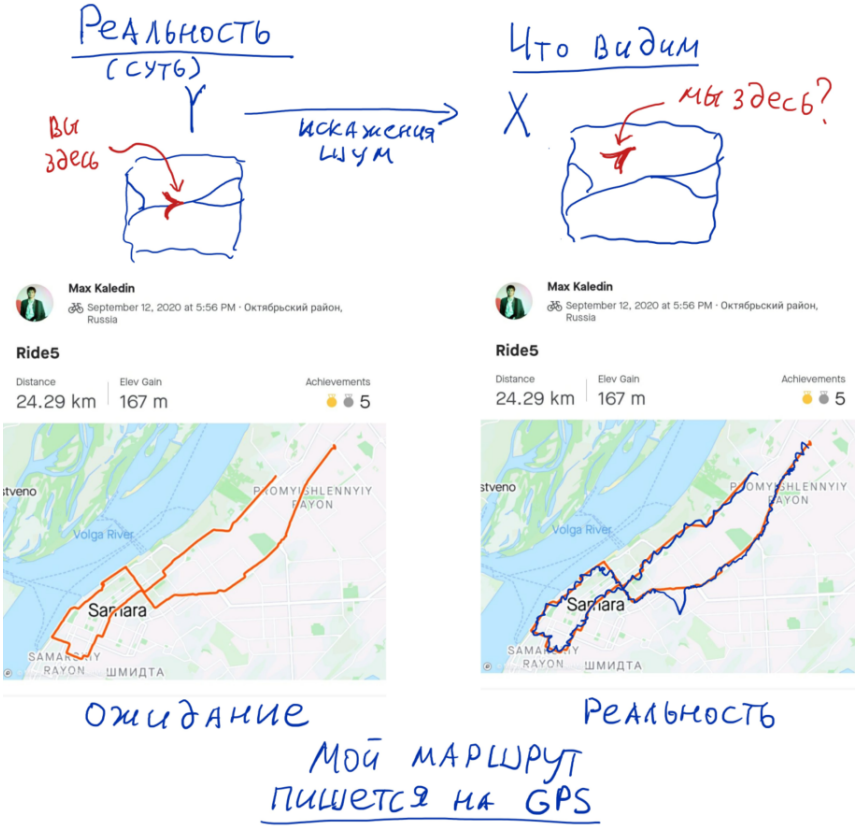
# Фильтр Калмана

*В ближайшие 2 главы мы рассмотрим ещё одну очень эффектную идею, вошедшую в практику почти сразу после того, как EM-алгоритм стал широко известен в 1970-х годах [4]. Речь пойдёт о фильтре Калмана [6], одном из центральных инструментов в обработке сигналов в робототехнике и различных приборах, включающих в себя датчики.*

## 4.1 Выделение целевого сигнала из шума




Мы сейчас живём в мире, наполненном огромным количеством разных устройств. В том числе есть огромное количество маленьких простых устройств, которые обслуживают самые разные ежедневные потребности. Огромная развивающаяся быстрыми темпами область – интернет вещей (Internet of Things, IoT), где конструируются различные технические решения, основанные на локальном использовании небольших вычислительных устройств, состоящих из вычислительной единицы (микроконтроллер или микропроцессор) и системы датчиков: для замера температуры, освещённости, датчкики геопозиционирования (GPS или ГЛОНАСС), акселерометры и многое другое.

Проблема в том, что сами по себе датчики несовершенны, даже на мощных смартфонах вы столкнётесь с тем, что в подземном переходе или тоннеле геопозиционирование может быть сильно сложнее. Но и в обычных обстоятельствах в поле рядом со Сколтехом точность определения геопозиции по GPS может составлять 5-10 метров в зависимости от используемых датчиков. По этой же причине приложения замера пройденного расстояния, работающие от GPS выдают всегда не совсем точные данные.

Но ситуация бы была гораздо хуже, если бы не использовались дополнительно алгоритмы фильтрации сигнала, которые в предположении некоторой динамики системы помогают из нескольких шумных наблюдений уточнить настоящий сигнал. Знаковое первенство (хотя, говорят, что это было сделано на плечах гигантов) в изобретении таких решений для технических приложений принадлежит Рудольфу Калману и Ричарду Бюси, которые решили задачу при условии известной динамики в дискретном времени в 1960 [6] и немного позже в непрерывном в 1963. Одновременно этой задачей также занимались датский математик и астроном Николай Тиле, американский инженер Петер Сверлинг и советский математик Руслан Стратонович. В 1961 году Калман встречался со Стратоновичем в Москве[11], когда Стратонович успел опубликовать некоторое количество работ по этой теме (из них, в частности, [12, 16]). Стратонович и Калман дальше много переписывались, а сам Калман приезжал в СССР и Россию. Фильтр Калмана является частным случаем фильтра Стратоновича, который строится для нелинейных систем.

## 4.2 Задача фильтрации и извлечение сигнала

Представим себе, что у нас есть линейная модель

$$Y_{t+1} = AY_t + U_{t+1},$$

описывающая некоторую техническую систему (например, движение автомобиля или робота) в дискретном времени с точностью до некоторого шума $U_t$. Такая модель оказывается вполне неплохой как минимум для небольших временных масштабов, а далее можно предполагать, что будут меняться матрица $A$ или параметры шума. Это реальный мир, но мы его наблюдаем через какие-то приборы (например, GPS-маячок), которые сигнал некоторым образом искажают и тоже добавляют свой шум:

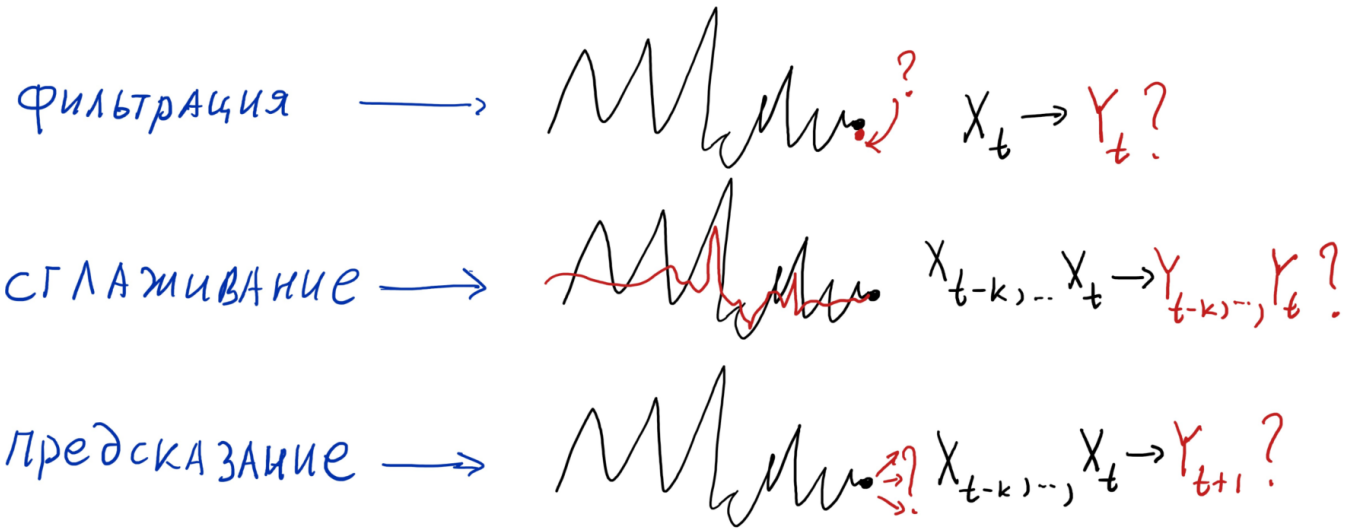
$$X_{t+1} = BY_{t+1} + W_{t+1}.$$

Мы наблюдаем только набор $X_t$ и хотели бы по максимуму избавиться от шума и попробовать оценить настоящий сигнал $Y_t$. При этом в зависимости от конкретного приложения нас может интересовать это оценивание в разном контексте:

1. *Задача фильтрации.* При известном $X_t$ на ходу оценивать $Y_t$.

2. *Задача сглаживания.* При известном $X_{t-n}, .., X_t$ оценить $Y_{t-n}, .., Y_t$.

3. *Задача предсказания.* При известном $X_{t-n}, .., X_t$ оценить $Y_{t+1}$.




Оказывается, что в данных условиях универсальным решением является алгоритм, который носит название *фильтра Калмана* и был впервые описан Калманом в 1960 году [6], а после распространения EM-алгоритма появились процедуры оценивания параметров этой и более общей нелинейной модели.

Как вообще поставить подобную задачу? Во всех трёх случаях нам надо

- при имеющейся информации (значения $X_{t_1}, .., X_{t_k}$) построить оценку $\hat{Y}_t$ значения $Y_t$;
- при этом, в каком-то смысле оптимальную.

Если говорить на языке вероятности, то прогноз (или оценка) – это функция от доступных данных $\hat{Y}_t(X_1(\omega), .., X_k(\omega))$. С другой стороны, если говорить о критерии оптимальности, давайте возьмём привычное MSE(Mean Squared Error)

$$\mathbb{E}\left[\left\|Y_t - \hat{Y}_t(X_{t_1}(\omega), .., X_{t_k}(\omega))\right\|_2^2\right],$$

то есть, мы попробуем минимизировать квадрат отклонения от истины. Вообще, это что-то очень напоминает.

## 4.3 Простая модель наблюдений

Давайте остановимся и попробуем понять, что мы вообще хотим сделать. Раньше мы в статистике говорили об оценках параметров (чисел, которые просто не даны но где-то там в мире фиксированно есть), а теперь мы оцениваем случайную величину?..

Рассмотрим упрощённый но очень показательный пример. В мире произошло великое событие: Саша выиграл $X$ рублей. Сколько конкретно, никто точно не знает и каждый в своих независимых суждениях основывается на разных слухах. Если пойти по людям и спросить их, как им кажется, сколько выиграл Саша , каждый человек $i$ скажет, что

$$Z_i = X + W_i.$$


По нашему дизайну $X, W_1, W_2, ...$ независимые, с нулевым матожиданием и дисперсиями $\mathbb{E}[X^2] = a^2$, $\mathbb{E}\left[W_j^2\right] = m^2$. Саша может и проиграть, но лотерея всегда устраивается так, что гарантированного выигрыша в среднем у участника нет. Как получить искомую оценку? Для начала попробуем сконструировать линейную оценку, то есть, оценку из класса

$$\mathcal{L}_k = \left\{ \sum_{i=1}^{k} c_i Z_i \ : \ c_i \in \mathbb{R} \right\}.$$

Это формально линейное подпространство большего Гильбертова пространства $L^2(P)$ (величины с конечным вторым моментом и нулевым матожиданием): там есть скалярное произведение через интеграл

$$\langle X, Y \rangle = \mathbb{E}[XY].$$

В этом смысле поиск наилучшей оценки при информации $Z_1, .., Z_k$ – это на самом деле задача поиска ортогональной проекции $\mathcal{P}_{\mathcal{L}_k}(X)$ на конечномерное подпространство $\mathcal{L}_k$.

Вообще $Z_j$ зависимы через $X$, поэтому в качестве базиса для ортогональной проекции не очень удобны, хотя они действительно линейно независимы. Чтобы упростить задачу поиска проекции, можно применить процедуру Грама-Шмидта. Так мы конструируем

$$A_1 = Z_1$$

и далее

$$A_j = Z_j - \sum_{i=1}^{j-1} \frac{\langle Z_j, A_i \rangle}{\langle A_i, A_i \rangle} A_i = Z_j - \sum_{i=1}^{j-1} \frac{\mathbb{E}[Z_j A_i]}{\mathbb{E}[A_i^2]} A_i.$$

В новом ортогональном базисе проекции считать сильно проще, проекцией будет

$$\hat{X}_k = \sum_{i=1}^{k} \frac{\mathbb{E}[X A_i]}{\mathbb{E}[A_i^2]} A_i,$$

где мы посчитали

$$c_i = \frac{\mathbb{E}\left[XA_i\right]}{\mathbb{E}\left[A_i^2\right]}.$$

Если вернуться к процессу ортогонализации, один шаг – это вычисление

$$A_j = Z_j - \mathcal{P}_{\mathcal{L}_{j-1}}(Z_j) = Z_j - \mathcal{P}_{\mathcal{L}_{j-1}}(X),$$

последнее происходит из-за того, что $A_i$ и $W_j$ независимы, если $i < j$ и поэтому $\mathbb{E}\left[A_i W_j\right] = 0$. Из этого также следует интересное наблюдение:

$$\mathbb{E}\left[XA_j\right] = \mathbb{E}\left[(X - \hat{X}_{j-1})^2\right].$$

То есть, оказывается, новую оценку (при данных $Z_1, .., Z_k$) можно получить путём некоторой линейной коррекции старой (при данных $Z_1, .., Z_{k-1}$) с помощью уже известной в момент $k$ невязки $Z_k - \hat{X}_{k-1}$:

$$\hat{X}_k = \hat{X}_{k-1} + \frac{\mathbb{E}\left[(X - \hat{X}_{k-1})^2\right]}{\mathbb{E}\left[(X - \hat{X}_{k-1})^2\right] + m^2}(Z_k - \hat{X}_{k-1}).$$

Проанализируем это матожидание, добавив $\pm\hat{X}_{j-1}$:

$$S_j = \mathbb{E}\left[(X - \hat{X}_j \pm \hat{X}_{j-1})^2\right] = S_{j-1} + \mathbb{E}\left[(\hat{X}_j - \hat{X}_{j-1})^2\right] - 2\mathbb{E}\left[(X - \hat{X}_{j-1})(\hat{X}_j - \hat{X}_{j-1})\right].$$

Почти всё мы уже видели, например,

$$\mathbb{E}\left[(\hat{X}_j - \hat{X}_{j-1})^2\right] = \frac{S_{j-1}^2}{(S_{j-1} + m^2)^2}\mathbb{E}\left[(Z_j - \hat{X}_{j-1})^2\right]$$

и

$$\mathbb{E}\left[(X - \hat{X}_{j-1})(\hat{X}_j - \hat{X}_{j-1})\right] = \mathbb{E}\left[X(\hat{X}_j - \hat{X}_{j-1})\right] = \frac{S_{j-1}}{S_{j-1} + m^2}\mathbb{E}\left[X(Z_j - \hat{X}_{j-1})\right] = \frac{S_{j-1}^2}{S_{j-1} + m^2},$$

и ещё

$$\mathbb{E}\left[(Z_j - \hat{X}_{j-1})^2\right] = \mathbb{E}\left[(X + W_j - \hat{X}_{j-1})^2\right] = S_{j-1} + m^2.$$

Если собрать это вместе, то мы получим уравнение на $S_k$

$$S_k = S_{k-1} - \frac{S_{k-1}^2}{S_{k-1} + m^2}.$$

На каждом шаге мы можем пересчитывать новое значение $S_k$ основываясь на предыдущем. Таким образом, мы получили совсем простую процедуру вычисления нового значения, уточнённого с помощью невязки предыдущего прогноза. Конкретнее, с приходом нового наблюдения $Z_j$ алгоритм действий такой:

1. Вычислить новый прогноз

$$\hat{X}_k = \hat{X}_{k-1} + \frac{S_{k-1}}{S_{k-1} + m^2}(Z_k - \hat{X}_{k-1});$$

2. Вычислить его ожидаемую квадратичную ошибку

$$S_k = S_{k-1} - \frac{S_{k-1}^2}{S_{k-1} + m^2}.$$

Оказывается, можно даже получить чёткую формулу $\hat{X}_k$, без рекуррентного отношения. Можно подумать, что стоит попробовать искать решение в виде $\hat{X}_k = \alpha_k \sum_{j=1}^k Z_j$. Это ортогональная проекция тогда и только тогда, когда $X - \hat{X}_k$ ортогонально $Z_i$ для всех $i \leq k$. В смысле $L^2$ это даёт

$$\mathbb{E}\left[(X - \hat{X}_k)Z_i\right] = \mathbb{E}\left[(X - \alpha_k \sum_{j=1}^k Z_j)Z_i\right] = a^2 - \alpha_k(ka^2 + m^2) = 0.$$

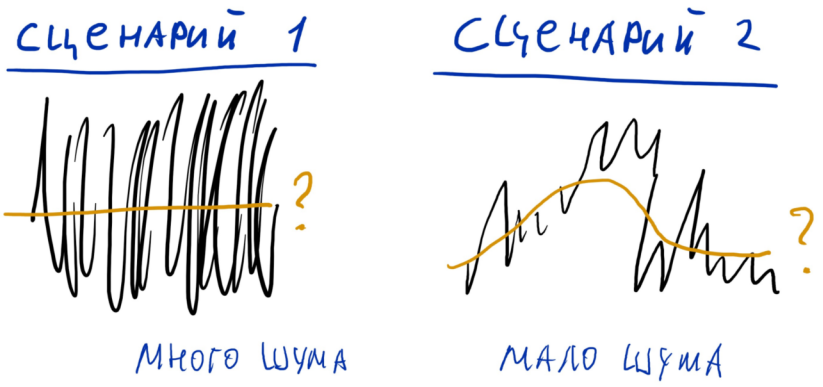
Это должно равняться нулю, поэтому

$$\alpha_k = \frac{a^2}{ka^2 + m^2},$$

что приводит к

$$\hat{X}_k = \frac{a^2}{a^2 + m^2/k} \frac{1}{k} \sum_{j=1}^k Z_j.$$

Видно, что это не совсем обычное предсказание выборочным средним. Если $m^2$ небольшое относительно $a^2$, то достаточно быстро мы получим, что прогноз примерно равен выборочному среднему наблюдений (верим в данные, потому что шум слабый, можем примерно прикинуть выигрыш), а если наоборот большое, то на начальных этапах наблюдения вообще игнорируются и прогноз близок к нулю (верим в матожидание, шум очень большой, из данных ничего особо не ясно, наверное, Саша ничего не выиграл). И в любом случае в силу закона больших чисел при $k \to \infty$ прогноз почти наверное сойдётся к.. $X$ – тому, что мы хотим оценивать (шумы усредняются в 0).




## 4.4 Найти условное матожидание

Итак, общая картина теперь более ясна: где-то там в мире есть настоящая динамика $Y_1(\omega), .., Y_t(\omega), ...$, мы её наблюдаем как $X_1(\omega), .., X_t(\omega), ..$ и хотели бы понять, чему примерно равняется $Y_t(\omega)$ если доступна информация о некоторых $X$.

На самом деле то, что мы искали, – это условное матожидание

$$\hat{Y}_t = \mathbb{E}\left[Y_t \mid X_1, .., X_t\right],$$

а матожидание квадратичной ошибки в общем случае – ковариационная матрица

$$P_t = \mathbb{E}\left[(Y_t - \hat{Y}_t)(Y_t - \hat{Y}_t)^T\right].$$

Попробуем обобщить идеи выше на наш случай. Вспомним модель:

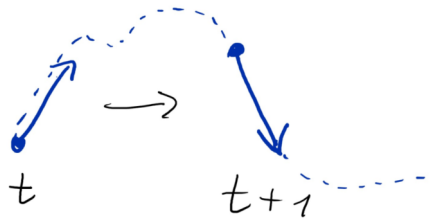
$$\begin{aligned} Y_{t+1} &= AY_t + U_{t+1} \quad \text{(мир)}, \\ X_{t+1} &= BY_{t+1} + W_{t+1} \quad \text{(наблюдения)}, \end{aligned}$$

а также зададим стартовое условие $Y_1 \sim N(\xi, \Lambda)$. Предположим, что все шумы $U_t \sim_{iid} N(0, R_Y)$ и $W_t \sim_{iid} N(0, R_X)$ и независимы. Давайте пока что как и раньше займёмся фильтрацией и будем оценивать $Y_t$ по значениям $X_1, .., X_t$. Метод для фильтрации был такой: получать новую оценку $\hat{Y}_t$ на основе старой $\hat{Y}_{t-1}$ с коррекцией на невязку.

Прежде чем переходить к записи алгоритма, попробуем понять его устройство. Алгоритм Калмана в том, чтобы сначала построить априорный прогноз на основе прошлого прогноза

$$\tilde{Y}_t = A\hat{Y}_{t-1}.$$

В отличие от предыдущего случая, вместо $X$ у нас теперь изменяющийся процесс $Y_t$, поэтому прогноз тоже делается с учётом того, что статичности нет и динамика поменяет $Y_{t-1}$ на $Y_t$.




Потом этот прогноз, как и раньше, нужно уточнить, потому что нам пришло новое наблюдение $X_t$. С помощью него производится фильтрация. Конкретнее, делается линейная коррекция с помощью невязки:

$$\hat{Y}_t = \tilde{Y}_t + K_t(X_t - B\tilde{Y}_t).$$

Матрица $K_t$ (её ещё называют *Kalman Gain*)выбирается на каждом шаге своя и такая, чтобы минимизировать квадрат ошибки апостериорного прогноза

$$K_t = \arg\min_K \mathbb{E}\left[\left\|\hat{Y}_t - Y_t\right\|_2^2 \mid X_1, .., X_t\right],$$

где условие в матожидании отражает тот факт, что мы используем информацию о значениях $X_1, .., X_t$. А как мы уже помним, решение этой задачи есть ортогональная проекция на подходящее подпространство (попробуйте подумать, с каким базисом!). Точно так же мы поступали в простом примере выше, подбирая правильный вес перед невязкой. Эта задача из-за полной линейности всего решается (если вы умеете брать производные по матрицам, можете попробовать сами, но мы поможем наметить путь). Конкретно нужно взять производную по матрице $K_t$, для этого удобно записать скалярное произведение через следы:

$$\partial_K \mathbb{E}\left[\|\tilde{Y}_t - Y_t + K(X_t - B\tilde{Y}_t)\|_2^2 \mid X_1, .., X_t\right] = \partial_K \mathbb{E}\left[\|\tilde{Y}_t - Y_t + K(BY_t + W_t - B\tilde{Y}_t)\|_2^2 \mid X_1, .., X_t\right] =$$
$$= \partial_K \mathbb{E}\left[\|(I - KB)(\tilde{Y}_t - Y_t) + KW_t)\|_2^2 \mid X_1, ..X_t\right] =$$
$$= \partial_K Tr\left(\mathbb{E}\left[\Big((I - KB)(\tilde{Y}_t - Y_t) + KW_t)\Big)\Big((I - KB)(\tilde{Y}_t - Y_t) + KW_t)\Big)^T \mid X_1, .., X_t\right]\right).$$

Далее, понемногу вычисляя производные и приравнивая выражение к 0, приходим к выражению

$$K_t = \mathbb{E}\left[(\tilde{Y}_t - Y_t)(\tilde{Y}_t - Y_t)^T \mid X_1, .., X_t\right] B^T (B\mathbb{E}\left[(\tilde{Y}_t - Y_t)(\tilde{Y}_t - Y_t)^T \mid X_1, .., X_t\right] B^T + R_x)^{-1} =$$
$$= \tilde{P}_t B^T (B\tilde{P}_t B^T + R_x)^{-1}.$$

Матрица

$$\tilde{P}_t = \mathbb{E}\left[(\tilde{Y}_t - Y_t)(\tilde{Y}_t - Y_t)^T \mid X_1, .., X_t\right] -$$

это ковариационная матрица ошибок априорного прогноза, через которую мы можем вычислить ковариационную матрицу апостериорных прогнозов (подставьте частично вычисленный $K_t$):

$$P_t = \mathbb{E}\left[(\hat{Y}_t - Y_t)(\hat{Y}_t - Y_t)^T \mid X_1, .., X_t\right] =$$
$$= \mathbb{E}\left[(\tilde{Y}_t - Y_t + K_t B(Y_t - \tilde{Y}_t))(\tilde{Y}_t - Y_t + K_t B(Y_t - \tilde{Y}_t))^T \mid X_1, .., X_t\right] =$$
$$... = \tilde{P}_t - K_t B\tilde{P}_t.$$

Матрицы $\tilde{P}_t$ и $P_t$ – важный технический ингредиент и побочный продукт, которые мы можем использовать для построения доверительных интервалов прогнозов.

Итак, общая схема такая: изначально задаётся $\hat{Y}_1 = \xi, P_1 = \Lambda$, а далее в цикле по $t$

$$\tilde{Y}_t = A\hat{Y}_{t-1} \quad \text{(априорный прогноз)},$$
$$\tilde{P}_t = AP_{t-1}A^T + R_y \quad \text{(ошибка априорного прогноза)},$$
$$K_t = \tilde{P}_t B^T (B\tilde{P}_t B^T + R_x)^{-1} \quad \text{(фильтр)}$$
$$\hat{Y}_t = \tilde{Y}_t + K_t(X_t - B\tilde{Y}_t) \quad \text{(апостериорный прогноз)},$$
$$P_t = \tilde{P}_t - K_t B\tilde{P}_t \quad \text{(ошибка апостериорного прогноза)}.$$

Что такое прогноз $\hat{Y}_t$? Это в точности условное матожидание, а $P_t$ – ковариационная матрица ошибок:

$$\hat{Y}_t = \mathbb{E}\left[Y_t \mid X_1, .., X_t\right], \quad P_t = \mathbb{E}\left[(Y_t - \hat{Y}_t)(Y_t - \hat{Y}_t)^T \mid X_1, .., X_t\right].$$

То есть, мы не только получили прогноз значения $Y_t$ в виде $\hat{Y}_t$, но ещё и оценку неопределённости, потому что по конструкции

$$Y_{t_{|X_1,..,X_t}} \sim N(\hat{Y}_t, P_t).$$

## 4.5 Универсальный ответ

Подытожим: мы решили задачу фильтрации, вычислив прогноз $\hat{Y}_t$ и какую-то оценку ошибки прогноза. Но фильтр Калмана – это на самом деле устройство, которое способно решать все три поставленные задачи. Более того, в гауссовской модели, как у нас, это будет оптимальным с точнки зрения дисперсии подходом.

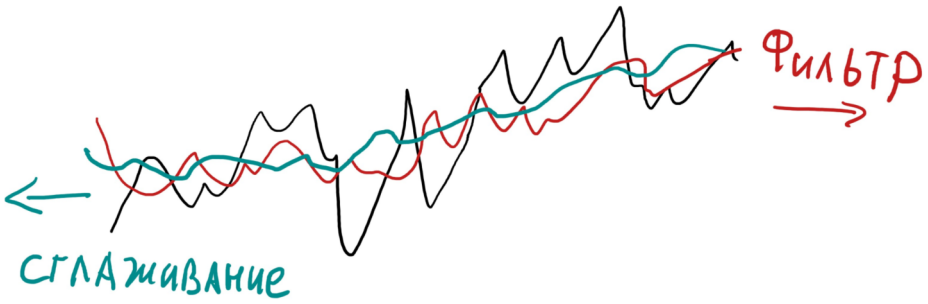
1. *Задача фильтрации.* Мы уже умеем: Двигаемся по времени и вычисляем по алгоритму $\hat{Y}_t$.

2. *Задача предсказания.* Можно догадаться, что предсказание – это тот самый шаг априорного прогноза: $\tilde{Y}_t = \mathbb{E}\left[Y_t \mid X_1, .., X_{t-1}\right]$. На самом деле, фильтр Калмана можно интерперетировать как Байесовскую модель (точнее Байесовскую векторную авторегрессию $VAR(1)$), потому что на выходе мы получаем ещё и оценку неопределённости.

3. *Задача сглаживания.* Вот тут сложнее. Но в какой-от момент была статья [10], где построили очень похожий алгоритм, но в обратную сторону по времени, который по именам авторов назвали RTSS(Rauch-Tung-Striebel smoother (RTSS)). Этот алгоритм применяется после фильтрации и уточняет все сделанные ранее оценки, получая сглаженные оценки $\overline{Y}_t$. Детальнее, алгоритм такой: задать $\overline{Y}_T = \hat{Y}_T$, $\overline{P}_T = P_T$, а далее в цикле $t = T-1, .., 1$ выполнять

$$\begin{aligned}
&S_t = P_t A^T (\widetilde{P}_{t+1})^{-1} \ \text{ (вычисление сглаживающего фильтра)},\\
&\overline{Y}_t = \hat{Y}_t + S_t(\overline{Y}_{t+1} - \widetilde{Y}_{t+1}) \ \text{ (сглаживание)},\\
&\overline{P}_t = P_t + S_t(\overline{P}_{t+1} - \widetilde{P}_{t+1})S_t^T \ \text{ (ошибка сглаживания)}.
\end{aligned}$$

По смыслу $\overline{Y}_t = \mathbb{E}\left[Y_t \mid X_1, .., X_T\right]$, а матрица $\overline{P}_t = \mathbb{E}\left[\right](Y_t - \overline{Y}_t)(Y_t - \overline{Y}_t)^T X$.


Фильтр
сглаживание

# Литература


[1] T. W. Anderson and D. A. Darling. Asymptotic Theory of Certain "Goodness of Fit"Criteria Based on Stochastic Processes. *The Annals of Mathematical Statistics*, 23(2):193 – 212, 1952.

[2] Peter C Austin, Muhammad M Mamdani, David N Juurlink, and Janet E Hux. Testing multiple statistical hypotheses resulted in spurious associations: a study of astrological signs and health. *J Clin Epidemiol*, 59(9):964–969, July 2006.

[3] Erika Cule, Paolo Vineis, and Maria De Iorio. Significance testing in ridge regression for genetic data. *BMC Bioinformatics*, 12(1):372, Sep 2011.

[4] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, 39(1):1–38, 1977.

[5] B. Efron. Bootstrap methods: Another look at the jackknife. *The Annals of Statistics*, 7(1):1–26, 1979.

[6] R. E. Kalman. A New Approach to Linear Filtering and Prediction Problems. *Journal of Basic Engineering*, 82(1):35–45, 03 1960.

[7] S.J. Levitt, S.D. Dubner. *Freakonomics*. NY: Harper Trophy, 2006.

[8] Richard Lockhart, Jonathan Taylor, Ryan J. Tibshirani, and Robert Tibshirani. A significance test for the lasso. *The Annals of Statistics*, 42(2):413–468, 2014.

[9] J. Scott Long and Laurie H. Ervin. Using heteroscedasticity consistent standard errors in the linear regression model. *The American Statistician*, 54(3):217–224, 2000.

[10] Herbert E Rauch, F Tung, and Charlotte T Striebel. Maximum likelihood estimates of linear dynamic systems. *AIAA journal*, 3(8):1445–1450, 1965.

[11] O. A. Stepanov. Kalman filtering: Past and present. an outlook from russia. (on the occasion of the 80th birthday of rudolf emil kalman). *Gyroscopy and Navigation*, 2(2):99–110, Apr 2011.

[12] R. L. Stratonovich. Conditional markov processes. *Theory of Probability & Its Applications*, 5(2):156–178, 1960.

[13] Halbert White. A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica*, 48(4):817–838, 1980.

[14] Платонов Е.Н. Горяинова Е.Р., Панков А.Р. *Прикладные методы анализа статистических данных.* Изд. дом Высшей школы экономики, 2012.

[15] Ю.М. Кельберт, М.Я. Сухов. *Вероятность и статистика в примерах и задачах. Т.3: теория информации и кодирования.* М.: МЦНМО, 2013.

[16] Р.Л. Стратонович. *Условные марковские процессы и их применение к теории оптимального управления.* Московский государственный университет, 1966.

[17] А.Н. Ширяев. *Основы стохастической финансовой математики.* МЦНМО, 2016.