Оптимальная фильтрация случайных процессов

Олег Николаевич Граничин

Санкт-Петербургский государственный университет,
математико-механический факультет

13 марта 2013

Оптимальная фильтрация

Под *оптимальной фильтрацией* понимаются алгоритмы обработки реализаций случайных процессов, направленные на максимальное в смысле некоторого критерия подавление помех, зашумляющих (обычно аддитивно) полезный сигнал. В фундаменте теории оптимальной фильтрации лежит *метод Винера–Колмогорова* и его рекуррентные модификации, известные под общим названием *фильтра Калмана–Бьюси*. Теория Винера–Колмогорова в существенной степени базируется на методе наименьших квадратов. Оценивание параметров в этой теории происходит на основе обработки последовательно поступающих входных данных, являющихся некоторой траекторией стохастического процесса. Это приводит к важным концепциям физической реализуемости и оптимальности синтезируемого фильтра. Здесь будет описываться только случай дискретного времени.

Фильтр Винера–Колмогорова

Ограничимся рассмотрением следующей постановки задачи: последовательность наблюдений удовлетворяет уравнению

$$y_n = \varphi^{\mathrm{T}} \mathbf{x}_n + v_n,\ n = \ldots, -1, 0, 1, \ldots,$$

в котором $\{\mathbf{x}_n\}$ и $\{v_n\}$ — вещественные векторные процессы: $\mathbf{x}_n \in \mathbb{R}^r$, $v_n \in \mathbb{R}^p$; φ — прямоугольная матрица размерности $r \times p$.
Требуется получить оценку $\widehat{\mathbf{x}}_n$ процесса $\mathbf{x}_n$ в момент времени n по наблюдениям за процессом $\{y_n\}$ до момента времени $n - l$, l — заданное целое число. Оценка ищется с помощью *линейного устойчивого стационарного фильтра*, уравнение которого имеет вид

$$\widehat{\mathbf{x}}_n = \sum_{i=l}^{\infty} \mathrm{H}(i - l) y_{n-i},$$

где $\mathrm{H}(i)$ — *весовая функция фильтра.*

Передаточная функция фильтра

Введём *передаточную функцию фильтра*

$$H(\lambda) = \lambda^l \sum_{i=0}^{\infty} \mathrm{H}(i)\lambda^i.$$

Для упрощения ограничимся рассмотрением фильтров с дробно–рациональными передаточными функциями. Дробно–рациональную функцию $\lambda^{-l}H(\lambda)$ будем называть *устойчивой*, если у неё нет полюсов, которые по абсолютной величине меньше либо равны единице. Свойство устойчивости фильтра равносильно устойчивости функции $\lambda^{-l}H(\lambda)$. Задача фильтрации называется по-разному в зависимости от числа l в уравнении фильтра. При $l > 0$ её называют задачей *экстраполяции (прогноза)* на l моментов времени, при $l < 0$ — задачей *интерполяции (сглаживания)*, при $l = 0$ — собственно *фильтрацией*. Таким образом, при сглаживании оценка может зависеть от некоторого числа “будущих” наблюдений, а передаточная функция фильтра имеет полюс порядка l в начале координат.

Функционал качества

В классической постановке задачи рассматриваются стационарные случайные процессы $\{y_n\}$ и $\{\mathbf{x}_n\}$, которые вдобавок стационарно связаны, и их матрицы ковариаций вместе со спектральными плотностями: $\mathrm{B}_{yy}(n)$, $S_{yy}(\lambda)$, $\mathrm{B}_{y\mathbf{x}}(n)$, $S_{y\mathbf{x}}(\lambda)$, $\mathrm{B}_{\mathbf{xx}}(n)$, $S_{\mathbf{xx}}(\lambda)$, существуют и известны. Оценка $\widehat{\mathbf{x}}_n$ должна быть оптимальной в смысле минимума среднеквадратичного функционала

$$f_l = \mathrm{E}\{\|\widehat{\mathbf{x}}_n - \mathbf{x}_n\|^2\} \; (= f_l(H(\lambda))).$$

В силу стационарности процессов этот функционал не зависит от времени.

Стационарные случайные процессы

Последовательность $\{\xi_n\}$ случайных векторов ξ_n называется *стационарным* (в широком смысле) *процессом*, если среднее значение и ковариация не зависят от сдвига по времени. Для стационарного в широком смысле случайного процесса справедливо представление в виде стохастического интеграла Ито

$$\xi_n = \frac{1}{\sqrt{2\pi}} \int_0^{2\pi} e^{i\,\mu\,n} d\zeta_\mu + \bar{\xi}, \; -\infty < n < \infty,$$

где $\bar{\xi}$ — константа, а $\{\zeta_\mu\}$ — случайный процесс с некоррелированными центрированными приращениями, т. е. при любых $\mu_1 \leq \mu_2 \leq \mu_3 \leq \mu_4$ из интервала $[0;2\pi]$ удовлетворяющий условиям:

$$\mathrm{E}(\zeta_{\mu_1} - \zeta_{\mu_2})(\zeta_{\mu_3} - \zeta_{\mu_4})^\star = 0, \mathrm{E}(\zeta_{\mu_1} - \zeta_{\mu_2})(\zeta_{\mu_2} - \zeta_{\mu_1})^\star = U_{\xi\,\xi}(\mu_2) - U_{\xi\,\xi}(\mu_1),$$

с монотонно неубывающей (в смысле квадратичных форм) симметричной матричной функцией $U_{\xi\,\xi}(\cdot)$, называемой *спектральной (структурной) функцией* процесса $\{\xi_n\}$.

Спектральная плотность

Будем рассматривать только регулярные стационарные процессы, для которых элементы матрицы $U_{\xi\xi}(\mu)$ — абсолютно непрерывные функции, т. е. при почти всех (по мере Лебега) $\mu \in [0; 2\pi]$ существует производная

$$\bar{S}_{\xi\xi}(\mu) = \frac{dU_{\xi\xi}(\mu)}{d\mu}.$$

Спектральную функцию $U_{\xi\xi}(\mu)$ можно представить в виде

$$U_{\xi\xi}(\mu) = \int_0^{\mu} \bar{S}_{\xi\xi}(\mu) d\mu.$$

Матрица $\bar{S}_{\xi\xi}(\mu)$ называется *матрицей спектральных плотностей* (или *спектральной плотностью*) процесса $\{\xi_n\}$. $\bar{S}_{\xi\xi}(\mu)$ — неотрицательно определённая матрица и:

$$\mathrm{cov}\{\xi_k, \xi_l\} = \mathrm{B}_{\xi\xi}(k-l) = \frac{1}{2\pi}\int_0^{2\pi} e^{i\mu(k-l)} dU_{\xi\xi}(\mu) = \frac{1}{2\pi}\int_0^{2\pi} e^{i\mu(k-l)} \bar{S}_{\xi\xi}(\mu)$$

Стационарно связанные случайные процессы

Стационарные процессы $\{\xi_n\}$ и $\{\eta_n\}$ называются *стационарно связанными*, если совокупный процесс $\{\xi_n, \eta_n\}$ — стационарный. Для стационарно связанных центрированных процессов $\{\xi_n\}$ и $\{\eta_n\}$, имеющих спектральные плотности, справедливы соотношения

$$\mathrm{cov}\{\xi_k, \xi_l\} = \mathrm{B}_{\xi\xi}(k-l) = \frac{1}{2\pi}\int_0^{2\pi} e^{i\mu(k-l)}\bar{S}_{\xi\xi}(\mu)d\mu$$

$$\mathrm{cov}\{\xi_k, \eta_l\} = \mathrm{B}_{\xi\eta}(k-l) = \frac{1}{2\pi}\int_0^{2\pi} e^{i\mu(k-l)}\bar{S}_{\xi\eta}(\mu)d\mu,$$

где $\bar{S}_{\xi\eta}(\mu)$ — совместная спектральная плотность, $\mathrm{B}_{\xi\eta}(n)$ — матрица ковариации процессов. Вводя $\lambda = e^{-i\mu}$, имеем:

$$\mathrm{B}_{\xi\xi}(n) = \frac{1}{2\pi i}\oint \lambda^{-n} S_{\xi\xi}(\lambda)\frac{d\lambda}{\lambda}, \mathrm{B}_{\xi\eta}(n) = \frac{1}{2\pi i}\oint \lambda^{-n} S_{\xi\eta}(\lambda)\frac{d\lambda}{\lambda},$$

где $\oint$ — интеграл по единичной окружности: $\oint d\lambda/\lambda = 2\pi i$, и

$$S_{\xi\xi}(\lambda) = \bar{S}_{\xi\xi}(\mu),\ S_{\xi\eta}(\lambda) = \bar{S}_{\xi\eta}(\mu).$$

Минимизация функционала качества

Перепишем на основании определений матриц ковариации функционал качества f_l в виде:

$$f_l = \mathrm{Tr}[\mathrm{E}\{(\widehat{\mathbf{x}}_n - \mathbf{x}_n)(\widehat{\mathbf{x}}_n - \mathbf{x}_n)^{\mathrm{T}}\}] = \mathrm{Tr}[\mathrm{E}\{(\sum_{i=l}^{\infty} \mathrm{H}(i-l) y_{n-i} - \mathbf{x}_n) \times$$

$$\times (\sum_{j=l}^{\infty} \mathrm{H}(j-l) y_{n-j} - \mathbf{x}_n)^{\mathrm{T}}\}] = \mathrm{Tr}[\mathrm{B}_{\mathbf{xx}}(0) + \sum_{i=l}^{\infty}\sum_{j=l}^{\infty} \mathrm{H}(i)\mathrm{B}_{yy}(j-i)\mathrm{H}(j)^{\mathrm{T}} -$$

$$- \sum_{i=l}^{\infty} \mathrm{H}(i)\mathrm{B}_{y\mathbf{x}}(-i-l) - \sum_{i=l}^{\infty} \mathrm{B}_{\mathbf{x}y}(i+l)\mathrm{H}(j)^{\mathrm{T}}].$$

Известно, что преобразование спектральной плотности в линейном фильтре позволяет записать функционал качества f_l как квадратичную форму от передаточной функции фильтра:

$$f_l = \mathrm{Tr}[\mathrm{B}_{\mathbf{xx}}(0) +$$

$$+ \frac{1}{2\pi i} \oint (H(\lambda) S_{yy}(\lambda) H(\lambda^{-1})^{\mathrm{T}} - H(\lambda) S_{y\mathbf{x}}(\lambda) - S_{\mathbf{x}y}(\lambda) H(\lambda^{-1})^{\mathrm{T}}) \frac{d\lambda}{\lambda}].$$

Факторизация матричных функций

Пусть $S(\lambda)$ *— дробно–рациональная (матричная) функция (д.–р.ф.) с вещественными коэффициентами в матричных элементах, определенная и неотрицательная при всех* $|\lambda| = 1$.

Тогда *существует устойчивая д.–р.ф.* $\Pi(\lambda)$ *такая, что справедливо представление*

$$S(\lambda) = \Pi(\lambda)\Pi(\lambda^{-1})^{\mathrm{T}}$$

при всех комплексных значениях λ.

При этом, если $\det S(\lambda) \neq 0$ *при* $|\lambda| = 1$,

то $\Pi(\lambda)^{-1}$ *— устойчивая д.–р.ф.*

Преобразование функционала качества

Предположим, что матрица $S_{yy}(\lambda)$ — положительно определенная при $|\lambda| = 1$. В этом случае можно выбрать такую матрицу $\Pi(\lambda)$, чтобы $\Pi(\lambda)^{-1}$ была устойчивой. С помощью формулы факторизации, учитывая выполнение при $|\lambda| = 1$ соотношения

$$S_{\mathbf{x}y}(\lambda) = S_{y\mathbf{x}}(\lambda^{-1})^{\mathrm{T}},$$

получаем формулу

$$f_l = \mathrm{Tr}[\mathrm{Q} + \frac{1}{2\pi i}\oint (H(\lambda)\Pi(\lambda) - R(\lambda))(H(\lambda)\Pi(\lambda) - R(\lambda))^{\mathrm{T}} \frac{d\lambda}{\lambda}],$$

в которой использованы обозначения:

$$R(\lambda) = S_{\mathbf{x}y}(\lambda)(\Pi(\lambda^{-1})^{\mathrm{T}})^{-1}, \;\; \mathrm{Q} = \mathrm{B}_{\mathbf{xx}}(0) - \frac{1}{2\pi i}\oint R(\lambda)R(\lambda^{-1})^{\mathrm{T}} \frac{d\lambda}{\lambda}.$$

Оптимальный фильтр?

Так как матрица Q не зависит от $H(\lambda)$, а второе слагаемое в правой части полученной для f_l формулы — неотрицательная матрица, то минимум функционала качества f_l достигается при

$$H(\lambda) = R(\lambda)\Pi(\lambda)^{-1} = S_{\mathbf{x}y}(\lambda)S_{yy}(\lambda)^{-1},$$

причём минимальное значение функционала качества равно

$$\min_{\{H(\lambda)\}} f_l = \mathrm{Tr}[\mathrm{Q}].$$

Однако, найденное решение неудовлетворительно, поскольку, вообще говоря, не выполняется условие устойчивости фильтра, так как матрица $S_{yy}(\lambda)^{-1}$ может иметь особенности при $|\lambda| \leq 1$ и это свойство передаётся $H(\lambda)$.

Оптимальный устойчивый фильтр

Для получения устойчивого фильтра надо произвести *сепарацию* функции $R(\lambda)$, т. е. представить её в виде

$$\lambda^{-l}R(\lambda) = R_+(\lambda) + R_-(\lambda),$$

в котором $R_+(\lambda), R_-(\lambda^{-1})$ — устойчивые матричные функции и

$$\lim_{|\lambda|\to\infty} R_-(\lambda) = 0.$$

Основной результат теории оптимальной фильтрации Винера–Колмогорова заключается в том, что *при сделанных предположениях передаточная функция оптимального устойчивого фильтра, минимизирующего функционал качества f_l, определяется по формуле*

$$H(\lambda) = \lambda^l R_+(\lambda)\Pi(\lambda)^{-1},$$

и соотв. минимальное значение функционала качества равно

$$\min_{\{H(\lambda)\}} f_l = \mathrm{Tr}[\mathrm{Q}] + \frac{1}{2\pi i}\oint \mathrm{Tr}[R_-(\lambda)R_-(\lambda^{-1})^{\mathrm{T}}]\frac{d\lambda}{\lambda}.$$

Скалярный случай

В случае скалярного процесса y_n с дробно–рациональной спектральной плотностью процедура построения функции $\Pi(\lambda)$ сводится, по существу, к нахождению корней и полюсов дробно–рациональной функции $S_{yy}(\lambda)$, которые по абсолютной величине больше единицы. Сепарация функции $\lambda^{-l}R(\lambda)$ в этом случае состоит в выделении целой части функции с последующим определением "устойчивых" и "неустойчивых" полюсов у полученной в результате дробно–рациональной функции.

Пример: оптимальный прогноз процесса

Предположим, что наблюдается скалярный процесс $\{y_n\}$:

$$y_n = \varphi x_n + v_n,$$

где $\{x_n\}$ и $\{v_n\}$ — стохастически независимые стационарно связанные процессы: $\mathrm{E}\{v_n\} = 0$, $\mathrm{E}\{v_n^2\} = \sigma_v^2$, причём $\{x_n\}$ определяется уравнением

$$x_{n+1} = a\,x_n + w^{n+1},\ n = 1,2,\ldots,\ \mathbf{x}_0 = 0,$$

в котором $0 < |a| < 1$, $\mathrm{E}\{w^n\} = 0$, $\mathrm{E}\{w^i w^j\} = \sigma_w^2 \delta_{ij}$.
В данном случае

$$S_{vv} = \sigma_v^2,\ S_{\mathbf{xx}}(\lambda) = \frac{\sigma_w^2}{(1-a\lambda)(1-a\lambda^{-1})},\ S_{yy}(\lambda) = S_{vv} + \varphi^2 S_{\mathbf{xx}}(\lambda).$$

Факторизация спектральной плотности

Для факторизации $S_{yy}(\lambda)$ найдём c_1 и c_2 из уравнения

$$\sigma_v^2(1-a\lambda)(1-a\lambda^{-1})+\varphi^2\sigma_w^2=(c_1+c_2\lambda)(c_1+c_2\lambda^{-1}).$$

Несложные расчёты дают $c_1=\frac{1}{2}(\rho_1+\rho_2)$, $c_2=\frac{1}{2}(\rho_1-\rho_2)$, где
$\rho_1=\sqrt{\varphi^2\sigma_w^2+\sigma_v^2(1-a)^2}$, $\rho_2=\sqrt{\varphi^2\sigma_w^2+\sigma_v^2(1+a)^2}$.
Обозначив

$$\Pi(\lambda)=\frac{c_1+c_2\lambda}{1-a\lambda},$$

с учётом введенных обозначений, имеем

$$S_{yy}(\lambda)=\Pi(\lambda)\Pi(\lambda^{-1})^{\mathrm{T}}.$$

При этом $\Pi(\lambda)$ и $\Pi(\lambda)^{-1}$ — устойчивые из-за $|c_1|>|c_2|$. Т. к.

$$S_{y\mathbf{x}}(\lambda)=\frac{\varphi\sigma_w^2}{(1-a\lambda)(1-a\lambda^{-1})},$$

то

$$R(\lambda)=\frac{\varphi\sigma_w^2\lambda}{(1-a\lambda)(c_1\lambda+c_2)}.$$

Оптимальный прогноз на один шаг ($l = 1$)

В результате сепарации функции

$$\lambda^{-1} R(\lambda) = \frac{\varphi \sigma_w^2}{(1 - a\lambda)(c_1 \lambda + c_2)}.$$

получаем

$$R_+(\lambda) = \frac{\varphi \sigma_w^2 a}{c_1 + c_2 a} \cdot \frac{1}{1 - a\lambda}, \; R_-(\lambda) = \frac{\varphi \sigma_w^2 c_1}{c_1 + c_2 a} \cdot \frac{1}{c_1 \lambda + c_2}.$$

Следовательно, передаточная функция оптимального фильтра

$$H(\lambda) = \frac{\varphi \sigma_w^2 a}{c_1 + c_2 a} \cdot \frac{\lambda}{c_1 + c_2 \lambda}.$$

Т. е.$\widehat{x}_{n+1}$ и $\widehat{x}_n$ связаны соотношением

$$c_1 \widehat{x}_{n+1} + c_2 \widehat{x}_n = \frac{\varphi \sigma_w^2 a}{c_1 + c_2 a} y_n, \text{ или } \widehat{x}_{n+1} = a\widehat{x}_n - a\alpha\varphi(\varphi\widehat{x}_n - y_n),$$

где $\alpha = \frac{\sigma_w^2}{c_1(c_1 + c_2 a)} \quad \left(\equiv \frac{1}{a\varphi^2} \left(\frac{c_2}{c_1} + a \right) \right).$

Фильтр Калмана–Бьюси

Предположим, что наблюдается случайный процесс

$$y_n = \varphi_n^{\mathrm{T}} \mathbf{x}_n + v_n,\ n = 1, 2, \dots,$$

представляющий собой смесь преобразованного векторного процесса $\{\mathbf{x}_n\}$ и векторной помехи $\{v_n\}$. Прямоугольные матрицы φ_n этого преобразования считаются известными и могут изменяться во времени. Векторный процесс $\{\mathbf{x}_n\}$ порождается соотношением

$$\mathbf{x}_{n+1} = \mathrm{A}_n \mathbf{x}_n + w^{n+1},$$

в котором $\mathbf{x}_0 = 0$ и A_n — известная матричная функция времени, а $\{w_n\}$ — последовательность центрированных независимых случайных векторов с известными матрицами ковариации: $\mathrm{E}\{w^n w^{j^{\mathrm{T}}}\} = \mathrm{Q}_w(n)\delta_{nj}$. Обычно считают, что помеха $\{v_n\}$ также представляет собой последовательность центрированных независимых случайных векторов с известными матрицами ковариации $\mathrm{E}\{v_n v_j^{\mathrm{T}}\} = \mathrm{B}_v(n)\delta_{nj}$, которые при всех n невырожденные, и $\{\varphi_n\}$ — детерминированная последовательность матриц.

Задача об оптимальном одношаговом прогнозе

Ограничимся рассмотрением задачи об оптимальном одношаговом прогнозе. Для $n = 1, 2, \ldots$ *требуется* по наблюдениям $y_1, y_2, \ldots, y_n$ найти линейные оценки $\widehat{\mathbf{x}}_{n+1}$ значений процесса $\{\mathbf{x}_n\}$ в моменты времени $n+1$, минимизирующие среднеквадратичные отклонения

$$f_n = \mathrm{E}\{\|\widehat{\mathbf{x}}_{n+1} - \mathbf{x}_{n+1}\|^2\}.$$

Необходимые и достаточные условия оптимальности

Предположим, что оптимальная оценка имеет вид: $\widehat{\mathbf{x}}_{n+1} = \sum_{i=1}^{n} \mathrm{H}_n(i) y_i$. Если оценка $\widehat{\mathbf{x}}_{n+1}$ минимизирует функционал качества f_n, то для любого $j = 1, 2, \ldots, n$ выполнено условие

$$\mathrm{E}\{(\widehat{\mathbf{x}}_{n+1} - \mathbf{x}_{n+1}) y_j^{\mathrm{T}}\} = 0$$

или

$$\mathrm{E}\{\mathbf{x}_{n+1} y_j^{\mathrm{T}}\} = \sum_{i=1}^{n} \mathrm{H}_n(i) \mathrm{E}\{y_i y_j^{\mathrm{T}}\}.$$

Последнее выражение представляет собой нестационарный вариант *уравнения Винера–Хопфа* (в дискретном времени) относительно весовых функций $\mathrm{H}_n(i)$. Это соотношение имеет простой геометрический смысл: случайная величина $\widehat{\mathbf{x}}_{n+1}$, являющаяся линейной комбинацией случайных величин $y_1, \ldots, y_n$, должна быть строго ортогональной проекцией вектора $\mathbf{x}_{n+1}$ на подпространство, натянутое на соответствующие векторы наблюдений.

Вывод рекуррентного соотношения I

Обозначив $\mathrm{B}_{ij} = \mathrm{E}\{y_i y_j^{\mathrm{T}}\}$ и $\mathrm{K}_n = \mathrm{H}_n(n)$, из последнего уравнения, записанного для двух последовательных значений времени n и $n+1$, с одной стороны, получаем

$$\mathrm{E}\{(\mathbf{x}_{n+1} - \mathbf{x}_n) y_j^{\mathrm{T}}\} = \sum_{i=1}^{n-1} (\mathrm{H}_n(i) - \mathrm{H}_{n-1}(i)) \mathrm{B}_{ij} + \mathrm{K}_n \mathrm{B}_{nj}.$$

С другой стороны, учитывая вид фильтра, порождающего процесс $\{\mathbf{x}_n\}$, имеем

$$\mathrm{E}\{(\mathbf{x}_{n+1} - \mathbf{x}_n) y_j^{\mathrm{T}}\} = (\mathrm{A}_n - \mathrm{I}) \sum_{i=1}^{n-1} \mathrm{H}_{n-1}(i) \mathrm{B}_{ij}.$$

Из последних двух уравнений следует, что

$$\sum_{i=1}^{n-1} \left(\mathrm{A}_n \mathrm{H}_{n-1}(i) - \mathrm{K}_n \varphi_n^{\mathrm{T}} \mathrm{H}_{n-1}(i) - \mathrm{H}_n(i) \right) \mathrm{B}_{ij} = 0, \; j = 1, 2, \ldots, n-1,$$

так как в силу уравнения наблюдений

$$\mathrm{B}_{nj} = \mathrm{E}\{y_n y_j^{\mathrm{T}}\} = \varphi_n^{\mathrm{T}} \mathrm{E}\{\mathbf{x}_n y_j^{\mathrm{T}}\} + \mathrm{E}\{v_n y_j^{\mathrm{T}}\} = \varphi_n^{\mathrm{T}} \sum_{i=1}^{n-1} \mathrm{H}_{n-1}(i) \mathrm{B}_{ij}.$$

А значит, оценка $\tilde{\mathbf{x}}^n = \sum_{i=1}^{n-1} (\mathrm{H}_{n-1}(i) - \mathrm{D}_n(i)) y_i$, где

$$\mathrm{D}_n(i) = \mathrm{A}_n \mathrm{H}_{n-1}(i) - \mathrm{K}_n \varphi_n^{\mathrm{T}} \mathrm{H}_{n-1}(i) - \mathrm{H}_n(i),$$

также является оптимальной в среднеквадратичном смысле оценкой вектора $\mathbf{x}_n$ по наблюдениям $y_1, y_2, \ldots, y_{n-1}$. Поэтому

$$\mathrm{E}\{\|\tilde{\mathbf{x}}_n - \widehat{\mathbf{x}}_n\|^2\} = 0 \text{ или } \mathrm{E}\{\| \sum_{i=1}^{n-1} \mathrm{D}_n(i) \varphi_i^{\mathrm{T}} \mathbf{x}_i \|^2\} + \sum_{i=1}^{n-1} \mathrm{D}_n(i)^{\mathrm{T}} \mathrm{B}_v(i) \mathrm{D}_n(i) = 0.$$

Вывод рекуррентного соотношения III

Так как $\mathrm{B}_v(i) > 0$ при $i = 1, 2, \ldots, n-1$, то $\mathrm{D}_n(i) = 0$, т. е.

$$\mathrm{H}_n(i) = \mathrm{A}_n \mathrm{H}_{n-1}(i) - \mathrm{K}_n \varphi_n^{\mathrm{T}} \mathrm{H}_{n-1}(i).$$

Это и есть искомое соотношение, которому должна удовлетворять весовая функция оптимального фильтра. Учитывая его, несложно найти разностное уравнение для последовательности оптимальных оценок $\{\widehat{\mathbf{x}}_n\}$:

$$\widehat{\mathbf{x}}_{n+1} = \mathrm{K}_n y_n + \sum_{i=1}^{n-1} \mathrm{H}_n(i) y_i = \mathrm{K}_n y_n + \sum_{i=1}^{n-1} \left(\mathrm{A}_n \mathrm{H}_{n-1}(i) - \mathrm{K}_n \varphi_n^{\mathrm{T}} \mathrm{H}_{n-1}(i) \right) y_i =$$

$$= \mathrm{K}_n y_n + (\mathrm{A}_n - \mathrm{K}_n \varphi_n^{\mathrm{T}}) \widehat{\mathbf{x}}_n = \mathrm{A}_n \widehat{\mathbf{x}}_n - \mathrm{K}_n (\varphi_n^{\mathrm{T}} \widehat{\mathbf{x}}_n - y_n).$$

Матричные функции K_n называются *калмановскими коэффициентами усиления*.

Рекуррентные формулы

K_n непосредственно связаны с ковариационными матрицами ошибок оценивания

$$\Gamma_n = \mathrm{E}\{(\widehat{\mathsf{x}}_n - \mathsf{x}_n)(\widehat{\mathsf{x}}_n - \mathsf{x}_n)^{\mathrm{T}}\},$$

так как из уравнения Винера–Хопфа следует:

$$0 = \mathrm{E}\{(\widehat{\mathsf{x}}_{n+1} - \mathsf{x}_{n+1})(y_n - \varphi_n^{\mathrm{T}}\widehat{\mathsf{x}}_n)^{\mathrm{T}}\} = -(\mathrm{A}_n - \mathrm{K}_n\varphi_n^{\mathrm{T}})\Gamma_n\varphi_n + \mathrm{K}_n\mathrm{B}_v(n).$$

Сформулируем окончательный результат.

Калмановский коэффициент усиления K_n определяется по формуле

$$\mathrm{K}_n = \mathrm{A}_n\Gamma_n\varphi_n(\mathrm{B}_v(n) + \varphi_n^{\mathrm{T}}\Gamma_n\varphi_n)^{-1},$$

где Γ_n — ковариационная матрица ошибки оценивания, последовательность которых удовлетворяет рекуррентному соотношению

$$\Gamma_{n+1} = (\mathrm{A}_n - \mathrm{K}_n\varphi_n^{\mathrm{T}})\Gamma_n(\mathrm{A}_n - \mathrm{K}_n\varphi_n^{\mathrm{T}})^{\mathrm{T}} + \mathrm{K}_n\mathrm{B}_v(n)\mathrm{K}_n^{\mathrm{T}} + \mathrm{Q}_w(n+1).$$

Последняя формула легко выводится из разностного уравнения, связывающего две последовательные оценки.

Фильтр Калмана–Бьюси

Используя матричное тождество рекуррентные соотношения для матриц Γ_n можно переписать в виде

$$\Gamma_{n+1} = \mathrm{A}_n \Gamma_n \mathrm{A}_n^{\mathrm{T}} - \mathrm{K}_n \varphi_n^{\mathrm{T}} \Gamma_n \mathrm{A}_n^{\mathrm{T}} + \mathrm{Q}_w(n+1)$$

или

$$\Gamma_{n+1} = \mathrm{A}_n(\Gamma_n - \Gamma_n \varphi_n (\mathrm{B}_v(n) + \varphi_n^{\mathrm{T}} \Gamma_n \varphi_n)^{-1} \varphi_n^{\mathrm{T}} \Gamma_n)\mathrm{A}_n^{\mathrm{T}} + \mathrm{Q}_w(n+1).$$

После задания начальных значений $\widehat{\mathrm{x}}_0$ и Γ_0 вместе с формулой для последовательного пересчета оценок

$$\widehat{\mathrm{x}}_{n+1} = \mathrm{A}_n \widehat{\mathrm{x}}_n - \mathrm{A}_n \Gamma_n \varphi_n (\mathrm{B}_v(n) + \varphi_n^{\mathrm{T}} \Gamma_n \varphi_n)^{-1} (\varphi_n^{\mathrm{T}} \widehat{\mathrm{x}}_n - y_n)$$

эти соотношения, называемые *фильтром Калмана–Бьюси*, определяют замкнутую систему для рекуррентного вычисления $\widehat{\mathrm{x}}_n$ и Γ_n во все моменты времени n. Такие же формулы можно получить и при рассмотрении не только детерминированной последовательности $\{\varphi_n\}$, но и считая её реализацией некоторого матричного независимого случайного процесса, некоррелированного с помехами $\{v_n\}$ и с порождающим процессом $\{w_n\}$.

Упрощенный фильтр Калмана–Бьюси

Стоит заметить, что при $\mathrm{A}_n \equiv \mathrm{I}, \mathrm{Q}_w(n) \equiv 0$ и выборе матрицы весовых коэффициентов $\mathrm{R}_n = \mathrm{B}_v(n)^{-1}$ фильтр Калмана–Бьюси в точности совпадает с обобщенным рекуррентным МНК, что и неудивительно. На практике полученные соотношения часто упрощают, используя для вычисления оценок формулу

$$\widehat{\mathbf{x}}_{n+1} = \mathrm{A}_n \widehat{\mathbf{x}}_n - \mathrm{A}_n \Gamma \varphi_n \alpha (\varphi_n^{\mathrm{T}} \widehat{\mathbf{x}}_n - y_n)$$

с заданными положительно определенными матрицами Γ и α. Дальнейшее упрощение возможно при выборе скалярного значения $\alpha > 0$.

Рандомизированный фильтр

В случае вырожденных помех наблюдения $\{v_n\}$, в частности, при задании их неизвестными детерминированными ограниченными функциями, о качестве оценок фильтра Калмана–Бьюси трудно что–либо утверждать. Если предположить, что $\{\varphi_n\}$ — последовательность независимых одинаково распределенных случайных величин с известным средним значением и положительной ограниченной дисперсией, то для решения задачи о прогнозировании можно воспользоваться рандомизированным алгоритмом вида

$$\widehat{\mathsf{x}}_{n+1} = \mathrm{A}_n \widehat{\mathsf{x}}_n - \alpha \mathrm{A}_n \Gamma (\varphi_n - \mathrm{E}\{\varphi_n\})(\varphi_n^{\mathrm{T}} \widehat{\mathsf{x}}_n - y_n).$$

В условиях скалярных наблюдений на фоне неизвестной, но ограниченной неслучайной помехи получаемые по этому алгоритму оценки могут давать достаточно хорошее качество предсказания при $\mathrm{A}_n \equiv \mathrm{A} : \|\mathrm{A}\| \leq 1$.

Пример: оптимальный прогноз процесса

Предположим, что наблюдается скалярный процесс $\{y_n\}$

$$y_n = \varphi_n x_n + v_n,$$

где $\{\varphi_n\}$, $\{x_n\}$ и $\{v_n\}$ — стохастически независимые процессы: $\mathrm{E}\{v_n\} = 0$, $\mathrm{E}\{v_n^2\} = \sigma_v^2 > 0$, $\{x_n\}$ определяется уравнением:

$$x_{n+1} = a\, x_n + w_{n+1},\ n = 1, 2, \ldots,\ x_0 = 0,$$

в котором $0 < |a| \leq 1$, $\mathrm{E}\{w_n\} = 0$, $\mathrm{E}\{w_{n^2}\} = \sigma_w^2 > 0$.

В данном случае

$$\mathrm{B}_v(n) \equiv \sigma_v^2,\ \mathrm{Q}_w(n) \equiv \sigma_w^2$$

и при задании $\widehat{x}_0 = 0, \Gamma_0 = 0$ оптимальная последовательность прогнозирующих оценок вычисляется по формулам

$$\widehat{x}_{n+1} = a\widehat{x}_n - a\frac{\Gamma_n}{\sigma_v^2 + \Gamma_n\varphi_n^2}\varphi_n(\varphi_n\widehat{x}_n - y_n),$$

$$\Gamma_{n+1} = \sigma_w^2 + \frac{a^2\sigma_v^2}{\varphi_n^2} - \frac{a^2\sigma_v^4}{\varphi_n^2(\sigma_v^2 + \Gamma_n\varphi_n^2)} \left(\equiv a^2\left(\Gamma_n - \frac{\Gamma_n^2\varphi_n^2}{\sigma_v^2 + \Gamma_n\varphi_n^2}\right) + \sigma_w^2\right).$$

Стационарный случай

При $\varphi_n \equiv \varphi$ или в той ситуации, когда $\{\varphi_n\}$ — бернуллиевский процесс: $\varphi_n = \pm\varphi$, $\mathrm{E}\{\varphi_n\} = 0$, последовательность $\{\Gamma_n\}$ сходится к пределу Γ_∞

$$\Gamma_\infty = \sigma_w^2 + \frac{a^2\sigma_v^2}{\varphi^2} - \frac{a^2\sigma_v^4}{\varphi^2(\sigma_v^2 + \Gamma_\infty\varphi^2)},$$

решение которого

$$\Gamma_\infty = \frac{\varphi^2\sigma_w^2 + (a^2-1)\sigma_v^2 + \sqrt{(\varphi^2\sigma_w^2 + (a^2-1)\sigma_v^2)^2 + 4\varphi^2\sigma_w^2\sigma_v^2}}{2\varphi^2}.$$

Обозначив $\alpha = \frac{\Gamma_\infty}{\sigma_v^2 + \Gamma_\infty\varphi^2} = \frac{c_1^2 + c_1c_2/a}{\varphi^2 c_1^2} \left(\equiv \frac{1}{a\varphi^2}\left(\frac{c_2}{c_1} + a\right)\right)$, где
$c_1 = \frac{1}{2}(\rho_1 + \rho_2)$, $c_2 = \frac{1}{2}(\rho_1 - \rho_2)$,
$\rho_1 = \sqrt{\varphi^2\sigma_w^2 + \sigma_v^2(1-a)^2}$, $\rho_2 = \sqrt{\varphi^2\sigma_w^2 + \sigma_v^2(1+a)^2}$, в пределе при $n \to \infty$ получаем

$$\widehat{x}_{n+1} \approx a\widehat{x}_n - a\alpha\varphi_n(\varphi_n\widehat{x}_n - y_n).$$

Заметим, что при $a \approx 1$ и $\sigma_w << \sigma_v$ имеем $\alpha \approx \frac{\sigma_w}{\varphi\sigma_v}$.