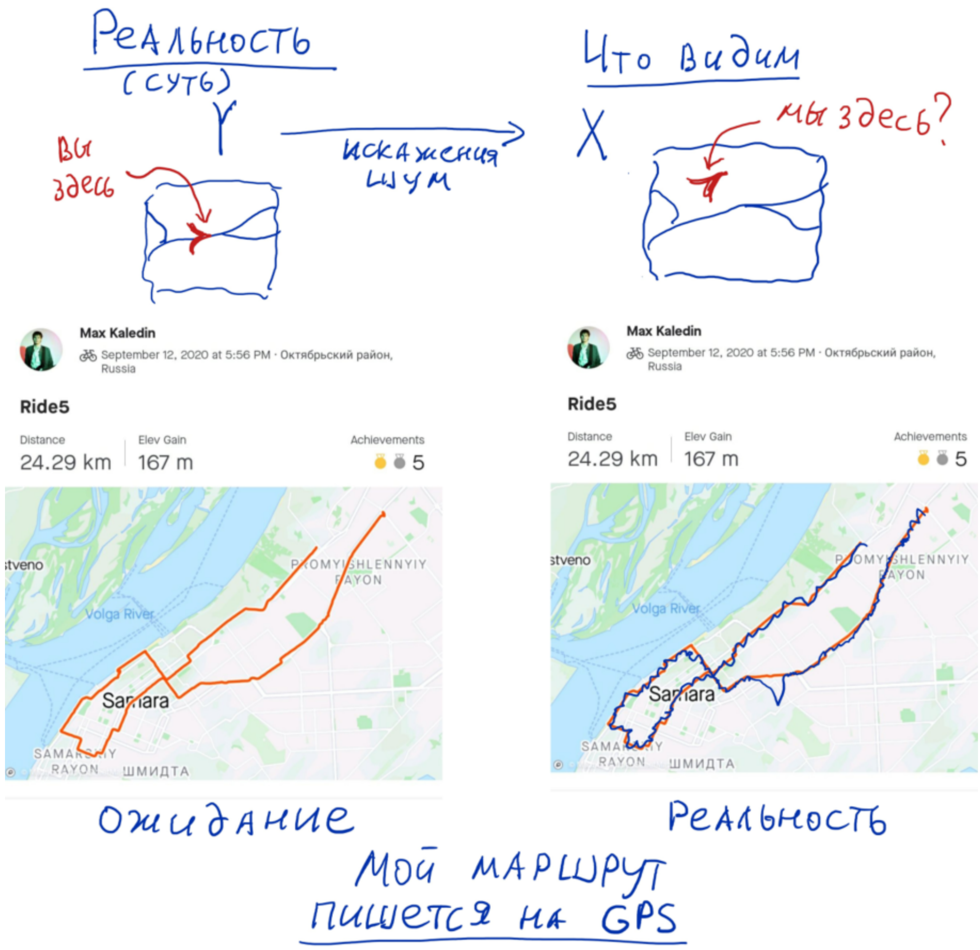
ЛЕКЦИЯ 13

Задача фильтрации

В этой лекции мы рассмотрим одну популярную инженерную задачу, актуальную для робототехники, интернета вещей и вообще работы с любыми датчиками. Мы поймём, как можно из наблюдаемых зашумлённых данных эффективно оценить скрытую истину. Мы не в силах покрыть все детали этой огромной области, но попробуем понять, чем нам в этой задаче может помочь знание стохастических дифференциальных уравнений.

13.1 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 [7] 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[12], когда Стратонович успел опубликовать некоторое количество работ по этой теме (из них, в частности, [13, 18]). 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.

13.2 Постановка задачи

Мы предполагаем, что система работает по некоторым своим законам, которые моделируются как дифференциальные уравнения типа Стратоновича

$$dY_t = b(t, Y_t)dt + \sigma(t, Y_t) \circ dU_t,$$

где U – многомерный Винеровский процесс. Сами Y_t ненаблюдаемые, вместо них наблюдается некоторая зашумлённая величина

$$H_t = c(t, Y_t) + \gamma(t, Y_t)\widetilde{W}_t,$$

где $\widetilde{W}$ – процесс гауссовского белого шума. Мы ранее видели проблемы, возникающие с такой постановкой, поэтому чтобы их обойти, вместо H_t будем рассматривать процесс аккумулированных наблюдений

$$X_t = \int_0^t H_s ds.$$

Это не искажает задачу, потому что эти процессы и фильтрация устроены так, что H_s измерима относительно $\mathcal{F}_s$ тогда и только тогда, когда X_s тоже измерима. Согласившись на такую постановку, мы получим уравнение для наблюдений

$$dX_t = c(t, Y_t)dt + \gamma(t, Y_t) \circ dV_t.$$

К счастью для нас, уравнения Ито и Стратоновича взаимозаменяемы, нужно только скорректировать функцию перед dt, поэтому дальше мы будем рассматривать только постановку в смысле Ито, про которую мы уже много знаем.

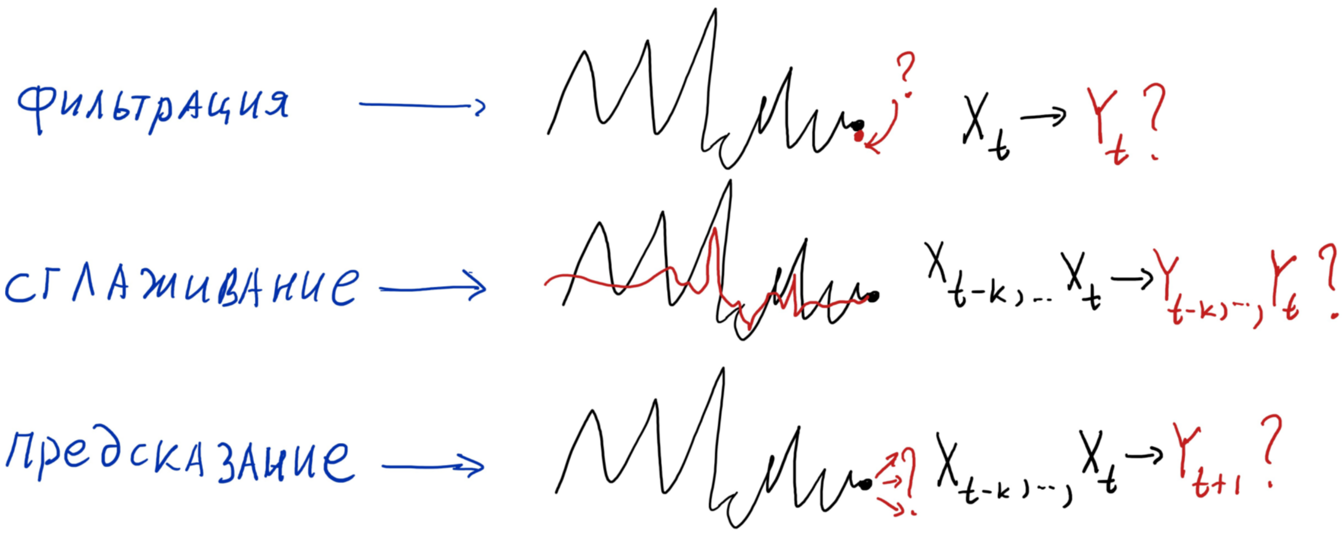
Итак, мы руководствуемся следующей моделью системы и наблюдений:

$$dY_t = b(t, Y_t)dt + \sigma(t, Y_t)dU_t, \text{ (модель системы)}$$
$$dX_t = c(t, Y_t)dt + \gamma(t, Y_t)dV_t, \text{ (модель наблюдений)},$$

где U_t и V_t независимые многомерные Винеровские процессы. Мы хотели бы решать фундаментально три задачи, используя только наблюдения

1. *Задача фильтрации.* При известном X_t на ходу оценивать Y_t.
2. *Задача сглаживания.* При известной истории X до t оценить значения Y_s при $s \le t$.
3. *Задача предсказания.* При известной истории X до t оценить будущий момент $Y_{t+\Delta}$.

Лучше всего их иллюстрирует задача в дискретном времени.



Формально *оценить с использованием истории* означает, что мы ищем оценку $\hat{Y}_t$ как согласованную с фильтрацией $(\mathcal{F}_t)$ функцию (мы используем фильтрацию, порождённую Винеровскими процессами). Более того, мы хотели бы оценить *наилучшим образом*, то есть, полученная оценка должна минимизировать

$$\mathbb{E}\left[\left|\hat{Y}_t - Y_t\right|^2\right],$$

иными словами, быть оптимальным L^2-прогнозом среди

$$\mathcal{K}_t = \left\{ Y : \Omega \to \mathbb{R}^n;\ Y \in L^2(P),\ Y \text{ измерима относительно } \mathcal{F}_t \right\}.$$

Оказывается, что в данных условиях универсальным решением является алгоритм, который носит название *фильтра Калмана* или *фильтра Калмана-Бюси*. Мы будем рассматривать задачу, где параметры модели и наблюдений известны, но вообще полезно знать, что существуют процедуры их оценки на основе EM-алгоритма, которые заслуживают отдельного курса.

Мы упростим себе задачу, рассматривая только одномерную постановку и линейную, то есть,

$$\begin{aligned} dY_t &= F(t)Y_t dt + C(t)dU_t,\ \text{(модель системы)} \\ dX_t &= G(t)Y_t dt + D(t)dV_t,\ \text{(модель наблюдений)}, \end{aligned}$$

где $F, C, G, D : \mathbb{R} \to \mathbb{R}$ предполагаются ограниченными на ограниченных интервалах. Уже здесь нужно будет сделать очень много усилий, чтобы получить результат. Детальный вывод с подробностями (они вполне нам по силам, но их очень много) можно найти в [10, Гл.6]. Технические подробности мы будем по возможности опускать, но попробуем хотя бы поверхностно понять идею позади конструкции фильтра Калмана.

13.3 Строим оценку сигнала(простой случай)

Чтобы понять лучше как можно подойти к решению задачи фильтрации, полезно рассмотреть очень простой случай. Пусть $X, W_1, W_2, \ldots$ независимые, с нулевым матожиданием и дисперсиями $\mathbb{E}\left[X^2\right] = a^2,\ \mathbb{E}\left[W_j^2\right] = m^2$. Пусть

$$Z_j = X + W_j,$$

как получить лучшую L^2-оценку X среди всех измеримых относительно $\mathcal{F}_k = \sigma(X_j, j \leq k)$? 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 линейное подпространство $L^2(P)$, а само пространство $L^2(P)$ является Гильбертовым: там есть скалярное произведение через интеграл и относительно метрики, порождённой скалярным произведением, оно полное метрическое пространство. В этом смысле поиск наилучшей в L^2 оценки – это на самом деле задача поиска ортогональной проекции $\mathcal{P}_{\mathcal{L}_k}$ 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, в противном случае $Z_j(\omega)$ выражалось бы линейно через остальные для всех $\omega \in \Omega$ и какого ненулевого α, что противоречит независимости W и X. Чтобы упростить задачу поиска проекции, можно применить процедуру Грама-Шмидта. Так мы конструируем

$$A_1 = Z_1$$

и далее

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

В новом ортогональном базисе проекции считать сильно проще, оптимальной оценкой в $L^2(P)$ будет

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

где мы посчитали

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

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 X и W_j независимы и поэтому $\mathbb{E}\left[X W_j\right] = 0$. Из этого также следует интересное наблюдение:

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

То есть, новую оценку можно получить путём некоторой линейной коррекции старой 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

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

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

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

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 $i \le k$. 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

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

Это не совсем обычное предсказание выборочным средним. Если m^2 небольшое относительно a^2, то в какой-то момент мы получим, что прогноз примерно равен выборочному среднему наблюдений, а если наоборот большое, то на начальных этапах наблюдения вообще игнорируются и прогноз близок к нулю. И в любом случае в силу закона больших чисел при $k \to \infty$ прогноз почти наверное сойдётся к среднему, то есть, к нулю.

13.4 Фильтр Калмана-Бюси (непрерывное время)

Что мы узнали из этого простого случая?

1. Мы умеем строить *линейные* оценки в дискретном случае;
2. Мы получаем прогноз, обновляя старый прогноз и *корректируя* его с помощью невязки $Z_k - \hat{X}_{k-1}$;
3. Z_j – процесс наблюдений, но он с ортогональными приращениями, что позволило так много упростить по пути;

4. Мы знали решение модели состояний, это был просто X.

Для того, чтобы перейти к дискретному случаю, нужно суметь обобщить эти наблюдения. Конкретнее:

1. Как соотнести линейную оценку с проекцией на пространство $\mathcal{K}_t$, которое больше не конечномерное и даже необязательно там есть хотя бы счётный базис.

2. Как сделать замену процесса наблюдений на удобный процесс с ортогональными приращениями.

3. Как получить итоговое соотношение для $\hat{X}_t$, конкретнее в непрерывном времени нас интересует дифференциальное уравнение, а не рекуррентное отношение.

13.4.1 Проекция – это условное матожидание

И начинаем мы с пункта 1, вспоминая по пути, что такое условное матожидание.

Лемма 13.1. *Пусть $\mathcal{H} \subset \mathcal{F}$ – это сигма-алгебра и $X \in L^2(P)$ (то есть, есть второй момент) измерима относительно $\mathcal{F}$ (то есть, является корректной случайной величиной). Обозначим за $\mathcal{P}_\mathcal{H}$ ортогональную проекцию на $\mathcal{K}_\mathcal{H}$, пространство измеримых относительно $\mathcal{H}$ функций. Тогда*

$$\mathcal{P}_\mathcal{H}(X) = \mathbb{E}\left[X \mid \mathcal{H}\right].$$

▷ Мы неоднократно вспоминаем это свойство, но полезно вернуться к нему в новом контексте. Вспомним, что условное матожижание

1. существует;

2. единственно (если есть другое, то разница почти наверное 0);

3. как функция от ω, измеримо относительно $\mathcal{H}$;

4. по определению это такая функция, что для всех $A \in \mathcal{H}$
$$\int_A \mathbb{E}\left[X \mid \mathcal{H}\right] dP = \int_A X dP.$$

С другой стороны, $\mathcal{P}_\mathcal{H}(X)$ – ортогональная проекция, это верно тогда и только тогда, когда для всех $Y \in \mathcal{K}_\mathcal{H}$
$$\int_\Omega (X - \mathcal{P}_\mathcal{H}(X)) Y dP = 0.$$

Индикаторная функция $\mathbb{1}_A(\omega)$ при $A \in \mathcal{H}$ будет измеримой относительно $\mathcal{H}$, поэтому для всех A
$$\int_\Omega (X - \mathcal{P}_\mathcal{H}(X)) \mathbb{1}_A dP = \int_A (X - \mathcal{P}_\mathcal{H}(X)) dP = 0.$$

Мы получили в точности определение условного матожидания, которое единственно. □

Этот факт позволяет нам уточнить, как считать проекцию, то есть, в качестве наилучшего L^2-прогноза мы должны искать

$$\hat{Y}_t = \mathcal{P}_{\mathcal{K}_t}(Y_t) = \mathbb{E}\left[Y_t \mid \mathcal{F}_t\right],$$

то есть, условное матожидание при условии сигма-алгебры из фильтрации.

13.4.2 Линейные оценки против измеримых

Далее, заметим, что процесс (Y_t, X_t) – гауссовский, мы явно выводили решение дифференциального уравнения с помощью формулы Ито и интегрирующего множителя в одномерном случае, к которому сводится данный путём замены базиса в $\mathbb{R}^2$. Это важный момент, потому что теперь мы можем сопоставить наилучшие в L^2 линейные и измеримые оценки, используя гауссовость.

Лемма 13.2. *Рассмотрим случайные величины Y и $X_s, s \leq t$ из $L^2(P)$ и такие, что*

$$Y, X_{s_1}, .., X_{s_n}$$

имеют совместное гауссовское распределение для любого конечного набора индексов $s_i \leq t$. Тогда

$$\mathcal{P}_{\mathcal{L}_t}(Y) = \mathcal{P}_{\mathcal{K}_t}(Y) = \mathbb{E}\left[Y \mid \mathcal{F}_t\right],$$

где $(\mathcal{F}_t)$ – фильтрация, с которой согласован процесс X, $\mathcal{K}_t$ – множество функций, измеримых относительно $\mathcal{F}_t$, а множество $\mathcal{L}_t$ – это замыкание множества

$$\left\{\sum_{i=1}^{k} c_i X_{t_i} \ : \ k \geq 1, \ c_i \in \mathbb{R}, \ t_i \leq t\right\}.$$

▷ Рассмотрим невязку с линейной проекцией $\widetilde{Y} = Y - \mathcal{P}_{\mathcal{L}}(Y)$ и покажем, что она независима от сигма-алгебры $\mathcal{F}_t$. Предел в L^2 нормальных величин тоже нормальная величина (включая константу), если он существует, поэтому проекция будет нормальной случайной величиной, как и невязка. Поэтому

$$\widetilde{Y}, X_{s_1}, .., X_{s_n}$$

тоже имеют совместное гауссовское распределение. Так как мы считали ортогональную проекцию,

$$\mathbb{E}\left[\widetilde{Y} X_{s_i}\right] = 0$$

и в силу гауссовости $\widetilde{Y}$ не зависит от любого набора $X_s, s \leq t$, которыми порождена $\mathcal{F}_t$, следовательно, невязка не зависит от $\mathcal{F}_t$. В этом случае для любого $A \in \mathcal{F}_t$

$$\mathbb{E}\left[\mathbb{1}_A \widetilde{Y}\right] = \mathbb{E}\left[\mathbb{1}_A\right] \mathbb{E}\left[\widetilde{Y}\right] = 0,$$

а значит, второе слагаемое равно нулю, и это в результате даёт определение условного матожидания. □

Таким образом, мы соединили результат про проекцию с конструкцией линейных оценок. Вообще, пространство $\mathcal{L}$ несмотря на свою общность кажется относительно представимым: это все конечные комбинации случайных величин до момента t и все L^2-пределы таких сумм. Немного знания про L^2-пространства(доказывать не будем) и можно получить его точное описание, которое не очень удивляет. Вернёмся в рамки изначальной модели с процессами Y, X.

Теорема 13.3. *Пространство $\mathcal{L}_T$, определённое как замыкание множества*

$$\left\{\sum_{i=1}^{k} c_i X_{t_i} \;:\; k \geq 1,\; c_i \in \mathbb{R},\; t_i \leq t\right\},$$

в точности

$$\mathcal{L}_T = \left\{c_0 + \int_0^T f(t) dX_t;\quad c_0 \in \mathbb{R},\; f \in L^2[0,T]\right\}.$$

13.4.3 Процесс невязок

Как и в дискретном случае, мы строим коррекцию, рассматривая невязки на предыдущих шагах, аналог процесса $Z_k - \hat{X}_{k-1}$ из начала – это процесс

$$N_t = X_t - \int_0^t \mathcal{P}_{\mathcal{K}_s}(G(s)Y_s) ds = X_t - \int_0^t G(s)\hat{Y}_s ds,$$

что в дифференциальной нотации эквивалентно записывается как

$$dN_t = G(t)(Y_t - \hat{Y}_t)dt + D(t)dV_t.$$

Оказывается, что этот процесс несильно отличается от Винеровского.

Лемма 13.4. *Процесс N_t*

- *гауссовский;*
- *имеет ортогональные приращения;*
- *имеет те же пространства $\mathcal{L}_t$, что и процесс Y.*

От Винеровского процесса его отличает только ковариационная функция и в частности дисперсия, потому что используя формулу Ито и взяв матожидание можно получить

$$\mathbb{E}\left[N_t^2\right] = \int_0^t D(s)^2 ds.$$

Если развить эту мысль, мы получим, что

$$dR_t = \frac{1}{D(t)} dN_t,\; R_0 = 0$$

будет настоящим Винеровским процессом с непрерывными траекториями. Более того, поскольку мы только поделили на детерминированную функцию пространства $\mathcal{L}_t$ останутся теми же, что и у N_t. В свою очередь строить проекцию на основе процесса R уже оказывается существенно проще и можно даже выписать точную формулу.

Теорема 13.5. *Проекцию можно вычислить как*

$$\hat{Y}_t = \mathbb{E}\left[Y_t\right] + \int_0^t \partial_s \mathbb{E}\left[Y_t R_s\right](s) dR_s.$$

▷ По описанию пространства, куда мы проектируем, проекция на пространство наблюдений процесса R имеет вид

$$\hat{Y}_t = c(t) + \int_0^t f(s) dR_s$$

для какой-нибудь $f \in L^2[0,t]$. Поскольку R – Винеровский процесс, интеграл справа – это интеграл Ито и его матожидание равно 0, поэтому

$$\mathbb{E}\left[\hat{Y}_t\right] = c(t)$$

и при этом $\mathbb{E}\left[Y_t\right] = \mathbb{E}\left[\hat{Y}_t\right]$ по свойству нашей проекции. Кроме того из ортогональности проекции следует, что

$$(Y_t - \hat{Y}_t) \text{ и } \int_0^t p(s) dR_s$$

ортогональны для любой $p \in L^2[0,t]$. Воспользуемся:

$$\mathbb{E}\left[Y_t \int_0^t p(s) dR_s\right] = \mathbb{E}\left[\hat{Y}_t \int_0^t p(s) dR_s\right] = \mathbb{E}\left[\int_0^t f(s) dR_s \int_0^t p(s) dR_s\right],$$

откуда получается с помощью изометрии Ито для двойных интегралов (см. ДЗ-3)

$$\mathbb{E}\left[\int_0^t f(s) dR_s \int_0^t p(s) dR_s\right] = \int_0^t f(s) p(s) ds.$$

Поскольку это верно для любых $p \in L^2[0,t]$, выберем $p(t) = \mathbb{1}(t \in [0,r])$ с $r \leq t$. Тогда получится

$$\mathbb{E}\left[Y_t R_r\right] = \int_0^r f(s) ds,$$

что изначально мы искали.

□

13.4.4 Финальные штрихи

На семинаре мы как-то выводили точную формулу для решения уравнения для Y_t, используя его и все теоремы до этого мы, наконец, можем подойти к финальному результату, для которого нужно проделать ещё достаточно дополнительных вычислений (в том числе, получить уравнение Рикатти). Финальный результат выглядит так.

Теорема 13.6. *(Фильтр Калмана-Бюси, 1d) Решение $\hat{X}_t$ задачи фильтрации для уравнений*

$$dY_t = F(t)Y_t dt + C(t)dU_t, \text{ (модель системы)}$$
$$dX_t = G(t)Y_t dt + D(t)dV_t, \text{ (модель наблюдений)},$$

где F, C, G, D ограничены на ограниченных интервалах, задаётся уравнением

$$d\hat{Y}_t = F(t)\hat{Y}_t dt + \frac{G(t)S(t)}{D(t)^2}\left(dX_t - G(t)\hat{Y}_t dt\right),$$

где функция $S(t) = \mathbb{E}\left[(Y_t - \hat{Y}_t)^2\right]$ берётся из уравнения Рикатти

$$\frac{dS}{dt} = 2F(t)S(t) - \frac{G(t)^2}{D(t)^2}S^2(t) + C(t)^2, \; S(0) = \mathbb{E}\left[(Y_0 - \mathbb{E}[Y_0])^2\right].$$

Величину

$$K(t) = \frac{G(t)S(t)}{D(t)^2}$$

часто называют *Kalman gain*, а множитель справа – невязкой предсказания $G(t)\hat{Y}_t dt$.

Наконец, для полноты сформулируем многомерную версию.

Теорема 13.7. *(Фильтр Калмана-Бюси, многомерная версия) Решение $\hat{X}_t$ задачи фильтрации для уравнений*

$$dY_t = F(t)Y_t dt + C(t)dU_t, \text{ (модель системы)}$$
$$dX_t = G(t)Y_t dt + D(t)dV_t, \text{ (модель наблюдений)},$$

где F, C, G, D ограничены на ограниченных интервалах, задаётся уравнением

$$d\hat{Y}_t = F(t)\hat{Y}_t dt + SG^T(DD^T)^{-1}\left(dX_t - G(t)\hat{Y}_t dt\right),$$

где матричная функция $S(t) = \mathbb{E}\left[(Y_t - \hat{Y}_t)(Y_t - \hat{Y}_t)^T\right]$ берётся из уравнения Рикатти

$$\frac{dS}{dt} = FS + SF^T - SG^T(DD^T)^{-1}GS + CC^T, \; S(0) = \mathbb{E}\left[(Y_0 - \mathbb{E}[Y_0])(Y_0 - \mathbb{E}[Y_0])^T\right].$$

Фильтр Калмана по своей конструкции при известных параметрах системы способен решать все три обозначенные в начале задачи сразу.

1. **Задача фильтрации.** Решение уравнения сразу даёт функцию от наблюдения X_t.

2. **Задача сглаживания.** Сглаживание обновляет оценки фильтрации. В момент фильтрации мы строили проекцию только основываясь на $\mathcal{F}_t$, то есть на истории. Теперь мы можем пройти назад и использовать всё накопившееся знание, чтобы обновить более ранние оценки. Это делается с помощью специального сглаживающего уравнения, которое получил Закриссон[15]. Обозначим сглажённую оценку $\widetilde{Y}_s = \mathbb{E}[Y_s \mid \mathcal{F}_t]$, для неё можно положить $\widetilde{Y}_t = \hat{Y}_t$, а для $s < t$ получить уравнение
$$\frac{d}{ds}\widetilde{Y}_s = F(s)\widetilde{Y}_s + C(s)C^T(s)S^{-1}(s)(\widetilde{Y}_s - \hat{Y}_s).$$

3. **Задача предсказания** Это делается достаточно просто, как раньше мы решали уравнения с помощью интегрирующих множителей, будто нет случайности:

$$\mathbb{E}\left[Y_{t+h} \mid \mathcal{F}_t\right] = e^{\int_t^T F(s)ds}\hat{Y}_t.$$

Задача фильтрации изначально была решена в постановке Стратоновича и в теории фильтрации, а также в оптимальном управлении можно встретить много отсылок к его интегралу.

[15] L.E. Zachrisson. On optimal smoothing of continuous time kalman processes. *Information Sciences*, 1(2):143–172, 1969.

[16] Ширяев А.Н. Булинский А.В. *Теория случайных процессов.* М:ФИЗМАТЛИТ, 2005.

[17] Синай Я.Г. Коралов Л.Б. *Теория вероятностей и случайные процессы.* МЦНМО, 2013.

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

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