

応用気象学 IB

榎本剛

2024-12-02

Table of contents

| | |
|---------------------------|-----------|
| はじめに | 5 |
| 第 1 章 データ同化の基礎 | 7 |
| 1.1 線型回帰 | 8 |
| 1.2 ガウス分布 | 10 |
| 1.3 ベイズの定理 | 11 |
| 1.4 最大後驗推定 | 12 |
| 1.5 簡単な例題 | 13 |
| 1.6 ベクトル場に対する解析 | 18 |
| 第 2 章 数値最適化 | 23 |
| 2.1 スカラー関数の最小化 | 23 |
| 2.2 データ同化のコスト関数 | 24 |
| 2.3 最急降下法 | 25 |
| 2.4 ニュートン法 | 25 |
| 2.5 ガウス・ニュートン法 | 26 |
| 2.6 準ニュートン法 | 26 |
| 2.7 共軛勾配法 | 27 |
| 2.8 線型探索 | 31 |
| 第 3 章 変分法 | 35 |
| 3.1 3次元変分法 | 35 |
| 3.2 4次元変分法 | 36 |

| | | |
|--------------|--------------------------------|-----------|
| 3.3 | 随伴法の理論 | 36 |
| 3.4 | 随伴モデル | 37 |
| 3.5 | 随伴モデルの作成 | 38 |
| 3.6 | Lorenz-63 モデル | 40 |
| 3.7 | Lorenz-96 モデル | 42 |
| 3.8 | 拡散方程式 | 42 |
| 3.9 | 4 次の Runge-Kutta 法 | 45 |
| 3.10 | 接線型及び随伴の動作確認 | 47 |
| 第 4 章 | アンサンブルカルマンフィルタ | 49 |
| 4.1 | カルマンフィルタ | 49 |
| 4.2 | アンサンブルカルマンフィルタ | 49 |
| 4.3 | 解析誤差共分散推定に対する標本誤差の影響 | 51 |
| 4.4 | 単一観測の同化 | 53 |
| 4.5 | 摂動法 | 54 |
| 4.6 | アンサンブル平方根フィルタ | 55 |
| 4.7 | 決定論的アンサンブルカルマンフィルタ | 55 |
| 4.8 | 準地衡流モデル | 57 |
| 第 5 章 | 感度解析 | 61 |
| 5.1 | 摂動の線型成長 | 61 |
| 5.2 | 随伴感度 | 62 |
| 5.3 | 特異ベクトル感度 | 63 |
| 5.4 | アンサンブル随伴感度解析 | 63 |
| 5.5 | 単回帰 | 64 |
| 5.6 | 多変数回帰 | 64 |
| 5.7 | アンサンブル SV | 65 |
| 5.8 | 日本域に対するアンサンブル感度 | 65 |
| 5.9 | 2003 年 3 月の爆弾低気圧 | 65 |
| 5.10 | 乾燥全エネルギーノルムの時間発展 | 66 |

| | | |
|-------------------|---|-----------|
| 5.11 | 欧州の切離低気圧と熱帯低気圧 Cristobal 2002 | 67 |
| 5.12 | アンサンブル SV の適用例 | 68 |
| 5.13 | 台風 Hagibis 2019 | 68 |
| 5.14 | まとめ | 69 |
| 第 6 章 | 機械学習 | 71 |
| 6.1 | 線型回帰 | 71 |
| 6.2 | ロジスティック回帰 | 73 |
| 6.3 | ナイーブベイズ | 73 |
| 6.4 | 決定木 | 74 |
| 6.5 | サポートベクトルマシン | 74 |
| 6.6 | カーネル法 | 76 |
| 6.7 | 線型回帰 | 77 |
| 6.8 | カーネル回帰 | 78 |
| 6.9 | ニューラルネットワーク | 79 |
| 6.10 | 自己組織化マップ | 80 |
| 6.11 | 畳み込みニューラルネットワーク | 80 |
| 6.12 | トランスフォーマ | 81 |
| 参考文献 | | 85 |
| Appendices | | 93 |
| 付録 A | 補遺 | 93 |
| A.1 | 線型問題 | 93 |
| A.2 | 様々な行列 | 93 |
| A.3 | 行列の定値性 | 94 |
| A.4 | ベクトル空間 | 94 |
| A.5 | 微分 | 94 |
| A.6 | 逆行列 | 95 |
| A.7 | 擬逆行列 | 95 |

はじめに



First of all he said to himself: "That buzzing-noise means something. You don't get a buzzing-noise like that, just buzzing and buzzing, without its meaning something. If there's a buzzing-noise, somebody's making a buzzing-noise, and the only reason for making a buzzing-noise that I know of is because you're a bee."

Then he thought another long time, and said: "And the only reason for being a bee that I know of is making honey."

And then he got up, and said: "And the only reason for making honey is so as I can eat it." So he began to climb the tree.

Winnie-the-Pooh (Milne 1926)

真理に近づく方法として、科学では演繹 (deduction) 法と帰納 (induction) 法がある。演繹は、南宋の儒学者朱熹による中庸への注釈、中庸章句にある「更互演繹，作為此書」に由来し、前提から論理を展開し結論を導き出す。気象学では、支配方程式に基づく気象力学や、それをコンピュータのプログラムとして書き起こしたモデルによるシミュレーションは演繹である。一方、帰納法は事実を積み重ねて一般的な法則を導き出す。気象学では、観測的研究や、データ解析が該当するだろう。

データ同化は、演繹と帰納を組み合わせた研究手法である。通常モデルにより第一推定値を演繹的に得る。統計的推論を基礎として観測によりモデルを修正する。数値天気予報の初期値を作成するという必要から研究が進んだため、単なる技術と認識され、予報精度を高める初期値が探求されてきた。しかし、研究手法の観点からは、力学及びモデル並びに観測及び解析の両方を利用した、真理を探究するためアプローチであると言える。

データ同化が基礎とする学術分野は、解析学 (変分法)、関数解析学、線型代数学、統計学、最適化理論、非線型力学 (カオス理論)、制御理論、情報理論など多岐に亘る。

本講義では、以下のようなデータ同化に関連した話題を取り上げる。

- 統計的推定: 最初二乗法、最尤推定
- 数値最適化: 最急降下法、ニュートン法、共軛勾配法
- 変分法: 随伴モデル、4DVar
- アンサンブルカルマンフィルタ: EAKF、DEnKF、インフレーション、局所化
- 感度解析: 随伴、特異ベクトル、アンサンブル
- 機械学習: ニューラルネットワーク、SOM、カーネル法
- 数学的補遺: 様々な行列、行列の微分、行列の分解

描画や簡単な計算には、R を用いている。一部のコードは Python と Fortran に翻訳して、[リポジトリ](#)に置いてある。

このページは次の QR コードからも接続できる。



第1章

データ同化の基礎

観測データは初期値の格子点上で取得されたとは限らず、取得された時刻も様々で、観測される変数、観測点の分布、測器の精度も多様である。数値天気予報 (numerical weather prediction) の初期値としては、規則的な格子上のデータが必要である Figure 1.1。観測データ (observation) を数値予報モデルに取り込むことをデータ同化 (data assimilation) という。客観的な一定の方法に従って、最適 (optimal) な場を推定したものを客観解析 (objective analysis) という。

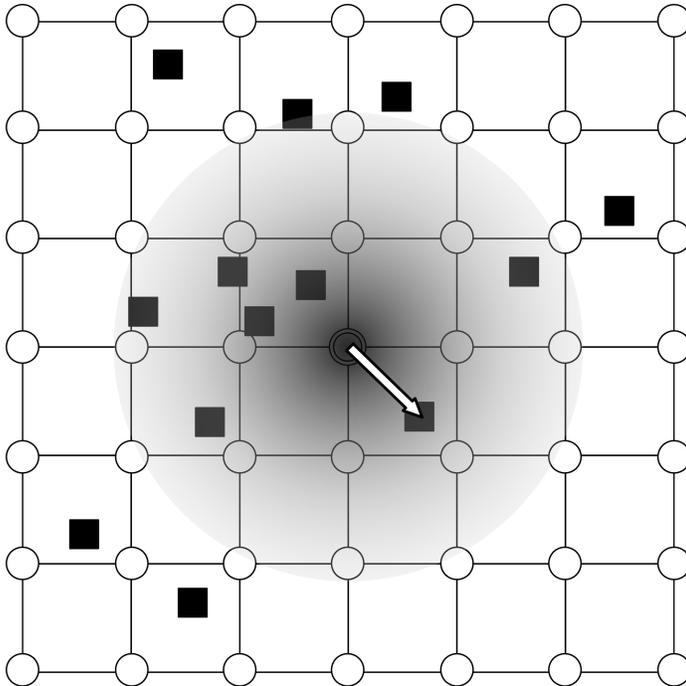


Figure1.1: 観測データの同化の概念図。観測点は■、モデルの格子点は○、円形の陰影は格子点(◎)に対して観測が影響する範囲。Kalnay (2003) を一部改変

1.1 線型回帰

回帰分析 (regression analysis) は、二つの変数の x と y のデータが与えられたときに、 $y = f(x)$ というモデルを当てはめる。 x を説明変数、 y を目的変数と呼び、 x がスカラーの場合を単回帰 (simple regression)、ベクトルの場合を重回帰 (multiple regression) という。モデルが線型 $f(x) = \beta_0 + \beta_1 x$ なら、線型回帰、 $f(x) = ax^b$, $f(x) = ae^{bx}$, $f(x) = a + b \log x$ のように非線型なら非線型回帰という。様々な当てはめ (curve-fitting) がある。

n 個の観測 y_1, \dots, y_n が与えられたとき、線型モデルでは次のように書ける。

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \quad (1.1)$$

ε_i は誤差を表している。最小分散推定 (minimum variance) は最小二乗法 (method of least squares) と呼ばれ、誤差の分散

$$S = \sum_i \varepsilon_i^2 \quad (1.2)$$

を最小にする β_0, β_1 を求める。

Equation 1.2 を β_0, β_1 で微分して 0 と置くと、正規方程式

$$\begin{aligned} \frac{\partial S}{\partial \beta_0} &= -2 \sum [y_i - (\beta_0 + \beta_1 x_i)] = 0 \\ \frac{\partial S}{\partial \beta_1} &= -2 \sum [y_i - (\beta_0 + \beta_1 x_i)] x_i = 0 \end{aligned} \quad (1.3)$$

平均

$$\bar{x} = \frac{1}{n} \sum x_i, \quad \bar{y} = \frac{1}{n} \sum y_i$$

を用いると、Equation 1.3 の最初の式から

$$\bar{y} = \beta_0 + \beta_1 \bar{x} \quad (1.4)$$

が得られる。

$$\begin{aligned} \sum (x_i - \bar{x}) \bar{x} &= \bar{x} \sum x_i - n \bar{x}^2 = 0 \\ \sum (y_i - \bar{y}) \bar{x} &= \bar{x} \sum y_i - n \bar{x} \bar{y} = 0 \end{aligned}$$

を用いると

$$\begin{aligned} \hat{\beta}_0 &= \bar{y} - \beta_1 \bar{x} \\ \hat{\beta}_1 &= \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2} = \frac{\text{cov}(x, y)}{\text{var}(x)} \end{aligned}$$

を得る。 $\hat{\beta}_1$ を回帰係数という。

1.1.1 例 1 予報と観測

2020 年 12 月の気象庁 GSM の 0000 UTC から 6 時間予報と京都のアメダスで観測された気温と風速を比較する。

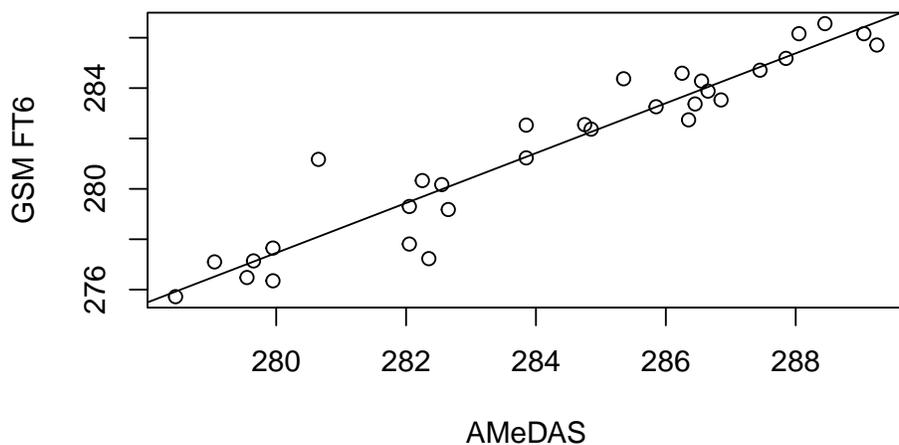
まず、平均を計算し、分散、共分散から、傾きと切片を計算する。

```
mean.x= 284.1565  var.x= 10.51396  mean.y= 281.5745  cov.xy= 10.4146
```

```
beta0 = 0.1032584  beta1= 0.9905503
```

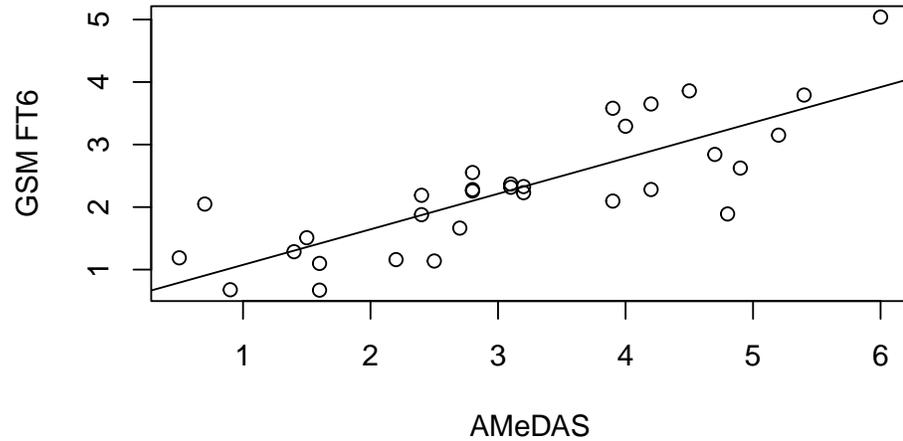
散布図を描いてみると、非常に対応が良いことが分かる。

T2m Kyoto 202012 15UTC bias=0.10 slope=0.99



次に 10 m 風速と京都のアメダスで観測された風速を比較する。予報は観測される強い風を表現しておらず、回帰直線の傾きが小さい。

|u| Kyoto 202012 15UTC bias=0.51 slope=0.57



1.2 ガウス分布

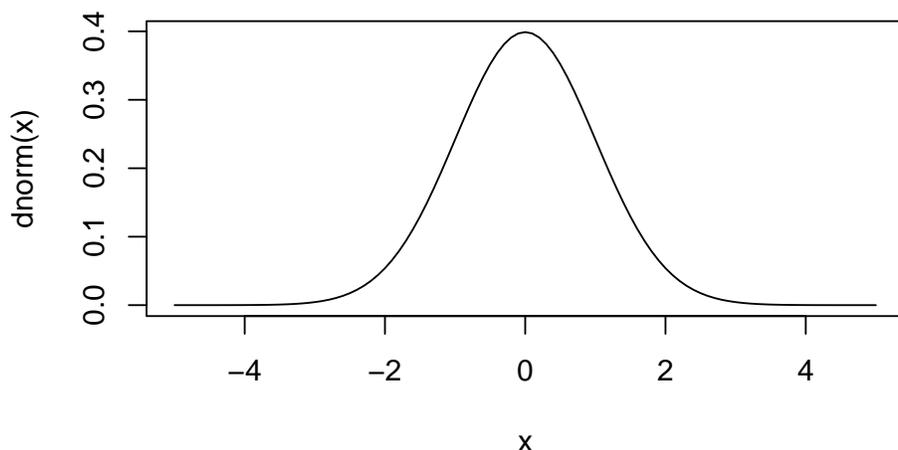
平均 \bar{x} 、標準偏差 σ の正規分布は

$$P(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(x - \bar{x})^2}{2\sigma^2}\right\} \quad (1.5)$$

n 次元のベクトル \mathbf{x} に対しては

$$P(\mathbf{x}) = \frac{1}{(2\pi)^{n/2} |\mathbf{S}_x|^{1/2}} \exp\left\{-(\mathbf{x} - \bar{\mathbf{x}})^T \mathbf{S}_x^{-1} (\mathbf{x} - \bar{\mathbf{x}})\right\} \quad (1.6)$$

と書ける。ここで \mathbf{S}_x は共分散行列である。



1.3 ベイズの定理

Rodgers (2000) に基づいて、状態 \mathbf{x} と観測 \mathbf{y} がスカラーの場合について、同時確率密度分布 (*pdf*: probability distribution function) の例を図示する。

- $P(\mathbf{x})$: 状態 \mathbf{x} の *pdf*。 $P(\mathbf{x})d\mathbf{x}$ は、測定をする前に多次元体積 $(\mathbf{x}, \mathbf{x} + d\mathbf{x})$ に \mathbf{x} が存在する確率。測する前において \mathbf{x} について分かっている。 $\int P(\mathbf{x})d\mathbf{x} = 1$ となるように規格化される。
- $P(\mathbf{y})$: 測定の先験 *pdf*。測定前の測定の *pdf* を表す。
- $P(\mathbf{x}, \mathbf{y})$: \mathbf{x} と \mathbf{y} との同時先験 *pdf*。 $P(\mathbf{x}, \mathbf{y})d\mathbf{x}d\mathbf{y}$ は、 \mathbf{x} が $(\mathbf{x}, \mathbf{x} + d\mathbf{x})$ に存在しかつ \mathbf{y} が $(\mathbf{y}, \mathbf{y} + d\mathbf{y})$ に存在する確率。
- $P(\mathbf{y}|\mathbf{x})$: \mathbf{x} が与えられたときの \mathbf{y} の条件付 *pdf*。 $P(\mathbf{y}|\mathbf{x})d\mathbf{y}$ は \mathbf{x} の値が与えられたとき \mathbf{y} が $(\mathbf{y}, \mathbf{y} + d\mathbf{y})$ に存在する確率。
- $P(\mathbf{x}|\mathbf{y})$: \mathbf{y} が与えられたときの \mathbf{x} の条件付 *pdf*。 $P(\mathbf{x}|\mathbf{y})d\mathbf{x}$ は \mathbf{y} の値が与えられたとき \mathbf{x} が $(\mathbf{x}, \mathbf{x} + d\mathbf{x})$ に存在する確率。

Figure 1.2 において、 $P(\mathbf{y}|\mathbf{x})$ は \mathbf{x} に対する \mathbf{y} の関数であり、 $\int P(\mathbf{y}|\mathbf{x})d\mathbf{y} = 1$ に規格化されるので、破線に沿った積分 $\int P(\mathbf{x}, \mathbf{y})d\mathbf{y} = P(\mathbf{x})$ で同時分布 $P(\mathbf{x}, \mathbf{y})$ を割ったものに等しい。

$$P(\mathbf{y}|\mathbf{x}) = \frac{P(\mathbf{x}, \mathbf{y})}{P(\mathbf{x})} \quad (1.7)$$

$P(\mathbf{x}|\mathbf{y})$ についても同様に表し、Equation 1.7 を用いて $P(\mathbf{x}, \mathbf{y})$ を消去すると、ベイズの

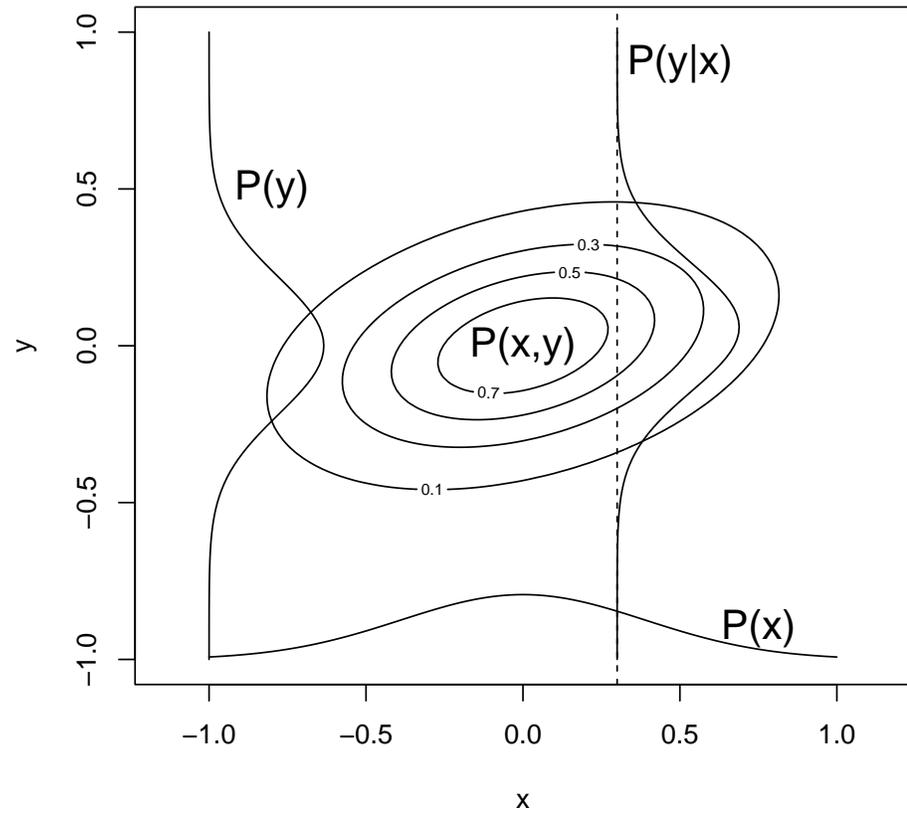


Figure1.2: 2次元における Bayes の定理の概念図

定理が得られる。

$$P(x|y) = \frac{P(x,y)}{P(y)} = \frac{P(y|x)P(x)}{P(y)} \quad (1.8)$$

Equation 1.8 から $P(x,y)$ をベクトルに拡張すると次のように書ける。

$$P(\mathbf{x}|y) = \frac{P(y|\mathbf{x})P(\mathbf{x})}{P(y)} \quad (1.9)$$

データ同化を含む逆問題では、 $P(\mathbf{x})$ 及び $P(y|\mathbf{x})$ から $P(\mathbf{x}|y)$ を推定する。Equation 1.9 の分母はスケーリングの定数で、通常は無視できる。

1.4 最大後驗推定

スカラーを一つ直接観測し

$$y = x + \epsilon$$

を得たとする。 x は第一推定値で分散は σ_f^2 、 ϵ は観測誤差、観測の分散は σ_o^2 とする。 y の先験分散は

$$\sigma_y^2 = \sigma_f^2 + \sigma_o^2 \quad (1.10)$$

である。

$P(y|x)$, $P(x)$, $P(y|x)$ がガウス分布 (Equation 1.5) であるとする、

$$\begin{aligned} -2 \ln P(y|x) &= \frac{y-x}{\sigma_o^2} + \text{const} \\ -2 \ln P(x) &= \frac{x-x^f}{\sigma_f^2} + \text{const} \\ -2 \ln P(x|y) &= \frac{x-x^a}{\sigma_a^2} + \text{const} \end{aligned}$$

と書ける。Bayes の定理 (Equation 1.9) を適用すると、 x^2 の係数から

$$\frac{1}{\sigma_a^2} = \frac{1}{\sigma_o^2} + \frac{1}{\sigma_f^2} \quad (1.11)$$

を得る。分散の逆数は精度の指標で、観測を第一推定値に同化することにより精度が向上することを示している。Equation 1.11 は

$$\sigma_a^2 = \frac{\sigma_o^2 \sigma_f^2}{\sigma_o^2 + \sigma_f^2} \quad (1.12)$$

と表すこともできる。一方、 x の係数から

$$\frac{x^a}{\sigma_a^2} = \frac{y}{\sigma_o^2} + \frac{x^f}{\sigma_f^2} \quad (1.13)$$

Equation 1.13 は Equation 1.11 を用いて

$$x^a = x^f + \frac{\sigma_f^2}{\sigma_o^2 + \sigma_f^2} (y - x^f) \quad (1.14)$$

と表すこともできる。後験確率密度が最大になるような推定であるので、最大後験推定 (MAP: Maximum a posteriori) と呼ばれている。最尤推定 (ML: maximum likelihood) 推定と呼ばれることもあるが、尤度は $P(\mathbf{y}|\mathbf{x})$ を指すので、厳密には先験情報がない場合にのみ、MAP 推定と ML 推定が一致する。

1.5 簡単な例題

ここで考える例題は、2つの気温の測定値 T_1, T_2 から解析値 T_a を求める問題を考える。真の値 (truth) を T_t とすると、測定値は

$$\begin{aligned} T_1 &= T_t + \varepsilon_1 \\ T_2 &= T_t + \varepsilon_2 \end{aligned}$$

と書ける。ここで $\varepsilon_1, \varepsilon_2$ は誤差を表す。真の値が分からないので、誤差 (error) も分からない。さらに、以下が仮定できるとする。

- バイアスがない。 $E(\varepsilon_1) = E(\varepsilon_2) = 0$ 。ここで、 $E(\cdot)$ は期待値を表す。
- 誤差の分散は既知。 $E(\varepsilon_1^2) = \sigma_1^2, E(\varepsilon_2^2) = \sigma_2^2$
- 誤差は無相関。 $E(\varepsilon_1\varepsilon_2) = 0$

$$T_a = a_1T_1 + a_2T_2 \quad (1.15)$$

- 解析誤差もバイアスがない。 $E(\varepsilon_a) = 0$ とすると、 $a_1 + a_2 = 1$ となる。

最小二乗法 (method of least squares) では、平均二乗誤差が最も小さくなるようにする。

$$\begin{aligned} \sigma_a^2 &= E[(T_a - T_t)^2] \\ &= a_1^2\sigma_1^2 + (1 - a_1)^2\sigma_2^2 \end{aligned}$$

微分をとると

$$\frac{\partial \sigma_a^2}{\partial a_1} = 2a_1\sigma_1^2 - 2(1 - a_1)\sigma_2^2$$

となる。最小値では微分が 0 になるので、

$$a_1 = \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2}, \quad a_2 = \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2} \quad (1.16)$$

と求まる。観測 1 の重みは、観測 2 の分散を分散の和で割ったものとして求まる。分子分母に $1/\sigma_1^2\sigma_2^2$ をかけると、

$$a_1 = \frac{1/\sigma_1^2}{1/\sigma_1^2 + 1/\sigma_2^2}, \quad a_2 = \frac{1/\sigma_2^2}{1/\sigma_1^2 + 1/\sigma_2^2} \quad (1.17)$$

とも書ける。ばらつき (分散) が大きいほど小さくなるので、分散の逆数は精度を示す。Equation 1.17 は、重みが精度に比例することを示す。分散が小さいものに、より大きな重みをつけることになる。

Equation 1.16 に Equation 1.16 を代入して、解析値の分散を求めてみる。

$$\begin{aligned} \sigma_a^2 &= \frac{\sigma_1^2\sigma_2^2}{\sigma_1^2 + \sigma_2^2} \\ \frac{1}{\sigma_a^2} &= \frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2} \end{aligned} \quad (1.18)$$

Equation 1.18 は、解析精度は全ての精度の和となる。

測定値 1 が第一推定値 ($T_1 = T_b$) 測定値 2 が観測値 ($T_2 = T_o$) であるとする。Equation 1.15 は

$$T_a = T_b + \frac{\sigma_b^2}{\sigma_b^2 + \sigma_o^2}(T_o - T_b) \quad (1.19)$$

と書くこともできる。 $T_o - T_b$ をイノベーション (innovation)、右辺第二項を解析インクリメント (increment) と呼ぶ。予測精度が悪い場合 (σ_b^2 が大きい場合) や第一推定値が観測値からずれている場合 (インクリメントが大きい場合)、Equation 1.19 の第 2 項が大きくなり、観測に近くなるように修正される。

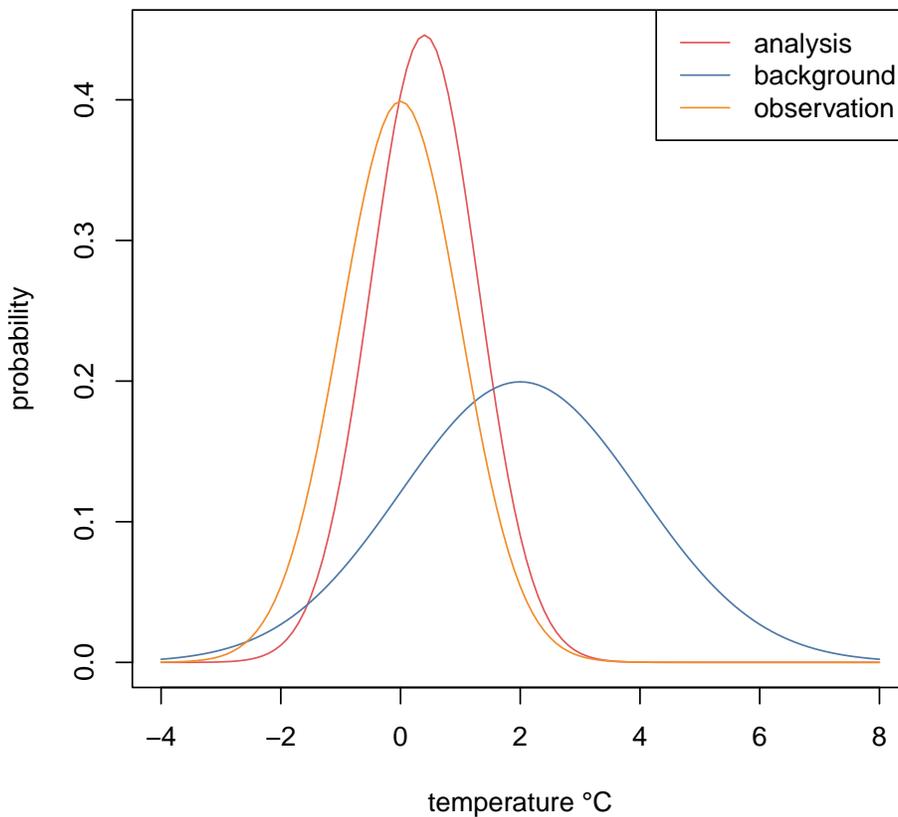


Figure1.3: 二つの測定に基づく解析

次に、最尤法に基づき T_1, T_2 の確率密度分布から最適な推定値を求めてみる。 T_1, T_2 の確率密度分布は正規分布で表すことができる。

$$\begin{aligned} p(T_1|T) &= \frac{1}{\sqrt{2\pi}\sigma_1} \exp\left[-\frac{(T_1-T)^2}{2\sigma_1^2}\right] \\ p(T_2|T) &= \frac{1}{\sqrt{2\pi}\sigma_2} \exp\left[-\frac{(T_2-T)^2}{2\sigma_2^2}\right] \end{aligned} \quad (1.20)$$

測定値が T_1, T_2 となる確率は、Equation 1.20 の二つの積で表すことができる。

$$L(T|T_1, T_2) = \frac{1}{2\pi\sigma_1\sigma_2} \exp\left[-\frac{(T_1-T)^2}{2\sigma_1^2} - \frac{(T_2-T)^2}{2\sigma_2^2}\right] \quad (1.21)$$

$L(\cdot)$ を尤度函数 (likelihood function) という。 T の尤もらしい値は、 $L(T|T_1, T_2)$ が最大になる値である。指数函数の引数に負号をかけた値

$$J(T) = \frac{1}{2} \left[\frac{(T_1-T)^2}{\sigma_1^2} + \frac{(T_2-T)^2}{\sigma_2^2} \right]$$

が最小になれば、 $L(T|T_1, T_2)$ の最大値が求まる。 $J(\cdot)$ をコスト函数 (cost function) と呼ぶ。このような推定法を最尤法 (maximum likelihood method) という。 $\partial J/\partial T = 0$ を計算すると、最尤法により求めた重みは最小二乗法で求めた重みと一致する。ただし、正規分布に従わない場合はこの限りではない。

さらに、リモートセンシングデータを同化する場合について考える。気温の代わりに遠隔測定された放射輝度 $y = h(T)$ が得られているとする。 $h(\cdot)$ を観測演算子 (observation operator) 呼ぶ。観測演算子は、気温から放射輝度のような変数変換だけでなく、空間内挿も含む。

気温の第一推定値は T_b 、予報誤差は ε_b で表すことにする。放射輝度の観測値は、 $y_o = h(T_b) + \varepsilon_o$ と表すことができる。 ε_o は観測誤差で、観測誤差と予報誤差は正規分布に従うと仮定する。イノベーションを誤差で表すと

$$y_o - h(T_b) = \varepsilon_o - H\varepsilon_b$$

となる。ここで

$$H = \frac{\partial h}{\partial T}$$

である。

最小二乗法で問題を解く。Equation 1.19 の形式で解析値 T_a を表すと

$$T_a = T_b + w(y_o - h(T_b)) = T_b + w(\varepsilon_o - H\varepsilon_b) \quad (1.22)$$

となる。すなわち、解析誤差は

$$\varepsilon_a = \varepsilon_b + w(\varepsilon_o - H\varepsilon_b)$$

解析値の分散は

$$\sigma_a^2 = \sigma_b^2 + w^2(\sigma_o^2 + \sigma_b^2 H^2) - 2w\sigma_b^2 H \quad (1.23)$$

となる。 w で微分して最適な重みを求めると

$$w = \sigma_b^2 H (\sigma_o^2 + \sigma_b^2 H^2)^{-1} \quad (1.24)$$

となる。 H でスケールされた重み

$$wH = \frac{\sigma_b^2 H^2}{\sigma_o^2 + \sigma_b^2 H^2} = \frac{H^2/\sigma_o^2}{1/\sigma_b^2 + H^2/\sigma_o^2}$$

は 0 と 1 の間の値を取る。観測の分散が大きい ($\sigma_o^2 \gg \sigma_b^2 H^2$) とき、解析値は第一推定値にほぼ等しくなる ($T_a \approx T_b$)。逆に予測の分散が大きい ($\sigma_b^2 H^2 \gg \sigma_o^2$) とき、解析値は観測値に近づく ($T_a \approx wy_o$)。

Equation 1.24 を Equation 1.23 に代入すると

$$\sigma_a^2 = (1 - wH)\sigma_b^2 \quad (1.25)$$

が得られる。また、

$$\frac{1}{\sigma_a^2} = \frac{1}{\sigma_b^2} + \frac{H^2}{\sigma_o^2} \quad (1.26)$$

と表すこともできる。この場合でも、解析精度は第一推定値の精度と観測の精度との和で表される。

最尤法において、観測演算子を含む場合のコスト関数は、

$$J(T) = \frac{1}{2} \left[\frac{(T_b - T)^2}{\sigma_b^2} + \frac{(y_o - h(T))^2}{\sigma_o^2} \right] \quad (1.27)$$

で表すことができる。

問 1

$T_b = 2.0^\circ\text{C}$, $\sigma_b = 2.0^\circ\text{C}$ という第一推定値と予報誤差の標準偏差を得たとする。

1. $T_o = 0.0^\circ\text{C}$, $\sigma_o = 1.0^\circ\text{C}$ のとき、 T_a と σ_a を求めよ。答: $T_a = 0.40^\circ\text{C}$,

$$\sigma_a = \sqrt{4/5} \approx 0.89^\circ\text{C}$$

2. $T_o = 0.0^\circ\text{C}$, $\sigma_o = 2.0^\circ\text{C}$ のとき、 T_a と σ_a を求めよ。答: $T_a = 1.0^\circ\text{C}$,
 $\sigma_a = \sqrt{2} \approx 1.4^\circ\text{C}$

1.6 ベクトル場に対する解析

Section 1.5 では1変数（気温）の1点について考えたが、ここでは多変数からなる複数の点に拡張する。モデルでは幾何形状に応じて、通常変数を2次元や3次元配列に格納する。データ同化では、全ての変数の全ての点が表す場を状態ベクトル（1次元） \mathbf{x} に格納する。

以下 Lorenc (1986) に基づいて解析方程式を導出する。Section 1.3 で学んだ Bayes の定理は

$$P(A|B) \propto P(B|A)P(A)$$

と表すことができる。事象 A と B は次のように表すことができる。

事象 A 場が実際に生じた \mathbf{x}_t となること、 $\mathbf{x} = \mathbf{x}_t$ である。ここで添字の t は真値を意味する。

事象 B 観測ベクトル \mathbf{y} が得られた観測 \mathbf{y}_o となること、 $\mathbf{y} = \mathbf{y}_o$ である。

「最良」推定は、

最小分散推定（平均）

$$\mathbf{x}_a = \int \mathbf{x} \mathbf{P}_a \, d\mathbf{x}$$

または最大後驗推定（モード）、 $\mathbf{P}_a(\mathbf{x})$ が最大になるような $\mathbf{x} = \mathbf{x}_a$ のいずれかである。ここで a は解析を表す。

先驗確率とは、観測を行う前の状態 \mathbf{x} についての知識である。スカラーの場合に誤差は分からないが、分散は既知とする仮定と同様に背景値は真値と同じ確率密度分布を持つと仮定する。そのような背景値が得られる予報モデルを完全モデル（perfect model）と呼ぶ。

$$P(\mathbf{x} = \mathbf{x}_t) = P(\mathbf{x} - \mathbf{x}_b)$$

観測は偶然誤差が含まれている。これは、真の観測値 $\mathbf{y}_1 = \mathbf{y}_t$ が与えられたときの誤差を表す。

$$P(\mathbf{y} = \mathbf{y}_o | \mathbf{y}_1 = \mathbf{y}_t) = P_o(\mathbf{y}_1 - \mathbf{y}_o)$$

Bayes 推定に必要な $P(B|A)$ つまり $P(\mathbf{y} = \mathbf{y}_o | \mathbf{x} = \mathbf{x}_t)$ を $P_{of}(\mathbf{y}_o - \mathbf{y}_f)$ で表す。ここで

$$\mathbf{y} = H(\mathbf{x}_t)$$

は観測演算子（観測の順モデル）である。

H が完全なら $P_{of}(\mathbf{y}_o - \mathbf{y}_f) = P_o(\mathbf{y}_1 - \mathbf{y}_o)$ であり、観測には偶然誤差だけが含まれる。実際には H はモデル誤差を含んでいる。モデル誤差の確率密度を

$$P(\mathbf{y}_1 = \mathbf{y}_t | \mathbf{x} = \mathbf{x}_t) = P_f(\mathbf{y}_1 - \mathbf{y}_t)$$

で表す。

B' を b' から $b' + db'$ の間にある確率とすると、 $P(A') = \int P(A'|B')P(B')db'$ を用いて

$$P(\mathbf{y} = \mathbf{y}_o | \mathbf{x} = \mathbf{x}_t) = \int P(\mathbf{y} = \mathbf{y}_o | \mathbf{x} = \mathbf{x}_t \text{ and } \mathbf{y}_1 = \mathbf{y}_t) P(\mathbf{y}_1 = \mathbf{y}_t | \mathbf{x} = \mathbf{x}_t) d\mathbf{y}_1$$

つまり

$$P_{of}(\mathbf{y}_o - \mathbf{y}_f) = \int P_o(\mathbf{y}_1 - \mathbf{y}_o) P_f(\mathbf{y}_1 - \mathbf{y}_f) d\mathbf{y}_1 \quad (1.28)$$

よって解析誤差共分散は

$$\begin{aligned} P_a(\mathbf{x}) &= P(\mathbf{x} = \mathbf{x}_t | \mathbf{y} = \mathbf{y}_o) \propto P(\mathbf{y} = \mathbf{y}_o | \mathbf{x} = \mathbf{x}_t) P(\mathbf{x} = \mathbf{x}_t) \\ P_a(\mathbf{x}) &= P_{of}(\mathbf{y}_o - H(\mathbf{x})) P_b(\mathbf{x} - \mathbf{x}_b) \\ &= \left\{ \int P_o(\mathbf{y}_1 - \mathbf{y}_o) P_f(\mathbf{y}_1 - H(\mathbf{x})) d\mathbf{y}_1 \right\} P_b(\mathbf{x} - \mathbf{x}_b) \end{aligned}$$

となる。

背景誤差 $P_b(\mathbf{x} - \mathbf{x}_b)$ と観測誤差 $P_{of}(\mathbf{y}_o - H(\mathbf{x}))$ とは無相関で、確率密度分布がガウス型である仮定する。

$$\begin{aligned} P_b(\mathbf{x} - \mathbf{x}_b) &\propto \exp \left[-\frac{1}{2} (\mathbf{x} - \mathbf{x}_b)^T \mathbf{B}^{-1} (\mathbf{x} - \mathbf{x}_b) \right] \\ P_o(\mathbf{y} - \mathbf{y}_o) &\propto \exp \left[-\frac{1}{2} (\mathbf{y} - \mathbf{y}_o)^T \mathbf{O}^{-1} (\mathbf{y} - \mathbf{y}_o) \right] \\ P_f(\mathbf{y} - \mathbf{y}_f) &\propto \exp \left[-\frac{1}{2} (\mathbf{y} - \mathbf{y}_f)^T \mathbf{F}^{-1} (\mathbf{y} - \mathbf{y}_f) \right] \end{aligned}$$

ここで

$$\mathbf{B} = \mathbb{E}[(\mathbf{x}_b - \mathbf{x}_t)(\mathbf{x}_b - \mathbf{x}_t)^T]$$

$$\mathbf{O} = \mathbb{E}[(\mathbf{y}_o - \mathbf{y}_t)(\mathbf{y}_o - \mathbf{y}_t)^T]$$

$$\mathbf{F} = \mathbb{E}[(\mathbf{y}_t - H(\mathbf{x}_t))(\mathbf{y}_t - H(\mathbf{x}_t))^T]$$

である。以下 $\mathbf{R} \equiv \mathbf{O} + \mathbf{F}$ と書く。

Equation 1.28 で定義した P_{of} は次のように表すことができる。

$$P_{\text{of}} \propto \exp \left[\frac{1}{2} (\mathbf{y}_o - \mathbf{y}_t)^T \mathbf{R}^{-1} (\mathbf{y}_o - \mathbf{y}_t) \right]$$

最小分散推定と最大後驗推定は一致する。

最大後驗推定において、 $P_a(\mathbf{x})$ の最大化は、コスト関数 $J = -\ln P_a(\mathbf{x})$ の最小化と等価である。

$$J = \frac{1}{2} [\mathbf{y}_o - H(\mathbf{x})]^T \mathbf{R}^{-1} [\mathbf{y}_o - H(\mathbf{x})] + \frac{1}{2} (\mathbf{x} - \mathbf{x}_b)^T \mathbf{B}^{-1} (\mathbf{x} - \mathbf{x}_b)$$

観測演算子を線型化する。

$$H(\mathbf{x} + d\mathbf{x}) \approx H(\mathbf{x}) + \mathbf{H}d\mathbf{x}$$

ここで $\mathbf{H} = \partial H / \partial \mathbf{x}$ は観測演算子のヤコビアンである。

J を最小にする解析値 $\mathbf{x} = \mathbf{x}_a$ は

$$0 = \mathbf{H}^T \mathbf{R}^{-1} [\mathbf{y}_o - H(\mathbf{x}_a)] + \mathbf{B}^{-1} (\mathbf{x}_b - \mathbf{x}_a)$$

より

$$\begin{aligned} \mathbf{x}_a &= \mathbf{x}_b + (\mathbf{H}^T \mathbf{R}^{-1} \mathbf{H} + \mathbf{B}^{-1})^{-1} \mathbf{H}^T \mathbf{R}^{-1} [\mathbf{y}_o - H(\mathbf{x}_b)] \\ &= \mathbf{x}_b + \mathbf{B} \mathbf{H}^T (\mathbf{H} \mathbf{B} \mathbf{H}^T + \mathbf{R})^{-1} [\mathbf{y}_o - H(\mathbf{x}_b)] \end{aligned}$$

が得られる。状態ベクトル \mathbf{x} の大きさを n 、観測ベクトルの大きさを m とすると、最初の式では $n \times n$ 、二つ目の式では $m \times m$ の行列の逆を計算する必要がある。

この二つは Woodbury の式 (Golub and Van Loan 2013) を用いて示すこともできるが、以下のようにすると確認できる (Rodgers 2000)。

$$\mathbf{H}^T \mathbf{R}^{-1} (\mathbf{H} \mathbf{B} \mathbf{H}^T + \mathbf{R}) = (\mathbf{H}^T \mathbf{R}^{-1} \mathbf{H} + \mathbf{B}^{-1}) \mathbf{B} \mathbf{H}^T$$

の両辺に左から $(\mathbf{H}^T \mathbf{R}^{-1} \mathbf{H} + \mathbf{B}^{-1})^{-1}$ 、右から $(\mathbf{H} \mathbf{B} \mathbf{H}^T + \mathbf{R})^{-1}$ を掛けると

$$(\mathbf{H}^T \mathbf{R}^{-1} \mathbf{H} + \mathbf{B}^{-1})^{-1} \mathbf{H}^T \mathbf{R}^{-1} = \mathbf{B} \mathbf{H}^T (\mathbf{H} \mathbf{B} \mathbf{H}^T + \mathbf{R})^{-1}$$

が得られる。

線型ガウス型の場合の解析誤差共分散の期待値は

$$\mathbb{E}[(\mathbf{x}_a - \mathbf{x}_t)(\mathbf{x}_a - \mathbf{x}_t)^T] = \mathbf{B} - \mathbf{B} \mathbf{H}^T (\mathbf{H} \mathbf{B} \mathbf{H}^T + \mathbf{R})^{-1} \mathbf{H} \mathbf{B} \quad (1.29)$$

となる。

第 2 章

数値最適化

最小解または最大解を数値的に求める手法を数値最適化 (numerical optimization) という。変分法データ同化や機械学習では、コスト関数を最小化して最も尤もらしい解を求める。最尤解を求めるには確率密度関数の最大値をとる解を求めても良いが、通常は負号をつけた対数尤度が最小を取る解を求める。

例えば、Equation 1.5 に対しては、

$$\frac{(x - \bar{x})^2}{2\sigma^2}$$

Equation 1.6 に対しては、

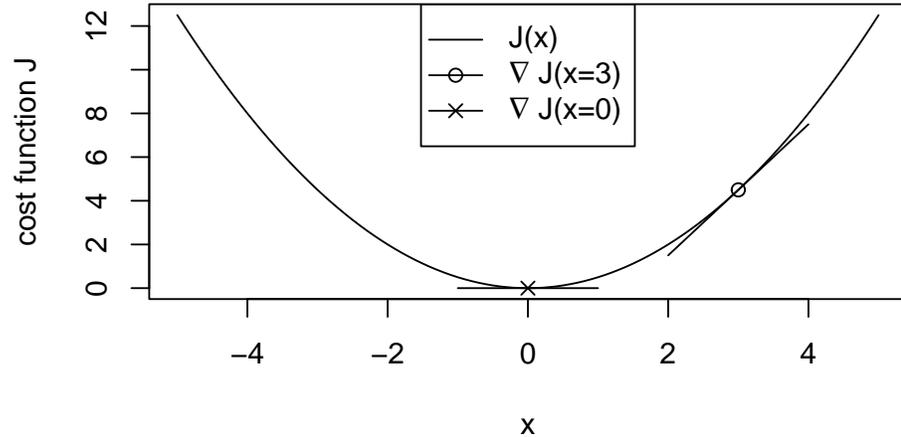
$$(\mathbf{x} - \bar{\mathbf{x}})^T \mathbf{S}_x^{-1} (\mathbf{x} - \bar{\mathbf{x}})$$

を最小化する。

最小解には、大域最小解と局所最小解がある。関数が凸であれば大域最小解が求まるが、凸でない場合は一般的には局所最小解が得られる。数値最適化手法は、勾配を用いるものと用いないものに大別される。ここでは、効率がよい勾配を用いる手法を学ぶ。

2.1 スカラー関数の最小化

最小化の概念や用語を学ぶために、スカラーの 2 次関数 $J(x) = x^2/2$ を考える。 $x = 3$ において勾配は $9/2$ で正なので、逆方向である負の方向に進めば J は小さくなる。 $dJ/dx = 0$ となる $x = 0$ は、勾配が 0 の停留点である (1 次の最適性必要条件)。 $d^2J/dx^2 = 1 > 0$ なので、関数は (下に) 凸 (convex) であり、 J は最小値である (2 次の最適性十分条件)。



2.2 データ同化のコスト関数

データ同化で用いる一般的なコスト関数は次のように書ける。

$$J(\mathbf{x}) = \frac{1}{2}(\mathbf{x} - \mathbf{x}_f)^T \mathbf{B}^{-1}(\mathbf{x} - \mathbf{x}_f) + \frac{1}{2}[\mathbf{y} - H(\mathbf{x})]^T \mathbf{R}^{-1}[\mathbf{y} - H(\mathbf{x})] \quad (2.1)$$

\mathbf{x} は状態ベクトルで、最適化の対象となる制御変数 (control variable) である。 \mathbf{x}_f は第一推定値、 \mathbf{y} は観測ベクトルを表す。 H は観測演算子 (observation operator) を表し、リトールでは順モデル (forward model) と呼ばれる。行列 \mathbf{B} と \mathbf{R} はそれぞれ背景及び観測誤差共分散である。右辺第 1 項は先験値、第 2 項は測定からの乖離を表し、それぞれ背景項、観測項と呼ぶ。イノベーションが線型 $\mathbf{y} - \mathbf{H}\mathbf{x}$ の場合は 2 次になる。

\mathbf{x} がベクトルのとき、最適性条件を確認する。

- 十分条件: $\nabla J(\mathbf{x}) = 0$ かつ $\nabla^2 J(\mathbf{x})$ が半正定値 (Section A.3) ならば \mathbf{x} は局所最小解である。
- 必要条件: \mathbf{x} は局所最小解ならば、 $\nabla J(\mathbf{x}) = 0$ かつ $\nabla^2 J(\mathbf{x})$ が半正定値でなければならない。

\mathbf{x} で微分すると、Equation 2.1 の勾配は次のように表される。

$$\nabla J(\mathbf{x}) = \mathbf{B}^{-1}(\mathbf{x} - \mathbf{x}_f) + \mathbf{H}^T \mathbf{R}^{-1}(\mathbf{y} - H(\mathbf{x}))$$

いくつかの最適化手法を概観する。スラカー函数の最適化で2次函数は容易に最小解が求められることをみた。そこで、ベクトル函数 $f(\mathbf{x})$ を2次近似する。

$$f(\mathbf{x} + \mathbf{d}) \approx f(\mathbf{x}) + \mathbf{g}^T \mathbf{d} + \mathbf{d}^T \mathbf{G} \mathbf{d} \quad (2.2)$$

ここで \mathbf{d} は降下ベクトル、 $\mathbf{g} = \nabla f$ 及び $\mathbf{G} = \nabla^2 f$ は \mathbf{x} における勾配ベクトル及びヘシアンである。

2.3 最急降下法

最急降下法 (steepest descent) は1次までの情報、すなわち勾配を用いて降下方向を $\mathbf{d} = -\mathbf{g}$ とする。

$$\mathbf{x}^{(n+1)} = \mathbf{x}^{(n)} + \alpha \mathbf{d}$$

ここで n は反復回数、 α はステップ幅を表す。機械学習では、最急降下法に基づくさまざまな手法が広く用いられている。

ステップ幅の計算は α に対する制約付最適化問題

$$\arg \min_{\alpha > 0} f(\mathbf{x}^{(n)} + \alpha \mathbf{d})$$

を解くことにより得られる。最適化におけるこの小問題は、線型探索 (Section 2.8) と呼ばれている。

2.4 ニュートン法

2次の情報、すなわち曲率を表すヘシアンを用いるのがニュートン法である。Equation 2.2 の停留点は連立一次方程式をなすニュートン方程式

$$\mathbf{G} \mathbf{d} = -\mathbf{g} \quad (2.3)$$

を解くことにより得られる。ニュートン法は二次収束する。すなわち \mathbf{x}^* を最小値、 C を定数としたときに

$$\|\mathbf{x}^{(n+1)} - \mathbf{x}^*\| \leq C \|\mathbf{x}^{(n)} - \mathbf{x}^*\|^2$$

が成り立つ。二次函数に近ければ、少ない反復回数で収束する。

ニュートン方程式 Equation 2.3 を厳密に解くと、ステップ幅は $\alpha = 1$ となる。ニュートン方程式は二次近似なので $\alpha = 1$ に近く、線型探索を省略することもできる。

一方、ヘシアンを格納するには $n \times n$ のメモリが必要となり計算量も $O(n^3)$ と多いので、大規模な最適化に事実上適用できない。

2.5 ガウス・ニュートン法

関数が複数の関数の二乗和

$$f(\mathbf{x}) = \frac{1}{2} \sum_i^k [f_k(\mathbf{x})]^2$$

で表される場合を考える。Equation 2.1 はこれに該当する。

残差 $f_k(\mathbf{x})$ が線型及び非線型の場合を、それぞれ線型及び非線型最小二乗法問題と呼ぶ。残差を並べたベクトルを

$$\mathbf{f} = (f_1(\mathbf{x}) \quad f_2(\mathbf{x}) \quad \dots \quad f_k(\mathbf{x}))^T$$

と置き、ヤコビアン

$$\mathbf{F} = \frac{\partial \mathbf{f}}{\partial \mathbf{x}}$$

を用いて勾配とヘシアンを近似する。

$$\mathbf{g} = \mathbf{F}^T \mathbf{f}$$

$$\mathbf{G} = \mathbf{F}^T \mathbf{F}$$

ニュートン方程式 Equation 2.3 の解は

$$\mathbf{d} = -\mathbf{G}^{-1} \mathbf{g} = -(\mathbf{F}^T \mathbf{F})^{-1} \mathbf{F}^T \mathbf{f}$$

となる。ガウス・ニュートン法は、二次近似と整合的で、コスト函数のような（非）線型最小二乗法の定番となる解法である。

2.6 準ニュートン法

準ニュートン法は、ニュートン方程式 Equation 2.3 を厳密に解かない方法である。代表的な BFGS 法 Goldfarb (1970) は、反復の中で曲率についての情報を構築して、ヘシアンを明示的に計算することなく勾配で近似する。函数評価と勾配の計算量は $O(n^2)$ であり、ヘシアンの計算量 $O(n^3)$ よりも少ない。メモリは $O(n^2)$ 必要であるが、 $O(n)$ となる省メモリの L-BFGS 法 (Nocedal 1980; Liu and Nocedal 1989) も提案されている。

Nocedal and Stephen J. Wright (2006) に基づいて概要を説明する。

$k+1$ 回目の反復におけるコスト函数の二次近似を

$$m_{k+1}(p) = f_{k+1} + \nabla f_{k+1}^T p + \frac{1}{2} p^T B_{k+1} p$$

と表す。コスト函数とその近似の勾配が x_{k+1} ($p=0$) 及び x_k ($p=-\alpha p_k$) で成り立つとすると、割線条件

$$B_{k+1} s_k = y_k$$

と書ける。ここで $s_k = x_{k+1} - x_k$ 、 $y_k = \nabla f_{k+1} - \nabla f_k$ である。 B_k が正定値対称行列であれば曲率条件

$$s_k^T y_k > 0$$

が成り立つ。曲率条件は Wolfe 条件 Section 2.8.4 や強 Wolfe 条件 Section 2.8.5 を満たせば成立が保証される。

ヘシアン逆行列に対する割線条件は

$$H_{k+1} y_k = s_k$$

と書ける。割線条件と曲率条件を満たす H_{k+1} は無数にあるので、一意に定めるために現在の H_k と近いものを選ぶと

$$H_{k+1} = V_k^T H_k V_k + \rho_k s_k s_k^T \quad (2.4)$$

が得られる。ここで $V_k = I - \rho_k y_k s_k^T$ 及び $\rho_k = 1/y_k^T s_k$ である。これを BFGS 公式という。Equation 2.4 は階数 2 の更新である。準ニュートン法の基本的な発想は、ヘシアンを一から計算せずに現在のヘシアンを微修正して更新することにある。L-BFGS は、 H_k を明示的に持たずに必要なベクトルを保持するものである。

2.7 共軛勾配法

共軛勾配法 (conjugate gradient, CG) は少ないメモリで効率的に数値最適化を行うことができる手法である。ここでは、まず共軛の概念と関連する定理を示した後、互いに共軛な方向を構成する方法について述べ、共軛勾配法の計算手順を示す。

共軛 (きょうやく, conjugacy) は、戦後の漢字制限の下で「共役」と書き換えられた。「きょうえき」は誤読である。「軛」は車の轅 (ながえ) の先端につけて、車を引く牛馬の後ろにかける横木を指す (スーパー大辞林)。

G: 正定値対称行列、**b** をベクトル、*c* をスカラーとすると、**x** の 2 次関数は次のように表される。

$$F(\mathbf{x}) = \frac{1}{2} \mathbf{x}^T \mathbf{G} \mathbf{x} + \mathbf{b}^T \mathbf{x} + c$$

二つのベクトル $\mathbf{u} \neq 0$ と $\mathbf{v} \neq 0$ は以下が成り立つとき **G** に関して共軛であると言う。

$$\mathbf{u}^T \mathbf{G} \mathbf{v} = 0$$

幾何学的には、異なる γ は共通の中心を持つ楕円体の表面を表す。

$$F(\mathbf{x}) = \gamma$$

2.7.1 定理 1

ベクトル \mathbf{d}_j が互いに共軛、すなわち全ての $i \neq j$ について $\mathbf{d}_i^T \mathbf{G} \mathbf{d}_j = 0$ が成り立つならば、これらのベクトルは線型独立である。

- \mathbf{G} に関して互いに共軛な n 個の独立なベクトルの組が少なくとも 1 組ある-
- \mathbf{G} の固有ベクトルの組はそのような組を構成する。

2.7.2 定理 2

\mathbf{x}_k と \mathbf{x}_{k+1} とを $F(\mathbf{x})$ の最小化において、連続した二つのベクトルとする。もし

1. \mathbf{x}_k が \mathbf{d}_l 方向に $F(\mathbf{x})$ を最小化する。
2. \mathbf{x}_{k+1} が \mathbf{d}_m 方向に $F(\mathbf{x})$ を最小化する。
3. \mathbf{d}_l と \mathbf{d}_m は共軛方向。

ならば \mathbf{x}_{k+1} は \mathbf{d}_l 方向にも $F(\mathbf{x})$ を最小化する。

$F(\mathbf{x}_k)$ の勾配を

$$\mathbf{g}_k = \nabla F(\mathbf{x}_k)$$

と表記する。1. と 2. から $\mathbf{g}_{k+1} \mathbf{d}_i = 0$ 及び

$$\mathbf{d}_l^T \mathbf{G} \mathbf{d}_m = 0$$

が $i = 0, \dots, l$ のそれぞれについて成り立つことが示唆される。2 次関数 $F(\mathbf{x})$ に対して $\mathbf{g}(\mathbf{x}) = \mathbf{G}\mathbf{x} + \mathbf{b}$ なので

$$\mathbf{g}_{k+1} - \mathbf{g}_k = \mathbf{G}(\mathbf{x}_{k+1} - \mathbf{x}_k)$$

2. から

$$\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha_k \mathbf{d}_k$$

で更新を行う。ここでステップ幅 α_k は線型探索 (1 次元最小化) で求める。

$$F(\mathbf{x}_k + \alpha_k \mathbf{d}_k) = \min_{\alpha} F(\mathbf{x}_k + \alpha \mathbf{d}_k)$$

2.7.3 定理 3

$\mathbf{d}_i, i = 1, \dots, m (m \leq n)$ を互いに共軛な方向とする。

- $F(\mathbf{x})$ の大域最小値は任意の初期位置 \mathbf{x}_0 から有限回の降下計算で求めることができる。
- 降下計算において各 \mathbf{d}_i は降下方向として 1 回のみ用いられる。

2.7.4 互いに共軛な方向の組を構成する方法

$\mathbf{v}_0, \dots, \mathbf{v}_{n-1}$ を線型独立なベクトルの組とする。 \mathbf{G} について互いに共軛 (\mathbf{G} -共軛) な方向 $\mathbf{d}_0, \dots, \mathbf{d}_{n-1}$ の組は以下の手順で構成できる。

$$\mathbf{d}_0 = \mathbf{v}_0$$

とおき、連続的に

$$\mathbf{d}_i = \mathbf{v}_i + \sum_{j=0}^{i-1} a_{ij} \mathbf{d}_j$$

と定義する。

係数 a_{ij} は \mathbf{d}_i がこれまでの降下方向 $\mathbf{d}_{i-1}, \mathbf{d}_{i-2}, \dots, \mathbf{d}_0$ に \mathbf{G} -共軛となるように $l = 0, \dots, i-1$ に対して

$$\mathbf{d}_i^T \mathbf{G} \mathbf{d}_l = \mathbf{v}_i^T \mathbf{G} \mathbf{d}_l + \sum_{j=0}^{i-1} a_{ij} \mathbf{d}_j^T \mathbf{G} \mathbf{d}_l = 0$$

を満たすように選ぶ。これまでの係数 a_{ij} が $\mathbf{d}_0, \dots, \mathbf{d}_{i-1}$ が \mathbf{G} -共軛になるように選ばれていれば

$$\mathbf{d}_j^T \mathbf{G} \mathbf{d}_l = 0, \quad j \neq l$$

なので

$$a_{ij} = -\frac{\mathbf{v}_i^T \mathbf{G} \mathbf{d}_j}{\mathbf{d}_j^T \mathbf{G} \mathbf{d}_j}, \quad i = 1, \dots, n-1, \quad j = 0, \dots, i-1$$

が得られる。降下方向の組 $\mathbf{d}_0, \dots, \mathbf{d}_{n-1}$ は \mathbf{G} -共軛で $\mathbf{d}_0, \dots, \mathbf{d}_i$ で張られた部分空間と $\mathbf{v}_0, \dots, \mathbf{v}_i$ で張られた部分空間とは同一である。

2.7.5 共軛勾配法の計算方法

最初のステップは最急降下方向。

$$\mathbf{v}_0 = -\mathbf{g}_0$$

更新は

$$\mathbf{x}_1 = \mathbf{x}_0 + \alpha_0 \mathbf{d}_0$$

で行い、 α_0 は線型探索で求める。

次の降下方向は

$$\mathbf{d}_1 = -\mathbf{g}_1 + \frac{\mathbf{g}_1^T \mathbf{G} \mathbf{d}_0}{\mathbf{d}_0^T \mathbf{G} \mathbf{d}_0} \mathbf{d}_0$$

式 (6) より

$$\mathbf{g}_1 - \mathbf{g}_0 = \mathbf{G}(\mathbf{x}_1 - \mathbf{x}_0) = \alpha_0 \mathbf{G} \mathbf{d}_0$$

なので (16) は \mathbf{G} を用いずに

$$\mathbf{d}_1 = -\mathbf{g}_1 + \frac{\mathbf{g}_1^T(\mathbf{g}_1 - \mathbf{g}_0)}{\mathbf{d}_0^T(\mathbf{g}_1 - \mathbf{g}_0)} \mathbf{d}_0$$

と書くことができる。

$k+1$ 番目のステップは

$$\mathbf{d}_k = -\mathbf{g}_k + \sum_{j=0}^{k-1} \frac{\mathbf{g}_k^T \mathbf{G} \mathbf{d}_j}{\mathbf{d}_j^T \mathbf{G} \mathbf{d}_j} \mathbf{d}_j$$

または

$$\mathbf{d}_k = -\mathbf{g}_k + \sum_{j=0}^{k-1} \frac{\mathbf{g}_k^T(\mathbf{g}_{j+1} - \mathbf{g}_j)}{\mathbf{d}_j^T(\mathbf{g}_{j+1} - \mathbf{g}_j)} \mathbf{d}_j \quad (2.5)$$

と書ける。

$\mathbf{d}_0, \dots, \mathbf{d}_{k-1}$ で張られた部分空間と $\mathbf{v}_0, \dots, \mathbf{v}_{k-1}$ で張られた部分空間とは同一なので、3. より

$$\mathbf{g}_k^T \mathbf{d}_j = 0 \quad j = 0, \dots, k-1$$

及び

$$\mathbf{g}_k^T \mathbf{g}_j = 0 \quad j = 0, \dots, k-1$$

となることを利用すれば (Equation 2.5) を簡単にすることができる。

$$\mathbf{d}_k = -\mathbf{g}_k + \beta_k \mathbf{d}_{k-1}$$

ここで

$$\beta_k = \frac{\mathbf{g}_k^T(\mathbf{g}_k - \mathbf{g}_{k-1})}{\mathbf{d}_{k-1}^T(\mathbf{g}_k - \mathbf{g}_{k-1})}$$

$$\mathbf{g}_k^T \mathbf{g}_j = \mathbf{g}_k^T \mathbf{d}_j = 0, \quad j = 0, \dots, k-1$$

$$\mathbf{d}_{k-1} = -\mathbf{g}_{k-1} + \beta_{k-1} \mathbf{d}_{k-2}$$

を用いると

$$\beta_k = \frac{\mathbf{g}_k^T(\mathbf{g}_k - \mathbf{g}_{k-1})}{\mathbf{g}_{k-1}^T \mathbf{g}_{k-1}} = \frac{\mathbf{g}_k^T \mathbf{g}_k}{\mathbf{g}_{k-1}^T \mathbf{g}_{k-1}}$$

- 最小化の過程で降下方向 \mathbf{d}_k を選択するには、現在と一つ前の勾配 \mathbf{g}_k と \mathbf{g}_{k-1} 及び一つ前の降下方向 \mathbf{d}_{k-1} の三つのベクトルだけが分かれば良い。 n が大きいときに効果的。
- n 次元の2次関数は最大 n 回で収束する。

2.7.6 Polak–Ribiere アルゴリズム

$$\beta_k = \frac{\mathbf{g}_{k+1}^T (\mathbf{g}_{k+1} - \mathbf{g}_k)}{\mathbf{g}_k^T \mathbf{g}_k}$$

とする。

2 次関数では $\mathbf{g}_{k+1}^T \mathbf{g}_k = 0$ なので Fletcher–Reeves アルゴリズムに一致する。非線型性や非厳密線型探索が原因で共軛性が失われると、最適化が停滞する。このとき \mathbf{d}_k は \mathbf{g}_k とほぼ直交する。

$$\mathbf{g}(\mathbf{x}_{k+1}) = \nabla F(\mathbf{x}_{k+1}) \approx \nabla F(\mathbf{x}_k) = \mathbf{g}_k$$

Polak–Ribiere アルゴリズムの β_{k+1} は 0 に近くなり最急降下方向に向くことで停滞を打破する。

2.8 線型探索

線型探索では、探索方向 p_k に進む幅 α_k を求める。反復は次の式で与えられる。

$$x_{k+1} = x_k + \alpha_k p_k$$

- 探索方向 p_k は通常降下方向にとる。 $p_k^T \nabla f_k < 0$

2.8.1 ステップ幅

線型探索では、 x_k や p_k を固定し、最適化の対象を α に関する 1 変数関数と見る。

$$\phi(\alpha) = f(x_k + \alpha_k p_k), \alpha > 0$$

- 最小を厳密に求めるには、多数の関数評価が必要となる。
- なるべく計算コストをかけずに、ある程度 f が減少するような厳密ではない線型探索を行う。

2.8.2 Armijo 条件

十分な減少するように α を選択する。

$$f(x_k + \alpha_k p_k) \leq f(x_k) + c_1 \alpha_k \nabla f_k^T p_k$$

- 左辺は線型探索の対象とする $\phi(\alpha_k)$
- 右辺は α_k の1次関数 $l(\alpha_k)$ で、傾き $c_1\alpha\nabla f_k^T p_k$ は負
- 減少量はステップ幅 α_k 及び方向微分 $\nabla f_k^T p_k$ に比例
- $c_1 \in (0, 1)$ で通常 $c_1 = 10^{-4}$ のような非常に小さい値を用いる。

2.8.3 曲率条件

十分な減少だけでは、ステップ幅が極端に短くなり、最適化が進まないことがある。極端に短いステップ幅を除くために二つ目の条件として曲率条件を与える。

$$\nabla f(x_k + \alpha_k p_k)^T p_k \geq c_2 \nabla f_k^T p_k$$

- 左辺 $\phi'(\alpha)$ が右辺 $\phi'(0)$ の c_2 倍以上であるとする条件
 - $\phi'(\alpha_k)$ が大きな負なら、ステップ幅を大きくとったほうが良い。
 - あまり大きくない負か正ならこれ以上この方向に進む必要はないので探索を打ち切る。
- $c_2 \in (c_1, 1)$ でニュートン法や準ニュートン法では 0.9、共軛勾配法では 0.1 が用いられることが多い。

2.8.4 Wolfe 条件

十分な減少と曲率条件を合わせて、Wolfe 条件と呼ぶ。

$$\begin{aligned} f(x_k + \alpha_k p_k) &\leq f(x_k) + c_1 \alpha_k \nabla f_k^T p_k \\ \nabla f(x_k + \alpha_k p_k)^T p_k &\geq c_2 \nabla f_k^T p_k \end{aligned}$$

ここで $0 < c_1 < c_2 < 1$ である。

2.8.5 強 Wolfe 条件

Wolfe 条件は最小から離れた点ならどこでも満たされうる。そこで以下の制約を設け、強 Wolfe 条件と呼ぶ。

- $\phi'(\alpha_k)$ が大きな正にならないように制約
- 最小や極小、停留点のから遠いところを排除

$$\begin{aligned} f(x_k + \alpha_k p_k) &\leq f(x_k) + c_1 \alpha_k \nabla f_k^T p_k \\ |\nabla f(x_k + \alpha_k p_k)^T p_k| &\leq c_2 |\nabla f_k^T p_k| \end{aligned}$$

以下の条件のとき Wolfe 条件, 強 Wolfe 条件が満たすステップ幅 α_k の区間が存在する。

- f が連続微分可能
- f は下に有界
- $0 < c_1 < c_2 < 1$
- p_k が降下方向

第 3 章

変分法

変分法は Sasaki (1958) により、物理法則による拘束を課す解析手法として提案され、Talagrand and Courtier (1987) に制御理論に基づく理論的な裏付けがされている。Lorenç (1986) は、数値天気予報における様々な解析手法を比較し、変分法がベイズ推定から導かれることを指摘している。ここでは、Tsuyuki and Miyoshi (2007) に基づいて、変分法データ同化について述べる。

3.1 3次元変分法

3次元変分法同化 (3DVar: three-dimensional variational assimilation) は、第一推定値 \mathbf{x}^f 及び観測 \mathbf{y}^o が与えられたときに、後驗 pdf を最大化するものを解析値 \mathbf{x}^a とする。

$$\begin{aligned}\mathbf{x}^a &= \arg \max_{\mathbf{x}} [p(\mathbf{x}|\mathbf{x}^f, \mathbf{y}^o)] \\ &= \arg \max_{\mathbf{x}} [p(\mathbf{x}^f|\mathbf{x})p(\mathbf{y}^o|\mathbf{x})p(\mathbf{x})]\end{aligned}\quad (3.1)$$

ガウス型の pdf を仮定し、負の対数 pdf でコスト関数 $J(\mathbf{x})$ を定義すると、

$$J(\mathbf{x}) = \frac{1}{2}(\mathbf{x} - \mathbf{x}^f)^T \mathbf{B}^{-1}(\mathbf{x} - \mathbf{x}^f) + \frac{1}{2}(\mathbf{h}(\mathbf{x}) - \mathbf{y}^o)^T \mathbf{R}^{-1}(\mathbf{h}(\mathbf{x}) - \mathbf{y}^o) + J_c(\mathbf{x}) \quad (3.2)$$

となる。ここで、 \mathbf{B} , \mathbf{R} はそれぞれ背景及び観測誤差共分散行列、 $\mathbf{h}(\mathbf{x})$ は状態変数を観測変数に変換する観測演算子である。式 (Equation 3.2) の右辺第一、二項はそれぞれ第一推定値、観測値からのずれを表す。式 (Equation 3.2) の右辺第三項は、式 (Equation 3.1) の右辺第三項に対応し、制約項と呼ばれている。

3DVar では、数値最適化により $J(\mathbf{x})$ の最小値として解析値 \mathbf{x}^a を求める。 $\mathbf{h}(\mathbf{x})$ の接線型演算子を \mathbf{H} で表すと、制約項を無視した解析値は

$$\mathbf{x}^a = \mathbf{x}^f + (\mathbf{B}^{-1} + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H})^{-1} \mathbf{H}^T \mathbf{R}^{-1} [\mathbf{y}^o - \mathbf{h}(\mathbf{x}^f)] \quad (3.3)$$

となる。

3.2 4次元変分法

4次元変分法同化 (4DVar: four-dimensional variational assimilation) では、同化窓と呼ばれる期間に対する MAP 推定を行う。時刻レベル $i-1$ から i までの数値モデルによる予報を

$$\mathbf{x}_i = M_{i-1,i}(\mathbf{x}_{i-1}) \quad i = 1, \dots, n \quad (3.4)$$

で表す。4DVar のコスト関数は

$$J(\mathbf{x}_0) = \frac{1}{2}(\mathbf{x}_0 - \mathbf{x}_0^f)^T \mathbf{B}^{-1}(\mathbf{x}_0 - \mathbf{x}_0^f) + \frac{1}{2} \sum_{i=0}^n (h_i(\mathbf{x}_i) - \mathbf{y}_i^o)^T \mathbf{R}_i^{-1} (h_i(\mathbf{x}_i) - \mathbf{y}_i^o) + J_c(\mathbf{x}_0, \dots, \mathbf{x}_n) \quad (3.5)$$

で表される。予報誤差共分散は、同化窓内において M の接線型演算子 \mathbf{M} で時間発展し、解析値は式 (Equation 3.3) の \mathbf{B} を $\mathbf{M}^T \mathbf{B} \mathbf{M}$ で置き換えたものとなる。

4DVar では、解析値を効率的に求めるために、 $J(\mathbf{x}_0)$ の勾配を用いた最適化を用いる (Talagrand and Courtier 1987)。初期値 \mathbf{x}_0 に対する勾配

$$\mathbf{p}_0 = \nabla_{\mathbf{x}_0} J(\mathbf{x}_0) \quad (3.6)$$

を随伴方程式を用いて求める。随伴変数 $\mathbf{p}_{n+1} = \mathbf{0}$ を初期値として、時間逆向きに

$$\mathbf{p}_i = \mathbf{M}_i^T \mathbf{p}_{i+1} + \frac{\partial J}{\partial \mathbf{x}_i} \quad i = n, \dots, 0 \quad (3.7)$$

右辺第一項は、接線型モデル \mathbf{M}_i^T で一つ前の時間レベルまで、随伴変数を積分すること表している。以下に随伴モデルの作り方を示す。

3.3 随伴法の理論

随伴法の理論 (Talagrand and Courtier 1987; Talagrand 1991) について述べる。

ヒルベルト空間 \mathcal{F} での内積を (\cdot, \cdot) で表す。 $\mathbf{v} \rightarrow \mathcal{J}(\mathbf{v})$ は、 \mathcal{F} 上の微分可能な函数とする。 \mathcal{F} 上の任意の点で \mathcal{F} の微分 $\delta \mathcal{F}$ は、内積により \mathbf{v} の微分 $\delta \mathbf{v}$ で表すことができる。

$$\delta \mathcal{J} = (\nabla_{\mathbf{v}} \mathcal{J}, \delta \mathbf{v}) \quad (3.8)$$

ここで $\nabla_{\mathbf{v}} \mathcal{J}$ は、 \mathbf{v} についての \mathcal{F} の勾配で唯一に定まる。 \mathcal{F} が有限次元で正規直交系 v_i で表されるとき、 $\nabla_{\mathbf{v}} \mathcal{J}$ の要素は $\partial \mathcal{J} / \partial v_i$ となるが、勾配の概念はより一般的なものを表している。

\mathcal{F} とは別のヒルベルト空間 \mathcal{E} での内積を $\langle \cdot, \cdot \rangle$ で表し、 \mathbf{L} を \mathcal{E} から \mathcal{F} への連続な線型演算子とする。任意の $\mathbf{u} \in \mathcal{E}$ と任意の $\mathbf{v} \in \mathcal{F}$ に対して内積が等値である、 \mathcal{F} から \mathcal{E} への連続な線型演算子の \mathbf{L}^* が唯一存在する。

$$(\mathbf{v}, \mathbf{L}\mathbf{u}) = \langle \mathbf{L}^*\mathbf{v}, \mathbf{u} \rangle \quad (3.9)$$

\mathbf{L}^* は \mathbf{L} の随伴演算子 (adjoint operator) と呼ばれる。 \mathcal{E} と \mathcal{F} が有限次元で正規直交系座標で表されるとき、 \mathbf{L}^* は \mathbf{L} を表す行列の転置 (要素が複素数なら共軛転置) である。

必ずしも線型とは限らない、 \mathcal{E} から \mathcal{F} への可能な函数 $\mathbf{u} \rightarrow \mathbf{v} = \mathbf{G}(\mathbf{u})$ を考えると、 $\mathcal{J}(\mathbf{v}) = \mathcal{J}[\mathbf{G}(\mathbf{u})]$ は \mathbf{u} の合成函数となる。 \mathbf{u} の微分は Equation 3.8、 $\delta\mathbf{v}$ の微分は

$$\delta\mathbf{v} = \mathbf{G}'\delta\mathbf{u} \quad (3.10)$$

で与えられる。ここで \mathbf{G}' は \mathbf{G} を微分した \mathcal{E} から \mathcal{F} への線型演算子である。Equation 3.10 を Equation 3.8 に代入し、 \mathbf{G} の随伴 \mathbf{G}'^* を導入すると

$$\delta\mathcal{J} = (\nabla_{\mathbf{v}}\mathcal{J}, \mathbf{G}'\delta\mathbf{u}) = \langle \mathbf{G}'^*\nabla_{\mathbf{v}}\mathcal{J}, \delta\mathbf{u} \rangle$$

が得られる。勾配の定義により、 \mathcal{J} の \mathbf{u} についての勾配 $\nabla_{\mathbf{u}}\mathcal{J}$ は

$$\nabla_{\mathbf{u}} = \mathbf{G}'^*\nabla_{\mathbf{v}}\mathcal{J} \quad (3.11)$$

Equation 3.11 は制御理論における随伴方程式の基本であり、数値的に $\nabla_{\mathbf{u}}\mathcal{J}$ を計算する非常に効率的な方法を与える。ここでは、 $\mathbf{u} \rightarrow \mathbf{v} = \mathbf{G}(\mathbf{u})$ を想定しているので、 \mathbf{v} は \mathbf{u} の明示的ではあるが複雑な関数になっているため、計算可能な $\nabla_{\mathbf{u}}\mathcal{J}$ の解析的な式を得ることは困難である。 $\nabla_{\mathbf{u}}\mathcal{J}$ の数値的に見積もるために、 \mathbf{u} の各要素に摂動を加えて、それぞれの摂動に対して $\mathbf{v} = \mathbf{G}(\mathbf{u})$ を計算するのは、大次元では現実的ではない。Equation 3.11 が示しているのは、与えられた \mathbf{w} に対して $\mathbf{G}'^*\mathbf{w}$ を計算するプログラムがあれば、 $\nabla_{\mathbf{u}}$ は $\nabla_{\mathbf{v}}$ から容易に計算できるということである。 $\nabla_{\mathbf{v}}$ 自体は \mathcal{J} が \mathbf{v} の簡単な函数であれば容易に求めることができる。て $\mathbf{G}'^*\mathbf{w}$ の計算は、 $\mathbf{G}(\mathbf{u})$ を計算と同程度の複雑さであり、 $\mathbf{v} = \mathbf{G}(\mathbf{u})$ を計算して $\nabla_{\mathbf{v}}\mathcal{J}$ を求め、Equation 3.11 により最終的に $\nabla_{\mathbf{u}}$ を求める計算は $\mathbf{G}(\mathbf{u})$ の 2~3 倍程度であり、直接明示的に \mathbf{u} の各要素に摂動を与えるよりも大幅に計算量を削減できる。

3.4 随伴モデル

次の形の常微分方程式を考える。

$$\frac{d\mathbf{w}}{dt} = N(\mathbf{w}) + \mathbf{L}\mathbf{w} + \mathbf{f} \quad (3.12)$$

状態変数を \mathbf{w} 、非線型項を $N(\mathbf{w})$ 、線型項を $\mathbf{L}\mathbf{w}$ 、強制を \mathbf{f} で表している。状態変数に摂動を与えた $\mathbf{w} + \delta\mathbf{w}$ を (Equation 3.12) に代入し、摂動の時間発展 $d\delta\mathbf{w}/dt =$

$d(\mathbf{w} + \delta\mathbf{w})/dt - d\mathbf{w}/dt$ を求めると接線型モデル (TLM: tangent linear model) が得られる。

$$\frac{d\delta\mathbf{w}}{dt} = \left(\left. \frac{\partial N}{\partial \mathbf{w}} \right|_{\mathbf{w}} + \mathbf{L} \right) \delta\mathbf{w} \quad (3.13)$$

ここで、摂動 $\delta\mathbf{w}$ が微小であると仮定し、その二次以上の項を無視した。強制項 \mathbf{f} は TLM にはない。

$\partial N/\partial \mathbf{w}|_{\mathbf{w}} + \mathbf{L}$ を改めて \mathbf{L} と置き、入力 $\delta\mathbf{w}$ 及び出力 $\delta\dot{\mathbf{w}} = d\delta\mathbf{w}/dt$ を縦に並べたベクトルを用いて行列形式で表す。

$$\begin{bmatrix} \delta\mathbf{w} \\ \delta\dot{\mathbf{w}} \end{bmatrix} = \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{L} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \delta\mathbf{w} \\ \delta\dot{\mathbf{w}} \end{bmatrix} \quad (3.14)$$

つまり

$$\begin{aligned} \delta\mathbf{w} &= \delta\mathbf{w} \\ \delta\dot{\mathbf{w}} &= \mathbf{L}\delta\mathbf{w} \end{aligned}$$

である。随伴変数を \mathbf{w}^a とすると、(Equation 3.14) の随伴モデル (ADM, adjoint model) は

$$\begin{bmatrix} \mathbf{w}^a \\ \dot{\mathbf{w}}^a \end{bmatrix} = \begin{bmatrix} \mathbf{I} & \mathbf{L}^T \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{w}^a \\ \dot{\mathbf{w}}^a \end{bmatrix} \quad (3.15)$$

と表すことができる。すなわち、

$$\begin{aligned} \mathbf{w}^a &= \mathbf{w}^a + \mathbf{L}^T \dot{\mathbf{w}}^a \\ \dot{\mathbf{w}}^a &= \mathbf{0} \end{aligned}$$

である。 $\dot{\mathbf{w}}^a = \mathbf{0}$ も ADM の一部であり、忘れるとバグの原因となりうる。

3.5 随伴モデルの作成

上述のように、TLM が行列 \mathbf{L} で表せる場合は、ADM はその転置 \mathbf{L}^T を取ればよい。明示的に行列で表さなくても、TLM のソースコードを行毎に逆順にたどりながら ADM を作成することもできる。ここでは、後で用いる、いくつかの簡単な操作に対する随伴を求める。

3.5.1 代入

代入 $A = B$ の接線型は

$$\begin{bmatrix} \delta B \\ \delta A \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} \delta B \\ \delta A \end{bmatrix} \quad (3.16)$$

で、その随伴は

$$\begin{bmatrix} B^a \\ A^a \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} B^a \\ A^a \end{bmatrix} \quad (3.17)$$

つまり

$$\begin{aligned} B^a &= A^a + B^a \\ A^a &= 0 \end{aligned}$$

である。

3.5.2 係数を掛けて和を取る操作

代入 $C = \alpha A + \beta B + \gamma D$ の接線型は

$$\begin{bmatrix} \delta A \\ \delta B \\ \delta D \\ \delta C \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ \alpha & \beta & \gamma & 0 \end{bmatrix} \begin{bmatrix} \delta A \\ \delta B \\ \delta D \\ \delta C \end{bmatrix} \quad (3.18)$$

で、その随伴は

$$\begin{bmatrix} A^a \\ B^a \\ D^a \\ C^a \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & \alpha \\ 0 & 1 & 0 & \beta \\ 0 & 0 & 1 & \gamma \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} A^a \\ B^a \\ D^a \\ C^a \end{bmatrix} \quad (3.19)$$

つまり

$$\begin{aligned} A^a &= A^a + \alpha C^a \\ B^a &= B^a + \beta C^a \\ D^a &= D^a + \gamma C^a \\ C^a &= 0 \end{aligned}$$

である。

3.5.3 ループ

TLM の最後の行から逆順に ADM を作る。ループは添字を逆に回す。 $i = 1, 2, \dots, m, 1 \leq j \leq m$ に対して

$$\delta x_j = c_1 \delta x_1 + c_2 \delta x_2 + \dots + c_i \delta x_i + \dots + c_m \delta x_m \quad (3.20)$$

の随伴は

$$\begin{aligned} x_m^a &= x_m^a + c_m x_j^a \\ &\vdots \\ x_i^a &= x_i^a + c_i x_j^a \\ &\vdots \\ x_2^a &= x_2^a + c_2 x_j^a \\ x_1^a &= x_1^a + c_1 x_j^a \\ x_j^a &= c_j x_j^a \end{aligned} \quad (3.21)$$

TLM の行の右辺に現れる摂動変数 δx_i に対応して、ADM の行では随伴変数 x_i^a が左辺に現れる。 x_i^a には x_i^a 自身に係数 c_i と TLM の左辺に現れる摂動変数の随伴 x_j^a との積を加えたものになる。左辺の変数と右辺の変数が場所が入れ替わっている。 x_j^a は最後に現れる。 $c_j = 0$ の場合のように、TLM の右辺に x_j^a がない場合でも、 x_j^a に 0 を割り当てないと、他の部分でエラーを引き起こしうる。

ADM における変分演算は、微分演算とは異なり、ある時刻における状態ベクトル全体に対する摂動なので、状態ベクトルの一部が変化しないからといって、それを省略したものは ADM としては正しくない。

3.6 Lorenz-63 モデル

Lorenz (1963) のモデル (以下 Lorenz-63) は熱対流を理想化したモデルで、パラメタ次第でカオスにふるまう。Lorenz-63 は次の 3 変数の常微分方程式で表される。

$$\begin{aligned}\dot{X} &= -\sigma X + \sigma Y \\ \dot{Y} &= -XZ + rX - Y \\ \dot{Z} &= XY - \beta Z\end{aligned}$$

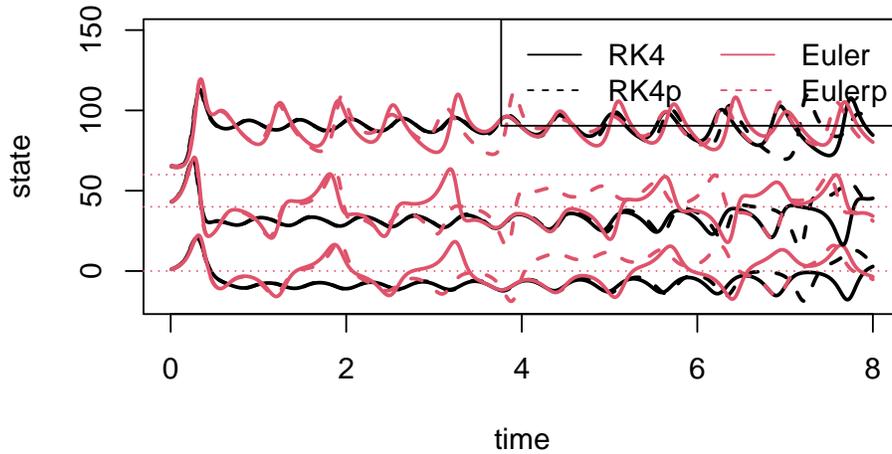
パラメタ σ, r, β はそれぞれ Prandtl 数, 変形された Rayleigh 数, アスペクト比を表す。予報変数 X, Y, Z はそれぞれ無次元化された対流の強さ, 最大温度差, 対流に伴う成層の変化を表す。左辺の変数の上の \cdot は時間微分 d/dt を表す記号である。

$\mathbf{x} = (1, 3, 5)$ (青) 及び $\mathbf{x} = (1.1, 3.3, 5.5)$ (橙) から無次元時刻 10 まで積分した結果の 3 次元可視化結果を示す。回転や拡大縮小 (右クリック) ができる。

../../../../private/var/folders/zr/fsq4_0pj2fsc52c5bbf1z9x80000gn/T/Rtmp-gyzzA4/file5bf5bb33ad9.png

Lorenz-63 の随伴モデルの作成方法は Huang and Yang (1996) に詳述されている。この報告書や [tenomoto/l63vda](#) では時間積分に Euler 法が用いられているが、4 次の Runge-Kutta 法 (RK4) を用いると、より安定するので、パラメタ r を大きめにする。青い実線及び破線は、時間積分に RK4 を用いて $\mathbf{x} = (1, 3, 5)$ 及び $(1.1, 3.3, 5.5)$ から積分したもので、橙色の実線及び破線は Euler 法で $\mathbf{x} = (1, 3, 5)$ 及び $(1.1, 3.3, 5.5)$ から積分したものである。点線は報告書に掲載されている同じ設定とよく一致している。RK4 を用いると、解は安定していて初期値の差が現れるのが遅い。

L63 free run



Lorenz-63 の接線型モデルは次のように書ける。

$$\begin{aligned}\delta\dot{X} &= -\sigma\delta X + \sigma\delta Y \\ \delta\dot{Y} &= -\delta X Z - X\delta Z + r\delta X - \delta Y \\ \delta\dot{Z} &= \delta X Y + X\delta Y - b\delta Z\end{aligned}$$

$$\begin{bmatrix} \delta\dot{X} \\ \delta\dot{Y} \\ \delta\dot{Z} \end{bmatrix} = \begin{bmatrix} -\sigma & \sigma & 0 \\ r-Z & -1 & -X \\ Y & X & -\beta \end{bmatrix} \begin{bmatrix} \delta X \\ \delta Y \\ \delta Z \end{bmatrix}$$

随伴モデルは行列を転置して次のように書ける。

$$\begin{bmatrix} \dot{X}^a \\ \dot{Y}^a \\ \dot{Z}^a \end{bmatrix} = \begin{bmatrix} X^a \\ Y^a \\ Z^a \end{bmatrix} + \begin{bmatrix} -\sigma & r-Z & Y \\ \sigma & -1 & X \\ 0 & -X & -\beta \end{bmatrix} \begin{bmatrix} \dot{X}^a \\ \dot{Y}^a \\ \dot{Z}^a \end{bmatrix}$$

ここで、 $(X^a, Y^a, Z^a)^T$ は随伴変数を表す。つまり

$$\begin{aligned}X^a &= X^a - \sigma\delta\dot{X}^a + (r-Z)\dot{Y}^a + Y\dot{Z}^a \\ Y^a &= Y^a + \sigma\dot{X}^a - \dot{Y}^a + X\dot{Z}^a \\ Z^a &= Z^a - X\dot{Y}^a - b\dot{Z}^a\end{aligned}$$

となる。

3.7 Lorenz-96 モデル

Lorenz-96 モデル (Lorenz and Emanuel 1998)

$$\frac{dX_i}{dt} = (X_{i+1} - X_{i-2})X_{i-1} - X_i + F \quad (3.22)$$

の TLM と ADM を作る。非線型項を $N(X) = (X_{i+1} - X_{i-2})X_{i-1}$ を線型化するには、掛け算の一方を順に摂動に置き換えるか、

$$\begin{aligned} \frac{\partial N}{\partial X_{i-2}} &= -X_{i-1} \\ \frac{\partial N}{\partial X_{i-1}} &= X_{i+1} - X_{i-2} \\ \frac{\partial N}{\partial X_{i+1}} &= X_{i-1} \end{aligned} \quad (3.23)$$

のように微分を計算すれば、次のように求められる。

$$\frac{d\delta X_i}{dt} = -X_{i-1}\delta X_{i-2} + (X_{i+1} - X_{i-2})\delta X_{i-1} + X_{i-1}\delta X_{i+1} - \delta X_i \quad (3.24)$$

(Equation 3.15) に基づいて、随伴モデルを作る。(Equation 3.24) の右辺が表す \mathbf{L} の随伴 \mathbf{L}^T は、 δX の添字が i となる行における係数を考えればよい。右辺第 1、2、3 項の δX_{i-2} , δX_{i-1} , δX_{i+1} なので、それぞれ +2, +1, -1 加えれば i 行目になる。さらに、左辺の変数 (時間微分) と右辺の変数を入れ替え $\mathbf{L}^T \dot{\mathbf{w}}^a$ が得られる。

L96 の随伴モデルを書き下すと

$$X_i^a = X_i^a - X_{i+1} \frac{dX_{i+2}^a}{dt} + (X_{i+2} - X_{i-1}) \frac{dX_{i+1}^a}{dt} + X_{i-2} \frac{dX_{i-1}^a}{dt} - \frac{dX_i^a}{dt} \quad (3.25)$$

及び

$$\frac{dX_i^a}{dt} = 0$$

となる。

3.8 拡散方程式

1 次元の拡散方程式

$$\frac{\partial u}{\partial t} = \sigma \frac{\partial^2 u}{\partial x^2}$$

を差分で離散化して、接線型モデルと随伴モデルを導出する (Kalnay 2003)。

時刻 (t , 添字 k) 方向に前方、空間 (x , 添字 i) 方向に中央差分すると

$$u_i^{k+1} = u_i^k + \alpha(u_{i+1}^k - 2u_i^k + u_{i-1}^k) \quad (3.26)$$

ここで $\alpha = \sigma \Delta t / \Delta x^2$ である。

拡散方程式は線型なので、 $u = u_b + \delta u$ のように基本場と摂動に分けると、接線型モデルは Equation 3.26 の変数を摂動に置き換えて次のように書ける。

$$\begin{aligned} \delta u_i^{k+1} &= \delta u_i^k + \alpha(\delta u_{i+1}^k - 2\delta u_i^k + \delta u_{i-1}^k) \\ &= (\alpha \quad 1 - 2\alpha \quad \alpha) \begin{pmatrix} \delta u_{i-1}^k \\ \delta u_i^k \\ \delta u_{i+1}^k \end{pmatrix} \end{aligned} \quad (3.27)$$

上式に現れる変数を列ベクトルで表し、行列ベクトル積で表すと次のようになる。

$$\begin{pmatrix} \delta u_{i-1}^k \\ \delta u_i^k \\ \delta u_{i+1}^k \\ \delta u_i^{k+1} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ \alpha & 1 - 2\alpha & \alpha & 0 \end{pmatrix} \begin{pmatrix} \delta u_{i-1}^k \\ \delta u_i^k \\ \delta u_{i+1}^k \\ \delta u_i^{k+1} \end{pmatrix} \quad (3.28)$$

摂動をベクトル $\delta \mathbf{u}^k$ で表すと

$$\begin{pmatrix} \delta \mathbf{u}^k \\ \delta \mathbf{u}^{k+1} \end{pmatrix} = \begin{pmatrix} \mathbf{I} & 0 \\ \mathbf{L} & 0 \end{pmatrix} \begin{pmatrix} \delta \mathbf{u}^k \\ \delta \mathbf{u}^{k+1} \end{pmatrix}$$

と表すことができる。ここで \mathbf{I} は単位行列で \mathbf{L} は、境界条件に合わせた（準）三重対角行列になる。例えば周期境界条件の場合

$$\mathbf{L} = \begin{pmatrix} 1 - 2\alpha & \alpha & 0 & \dots & 0 & 0 & \alpha \\ \alpha & 1 - 2\alpha & \alpha & 0 & \dots & 0 & 0 \\ 0 & \alpha & 1 - 2\alpha & \alpha & 0 & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \dots & 0 & \alpha & 1 - 2\alpha & \alpha & 0 \\ 0 & \dots & 0 & 0 & \alpha & 1 - 2\alpha & \alpha \\ \alpha & 0 & 0 & \dots & 0 & \alpha & 1 - 2\alpha \end{pmatrix}$$

Equation 3.28 を転置すると、随伴モデル

$$\begin{pmatrix} \delta^* u_{i-1}^k \\ \delta^* u_i^k \\ \delta^* u_{i+1}^k \\ \delta^* u_i^{k+1} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & \alpha \\ 0 & 1 & 0 & 1 - 2\alpha \\ 0 & 0 & 1 & \alpha \\ 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} \delta^* u_{i-1}^k \\ \delta^* u_i^k \\ \delta^* u_{i+1}^k \\ \delta^* u_i^{k+1} \end{pmatrix} \quad (3.29)$$

あるいは

$$\begin{pmatrix} \delta^* \mathbf{u}^k \\ \delta^* \mathbf{u}^{k+1} \end{pmatrix} = \begin{pmatrix} \mathbf{I} & \mathbf{L} \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \delta \mathbf{u}^k \\ \delta \mathbf{u}^{k+1} \end{pmatrix}$$

となる。この例では $\mathbf{L}^* = \mathbf{L}$ である。

Equation 3.29 を書き下すと

$$\begin{aligned}\delta^* u_{i-1}^k &= \delta^* u_{i-1}^k + \alpha \delta^* u_i^{k+1} \\ \delta^* u_i^k &= \delta^* u_i^k + (1 - 2\alpha) \delta^* u_i^{k+1} \\ \delta^* u_{i+1}^k &= \delta^* u_{i+1}^k + \alpha \delta^* u_i^{k+1} \\ \delta^* u_i^{k+1} &= 0\end{aligned}$$

となる。 $k+1$ における摂動の更新 (0 にする) は最後に行う。

u のフーリエ変換と逆変換を

$$\begin{aligned}u_m &= \frac{1}{\sqrt{2\pi}} \int_0^{2\pi} u(x) \exp(-imx) dx \\ u(x) &= \frac{1}{\sqrt{2\pi}} \sum_{-\infty}^{\infty} u_m \exp(imx)\end{aligned}$$

とすると、2階微分は次のように書ける。

$$\frac{\partial^2 u}{\partial x^2} = -\frac{m^2}{\sqrt{2\pi}} \sum_{-\infty}^{\infty} u_m \exp(imx)$$

拡散方程式は波数空間で次のように表される。

$$\mathbf{u}^{k+1} = \alpha \mathbf{u}^k$$

ここで、 $\alpha = 1 - \sigma m^2 \Delta t$ である。

接線型モデルは

$$\delta \mathbf{u}^{k+1} = \alpha \delta \mathbf{u}^k$$

$$\begin{pmatrix} \delta \mathbf{u}^k \\ \delta \mathbf{u}^{k+1} \end{pmatrix} = \begin{pmatrix} \mathbf{I} & 0 \\ \mathbf{L} & 0 \end{pmatrix} \begin{pmatrix} \delta \mathbf{u}^k \\ \delta \mathbf{u}^{k+1} \end{pmatrix}$$

となる。ここで $\mathbf{L} = \alpha \mathbf{I}$ である。

随伴モデルは

$$\begin{pmatrix} \delta^* \mathbf{u}^k \\ \delta^* \mathbf{u}^{k+1} \end{pmatrix} = \begin{pmatrix} \mathbf{I} & \mathbf{L} \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \delta^* \mathbf{u}^k \\ \delta^* \mathbf{u}^{k+1} \end{pmatrix}$$

すなわち

$$\begin{aligned}\delta^* \mathbf{u}^k &= (\mathbf{I} + \mathbf{L}) \delta^* \mathbf{u}^{k+1} = (1 + \alpha) \mathbf{I} \delta^* \mathbf{u}^{k+1} \\ \delta^* \mathbf{u}^{k+1} &= 0\end{aligned}$$

となる。

3.9 4 次の Runge-Kutta 法

$$\dot{y} = f(t, y) \quad (3.30)$$

を 4 次の Runge-Kutta (RK4) 法で積分する。

$$\begin{aligned} y_{n+1} &= y_n + \frac{h}{6}(k_1 + 2k_2 + 2k_3 + k_4) \\ k_1 &= f(t_n, y_n) \\ k_2 &= f\left(t_n + \frac{h}{2}, y_n + \frac{h}{2}k_1\right) \\ k_3 &= f\left(t_n + \frac{h}{2}, y_n + \frac{h}{2}k_2\right) \\ k_4 &= f(t_n + h, y_n + hk_3) \end{aligned} \quad (3.31)$$

k の係数は Butcher の表 (Butcher and Wanner 1996)

$$\begin{array}{c|c} \mathbf{c} & \mathbf{A} \\ \hline & \mathbf{b}^T \end{array} \quad (3.32)$$

に整理できる。 s 次の陽的 Runge-Kutta 法は

$$\begin{aligned} y_{n+1} &= y_n + h \sum_{i=1}^s b_i k_i \\ k_i &= f\left(t_n + c_i h, y_n + h \sum_{j=1}^{i-1} a_{ij} k_j\right) \end{aligned}$$

と書ける。 \mathbf{A} は対角成分が 0 である下三角行列になる。

RK4 に対する Butcher の表は次のようになる。

$$\begin{array}{c|ccc} 0 & & & \\ 1/2 & 1/2 & & \\ 1/2 & 0 & 1/2 & \\ 1 & 0 & 0 & 1 \\ \hline & 1/6 & 1/3 & 1/3 & 1/6 \end{array} \quad (3.33)$$

RK4 の接線型は次の通りである。

$$\begin{aligned} \delta y_{n+1} &= \delta y_n + \frac{h}{6}(\delta k_1 + 2\delta k_2 + 2\delta k_3 + \delta k_4) \\ \delta k_1 &= \delta f(t_n, y_n, \delta y_n) \\ \delta k_2 &= \delta f\left(t_n + \frac{h}{2}, y_n + \frac{h}{2}k_1, \delta y_n + \frac{h}{2}\delta k_1\right) \\ \delta k_3 &= \delta f\left(t_n + \frac{h}{2}, y_n + \frac{h}{2}k_2, \delta y_n + \frac{h}{2}\delta k_2\right) \\ \delta k_4 &= \delta f(t_n + h, y_n + hk_3, \delta y_n + h\delta k_3) \end{aligned} \quad (3.34)$$

ここで $\delta f = \partial f / \partial y$ は接線型モデルである。 $\delta k_1, \delta k_2, \delta k_3, \delta k_4$ は出力しないので、 δy を更新した後

$$\delta \mathbf{k} \equiv [\delta k_1, \delta k_2, \delta k_3, \delta k_4]^T = 0$$

と置くと、行列形式では

$$\begin{bmatrix} \delta \mathbf{k} \\ \delta y \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ h\mathbf{b}^T & 1 \end{bmatrix} \begin{bmatrix} \delta \mathbf{k} \\ \delta y \end{bmatrix} \quad (3.35)$$

となる。

随伴は

$$\begin{bmatrix} \mathbf{k}^a \\ y^a \end{bmatrix} = \begin{bmatrix} 0 & h\mathbf{b} \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \mathbf{k}^a \\ y^a \end{bmatrix} \quad (3.36)$$

つまり、 $k_1^a = hy^a/6, k_2^a = hy^a/3, k_3^a = hy^a/3, k_4^a = hy^a/6$ と初期化される。

L を線型演算子、 α をスカラ係数として $C = L(A + \alpha B)$ は

$$\begin{bmatrix} A \\ B \\ C \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ L & \alpha L & 0 \end{bmatrix} \begin{bmatrix} A \\ B \\ C \end{bmatrix}$$

と書けるので、その随伴は

$$\begin{bmatrix} A^a \\ B^a \\ C^a \end{bmatrix} = \begin{bmatrix} 1 & 0 & L^T \\ 0 & 1 & \alpha L^T \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} A^a \\ B^a \\ C^a \end{bmatrix}$$

つまり $A^a = A^a + LC^a, B^a = B^a + \alpha LC^a, C^a = 0$ である。これを用いると RK4 の随伴は次の順序で計算される。

$$f_4^a \equiv f^a(t_n + h, y_n + hk_3, k_4^a)$$

$$y^a = y^a + f_4^a$$

$$k_3^a = k_3^a + hf_4^a$$

$$k_4^a = 0$$

$$f_3^a \equiv f^a\left(t_n + \frac{h}{2}, y_n + \frac{h}{2}k_2, k_3^a\right)$$

$$y^a = y^a + f_3^a$$

$$k_2^a = k_2^a + \frac{h}{2}f_3^a$$

$$k_3^a = 0$$

$$f_2^a \equiv \delta f^a\left(t_n + \frac{h}{2}, y_n + \frac{h}{2}k_1, k_2^a\right)$$

$$y^a = y^a + f_2^a$$

$$k_1^a = k_1^a + \frac{h}{2}f_2^a$$

$$k_2^a = 0$$

$$y^a = y^a + f^a(t_n, y_n, k_1^a)$$

$$k_1^a = 0$$

3.10 接線型及び随伴の動作確認

接線型は

$$N(\mathbf{w} + \Delta\mathbf{w}) - N(\mathbf{w}) \approx \mathbf{L}(\mathbf{w})\Delta\mathbf{w} \quad (3.37)$$

が近似的に成り立つことを確認する。ここで N は非線型モデル、 \mathbf{L} は接線型モデル、 \mathbf{w} は状態、 $\Delta\mathbf{w}$ は摂動を表す。精度は \mathbf{w} における N の非線型の強さや摂動 $\Delta\mathbf{w}$ の振幅に依存する。

随伴は

$$(\mathbf{L}(\mathbf{w})\Delta\mathbf{w})^T(\mathbf{L}(\mathbf{w})\Delta\mathbf{w}) = \Delta\mathbf{w}^T[\mathbf{L}(\mathbf{w})^T(\mathbf{L}(\mathbf{w})\Delta\mathbf{w})] \quad (3.38)$$

が成り立つことを確認する。厳密に成り立つか、機械精度以下の誤差（下 1~2 桁の差異）でなければバグが存在していることを示す。

第 4 章

アンサンブルカルマンフィルタ

この章では、アンサンブルカルマンフィルタについて述べる。

4.1 カルマンフィルタ

Section 1.6 で求めた解析の式は、

$$\mathbf{x}^a = \mathbf{x}^b + \mathbf{K}(\mathbf{y}^o - \mathbf{H}\mathbf{x}^b)$$

と書くことができる。ここで

$$\mathbf{K} = \mathbf{B}\mathbf{H}^T(\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R})^{-1} \quad (4.1)$$

はカルマンゲインである。

カルマンゲインを用いると、解析誤差共分散 Section 1.6 は

$$\mathbf{P}_a = (\mathbf{I} - \mathbf{K}\mathbf{H})\mathbf{B}(\mathbf{I} - \mathbf{K}\mathbf{H})^T + \mathbf{K}\mathbf{R}\mathbf{K}^T = (\mathbf{I} - \mathbf{K}\mathbf{H})\mathbf{B} \quad (4.2)$$

となり、背景誤差共分散よりも縮小する。

4.2 アンサンブルカルマンフィルタ

アンサンブルカルマンフィルタ (EnKF: Ensemble Kalman Filter Evensen (1994)) では背景誤差共分散 \mathbf{B} をモデル予報のアンサンブルからの標本共分散 \mathbf{P}_b で近似する。

$$\begin{aligned}\bar{\mathbf{x}}^b &= \frac{1}{k} \sum_{i=1}^k \mathbf{x}_i^b \\ \mathbf{P}^b &= \frac{1}{k-1} \sum_{i=1}^k (\mathbf{x}_i^b - \bar{\mathbf{x}}^b)(\mathbf{x}_i^b - \bar{\mathbf{x}}^b)^T\end{aligned}\quad (4.3)$$

同様に解析値の平均と誤差共分散は次の式で定義される。

$$\begin{aligned}\bar{\mathbf{x}}^a &= \frac{1}{k} \sum_{i=1}^k \mathbf{x}_i^a \\ \mathbf{P}_a &= \frac{1}{k-1} \sum_{i=1}^k (\mathbf{x}_i^a - \bar{\mathbf{x}}^a)(\mathbf{x}_i^a - \bar{\mathbf{x}}^a)^T\end{aligned}\quad (4.4)$$

EnKF の更新式は次のように書ける。

$$\begin{aligned}\bar{\mathbf{x}}^a &= \bar{\mathbf{x}}^b + \mathbf{K}(\bar{\mathbf{y}}^o - \mathbf{H}\bar{\mathbf{x}}^b) \\ \mathbf{x}'^a &= \mathbf{x}'^b + \tilde{\mathbf{K}}(\mathbf{y}'^o - \mathbf{H}\mathbf{x}'^b)\end{aligned}\quad (4.5)$$

ここで、 \mathbf{y}'^o は観測誤差の確率密度分布からランダムに選んだ摂動、 $\tilde{\mathbf{K}}$ は偏差の更新に用いるカルマンゲインである。EnKF では $\tilde{\mathbf{K}} = \mathbf{K}$ を使い、 (\mathbf{y}'^o) は観測誤差の確率密度分布からランダムに抽出する。これを摂動観測法 (PO: Perturbed Observations, Burgers et al. 1998; Houtekamer and Mitchell 1998) と呼ぶ (Section 4.5)。

全てのアンサンブルメンバーが同じ観測を同化すると $\mathbf{y}'^o = 0$ となるので解析誤差共分散は Equation 4.5 より

$$\mathbf{P}_a = (\mathbf{I} - \mathbf{K}\mathbf{H})\mathbf{P}_b(\mathbf{I} - \mathbf{K}\mathbf{H})^T \quad (4.6)$$

となる。

Equation 4.2 と比較すると Equation 4.6 は、 $\mathbf{K}\mathbf{R}\mathbf{K}^T$ の項がなく、 $\mathbf{I} - \mathbf{K}\mathbf{H}$ 倍、解析誤差共分散が過小評価されている。

観測誤差が $\mathbf{y}'^o \neq 0$ の場合、解析誤差共分散は次のようになる。

$$\begin{aligned}\mathbf{P}_a &= (\mathbf{I} - \mathbf{K}\mathbf{H})\mathbf{P}_b(\mathbf{I} - \mathbf{K}\mathbf{H})^T \\ &\quad + \mathbf{K} \left(\overline{\mathbf{y}'\mathbf{y}'^T} - \overline{\mathbf{H}\mathbf{x}'^b\mathbf{y}'^T} - \overline{\mathbf{y}'\mathbf{x}'^b\mathbf{H}^T} \right) \mathbf{K}^T \\ &\quad + \overline{\mathbf{x}'^b\mathbf{y}'^T} \mathbf{K}^T + \mathbf{K}\overline{\mathbf{y}'\mathbf{x}'^b\mathbf{H}^T}\end{aligned}\quad (4.7)$$

観測ノイズが $\mathbf{R} = E(\overline{\mathbf{y}'\mathbf{y}'^T})$ で与えられ、 $E(\overline{\mathbf{x}'^b\mathbf{y}'^T}) = 0$ ならば \mathbf{P}_a の期待値は Equation 4.2 と等しくなる。

4.3 解析誤差共分散推定に対する標本誤差の影響

有限のアンサンブル数では、背景誤差共分散の推定において標本誤差が生じる。観測誤差に摂動を与えると、観測誤差共分散にも標本誤差が含まれる。これらの誤差が解析誤差共分散にどのように影響を与えるか簡単な実験で示す (Whitaker and Hamill 2002)。摂動観測法ありとなし（解析誤差共分散を Equation 4.2 で計算）とを比較する。

- 状態ベクトル \mathbf{x}^b 及び観測 \mathbf{y}
- \mathbf{x}^b と \mathbf{y} とは同じ量つまり $\mathbf{H} = 1$
- 試行回数 1,000,000 回
- アンサンブルメンバー数 5
- \mathbf{x}^b 正規分布 $N(0, 1)$ からランダムに選ぶ。
- 背景誤差分散は自由度 4 の χ^2 分布で平均は 1
- 観測誤差分散 1 の EnKF では観測を $N(0, 1)$ から選択し平均 0 分散 1 となるように調整する。

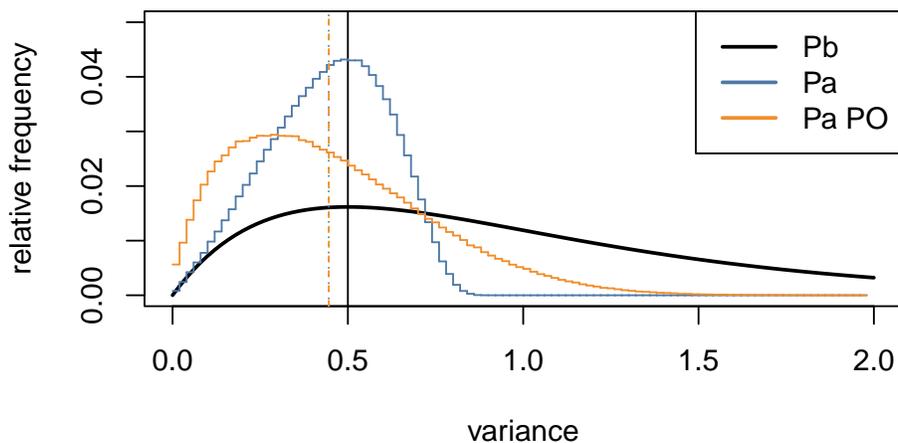


Figure 4.1: 背景及び解析誤差共分散の確率密度分布

背景誤差共分散 (Figure 4.1 黒) は χ^2 分布に従う。観測を同化すると、小さい値の振幅が大きくなり分散は小さくなる (青)。摂動観測法の解析誤差分散 (橙) は小さい値の確率が高い。

Equation 4.1 と Equation 4.2 より解析誤差共分散は背景誤差共分散に非線型に依存する

ことが分かる。

$$\mathbf{P}_a = (1 - \mathbf{K})\mathbf{P}_b = \left(1 - \frac{\mathbf{P}_b}{\mathbf{P}_b + \mathbf{R}}\right) \mathbf{P}_b$$

観測に対する摂動の有無は関係なく、標本誤差のため \mathbf{P}^b は確率変数であるため、 \mathbf{P}^a も確率変数である。 $\mathbf{P}_b = 1$ に対する \mathbf{P}_a の期待値は 0.5 であるが、 \mathbf{P}_a の平均は約 0.44 である。 \mathbf{P}_b にバイアスがなくても、 \mathbf{P}_a にバイアスが生じるのは同系交配 (inbreeding) の影響である (Houtekamer and Mitchell 1998)。

非線型性に起因す誤差分散の過小バイアスは、アンサンブル数を増やすと減少する。 \mathbf{P}^b が平均の周りに狭く対称に分布し、カルマンゲインで非線型変換された \mathbf{P}^a も対称に近づくためである。

観測摂動法 EnKF に対しては、Equation 4.7 で与えられる解析誤差分散 \mathbf{P}_{PO}^a は、この実験において以下のように簡単になる。

$$\mathbf{P}_{PO}^a = \underbrace{(1 - \mathbf{K})^2 \mathbf{P}^b}_{\zeta_1} + \underbrace{\mathbf{K}^2 \mathbf{R}}_{\zeta_2} + \underbrace{2\mathbf{K}\mathbf{y}'\mathbf{x}'^b(1 - \mathbf{K})}_{\zeta_3}$$

\mathbf{P}_{PO}^a の平均は \mathbf{P}^a と同じになるが、標本誤差の影響で分布は幅広く、誤差は大きくなる。

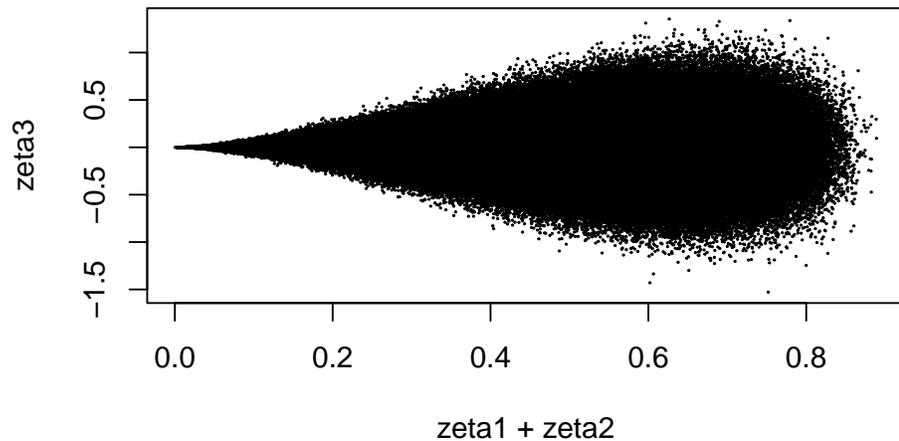


Figure4.2: 摂動観測法における背景と観測の偽相関

$\zeta_1 + \zeta_2$ が大きくなると ζ_3 の分散が大きくなる。ノイズ ζ_3 は $\zeta_1 + \zeta_2$ が大きくなるほど分布を両側に広げるために、 \mathbf{P}_{PO}^a の分布が左に歪む。

4.4 単一観測の同化

個々の観測が独立であれば、 \mathbf{R} は対角で、一要素ずつ順番に同化しても結果は変わらない (Anderson 2003)。これを利用して、演算を簡略化し、同化を 2 段階で行う。

1. 単一の測定値をアンサンブル予報から推定した観測変数に同化し、インクリメント (変化量) Δy_i を求める。ここで i はアンサンブルの番号 $i = 1, \dots, N$ 。
2. 各状態変数の各アンサンブル標本について、対応するインクリメントを線型回帰で求める。状態変数及び観測変数との間に、先験関係がガウス分布であることを仮定し、最小二乗法で直線を当てはめている。非線型であることも考えられる観測演算子に対して、線型化を施し逆関数を求めていることに相当する。

y の変化に対する x の変化の回帰は次のように書ける。

$$\Delta x = \frac{\text{cov}(x, y)}{\text{var}(y)} \Delta y \quad (4.8)$$

長さ n の状態ベクトル \mathbf{x} 及び、それから求めた長さ m の観測相当量 $\mathbf{y} = h(\mathbf{x})$ を結合した長さ $k + m$ のベクトル

$$\mathbf{z} = \begin{bmatrix} \mathbf{x} \\ h(\mathbf{x}) \end{bmatrix}$$

を考える。これを結合状態・観測空間とよぶ。その共分散は

$$\Sigma_a = [\Sigma_f^{-1} + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H}]^{-1} \quad (4.9)$$

アンサンブル平均は

$$\bar{\mathbf{z}}^a = \Sigma_a [\Sigma_f^{-1} \bar{\mathbf{z}}^f + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{y}^o] \quad (4.10)$$

と表される。

1 個のスカラ観測に対しては、 Σ_f , Σ_a の左上から $n \times n$ の部分は、それぞれ \mathbf{P}_f , \mathbf{P}_a であり、 k 列目は変数間の共分散 \mathbf{s}^f , \mathbf{s}^a , $\mathbf{s}_k^f = s^f$, $\mathbf{s}_k^a = s^a$ は $\mathbf{y}^f = h(\mathbf{x}^f)$, $\mathbf{y}^a = h(\mathbf{x}^a)$ の分散を表す。線型演算子 $\mathbf{H} = [0 \quad 0 \quad \dots \quad 1]$ は最後の要素 y を取り出す $1 \times k$ の行列である。観測誤差共分散行列 \mathbf{R} はスカラ分散 r で表される。

式 (Equation 4.9), (Equation 4.10) はそれぞれ、

$$\Sigma_a = \left(\mathbf{I} - \frac{1}{r + s^f} \Sigma_f^{0k} \right) \Sigma_f, \quad (4.11)$$

$$\Delta \bar{\mathbf{z}} = \mathbf{z}^a - \mathbf{z}^f = \frac{y^o - \bar{y}^f}{r + s^f} \mathbf{s}^f \quad (4.12)$$

と簡単になる。ここで

$$\Sigma_f^{0k} = \Sigma_f \mathbf{H}^T \mathbf{H}$$

は、 Σ_f の 1 から n 列目を 0 とし、 k 列目だけを残した行列である。観測変数の分散は

$$s^a = \frac{1}{1/s^f + 1/r} = \frac{s^f r}{s^f + r} \quad (4.13)$$

解析値は

$$\bar{y}^a = s^a (\bar{y}^f / s^f + y^o / r) = \frac{1}{s^f + r} (r \bar{y}^f + s^f y^o) \quad (4.14)$$

観測変数のインクリメントは

$$\Delta \bar{y} = \bar{y}^a - \bar{y}^f = \frac{s^f}{s^f + r} (y^o - \bar{y}^f) \quad (4.15)$$

となる。一方、状態変数は

$$\Delta \bar{x} = \bar{x}^a - \bar{x}^f = \frac{[s_1^f, \dots, s_n^f]^T}{s^f + r} (y^o - \bar{y}^f) = \frac{[s_1^f, \dots, s_n^f]^T}{s^f + r} \Delta \bar{y} \quad (4.16)$$

となる。これは、それぞれの状態変数に対する線型回帰 (Equation 4.8) に他ならない。

4.5 摂動法

確率的 (stochastic) もしくはモンテカルロ法に基づくアンサンブルカルマンフィルタ (Evensen 1994; Houtekamer and Mitchell 1998; Burgers et al. 1998) では、観測に摂動を加えるため摂動法 (PO: perturbed observation) と呼ばれている。Vetra-Carvalho et al. (2018) は、観測 y^o には測定誤差があり、それに伴う確率分布を持っているので、予測の観測推定量 \mathbf{y} に摂動を加えると解釈する方が適切であるとしているが、ここでは前者に従い、平均と分散が y^o, r となる正規分布で誤差を与える。

$$\mathbf{y}^o \sim N(y^o, r)$$

各メンバーに対して線型回帰 (Equation 4.16) を適用して解析値を求める。

4.6 アンサンブル平方根フィルタ

観測に摂動を与えることに伴う標本誤差 (Whitaker and Hamill 2002) を回避するアンサンブル平方根フィルタ (EnSRF: ensemble square-root filter) が複数考案された (Anderson 2001; Bishop et al. 2001; Whitaker and Hamill 2002; Tippett et al. 2003)。ここでは、スカラー観測を逐次に同化するアンサンブル調節カルマンフィルタ (EAKF: ensemble adjustment filter, Anderson 2003) について述べる。

アンサンブル平均からの偏差を δ で表す。観測変数の予報分散は s^f 、解析分散は (Equation 4.13) なので、その比

$$\alpha = \sqrt{\frac{r}{r + s^f}} \quad (4.17)$$

を用いると、メンバー i の解析インクリメントは

$$\delta \mathbf{y}_i = \delta \mathbf{y}_i^a - \delta \mathbf{y}_i^f = (\alpha - 1) \delta \mathbf{y}_i^f \quad (4.18)$$

となる。摂動のインクリメントは、線型回帰により

$$\Delta \mathbf{z}_i = \frac{\mathbf{s}^f}{s^f} \Delta \mathbf{y}_i = \frac{\mathbf{s}^f}{s^f} (\alpha - 1) \delta \mathbf{y}_i^f \quad (4.19)$$

平均のインクリメント (Equation 4.12) を加えてアンサンブルを更新する。

$$\mathbf{z}_i^a = \mathbf{z}_i^f + \frac{\mathbf{s}^f}{r + s^f} (y^o - \bar{y}^f) + \frac{\mathbf{s}^f}{s^f} (\alpha - 1) (\mathbf{y}_i^f - \bar{y}^f) \quad (4.20)$$

4.7 決定論的アンサンブルカルマンフィルタ

通常、背景場に対する修正が微小である。このことに着目し平方根フィルタを線型近似したものが、決定論的アンサンブルカルマンフィルタ (DEnKF: deterministic EnKF, Sakov and Oke 2008) である。

\mathbf{KH} が小さいとき、アンサンブルカルマンフィルタの解析誤差共分散 Equation 4.6 は次のように書ける。

$$\mathbf{P}_a = \mathbf{P}_b - 2\mathbf{KHP}_b + \mathbf{KHP}_b \mathbf{H}^T \mathbf{K}^T \approx (\mathbf{I} - 2\mathbf{KH}) \mathbf{P}_b \quad (4.21)$$

ここで

$$\mathbf{P}_b \mathbf{H}^T \mathbf{K}^T = \mathbf{P}_b \mathbf{H}^T (\mathbf{HP}_b \mathbf{H}^T + \mathbf{R})^{-1} \mathbf{HP}_b = \mathbf{KHP}_b$$

を用いた。

Equation 4.21 は、カルマンゲインを半分にすると理想的な解析誤差共分散 Equation 4.2 に近づくことを示唆している。

アンサンブル偏差を

$$\delta \mathbf{x}_i = \mathbf{x}_i - \bar{\mathbf{x}} \quad (4.22)$$

と表す。

DEnKF の計算手順は次のとおりである。

1. 与えられたアンサンブル予測に対して、Equation 4.3 の最初の式を用いて、アンサンブル平均 $\bar{\mathbf{x}}_b$ とアンサンブル偏差 $\delta \mathbf{x}^b$ Equation 4.22 を計算する。
2. Equation 4.5 の最初の式でアンサンブル平均に対して $\bar{\mathbf{x}}^a$ を解析値を求める。
3. 解析のアンサンブル偏差は、Equation 4.5 の二番目の式で $\tilde{\mathbf{K}} = \mathbf{K}/2$ として計算する。

$$\delta \mathbf{x}^a = \delta \mathbf{x}^b - \frac{1}{2} \mathbf{K} \mathbf{H} \delta \mathbf{x}^b \quad (4.23)$$

4. アンサンブル偏差に解析値を加えて、アンサンブルの解析値 $\mathbf{x}_i^a = \delta \mathbf{x}_i^a + \bar{\mathbf{x}}_a$ を計算する。

この手法では、半分のカルマンゲインを用いたカルマンフィルタの解析方程式を各アンサンブルメンバに適用し、アンサンブル平均に対する解析を加えている。EnSRF 同様に摂動観測を用いていないので、決定論的フィルタである。

Equation 4.23 は、EnSRF のアンサンブル偏差の線型近似となっている。Equation 4.23 を用いて解析誤差共分散を求めると

$$\mathbf{P}_a = (\mathbf{I} - \mathbf{K} \mathbf{H}) \mathbf{P}_b + \frac{1}{4} \mathbf{K} \mathbf{H} \mathbf{P}_b \mathbf{H}^T \mathbf{K}^T \quad (4.24)$$

となる。 $\mathbf{K} \mathbf{H}$ が小さいとき、理想的な解析誤差共分散 Equation 4.2 の良い近似となっている。 $\mathbf{K} \mathbf{H}$ が大きいときは、解析誤差共分散が Equation 4.24 の右辺第 2 項の分だけ大きくなるので、適応インフレーションとして作用する。

Equation 4.23 は

$$\delta \mathbf{x}^a = \frac{1}{2} \delta \mathbf{x}^b + \frac{1}{2} (\delta \mathbf{x}^b - \mathbf{K} \mathbf{H} \delta \mathbf{x}^b) \quad (4.25)$$

と表すことができる。Equation 4.25 の右辺右辺第 1、2 項は、それぞれアンサンブル偏差の背景値 (先験値)、解析値なので、両者の平均をとっている。すなわち、DEnKF は重み 1/2 の先験偏差への緩和 (RTPP: relaxation-to-prior-perturbation Zhang et al. 2004) に相当する (Bowler et al. 2013)。

4.8 準地衡流モデル

Sakov and Oke (2008) を追試した 榎本 and 中下 (2022) の概要を紹介する。

2次元準地衡流 (quasi-geostrophic, QG) 方程式は中緯度における大気や海洋の大規模な流れの特徴を表すことができるため、計算量を抑えつつトイモデルよりも現実的なデータ同化の問題を検討することができる。

ここでは中緯度海洋における表層流を模した流れに対するデータ実験を行う。

QG 方程式は次の式で表される。

$$\frac{\partial q}{\partial t} = -\frac{\partial \psi}{\partial x} - \varepsilon J(\psi, q) - A\Delta^3 \psi + \nabla \times \tau \quad (4.26)$$

t, x, y ははそれぞれ時刻と東西及び南北座標, ψ は流線函数, $J(p, q)$ はヤコビアン, $\Delta = \partial^2/\partial x^2 + \partial^2/\partial y^2$ はラプラシアンを表す。

渦位と流線函数は

$$q = \Delta \psi - F\psi \quad (4.27)$$

で関係付けられる。

理想的な風応力 $\tau = -\cos(2\pi y)$ を考えると、そのカールは

$$\nabla \times \tau = -2\pi \sin(2\pi y) \quad (4.28)$$

となる。

モデルの設定と計算手法の概要は次通りである。正方形領域 $0 \leq x \leq 1, 0 \leq y \leq 1$ を 129×129 点に離散化し、境界条件は $\psi = \Delta \psi = \Delta^2 \psi = 0$ とする。内部領域のみ4次の Runge-Kutta 法を用いて、時間ステップ $\Delta t = 1.25$ で時間発展させる。Equation 4.27 の右辺第1項は、惑星渦度を表し中央差分を用いる。右辺第2項は、(Arakawa 1966) ヤコビアンを用いて非線型不安定を抑制する。ポワソン方程式 Equation 4.27 は多重格子法を用いて解く。標準パラメタは $\varepsilon = 10^{-5}, A = 2 \times 10^{-12}, F = 1600$ とする。

データ同化実験の一例を示す。長期積分から25メンバをランダムに選択して、アンサンブル予報の初期値とした。観測は真値に分散 $\sigma^2 = 4.0$ ガウス型のランダム誤差を付加した ψ とし、4ステップ毎に $m = 300$ 個を同化した。観測網は海面高度計を模したものとする。

サンプル数不足による偽相関を抑制するため、共分散行列に局所化を適用する。DEnKFの特徴はETKF等とは異なり、予報共分散行列に対してSchur積（要素毎の積）による局所化が容易に行えることである。局所化関数は、局所化半径 $r_0 = 15$ 格子のガウス型とする。共分散膨張係数は $\alpha = 1.06$ とする。

実験の結果を以下に示す。境界の渦やその周辺から観測により誤差が減少し観測誤差を下回るが、渦の活動が活発な西側の南北の中央付近では最終時刻でも観測誤差程度の誤差が残る。予報誤差標準偏差の大きな領域ではインクリメントが入りやすいことも分かる。

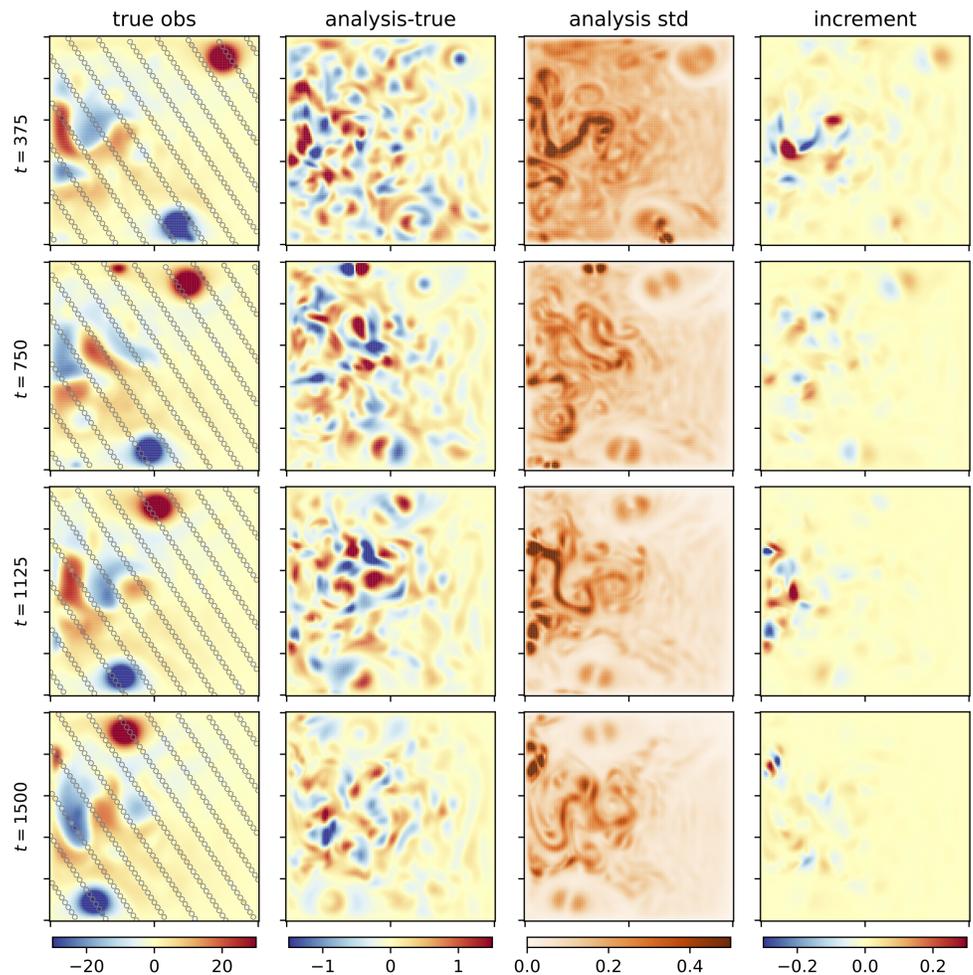


Figure4.3: 真の流線函数、解析誤差、解析標準偏差、インクリメント（右から左）の $t = 375, 750, 1125, 1500$ （データ同化の75, 150, 225, 300サイクル、上から下）におけるスナップショット。真値の上に観測位置を \circ で示す。

領域平均では、データ同化の初期に解析誤差は急減して観測誤差を下回る。解析標準偏差は解析誤差を下回っており、予報誤差共分散を過小評価が示唆される。渦位と流線函数を予報した場合も、渦位のみを予報した場合も同一の結果であり、多重格子法により流線函

数が精度良く推定できている。流線函数のみ予報し、渦位を流線函数から診断すると、渦の振幅が小さくなり適切に同化ができなかった。局所化を用いない場合は、フィルタは発散する。

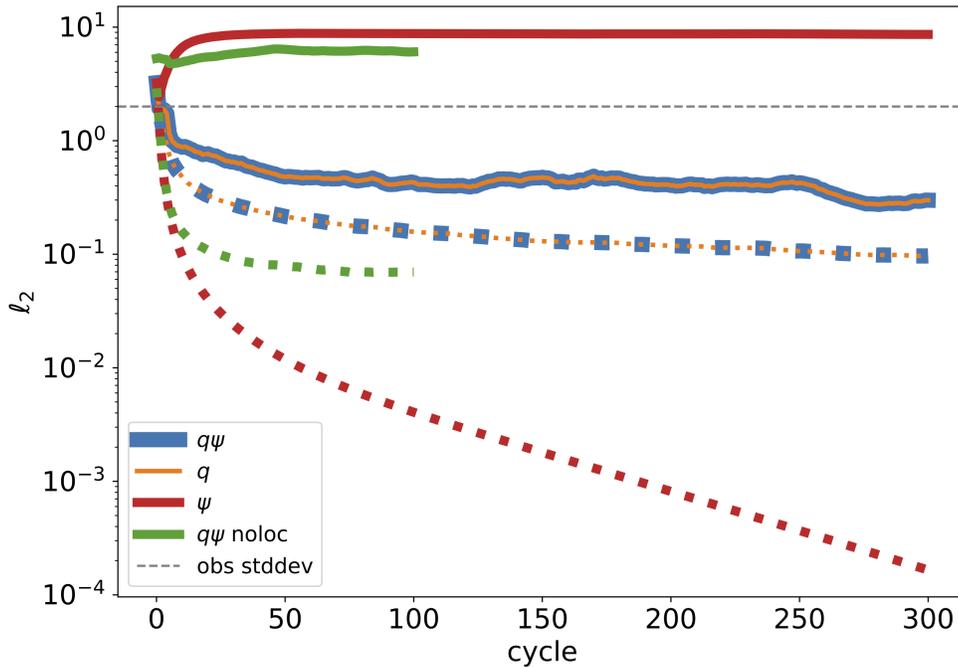


Figure4.4: 準地衡モデルを用いた同化実験における二乗平均平方根誤差（実線）及び解析標準偏差（点線） q, ψ を予報（青）、 q のみ（橙）、 ψ のみ（赤）、 q, ψ 局所化なし（緑、100サイクルで中断）。灰色の破線は与えた観測誤差

🔥 [@Enomoto-Nakashita:2022] の誤植

- p. 131 右段 9 行目: Fig. 7

第 5 章

感度解析

入力パラメタに対する出力パラメタの感度の問題には以下のようなものがある。(Tala-grand 1991).

感度実験 (sensitivity experiments) 入力パラメタの変化に対する系の物理応答を調べる。
不確実性の定量化 (Uncertainty quantification) 入力パラメタの不確実性により生じる出力パラメタの不確実性を測る。

感度解析 (sensitivity analysis) モデルの入力パラメタの何が、出力パラメタにおいて後験的に観測された、特徴の源であったのかを特定する。

パラメタ推定 (parameter estimation) 及びデータ同化 (data assimilation) 出力パラメタが与えた条件を満たすような入力パラメタの値を求める。

予報感度解析は予報に含まれる特徴の源を見つける手法で、ターゲット観測に応用されてきた (Snyder 1996; Lorenz and Emanuel 1998; Bishop and Toth 1999)。感度は、予報モデルの随伴 (Rabier et al. 1996; Langland et al. 2000) やアンサンブル予報 (Ansell and Hakim 2007; Enomoto et al. 2007) から計算できるが、ラグランジュの未定乗数法を用いると一貫した導出が可能である (Enomoto et al. 2015)。

5.1 摂動の線型成長

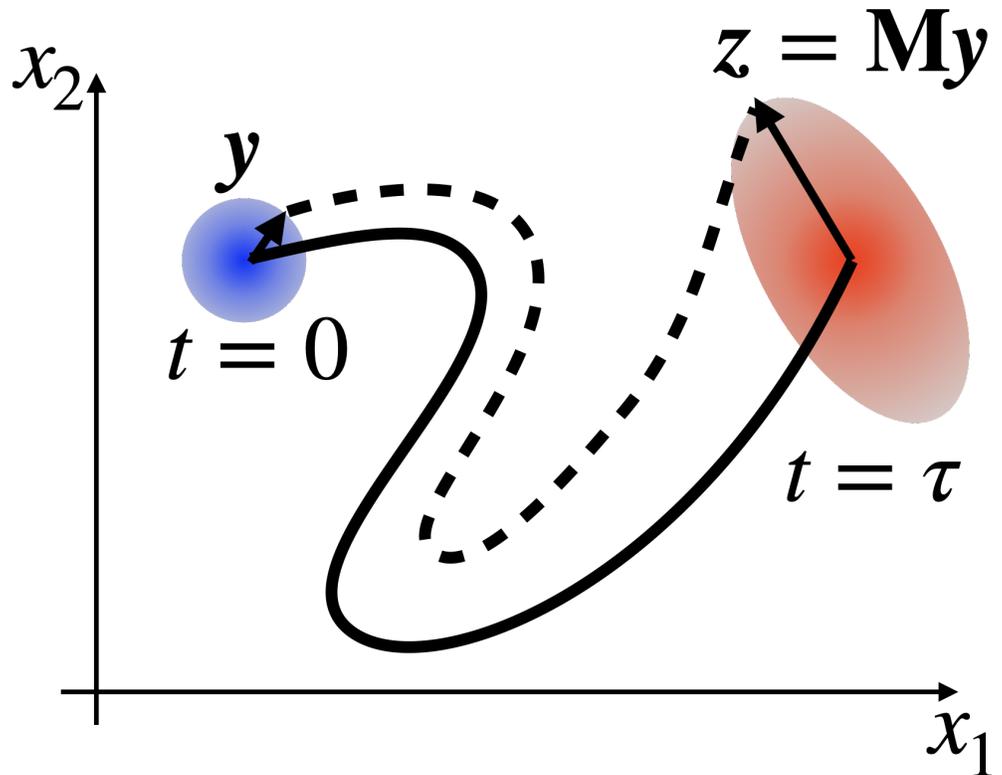
初期 $t = 0$ における状態 \mathbf{x}_0 からの非線型時間発展を考える。

$$\mathbf{x}_\tau = M(\mathbf{x}_0) \quad (5.1)$$

初期摂動 \mathbf{y} 次の式に従って線型に成長するものと仮定し、時刻 τ での摂動を \mathbf{z} で表す。

$$\mathbf{z} = M(\mathbf{x}_0 + \mathbf{y}) - M(\mathbf{x}_0) \approx \mathbf{M}\mathbf{y} \quad (5.2)$$

ここで $M = \partial M / \partial \mathbf{x}$ は接線型モデルである。



5.2 随伴感度

$\|\mathbf{y}\| \equiv \sqrt{\mathbf{y}^T \mathbf{G}_0 \mathbf{y}} = 1$ という条件の下で、 δJ を最大にする \mathbf{y} を求める。

$$F(\mathbf{y}, \lambda) = \delta J + \lambda(1 - \mathbf{y}^T \mathbf{G}_0 \mathbf{y})$$

を最小化する。ここで、 λ はラグランジュの未定乗数で δJ は

$$\delta J = \left[\frac{\partial J}{\partial \mathbf{x}_\tau} \right]^T \mathbf{z} = \left[\frac{\partial J}{\partial \mathbf{x}_\tau} \right]^T \mathbf{M} \mathbf{y} = \left[\mathbf{M}^T \frac{\partial J}{\partial \mathbf{x}_\tau} \right]^T \mathbf{y} = \left[\frac{\partial J}{\partial \mathbf{x}_0} \right]^T \mathbf{y} \quad (5.3)$$

で表される。

つまり、

$$F(\mathbf{y}, \lambda) = \left[\frac{\partial J}{\partial \mathbf{x}_0} \right]^T \mathbf{y} + \lambda(1 - \mathbf{y}^T \mathbf{G}_0 \mathbf{y})$$

$\partial F / \partial \mathbf{y} = 0$ を計算すると

$$2\lambda\mathbf{y} = \mathbf{G}_0^{-1}\mathbf{M}^T \frac{\partial J}{\partial \mathbf{x}_\tau} = \mathbf{G}_0^{-1} \frac{\partial J}{\partial \mathbf{x}_0}. \quad (5.4)$$

が得られる。

5.3 特異ベクトル感度

特異ベクトル (SV: singular vector) 感度は $\|\mathbf{y}\| = 1$ の条件の下で $\|\mathbf{z}\| \equiv \sqrt{\mathbf{z}^T \mathbf{G}_\tau \mathbf{z}}$ を最大化する \mathbf{y} を求めるには、

$$F(\mathbf{y}, \lambda) = \mathbf{z}^T \mathbf{G}_\tau \mathbf{z} + \lambda(1 - \mathbf{y}^T \mathbf{G}_0 \mathbf{y})$$

を最小化する。 $\partial F / \partial \mathbf{y} = 0$ を計算すると一般化固有値問題

$$\mathbf{M}^T \mathbf{G}_\tau \mathbf{M} \mathbf{y} = \lambda \mathbf{G}_0 \mathbf{y}. \quad (5.5)$$

が得られる。左から \mathbf{y}^T を掛けると、

$$\lambda = \|\mathbf{z}\|^2 / \|\mathbf{y}\|^2 \quad (5.6)$$

つまり成長率の二乗となる。 $n \times n$ の \mathbf{G}_0 及び $\mathbf{G}_0^{-1} \mathbf{M}^T \mathbf{G}_\tau \mathbf{M}$ の逆行列を計算する必要がある。

5.4 アンサンブル随伴感度解析

\mathbf{p} で重みづけした平均で最適な摂動を $\mathbf{y} = \mathbf{Y}\mathbf{p}$ 及び $\delta J = \delta \mathbf{J}^T \mathbf{p}$ と表す。ここで $\mathbf{Y} = (\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_m)^T$ 及び $\delta \mathbf{J} = (\delta J_1, \delta J_2, \dots, \delta J_m)^T$ である。 $\|\mathbf{y}\| = 1$ の条件下で δJ を最大化する \mathbf{p} を求めるには、

$$F(\mathbf{p}, \lambda) = \delta \mathbf{J}^T \mathbf{p} + \lambda(1 - \mathbf{p}^T \mathbf{Y}^T \mathbf{G}_0 \mathbf{Y} \mathbf{p})$$

を最小化すればよい。 $\partial F / \partial \mathbf{p} = 0$ を計算すると

$$2\lambda \mathbf{p} = (\mathbf{Y}^T \mathbf{G}_0 \mathbf{Y})^{-1} \delta \mathbf{J}$$

すなわち

$$\mathbf{y} \propto \mathbf{Y} (\mathbf{Y}^T \mathbf{G}_0 \mathbf{Y})^{-1} \delta \mathbf{J}$$

を得る。 $\|\mathbf{y}\|_i = 1$ かつ $\mathbf{y}_i^T \mathbf{G}_0 \mathbf{y}_j = 0$, $i \neq j$ ならば $\mathbf{Y}^T \mathbf{G}_0 \mathbf{Y} \propto \mathbf{I}$ 及び $\mathbf{y} = \mathbf{Y} \delta \mathbf{J}$ 。

i Note

- $\partial J / \partial \mathbf{x}_t = 2 \mathbf{G}_t \mathbf{z}$ なら、Equation 5.6 同様に指数函数的成長を示唆する。
- メンバ数が状態数に近づく ($m \rightarrow n$) につれて、アンサンブル随伴感度は随伴感度に漸近する。

5.5 単回帰

各アンサンブルメンバが $\delta J = [\partial J / \partial \mathbf{x}_0]^T \mathbf{y}$ であるとき、これを集めて $\delta \mathbf{J} = \mathbf{Y}^T \partial J / \partial \mathbf{x}_0$ と表す。

左から \mathbf{Y} を掛けると最小分散解

$$\frac{\partial J}{\partial \mathbf{x}_0} = (\mathbf{Y} \mathbf{Y}^T)^{-1} \mathbf{Y} \delta \mathbf{J} \quad (5.7)$$

Ancell and Hakim (2007) はさらに $\mathbf{A} = \mathbf{Y} \mathbf{Y}^T$ に対して対角近似 $\mathbf{D} = \text{diag}(\mathbf{A})$ を仮定し、単回帰を得た。

5.6 多変数回帰

Hacker and Lei (2015) は感度解析を次のように定式化した。

$$\delta \mathbf{J} = \mathbf{Y}^T \beta + \varepsilon \quad (5.8)$$

を考えた。ここで $\beta = \partial J / \partial \mathbf{x}_0$ である。状態の自由度が標本数よりも大きい場合、Equation 5.8 は劣決定となる。

解を一つに定めるため最小ノルム解を考え、 $\delta \mathbf{J} = \mathbf{Y}^T \beta$ の条件の下で $\beta^T \beta$ を最小化した。

$$\frac{\partial J}{\partial \mathbf{x}_0} = \mathbf{Y} (\mathbf{Y}^T \mathbf{Y})^{-1} \delta \mathbf{J}$$

これは多変数回帰である。

Duc et al. (2023) は正則化と Moore–Penrose 逆行列を用いて主双対解を得た。Enomoto et al. (2015) は多変量で、アンサンブル空間で逆行列を求めており、初期ノルム \mathbf{G}_0 を考慮している。

5.7 アンサンブル SV

重み \mathbf{p} を用いて検証時刻の摂動を $\mathbf{z} = \mathbf{Z}\mathbf{p}$ 表す。 $\|\mathbf{y}\| = 1$ の下で $\mathbf{p}^T \mathbf{Z}^T \mathbf{G}_\tau \mathbf{Z} \mathbf{p}$ を最大化する \mathbf{p} を求める。

$$F(\mathbf{p}, \lambda) = \mathbf{p}^T \mathbf{Z}^T \mathbf{G}_\tau \mathbf{Z} \mathbf{p} + \lambda(1 - \mathbf{p}^T \mathbf{Y}^T \mathbf{G}_0 \mathbf{Y} \mathbf{p})$$

を最小化する。 $\partial F / \partial \mathbf{p} = 0$ を計算し、一般化固有値問題を

$$\mathbf{Z}^T \mathbf{G}_\tau \mathbf{Z} \mathbf{p} = \lambda \mathbf{Y}^T \mathbf{G}_0 \mathbf{Y} \mathbf{p}.$$

を得る。 $\mathbf{Y}^T \mathbf{G}_0 \mathbf{Y} \propto \mathbf{I}$ なら $\mathbf{Z}^T \mathbf{G}_\tau \mathbf{Z} \mathbf{p} = \lambda \mathbf{p}$ に帰着する。アンサンブル変換 (ET: ensemble transform) (Bishop and Toth 1999) 及び解析誤差共分散 SV (Hamill et al. 2003) と同じである。

5.8 日本域に対するアンサンブル感度

Enomoto, Yamane, Ohfuchi (2015)

- データ: 25 メンバの気象庁週間アンサンブル予報
- 初期時刻: 2003 年 1 月 2 日 1200 UTC
- 検証領域: 125°N–150°E, 25°N–50°N
- 検証時刻: 2003 年 1 月 5 日 1200 UTC
- 乾燥全エネルギーノルム $TE = \frac{1}{2} \int_{p_r}^{p_t} \int_A u'^2 + v'^2 + \frac{c_p}{T_r} T'^2 + RT_r \left(\frac{p'_s}{p_r} \right)^2 dAdp$

5.9 2003 年 3 月の爆弾低気圧

Enomoto, Yamane, Ohfuchi (2015)

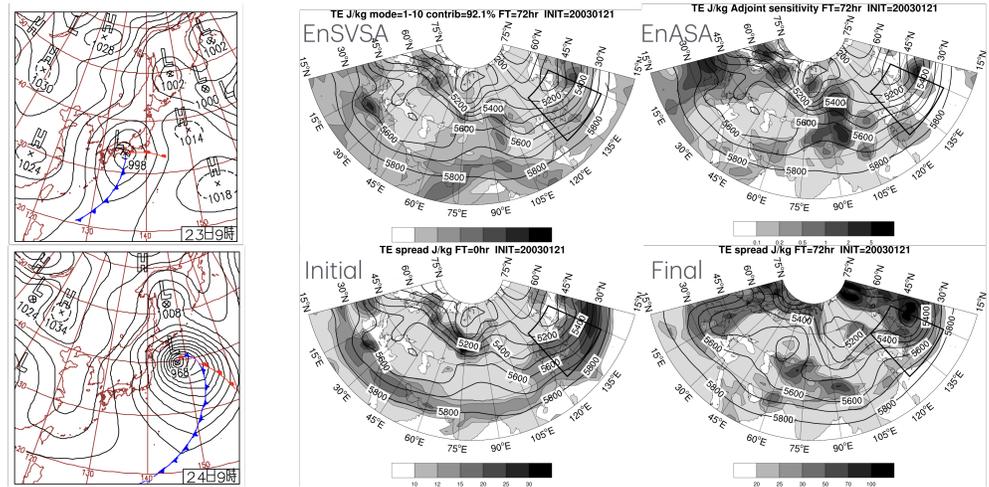


Figure5.1: 左列: 2003年1月23日及び24日9時(日本時)の気象庁天気図。右2列: 1月21日1200 UTCからの気象庁週間アンサンブル予報を用いたアンサンブル感度解析(上段、左からEnSVSA及びEnASA)、初期及び検証時刻のアンサンブルスプレッド。検証領域を矩形で示す。

5.10 乾燥全エネルギーノルムの時間発展

Enomoto, Yamane, Ohfuchi (2015)

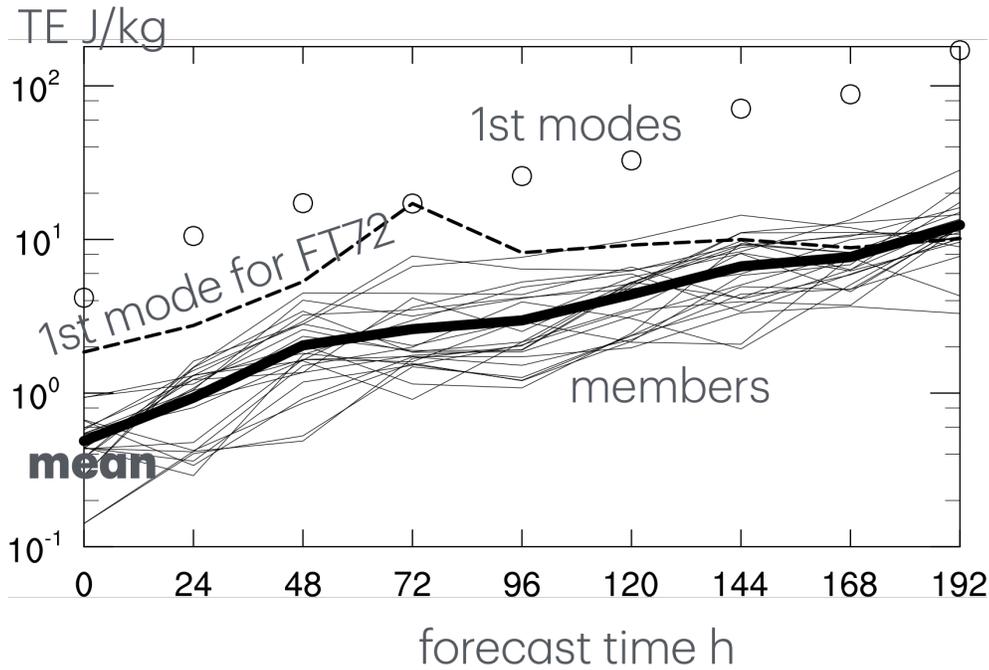


Figure5.2: 乾燥全エネルギーノルムの時間発展。2003年1月2日1200 UTCからの気象庁週間アンサンブル予報。検証領域は日本域(120°—150°E、20°N—50°N)。横軸は予報時間(h)、縦軸は鉛直に積算した単位面積あたりの乾燥全エネルギーノルム。細実線はアンサンブルメンバ、太実線はアンサンブル平均、破線は予報時間72 hに対するアンサンブル特異ベクトル第一モード。○は各予報時間に対する第一モード

5.11 欧州の切離低気圧と熱帯低気圧 Cristobal 2002

Enomoto, Ohfuchi, Nakamura, Shapiro (2007)

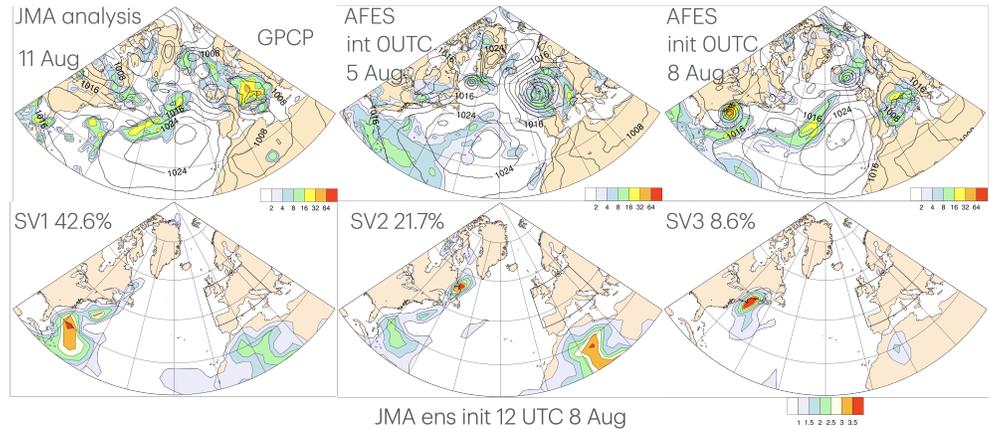


Figure5.3: 上段: 2002年8月11日0000UTCにおける気象庁海面気圧解析 hPa 及び GPCP 降水量 mm/d (左)、AFES による8月5日(中)及び8日(右)0000UTCからの予報。下段: 8月8日1200 UTCからの気象庁アンサンブル予報を用いたアンサンブル特異ベクトル第一〜三(右から左), Enomoto et al. 2007

5.12 アンサンブル SV の適用例

- Nishii and Nakamura (2010) は 2006 年 1 月下旬の成層圏突然昇温は北大西洋のブロッキングに感度があり、ブロッキングは北東太平洋の低気圧に感度があることを見出した。
- Matsueda et al. (2011) は 2005 年 12 月 15 日の NCEP の予報が外れ、ブロッキングが強すぎた原因は中部北東太平洋の不確実性にあると特定した。
- Takemura et al. (2021) はモンスーントラフの予報に影響を与えた、上流の気圧の峰と下流の渦位偏差を検出した。

5.13 台風 Hagibis 2019

Nakashita and Enomoto (2021)

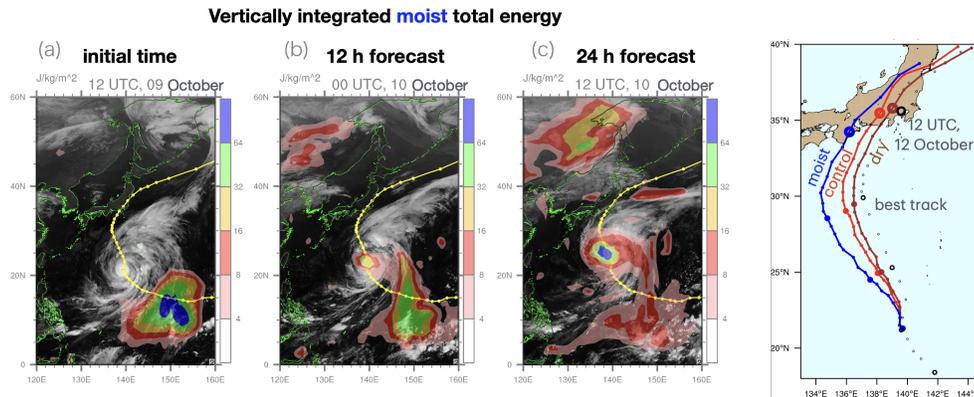


Figure5.4: 左: 鉛直に積分された湿潤全エネルギーで示された感度。左から初期時刻、12時間予報、24時間予報) (Nakashita and Enomoto (2021), Fig. 4 にひまわり画像を追加)、右: 気象庁 GSM を用いて摂動を加えた実験 (中下他 2025, 気象研究ノート 249

5.14 まとめ

- 予報感度解析は注目している現象に影響を与える領域や変数を特定する有用な道具である。
- 随伴及びアンサンブルに基づく定式化は、どちらもラグランジュの未定乗数法で一貫して導出できる。
- アンサンブル SV は、爆弾低気圧や台風など顕著な気象に適用され、メカニズムや予測可能性に関する要因を特定できただけでなく、アンサンブルカルマンフィルタのインフレーションとして有効である。
- 偽りの感度は局所化等を用いて特定する必要がある。(Griewank et al. 2023).

第6章

機械学習

機械学習は、データを学習して推定を改善する手法である。

まず、Chase et al. (2022) に基づいて、統計的学習と呼ばれる古典的な手法について述べる。

機械学習は、物理法則に依らずデータに基づく経験的な手法で、訓練データセットに対してパラメタを適合（学習）し、あらかじめ定めた損失（コスト）を最適化する。

機械学習は、教師あり（supervised）と教師なし（unsupervised）に大別される。教師あり学習には入力の特徴量（feature）に対する出力のラベル（label）を用いるが、教師なし学習にはラベルを用いない。

特徴量は説明変数、予測因子、予測変数（predictor）とも呼ばれ、ベクトルまたは行列 \mathbf{X} で表される。ラベルは目的変数（target）や予測対象、被予測変数（predictand）とも呼ばれ、スカラーまたはベクトル y で表される。

教師あり学習は、回帰（regression）と分類（classification）があり、回帰では連続した出力、分類では離散的な出力を得る。どちらも訓練データを用いて、予測するために最適な重み（閾値）を求める。重みは、誤差（error）を最小に、ラベルの確率を最大にするように決める。

6.1 線型回帰

最も簡単な回帰は線型回帰 Section 1.1 である。予測したい量 \hat{y} を重付き平均で求める。

$$\hat{y} = \sum_{i=1}^{i=D} w_i x_i + b \quad (6.1)$$

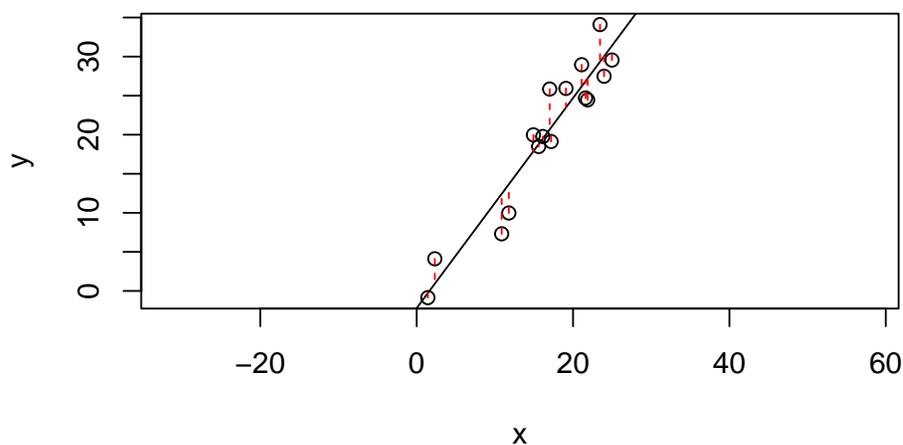
D は特徴量の数で b はバイアス項である。

重み w_i を求めるため、損失関数（loss function）を最小化する。

残差和二乗 (RSS: residual summed squared) 損失は

$$\text{RSS} = \sum_{j=1}^N (y_j - \hat{y}_j)^2 \quad (6.2)$$

で定義される。 y_i はデータ点、 \hat{y}_j は予報されたデータ点、 N は訓練データセットにあるデータ点の数である。



データにノイズが含まれている場合、学習が不安定になる。これを回避する方法が、正則化 (regularization) である。

リッジ (ridge) 回帰では、L2 損失

$$\text{RSS}_{L2} = \sum_{j=1}^N (y_j - \hat{y}_j)^2 + \lambda \sum_{i=0}^D w_i^2 \quad (6.3)$$

ラッソ (Lasso) 回帰では (L1) 損失

$$\text{RSS}_{L2} = \sum_{j=1}^N (y_j - \hat{y}_j)^2 + \lambda \sum_{i=0}^D |w_i| \quad (6.4)$$

を用いる。二つを組み合わせたものがエラスティックネット (elastic net) である。

$$\text{RSS}_{L2} = \sum_{j=1}^N (y_j - \hat{y}_j)^2 + \sum_{i=0}^D [\alpha w_i^2 + (1 - \alpha)|w_i|] \quad (6.5)$$

6.2 ロジスティック回帰

ロジスティック回帰は分類に用いられ、出力にシグモイド (sigmoid) 関数を適用する。

$$S(y) = \frac{1}{1 + e^{-y}} \quad (6.6)$$

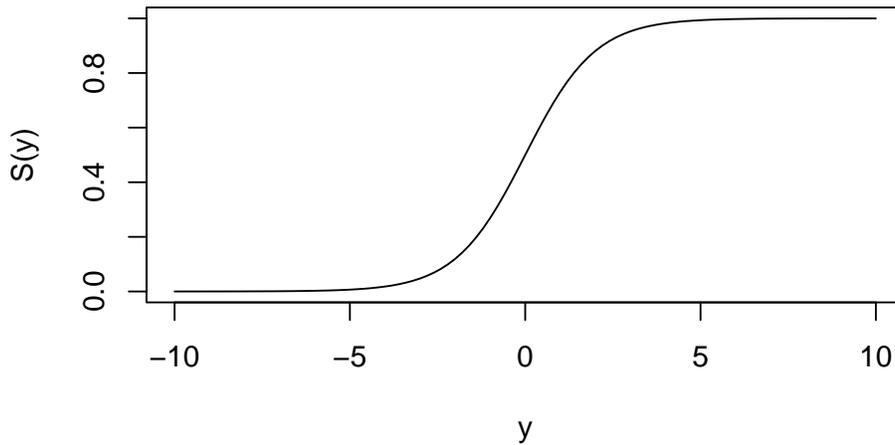


Figure6.1

ロジスティック回帰を用いた分類の損失関数は次のようになる。

$$\text{loss} = \sum_{i=0}^{i=D} -y_i \log[S(\hat{y})] + (1 + y_i) \log[1 - S(\hat{y})] \quad (6.7)$$

6.3 ナイーブベイズ

ナイーブベイズは、ベイズの定理 Section 1.3 に由来する分類である。入力特徴量 x が与えられたとき、ラベル y の確率密度 $P(y|x)$ を求める。(Section 1.3 とは y と x が入れ替わっていることに注意。)

$$P(y|x) = \frac{P(y)P(x|y)}{P(x)} \quad (6.8)$$

ナイーブと呼ばれるのは、現実には必ずしも当てはまらない、全ての入力特徴量 x が独立で、 $P(x|y)$ が正規分布のようなモデル化されたもので表されるという仮定を用いているためである。

予測は次の式で与えられる。

$$\hat{y} = \operatorname{argmax} \left\{ \log[P(y)] + \sum_{i=0}^N \log[P(x_i|y)] \right\} \quad (6.9)$$

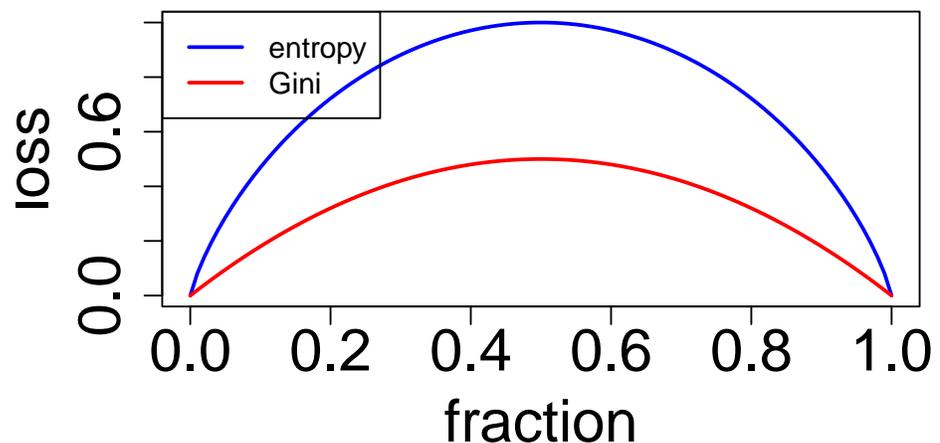
6.4 決定木

決定木（けっていぎ）を用いた分類では、ジニ不純度やエントロピーを最小化するようにデータを分ける。

$$\text{Gini} = \sum_{i=0}^{i=k} p_i(1 - p_i) \quad (6.10)$$

$$\text{entropy} = \sum_{i=0}^{i=k} p_i \log_2(p_i) \quad (6.11)$$

tree-based loss



6.5 サポートベクトルマシン

サポートベクトルマシン（SVM: Support Vector Machine）は類似したデータの集まり（クラス）を分けるマージン（margin）を最大化する手法である。

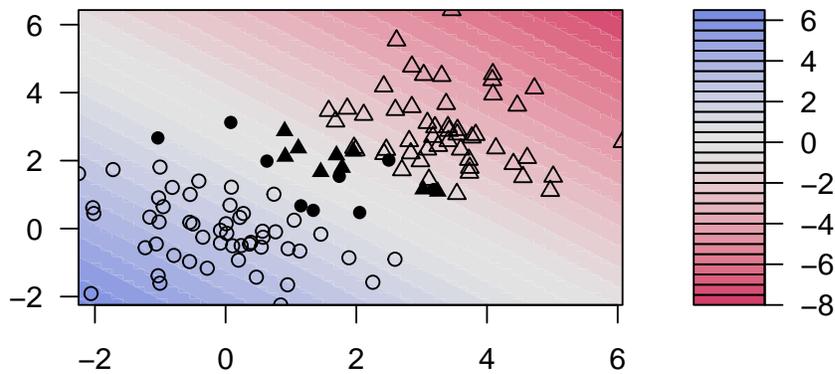
$$\hat{y} = \mathbf{w}^T \mathbf{x} + b \quad (6.12)$$

ここで \mathbf{w} は重みベクトル、 \mathbf{x} は特徴量、 b はバイアス。分類では、右辺の符号のみを用いる。線型手法とは異なり、線型境界にマージンが最大になるように \mathbf{w}^T を決める。

$$\text{margin} = \frac{1}{\mathbf{w}^T \mathbf{w}}$$

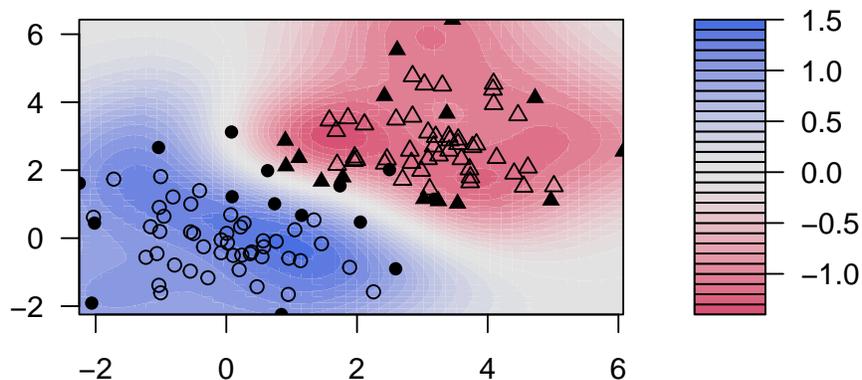
6.5.1 vanilladot

SVM classification plot



6.5.2 rbfdot

SVM classification plot



6.6 カーネル法

カーネル法は、函数近似や内挿のような数値解析や、偏微分方程式のメッシュレス解法、ニューラルネットワーク及び機械学習に用いられている (Schaback and Wendland 2006)。

カーネルは、の特徴マップにより学習入力の空間 Ω を解析が容易な高次元 (無限次元) の空間 (再生核ヒルベルト空間) \mathcal{F} に写像する。 \mathcal{F} では線型手法を用いる。

カーネルは一般的に

$$K : \Omega \times \Omega \rightarrow \mathbb{R} \quad (6.13)$$

と定義される。

特徴マップ $\Phi : \Omega \rightarrow \mathcal{F}$ について、カーネルは全ての $x, y \in \Omega$ に対し、内積を用いて

$$K(x, y) := (\Phi(x), \Phi(y))_{\mathcal{F}} \quad (6.14)$$

と定義される。

正定値対称カーネルのみを扱う。対称性は次を満たすことをいう。

$$K(x, y) = K(y, x) \quad (6.15)$$

任意の n に対して要素が $K(x_i, x_j)$ であるグラム行列 \mathbf{K} が正定値 Section A.3 である。

例えば、ランダム変数の共分散は標準的なカーネルである。 Ω からの標本 x と y とが類似していれば、相関が高く、カーネルの値は大きな正の値を取る。

よく用いられる、ガウス型カーネルは

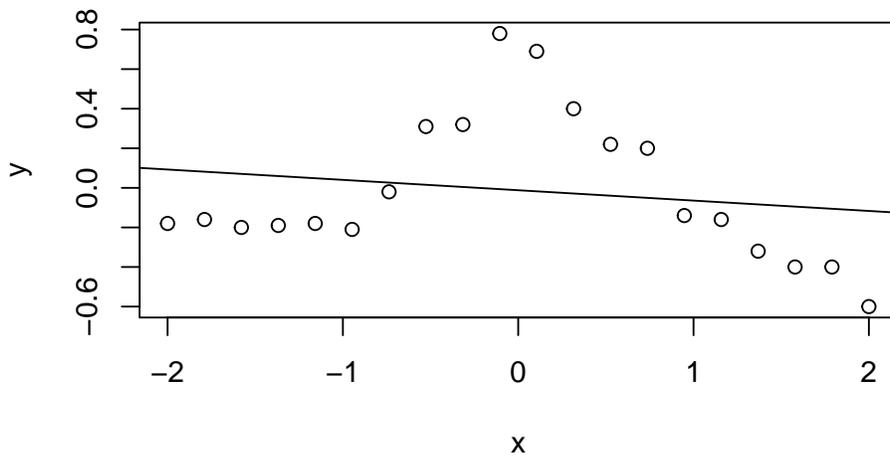
$$K(x, y) := \exp(-\|x - y\|_2^2) \quad (6.16)$$

と表される。

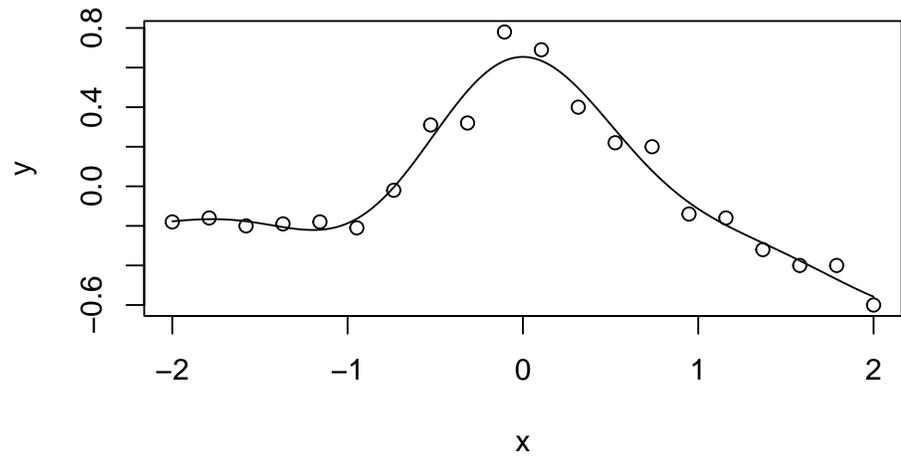
カーネル法は、回帰、サポートベクトルマシン、主成分分析、正準相関解析等に適用されている。

まず、線型回帰とカーネル回帰を比較する (赤穂昭太郎 2008)。

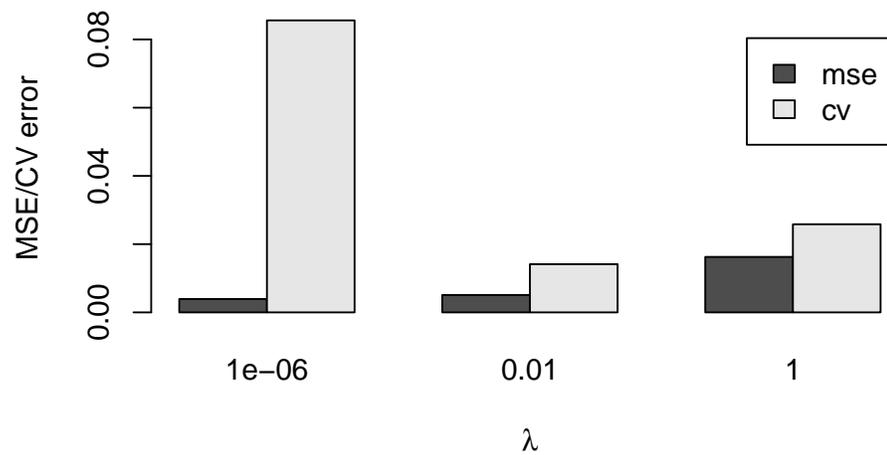
6.7 線型回帰



6.8 カーネル回帰



交差検証



カーネル密度推定は確率密度分布を推定する手法である。

独立同分布 (i.i.d.: independent and identically distributed) の標本から確率密度分布を推定する。

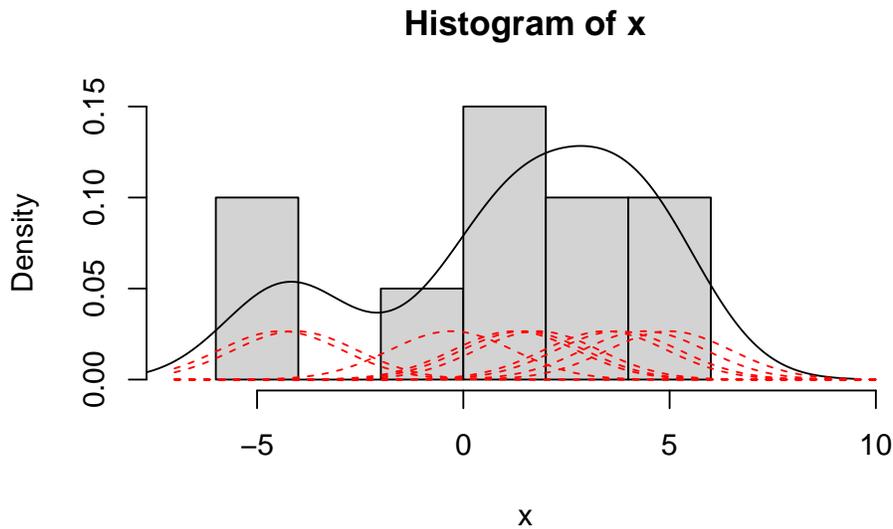
$$\hat{f}_h(x) = \frac{1}{n} \sum_{i=1}^n K_h(x - x_i) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - x_i}{h}\right) \quad (6.17)$$

ここで、 K : 非負のカーネル、 $K_h(x) = K(x/h)/h$: スケールされたカーネル、 h : バンド幅

ガウシアンカーネルを用いると次のように書ける。

$$\hat{f}_h(x) = \frac{1}{n} \frac{1}{h\sigma\sqrt{2\pi}} \sum_{i=1}^n \exp\left[\frac{-(x - x_i)^2}{2(h\sigma)^2}\right] \quad (6.18)$$

ヒストグラムと比較する。



6.9 ニューラルネットワーク

Chase et al. (2023) に基づいて、ニューラルネットワークについて述べる。

ニューロンは脳の基礎単位で、電気で信号を伝達する機能を持つ。

単純パーセプトロンはニューロンをモデル化したものである。最も簡単なパーセプトロンは、入力と出力の2層からなり、二つの出力 (0 と 1) から一つを選択して出力する。

n 個の入力に対し、 m 個の出力をするパーセプトロンは次の式で表すことができる。

$$y_j = \sum_{i=1}^n w_{ij}x_i + b_i, \quad j = 1, \dots, m$$

ここで、 x_i は入力、 y_j は出力、 w_{ij} は重み、 b_i はバイアス、 $\sigma()$ は活性化関数である。
ベクトルや行列形式では

$$\mathbf{y} = \mathbf{W}\mathbf{x} + \mathbf{b}$$

と表すことができる。

データから重み \mathbf{W} を決めることにより、単純パーセプトロンは線型分離可能な問題を解くことができるが、非線型で分離不可能な問題は解けない。

単純パーセプトロンの出力に活性化関数 $\sigma()$ を作用させ、中間層（隠れ層）を増やして多層化すると、より複雑な問題を解くことができる。これを多層パーセプトロンまたは人工ニューラルネットワーク（ANN: artificial neural network）という。

$$\mathbf{x}_k = \sigma_{k-1}(\mathbf{W}_{k-1}\mathbf{x}_{k-1} + \mathbf{b}_{k-1}) \quad (6.19)$$

ANN の重みは、確率降下学習法 (Amari 1967; 甘利 1968) と誤差逆伝播法 (Rumelhart et al. 1986) によって求めることができる。その際、自然勾配を用いることにより速く収束させることができる (Amari 1998; 甘利 2001)。

活性化関数がシグモイド関数であれば、各層はロジスティック回帰 Section 6.2 になる。

6.10 自己組織化マップ

自己組織化マップ (SOM: self-organizing map, Kohonen 1982) は、教師なし学習のクラスタリングに用いられる手法で、類似したデータを同一のカテゴリにする。SOM は隠れ層を持たず、出力は 2 次元に配置された節であり、類似した節が隣接する。そのため、パターンの遷移が解釈しやすい。

Eckert et al. (1996) は、アンサンブル予報に SOM を用いた。榎本 (2021) は日本域に適用して EOF との比較を行った。

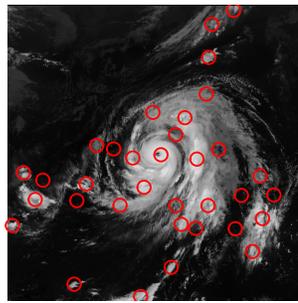
6.11 畳み込みニューラルネットワーク

畳み込みニューラルネットワーク (CNN: convolutional neural network) は、画像処理に用いられる畳み込み（フィルタ）を用いて特徴を自動的に抽出する (Fukushima 1980)。畳み込み層とともに用いられるのが、プーリング（pooling）層である。プーリング層は低解像度化など次元削減を行い、特徴の抽出を助ける。

2019 年 10 月 9 日 0000UTC における台風第 19 号のひまわり赤外画像に以下の処理を施した。

- 標準偏差 30 のガウシアンフィルタでぼかす。
- 次の式で画像 I のヘシアン H の行列式の上位 95% の座標を抽出。
- 境界上の座標を除外。
- 座標の中心を赤い丸で表示。

$$\det(H) = I_{xx}I_{yy} - I_{xy}^2$$



U-net は画像から画像を出力する CNN である。逆畳み込み (deconvolution) 層と逆プーリング (unpooling) 層により、次元を復元した出力を得る。畳み込みやプーリングをエンコーディング、逆畳み込みや逆プーリングをデコーディングといい、全体の流れが U 字となることから、名付けられている。U-net には、スペクトル変換や多重格子法と類似点がある。

6.12 トランスフォーマ

カーネル回帰から画像認識、大規模言語モデルへの発展をレビューした Milanfar の [ツイート](#) や Kassinos and Alexiadis (2024) を参考に大規模言語モデル (LLM: large language model) に用いられるトランスフォーマ (transformer) について述べる。

トランスフォーマはアテンション機構 (Vaswani et al. 2023) が重要な役割を果たしている。アテンション機構は、カーネルをデータで推定する。

トランスフォーマは言語に特化した機械学習手法と思われがちだが、処理されているのはベクトルであり、トランスフォーマで微分方程式を解くこともできる (Kassinos and

Alexiadis 2024)。

アテンション機構は、フーリエ変換などの畳み込みと似ており、単語の頻度を振動数に変換するようなものである。LLM は学習を元に、入力に対してもっともらしい出力を生成する。

アテンション機構に絞って見ていこう。 N 個のデータ列のうち i 番目の入力を $\mathbf{x}_i \in \mathbb{R}^d$ とする。3つの重み行列 $\mathbf{W}_K, \mathbf{W}_Q, \mathbf{W}_V \in \mathbb{R}^{d' \times d}$ により線型変換を行う。重みは学習して決める。

- キー: $\mathbf{K}_i = \mathbf{W}_K \mathbf{x}_i$
- クエリ: $\mathbf{Q}_i = \mathbf{W}_Q \mathbf{x}_i$
- バリュウ: $\mathbf{V}_i = \mathbf{W}_V \mathbf{x}_i$

アテンションスコアとは、入力ベクトル列の中の各対 $(\mathbf{x}_i, \mathbf{x}_j)$ に対するスコアである。

$$s_{ij} = \frac{\mathbf{Q}_i^T \mathbf{K}_j}{\sqrt{d'}} \quad (6.20)$$

これを要素とする行列がスコア行列 $\mathbf{s}_{ij} \in \mathbb{R}^{N \times N}$ である。

ソフトマックス関数は、入力 $\mathbf{z} = (z_1, \dots, z_K) \in \mathbb{R}^K$ を出力 $\sigma: \mathbb{R}^K \rightarrow (0, 1)^K$ に射影する関数である。シグモイド関数 ($K = 2$) Equation 6.6 を一般化したもので、統計力学ではボルツマン (ギブス) 分布である。

$$\sigma(\mathbf{z})_i = \frac{\exp(z_i)}{\sum_j^K \exp(z_j)}$$

ボルツマン分布をおさらいする。理想気体は、粒子 (分子) 間の相互作用が無視できるほど小さい気体である。 k 番目の量子状態にある気体の粒子数 (占拠数) $\bar{n}_k \ll 1$ とする。

$$\bar{n}_k = a \exp\left(-\frac{\varepsilon_k}{T}\right) \quad (6.21)$$

a は規格化の条件 $\sum_k \bar{n}_k = N$ によって決まる定数。Equation 6.21 で与えられる分布がボルツマン分布である。

スコア行列の各行にソフトマックス関数を適用し、アテンション重みを得る。

$$\mathbf{S}_{ij} = \frac{\exp(s_{ij})}{\sum_k^N \exp(s_{ik})} \quad (6.22)$$

各ベクトルに対するアテンション重みの和は、ソフトマックスにより 1 になるので、確率は比較可能となる。

各入力ベクトル \mathbf{x}_i に対して、アテンション重み \mathbf{S}_{ij} を用いて、バリューベクトル \mathbf{V}_j の加重合計として、出力ベクトル \mathbf{A}_i を計算する。

$$\mathbf{A}_i = \sum_{j=0}^{N-1} \mathbf{S}_{ij} \mathbf{V}_j \quad (6.23)$$

自己アテンション機構の最終出力は、長さ d' のベクトル \mathbf{A}_i が N 個連なる列である。各出力ベクトルは対応するアテンションスコアに重みづけされ、入力列全体の影響を受ける。

一見アテンションは恣意的に見えるが、数学や物理学の類似した概念と関連づけることができる。Equation 6.23 は、離散積分変換と同形であり、カーネル \mathbf{S}_{ij} により \mathbf{V}_j を \mathbf{A}_i に変化する。Equation 6.23 はフーリエ級数と類似しているが、データから学習したカーネルを用いることが異なる。

離散フーリエ変換では、カーネル $\exp(-i2\pi kn/N)$ 時系列 x_n は振動数列 X_k に変換される。

$$X_k = \sum_{n=0}^{N-1} x_n \exp\left(\frac{-i2\pi kn}{N}\right) \quad (6.24)$$

アテンションスコア Equation 6.20 s_{ij} は、 $\sqrt{d'}$ でスケールされた \mathbf{Q}_i と \mathbf{K}_j の共分散の要素である。 \mathbf{Q}_i と \mathbf{K}_j は共に \mathbf{x}_i の線型変換なので、カーネル \mathbf{S}_{ij} Equation 6.22 は自己共分散行列の情報を与える。カーネルが入力 \mathbf{x}_i の自己相関行列である積分変換は、主成分分析と関連する Karhunen-Loève 変換である。カーネルが固定されているフーリエ変換とは異なり、Karhunen-Loève 変換はトランスフォーマ同様にデータに依存する。

参考文献

- Amari, S., 1967: A theory of adaptive pattern classifiers. *IEEE Transactions on Electronic Computers*, **EC-16**, 299–307, <https://doi.org/10.1109/PGEC.1967.264666>.
- Amari, S., 1998: Natural gradient works efficiently in learning. *Neural Comput.*, **10**, 251–276, <https://doi.org/10.1162/089976698300017746>.
- Ancell, B., and G. J. Hakim, 2007: Comparing adjoint- and ensemble-sensitivity analysis with applications to observation targeting. *Mon. Wea. Rev.*, **135**, 4117–4134, <https://doi.org/10.1175/2007MWR1904.1>.
- Anderson, J. L., 2001: An ensemble adjustment Kalman filter for data assimilation. *Mon. Wea. Rev.*, **129**, 2884–2903, [https://doi.org/10.1175/1520-0493\(2001\)129%3C2884:AEAKFF%3E2.0.CO;2](https://doi.org/10.1175/1520-0493(2001)129%3C2884:AEAKFF%3E2.0.CO;2).
- , 2003: A local least squares framework for ensemble filtering. *Mon. Wea. Rev.*, **131**, 634–642, [https://doi.org/10.1175/1520-0493\(2003\)131%3C0634:ALLSFF%3E2.0.CO;2](https://doi.org/10.1175/1520-0493(2003)131%3C0634:ALLSFF%3E2.0.CO;2).
- Arakawa, A., 1966: Computational design for long-term numerical integration of the equations of fluid motion: Two-dimensional incompressible flow Part I. *J. Comput. Phys.*, **1**, 119–143.
- Bishop, C. H., and Z. Toth, 1999: Ensemble transformation and adaptive observations. *J. Atmos. Sci.*, **56**, 1748–1765, [https://doi.org/10.1175/1520-0469\(1999\)056%3C1748:ETAAO%3E2.0.CO;2](https://doi.org/10.1175/1520-0469(1999)056%3C1748:ETAAO%3E2.0.CO;2).
- , J. Etherton, and S. J. Majumdar, 2001: Adaptive sampling with the ensemble transform Kalman filter. Part I: Theoretical aspects. *Mon. Wea. Rev.*, **129**, 420–436, [https://doi.org/10.1175/1520-0493\(2001\)129%3C0420:ASWTET%3E2.0.CO;2](https://doi.org/10.1175/1520-0493(2001)129%3C0420:ASWTET%3E2.0.CO;2).

- Bowler, N. E., J. Flowerdew, and S. R. Pring, 2013: Tests of different flavours of EnKF on a simple model. *Quart. J. Roy. Meteor. Soc.*, **139**, 1505–1519, <https://doi.org/10.1002/qj.2055>.
- Broyden, C. G., 1970: The convergence of a class of double-rank minimization algorithms 1. General considerations. *IMA J. Appl. Math.*, **6**, 76–90, <https://doi.org/10.1093/imamat/6.1.76>.
- Burgers, G., P. J. van Leeuwen, and G. Evensen, 1998: Analysis scheme in the ensemble Kalman filter. *Mon. Wea. Rev.*, **126**, 1719–1724, [https://doi.org/10.1175/1520-0493\(1998\)126%3C1719:ASITEK%3E2.0.CO;2](https://doi.org/10.1175/1520-0493(1998)126%3C1719:ASITEK%3E2.0.CO;2).
- Butcher, J. C., and G. Wanner, 1996: Runge-Kutta methods: Some historical notes. *Appl. Numer. Math.*, **22**, 113–151, [https://doi.org/10.1016/S0168-9274\(96\)00048-7](https://doi.org/10.1016/S0168-9274(96)00048-7).
- Chase, R. J., D. R. Harrison, A. Burke, G. M. Lackmann, and A. McGovern, 2022: A machine learning tutorial for operational meteorology. Part I: Traditional machine learning. *Wea. Forecast.*, **37**, 1509–1529, <https://doi.org/10.1175/WAF-D-22-0070.1>.
- , ———, G. M. Lackmann, and A. McGovern, 2023: A machine learning tutorial for operational meteorology. Part II: Neural networks and deep learning. *Wea. Forecast.*, **38**, 1271–1293, <https://doi.org/10.1175/WAF-D-22-0187.1>.
- Duc, L., T. Kawabata, and D. Hotta, 2023: On the foundation and different interpretations of ensemble sensitivity. *Mon. Wea. Rev.*, **151**, 1689–1697, <https://doi.org/10.1175/MWR-D-22-0273.1>.
- Eckert, P., D. Cattani, and J. Ambühl, 1996: Classification of ensemble forecasts by means of an artificial neural network. *Meteor. Appl.*, **3**, 169–178, <https://doi.org/https://doi.org/10.1002/met.5060030207>.
- Enomoto, T., W. Ohfuchi, H. Nakamura, and M. A. Shapiro, 2007: Remote effects of tropical storm Cristobal upon a cut-off cyclone over Europe in August 2002. *Meteor. Atmos. Phys.*, **96**, 29–42, <https://doi.org/10.1007/s00703-006-0219-2>.
- , S. Yamane, and W. Ohfuchi, 2015: Simple sensitivity analysis using ensemble forecasts. *J. Meteor. Soc. Japan*, **93**, 199–213, <https://doi.org/10.2151/jmsj>.

2015-011.

- Evensen, G., 1994: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.: Oceans*, **99**, 10143–10162, <https://doi.org/10.1029/94JC00572>.
- Fletcher, R., 1970: A new approach to variable metric algorithms. *The Computer Journal*, **13**, 317–322, <https://doi.org/10.1093/comjnl/13.3.317>.
- Fukushima, K., 1980: Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biological Cybernetics*, **36**, 193–202, <https://doi.org/10.1007/BF00344251>.
- Goldfarb, D., 1970: A family of variable-metric methods derived by variational means. *Math. Comput.*, **24**, 23–26, <https://doi.org/10.1090/S0025-5718-1970-0258249-6>.
- Golub, G. H., and C. F. Van Loan, 2013: *Matrix computations - 4th edition*. fourth. Johns Hopkins University Press,.
- Griewank, P. J., M. Weissmann, T. Necker, T. Nomokonova, and U. Löhnert, 2023: Ensemble-based estimates of the impact of potential observations. *Quart. J. Roy. Meteor. Soc.*, **149**, 1546–1571, <https://doi.org/10.1002/qj.4464>.
- Hacker, J. P., and L. Lei, 2015: Multivariate ensemble sensitivity with localization. *Mon. Wea. Rev.*, **143**, 2013–2027, <https://doi.org/10.1175/MWR-D-14-00309.1>.
- Hamill, T. M., C. Snyder, and J. S. Whitaker, 2003: Ensemble forecasts and the properties of flow-dependent analysis-error covariance singular vectors. *Mon. Wea. Rev.*, **131**, 1741–1758, <https://doi.org/10.1175//2559.1>.
- Houtekamer, P. L., and H. L. Mitchell, 1998: Data assimilation using an ensemble Kalman filter technique. *Mon. Wea. Rev.*, **126**, 796–811, [https://doi.org/10.1175/1520-0493\(1998\)126%3C0796:DAUAEK%3E2.0.CO;2](https://doi.org/10.1175/1520-0493(1998)126%3C0796:DAUAEK%3E2.0.CO;2).
- Huang, X.-Y., and X. Yang, 1996: *Variational data assimilation with the Lorenz model*. HIRLAM,.
- Kalnay, E., 2003: *Atmospheric modeling, data assimilation and predictability*. Cambridge University Press,.

- Kassinis, S., and A. Alexiadis, 2024: Beyond language: Applying MLX transformers to engineering physics. <https://doi.org/10.48550/arXiv.2410.04167>.
- Kohonen, T., 1982: Self-organized formation of topologically correct feature maps. *Biol. Cybern.*, **43**, 59–69, <https://doi.org/10.1007/BF00337288>.
- Langland, R. H., M. A. Shapiro, and R. Gelaro, 2000: Initial condition sensitivity and error growth in forecasts of the 25 January 2000 East Coast snowstorm. *Mon. Wea. Rev.*, **130**, 957–974, [https://doi.org/10.1175/1520-0493\(2002\)130%3C0957:ICSAEG%3E2.0.CO;2](https://doi.org/10.1175/1520-0493(2002)130%3C0957:ICSAEG%3E2.0.CO;2).
- Liu, D. C., and J. Nocedal, 1989: On the limited memory BFGS method for large scale optimization. *Math. Programming*, **45**, 503–528, <https://doi.org/10.1007/BF01589116>.
- Lorenc, A. C., 1986: Analysis methods for numerical weather prediction. *Quart. J. Roy. Meteor. Soc.*, **112**, 1177–1194.
- Lorenz, E. N., 1963: Deterministic nonperiodic flow. *J. Atmos. Sci.*, **20**, 130–141, [https://doi.org/10.1175/1520-0469\(1963\)020%3C0130:DNF%3E2.0.CO;2](https://doi.org/10.1175/1520-0469(1963)020%3C0130:DNF%3E2.0.CO;2).
- , and K. A. Emanuel, 1998: Optimal sites for supplementary weather observations: Simulation with a small model. *J. Atmos. Sci.*, **55**, 399–414, [https://doi.org/10.1175/1520-0469\(1998\)055%3C0399:OSFSWO%3E2.0.CO;2](https://doi.org/10.1175/1520-0469(1998)055%3C0399:OSFSWO%3E2.0.CO;2).
- Matsueda, M., M. Kyouda, Z. Toth, H. L. Tanaka, and T. Tsuyuki, 2011: Predictability of an atmospheric blocking event that occurred on 15 December 2005. *Mon. Wea. Rev.*, **139**, 2455–2470, <https://doi.org/10.1175/2010MWR3551.1>.
- Milne, A. A., 1926: *Winnie-the-pooh*. Project Gutenberg,.
- Nakashita, S., and T. Enomoto, 2021: Factors for an abrupt increase in track forecast error of Typhoon Hagibis (2019). *SOLA*, **17A**, 33–37, <https://doi.org/10.2151/sola.17A-006>.
- Nishii, K., and H. Nakamura, 2010: Three-dimensional evolution of ensemble forecast spread during the onset of a stratospheric sudden warming event in January 2006. *Quart. J. Roy. Meteor. Soc.*, **136**, 894–905, <https://doi.org/10.1002/qj.607>.

- Nocedal, J., 1980: Updating quasi-Newton matrices with limited storage. *Math. Comput.*, **35**, 773–782, <https://doi.org/10.2307/2006193>.
- , and Stephen J. Wright, 2006: *Numerical Optimization*. 2nd ed. Springer,.
- Rabier, F., E. Klinker, P. Courtier, and A. Hollingsworth, 1996: Sensitivity of forecast errors to initial conditions. *Quart. J. Roy. Meteor. Soc.*, **112**, 121–150, <https://doi.org/10.1002/qj.49712252906>.
- Rodgers, C. D., 2000: *Inverse methods for atmospheric sounding: Theory and practice*. World Scientific,.
- Rumelhart, D. E., G. E. Hinton, and R. J. Williams, 1986: Learning representations by back-propagating errors. *Nature*, **323**, 533–536, <https://doi.org/10.1038/323533a0>.
- Sakov, P., and P. R. Oke, 2008: A deterministic formulation of the ensemble Kalman filter: An alternative to ensemble square root filters. *Tellus A*, **60**, 361–371, <https://doi.org/10.1111/j.1600-0870.2007.00299.x>.
- Sasaki, Y., 1958: An objective analysis based on the variational method. *J. Meteor. Soc. Japan*, **36**, 77–88, https://doi.org/10.2151/jmsj1923.36.3_77.
- Schaback, R., and H. Wendland, 2006: Kernel techniques: From machine learning to meshless methods. *Acta Numerica*, **15**, 543–639, <https://doi.org/10.1017/S0962492906270016>.
- Shanno, D. F., 1970: Conditioning of quasi-Newton methods for function minimization. *Math. Comput.*, **24**, 647–656, <https://doi.org/10.1090/S0025-5718-1970-0274029-X>.
- Snyder, C., 1996: Summary of an informal workshop on adaptive observations and FASTEX. *Bull. Amer. Meteor. Soc.*, **77**, 953–961, <https://doi.org/10.1175/1520-0477-77.5.953>.
- Takemura, K., T. Enomoto, and H. Mukougawa, 2021: Predictability of Enhanced Monsoon Trough Related to the Meandered Asian Jet and Consequent Rossby Wave Breaking in Late August 2016. *J. Meteor. Soc. Japan*, **99**, 339–356, <https://doi.org/10.2151/jmsj.2021-016>.

- Talagrand, O., 1991: Adjoint models. Reading, UK, European Centre for Medium-Range Weather Forecasts, 73–91 <https://www.ecmwf.int/sites/default/files/elibrary/1991/12548-adjoint-models.pdf>.
- , and P. Courtier, 1987: Variational assimilation of meteorological observations with the adjoint vorticity equation. I: Theory. *Quart. J. Roy. Meteor. Soc.*, **113**, 1311–1328, <https://doi.org/10.1002/qj.49711347812>.
- Tippett, M. K., J. L. Anderson, C. H. Bishop, T. M. Hamill, and J. S. Whitaker, 2003: Ensemble square root filters. *Mon. Wea. Rev.*, **131**, 1485–1490, [https://doi.org/10.1175/1520-0493\(2003\)131%3C1485:ESRF%3E2.0.CO;2](https://doi.org/10.1175/1520-0493(2003)131%3C1485:ESRF%3E2.0.CO;2).
- Tsuyuki, T., and T. Miyoshi, 2007: Recent progress of data assimilation methods in Meteorology. *J. Meteor. Soc. Japan.*, **85B**, 331–361, <https://doi.org/10.2151/jmsj.85B.331>.
- Vaswani, A., N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, 2023: Attention Is All You Need. <https://doi.org/10.48550/arXiv.1706.03762>.
- Vetra-Carvalho, S., P. J. V. Leeuwen, L. Nerger, A. Barth, M. U. Altaf, P. Brasseur, P. Kirchgessner, and J.-M. Beckers, 2018: State-of-the-art stochastic data assimilation methods for high-dimensional non-Gaussian problems. *Tellus A*, **70**, 1–43.
- Whitaker, J. S., and T. M. Hamill, 2002: Ensemble data assimilation without perturbed observations. *Mon. Wea. Rev.*, **130**, 1913–1924, [https://doi.org/10.1175/1520-0493\(2002\)130%3C1913:EDAWPO%3E2.0.CO;2](https://doi.org/10.1175/1520-0493(2002)130%3C1913:EDAWPO%3E2.0.CO;2).
- Zhang, F., C. Snyder, and J. Sun, 2004: Impacts of initial estimate and observation availability on convective-scale data assimilation with an ensemble Kalman filter. *Mon. Wea. Rev.*, **132**, 1238–1253, [https://doi.org/10.1175/1520-0493\(2004\)132%3C1238:IOIEAO%3E2.0.CO;2](https://doi.org/10.1175/1520-0493(2004)132%3C1238:IOIEAO%3E2.0.CO;2).
- 榎本剛., 2021: 自己組織化マップを用いた大気循環パターンのクラスタ解析. 京都大学防災研究所年報. *B*, **64**, 313–316.
- , and 中下早織., 2022: 準地衡流モデルへの決定論的アンサンブルデータ同化. 京都大学防災研究所年報. *B*, **65**.

甘利俊一., 1968: パターン認識の理論. 計測と制御, 7, 180–189, <https://doi.org/10.11499/sicejl1962.7.180>.

——, 2001: 自然勾配学習法-学習空間の幾何学. 計測と制御, 40, 735–739, <https://doi.org/10.11499/sicejl1962.40.735>.

赤穂昭太郎, 2008: カーネル多変量解析—非線形データ解析の新しい展開. 岩波書店,.

付録 A

補遺

数学的な補足を記す。

A.1 線型問題

線型問題は $\mathbf{y} = \mathbf{H}\mathbf{x}$ と表せる。測定ベクトル \mathbf{y} の長さ m で、状態ベクトル \mathbf{x} の長さは n である。線型演算子 \mathbf{H} は状態空間を測定空間に射影する $m \times n$ の実行列である。

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix} = \begin{bmatrix} H_{11} & H_{12} & \dots & H_{1n} \\ H_{21} & H_{22} & \dots & H_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ H_{m1} & H_{m2} & \dots & H_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \quad (\text{A.1})$$

問題は、情報量に応じて分類される。 $m < n$ のとき劣決定 (underdetermined)、 $m > n$ のとき優決定 (overdetermined) という。劣決定と優決定は背反ではなく、あるパラメタに対しては優決定でも別のパラメタに対しては劣決定であることもあり、これを混合決定 (mixed-determined) という。

A.2 様々な行列

正則 (regular) 行列: $\mathbf{A}\mathbf{A}^{-1} = \mathbf{A}^{-1}\mathbf{A} = \mathbf{I}_n$ となる行列

正規 (normal) 行列: $\mathbf{A}\mathbf{A}^* = \mathbf{A}^*\mathbf{A}$

ユニタリ (unitary) 行列: $\mathbf{U}\mathbf{U}^* = \mathbf{U}^*\mathbf{U} = \mathbf{I}_n$ 、 $\mathbf{U}^{-1} = \mathbf{U}^*$ 。実数の場合は直交 (orthogonal) 行列という。

エルミート (Hermitian) 行列: $\mathbf{A}^* = \mathbf{A}$ 。実数の場合は対称 (symmetric) 行列という。

グラム (Gram) 行列: ベクトル $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n \in \mathbb{C}$ に対して要素がベクトルの内積で表さ

れる行列。

$$\mathbf{G} = \begin{pmatrix} \mathbf{u}_1^* \mathbf{u}_1 & \cdots & \mathbf{u}_n^* \mathbf{u}_1 \\ \vdots & \ddots & \vdots \\ \mathbf{u}_1^* \mathbf{u}_n & \cdots & \mathbf{u}_n^* \mathbf{u}_n \end{pmatrix} = \mathbf{U}^* \mathbf{U}$$

ここで $\mathbf{U} = (\mathbf{u}_1 \ \mathbf{u}_2 \ \cdots \ \mathbf{u}_n)$ である。グラム行列はエルミート行列で、半正定値、 $\det \mathbf{G} \geq 0$ である。

A.3 行列の定値性

正定値

$$\mathbf{x}^T \mathbf{A} \mathbf{x} > 0$$

半正定値

$$\mathbf{x}^T \mathbf{A} \mathbf{x} \geq 0$$

A.4 ベクトル空間

バナッハ (Banach) 空間 ノルムが定める距離が完備である線型空間。

ソボレフ (Sobolev) 空間 関数と k 階までの弱微分が L^p ノルムに属する空間でバナッハ空間の一部。

ヒルベルト (Hilbert) 空間 バナッハ空間の特別な場合で内積が定義できる。 L^2 のソボレフ空間。内積から導かれるノルムで完備であり、ベクトル同士の角度を考慮することができる。ユークリッド空間とは異なり、無限次元。

ヘルダー (Hölder) 空間 関数が k 階微分可能で、以下のヘルダー条件を満たす。

$$|f(x) - f(y)| \leq C|x - y|^\alpha$$

ここで $0 < \alpha \leq 1$ 。

A.5 微分

弱微分 (weak derivative) ソボレフ空間や偏微分方程式の理論に用いられる微分の一般化。積分等式

$$\int u \frac{\partial \phi}{\partial x} dx = - \int v \phi dx$$

が成り立つとき、 v は u の x に関する弱微分という。

劣微分 (subdifferential) 凸解析における概念で最適化に用いられる。関数 $f: \mathbb{R}^n \rightarrow \mathbb{R}$ が凸であれば、

$$f(\mathbf{y}) \geq f(\mathbf{x}) + \langle \mathbf{v}, \mathbf{y} - \mathbf{x} \rangle \quad \mathbf{y} \in \mathbb{R}^n$$

を満たす $\mathbf{v} \in \mathbb{R}^n$ の集合を \mathbf{x} における劣微分 $\partial f(\mathbf{x})$ という。

ガトー (Gâteaux) 微分 バナッハ空間等のノルム空間に拡張した方向微分。方向 v に対して

$$\lim_{t \rightarrow 0} \frac{f(x + tv) - f(x)}{t}$$

を函数 $f: X \rightarrow Y$ の x における方向 v へのガトー微分という。全ての方向に微分可能とは限らず、線型とは限らない。

フレシェ (Fréchet) 微分 次を満たす線型写像 $A: X \rightarrow Y$ を、函数 $f: X \rightarrow Y$ の x におけるフレシェ (Fréchet) 微分という。

$$\lim_{\|h\| \rightarrow 0} \frac{\|f(x+h) - f(x) - A(h)\|}{\|h\|} = 0$$

$\partial \mathbf{y} / \partial \mathbf{x}$ は、通常次の規約が用いられる。

$$\left[\frac{\partial \mathbf{y}}{\partial \mathbf{x}} \right]_{ij} = \left(\frac{\partial y_i}{\partial x_j} \right)$$

\mathbf{y} は行の添字、 \mathbf{x} は列の添字に関連付けられ、テイラー展開は「右を見て」

$$\mathbf{y} = \mathbf{y}_0 + \frac{\partial \mathbf{y}}{\partial \mathbf{x}} (\mathbf{x} - \mathbf{x}_0) + \dots$$

のように書ける。

A.6 逆行列

Woodbury の公式 (Golub and Van Loan 2013)

$$(\mathbf{A} + \mathbf{UBV})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1} \mathbf{U} (\mathbf{B}^{-1} + \mathbf{VA}^{-1} \mathbf{U})^{-1} \mathbf{VA}^{-1}$$

A.7 擬逆行列

- Hingham, N. 2023: [What Is the Pseudoinverse of a Matrix](#)

一般化行列 (generalized inverse) は非特異な正方行列の逆行列を拡張したもので、多数存在する。一般化行列の一つで有用なものが擬逆行列 (pseudoinverse) である。行列 $\mathbf{A} \in \mathbb{C}^{m \times n}$ の擬逆行列 \mathbf{X} は、以下の Moore-Penrose 条件を満たすものである。

1. $\mathbf{AXA} = \mathbf{A}$
2. $\mathbf{XAX} = \mathbf{X}$
3. $(\mathbf{AX})^* = \mathbf{AX}$
4. $(\mathbf{XA})^* = \mathbf{XA}$

ここで上付の $*$ は複素共軛を表す。 \mathbf{X} は唯一に定まることが示されている。擬逆行列を \mathbf{A}^+ または \mathbf{A}^\dagger で表す。

優決定または劣決定の線型方程式系 $\mathbf{A}\mathbf{x} = \mathbf{b}$ の最小 2 ノルム最小二乗解は擬逆行列で表され、 $\mathbf{x} = \mathbf{A}^+\mathbf{b}$ は $\|\mathbf{A}\mathbf{x} - \mathbf{b}\|$ を最小化する。

擬逆行列は特異値分解 $\mathbf{A} = \mathbf{U}\Sigma\mathbf{V}^*$ で表すことができる。ここで $m \times m$ の \mathbf{U} 及び $n \times n$ の \mathbf{V} はユニタリ行列である。 $m \times n$ 行列 $\Sigma = \text{diag}(\sigma_1, \dots, \sigma_p)$, $p = \min(m, n)$ で、 $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r > 0$, $\sigma_{r+1} = \dots = \sigma_p = 0$ である。擬逆行列は特異値分解を用いて次のように表される。

$$\mathbf{A}^+ = \mathbf{V}\Sigma^+\mathbf{U}^* \quad (\text{A.2})$$

ここで、 $n \times m$ 行列 $\Sigma^+ = \text{diag}(\sigma_1^{-1}, \dots, \sigma_r^{-1}, 0, \dots, 0)$ である。特異値分解を用いると、Moore-Penrose の条件を確認することができる。

Equation A.2 より $(\mathbf{A}^+)^+ = \mathbf{A}$ 及び $(\mathbf{A}^*)^+ = (\mathbf{A}^+)^*$ 導かれる。

\mathbf{A} が列完全階数 (full rank) すなわち $\text{rank}(A) = n \leq m$ (縦長) のとき

$$\mathbf{A}^+ = (\mathbf{A}^*\mathbf{A})^{-1}\mathbf{A}^*$$

となる。このとき $\mathbf{A}^+\mathbf{A} = \mathbf{I}_n$ である。これは、標本数 m が自由度 n よりも大きいときの最小二乗解に対応する。

同様に \mathbf{A} が行完全階数 (full rank) すなわち $\text{rank}(A) = m \leq n$ (横長) のとき

$$\mathbf{A}^+ = \mathbf{A}^*(\mathbf{A}\mathbf{A}^*)^{-1}$$

となる。このとき $\mathbf{A}\mathbf{A}^+ = \mathbf{I}_m$ である。これは、標本数 m が自由度 n よりも小さいときの最小ノルム解に対応する。