

Лекция 10

Многомерные гауссовские системы.

Калмановская фильтрация

Артемов А. В.
мФТиАД ФКН ВШЭ

19 апреля 2018

В качестве пререквизита к лекции см. Лекцию 2 о многомерных гауссовских системах.

1 Фильтр Калмана

Перейдём к статистике. Для начала введём понятие *оптимальной в среднеквадратическом смысле* (MMSE — Minimum Mean Square Error) оценки.

Определение 1. *Среднеквадратическая ошибка* оценки $\hat{\xi}$ случайного вектора ξ равна

$$\text{MSE} = \text{tr } \mathbb{E}[(\xi - \hat{\xi})(\xi - \hat{\xi})^\top] = \mathbb{E}[(\xi - \hat{\xi})^\top(\xi - \hat{\xi})].$$

Определение 2. Будем называть $\hat{\xi}$ *MMSE-оценкой* случайного вектора ξ , если для неё среднеквадратическая ошибка минимальна.

Теперь приступим к так называемым фильтрам Калмана для линейных систем с дискретным временем. Его вывод будет основан на двух предположениях:

- В гауссовском случае фильтр Калмана является оптимальной оценкой состояния в среднеквадратичном смысле (MMSE state estimator).
- В остальных случаях фильтр Калмана является оптимальной *линейной* оценкой состояния в среднеквадратичном смысле (LMMSE state estimator).

Также далее мы опишем детерменистическую модель (метод наименьших квадратов). Но начнём с описания базового пространства состояний.

1.1 Стохастическое пространство состояний

И сразу же дадим определение.

Определение 3. Линейное (изменяющееся во времени) пространство состояний с дискретным временем задаётся парой уравнений:

$$\begin{aligned} \mathbf{x}_{k+1} &= \mathbf{F}_k \mathbf{x}_k + \mathbf{G}_k \mathbf{w}_k && \text{(уравнение эволюции системы)} \\ \mathbf{z}_{k+1} &= \mathbf{H}_k \mathbf{x}_k + \mathbf{v}_k && \text{(уравнение измерения)} \end{aligned}$$

где

- $\mathbf{F}_k \in \text{Mat}_{n \times n}(\mathbb{R})$, $\mathbf{G}_k \in \text{Mat}_{n \times n_w}(\mathbb{R})$, $\mathbf{H}_k \in \text{Mat}_{m \times n}(\mathbb{R})$ — известные матрицы,
- $\mathbf{x}_k \in \mathbb{R}^n$ — вектор состояния,
- $\mathbf{w}_k \in \mathbb{R}^{n_w}$ — шум состояния,
- $\mathbf{z}_k \in \mathbb{R}^m$ — вектор наблюдений,
- $\mathbf{v}_k \in \mathbb{R}^m$ — шум наблюдений,

Начальным условием для такой системы является \mathbf{x}_0 , которая обычно считается за случайную величину.

Теперь докажем одно свойство:

Свойство 1. Если $\{\mathbf{w}_n\}_{n \in \mathbb{N}}$ является последовательностью независимых случайных величин и она не зависит от \mathbf{x}_0 , то процесс $\{\mathbf{x}_k\}_{k \in \mathbb{N}}$ является марковским.

Доказательство. Раскрыв формулу для \mathbf{x}_{k+1} , получим, что это линейная функция от независимых случайных величин $\mathbf{x}_0, \mathbf{w}_1, \dots, \mathbf{w}_k$. Из этого следует, что \mathbf{w}_k не зависит от \mathbf{x}_i при $i \leq k$. Отсюда следует, что \mathbf{x}_{k+1} разбивается на две части: на $\mathbf{F}_k \mathbf{x}_k$, которая однозначно задаётся значением \mathbf{x}_k , и на $\mathbf{G}_k \mathbf{w}_k$, которая не зависит от $\mathbf{x}_0, \mathbf{x}_1, \dots, \mathbf{x}_k$. Следовательно,

$$P(\mathbf{x}_{k+1} = \mathbf{a}_{k+1} \mid \mathbf{x}_0 = \mathbf{a}_0, \dots, \mathbf{x}_k = \mathbf{a}_k) = P(\mathbf{x}_{k+1} = \mathbf{a}_{k+1} \mid \mathbf{x}_k = \mathbf{a}_k). \quad \square$$

Примечание. На самом деле тут не обязательна линейность — достаточно, что \mathbf{x}_{k+1} есть функция от \mathbf{x}_k и \mathbf{w}_k .

Примечание. Процесс $(\mathbf{z}_n)_{n \in \mathbb{N}}$ обычно не является марковским.

Следствие. В принципе, плотность распределения вектора \mathbf{x}_{k+1} можно считать с помощью аналога обобщённого уравнения Маркова:

$$p_{\mathbf{x}_{k+1}}(\mathbf{y}_{k+1}) = \int \cdots \int_{\mathbb{R}^n} p_{\mathbf{x}_{k+1}|\mathbf{x}_k}(\mathbf{y}_{k+1} \mid \mathbf{y}_k) p_{\mathbf{x}_k}(\mathbf{y}_k) d\mathbf{y}_k,$$

где $p_{\mathbf{x}_{k+1}|\mathbf{x}_k}(\mathbf{y}_{k+1} \mid \mathbf{y}_k)$ задаётся с помощью $p_{\mathbf{w}_k}(\mathbf{y}_{k+1})$.

Теперь рассмотрим так называемые *гауссовские пространства событий*. Они характеризуются тем, что последовательности шумовые последовательности $\{\mathbf{w}_n\}$, $\{\mathbf{v}_n\}$ и начальное условие \mathbf{x}_0 образуют гауссовскую последовательность, то есть имеют совместное нормальное распределение. Отсюда сразу же получаем, что тогда процессы $\{\mathbf{x}_n\}$ и $\{\mathbf{z}_n\}$ тоже являются гауссовскими (как линейное преобразование). Если же выполнено **свойство марковости**, то такие пространства называют *моделями Гаусса-Маркова*.

Требование независимости весьма сильное, поэтому иногда его ослабляют следующим образом:

- Считают, что $\{\mathbf{w}_n\}$ — это белый шум в широком смысле, то есть $E[\mathbf{w}_k] = \mathbf{0}$ и $\text{cov}(\mathbf{w}_k, \mathbf{w}_l) = \mathbf{Q}_k \delta_{kl}$, где δ_{kl} — дельта Кронекера.
- Аналогичное предположение делается относительно $\{\mathbf{v}_n\}$: $E[\mathbf{v}_k] = \mathbf{0}$ и $\text{cov}(\mathbf{v}_k, \mathbf{v}_l) = \mathbf{R}_k \delta_{kl}$.
- Шумы некоррелированы: $\text{cov}(\mathbf{w}_k, \mathbf{v}_l) = \mathbf{0}$.
- \mathbf{x}_0 некоррелирован с шумовыми последовательностями. При этом про \mathbf{x}_0 известно, что $D[\mathbf{x}_0] = \mathbf{P}_0$.

Если выполнены эти условия, то пространство событий называют *стандартным пространством второго порядка*. Иногда бывает полезно позволять корреляцию между шумами:

$$\text{cov}(\mathbf{w}_k, \mathbf{v}_l) = \mathbf{S}_k \delta_{kl}.$$

В таком случае пространство событий называют *пространством второго порядка с коррелированным шумом*.

Всё вышесказанное можно записать одним уравнением:

$$\text{cov} \left(\begin{pmatrix} \mathbf{w}_k \\ \mathbf{v}_k \\ \mathbf{x}_0 \end{pmatrix}, \begin{pmatrix} \mathbf{w}_l \\ \mathbf{v}_l \\ \mathbf{x}_0 \end{pmatrix} \right) = \begin{pmatrix} \mathbf{Q}_k \delta_{kl} & \mathbf{S}_k \delta_{kl} & \mathbf{0} \\ \mathbf{S}_k^\top \delta_{kl} & \mathbf{R}_k \delta_{kl} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{P}_0 \end{pmatrix}$$

Стоит заметить, что модели Гаусса-Маркова являются частным случаем моделей второго порядка с коррелированным шумом.

Для стандартной нормальной модели несложно рекурсивно задать матожидание и дисперсию \mathbf{x}_k . Действительно, по линейности матожидания

$$\mathbf{E}[\mathbf{x}_{k+1}] = \mathbf{F}_k \mathbf{E}[\mathbf{x}_k] + \mathbf{G}_k \mathbf{E}[\mathbf{w}_k] = \mathbf{F}_k \mathbf{E}[\mathbf{x}_k].$$

Далее, введём обозначение $\mathbf{P}_k = \mathbf{D}[\mathbf{x}_k] = \mathbf{E}[(\mathbf{x}_k - \mathbf{E}[\mathbf{x}_k])(\mathbf{x}_k - \mathbf{E}[\mathbf{x}_k])^\top]$. Заметим, что $\mathbf{x}_{k+1} - \mathbf{E}[\mathbf{x}_{k+1}] = \mathbf{F}_k(\mathbf{x}_k - \mathbf{E}[\mathbf{x}_k]) + \mathbf{G}_k \mathbf{w}_k$. Далее, вспомним, что \mathbf{x}_k есть линейная функция от $\mathbf{x}_0, \mathbf{w}_0, \mathbf{w}_1, \dots, \mathbf{w}_{k-1}$. Отсюда следует, что \mathbf{w}_k некоррелировано с \mathbf{x}_k , а, следовательно, и с $\mathbf{x}_{k+1} - \mathbf{E}[\mathbf{x}_{k+1}]$. Следовательно,

$$\begin{aligned} \mathbf{P}_{k+1} &= \mathbf{E}[(\mathbf{F}_k(\mathbf{x}_k - \mathbf{E}[\mathbf{x}_k]) + \mathbf{G}_k \mathbf{w}_k)(\mathbf{F}_k(\mathbf{x}_k - \mathbf{E}[\mathbf{x}_k]) + \mathbf{G}_k \mathbf{w}_k)^\top] = \\ &= \mathbf{E}[\mathbf{F}_k(\mathbf{x}_k - \mathbf{E}[\mathbf{x}_k])(\mathbf{x}_k - \mathbf{E}[\mathbf{x}_k])^\top \mathbf{F}_k^\top] + \mathbf{E}[\mathbf{G}_k \mathbf{w}_k \mathbf{w}_k^\top \mathbf{G}_k^\top] = \\ &= \mathbf{F}_k \mathbf{P}_k \mathbf{F}_k^\top + \mathbf{G}_k \mathbf{Q}_k \mathbf{G}_k^\top. \end{aligned}$$

Полученное уравнение относится к так называемым *разностным уравнениям Ляпунова*.

Теперь посмотрим на \mathbf{z}_k . Так как по определению $\mathbf{z}_{k+1} = \mathbf{H}_k \mathbf{x}_k + \mathbf{v}_k$, где \mathbf{x}_k и \mathbf{v}_k некоррелированы, то несложно посчитать и матожидание, и дисперсию:

$$\begin{aligned} \mathbf{E}[\mathbf{z}_{k+1}] &= \mathbf{H}_k \mathbf{E}[\mathbf{x}_k] + \mathbf{E}[\mathbf{v}_k] = \mathbf{H}_k \mathbf{E}[\mathbf{x}_k] \\ \mathbf{D}[\mathbf{z}_k] &= \mathbf{E}[(\mathbf{H}_k(\mathbf{x}_k - \mathbf{E}[\mathbf{x}_k]) + \mathbf{v}_k)(\mathbf{H}_k(\mathbf{x}_k - \mathbf{E}[\mathbf{x}_k]) + \mathbf{v}_k)^\top] = \\ &= \mathbf{H}_k \mathbf{P}_k \mathbf{H}_k^\top + \mathbf{R}_k. \end{aligned}$$

1.2 Фильтр Калмана в гауссовском случае

Рассмотрим следующую модель Гаусса-Маркова:

$$\begin{aligned} \mathbf{x}_{k+1} &= \mathbf{F}_k \mathbf{x}_k + \mathbf{G}_k \mathbf{w}_k \\ \mathbf{z}_{k+1} &= \mathbf{H}_k \mathbf{x}_k + \mathbf{v}_k \end{aligned}$$

для которой

- $\{\mathbf{w}_n\}$ и $\{\mathbf{v}_n\}$ — независимые белые гауссовские шумы, ковариации которых равны $\text{cov}(\mathbf{w}_k, \mathbf{w}_l) = \mathbf{Q}_k \delta_{kl}$ и $\text{cov}(\mathbf{v}_k, \mathbf{v}_l) = \mathbf{R}_k \delta_{kl}$.
- Начальное состояние системы \mathbf{x}_0 не зависит от шумов и имеет нормальное распределение с дисперсией \mathbf{P}_0 .

Далее, пусть $\mathbf{Z}_k = (\mathbf{z}_0, \dots, \mathbf{z}_k)$. Наша цель — найти рекурсивную формулу для следующей *оптимальной* в среднеквадратичном смысле оценки \mathbf{x}_k :

$$\hat{\mathbf{x}}_k^+ \equiv \hat{\mathbf{x}}_{k|k} = \mathbb{E}[\mathbf{x}_k | \mathbf{Z}_k].$$

Далее, введём *одношаговую оценку* (one-step predictor):

$$\hat{\mathbf{x}}_k^- \equiv \hat{\mathbf{x}}_{k|k-1} = \mathbb{E}[\mathbf{x}_k | \mathbf{Z}_{k-1}].$$

Для них вводятся соответствующие матрицы условных ковариаций:

$$\mathbf{P}_k^+ \equiv \mathbf{P}_{k|k} = \mathbb{D}[\mathbf{x}_k | \mathbf{Z}_k], \quad \mathbf{P}_k^- \equiv \mathbf{P}_{k|k-1} = \mathbb{D}[\mathbf{x}_k | \mathbf{Z}_{k-1}]$$

Примечание. На матрицу \mathbf{P}_k^+ (и, аналогично, на матрицу \mathbf{P}_k^-) можно смотреть, как на

1. матрицу ковариации *постериорной ошибки измерения* $\mathbf{e}_k = \mathbf{x}_k - \hat{\mathbf{x}}_k^+$. Стоит заметить, что

$$\text{MMSE} = \arg \min_{\hat{\mathbf{x}}_k^+} \text{tr} \mathbb{E}[(\mathbf{x} - \hat{\mathbf{x}}_k^+)^{\top} (\mathbf{x} - \hat{\mathbf{x}}_k^+)] = \text{tr} \mathbf{P}_k^+.$$

2. матрицу ковариаций *условной случайной величины* $(\mathbf{x}_k | \mathbf{Z}_k)$.

Напоследок введём два обозначения: $\mathbf{P}_0^- = \mathbf{P}_0$ и $\hat{\mathbf{x}}_0^- = \mathbb{E}[\mathbf{x}_0]$.

Теперь вспомним теорему о нормальной корреляции. Она гласит следующее:

- Если ξ и η имеют совместное нормальное распределение, то $(\xi | \eta)$ “имеет нормальное распределение” $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, где

$$\begin{aligned} \boldsymbol{\mu} &= \boldsymbol{\mu}_{\xi} + \boldsymbol{\Sigma}_{\xi\eta} \boldsymbol{\Sigma}_{\eta\eta}^{-1} (\eta - \boldsymbol{\mu}_{\eta}), \\ \boldsymbol{\Sigma} &= \boldsymbol{\Sigma}_{\xi\xi} - \boldsymbol{\Sigma}_{\xi\eta} \boldsymbol{\Sigma}_{\eta\eta}^{-1} \boldsymbol{\Sigma}_{\eta\xi}. \end{aligned}$$

- Более того, эта теорема корректна и в том случае, когда на всё навешивается условие на какой-то другой случайный вектор.

Теперь запишем одно простое свойство описанной выше модели:

Свойство 1. Все случайные процессы описанной выше модели, то есть шумы, \mathbf{x}_k и \mathbf{z}_k , являются гауссовскими.

Доказательство. Следует из того, что начальные условия и шумы имеют совместное гауссовское распределение (так как независимость и каждый элемент имеет нормальное распределение), и того, что линейное преобразование гауссовского вектора тоже является гауссовским вектором. \square

Отсюда, недолго думая, получаем гауссовость условных случайных величин $(\mathbf{x}_k | \mathbf{Z}_m)$ при всех $k, m \in \mathbb{N}$. Заметим, что

$$(\mathbf{x}_k | \mathbf{Z}_k) \sim \mathcal{N}(\hat{\mathbf{x}}_k^+, \mathbf{P}_k^+), \quad (\mathbf{x}_k | \mathbf{Z}_{k-1}) \sim \mathcal{N}(\hat{\mathbf{x}}_k^-, \mathbf{P}_k^-).$$

Теперь приступим к выводу. Допустим, что нам известна $(\hat{\mathbf{x}}_k^-, \mathbf{P}_k^-)$. Как нам оценить $(\hat{\mathbf{x}}_k^+, \mathbf{P}_k^+)$ и $(\hat{\mathbf{x}}_{k+1}^-, \mathbf{P}_{k+1}^-)$? Это делается следующим образом.

Сначала проходит *шаг обновления измерений*. Вспомним, что $\mathbf{z}_k = \mathbf{H}_k \mathbf{x}_k + \mathbf{v}_k$. Следовательно (почему?),

$$\left(\begin{pmatrix} \mathbf{x}_k \\ \mathbf{z}_k \end{pmatrix} \middle| \mathbf{Z}_{k-1} \right) \sim \mathcal{N} \left(\begin{pmatrix} \hat{\mathbf{x}}_k^- \\ \mathbf{H}_k \hat{\mathbf{x}}_k^- \end{pmatrix}, \begin{pmatrix} \mathbf{P}_k^- & \mathbf{P}_k^- \mathbf{H}_k^{\top} \\ \mathbf{H}_k \mathbf{P}_k^- & \mathbf{M}_k \end{pmatrix} \right),$$

где $\mathbf{M}_k \equiv \mathbf{H}_k \mathbf{P}_k^- \mathbf{H}_k^\top + \mathbf{R}_k$.

Теперь надо определить распределение $(\mathbf{x}_k | \mathbf{Z}_k) = (\mathbf{x}_k | \mathbf{z}_k, \mathbf{Z}_{k-1})$. Для этого воспользуемся теоремой о нормальной корреляции. Тогда $(\mathbf{x}_k | \mathbf{Z}_k)$ имеет нормальное распределение с параметрами

$$\begin{aligned}\widehat{\mathbf{x}}_k^+ &= \mathbb{E}[\mathbf{x}_k | \mathbf{Z}_k] = \widehat{\mathbf{x}}_k^- + \mathbf{P}_k^- (\mathbf{M}_k)^{-1} (\mathbf{z}_k - \mathbf{H}_k \widehat{\mathbf{x}}_k^-) \\ \mathbf{P}_k^+ &= \mathbb{D}[\mathbf{x}_k | \mathbf{Z}_k] = \mathbf{P}_k^- - \mathbf{P}_k^- \mathbf{H}_k^\top (\mathbf{M}_k)^{-1} \mathbf{H}_k \mathbf{P}_k^-.\end{aligned}$$

Дальше идёт *шаг обновления времени*. Теперь вспомним, что $\mathbf{x}_{k+1} = \mathbf{F}_k \mathbf{x}_k + \mathbf{G}_k \mathbf{w}_k$. Далее, заметим, что \mathbf{x}_k и \mathbf{w}_k независимы при условии \mathbf{Z}_k (почему?). Тогда

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

Теперь выпишем несколько примечаний к этому:

- Фильтр Калмана считает и оценку $\widehat{\mathbf{x}}_k^+$, и её матрицу ковариаций \mathbf{P}_k^+ (аналогично для $\widehat{\mathbf{x}}_k^-$). Заметьте, что вычисление ковариации необходимо, так как оно является частью вычисления оценки. Впрочем, оно важно само по себе, так как оно задаёт меру неопределённости (или достоверности) оценки.
- Стоит заметить, что матрицы \mathbf{P}_k^\pm не зависят от измерений $\{\mathbf{z}_k\}$. Следовательно, их можно посчитать заранее, если известны матрицы ковариаций для шума и *системные матрицы*: \mathbf{F}_k , \mathbf{G}_k , \mathbf{H}_k , \mathbf{Q}_k , \mathbf{R}_k и \mathbf{P}_0 .
- В гауссовском случае матрица \mathbf{P}_k^+ является безусловной ковариационной матрицей: $\mathbf{P}_k^+ = \mathbb{D}[\mathbf{x}_k - \widehat{\mathbf{x}}_k^+]$.
В общем случае безусловная ковариация будет играть ключевую роль в выведении линейной оптимальной в среднеквадратичном смысле оценки.
- Допустим, что мы хотим оценить $\mathbf{s}_k = \mathbf{C} \mathbf{x}_k$. В таком случае оптимальной оценкой будет $\widehat{\mathbf{s}}_k = \mathbb{E}[\mathbf{C} \mathbf{x}_k | \mathbf{Z}_k] = \mathbf{C} \widehat{\mathbf{x}}_k^+$.
- Отклонение наблюдения от ожидаемого

$$\tilde{\mathbf{z}}_k \equiv \mathbf{z}_k - \mathbf{H}_k \widehat{\mathbf{x}}_k^- = \mathbf{z}_k - \mathbb{E}[\mathbf{z}_k | \mathbf{Z}_{k-1}]$$

называется *инновацией*, а процесс $\{\tilde{\mathbf{z}}_k\}$ называется *процессом инноваций*. Этот процесс сыграет немалую роль в дальнейшем. К слову: $\mathbf{M}_k = \mathbb{D}[\tilde{\mathbf{z}}_k]$.

1.3 Линейные оценки. Инновационный подход

Теперь перейдём к общему случаю. Но для начала немного пробежимся по линейным оценкам. Докажем одно утверждение:

Теорема 1. Пусть \mathbf{x} , \mathbf{y} — случайные векторы. Тогда оптимальная в среднеквадратичном смысле линейная (LMMSE) оценка \mathbf{x} по \mathbf{y} равна

$$\widehat{\mathbf{x}} = \mathbf{m}_x + \Sigma_{xy} \Sigma_{yy}^{-1} (\mathbf{y} - \mathbf{m}_y),$$

где $\mathbf{m}_x = \mathbb{E}[\mathbf{x}]$, $\Sigma_{xy} = \text{cov}(\mathbf{x}, \mathbf{y})$.

Доказательство. Так как мы ищем линейную оценку, то скажем, что $\hat{\mathbf{x}} = \mathbf{A}^\top \mathbf{y} + \mathbf{b}$ (смысл транспонирования станет ясен немного позднее). Покажем, что оценка должна быть несмещённой. Допустим, что это не так, и $\mathbf{x} - \hat{\mathbf{x}} = \mathbf{a} = (a_1, \dots, a_n)$. Распишем MSE:

$$\text{MSE} = \mathbb{E}[(\mathbf{x} - \hat{\mathbf{x}})^\top (\mathbf{x} - \hat{\mathbf{x}})] = \sum_{k=1}^n \mathbb{E}[(x_k - \hat{x}_k)^2].$$

Отсюда получаем, что достаточно рассмотреть одну компоненту — например, k -ю. Заметим, что $\mathbb{E}[x_k] - \mathbb{E}[\hat{x}_k] = a_k$. Пусть $\tilde{x}_k = (x_k - \mathbb{E}[x_k]) - (\hat{x}_k - \mathbb{E}[\hat{x}_k])$, то есть разность между оцениваемым параметром и несмещённой оценкой. Тогда $x_k - \hat{x}_k = \tilde{x}_k + a_k$. Следовательно,

$$\mathbb{E}[(x_k - \hat{x}_k)^2] = \mathbb{E}[(\tilde{x}_k + a_k)^2] = \mathbb{E}[\tilde{x}_k^2] + a_k^2 + 2a_k \mathbb{E}[\tilde{x}_k] = \mathbb{E}[\tilde{x}_k^2] + a_k^2.$$

Следовательно, $a_k = 0$ и оценка должна быть несмещённой. Тогда можно сказать, что $\mathbb{E}[\mathbf{x}] = \mathbb{E}[\hat{\mathbf{x}}] = \mathbf{0}$.

Распишем среднеквадратическую ошибку, как функцию от матрицы \mathbf{A} :

$$\begin{aligned} \text{MSE}(\mathbf{A}) &= \text{tr} \mathbb{E}[(\mathbf{x} - \hat{\mathbf{x}})(\mathbf{x} - \hat{\mathbf{x}})^\top] = \\ &= \text{tr} (\mathbb{E}[\mathbf{x}\mathbf{x}^\top] - \mathbb{E}[\hat{\mathbf{x}}\mathbf{x}^\top] - \mathbb{E}[\mathbf{x}\hat{\mathbf{x}}^\top] + \mathbb{E}[\hat{\mathbf{x}}\hat{\mathbf{x}}^\top]) = \\ &= \text{tr} (\Sigma_{\mathbf{x}\mathbf{x}} - \mathbf{A}^\top \Sigma_{\mathbf{y}\mathbf{x}} - \Sigma_{\mathbf{x}\mathbf{y}} \mathbf{A} + \mathbf{A}^\top \Sigma_{\mathbf{y}\mathbf{y}} \mathbf{A}) \end{aligned}$$

Теперь продифференцируем по \mathbf{A} . Для этого вспомним три правила матричного дифференцирования (если не знали их ранее, то их несложно вывести покомпонентно):

$$\frac{\partial \text{tr}(\mathbf{X}^\top \mathbf{A})}{\partial \mathbf{X}} = \mathbf{A}, \quad \frac{\partial \text{tr}(\mathbf{A}\mathbf{X})}{\partial \mathbf{X}} = \mathbf{A}^\top, \quad \frac{\partial \text{tr} \mathbf{X}^\top \mathbf{A} \mathbf{X}}{\partial \mathbf{X}} = (\mathbf{A} + \mathbf{A}^\top) \mathbf{X}.$$

Следовательно,

$$\frac{\partial}{\partial \mathbf{A}} \text{MSE}(\mathbf{A}) = -2\Sigma_{\mathbf{y}\mathbf{x}} + 2\Sigma_{\mathbf{y}\mathbf{y}} \mathbf{A} = 0 \implies \mathbf{A} = \Sigma_{\mathbf{y}\mathbf{y}}^{-1} \Sigma_{\mathbf{y}\mathbf{x}}.$$

Отсюда получаем, что $\hat{\mathbf{x}} = \Sigma_{\mathbf{x}\mathbf{y}} \Sigma_{\mathbf{y}\mathbf{y}}^{-1} \mathbf{y}$. Переходя к произвольным значениям матожиданий, получаем желаемое. \square

Рассуждая аналогично доказательству теоремы о нормальной компоненте, можно получить, что

$$\mathbf{D}[\mathbf{x} - \hat{\mathbf{x}}] = \Sigma_{\mathbf{x}\mathbf{x}} - \Sigma_{\mathbf{x}\mathbf{y}} \Sigma_{\mathbf{y}\mathbf{y}}^{-1} \Sigma_{\mathbf{y}\mathbf{x}}.$$

Для удобства введём обозначение $\mathbb{E}^L[\mathbf{x} | \mathbf{y}] \equiv \hat{\mathbf{x}}$. Стоит заметить, что это *не* стандартное обозначение.

Теперь покажем одно свойство этой оценки:

Свойство 1 (принцип ортогональности). Для любой линейной функции $L(\mathbf{y})$ выполнено

$$\mathbb{E}[(\mathbf{x} - \mathbb{E}^L[\mathbf{x} | \mathbf{y}])L(\mathbf{y})^\top] = 0.$$

Доказательство. Так как мы рассматриваем линейные функции от \mathbf{y} , то скажем, что $L(\mathbf{y}) = \mathbf{A}\mathbf{y} + \mathbf{b}$. Далее, заметим, что $\mathbb{E}[\mathbb{E}^L[\mathbf{x} | \mathbf{y}]] = \mathbb{E}[\mathbf{x}]$. Тогда от значения \mathbf{b} ничего не зависит и можно считать, что

$$L(\mathbf{y}) = \mathbf{A}(\mathbf{y} - \mathbb{E}[\mathbf{y}]).$$

Теперь можно сказать, что достаточно смотреть на величины \mathbf{x} и \mathbf{y} с нулевым матожиданием. Теперь же можно просто аккуратно расписать:

$$\begin{aligned} \mathbb{E}[(\mathbf{x} - \Sigma_{\mathbf{x}\mathbf{y}} \Sigma_{\mathbf{y}\mathbf{y}}^{-1} \mathbf{y})(\mathbf{A}\mathbf{y})^\top] &= \mathbb{E}[\mathbf{x}\mathbf{y}^\top] \mathbf{A}^\top - \Sigma_{\mathbf{x}\mathbf{y}} \Sigma_{\mathbf{y}\mathbf{y}}^{-1} \mathbb{E}[\mathbf{y}\mathbf{y}^\top] \mathbf{A}^\top = \\ &= \Sigma_{\mathbf{x}\mathbf{y}} \mathbf{A}^\top - \Sigma_{\mathbf{x}\mathbf{y}} \Sigma_{\mathbf{y}\mathbf{y}}^{-1} \Sigma_{\mathbf{y}\mathbf{y}} \mathbf{A}^\top = \mathbf{0}. \end{aligned}$$

Тем самым мы получили желаемое. \square

Следующее свойство будет крайне полезно в дальнейшем описании. Оно получается, если положить $\mathbf{y} = (\mathbf{y}_1, \mathbf{y}_2)$.

Свойство 2. Пусть $\text{cov}(\mathbf{y}_1, \mathbf{y}_2) = \mathbf{0}$. Тогда

$$\mathbb{E}^L[\mathbf{x} \mid \mathbf{y}_1, \mathbf{y}_2] = \mathbb{E}^L[\mathbf{x} \mid \mathbf{y}_1] + \mathbb{E}^L[\mathbf{x} \mid \mathbf{y}_2] - \mathbb{E}[\mathbf{x}].$$

Более того,

$$\mathbf{D}[\mathbf{x} - \mathbb{E}^L[\mathbf{x} \mid \mathbf{y}_1, \mathbf{y}_2]] = \Sigma_{\mathbf{x}\mathbf{x}} - \Sigma_{\mathbf{x}\mathbf{y}_1} \Sigma_{\mathbf{y}_1\mathbf{y}_1}^{-1} \Sigma_{\mathbf{y}_1\mathbf{x}} - \Sigma_{\mathbf{x}\mathbf{y}_2} \Sigma_{\mathbf{y}_2\mathbf{y}_2}^{-1} \Sigma_{\mathbf{y}_2\mathbf{x}}.$$

Доказательство. Докажем только первое утверждение, оставив второе в качестве упражнения. Заметим, что матрица ковариаций имеет следующий вид:

$$\left(\begin{array}{c|c} \Sigma_{\mathbf{x}\mathbf{x}} & \Sigma_{\mathbf{x}\mathbf{y}} \\ \hline \Sigma_{\mathbf{y}\mathbf{x}} & \Sigma_{\mathbf{y}\mathbf{y}} \end{array} \right) = \left(\begin{array}{c|c} \Sigma_{\mathbf{x}\mathbf{x}} & \Sigma_{\mathbf{x}\mathbf{y}_1} \quad \Sigma_{\mathbf{x}\mathbf{y}_2} \\ \hline \Sigma_{\mathbf{y}_1\mathbf{x}} & \Sigma_{\mathbf{y}_1\mathbf{y}_1} \quad \mathbf{0} \\ \Sigma_{\mathbf{y}_2\mathbf{x}} & \mathbf{0} \quad \Sigma_{\mathbf{y}_2\mathbf{y}_2} \end{array} \right).$$

Теперь посмотрим на $\mathbb{E}^L[\mathbf{x} \mid \mathbf{y}_1, \mathbf{y}_2]$. Оно равно

$$\mathbb{E}^L[\mathbf{x} \mid \mathbf{y}_1, \mathbf{y}_2] = \mathbb{E}[\mathbf{x}] + \Sigma_{\mathbf{x}\mathbf{y}} \Sigma_{\mathbf{y}\mathbf{y}}^{-1} (\mathbf{y} - \mathbb{E}[\mathbf{y}]).$$

Подставим полученные выше блочные матрицы:

$$\mathbb{E}^L[\mathbf{x} \mid \mathbf{y}_1, \mathbf{y}_2] = \mathbb{E}[\mathbf{x}] + (\Sigma_{\mathbf{x}\mathbf{y}_1} \quad \Sigma_{\mathbf{x}\mathbf{y}_2}) \begin{pmatrix} \Sigma_{\mathbf{y}_1\mathbf{y}_1} & \mathbf{0} \\ \mathbf{0} & \Sigma_{\mathbf{y}_2\mathbf{y}_2} \end{pmatrix}^{-1} (\mathbf{y} - \mathbb{E}[\mathbf{y}]).$$

Теперь вспомним, что обратная к диагональной блочной матрице равна матрице с обратными блоками. Следовательно, это равно

$$\mathbb{E}^L[\mathbf{x} \mid \mathbf{y}_1, \mathbf{y}_2] = \mathbb{E}[\mathbf{x}] + (\Sigma_{\mathbf{x}\mathbf{y}_1} \quad \Sigma_{\mathbf{x}\mathbf{y}_2}) \begin{pmatrix} \Sigma_{\mathbf{y}_1\mathbf{y}_1}^{-1} & \mathbf{0} \\ \mathbf{0} & \Sigma_{\mathbf{y}_2\mathbf{y}_2}^{-1} \end{pmatrix} (\mathbf{y} - \mathbb{E}[\mathbf{y}]).$$

Осталось заметить, что произведение матриц ведёт себя, как квадратичная форма (и такое рассмотрение корректно из-за совпадений размеров матриц). Тогда

$$\begin{aligned} (\Sigma_{\mathbf{x}\mathbf{y}_1} \quad \Sigma_{\mathbf{x}\mathbf{y}_2}) \begin{pmatrix} \Sigma_{\mathbf{y}_1\mathbf{y}_1}^{-1} & \mathbf{0} \\ \mathbf{0} & \Sigma_{\mathbf{y}_2\mathbf{y}_2}^{-1} \end{pmatrix} \begin{pmatrix} \mathbf{y}_1 - \mathbb{E}[\mathbf{y}_1] \\ \mathbf{y}_2 - \mathbb{E}[\mathbf{y}_2] \end{pmatrix} &= \Sigma_{\mathbf{x}\mathbf{y}_1} \Sigma_{\mathbf{y}_1\mathbf{y}_1}^{-1} (\mathbf{y}_1 - \mathbb{E}[\mathbf{y}_1]) + \\ &+ \Sigma_{\mathbf{x}\mathbf{y}_2} \Sigma_{\mathbf{y}_2\mathbf{y}_2}^{-1} (\mathbf{y}_2 - \mathbb{E}[\mathbf{y}_2]) \end{aligned}$$

Отсюда получаем желаемое. \square

Теперь рассмотрим случайную последовательность $\{\mathbf{z}_k\}_{k \geq 0}$. По ней можно построить так называемую *последовательность инноваций* (в широком смысле):

$$\tilde{\mathbf{z}}_k = \mathbf{z}_k - \mathbb{E}^L[\mathbf{z}_k \mid \mathbf{Z}_{k-1}].$$

На условную случайную величину $\tilde{\mathbf{z}}_k$ можно смотреть, как на случайную величину, содержащую только новую статистическую информацию, которой нет в \mathbf{Z}_{k-1} .

Пользуясь свойствами LMMSE-оценки, можно сразу же выписать три свойства:

- $\mathbb{E}[\tilde{\mathbf{z}}] = \mathbf{0}$.
- $\tilde{\mathbf{z}}_k$ есть линейная функция от \mathbf{Z}_k .

- Из второго пункта следует, что $\text{cov}(\tilde{\mathbf{z}}_k, \tilde{\mathbf{z}}_l) = \mathbf{0}$ при $k \neq l$.

Тогда можно сказать, что $\{\mathbf{z}_k\}_{k \geq 0}$ является *белым шумом в широком смысле*.

Теперь введём обозначение $\tilde{\mathbf{Z}}_k = (\tilde{\mathbf{z}}_1, \dots, \tilde{\mathbf{z}}_k)$. Сразу же заметим, что $\tilde{\mathbf{Z}}_k$ есть линейная функция от \mathbf{Z}_k (так как при построении берутся только линейные преобразования). Следовательно, $\mathbb{E}^L[\mathbf{x} \mid \mathbf{Z}_{k-1}] = \mathbb{E}^L[\mathbf{x} \mid \tilde{\mathbf{Z}}_{k-1}]$ для любой случайной величины \mathbf{x} . Действительно, если $\tilde{\mathbf{Z}}_k = \mathbf{A}\mathbf{Z}_k + \mathbf{b}$, то

$$\begin{aligned} \mathbb{E}^L[\mathbf{x} \mid \tilde{\mathbf{Z}}_k] &= \mathbb{E}[\mathbf{x}] + \text{cov}(\mathbf{x}, \mathbf{A}\mathbf{Z}_k + \mathbf{b}) \mathbf{D}[\mathbf{A}\mathbf{Z}_k + \mathbf{b}]^{-1} (\mathbf{A}\mathbf{Z}_k + \mathbf{b} - \mathbb{E}[\mathbf{A}\mathbf{Z}_k + \mathbf{b}]) = \\ &= \mathbb{E}[\mathbf{x}] + \text{cov}(\mathbf{x}, \mathbf{Z}_k) \mathbf{A}^\top (\mathbf{A} \mathbf{D}[\mathbf{Z}_k] \mathbf{A}^\top)^{-1} \mathbf{A} (\mathbf{Z}_k - \mathbb{E}[\mathbf{Z}_k]) = \\ &= \mathbb{E}[\mathbf{x}] + \Sigma_{\mathbf{x}\mathbf{Z}_k} \mathbf{A}^\top (\mathbf{A}^\top)^{-1} \Sigma_{\mathbf{Z}_k\mathbf{Z}_k}^{-1} \mathbf{A}^{-1} \mathbf{A} (\mathbf{Z}_k - \mathbb{E}[\mathbf{Z}_k]) = \\ &= \mathbb{E}[\mathbf{x}] + \Sigma_{\mathbf{x}\mathbf{Z}_k} \Sigma_{\mathbf{Z}_k\mathbf{Z}_k}^{-1} (\mathbf{Z}_k - \mathbb{E}[\mathbf{Z}_k]) = \mathbb{E}^L[\mathbf{x} \mid \mathbf{Z}_k] \end{aligned}$$

Из этого следует, что (если $\mathbb{E}[\mathbf{x}] = 0$)

$$\mathbb{E}^L[\mathbf{x} \mid \mathbf{Z}_k] = \mathbb{E}^L[\mathbf{x} \mid \tilde{\mathbf{Z}}_k] = \mathbb{E}^L[\mathbf{x} \mid \tilde{\mathbf{Z}}_{k-1}] + \mathbb{E}^L[\mathbf{x} \mid \tilde{\mathbf{z}}_k] = \sum_{i=0}^k \mathbb{E}^L[\mathbf{x} \mid \tilde{\mathbf{z}}_i]$$

В принципе, вся необходимая теория была введена, так что можно начинать вводить фильтр Калмана для линейных негауссовских моделей. Мы немного обобщим задачу, решив корреляцию между шумами. Рассмотрим следующую модель:

$$\begin{aligned} \mathbf{x}_{k+1} &= \mathbf{F}_k \mathbf{x}_k + \mathbf{G}_k \mathbf{w}_k \\ \mathbf{z}_{k+1} &= \mathbf{H}_k \mathbf{x}_k + \mathbf{v}_k \end{aligned}$$

Скажем, что $\{\mathbf{w}_k\}$ и $\{\mathbf{v}_k\}$ — белые шумы в широком смысле с нулевым матожиданием и ковариацией

$$\text{cov} \left(\begin{pmatrix} \mathbf{w}_k \\ \mathbf{v}_k \end{pmatrix}, \begin{pmatrix} \mathbf{w}_l \\ \mathbf{v}_l \end{pmatrix} \right) = \begin{pmatrix} \mathbf{Q}_k & \mathbf{S}_k \\ \mathbf{S}_k^\top & \mathbf{R}_k \end{pmatrix} \delta_{kl}$$

Далее, начальное условие \mathbf{x}_0 имеет матрицу ковариаций \mathbf{P}_0 и не коррелирует с шумами.

Теперь введём несколько обозначений:

$$\begin{aligned} \mathbf{Z}_k &= (\mathbf{z}_1, \dots, \mathbf{z}_k), \\ \hat{\mathbf{x}}_{k|k} &= \mathbb{E}^L[\mathbf{x}_k \mid \mathbf{Z}_k], & \hat{\mathbf{x}}_{k|k-1} &= \mathbb{E}^L[\mathbf{x}_k \mid \mathbf{Z}_{k-1}], \\ \tilde{\mathbf{x}}_{k|k} &= \mathbf{x}_k - \hat{\mathbf{x}}_{k|k}, & \tilde{\mathbf{x}}_{k|k-1} &= \mathbf{x}_k - \hat{\mathbf{x}}_{k|k-1}, \\ \mathbf{P}_{k|k} &= \mathbf{D}[\hat{\mathbf{x}}_{k|k}], & \mathbf{P}_{k|k-1} &= \mathbf{D}[\hat{\mathbf{x}}_{k|k-1}]. \end{aligned}$$

Инновации будем вводить, как и раньше:

$$\tilde{\mathbf{z}}_k = \mathbf{z}_k - \mathbb{E}^L[\mathbf{z}_k \mid \mathbf{Z}_{k-1}].$$

Заметим, что второй член можно преобразовать:

$$\begin{aligned} \mathbb{E}^L[\mathbf{z}_k \mid \mathbf{Z}_{k-1}] &= \mathbb{E}[\mathbf{z}_k] + \Sigma_{\mathbf{z}_k\mathbf{Z}_{k-1}} \Sigma_{\mathbf{Z}_{k-1}\mathbf{Z}_{k-1}}^{-1} (\mathbf{Z}_{k-1} - \mathbb{E}[\mathbf{Z}_{k-1}]) = \\ &= \mathbf{H}_k \mathbb{E}[\mathbf{x}_k] + \mathbf{H}_k \Sigma_{\mathbf{x}_k\mathbf{Z}_{k-1}} \Sigma_{\mathbf{Z}_{k-1}\mathbf{Z}_{k-1}}^{-1} (\mathbf{Z}_{k-1} - \mathbb{E}[\mathbf{Z}_{k-1}]) = \\ &= \mathbf{H}_k \mathbb{E}^L[\mathbf{x}_k \mid \mathbf{Z}_{k-1}] = \mathbf{H}_k \hat{\mathbf{x}}_{k|k-1}. \end{aligned}$$

Пользуясь этим, получаем, что

$$\tilde{\mathbf{z}}_k = \mathbf{z}_k - \mathbf{H}_k \hat{\mathbf{x}}_{k|k-1} = \mathbf{H}_k \tilde{\mathbf{x}}_{k|k-1} + \mathbf{v}_k.$$

Приступим к описанию самого фильтра. Он состоит из шага обновления измерения, шага обновления времени. Иногда вводят шаг обновления ковариации.

1. Начнём с шага обновления измерения, то есть с перехода от $\hat{\mathbf{x}}_{k|k-1}$ к $\hat{\mathbf{x}}_{k|k}$. Пользуясь свойством 2 и тем, что $E^L[\mathbf{x} | \mathbf{Z}_k] = E^L[\mathbf{x} | \tilde{\mathbf{Z}}_k]$, получаем, что

$$\hat{\mathbf{x}}_{k|k} = E^L[\mathbf{x}_k | \mathbf{Z}_k] = E^L[\mathbf{x}_k | \tilde{\mathbf{Z}}_k] = E^L[\mathbf{x}_k | \tilde{\mathbf{Z}}_{k-1}] + E^L[\mathbf{x}_k | \tilde{\mathbf{z}}_k] - E[\mathbf{x}_k].$$

Это выражение и лежит в основе инновационного подхода. Остальное следует из прямых вычислений и принципа ортогональности. Для начала заметим, что

$$E^L[\mathbf{x}_k | \tilde{\mathbf{z}}_k] - E[\mathbf{x}_k] = \text{cov}(\mathbf{x}_k, \tilde{\mathbf{z}}_k) D[\tilde{\mathbf{z}}_k]^{-1} \tilde{\mathbf{z}}_k.$$

Теперь считаем эти ковариации. Начнём с первой:

$$\text{cov}(\mathbf{x}_k, \tilde{\mathbf{z}}_k) = \text{cov}(\mathbf{x}_k, \mathbf{H}_k \tilde{\mathbf{x}}_{k|k-1} + \mathbf{v}_k) = \text{cov}(\mathbf{x}_k, \tilde{\mathbf{x}}_{k|k-1}) \mathbf{H}_k^\top + \text{cov}(\mathbf{x}_k, \mathbf{v}_k).$$

Распишем первый член подробнее. Заметим, что $\hat{\mathbf{x}}_k$ есть несмещённая оценка \mathbf{x}_k . Тогда, добавив принцип ортогональности, получаем, что

$$\begin{aligned} \text{cov}(\mathbf{x}_k, \tilde{\mathbf{x}}_{k|k-1}) &= E[(\mathbf{x}_k - E[\mathbf{x}_k])(\mathbf{x}_k - \hat{\mathbf{x}}_{k|k-1})^\top] = E[\mathbf{x}_k(\mathbf{x}_k - \hat{\mathbf{x}}_{k|k-1})^\top] = \\ &= E[\mathbf{x}_k(\mathbf{x}_k - \hat{\mathbf{x}}_{k|k-1})^\top] + (E[(\mathbf{x}_k - \hat{\mathbf{x}}_k)\hat{\mathbf{x}}_{k|k-1}^\top])^\top = \\ &= D[\mathbf{x}_k - \hat{\mathbf{x}}_{k|k-1}] = \mathbf{P}_{k|k-1}. \end{aligned}$$

Следовательно, так как \mathbf{x}_k не коррелирует с \mathbf{v}_k , то $\text{cov}(\mathbf{x}_k, \tilde{\mathbf{z}}_k) = \mathbf{P}_{k|k-1} \mathbf{H}_k^\top$. Далее заметим, что $\hat{\mathbf{x}}_{k-1}$ тоже не коррелирует с \mathbf{v}_k , так как зависит только от \mathbf{Z}_{k-1} и \mathbf{x}_k . Следовательно,

$$D[\tilde{\mathbf{z}}_k] = D[\mathbf{H}_k \tilde{\mathbf{x}}_{k|k-1} + \mathbf{v}_k] = \mathbf{H}_k \mathbf{P}_{k|k-1} \mathbf{H}_k^\top + \mathbf{R}_k \equiv \mathbf{M}_k.$$

Собирая результаты выше в один большой результат, получаем, что

$$\hat{\mathbf{x}}_{k|k} = \hat{\mathbf{x}}_{k|k-1} + \mathbf{P}_{k|k-1} \mathbf{H}_k^\top \mathbf{M}_k^{-1} \tilde{\mathbf{z}}_k.$$

2. Теперь приступим к шагу обновления времени, то есть к переходу от $\hat{\mathbf{x}}_{k|k}$ к $\hat{\mathbf{x}}_{k+1|k}$. Распишем:

$$\begin{aligned} \hat{\mathbf{x}}_{k+1|k} &= E^L[\mathbf{x}_{k+1} | \mathbf{Z}_k] = E^L[\mathbf{F}_k \mathbf{x}_k + \mathbf{G}_k \mathbf{w}_k | \mathbf{Z}_k] = \\ &= E[\mathbf{F}_k \mathbf{x}_k + \mathbf{G}_k \mathbf{w}_k] + \text{cov}(\mathbf{F}_k \mathbf{x}_k + \mathbf{G}_k \mathbf{w}_k, \mathbf{Z}_k) D[\mathbf{Z}_k]^{-1} (\mathbf{Z}_k - E[\mathbf{Z}_k]) = \\ &= \mathbf{F}_k E^L[\mathbf{x}_k | \mathbf{Z}_k] + \mathbf{G}_k E^L[\mathbf{w}_k | \mathbf{Z}_k]. \end{aligned}$$

Дальше воспользуемся тем, что \mathbf{w}_k не коррелирует с \mathbf{Z}_{k-1} , а, следовательно, и с $\tilde{\mathbf{Z}}_{k-1}$. Тогда это равно

$$\begin{aligned} \hat{\mathbf{x}}_{k+1|k} &= \mathbf{F}_k E^L[\mathbf{x}_k | \mathbf{Z}_k] + \mathbf{G}_k E^L[\mathbf{w}_k | \tilde{\mathbf{Z}}_k] = \\ &= \mathbf{F}_k \hat{\mathbf{x}}_{k|k} + \mathbf{G}_k (E^L[\mathbf{w}_k | \tilde{\mathbf{z}}_k] + E^L[\mathbf{w}_k | \tilde{\mathbf{Z}}_{k-1}]) = \\ &= \mathbf{F}_k \hat{\mathbf{x}}_{k|k} + \mathbf{G}_k E^L[\mathbf{w}_k | \tilde{\mathbf{z}}_k]. \end{aligned}$$

Осталось посчитать $E^L[\mathbf{w}_k | \tilde{\mathbf{z}}_k] = E[\mathbf{w}_k] + \text{cov}(\mathbf{w}_k, \tilde{\mathbf{z}}_k) D[\tilde{\mathbf{z}}_k]^{-1} \tilde{\mathbf{z}}_k$. Распишем ковариацию подробнее, пользуясь тем, что и \mathbf{x}_k , и $\hat{\mathbf{x}}_{k|k-1}$ не коррелируют с \mathbf{w}_k :

$$\begin{aligned} \text{cov}(\mathbf{w}_k, \tilde{\mathbf{z}}_k) &= \text{cov}(\mathbf{w}_k, \mathbf{H}_k \tilde{\mathbf{x}}_{k|k-1} + \mathbf{v}_k) = \text{cov}(\mathbf{w}_k, \tilde{\mathbf{x}}_{k|k-1}) \mathbf{H}_k^\top + \mathbf{S}_k = \\ &= \text{cov}(\mathbf{w}_k, \mathbf{x}_k) - \text{cov}(\mathbf{w}_k, \hat{\mathbf{x}}_k) + \mathbf{S}_k = \mathbf{S}_k. \end{aligned}$$

Следовательно, шаг обновления времени записывается так:

$$\hat{\mathbf{x}}_{k+1|k} = \mathbf{F}_k \hat{\mathbf{x}}_{k|k} + \mathbf{G}_k \mathbf{S}_k \mathbf{M}_k^{-1} \tilde{\mathbf{z}}_k.$$

3. Предыдущие шаги можно скомпоновать в один:

$$\hat{\mathbf{x}}_{k+1|k} = \mathbf{F}_k \hat{\mathbf{x}}_{k|k-1} + \mathbf{K}_k \tilde{\mathbf{z}}_k,$$

$$\text{где } \mathbf{K}_k = (\mathbf{F}_k \mathbf{P}_{k|k-1} \mathbf{H}_k^\top + \mathbf{G}_k \mathbf{S}_k) \mathbf{M}_k^{-1}.$$

4. Теперь рассмотрим шаг обновления ковариации, то есть переход от $\mathbf{P}_{k|k-1}$ к $\mathbf{P}_{k+1|k}$. Стоит сказать, что переход от $\mathbf{P}_{k|k-1}$ к $\mathbf{P}_{k|k}$ полностью соответствует гауссовскому случаю, поэтому мы сразу смотрим комбинацию переходов. Заметим, что

$$\begin{aligned} \tilde{\mathbf{x}}_{k+1|k} &= \mathbf{x}_{k+1} - \hat{\mathbf{x}}_{k+1|k} = \mathbf{x}_{k+1} - \mathbf{F}_k \hat{\mathbf{x}}_{k|k-1} - \mathbf{K}_k \tilde{\mathbf{z}}_k = \\ &= \mathbf{F}_k \mathbf{x}_k + \mathbf{G}_k \mathbf{w}_k - \mathbf{F}_k \hat{\mathbf{x}}_{k|k-1} - \mathbf{K}_k (\mathbf{H}_k \tilde{\mathbf{x}}_{k|k-1} + \mathbf{v}_k) = \\ &= (\mathbf{F}_k - \mathbf{K}_k \mathbf{H}_k) \tilde{\mathbf{x}}_{k|k-1} + \mathbf{G}_k \mathbf{w}_k - \mathbf{K}_k \mathbf{v}_k. \end{aligned}$$

Теперь, пользуясь некоррелированностью $\tilde{\mathbf{x}}_{k|k-1}$ с \mathbf{w}_k и \mathbf{v}_k , получаем, что

$$\begin{aligned} \mathbf{P}_{k+1|k} &= \mathbf{D}[(\mathbf{F}_k - \mathbf{K}_k \mathbf{H}_k) \tilde{\mathbf{x}}_{k|k-1}] + \mathbf{D}[\mathbf{G}_k \mathbf{w}_k - \mathbf{K}_k \mathbf{v}_k] = \\ &= (\mathbf{F}_k - \mathbf{K}_k \mathbf{H}_k) \mathbf{P}_{k|k-1} (\mathbf{F}_k - \mathbf{K}_k \mathbf{H}_k)^\top + \mathbf{D}[\mathbf{G}_k \mathbf{w}_k - \mathbf{K}_k \mathbf{v}_k]. \end{aligned}$$

Значение второго члена получить несложно, но формула для него достаточно длинная. Запишем её без доказательства:

$$\mathbf{D}[\mathbf{G}_k \mathbf{w}_k - \mathbf{K}_k \mathbf{v}_k] = \mathbf{G}_k \mathbf{Q}_k \mathbf{G}_k^\top + \mathbf{K}_k \mathbf{R}_k \mathbf{K}_k^\top - \mathbf{G}_k \mathbf{S}_k \mathbf{K}_k^\top - \mathbf{K}_k \mathbf{S}_k^\top \mathbf{G}_k^\top.$$

Определения и результаты, связанные с линейными оценками, можно хорошо интерпретировать с помощью гильбертовых пространств.

Пусть, для простоты, все случайные величины (то есть, \mathbf{w}_k , \mathbf{v}_k и \mathbf{x}_0) имеют нулевое матожидание.

Напомним, что гильбертово пространство — это полное пространство со скалярным произведением. Другими словами, это линейное пространство V (для простоты скажем, что над \mathbb{R}) с определённой на нём операцией скалярного произведения $\langle \cdot, \cdot \rangle : V \times V \mapsto \mathbb{R}$, которая обладает следующими свойствами:

1. Оно линейно по первому аргументу: для любых $\alpha, \beta \in \mathbb{R}$, $x, y, z \in V$

$$\langle \alpha x + \beta y, z \rangle = \alpha \langle x, z \rangle + \beta \langle y, z \rangle.$$

2. Оно симметрично: для любых $x, y \in V$ $\langle x, y \rangle = \langle y, x \rangle$.

3. Оно положительно определено: для любого $x \in V$ $\langle x, x \rangle \geq 0$, причём $\langle x, x \rangle = 0 \iff x = 0$.

Полнота означает, что любая фундаментальная (т.е. для которой выполнено условие Коши) последовательность имеет предел. Далее, это скалярное произведение порождает норму: $\|x\| = \sqrt{\langle x, x \rangle}$. Теперь выпишем несколько стандартных утверждений из линейной алгебры:

- *Подпространство* S — это замкнутое относительно линейных преобразований подмножество V . Другими словами, оно является линейной оболочкой каких-то векторов $\{v_\alpha\}$.
- *Ортогональная проекция* $\Pi_S v$ вектора v на подпространство S — это ближайший к v элемент из S , то есть это вектор $w \in S$, который минимизирует $\|v - w\|$. Такой вектор действительно существует и для него верно, что $v - \Pi_S v \perp S$, то есть $\langle v - \Pi_S v, w \rangle = 0$ для любого $w \in S$.

- Если $S = \text{span}(s_1, \dots, s_n)$, то

$$\Pi_S v = \sum_{k=1}^n \alpha_k s_k, \text{ где } (\alpha_1 \dots \alpha_n) = (\langle v, s_1 \rangle \dots \langle v, s_n \rangle) \begin{pmatrix} \langle s_1, s_1 \rangle & \dots & \langle s_1, s_n \rangle \\ \vdots & \ddots & \vdots \\ \langle s_n, s_1 \rangle & \dots & \langle s_n, s_n \rangle \end{pmatrix}^{-1}$$

Если (s_1, \dots, s_n) — это ортогональный базис S , то

$$\Pi_S v = \sum_{k=1}^n \langle v, s_k \rangle \langle s_k, s_k \rangle^{-1} s_k.$$

- Если $S = S_1 \oplus S_2$ (то есть, S есть прямая сумма двух ортогональных подпространств S_1 и S_2), то $\Pi_S v = \Pi_{S_1} v + \Pi_{S_2} v$.
- Если есть набор линейно независимых векторов $\{v_1, v_2, \dots\}$, то его можно превратить в ортогональный базис, используя *процесс Грама-Шмидта*:

$$\tilde{v}_k = v_k - \Pi_{\text{span}(v_1, \dots, v_{k-1})} v_k = v_k - \sum_{i=1}^{k-1} \langle v_k, \tilde{v}_i \rangle \langle \tilde{v}_i, \tilde{v}_i \rangle^{-1} \tilde{v}_i.$$

Теперь можно провести аналогию между полученными ранее результатами, связанными с линейными оценками, с этими фактами, заметив следующее:

- В качестве гильбертова пространства возьмём пространство случайных векторов, для которых $E[\mathbf{x}] = \mathbf{0}$ и $E[\mathbf{x}^\top \mathbf{x}] < +\infty$. В таком случае скалярное произведение задаётся следующим образом: $\langle \mathbf{x}, \mathbf{y} \rangle = E[\mathbf{x}^\top \mathbf{y}] = \text{tr } E[\mathbf{x} \mathbf{y}^\top]$.
- Оптимальная линейная оценка $E^L[\mathbf{x} | \mathbf{Z}_k]$ является ничем иным, как ортогональная проекция \mathbf{x} на линейную оболочку векторов $(\mathbf{z}_1, \dots, \mathbf{z}_k)$ (почему?).
- Процесс инноваций $\{\tilde{\mathbf{z}}_k\}$ есть ортогонализованная версия процесса $\{\mathbf{z}_k\}$.

Формулировка через гильбертовы пространства может дать несколько хороших идей и результатов (особенно в случае с непрерывным временем, так называемым фильтром Калмана-Бьюси). Но мы остановимся на этом.

1.4 Фильтр Калмана и метод наименьших квадратов

Рассмотрим следующую задачу оптимизации: на переменные $\mathbf{x}_0, \dots, \mathbf{x}_k$ и $\mathbf{w}_0, \dots, \mathbf{w}_{k-1}$ наложены следующие ограничения:

$$\mathbf{x}_{i+1} = \mathbf{F}_i \mathbf{x}_i + \mathbf{G}_i \mathbf{w}_i, \quad i \in \{0, 1, \dots, k-1\}.$$

Далее, платёжная функция, которую нужно минимизировать, равна

$$J_k = \frac{1}{2} (\mathbf{x}_0 - \bar{\mathbf{x}}_0)^\top \mathbf{P}_0^{-1} (\mathbf{x}_0 - \bar{\mathbf{x}}_0) + \frac{1}{2} \sum_{i=0}^k (\mathbf{z}_i - \mathbf{H}_i \mathbf{x}_i)^\top \mathbf{R}_i^{-1} (\mathbf{z}_i - \mathbf{H}_i \mathbf{x}_i) + \frac{1}{2} \sum_{i=0}^{k-1} \mathbf{w}_i^\top \mathbf{Q}_i^{-1} \mathbf{w}_i.$$

В этой задаче $\bar{\mathbf{x}}_0, \{\mathbf{z}_k\}$ — это известные векторы, а $\mathbf{P}_0, \mathbf{R}_k$ и \mathbf{Q}_k — положительно определённые симметричные матрицы.

Пусть $\mathbf{x}_0^{(k)}, \dots, \mathbf{x}_k^{(k)}$ — это оптимальное решение задачи. Утверждается, что $\mathbf{x}_k^{(k)}$ можно вычислить точно так же, как вычисляется $\hat{\mathbf{x}}_{k|k}$ в фильтре Калмана.

Этот результат можно получить в лоб, расписывая решение методом наименьших квадратов для $k - 1$ и k и преобразованиями матриц. Но мы пойдём проще, используя гауссовский подход.

Теорема 2. *Оптимальное решение $\mathbf{x}_0^{(k)}, \dots, \mathbf{x}_k^{(k)}$ описанной выше задачи оптимизации является максимизатором условной вероятности (другими словами, это оценка апостериорного максимума, или MAP-оценка)*

$$(\mathbf{x}_0^{(k)}, \dots, \mathbf{x}_k^{(k)}) = \arg \max_{\mathbf{x}_0, \mathbf{x}_1, \dots, \mathbf{x}_k} p(\mathbf{x}_0, \mathbf{x}_1, \dots, \mathbf{x}_k \mid \mathbf{Z}_k),$$

которая связана с следующей гауссовской моделью:

$$\begin{aligned} \mathbf{x}_{k+1} &= \mathbf{F}_k \mathbf{x}_k + \mathbf{G}_k \mathbf{w}_k, & \mathbf{x}_0 &\sim \mathcal{N}(\mathbf{0}, \mathbf{P}_0), \mathbf{w}_k \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}_k) \\ \mathbf{z}_{k+1} &= \mathbf{H}_k \mathbf{x}_k + \mathbf{v}_k & \mathbf{v}_k &\sim \mathcal{N}(\mathbf{0}, \mathbf{R}_k) \end{aligned}$$

Для этой модели $\{\mathbf{w}_k\}$ и $\{\mathbf{v}_k\}$ — белые шумы, не зависящие от начального условия \mathbf{x}_0 , матожидание которого равно $\bar{\mathbf{x}}_0$.

Указание. Распишите $p(\mathbf{x}_0, \mathbf{x}_1, \dots, \mathbf{x}_k, \mathbf{Z}_k)$.

Теперь нужно указать ещё один факт (без доказательства):

Теорема 3. *Для таких гауссовских моделей $\text{MAP} = \text{MMSE}$.*

Из указанных выше фактов следует то, что $\mathbf{x}_k^{(k)} = \mathbf{x}_k^+$, что и требовалось доказать.