

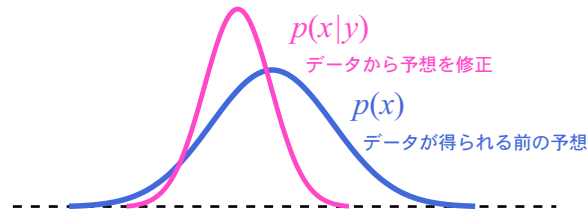
3 カルマンフィルタ

逐次データ同化の考え方

ベイズの定理

$$p(\mathbf{x}|\mathbf{y}) = \frac{p(\mathbf{y}|\mathbf{x})p(\mathbf{x})}{\int p(\mathbf{y}|\mathbf{x})p(\mathbf{x}) d\mathbf{x}}.$$

を用いると、事前の予想と実データの情報を組み合わせた推定ができる。このとき、なるべくよい事前分布を使った方が、よい結果が得られる。そこで、推定したい時刻よりも少し前に取得された観測データも活用してよい事前分布を作ること考える。異なる時刻に取得されたデータをつなぎ合わせるためにシミュレーションモデルを活用する。



改めて状態空間モデル

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

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

観測モデルだけでなく、システムモデル $\mathbf{x}_k = \mathbf{f}_k(\mathbf{x}_{k-1}) + \mathbf{v}_k$ を考慮することで、その時点のデータだけでなく、時間方向のつながりを考え、よい推定を得る。

ベイズの定理

$$p(\mathbf{x}|\mathbf{y}) = \frac{p(\mathbf{y}|\mathbf{x})p(\mathbf{x})}{\int p(\mathbf{y}|\mathbf{x})p(\mathbf{x}) d\mathbf{x}}.$$

で、事前分布 $p(\mathbf{x}_k)$ に時刻 t_1 から $t_{k-1} (< t_k)$ までの観測 $\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_{k-1}$ の情報を取り入れることを考えよう。以下では、この観測データの時系列 $\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_{k-1}$ をまとめて $\mathbf{y}_{1:k-1}$ と表記することにする。時刻 t_{k-1} までの観測データ $\mathbf{y}_{1:k-1}$ の情報を踏まえた事前分布は、 $p(\mathbf{x}_k|\mathbf{y}_{1:k-1})$ という条件付き分布で表される。この条件付き分布のことを予測分布と呼ぶ。

予測分布 $p(\mathbf{x}_k|\mathbf{y}_{1:k-1})$ を事前分布として、ベイズの定理を適用すれば

$$p(\mathbf{x}_k|\mathbf{y}_{1:k}) = p(\mathbf{x}_k|\mathbf{y}_{1:k-1}, \mathbf{y}_k) = \frac{p(\mathbf{y}_k|\mathbf{x}_k, \mathbf{y}_{1:k-1})p(\mathbf{x}_k|\mathbf{y}_{1:k-1})}{\int p(\mathbf{y}_k|\mathbf{x}_k, \mathbf{y}_{1:k-1})p(\mathbf{x}_k|\mathbf{y}_{1:k-1}) d\mathbf{x}_k} \quad (3)$$

となる。ここで、式 (2) の観測モデル $\mathbf{y}_k = \mathbf{H}_k \mathbf{x}_k + \mathbf{w}_k$ を見ると、 \mathbf{y}_k は \mathbf{x}_k が与えられれば、観測ノイズ \mathbf{w}_k 以外のものからの影響を受けないので、 \mathbf{w}_k が過去の観測 $\mathbf{y}_{1:k-1}$ や \mathbf{x}_k と独立であるとすれば、

$$p(\mathbf{y}_k|\mathbf{x}_k, \mathbf{y}_{k-1}) = p(\mathbf{y}_k|\mathbf{x}_k) \quad (4)$$

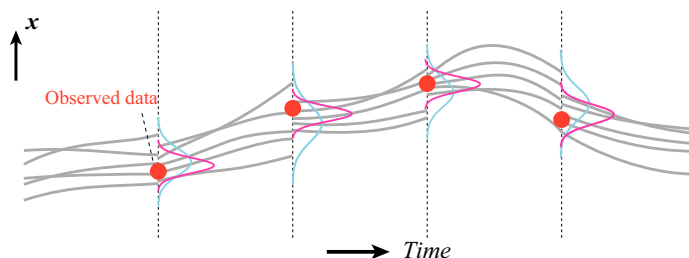
が言える。(\mathbf{x}_k が与えられたもとで \mathbf{y}_k と \mathbf{y}_{k-1} とが条件つき独立であるという。) したがって、式 (3) は、

$$p(\mathbf{x}_k|\mathbf{y}_{1:k}) = \frac{p(\mathbf{y}_k|\mathbf{x}_k)p(\mathbf{x}_k|\mathbf{y}_{1:k-1})}{\int p(\mathbf{y}_k|\mathbf{x}_k)p(\mathbf{x}_k|\mathbf{y}_{1:k-1}) d\mathbf{x}_k} \quad (5)$$

となる。 $p(\mathbf{y}_k|\mathbf{x}_k)$ は観測モデル $\mathbf{y}_k = \mathbf{H}_k \mathbf{x}_k + \mathbf{w}_k$ から得られる。あとは、システムモデル $\mathbf{x}_k = \mathbf{f}_k(\mathbf{x}_{k-1}) + \mathbf{v}_k$ を利用して、 $p(\mathbf{x}_k|\mathbf{y}_{1:k-1})$ を求めればよい。

逐次データ同化では、

1. 前の時間ステップ (*) から今の状態 \mathbf{x}_k を予測する分布 $p(\mathbf{x}_k|\mathbf{y}_1, \dots, \mathbf{y}_{k-1})$ の計算,
 2. $p(\mathbf{x}_k|\mathbf{y}_1, \dots, \mathbf{y}_{k-1})$ を事前分布とし, 観測 \mathbf{y}_k を用いて得た事後分布 $p(\mathbf{x}_k|\mathbf{y}_1, \dots, \mathbf{y}_{k-1}, \mathbf{y}_k)$ の計算,
- を繰り返すことで, 各時刻の \mathbf{x}_k を順々に推定していく.
 (* 最初の時間ステップの分布 $p(\mathbf{x}_0)$ だけは, 適当に決める必要がある.)



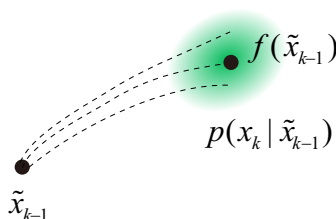
- $p(\mathbf{x}_k|\mathbf{y}_{1:k-1})$ は, 1 つ前の時間ステップまでの観測から \mathbf{x}_k を予測したもので, 予測分布と呼ぶ.
- $p(\mathbf{x}_k|\mathbf{y}_{1:k})$ は, その時間までの観測を使って \mathbf{x}_k を推定したもので, フィルタ分布と呼ぶ.
- $p(\mathbf{x}_k|\mathbf{y}_{1:k-1})$ から $p(\mathbf{x}_k|\mathbf{y}_{1:k})$ を求める操作を, フィルタリングなどと呼ぶ.

どのように, 各時間ステップで予測分布, フィルタ分布を求めるかが問題.

逐次データ同化の定式化

予測分布の計算

システムモデル $\mathbf{x}_k = \mathbf{f}_k(\mathbf{x}_{k-1}) + \mathbf{v}_k$ に基づいて, \mathbf{x}_{k-1} が与えられたときの \mathbf{x}_k の確率密度分布 $p(\mathbf{x}_k|\mathbf{x}_{k-1})$ を考える. \mathbf{x}_{k-1} が与えられる場合, $\mathbf{f}_k(\mathbf{x}_{k-1})$ の値と, \mathbf{v}_k の分布がわかれば, $p(\mathbf{x}_k|\mathbf{x}_{k-1})$ も決まる.



実際には, \mathbf{x}_{k-1} も不確かである. そこで, \mathbf{x}_{k-1} の確率分布 $p(\mathbf{x}_{k-1})$ をまず考える. 様々な \mathbf{x}_{k-1} に対して $p(\mathbf{x}_k|\mathbf{x}_{k-1})$ を求め, 確率 $p(\mathbf{x}_{k-1})$ に応じて重み付けして足し合わせると, \mathbf{x}_{k-1} が不確かな場合でも $p(\mathbf{x}_k)$ が得られる.

様々な \mathbf{x}_{k-1} に対して $p(\mathbf{x}_k|\mathbf{x}_{k-1})$ を求め, 確率 $p(\mathbf{x}_{k-1})$ に応じて重み付けして足し合わせる, という操作は, 以下の積分で書ける:

$$p(\mathbf{x}_k) = \int p(\mathbf{x}_k|\mathbf{x}_{k-1}) p(\mathbf{x}_{k-1}) d\mathbf{x}_{k-1}. \quad (6)$$

こうして, $p(\mathbf{x}_{k-1})$, $p(\mathbf{x}_k|\mathbf{x}_{k-1})$ が与えられたときの $p(\mathbf{x}_k|\mathbf{x}_{k-1})$ が計算できる.

フィルタ分布の計算

$p(\mathbf{x}_k)$ が得られると、ベイズの定理

$$p(\mathbf{x}_k | \mathbf{y}_k) = \frac{p(\mathbf{y}_k | \mathbf{x}_k) p(\mathbf{x}_k)}{\int p(\mathbf{y}_k | \mathbf{x}_k) p(\mathbf{x}_k) d\mathbf{x}_k}. \quad (7)$$

より、その時刻の観測 \mathbf{y}_k を使って事後分布 $p(\mathbf{x}_k | \mathbf{y}_k)$ が計算できる。このようにすれば、少し前の時刻 t_{k-1} の状態 \mathbf{x}_{k-1} についての情報と、時刻 t_k の観測 \mathbf{y}_k の両方を踏まえた \mathbf{x}_k の推定ができる。

さらに次の予測

$p(\mathbf{x}_k | \mathbf{y}_k)$ は、そのまま次の時刻 t_{k+1} の予測 $p(\mathbf{x}_{k+1} | \mathbf{y}_k)$ を得るのにも使える。ここで、 $p(\mathbf{x}_{k+1} | \mathbf{y}_k)$ は、時刻 t_k の観測 \mathbf{y}_k の情報を考慮した \mathbf{x}_{k+1} の分布を意味し、以下の積分から得られる：

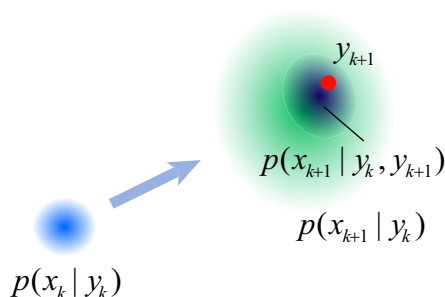
$$\begin{aligned} p(\mathbf{x}_{k+1} | \mathbf{y}_k) &= \int p(\mathbf{x}_{k+1} | \mathbf{x}_k, \mathbf{y}_k) p(\mathbf{x}_k | \mathbf{y}_k) d\mathbf{x}_k \\ &= \int p(\mathbf{x}_{k+1} | \mathbf{x}_k) p(\mathbf{x}_k | \mathbf{y}_k) d\mathbf{x}_k. \end{aligned} \quad (8)$$

ここで、1 行目から 2 行目に行くときに、

$$p(\mathbf{x}_{k+1} | \mathbf{x}_k, \mathbf{y}_k) = p(\mathbf{x}_{k+1} | \mathbf{x}_k) \quad (9)$$

を仮定した。これは、 \mathbf{x}_k が与えられれば、 \mathbf{y}_k の情報は \mathbf{x}_{k+1} に影響しない (\mathbf{x}_k を介してのみ、 \mathbf{y}_k の情報が \mathbf{x}_{k+1} の推定に寄与する) ことを意味する。

$p(\mathbf{x}_k | \mathbf{y}_k)$ を使って得た \mathbf{x}_{k+1} の予測、 $p(\mathbf{x}_{k+1} | \mathbf{y}_k)$ を事前分布として使うと、時刻 t_k の観測 \mathbf{y}_k と時刻 t_{k+1} の観測 \mathbf{y}_{k+1} の両方を考慮した推定 $p(\mathbf{x}_{k+1} | \mathbf{y}_k, \mathbf{y}_{k+1})$ が得られる。



逐次データ同化

時刻 t_{k-1} までの観測 $\mathbf{y}_1, \dots, \mathbf{y}_{k-1}$ を使った \mathbf{x}_k の予測 $p(\mathbf{x}_k | \mathbf{y}_{1:k-1})$ ：

$$p(\mathbf{x}_k | \mathbf{y}_{1:k-1}) = \int \underbrace{p(\mathbf{x}_k | \mathbf{x}_{k-1})}_{\text{Given}} \underbrace{p(\mathbf{x}_{k-1} | \mathbf{y}_{1:k-1})}_{\text{Given}} d\mathbf{x}_{k-1}.$$

時刻 t_k の観測 \mathbf{y}_k の情報を追加して \mathbf{x}_k の予測を修正 (フィルタリング)：

$$p(\mathbf{x}_k | \mathbf{y}_{1:k}) = \frac{\underbrace{p(\mathbf{y}_k | \mathbf{x}_k)}_{\text{Given}} \underbrace{p(\mathbf{x}_k | \mathbf{y}_{1:k-1})}_{\text{Given}}}{\int \underbrace{p(\mathbf{y}_k | \mathbf{x}_k)}_{\text{Given}} \underbrace{p(\mathbf{x}_k | \mathbf{y}_{1:k-1})}_{\text{Given}} d\mathbf{x}_k}.$$

- 基本は、上の計算を繰り返すことなのだが、カルマンフィルタやアンサンブルカルマンフィルタは、各ステップの事前分布、事後分布をガウス分布と仮定して大体の話が進む。
- そのため、ほぼ平均ベクトルと分散共分散行列の計算で済む。
- あまり分布の積分計算を意識する必要はない。

カルマンフィルタ

線型システムの場合

$$\mathbf{x}_k = \mathbf{f}_k(\mathbf{x}_{k-1}) + \mathbf{v}_k, \quad (10a)$$

$$\mathbf{y}_k = \mathbf{H}_k \mathbf{x}_k + \mathbf{w}_k \quad (10b)$$

において、まずは観測モデルだけでなく、システムモデルも線型と仮定し、

$$\mathbf{x}_k = \mathbf{F}_k \mathbf{x}_{k-1} + \mathbf{v}_k, \quad (11a)$$

$$\mathbf{y}_k = \mathbf{H}_k \mathbf{x}_k + \mathbf{w}_k \quad (11b)$$

とした場合の予測分布 $p(\mathbf{x}_k | \mathbf{y}_{1:k-1})$ 、フィルタ分布 $p(\mathbf{x}_k | \mathbf{y}_{1:k})$ を求める。

線型システムにおける予測

とりあえず、線型システム

$$\mathbf{x}_k = \mathbf{F}_k \mathbf{x}_{k-1} + \mathbf{v}_k,$$

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

において、時刻 t_{k-1} の状態 \mathbf{x}_{k-1} の確率分布が t_{k-1} までに得られた観測 $\mathbf{y}_{1:k-1}$ で条件付けられた分布 $p(\mathbf{x}_{k-1} | \mathbf{y}_{1:k-1})$ で与えられているものとする。また、 $p(\mathbf{x}_{k-1} | \mathbf{y}_{1:k-1})$ はガウス分布であるとする。また、システムノイズ \mathbf{v}_k 、観測ノイズ \mathbf{w}_k の確率分布も

$$p(\mathbf{v}_k) = \mathcal{N}(\mathbf{0}, \mathbf{Q}_k), \quad (12)$$

$$p(\mathbf{w}_k) = \mathcal{N}(\mathbf{0}, \mathbf{R}_k) \quad (13)$$

このとき、

$$\begin{aligned} p(\mathbf{x}_k | \mathbf{x}_{k-1}) &= p(\mathbf{F}_k \mathbf{x}_{k-1} + \mathbf{v}_k | \mathbf{x}_{k-1}) \\ &= \mathcal{N}(\mathbf{F}_k \mathbf{x}_{k-1}, \mathbf{Q}_k) \end{aligned} \quad (14)$$

のように、 $p(\mathbf{x}_k | \mathbf{x}_{k-1})$ がガウス分布となる。これと、やはりガウス分布で与えられる $p(\mathbf{x}_{k-1} | \mathbf{y}_{1:k-1})$ を用いて、

$$\begin{aligned} p(\mathbf{x}_k | \mathbf{y}_{1:k-1}) &= \int p(\mathbf{x}_k | \mathbf{x}_{k-1}) p(\mathbf{x}_{k-1} | \mathbf{y}_{1:k-1}) d\mathbf{x}_{k-1} \\ &= \int p(\mathbf{F}_k \mathbf{x}_{k-1} + \mathbf{v}_k | \mathbf{x}_{k-1}) p(\mathbf{x}_{k-1} | \mathbf{y}_{1:k-1}) d\mathbf{x}_{k-1} \end{aligned} \quad (15)$$

を計算すれば、 $p(\mathbf{x}_k | \mathbf{y}_{1:k-1})$ が求まることになる。以下では、

- $\bar{\mathbf{x}}_{k-1|k-1}$: $p(\mathbf{x}_{k-1}|\mathbf{y}_{1:k-1})$ の平均.
- $\mathbf{P}_{k-1|k-1}$: $p(\mathbf{x}_{k-1}|\mathbf{y}_{1:k-1})$ の分散共分散行列.
- $\bar{\mathbf{x}}_{k|k-1}$: $p(\mathbf{x}_k|\mathbf{y}_{1:k-1})$ の平均.
- $\mathbf{P}_{k|k-1}$: $p(\mathbf{x}_k|\mathbf{y}_{1:k-1})$ の分散共分散行列.
- \vdots

とする. つまり,

$$p(\mathbf{x}_{k-1}|\mathbf{y}_{1:k-1}) = \mathcal{N}(\bar{\mathbf{x}}_{k-1|k-1}, \mathbf{P}_{k-1|k-1}),$$

$$p(\mathbf{x}_k|\mathbf{y}_{1:k-1}) = \mathcal{N}(\bar{\mathbf{x}}_{k|k-1}, \mathbf{P}_{k|k-1}).$$

確率分布 $p(\mathbf{x}_{k-1}|\mathbf{y}_{1:k-1})$ が平均ベクトル $\mathbf{x}_{k-1|k-1}$, 分散共分散行列 $\mathbf{P}_{k-1|k-1}$ のガウス分布である場合,

$$p(\mathbf{x}_k|\mathbf{y}_{1:k-1}) = \int p(\mathbf{F}_k \mathbf{x}_{k-1} + \mathbf{v}_k | \mathbf{x}_{k-1}) p(\mathbf{x}_{k-1}|\mathbf{y}_{1:k-1}) d\mathbf{x}_{k-1}$$

の積分をまじめに計算する必要はない.

$$\mathbf{x}_k = \mathbf{F}_k \mathbf{x}_{k-1} + \mathbf{v}_k$$

において, \mathbf{x}_{k-1} と \mathbf{v}_k が独立な分布に従うとし, ガウス分布に従う 2 つの確率変数の和がやはりガウス分布に従うという性質を使って $p(\mathbf{x}_k|\mathbf{y}_{1:k-1})$ が得られる.

- \mathbf{x}_{k-1} が $\mathcal{N}(\bar{\mathbf{x}}_{k-1|k-1}, \mathbf{P}_{k-1|k-1})$ に従う
 $\Rightarrow \mathbf{F}_k \mathbf{x}_{k-1}$ の分布は, ガウス分布 $\mathcal{N}(\mathbf{F}_k \bar{\mathbf{x}}_{k-1|k-1}, \mathbf{F}_k \mathbf{P}_{k-1|k-1} \mathbf{F}_k^T)$.
- \mathbf{v}_k はガウス分布 $\mathcal{N}(\mathbf{0}, \mathbf{Q}_k)$ に従う.

従って, 予測分布 $p(\mathbf{x}_k|\mathbf{y}_{1:k-1})$ は

- ガウス分布で,
- 平均ベクトルは $\mathbf{F}_k \bar{\mathbf{x}}_{k-1|k-1}$,
- 分散共分散行列 $\mathbf{F}_k \mathbf{P}_{k-1|k-1} \mathbf{F}_k^T + \mathbf{Q}_k$ となる.

つまり, $p(\mathbf{x}_k|\mathbf{y}_{1:k-1}) = \mathcal{N}(\mathbf{F}_k \bar{\mathbf{x}}_{k-1|k-1}, \mathbf{F}_k \mathbf{P}_{k-1|k-1} \mathbf{F}_k^T + \mathbf{Q}_k)$. このように予測分布 $p(\mathbf{x}_k|\mathbf{y}_{1:k-1})$ を求める手続きは一期先予測と呼ばれる.

フィルタリング

$p(\mathbf{x}_k|\mathbf{y}_{1:k-1}) = \mathcal{N}(\mathbf{F}_k \bar{\mathbf{x}}_{k-1|k-1}, \mathbf{F}_k \mathbf{P}_{k-1|k-1} \mathbf{F}_k^T + \mathbf{Q}_k)$ は, 前の時刻までの観測の情報を取り入れた \mathbf{x}_k の予測. この予測 $p(\mathbf{x}_k|\mathbf{y}_{1:k-1})$ を事前分布とし, 観測 \mathbf{y}_k の情報を取り入れて, 事後分布 $p(\mathbf{x}_k|\mathbf{y}_{1:k})$ を求める.

状態 \mathbf{x}_k と観測 \mathbf{y}_k の関係を

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

と表し、 $p(\mathbf{y}_k|\mathbf{x}_k)$ はガウス分布になるものとする。 $p(\mathbf{x}_k|\mathbf{y}_{1:k-1})$ がガウス分布として得られているので、事後分布の計算は、前に得た結果がそのまま使える。すなわち、事後分布はガウス分布で、平均ベクトル $\bar{\mathbf{x}}_{k|k}$ 、分散共分散行列 $\mathbf{P}_{k|k}$ はそれぞれ

$$\bar{\mathbf{x}}_{k|k} = \bar{\mathbf{x}}_{k|k-1} + \mathbf{P}_{k|k-1} \mathbf{H}_k^T (\mathbf{H}_k \mathbf{P}_{k|k-1} \mathbf{H}_k^T + \mathbf{R}_k)^{-1} (\mathbf{y}_k - \mathbf{H}_k \bar{\mathbf{x}}_{k|k-1}), \quad (16a)$$

$$\mathbf{P}_{k|k} = \mathbf{P}_{k|k-1} - \mathbf{P}_{k|k-1} \mathbf{H}_k^T (\mathbf{H}_k \mathbf{P}_{k|k-1} \mathbf{H}_k^T + \mathbf{R}_k)^{-1} \mathbf{H}_k \mathbf{P}_{k|k-1} \quad (16b)$$

となる。

カルマンゲイン

式 (16a), (16b) は,

$$\mathbf{K}_k = \mathbf{P}_{k|k-1} \mathbf{H}_k^T (\mathbf{H}_k \mathbf{P}_{k|k-1} \mathbf{H}_k^T + \mathbf{R}_k)^{-1} \quad (17)$$

で定義される行列 \mathbf{K}_k を導入すると,

$$\bar{\mathbf{x}}_{k|k} = \bar{\mathbf{x}}_{k|k-1} + \mathbf{K}_k (\mathbf{y}_k - \mathbf{H}_k \bar{\mathbf{x}}_{k|k-1}), \quad (18a)$$

$$\mathbf{P}_{k|k} = (\mathbf{I} - \mathbf{K}_k \mathbf{H}_k) \mathbf{P}_{k|k-1} \quad (18b)$$

というすっきりとした形で書ける。カルマンフィルタでは通常こちらの書き方をする。行列 \mathbf{K}_k はカルマンゲインと呼ばれる。

カルマンフィルタのアルゴリズム (まとめ)

- 時間ステップ $k = 1, \dots, K$ について、以下を実行する。

- 一期先予測:

$$\bar{\mathbf{x}}_{k|k-1} = \mathbf{F}_k \bar{\mathbf{x}}_{k-1|k-1},$$

$$\mathbf{P}_{k|k-1} = \mathbf{F}_k \mathbf{P}_{k-1|k-1} \mathbf{F}_k^T + \mathbf{Q}_k.$$

- フィルタリング:

$$\mathbf{K}_k = \mathbf{P}_{k|k-1} \mathbf{H}_k^T (\mathbf{H}_k \mathbf{P}_{k|k-1} \mathbf{H}_k^T + \mathbf{R}_k)^{-1},$$

$$\bar{\mathbf{x}}_{k|k} = \bar{\mathbf{x}}_{k|k-1} + \mathbf{K}_k (\mathbf{y}_k - \mathbf{H}_k \bar{\mathbf{x}}_{k|k-1}),$$

$$\mathbf{P}_{k|k} = (\mathbf{I} - \mathbf{K}_k \mathbf{H}_k) \mathbf{P}_{k|k-1}.$$

カルマンフィルタの例

1 変数の場合

一期先予測:

$$\bar{x}_{k|k-1} = F_k \bar{x}_{k-1|k-1},$$

$$P_{k|k-1} = F_k P_{k-1|k-1} F_k^T + Q_k.$$

は、1 変数の場合、

$$\bar{x}_{k|k-1} = F_k \bar{x}_{k-1|k-1}, \quad (19a)$$

$$P_{k|k-1} = F_k^2 P_{k-1|k-1} + Q_k. \quad (19b)$$

フィルタリング:

$$\mathbf{K}_k = \mathbf{P}_{k|k-1} \mathbf{H}_k^T (\mathbf{H}_k \mathbf{P}_{k|k-1} \mathbf{H}_k^T + \mathbf{R}_k)^{-1},$$

$$\bar{\mathbf{x}}_{k|k} = \bar{\mathbf{x}}_{k|k-1} + \mathbf{K}_k (\mathbf{y}_k - \mathbf{H}_k \bar{\mathbf{x}}_{k|k-1}),$$

$$\mathbf{P}_{k|k} = (\mathbf{I} - \mathbf{K}_k \mathbf{H}_k) \mathbf{P}_{k|k-1}.$$

は、1 変数の場合、

$$K_k = \frac{H_k P_{k|k-1}}{H_k^2 P_{k|k-1} + R_k}, \quad (20a)$$

$$\bar{x}_{k|k} = \bar{x}_{k|k-1} + K_k (y_k - H_k \bar{x}_{k|k-1}) = \bar{x}_{k|k-1} + \frac{H_k P_{k|k-1} (y_k - H_k \bar{x}_{k|k-1})}{H_k^2 P_{k|k-1} + R_k}, \quad (20b)$$

$$P_{k|k} = (1 - K_k H_k) P_{k|k-1} = \frac{R_k P_{k|k-1}}{H_k^2 P_{k|k-1} + R_k}. \quad (20c)$$

単純な応用例

時系列データ y_1, \dots, y_K が得られていたとして、 y_k を以下のように長期トレンド x_k とノイズ w_k の和で表現する:

$$y_k = x_k + w_k.$$

長期トレンド x_k は、あまり時間的に変化しないものと考え、

$$x_k = x_{k-1} + v_k$$

のように x_k と x_{k-1} の関係を表現する。これは 1 次のトレンドモデルと呼ばれる。まとめると、

$$x_k = x_{k-1} + v_k,$$

$$y_k = x_k + w_k.$$

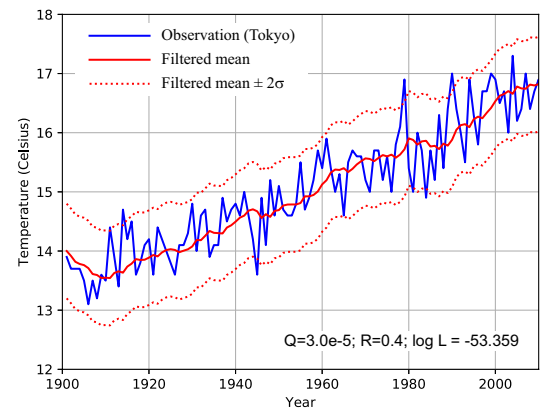
これは、状態空間モデル

$$\mathbf{x}_k = \mathbf{F}_k \mathbf{x}_{k-1} + \mathbf{v}_k,$$

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

で、

$$\mathbf{x}_k = (x_k), \mathbf{F}_k = (1), \mathbf{H}_k = (1)$$



とおいたものに相当する.

$$\mathbf{x}_k = (x_k), \mathbf{F}_k = (1), \mathbf{H}_k = (1)$$

を先程の 1 変数のカルマンフィルタの式に代入すると,

- 一期先予測:

$$\begin{aligned}\bar{x}_{k|k-1} &= \bar{x}_{k-1|k-1}, \\ P_{k|k-1} &= P_{k-1|k-1} + Q_k.\end{aligned}$$

- フィルタリング:

$$\begin{aligned}K_k &= P_{k|k-1} / (P_{k|k-1} + R_k), \\ \bar{x}_{k|k} &= \bar{x}_{k|k-1} + K_k(y_k - \bar{x}_{k|k-1}), \\ P_{k|k} &= (1 - K_k)P_{k|k-1}.\end{aligned}$$

$Q_k \rightarrow \infty$ または $R_k \rightarrow 0$ のとき, $K_k \rightarrow 1$ となるので, $\bar{x}_{k|k} \rightarrow y_k$ となる.

実装上の要点

- カルマンフィルタなど, 状態空間モデルに基づく手法では, $\bar{\mathbf{x}}_{0|0}, \mathbf{P}_{0|0}, \mathbf{Q}_k, \mathbf{R}_k$ を最初に与える必要がある.
- $\bar{\mathbf{x}}_{0|0}, \mathbf{P}_{0|0}$ は, $\mathbf{P}_{0|0}$ の対角要素 (\mathbf{x}_0 の各要素の分散) を大きめに与えておき, ある程度, カルマンフィルタのステップを進めれば, 結果には大きく影響しない.
- $\mathbf{Q}_k, \mathbf{R}_k$ は, 結果に大きく影響しがちだが, データ同化のような大規模な問題では, 場当たりの設定していることも多い.
- また, カルマンフィルタを \mathbf{x}_k が高次元の大規模な問題に適用する際には, 分散共分散行列を計算するためにメモリを多く必要するという問題もある.
- なお, カルマンフィルタで \mathbf{R}_k を a^2 倍したとき,

- 一期先予測:

$$\begin{aligned}\bar{\mathbf{x}}_{k|k-1} &= \mathbf{F}_k \bar{\mathbf{x}}_{k-1|k-1}, \\ \mathbf{P}_{k|k-1} &= \mathbf{F}_k \mathbf{P}_{k-1|k-1} \mathbf{F}_k^\top + \mathbf{Q}_k.\end{aligned}$$

- フィルタリング:

$$\begin{aligned}\mathbf{K}_k &= \mathbf{P}_{k|k-1} \mathbf{H}_k^\top (\mathbf{H}_k \mathbf{P}_{k|k-1} \mathbf{H}_k^\top + \mathbf{R}_k)^{-1}, \\ \bar{\mathbf{x}}_{k|k} &= \bar{\mathbf{x}}_{k|k-1} + \mathbf{K}_k (y_k - \mathbf{H}_k \bar{\mathbf{x}}_{k|k-1}), \\ \mathbf{P}_{k|k} &= (\mathbf{I} - \mathbf{K}_k \mathbf{H}_k) \mathbf{P}_{k|k-1}.\end{aligned}$$

を見ると分かるように, $P_{k|k-1}, \mathbf{Q}_k$ も a^2 倍すれば, 得られる $\bar{\mathbf{x}}_{k|k}$ は同じになる.

- 1 変数の問題なら, $Q_k/R_k, P_{0|0}/R_0$ が同じなら同じ結果が得られる.

パラメータの評価

$\mathbf{Q}_k, \mathbf{R}_k$ のよさを評価する指標としては、周辺尤度と呼ばれるものがある。 $\mathbf{Q}_k, \mathbf{R}_k$ などのパラメータをまとめて θ とおくと、

$$p(\mathbf{y}_{1:K}|\theta) = \prod_{k=1}^K p(\mathbf{y}_k|\mathbf{y}_{1:k-1}, \theta), \quad (21)$$

$$\text{where } p(\mathbf{y}_k|\mathbf{y}_{1:k-1}, \theta) = \int p(\mathbf{y}_k|\mathbf{x}_k, \theta) p(\mathbf{x}_k|\mathbf{y}_{1:k-1}, \theta) d\mathbf{x}_k$$

通常、周辺尤度は計算機で扱えないくらい小さな値になるので、対数を取った

$$\log p(\mathbf{y}_{1:K}|\theta) = \sum_{k=1}^K \log p(\mathbf{y}_k|\mathbf{y}_{1:k-1}, \theta) \quad (22)$$

で評価することが多い。 $p(\mathbf{y}_k|\mathbf{x}_k, \theta), p(\mathbf{x}_k|\mathbf{y}_{1:k-1}, \theta)$ が

$$p(\mathbf{y}_k|\mathbf{x}_k, \theta) = \frac{1}{\sqrt{(2\pi)^n |\mathbf{R}_k|}} \exp\left(-\frac{1}{2}(\mathbf{y}_k - \mathbf{H}_k \mathbf{x}_k)^\top \mathbf{R}_k^{-1} (\mathbf{y}_k - \mathbf{H}_k \mathbf{x}_k)\right),$$

$$p(\mathbf{x}_k|\mathbf{y}_{1:k-1}, \theta) = \frac{1}{\sqrt{(2\pi)^m |\mathbf{P}_{k|k-1}|}} \exp\left(-\frac{1}{2}(\mathbf{x}_k - \bar{\mathbf{x}}_{k|k-1})^\top \mathbf{P}_{k|k-1}^{-1} (\mathbf{x}_k - \bar{\mathbf{x}}_{k|k-1})\right).$$

のように正規分布で書けるとすると、

$$p(\mathbf{y}_k|\mathbf{y}_{1:k-1}, \theta) = \frac{1}{\sqrt{(2\pi)^n |\mathbf{H}_k \mathbf{P}_{k|k-1} \mathbf{H}_k^\top + \mathbf{R}_k|}} \exp\left[-\frac{1}{2}(\mathbf{y}_k - \mathbf{H}_k \bar{\mathbf{x}}_{k|k-1})^\top (\mathbf{H}_k \mathbf{P}_{k|k-1} \mathbf{H}_k^\top + \mathbf{R}_k)^{-1} (\mathbf{y}_k - \mathbf{H}_k \bar{\mathbf{x}}_{k|k-1})\right].$$

となる。その大雑把な説明としては、 $\mathbf{y}'_k = \mathbf{H}_k \mathbf{x}_k$, $\mathbf{w}_k = \mathbf{y}_k - \mathbf{H}_k \mathbf{x}_k$ とおくとよい。

1. 明らかに \mathbf{w}_k は正規分布 $\mathcal{N}(\mathbf{0}, \mathbf{R}_k)$ に従う。
2. また、 \mathbf{x}_k が $\mathcal{N}(\bar{\mathbf{x}}_{k|k-1}, \mathbf{P}_{k|k-1})$ に従うなら、正規分布の性質から \mathbf{y}'_k は正規分布 $\mathcal{N}(\mathbf{H}_k \bar{\mathbf{x}}_{k|k-1}, \mathbf{H}_k \mathbf{P}_{k|k-1} \mathbf{H}_k^\top)$ に従う。

$\mathbf{y}_k = \mathbf{H}_k \mathbf{x}_k + \mathbf{w}_k = \mathbf{y}'_k + \mathbf{w}_k$ から、 \mathbf{y}_k を \mathbf{y}'_k と \mathbf{w}_k の和の確率分布に従うと見なすと、 \mathbf{y}_k は正規分布 $\mathcal{N}(\mathbf{H}_k \bar{\mathbf{x}}_{k|k-1}, \mathbf{H}_k \mathbf{P}_{k|k-1} \mathbf{H}_k^\top + \mathbf{R}_k)$ に従うと考えられる。

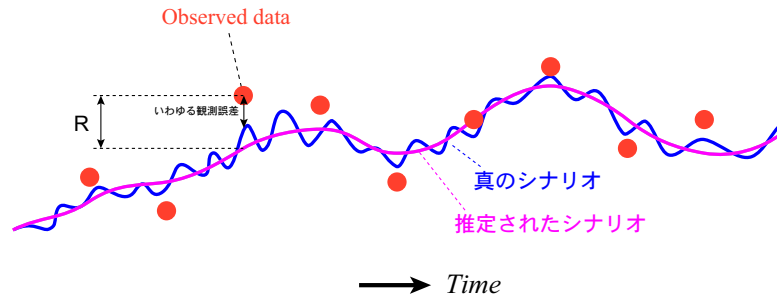
$p(\mathbf{y}_k|\mathbf{y}_{1:k-1}, \theta)$ の対数を取ると、

$$\log p(\mathbf{y}_k|\mathbf{y}_{1:k-1}, \theta) = -\frac{1}{2} \log [(2\pi)^n |\mathbf{H}_k \mathbf{P}_{k|k-1} \mathbf{H}_k^\top + \mathbf{R}_k|] - \frac{1}{2} (\mathbf{y}_k - \mathbf{H}_k \bar{\mathbf{x}}_{k|k-1})^\top (\mathbf{H}_k \mathbf{P}_{k|k-1} \mathbf{H}_k^\top + \mathbf{R}_k)^{-1} (\mathbf{y}_k - \mathbf{H}_k \bar{\mathbf{x}}_{k|k-1})$$

なので、式 (22) の対数周辺尤度は

$$\log p(\mathbf{y}_{1:K}|\theta) = -\frac{1}{2} \sum_{k=1}^K \log [(2\pi)^n |\mathbf{H}_k \mathbf{P}_{k|k-1} \mathbf{H}_k^\top + \mathbf{R}_k|] - \frac{1}{2} \sum_{k=1}^K (\mathbf{y}_k - \mathbf{H}_k \bar{\mathbf{x}}_{k|k-1})^\top (\mathbf{H}_k \mathbf{P}_{k|k-1} \mathbf{H}_k^\top + \mathbf{R}_k)^{-1} (\mathbf{y}_k - \mathbf{H}_k \bar{\mathbf{x}}_{k|k-1}) \quad (23)$$

- 観測ノイズ \mathbf{R}_k は、必ずしも観測機器の測定誤差を意味するものではないことに注意する。
- 通常、シミュレーションモデルには、方程式系を単純化したり、細かいスケールの現象を無視したりしたことによる誤差を持っており、完全に真のシナリオを再現できるわけではない。
- \mathbf{R}_k には、シミュレーションモデルで合わせきれない観測との差も含まれる。



練習問題

- 適当な気象時系列データ (例えば京都の各年の平均気温など) を一つ選び、以下の簡単な 1 変数状態空間モデル

$$x_k = x_{k-1} + v_k,$$

$$y_k = x_k + w_k$$

のもと、カルマンフィルタを適用し、各時点の x_k を推定せよ。 $\mathbf{Q}_k, \mathbf{R}_k$ はデータの特徴に応じて適宜設定すること。

- 気象の時系列データは、例えば、気象庁の HP

<https://www.data.jma.go.jp/obd/stats/etrn/index.php>

にある。(地点などを選ぶとテーブルの形式で表示されるが、例えば、マウスで選んでコピー・アンド・ペーストすることで、Excel に取り込むことができる。

状態空間モデルの構成

状態空間モデルの構成方法の例として、単振動の方程式

$$\frac{d^2 z}{dt^2} = -\alpha^2 z \quad (24)$$

を修正オイラー法で解くシミュレーションモデルから、状態空間モデルを構成してみる。修正オイラー法とは、

$$\frac{dx}{dt} = g(x, t) \quad (25)$$

の形で与えられた常微分方程式を数値的に解くために、時刻 t_k から $t_{k+1}(= t_k + \Delta t)$ の間の \mathbf{x} の値の時間発展を以下のように計算する方法である:

$$\mathbf{x}_{k+1/2} = \mathbf{x}_k + \frac{1}{2} \mathbf{g}(\mathbf{x}_k, t_k) \Delta t, \quad (26)$$

$$\mathbf{x}_{k+1} = \mathbf{x}_k + \mathbf{g}\left(\mathbf{x}_{k+1/2}, t_k + \frac{\Delta t}{2}\right) \Delta t. \quad (27)$$

式 (24) で, $\dot{z} = dz/dt$ とおくと

$$\frac{dz}{dt} = \dot{z}, \quad \frac{d\dot{z}}{dt} = -\alpha^2 z$$

となる. これを修正オイラー法で解くには, 各時間ステップ k において

$$z_{k+1/2} = z_k + \frac{1}{2} \dot{z}_k \Delta t, \quad (28)$$

$$\dot{z}_{k+1/2} = \dot{z}_k - \frac{1}{2} \alpha^2 z_k \Delta t \quad (29)$$

を求めた上で,

$$z_{k+1} = z_k + \dot{z}_{k+1/2} \Delta t, \quad (30)$$

$$\dot{z}_{k+1} = \dot{z}_k - \alpha^2 z_{k+1/2} \Delta t \quad (31)$$

を計算する.

カルマンフィルタを適用するには, 先程の式を $\mathbf{x}_k = \mathbf{F}_k \mathbf{x}_{k-1}$ の形に書き直すと見通しがよくなる. 式 (30), (31) に式 (28), (29) を代入して,

$$z_{k+1} = z_k + \left(\dot{z}_k - \frac{1}{2} \alpha^2 z_k \Delta t \right) \Delta t = \left(1 - \frac{\alpha^2 \Delta t^2}{2} \right) z_k + \dot{z}_k \Delta t, \quad (32)$$

$$\dot{z}_{k+1} = \dot{z}_k - \alpha^2 \left(z_k + \frac{1}{2} \dot{z}_k \Delta t \right) \Delta t = -(\alpha^2 \Delta t) z_k + \left(1 - \frac{\alpha^2 \Delta t^2}{2} \right) \dot{z}_k. \quad (33)$$

まとめると,

$$\begin{pmatrix} z_{k+1} \\ \dot{z}_{k+1} \end{pmatrix} = \begin{pmatrix} 1 - \frac{\alpha^2 \Delta t^2}{2} & \Delta t \\ -\alpha^2 \Delta t & 1 - \frac{\alpha^2 \Delta t^2}{2} \end{pmatrix} \begin{pmatrix} z_k \\ \dot{z}_k \end{pmatrix} \quad (34)$$

となる. これにシステムノイズを付加すれば, システムモデルが得られる. \dot{z}_k に対してのみ, システムノイズ δ_k が加わると仮定して

$$\begin{pmatrix} z_{k+1} \\ \dot{z}_{k+1} \end{pmatrix} = \begin{pmatrix} 1 - \frac{\alpha^2 \Delta t^2}{2} & \Delta t \\ -\alpha^2 \Delta t & 1 - \frac{\alpha^2 \Delta t^2}{2} \end{pmatrix} \begin{pmatrix} z_k \\ \dot{z}_k \end{pmatrix} + \begin{pmatrix} 0 \\ \delta_k \end{pmatrix} \quad (35)$$

とおくことにする. これはシステムモデル $\mathbf{x}_k = \mathbf{F}_k \mathbf{x}_{k-1} + \mathbf{v}_k$ において, 以下のようにおいたことに相当する:

$$\mathbf{x}_k = \begin{pmatrix} z_k \\ \dot{z}_k \end{pmatrix}, \quad \mathbf{F}_k = \begin{pmatrix} 1 - \frac{\alpha^2 \Delta t^2}{2} & \frac{\Delta t}{2} \\ -\alpha^2 \Delta t & 1 - \frac{\alpha^2 \Delta t^2}{2} \end{pmatrix}, \quad \mathbf{v}_k = \begin{pmatrix} 0 \\ \delta_k \end{pmatrix}. \quad (36)$$

観測については, 式 (24) に従う z を一定時間間隔 Δt ごとに観測ノイズ付きで観測できたものとし, 時刻 t_1, \dots, t_K ($t_k = k\Delta t$) に得られた観測データを y_1, \dots, y_K とする. 観測データ y_k と z_k との関係を, 観測ノイズを w_k として,

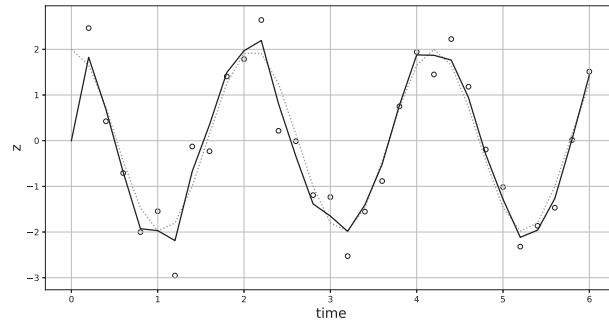
$$y_k = z_k + w_k \quad (37)$$

と書くことにすると,

$$\mathbf{H}_k = \begin{pmatrix} 1 & 0 \end{pmatrix} \quad (38)$$

とおけば,

$$y_k = \mathbf{H}_k \mathbf{x}_k + w_k \quad (39)$$

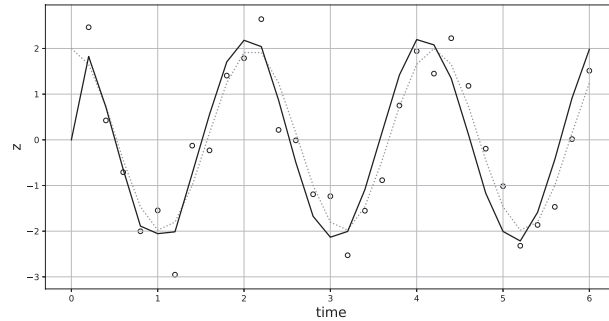


のように観測モデルの形になるので，システムモデルと組み合わせて，カルマンフィルタのアルゴリズムが適用できる．

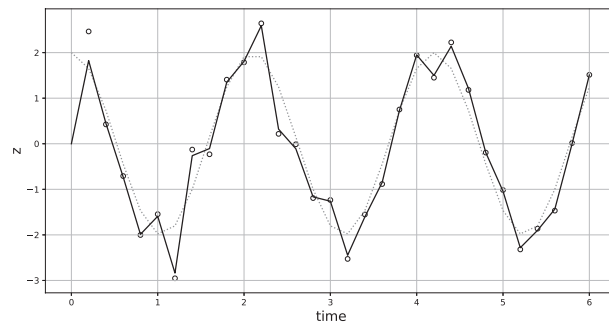
「観測データ」 y_1, \dots, y_K として，式 (24) で初期値を $z = 2, \dot{z} = 0$ とおいたときの解析解

$$z = 2 \cos \frac{5\pi t}{3} \quad (40)$$

から，時間 $\Delta t = 0.1$ ごとに正規分布 $\mathcal{N}(0, 0.3^2)$ にしたがう乱数を観測ノイズとして加えて得た擬似データを与えたときの推定結果を示す．但し， $\sigma_Q = 1$ としている．



カルマンフィルタによる単振動の推定 ($\sigma_Q = 0.1$ の場合)．



カルマンフィルタによる単振動の推定 ($\sigma_Q = 10$ の場合)．