}} \check{\mathbf{V}}^\top$$

なので,

$$P_z = \frac{1}{n-1} \mathbf{d}^\top \mathbf{d} = \frac{1}{n-1} \mathbf{a}^\top \check{\mathbf{U}} \check{\boldsymbol{\Lambda}}^2 \check{\mathbf{U}}^\top \mathbf{a}. \quad (77)$$

主成分分析について

$$\boldsymbol{a} = \mathbf{U}\boldsymbol{b} = \sum_{j=1}^m b_j \boldsymbol{u}_j$$

とおくと、 $\|\boldsymbol{a}\|^2 = 1$ なので $\|\boldsymbol{b}\|^2 = 1$.

これを用いると、

$$P_z = \frac{1}{n-1} \boldsymbol{a}^\top \check{\mathbf{U}} \check{\mathbf{\Lambda}}^2 \check{\mathbf{U}}^\top \boldsymbol{a} = \frac{1}{n-1} \boldsymbol{b}^\top \mathbf{\Lambda}^2 \boldsymbol{b} = \sum_{j=1}^r b_j^2 \lambda_j^2. \quad (78)$$

$\lambda_1 \geq \cdots \geq \lambda_r > 0$ とすると、 $\|\boldsymbol{b}\|^2 = 1$ の下で P_z を最大化するには

$$b_1 = 1, \quad b_j = 0 \quad (j \neq 1) \quad (79)$$

とすればよいことが分かる。したがって、 $\boldsymbol{a} = \boldsymbol{u}_1$ のとき、 P_z が最大となる。

主成分分析について

以上の議論から、

$$z^{(i)} = \mathbf{a}^\top (\mathbf{x}^{(i)} - \boldsymbol{\mu}) \quad (i = 1, \dots, n)$$

の形で与えられる $\{z^{(i)}\}$ の分散 P_z を最大にするには、 $\mathbf{a} = \mathbf{u}_1$ とすればよい。

これは、 $\{\mathbf{x}^{(i)}\}$ が \mathbf{u}_1 の方向に大きくばらついていることを意味する。この \mathbf{u}_1 を第 1 主成分と呼ぶ。

\mathbf{u}_1 の次にばらつきの大きい方向を調べるのは、先程の議論を繰り返せばよい。すなわち、

$$\mathbf{a} = \mathbf{U}\mathbf{b} = \sum_{j=1}^m b_j \mathbf{u}_j, \quad P_z = \sum_{j=1}^r b_j^2 \lambda_j^2$$

と、 $\lambda_1 \geq \dots \geq \lambda_r > 0$ から、 \mathbf{u}_1 に直交で $\|\mathbf{a}\|^2 = 1$ という条件の下で P_z を最大化するベクトル \mathbf{a} は、 $\mathbf{a} = \mathbf{u}_2$ となる。 \mathbf{u}_2 を第 2 主成分と呼ぶ。

その次にばらつきの大きい方向 (第 3 主成分, 第 4 主成分, ...) も、同様にして順次得られる。

主成分分析について

■ 今,

$$\mathbf{X} = \begin{pmatrix} \mathbf{x}^{(1)} - \boldsymbol{\mu} & \cdots & \mathbf{x}^{(n)} - \boldsymbol{\mu} \end{pmatrix}$$

の特異値分解

$$\mathbf{X} = \check{\mathbf{U}} \check{\boldsymbol{\Lambda}} \check{\mathbf{V}}^{\top} = (\mathbf{u}_1 \quad \cdots \quad \mathbf{u}_r) \begin{pmatrix} \lambda_1 & & \mathbf{O} \\ & \ddots & \\ \mathbf{O} & & \lambda_r \end{pmatrix} \begin{pmatrix} \mathbf{v}_1^{\top} \\ \vdots \\ \mathbf{v}_r^{\top} \end{pmatrix}$$

から, 主成分分析における各主成分が得られた.

■ なお, 通常の主成分分析では, $\{\mathbf{x}^{(i)}\}$ の分散共分散行列

$$\mathbf{P} = \frac{1}{n-1} \sum_{i=1}^n \left(\mathbf{x}^{(i)} - \boldsymbol{\mu} \right) \left(\mathbf{x}^{(i)} - \boldsymbol{\mu} \right)^{\top} = \frac{1}{n-1} \mathbf{X} \mathbf{X}^{\top} \quad (80)$$

の固有値分解を行う.

主成分分析について

- 実際には、 \mathbf{X} の特異値分解

$$\mathbf{X} = \mathbf{U}\mathbf{\Lambda}\mathbf{V}^T$$

を用いると、分散共分散行列は、

$$\mathbf{P} = \frac{1}{n-1} \mathbf{X}\mathbf{X}^T = \frac{1}{n-1} \mathbf{U}\mathbf{\Lambda}\mathbf{V}^T\mathbf{V}\mathbf{\Lambda}\mathbf{U}^T = \frac{1}{n-1} \mathbf{U}\mathbf{\Lambda}^2\mathbf{U}^T \quad (81)$$

と変形できる．これは \mathbf{P} の固有値分解となっており， \mathbf{U} を構成する縦ベクトル u_1, \dots, u_m は， \mathbf{P} の固有ベクトルとなる．したがって，特異値分解を使っても，固有値分解を使っても同じ u_1, u_2, \dots が得られるので，どちらでも問題ない．

まとめ

- 数値シミュレーションには、初期条件、境界条件など色々と不確かな要素が含まれている。
- データ同化は、実際の観測データとシミュレーションを組み合わせ、初期条件、境界条件などを推定し、観測をうまく説明する尤もらしいシナリオを提示する。
- ここでは、特異値分解や多変量正規分布など、データ同化のみならず、多変量データ解析全般で用いられる基礎的な事項について概観した。