

1 導入と数学的準備

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。
- 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

- \mathbf{a} : ベクトル (確率変数)
- \mathbf{A} : 行列
- $\mathbf{a}^T, \mathbf{A}^T$: ベクトル, 行列の転置
- \mathbf{x} : 状態変数
- \mathbf{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
 - $\mathbf{y}_1, \dots, \mathbf{y}_k$ をまとめて, $\mathbf{y}_{1:k}$ などと表記することもある.
 - $p(\mathbf{x}_k | \mathbf{y}_{1:k-1}) = p(\mathbf{x}_k | \mathbf{y}_1, \dots, \mathbf{y}_{k-1})$ など.
- $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$: 平均 $\boldsymbol{\mu}$, 分散共分散行列 $\boldsymbol{\Sigma}$ のガウス分布 (正規分布)

定式化

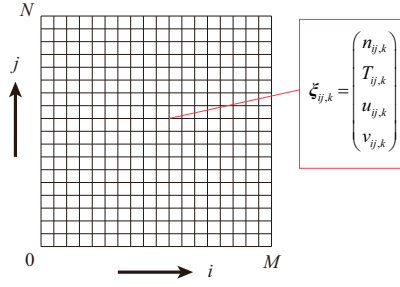
まず, シミュレーションの中で扱っている全ての変数の時刻 t_k における値を一つのベクトルにまとめて \mathbf{x}_k とおく. 各グリッドに種々の変数が割り当てられている場合, ある時刻 t_k , あるグリッド (i, j) での全変数の値をまとめたベクトルを ξ_{ijk} とすると,

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

時刻 t_{k-1} の状態 \mathbf{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 \mathbf{x} の時間発展を

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

のように \mathbf{f}_k で表現する. ここで, $\mathbf{x}_k = \mathbf{f}_k(\mathbf{x}_{k-1})$ 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. \mathbf{v}_k をシステムノイズと呼ぶ.



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

$$\mathbf{y}_k = \mathbf{H}_k \mathbf{x}_k + \mathbf{w}_k. \quad (2)$$

観測モデル $\mathbf{y}_k = \mathbf{H}_k \mathbf{x}_k + \mathbf{w}_k$ の行列 \mathbf{H}_k は、例えば、シミュレーション変数 \mathbf{x}_k の中の x_1, x_2, x_3 のみが直接観測できた場合、

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

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。

シミュレーション変数と観測データとの関係を

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

と書いたが、 $\mathbf{y}_k = \mathbf{H}_k \mathbf{x}_k$ 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、 \mathbf{w}_k を観測ノイズと呼ぶ。

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。

$$x_k \approx f_k(x_{k-1})$$

の $f_k(x_{k-1})$ は、近似式から得られた x_k の予測であることを考慮し、(未知の) 予測誤差 v_k を導入して、

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

観測については、 x_k の各成分がノイズ w_k つきで観測されると仮定すると、

$$y_k = x_k + w_k.$$

これは H_k を単位行列 I とすれば、

$$y_k = H_k x_k + w_k.$$

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。

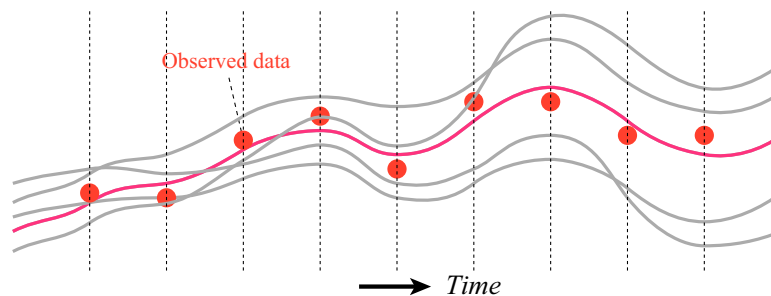
主なデータ同化手法

- 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

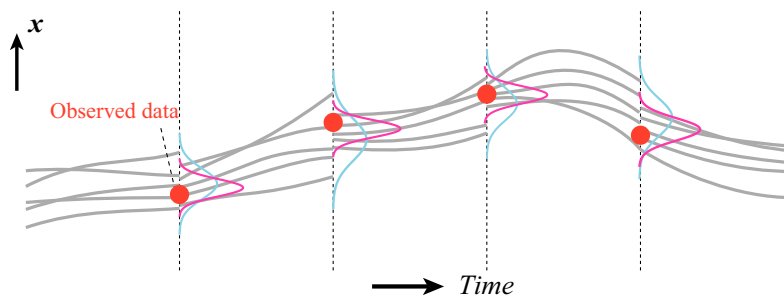
4次元変分法

- 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、 $\mathbf{x}_k = \mathbf{f}_k(\mathbf{x}_{k-1})$ という拘束条件の下で $\sum_{i=1}^N \|\mathbf{y}_k - \mathbf{H}_k \mathbf{x}_k\|^2$ を最小化する \mathbf{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。



確率分布についての基本的事項

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 \mathbf{x} の平均 $\boldsymbol{\mu}$ 、分散共分散行列 \mathbf{P} は以下のように定義される:

$$\boldsymbol{\mu} = E[\mathbf{x}], \quad (7)$$

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

但し、 E は期待値を表し、上付きの T は転置行列を表す。 \mathbf{x} の確率密度関数 $p(\mathbf{x})$ が与えられているとき、 $\boldsymbol{\mu}$ 、 \mathbf{P} はそれぞれ、

$$\boldsymbol{\mu} = \int \mathbf{x} p(\mathbf{x}) d\mathbf{x}, \quad (9)$$

$$\mathbf{P} = \int (\mathbf{x} - \boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})^T p(\mathbf{x}) d\mathbf{x} \quad (10)$$

のように得られる。但し、積分領域は \mathbf{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。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 \mathbf{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)、 \mathbf{P} は正則行列となる(逆行列 \mathbf{P}^{-1} を持つ)。このとき、 \mathbf{P} を

$$\mathbf{P} = \mathbf{U} \boldsymbol{\Lambda} \mathbf{U}^T \quad (12)$$

のように固有値分解すると、 \mathbf{P}^{-1} は

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

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

同時分布と周辺分布

2つの確率ベクトル $\mathbf{x}_a, \mathbf{x}_b$ を考えたとき, $\mathbf{x}_a, \mathbf{x}_b$ をまとめた

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

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) $p(\mathbf{x}_a, \mathbf{x}_b)$ が与えられているとき,

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

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

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 $p(\mathbf{x}_a), p(\mathbf{x}_b)$ を, それぞれ $\mathbf{x}_a, \mathbf{x}_b$ 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

$\mathbf{x}_a, \mathbf{x}_b$ のうち, \mathbf{x}_a の値が与えられたときの \mathbf{x}_b の分布を, \mathbf{x}_a が与えられたもとの \mathbf{x}_b の条件付き分布と呼び, その確率密度関数は $p(\mathbf{x}_b|\mathbf{x}_a)$ 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,

$$p(\mathbf{x}_a, \mathbf{x}_b) = p(\mathbf{x}_b|\mathbf{x}_a)p(\mathbf{x}_a) = p(\mathbf{x}_a|\mathbf{x}_b)p(\mathbf{x}_b) \quad (17)$$

が成り立つ。式 (16) の周辺化の操作は, 式 (17) を適用した

$$\int p(\mathbf{x}_b|\mathbf{x}_a)p(\mathbf{x}_a) d\mathbf{x}_a = p(\mathbf{x}_b), \quad \int p(\mathbf{x}_a|\mathbf{x}_b)p(\mathbf{x}_b) d\mathbf{x}_b = p(\mathbf{x}_a) \quad (18)$$

の形で書かれることも多い。

同時確率密度関数 $p(\mathbf{x}_a, \mathbf{x}_b)$ が任意の $\mathbf{x}_a, \mathbf{x}_b$ に対して,

$$p(\mathbf{x}_a, \mathbf{x}_b) = p(\mathbf{x}_a)p(\mathbf{x}_b) \quad (19)$$

を満たすとき, \mathbf{x}_a と \mathbf{x}_b が独立であると言う。 \mathbf{x}_a と \mathbf{x}_b が独立のとき, 式 (17) より,

$$p(\mathbf{x}_b|\mathbf{x}_a) = p(\mathbf{x}_b), \quad p(\mathbf{x}_a|\mathbf{x}_b) = p(\mathbf{x}_a) \quad (20)$$

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 \mathbf{g} を用いて, m 次元の確率変数 \mathbf{x} と m 次元確率変数 \mathbf{u} が $\mathbf{x} = \mathbf{g}(\mathbf{u})$ のように対応付けられているとする。 \mathbf{x} の確率密度関数 $p(\mathbf{x})$ が

$$p(\mathbf{x}) = f(\mathbf{x}) \quad (21)$$

の形で与えられたとき、 \mathbf{u} の確率密度関数は、

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

のように書くことができる。但し、 $\left| \frac{\partial \mathbf{g}}{\partial \mathbf{u}^\top} \right|$ は関数 \mathbf{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、 \mathbf{u} が定義される任意の点で $\left| \frac{\partial \mathbf{g}}{\partial \mathbf{u}^\top} \right| \neq 0$ が成り立つ必要がある。

$p(\mathbf{x})$ から $p(\mathbf{u})$ を得るときにヤコビ行列式を掛けるのは、

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

のように、 \mathbf{x} が領域 A に含まれる確率を領域 A の \mathbf{g} による像 $\mathbf{g}(A)$ に \mathbf{u} が含まれる確率と対応づけるためである。式 (22) のように \mathbf{x} の確率密度関数 $p(\mathbf{x})$ を \mathbf{u} の確率密度関数 $p(\mathbf{u})$ に変換する操作は、 \mathbf{x} から \mathbf{u} への変数変換と呼ばれる。

正規分布

1 次元正規分布

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。) 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

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

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。

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。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 \mathbf{z} を

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

のように変換した m 次元確率変数 \mathbf{x} がしたがう確率分布のことを言う。但し, $\boldsymbol{\mu}$ 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, \mathbf{z} に \mathbf{T} による線型変換と $\boldsymbol{\mu}$ による平行移動を適用すると \mathbf{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。

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

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

分散共分散行列は,

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

となる。一方, 任意の m 次元の平均ベクトル $\boldsymbol{\mu}$ と m 次の分散共分散行列 \mathbf{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,

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

を満たす行列 \mathbf{T} を求め, 式 (32) にしたがって, 標準正規分布をアフィン変換する。

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

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

で定義される行列 \mathbf{T} は式 (35) を満たす。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。

式 (32) では、標準正規分布にしたがう \mathbf{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 \mathbf{x}_a のアフィン変換

$$\mathbf{x}_b = \mathbf{b} + \mathbf{A}\mathbf{x}_a \quad (38)$$

も正規分布 $\mathcal{N}(\mathbf{b} + \mathbf{A}\mu_a, \mathbf{A}\mathbf{P}_a\mathbf{A}^\top)$ 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。 \mathbf{x}_a は、 n 次元標準正規分布にしたがう確率変数 \mathbf{z} と行列 \mathbf{T}_a を用いて、

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

と表せる。これを式 (38) に代入すると、

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

と書けるので、 \mathbf{x}_b は \mathbf{z} のアフィン変換であることが分かる。このとき、 \mathbf{x}_b の平均は $\mathbf{b} + \mathbf{A}\mu_a$ 、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、 \mathbf{x}_b は正規分布 $\mathcal{N}(\mathbf{b} + \mathbf{A}\mu_a, \mathbf{A}\mathbf{P}_a\mathbf{A}^\top)$ 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 $\mathbf{x}_a, \mathbf{x}_b$ が互いに独立でそれぞれ正規分布 $\mathcal{N}(\mu_a, \mathbf{P}_a), \mathcal{N}(\mu_b, \mathbf{P}_b)$ にしたがうとき、その和 $\mathbf{x}_a + \mathbf{x}_b$ は正規分布 $\mathcal{N}(\mu_a + \mu_b, \mathbf{P}_a + \mathbf{P}_b)$ 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 $\mathbf{P}_a = \mathbf{T}_a\mathbf{T}_a^\top$ 、 $\mathbf{P}_b = \mathbf{T}_b\mathbf{T}_b^\top$ を満たすように行列 $\mathbf{T}_a, \mathbf{T}_b$ を選ぶ。このとき、確率変数 $\mathbf{x}_a, \mathbf{x}_b$ は、独立に標準正規分布にしたがう確率変数 $\mathbf{z}_a, \mathbf{z}_b$ を用いて、

$$\mathbf{x}_a = \mu_a + \mathbf{T}_a\mathbf{z}_a, \quad \mathbf{x}_b = \mu_b + \mathbf{T}_b\mathbf{z}_b \quad (41)$$

のように書ける．これを1つにまとめると，

$$\begin{pmatrix} \mathbf{x}_a \\ \mathbf{x}_b \end{pmatrix} = \begin{pmatrix} \boldsymbol{\mu}_a \\ \boldsymbol{\mu}_b \end{pmatrix} + \begin{pmatrix} \mathbf{T}_a & \mathbf{O} \\ \mathbf{O} & \mathbf{T}_b \end{pmatrix} \begin{pmatrix} \mathbf{z}_a \\ \mathbf{z}_b \end{pmatrix}. \quad (42)$$

$\mathbf{z}_a, \mathbf{z}_b$ が独立に m 次元標準正規分布にしたがうので， $\begin{pmatrix} \mathbf{z}_a \\ \mathbf{z}_b \end{pmatrix}$ は $2m$ 次元標準正規分布にしたがう．したがって，

$$\begin{pmatrix} \mathbf{x}_a \\ \mathbf{x}_b \end{pmatrix} = \begin{pmatrix} \boldsymbol{\mu}_a \\ \boldsymbol{\mu}_b \end{pmatrix} + \begin{pmatrix} \mathbf{T}_a & \mathbf{O} \\ \mathbf{O} & \mathbf{T}_b \end{pmatrix} \begin{pmatrix} \mathbf{z}_a \\ \mathbf{z}_b \end{pmatrix}$$

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

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

となる．和 $\mathbf{x}_a + \mathbf{x}_b$ は，

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

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

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

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

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

多次元正規分布の確率密度関数

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， \mathbf{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 $\mathbf{x} = \boldsymbol{\mu} + \mathbf{T}\mathbf{z}$ (32) から

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

のように書き直せる． \mathbf{z} および \mathbf{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，

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

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,

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

今, \mathbf{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 \mathbf{P} が正則行列であれば, $\mathbf{P} = \mathbf{T}\mathbf{T}^\top$ を満たす正則行列 \mathbf{T} を得ることは可能である. 例えば, \mathbf{P} が正則行列のとき, 式 (12) の固有値分解 $\mathbf{P} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^\top$ を行くと, $\boldsymbol{\Lambda}$ 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,

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

のように \mathbf{T} をおくと, $\mathbf{T}^{-1} = (\boldsymbol{\Lambda}^{\frac{1}{2}})^{-1}\mathbf{U}^\top$ のように \mathbf{T} の逆行列が存在し, \mathbf{T} は正則行列となる. 一般に, 平均 $\boldsymbol{\mu}$, 分散共分散行列 \mathbf{P} の m 次元正規分布は, その \mathbf{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.

(* \mathbf{T} が非正則であっても, $\mathbf{P} = \mathbf{T}\mathbf{T}^\top$ が正則行列になる場合がある. それは, \mathbf{x} の次元 m より \mathbf{z} の次元 n が大きく, 行列 \mathbf{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.)

一方, \mathbf{P} が正則ではない場合, \mathbf{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, \mathbf{x} はある部分空間 (アフィン部分空間) 上に分布しており, \mathbf{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 $p(\mathbf{x})$ は定義できない. ただ, そのような場合でも, $p(\mathbf{x})$ 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, \mathbf{P} が正則でなくても, 確率密度関数 $p(\mathbf{x})$ で \mathbf{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 $\mathcal{N}(\mathbf{0}, \mathbf{I})$ 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 \mathbf{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 $\mathcal{N}(\boldsymbol{\mu}, \mathbf{P})$ 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,

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

と変換すればよい.

特異値分解と確率変数の変換

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 $n \times n$ の正規直交行列であり、また、

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

$\mathbf{U} = [\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_m]$, $\mathbf{V} = [\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n]$ 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。

\mathbf{A} をさらに変形すると以下ようになる。

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

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 $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_r > 0$ 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。

後で便利なように、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、

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

特異値分解定理の証明

$n \times m$ 行列 \mathbf{A} から m 次の正方行列 $\mathbf{A}^T \mathbf{A}$ を作ると、 $\mathbf{A}^T \mathbf{A}$ は明らかに対称行列となる．また任意の m 次元ベクトル \mathbf{x} に対して $\mathbf{x}^T \mathbf{A}^T \mathbf{A} \mathbf{x} = \|\mathbf{A} \mathbf{x}\|^2 \geq 0$ が成り立つので、 $\mathbf{A}^T \mathbf{A}$ は半正定値行列である．したがって、

$$\mathbf{A}^T \mathbf{A} = \mathbf{V} \mathbf{D} \mathbf{V}^T \quad (55)$$

と固有値分解すると、 \mathbf{V} は直交行列となり、対角行列 $\mathbf{D} = \text{diag}(d_1, \dots, d_m)$ 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 \mathbf{A} の階数を r とおくと、 $\mathbf{A}^T \mathbf{A}$ の階数も r となるので、 \mathbf{D} の 0 でない対角要素の数も r となる． $d_1 \geq d_2 \geq \dots \geq d_r > 0, d_i = 0 (i > r)$ 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

$$\check{\mathbf{D}} = \text{diag}(d_1, \dots, d_r) = \begin{pmatrix} d_1 & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & d_r \end{pmatrix} \quad (56)$$

を定義する．また、行列 \mathbf{V} を $\mathbf{V} = (\mathbf{v}_1 \dots \mathbf{v}_m)$ 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

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

を定義する．このとき、式 (55) は以下のように書き直せる：

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

各 \mathbf{v}_i は行列 $\mathbf{A}^T \mathbf{A}$ の固有値 d_i 固有ベクトルとなるので、

$$\mathbf{A}^T \mathbf{A} \mathbf{v}_i = d_i \mathbf{v}_i \quad (59)$$

を満たす．この両辺に \mathbf{A} を掛けると

$$\mathbf{A} \mathbf{A}^T \mathbf{A} \mathbf{v}_i = d_i \mathbf{A} \mathbf{v}_i \quad (60)$$

となり、ベクトル $\mathbf{A} \mathbf{v}_i$ が行列 $\mathbf{A} \mathbf{A}^T$ 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

$$\mathbf{u}_i = d_i^{-\frac{1}{2}} \mathbf{A} \mathbf{v}_i \quad (61)$$

とおき、 $\mathbf{u}_1, \dots, \mathbf{u}_r$ を並べた行列

$$\check{\mathbf{U}} = (\mathbf{u}_1 \mathbf{u}_2 \dots \mathbf{u}_r) = \mathbf{A} \check{\mathbf{V}} \check{\mathbf{D}}^{-\frac{1}{2}} \quad (62)$$

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

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

を定義する。また、 $n \times m$ 行列 $\mathbf{\Lambda}$ を

$$\mathbf{\Lambda} = \begin{pmatrix} \sqrt{d_1} & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \sqrt{d_r} \\ & \mathbf{0}_{(n-r) \times r} & \mathbf{0}_{(n-r) \times (m-r)} \end{pmatrix}$$

と定義する。式 (61) などより、

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

が成り立つので、まとめると、

$$\mathbf{A}\mathbf{V} = \mathbf{U}\mathbf{\Lambda}.\tag{64}$$

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

$$\mathbf{A} = \mathbf{U}\mathbf{\Lambda}\mathbf{V}^T.\tag{65}$$

一般逆行列

行列 \mathbf{A} が正則な場合、これを

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

のように特異値分解すると、 $\mathbf{\Lambda}$ 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、 \mathbf{A} の逆行列は以下のように得られる：

$$\mathbf{A}^{-1} = \mathbf{V}\mathbf{\Lambda}^{-1}\mathbf{U}^T.\tag{66}$$

\mathbf{A} が (半) 正定値実対称行列 (固有値が全部非負) なら、 \mathbf{A} は

$$\mathbf{A} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^T.\tag{67}$$

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 \mathbf{A} についても特異値分解 $\mathbf{A} = \check{\mathbf{U}}\check{\mathbf{\Lambda}}\check{\mathbf{V}}^T$ を利用して,

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

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。 $\mathbf{y} = \mathbf{Ax}$ という方程式があり、 $\dim \mathbf{y} > \dim \mathbf{x}$ のとき、この一般逆行列 \mathbf{A}^- を使うと、 \mathbf{y} と \mathbf{Ax} との差のユークリッドノルムを最小とする \mathbf{x} が

$$\mathbf{x}_{\text{est}} = \mathbf{A}^- \mathbf{y} \quad (69)$$

の形で得られる。

- 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。

主成分分析への応用

応用例として、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 of 降水量	$x_{1,11}$...	$x_{12,11}$
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	$x_{1,14}$...	$x_{12,14}$

データ行列の例 (2016 年の各月の降水量)

$$\mathbf{x}_i = \begin{pmatrix} x_{i,1} \\ \vdots \\ x_{i,m} \end{pmatrix} (i = 1, \dots, n), \quad \boldsymbol{\mu} = \sum_i \mathbf{x}_i / n \quad \text{として},$$

$$\mathbf{X} = (\mathbf{x}_1 - \boldsymbol{\mu} \quad \cdots \quad \mathbf{x}_n - \boldsymbol{\mu}).$$

とデータ行列を定義し、その特異値分解 $\mathbf{X} = \check{\mathbf{U}}\check{\mathbf{\Lambda}}\check{\mathbf{V}}^\top$ を考える。 $\check{\mathbf{\Lambda}}\check{\mathbf{V}}^\top$ を成分で書くと、

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

これを用いると、

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

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。

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

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

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

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、

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

が言える。(最後の等号は \mathbf{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。

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

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

で得られる $\{z^{(i)}\}$ の分散が最大になるように選ぶというものである。

$$\mathbf{d} = \begin{pmatrix} z^{(1)} \\ \vdots \\ z^{(N)} \end{pmatrix}$$

とおくと、その $\{z^{(i)}\}$ の分散は以下のように書ける。

$$P_z = \frac{1}{n-1} \mathbf{d}^\top \mathbf{d}. \quad (74)$$

これを最大化する \mathbf{a} を探せばよい。平均を引いたデータ行列

$$\mathbf{X} = (\mathbf{x}^{(1)} - \boldsymbol{\mu} \quad \cdots \quad \mathbf{x}^{(N)} - \boldsymbol{\mu}) \quad (75)$$

を考えると、 $\mathbf{d}^\top = \mathbf{a}^\top \mathbf{X}$ となる。 \mathbf{X} の特異値分解

$$\mathbf{X} = \check{\mathbf{U}} \check{\boldsymbol{\Lambda}} \check{\mathbf{V}}^\top \quad (76)$$

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、

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

ここで、 $\mathbf{a} = \mathbf{U} \mathbf{b} = \sum_{j=1}^m b_j \mathbf{u}_j$ とおくと、 $\|\mathbf{a}\|^2 = 1$ なので $\|\mathbf{b}\|^2 = 1$ 。これを用いると、

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

$\lambda_1 \geq \cdots \geq \lambda_r > 0$ とすると、 $\|\mathbf{b}\|^2 = 1$ 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、 $\mathbf{a} = \mathbf{u}_1$ のとき、 P_z が最大となる。

以上の議論から、

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

で定義される $\{z^{(i)}\}$ 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、 $\lambda_1 \geq \cdots \geq \lambda_r > 0$ 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。

今、

$$\mathbf{X} = (\mathbf{x}^{(1)} - \boldsymbol{\mu} \quad \cdots \quad \mathbf{x}^{(N)} - \boldsymbol{\mu})$$

の特異値分解から、主成分分析における各主成分が得られた。一方、通常、主成分分析は、 $\{\mathbf{x}^{(i)}\}$ の分散共分散行列

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

の固有値分解

$$\mathbf{P} = \frac{1}{n-1} \mathbf{X} \mathbf{X}^\top = \frac{1}{n-1} \mathbf{U} \boldsymbol{\Lambda}^2 \mathbf{U}^\top$$

から得られる。実際には、 \mathbf{X} の特異値分解 $\mathbf{X} = \mathbf{U} \mathbf{\Lambda} \mathbf{V}^\top$ を用いると、分散共分散行列は、

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

となるため、特異値分解を使っても、固有値分解を使っても同じことになる。