1 Напоминание из линала. Сопряженный оператор

Определение Пусть X, Y – линейные пространства над $\mathbb{C}$, $A : X \to Y$ – линейный оператор. Сопряженным к нему называется оператор $A^* : Y^* \to X^*$, который действует по формуле: $(A^* y^*)x = y^*(Ax)$

Определение Пусть X – конечномерное векторное пространство над $\mathbb{C}$. Оно называется эрмитовым, если на нем задана форма $\langle \cdot, \cdot \rangle : X \times X \to \mathbb{C}$, обладающая следующими свойствами:
1. $\forall \alpha, \beta \in \mathbb{C},\ \forall x_1, x_2, y \in X\ \ \langle \alpha x_1 + \beta x_2, y \rangle = \alpha \langle x_1, y \rangle + \beta \langle x_2, y \rangle$;
2. $\forall x_1, x_2 \in X\ \ \langle x_1, x_2 \rangle = \overline{\langle x_2, x_1 \rangle}$;
3. $\forall x \in X\ \ \langle x, x \rangle \geqslant 0$, при этом $\langle x, x \rangle = 0 \iff x = 0$

Для эрмитова пространства есть другое понятие сопряженного оператора. Сначала докажем важное утверждение.

Теорема (**Каноническое отождествление**)
Пусть X – эрмитово пространство. Тогда отображение $\varphi : X \to X^*$, заданное формулой $(\varphi(x))(v) = \langle v, x \rangle$ – биекция.

Доказательство
Покажем, что оно инъективно. Пусть $\varphi(x_1) = \varphi(x_2)$, тогда $\langle v, x_1 \rangle = \langle v, x_2 \rangle$ для всех $v \in X$. То есть для всех v получаем $\langle v, x_1 - x_2 \rangle = 0$. Отсюда конечно же следует, что $x_1 - x_2 = 0$.
Теперь докажем сюръективность этого отображения. Зафиксируем ортонормированный базис $e_1, \dots, e_n$, а тогда всякий элемент $f \in X^*$ имеет вид $f(v) = \sum\limits_{j=1}^{n} \lambda_j v_j$, где v_j – коэффициенты разложения v по нашему базису. Теперь положим $x = \sum\limits_{j=1}^{n} \overline{\lambda_j} e_j$ и посчитаем $\varphi(x)$ на произвольном элементе v.

$$(\varphi(x))(v) = \langle v, x \rangle = \langle v, \sum_{j=1}^{n} \overline{\lambda_j} e_j \rangle = \sum_{j=1}^{n} \lambda_j \langle v, e_j \rangle = \sum_{j=1}^{n} \lambda_j v_j = f(v)$$

То есть наш функционал оказался в образе. Таким образом, мы проверили инъективность и сюръективность. ■

Определение Отображение из теоремы называется каноническим отождествлением.

Именно благодаря этому отображению мы и можем определить эрмитов сопряженный оператор. Это мы и сделаем.

Определение Пусть X, Y – эрмитовы пространства, $\varphi : X \to X^*$ – каноническое отождествление X и X^*, $\psi : Y \to Y^*$ – каноническое отождествление Y и Y^*, $A : X \to Y$ – линейный оператор.
Обозначим через $A^* : Y^* \to X^*$ – сопряженный оператор. Эрмитово сопряженным оператором к A называется линейный оператор $\hat{A} : Y \to X$, такой что следующая диаграмма коммутативна:

$$\begin{tikzcd}
Y \arrow[r, "\hat{A}"] \arrow[d, "\psi"'] & X \arrow[d, "\varphi"] \\
Y^* \arrow[r, "A^*"'] & X^*
\end{tikzcd}$$

Утверждение Эрмитов сопряженный оператор существует и единственен.

Доказательство
Мы хотим, чтобы было выполнено тождество $A^* \circ \psi = \varphi \circ \hat{A}$, отсюда сразу же следует, что $\hat{A} = \varphi^{-1} \circ A^* \circ \psi$, то есть это отображение единственно. Надо лишь проверить что оно линейно. Распишем $A^* \circ \psi = \varphi \circ \hat{A}$:
1. $((A^* \circ \psi)(y))(x) = ((A^*(\psi(y))(x) = (\psi(y))(Ax) = \langle Ax, y \rangle_Y$;
2. $((\varphi \circ \hat{A})(y))(x) = ((\varphi(\hat{A}y))(x) = \langle x, \hat{A}y \rangle_X$.
Отсюда получаем, что выполнено тождество $\langle x, \hat{A}y \rangle_X = \langle Ax, y \rangle_Y$ из которого сразу же вытекает линейность $\hat{A}$. ■

Отметим в частности, что из доказательства вытекает и другое равносильное определение. Впрочем, из него так просто не вытекает существование такого оператора.

Определение Пусть X, Y – эрмитовы пространства, $A : X \to Y$ – линейный оператор. Эрмитовым сопряженный к нему называют оператор $\hat{A} : Y \to X$, такой что выполнено тождество $\langle x, \hat{A}y \rangle_X = \langle Ax, y \rangle_Y$ для всяких $x \in X,\ y \in Y$.

Далее вместо $\hat{A}$ будем писать A^* в случае, если речь идет об эрмитовом пространстве.

Утверждение Пусть X, Y – эрмитовы пространства, $A : X \to Y$ – линейный оператор с матрицей A в ортонормированных базисах пространств X, Y. Тогда матрица сопряженного оператора в этих же базисах имеет вид $\overline{A^T}$.

Доказательство
Обозначим через a^*_{ij} элемент матрицы сопряженного оператора. Пусть вектора $\{x_i\}_i$ – ортонормированный базис X, а $\{y_j\}_j$ – ортонормированный базис в Y. Так как базис ортонормирован, то:

$$a^*_{ij} = (A^* y_j)_i = \langle A^* y_j, x_i\rangle_X = \overline{\langle x_i, A^* y_j\rangle_X} = \overline{\langle A x_i, y_j\rangle_Y} = \overline{a_{ji}}$$

Это и есть нужное нам тождество. ■

Определение Если A – квадратная матрица, то эрмитово сопряженной к ней называется матрица $\overline{A^T}$, которая обозначается A^*. Мы увидели, что это матрица сопряженного оператора в ортонормированном базисе.

2 Задача линейного оценивания и ее корректность

Определение Пусть $(\Omega, \mathcal{F}, \mathbb{P})$ – вероятностное пространство. Отображение $\xi : \Omega \to \mathbb{C}$ будем называть комплексной случайной величиной, если $Re\xi$ и $Im\xi$ являются случайными величинами. При этом считаем, что $\mathbb{E}\xi$ определено, если вещественная и мнимая части ξ интегрируемы. В этом случае $\mathbb{E}\xi := \mathbb{E}Re\xi + i\mathbb{E}Im\xi$.

Определение По аналогии со стандартными пространствами Лебега определим $L^p_{\mathbb{C}}(\mathbb{P})$, состоящее из классов эквивалентности комплексных случайных величин, для которых $\mathbb{E}|\xi|^p < \infty$.

Определение Пусть $(\Omega, \mathcal{F}, \mathbb{P})$ – вероятностное пространство. Отображение $\xi : \Omega \to \mathbb{C}^n$ будем называть комплексным случайным вектором, если $\xi = \begin{pmatrix} \xi_1 \\ \xi_2 \\ \dots \\ \xi_n \end{pmatrix}$, где все ξ_i являются комплексными случайными величинами. Математическим ожиданием вектора назовем вектор из математических ожиданий компонент, требуя, что все компоненты интегрируемы.

Далее мы будем считать, что все векторы имеют компоненты из $L^2_{\mathbb{C}}(\mathbb{P})$ – это гильбертово пространство со скалярным произведением $\langle \xi, \eta\rangle = \mathbb{E}\xi\overline{\eta}$. Тогда с каждым вектором можно связать ковариационную матрицу.

Определение Пусть $\xi : \Omega \to \mathbb{C}^n$ – комплексный случайный вектор. Его ковариационной матрицей назовем квадратную матрицу R_ξ размера $n \times n$, имеющую следующий вид:

$$R_\xi = \begin{pmatrix} cov(\xi_1, \xi_1) & \dots & cov(\xi_1, \xi_n) \\ \dots & \dots & \dots \\ cov(\xi_n, \xi_1) & \dots & cov(\xi_n, \xi_n) \end{pmatrix}$$

Здесь $cov(\zeta, \eta) = \mathbb{E}[(\zeta - \mathbb{E}\zeta)\overline{(\eta - \mathbb{E}\eta)}]$.
Заметим, что если величины центрированы, то ковариация – скалярное произведение в $L^2_{\mathbb{C}}(\mathbb{P})$.

Утверждение Матрица R_ξ – является эрмитовой и неотрицательно определенной.

Доказательство
То, что матрица эрмитова следует из того, что $\overline{cov(\xi_i, \xi_j)} = \overline{\mathbb{E}[(\xi_i - \mathbb{E}\xi_i)\overline{(\xi_j - \mathbb{E}\xi_j)}]} = \mathbb{E}[\overline{(\xi_i - \mathbb{E}\xi_i)}(\xi_j - \mathbb{E}\xi_j)] = cov(\xi_j, \xi_i)$. Теперь проверим, что она неотрицательно определена. Сделаем это по определению. Возьмем произвольный вектор $v = \begin{pmatrix} v_1 \\ v_2 \\ \dots \\ v_n \end{pmatrix}$ и рассмотрим выражение $v^* R_\xi v = \sum\limits_{\alpha,\beta=1}^{n} v_\alpha cov(\xi_\alpha, \xi_\beta)\overline{v_\beta} = cov(\sum\limits_\alpha \overline{v_\alpha}\xi_\alpha, \sum\limits_\alpha \overline{v_\alpha}\xi_\alpha) = \mathbb{E}\left|\sum\limits_{\alpha=1}^{n} \overline{v_\alpha}\xi_\alpha - \mathbb{E}\sum\limits_{\alpha=1}^{n} \overline{v_\alpha}\xi_\alpha\right|^2 \geqslant 0$.
Это и означает, что матрица неотрицательно определена. ■

Определение Рассмотрим два комплексных случайных вектора с компонентами из $L^2_{\mathbb{C}}(\mathbb{P})$, $\xi : \Omega \to \mathbb{C}^n, \eta : \Omega \to \mathbb{C}^p$. Их ковариационной матрицей называется прямоугольная матрица $R_{\xi\eta}$ размера $n \times p$, имеющая следующий вид:

$$R_{\xi\eta} = \begin{pmatrix} cov(\xi_1, \eta_1) & \dots & cov(\xi_1, \eta_p) \\ \dots & \dots & \dots \\ cov(\xi_n, \eta_1) & \dots & cov(\xi_n, \eta_p) \end{pmatrix}$$

Теперь мы введем частичный порядок на множестве эрмитовых матриц.

Определение Пусть A, B – эрмитовы матрицы одного размера. Будем говорить, что $A \leqslant B$, если матрица $B - A$ неотрицательно определена.

Определение Поставим математически задачу линейного оценивания. Пусть имеются случайные векторы $\xi : \Omega \to \mathbb{C}^n, \eta : \Omega \to \mathbb{C}^p$ с компонентами из $L^2_{\mathbb{C}}(\mathbb{P})$. Потребуем, что они имеют нулевые матожидания. Мы хотим найти $\hat{\xi}$ вида $K\eta$, где K – комплексная матрица размера $n \times p$.
Назовем решение $\hat{\xi} = K\eta$ сильным, если оно минимизирует $R_{\xi-\hat{\xi}}$ в смысле порядка, определенного выше.
Назовем решение $\hat{\xi} = K\eta$ слабым, если для всякого индекса j минимизируется $\mathbb{E}|\hat{\xi}_j - (K\eta)_j|^2$

Утверждение Сильное решение задачи является слабым решением.

Доказательство
В самом деле, пусть $\hat{\xi} = K_0\eta$ – сильное решение нашей задачи. Тогда для всякой матрицы K получим, что $R_{\xi-K_0\eta} \leqslant R_{\xi-K\eta}$, то есть матрица $R_{\xi-K\eta} - R_{\xi-K_0\eta}$ – неотрицательно определена. Рассмотрим $e_j^*(R_{\xi-K\eta} - R_{\xi-K_0\eta})e_j$, где e_j – вектор стандартного базиса. Тогда:

$$0 \leqslant e_j^*(R_{\xi-K\eta} - R_{\xi-K_0\eta})e_j = cov(\xi_j - (K\eta)_j, \xi_j - (K\eta)_j) - cov(\xi_j - (K_0\eta)_j, \xi_j - (K_0\eta)_j) = \mathbb{E}|\xi_j - (K\eta)_j|^2 - \mathbb{E}|\xi_j - (K_0\eta)_j|^2$$

То есть мы получаем, что $\mathbb{E}|\xi_j - \hat{\xi}_j|^2 \leqslant \mathbb{E}|\xi_j - (K\eta)_j|^2$. ■

Мы будем решать сильную задачу и переформулируем ее на языке евклидовых модулей над некоммутативными кольцами. Напомним несколько алгебраических понятий.

Определение Вещественной ассоциативной алгеброй с единицей A назовем векторное пространство над $\mathbb{R}$ вместе с билинейной ассоциативной операцией $\cdot : A \times A \to A$. При этом требуем, что в A есть единица относительно этой операции.

Пример
1. $M_n(\mathbb{R})$ является вещественной ассоциативной алгеброй над $\mathbb{R}$ размерности n^2 с обычным матричным умножением.
2. $M_n(\mathbb{C})$ является вещественной ассоциативной алгеброй над $\mathbb{R}$ размерности $2n^2$ с обычным матричным умножением.
3. Множество эрмитовых матриц фиксированного размера уже не является алгеброй, поскольку произведение эрмитовых матриц не всегда является эрмитовой матрицей.

Отметим, что вещественная ассоциативная алгебра с единицей конечно же является ассоциативным кольцом с единицей.

Определение Вещественной ассоциативной $*$-алгеброй с единицей A назовем вещественную ассоциативную алгебру с единицей, в которой есть дополнительная операция $* : A \to A$, обладающая следующими свойствами:
1. Она инволютивна, то есть $\forall a \in A \ \ (a^*)^* = a$;
2. $\forall a, b \in A \ \ (ab)^* = b^*a^*$;
3. $\forall \lambda \in \mathbb{R}, \ \forall a \in A \ \ (\lambda a)^* = \lambda a^*$

В примерах 1 и 2 мы можем взять в качестве операции $*$ транспонирование или эрмитово сопряжение и получить примеры $*$-алгебр. В дальшнейшем в качестве A у нас всегда будет выступать $M_n(\mathbb{C})$ с операцией эрмитова сопряжения.

Определение Пусть R – ассоциативное кольцо с единицей. Левым модулем над R мы назовем абелеву группу M вместе с операцией умножения $\cdot : R \times M \to M$, обладающей следующими свойствами:
1. Унитальность: $\forall m \in M \ \ 1m = m$;
2. $\forall r_1, r_2 \in R, \ \forall m \in M \ \ (r_1r_2)m = r_1(r_2m)$;
3. $\forall r_1, r_2 \in R, \ \forall m \in M \ \ (r_1 + r_2)m = r_1m + r_2m$;
4. $\forall m_1, m_2 \in M, \ \forall r \in R \ \ r(m_1 + m_2) = rm_1 + rm_2$;
Если R – это вещественная ассоциативная $*$-алгебра с единицей, то M мы называем левым $*$-модулем.

Определение Евклидовым левым $*$-модулем M назовем левый $*$-модуль над вещественная ассоциативная $*$-алгебра с единицей A, на котором задано отображение $\langle \cdot, \cdot \rangle : M \times M \to A$ со следующими свойствами:
1. $\forall a, b \in A, \ \forall u, v, w \in M \ \ \langle au + bv, w \rangle = a\langle u, w \rangle + b\langle v, w \rangle$;
2. $\forall u, v \in M \ \ \langle u, v \rangle^* = \langle v, u \rangle$;
3. $\|u\|^2 := \langle u, u \rangle = 0 \iff u = 0$
Рассмотрим теперь самый важный пример, с которым мы и будем работать.

Пример Рассмотрим два случайных вектора с нулевыми средними и компонентами из $L^2_{\mathbb{C}}(\mathbb{P})$ $\xi : \Omega \to \mathbb{C}^n, \eta : \Omega \to \mathbb{C}^p$. Построим по ним $M = \{B\xi + K\eta \mid B \in M_n(\mathbb{C}),\ K \in M_{n\times p}(\mathbb{C})\}$ – это какое-то множество классов эквивалентности n-мерных случайных векторов. Теперь заметим, что если в качестве вещественной алгебры A взять $M_n(\mathbb{C})$ с операцией эрмитового сопряжения, то M становится левым $*$-модулем над A. Определим на нем евклидову структуру, с помощью формы $\langle u, v\rangle = R_{uv}$. Тогда получаем на M как раз структуру евклидова $*$-модуля.

Что тут неочевидно? Почему выполнено самое последнее условие, то есть, если ξ имеет нулевое среднее и $R_\xi = 0$, то $\xi = 0$ почти всюду. Дело в том, что тогда $\mathbb{E}|\xi_j|^2 = 0$, ибо это элемент на диагонали, а тогда, конечно, $\xi_j = 0$ почти всюду.

Определение Модуль из примера выше назовем основным модулем для задачи линейного оценивания ξ по η и будем обозначать его $M_{\xi\eta}$

Теорема (Об ортогональной проекции)
Пусть $M_{\xi\eta}$ – основной модуль. Обозначим $L = \{K\eta \mid K \in M_{n\times p}(\mathbb{C})\}$. Тогда существует и единственна ортогональная проекция ξ на L, обозначаемая $\hat{\xi}$, то есть $\hat{\xi} \in L$ и для всякого элемента $l \in L$ выполнено $\langle \xi - \hat{\xi}, l\rangle = 0$.

Доказательство
Рассмотрим $V = span\{\xi_1, \xi_2, \dots, \xi_n, \eta_1, \eta_2, \dots, \eta_p\}$, $W = span\{\eta_1, \eta_2, \dots, \eta_p\}$. Это линейные подпространства в $L^2_{\mathbb{C}}(\mathbb{P})$. Выберем ортонормированный базис $\zeta_1, \dots, \zeta_M$ в W и дополним его до ортонормированного базиса всего V векторами $\zeta_{M+1}, \dots, \zeta_N$. Тогда каждую компоненту ξ_j вектора ξ можно разложить по этому базису. Если $\xi_j = \sum\limits_{k=1}^{N} \alpha_k\zeta_k$, то положим $\hat{\xi}_j = \sum\limits_{k=1}^{M} \alpha_k\zeta_k$. Покажем, что вектор $\hat{\xi} = \begin{pmatrix} \hat{\xi}_1 \\ \hat{\xi}_2 \\ \dots \\ \hat{\xi}_n \end{pmatrix}$ является искомой ортогональной проекцией. Прежде всего отметим, что $\hat{\xi} \in L$, поскольку $\hat{\xi}_j$ выражается через $\zeta_1, \dots, \zeta_M$, а тогда и через $\eta_1, \eta_2, \dots, \eta_p$. Это и означает, что $\hat{\xi} = K\eta$, то есть $\hat{\xi} \in L$. Теперь покажем, что $\hat{\xi}$ действительно ортогонально L. Возьмем произвольный элемент $l \in L$, он имеет вид $l = K\eta$. Посчитаем $\langle \xi - \hat{\xi}, K\eta\rangle$. Произвольная компонента этой матрицы с индексами α, β имеет вид:

$$cov(\xi_\alpha - \hat{\xi}_\alpha, (K\eta)_\beta) = cov\left(\xi_\alpha - \hat{\xi}_\alpha, \sum_{j=1}^{p} k_{\beta j}\eta_j\right)$$

Осталось заметить, что элемент $\xi_\alpha - \hat{\xi}_\alpha$ лежит в $span\{\zeta_{M+1}, \dots, \zeta_N\}$, а элемент $\sum\limits_{j=1}^{p} k_{\beta j}\eta_j$ лежит в $span\{\zeta_1, \dots, \zeta_M\}$.
В силу того, что базис ортонормированный, эти линейные оболочки ортогональны друг к другу, а поэтому $cov\left(\xi_\alpha - \hat{\xi}_\alpha, \sum\limits_{j=1}^{p} k_{\beta j}\eta_j\right) = 0$. Таким образом, $\langle \xi - \hat{\xi}, K\eta\rangle = 0$.
Мы доказали, что проекция существует. Почему она единственна? Предположим, что ξ_1, ξ_2 – две различные проекции. Тогда получаем, что:

$$\forall l \in L \ \ \langle \xi - \xi_1, l\rangle = \langle \xi - \xi_2, l\rangle = 0 \implies \forall l \in L \ \ \langle \xi_1 - \xi_2, l\rangle = 0 \implies \xi_1 = \xi_2$$

$\blacksquare$

Теперь мы можем доказать, что сильная задача линейного оценивания корректна, то есть ее решение существует и единственно. Заметим, что в модуле $M_{\xi\eta}$ имеем $\|\xi - K\eta\|^2 = R_{\xi - K\eta}$. Поэтому сильная задача – это минимизация квадрата нормы в модуле $M_{\xi\eta}$

Теорема Рассмотрим два случайных вектора с нулевыми средними и компонентами из $L^2_{\mathbb{C}}(\mathbb{P})$ $\xi : \Omega \to \mathbb{C}^n, \eta : \Omega \to \mathbb{C}^p$. Рассмотрим задачу $\|\xi - l\|^2 \to \min\limits_{l \in L}$ в модуле $M_{\xi\eta}$. Тогда верно следующее:
1. Эта задача имеет единственное решение;
2. Решение этой задачи – это ортогональная проекция ξ на L;
3. Элемент $\hat{K}\eta$ является решением задачи тогда и только тогда, когда $R_{\xi\eta} = \hat{K}R_\eta$

Доказательство

Итак, мы знаем уже, что существует ортогональная проекция $\hat{\xi}$. Возьмем произвольный элемент $K\eta \in L$ и сравним $\|\xi - \hat{\xi}\|$ с $\|\xi - K\eta\|$. Обозначим $\hat{\xi} = \hat{K}\eta$. Тогда получим:

$$\|\xi - K\eta\|^2 = \|\xi - \hat{K}\eta + (\hat{K} - K)\eta\|^2 = \langle \xi - \hat{K}\eta + (\hat{K} - K)\eta,\ \xi - \hat{K}\eta + (\hat{K} - K)\eta \rangle = \|\xi - \hat{\xi}\|^2 + \|(\hat{K} - K)\eta\|^2$$

То есть матрица $R_{\xi - K\eta} - R_{\xi - \hat{\xi}} = R_{(\hat{K} - K)\eta} \geqslant 0$, поскольку это ковариационная матрица. В самом деле мы получили, что ортогональная проекция является решением. Из этого же вычисления видно, что решение единственно. Таким образом, решение – это в точности ортогональная проекция. Осталось доказать последнюю часть:

$$\forall K \quad \langle \xi - \hat{K}\eta, K\eta \rangle = 0 \iff \forall K \quad \langle \xi, K\eta \rangle = \langle \hat{K}\eta, K\eta \rangle \iff \forall K \quad R_{\xi\eta}K^* = \hat{K}R_\eta K^* \iff R_{\xi\eta} = \hat{K}R_\eta$$

■

3 Оценка интеграла случайной функции по значениям в концах

Продемонстрируем пример применения развитой нами техники. Пусть $X_t,\ t \in [0, T]$ – действительный случайный процесс. Потребуем от него, что $X \in L^2(\lambda \otimes \mathbb{P})$, а также непрерывность его траекторий, центрированность ($\forall t\ \ \mathbb{E}X_t = 0$) и стационарность ($\mathbb{E}[X_t X_s] = \rho(t - s),\ t \geqslant s$). Обозначим $\xi = \int\limits_0^T X_t dt,\ \eta = \begin{pmatrix} X_0 \\ X_T \end{pmatrix}$.

Теорема Определенные отображения ξ, η являются случайными векторами с нулевыми средними и компонентами из $L^2(\mathbb{P})$. При этом $R_{\xi\eta} = \begin{pmatrix} \int\limits_0^T \rho(t)dt & \int\limits_0^T \rho(t)dt \end{pmatrix},\ R_\eta = \begin{pmatrix} \rho(0) & \rho(T) \\ \rho(T) & \rho(0) \end{pmatrix}$

Доказательство

Покажем сначала, что ξ случайная величина. В самом деле, интеграл от непрерывной функции есть предел своих римановых сумм, которые являются измеримыми. Дальше надо воспользоваться известным утверждением о том, что поточечный предел измеримых функций – измерим. Теперь покажем, что $\mathbb{E}|\xi|^2 < \infty$. В самом деле:

$$\mathbb{E}|\xi|^2 = \mathbb{E}\left(\int\limits_0^T X_t dt\right)^2 \leqslant \mathbb{E}\left[T\int_0^T X_t^2 dt\right] = T\int\limits_{[0,T]\times\Omega} |X|^2 d(\lambda \otimes \mathbb{P}) < \infty$$

Тот факт, что η – случайная величина с компонентами из $L^2(\mathbb{P})$ и нулевым средним очевиден. В силу наших предположений X интегрируем на квадрате, а тогда по теореме Фубини:

$$\mathbb{E}\xi = \mathbb{E}\int\limits_0^T X_t dt = \int_0^T \mathbb{E}X_t dt = 0$$

Теперь посчитаем ковариационные матрицы:

$$cov(\xi, \eta_0) = \mathbb{E}[\xi X_0] = \mathbb{E}\left[X_0 \int\limits_0^T X_t dt\right] = \mathbb{E}\left[\int\limits_0^T X_0 X_t dt\right] = \int\limits_0^T \mathbb{E}[X_0 X_t]dt = \int\limits_0^T \rho(t)dt$$

По аналогии считается $cov(\xi, \eta_1) = \int\limits_0^T \rho(t)dt$. Теперь найдем ковариации вектора η:

$$cov(\eta_0, \eta_0) = \mathbb{E}[X_0 X_0] = \rho(0),\ cov(\eta_0, \eta_1) = \mathbb{E}[X_0 X_T] = \rho(T)$$

Итого получаем, что $R_{\xi\eta} = \begin{pmatrix} \int\limits_0^T \rho(t)dt & \int\limits_0^T \rho(t)dt \end{pmatrix}, R_\eta = \begin{pmatrix} \rho(0) & \rho(T) \\ \rho(T) & \rho(0) \end{pmatrix}$ ■

4 Свойства решения сильной задачи

Определение Для двух случайных векторов ξ, η с нулевыми средними и компонентами из $L^2_{\mathbb{C}}(\mathbb{P})$ обозначим через $S(\xi, \eta)$ – решение сильной задачи линейного оценивания ξ по η. Это опредление корректно, поскольку мы уже знаем, что решение существует и единственно.

Определение Введем операцию склейки. Если $\eta_1 : \Omega \to \mathbb{C}^m, \eta_2 : \Omega \to \mathbb{C}^n$ – произвольные отображения, то их склейкой назовем отображение $\eta_1 \oplus \eta_2 : \Omega \to \mathbb{C}^{m+n},\ (\eta_1 \oplus \eta_2)(\omega) = \begin{pmatrix} \eta_1(\omega) \\ \eta_2(\omega) \end{pmatrix}$

Утверждение (**Свойства решения**)
1. $S(\xi_1 + \xi_2, \eta) = S(\xi_1, \eta) + S(\xi_2, \eta)$;
2. Если ξ имеет размерность n, то для всякой матрицы размера $m \times n$ выполнено: $S(L\xi, \eta) = LS(\xi, \eta)$;
3. Если η_1, η_2 некоррелированные (то есть $R_{\eta_1\eta_2} = 0$), то $S(\xi, \eta_1 \oplus \eta_2) = S(\xi, \eta_1) + S(\xi, \eta_2)$.

Доказательство
Мы будем пользоваться интерпретацией с ортогональной проекцией.
1. Хотим показать, что для всякой матрицы K выполнено $\langle (\xi_1 + \xi_2) - (S(\xi_1, \eta) + S(\xi_2, \eta)), K\eta \rangle = 0$. В самом деле:

$$\langle (\xi_1 + \xi_2) - (S(\xi_1, \eta) + S(\xi_2, \eta)), K\eta \rangle = \langle \xi_1 - S(\xi_1, \eta), K\eta \rangle + \langle \xi_2 - S(\xi_2, \eta), K\eta \rangle = 0$$

Это и означает, что $S(\xi_1, \eta) + S(\xi_2, \eta)$ является ортогональной проекцией $\xi_1 + \xi_2$.
2. Тут поступим немного иначе, а именно воспользуемся матричным уравнением. Итак, пусть $S(\xi, \eta) = K\eta$. Тогда $R_{\xi\eta} = KR_\eta$. Покажем, что $S(L\xi, \eta) = LK\eta$. В самом деле, $R_{L\xi,\eta} = \mathbb{E}[L\xi\eta^*] = L\mathbb{E}[\xi\eta^*] = LR_{\xi\eta} = LKR_\eta$.
По основной теореме мы и получаем требуемое утверждение.
3. Доказательство основано на наблюдении, что $K(\eta_1 \oplus \eta_2) = K_1\eta_1 + K_2\eta_2$ для некоторых матриц K_1, K_2, и наоборот. Из этого замечания ясно, что $S(\xi, \eta_1) + S(\xi, \eta_2)$ в самом деле лежит в $\{K(\eta_1 \oplus \eta_2)\}_K$. Осталось показать, что в самом деле выполнено условие ортогональности. Обозначим для удобства $S(\xi, \eta_1) = \hat{K}_1\eta_1,\ S(\xi, \eta_2) = \hat{K}_2\eta_2$. Тогда:

$$\langle \xi - S(\xi, \eta_1) - S(\xi, \eta_2), K(\eta_1 \oplus \eta_2) \rangle = \langle \xi - \hat{K}_1\eta_1 - \hat{K}_2\eta_2, K_1\eta_1 + K_2\eta_2 \rangle = \langle \xi - \hat{K}_1\eta_1 - \hat{K}_2\eta_2, K_1\eta_1 \rangle + \langle \xi - \hat{K}_1\eta_1 - \hat{K}_2\eta_2, K_2\eta_2 \rangle =$$

$$= \langle \xi - \hat{K}_1\eta_1, K_1\eta_1 \rangle - \langle \hat{K}_2\eta_2, K_1\eta_1 \rangle + \langle \xi - \hat{K}_2\eta_2, K_2\eta_2 \rangle - \langle \hat{K}_1\eta_1, K_2\eta_2 \rangle = 0$$

Здесь первое и третье слагаемые обнулились в силу того, что $\hat{K}_1\eta_1, \hat{K}_2\eta_2$ являются ортогональными проекциями, а второе и четвертое слагаемое равны нулю в силу некоррелированности η_1, η_2. ■

Теперь мы можем доказать, что сильная задача эквивалентна слабой. Воспользуемся следующей леммой:

Лемма Пусть имеется неотрицательно опредленная эрмитова матрица A, такая что на главной диагонали у нее только нули. Тогда $A = 0$.

Доказательство
Рассмотрим элемент матрицы a_{pq}. Применим к квадратичной форме $v = e_p - te_q,\ t \in \mathbb{R}$, где $\{e_k\}_k$ - стандартный базис. Тогда $v^*Av = a_{pp} - ta_{pq} - ta_{qp} + t^2a_{qq} = -t(a_{pq} + a_{qp}) = -2tRe(a_{pq})$. Это выражение становится отрицательным при некоторых t, если $Re(a_{pq}) \neq 0$. Итого мы показали, что все компоненты матрицы чисто мнимые. Теперь рассмотрим вектор $v = e_p - ite_q,\ t \in \mathbb{R}$. Применяя его к матрице A, мы получим $v^*Av = a_{pp} + ita_{qp} - ita_{pq} + t^2a_{qq} = it(a_{qp} - a_{pq})$. Осталось заметить, что в силу эрмитовости матрицы $it(a_{qp} - a_{pq}) = 2tIm(a_{pq})$. Отсюда следует, что и мнимые части равны нулю. ■

Теорема Сильная задача эквивалентна слабой.

Доказательство
Мы уже видели что сильное решение является слабым. Докажем обратное. Покажем, что решение слабой задачи является решением сильной. Обозначим $K\eta$ – решение слабой задачи. Мы уже знаем, что существует сильное решение $\hat{\xi}$, а поэтому $R_{\xi-\hat{\xi}} \leqslant R_{\xi-K\eta} \implies R_{\xi-K\eta} - R_{\xi-\hat{\xi}} \geqslant 0$. Осталось заметить, что матрица $R_{\xi-K\eta} - R_{\xi-\hat{\xi}}$ имеет нули на главной диагонали, значит она нулевая. Итого получаем, что $R_{\xi-K\eta} = R_{\xi-\hat{\xi}}$, а значит $K\eta$ – сильное решение. ■

В силу этой теоремы мы не будем теперь разделять сильные и слабые решения. Будем говорить просто о решениях задачи оценивания.

5 Алгебраическое доказательство существования и единственности

Мы покажем другой способ доказательства существования и единственности решения задачи линейного оценивания, который не использует модули и проекции. Вместо этого мы будем использовать соображения из теории операторов.

Лемма Пусть имеются эрмитовы пространства X, Y и линейный оператор $A : X \to Y$. Пусть A^* сопряженный оператор. Тогда верны следующие соотношения:
1. $KerA = (ImA^*)^\perp$;
2. $ImA = (KerA^*)^\perp$

Доказательство
Это следует сразу же из определений. Докажем сначала первую часть:

$$x \in KerA \iff Ax = 0 \iff \forall y \in Y \quad \langle Ax, y\rangle_Y = 0 \iff \forall y \in Y \quad \langle x, A^*y\rangle_X = 0 \iff x \in (ImA^*)^\perp$$

Теперь докажем вторую часть. Для этого воспользуемся уже доказанной первой частью для оператора A^*. Получим: $KerA^* = (ImA^{**})^\perp = (ImA)^\perp$. Навесим на все это ортогональное дополнение: $(KerA^*)^\perp = ImA$, поскольку $(L^\perp)^\perp = L$ в конечномерном пространстве. Таким образом, все доказано. ∎

Из этой леммы сразу же вытекает, что $\mathbb{C}^q = KerA \oplus (KerA)^\perp = KerA \oplus ImA^*$, поскольку по лемме $ImA^* = (KerA)^\perp$

Теорема Рассмотрим два случайных вектора с нулевыми средними и компонентами из $L^2_{\mathbb{C}}(\mathbb{P})$ $\xi : \Omega \to \mathbb{C}^n, \eta : \Omega \to \mathbb{C}^p$. Тогда уравнение $R_{\xi\eta} = XR_\eta$ имеет решение, при этом если X_1, X_2 – различные решения, то тогда $X_1\eta = X_2\eta$.

Доказательство
Сначала докажем единственность. Предположим, что найдутся различные решения X_1, X_2. Тогда $X_1R_\eta = X_2R_\eta \implies (X_1 - X_2)R_\eta = 0$. Тогда заметим, что $R_{(X_1-X_2)\eta} = (X_1 - X_2)R_\eta(X_1 - X_2)^* = 0 \implies X_1\eta = X_2\eta$.
Теперь перейдем к сущестованию. Зафиксируем стандартные базисы и стандартное эрмитово произведение в $\mathbb{C}^n, \mathbb{C}^p$ и отождествим матрицы $R_\eta, R_{\xi\eta}$ с линейыми операторами. Будем доказывать, что следующие утверждения эквивалентны:
1. Существует решение $R_{\xi\eta} = XR_\eta$;
2. $ImR_{\eta\xi} \subset ImR_\eta$;
3. $KerR_\eta \subset KerR_{\xi\eta}$.

Итак, $1 \implies 2$, поскольку существование решение у нашего уравнения равносильно существованию решения у сопряженного уравнения $R_{\eta\xi} = R_\eta X^*$, а тогда $R_{\eta\xi}v = R_\eta X^*v = R_\eta(X^*v) \in ImR_\eta$.
Для доказательства $2 \implies 1$ найдем для каждого базисного вектора $e_k \in \mathbb{C}^n$ вектор v_k, такой что $R_{\eta\xi}e_k = R_\eta v_k$. Это можно сделать в силу условия 2. Тогда оператор X^*, определенный на базисных элементах по формуле $X^*e_k = v_k$ является решением уравнения. Поэтому доказано $1 \iff 2$
Воспользуемся теперь разложением пространства в прямую сумму: $\mathbb{C}^p = KerR_\eta \oplus ImR_\eta = KerR_{\xi\eta} \oplus ImR_{\eta\xi}$.
Заметим, что прямая сумма тут ортогональна, поэтому мы моментально получаем $2 \iff 3$. Итого, нам осталось доказать, что если $R_\eta v = 0$, то $R_{\xi\eta}v = 0$. Действительно:

$$R_\eta v = 0 \implies \forall k \quad \mathbb{E}[\eta_k \sum_{i=1}^p v_i\overline{\eta_j}] = 0 \implies \mathbb{E}|\sum_{i=1}^p \overline{v_i}\eta_i|^2 = 0 \implies \sum_{i=1}^p \overline{v_i}\eta_i = 0 \implies \forall k \quad \mathbb{E}[\xi_k \sum_{i=1}^p v_i\overline{\eta_j}] = 0 \implies R_{\xi\eta}v = 0$$

∎

6 Псевдообратная матрица

Начнем с сингулярного разложения. Имеется теорема о том, что оно всегда существует.

Определение Квадратная комплексная матрица A называется унитарной, если $A^{-1} = A^*$.

Теорема (Сингулярное разложение)
Пусть $A \in M_{m\times n}(\mathbb{C})$. Тогда существуют унитарные матрицы $U \in M_m(\mathbb{C}), V \in M_n(\mathbb{C})$ и матрица $\Lambda \in M_{m\times n}(\mathbb{C})$, у которой все элементы вне главной диагонали нулевые, а на главной диагонали неотрицательные вещественные числа, такие что:

$$A = U\Lambda V^*$$

При этом, если матрица A была вещественной, то матрицы U, V также можно выбрать вещественными.

Определение Представление матрицы A в виде $A = U\Lambda V^*$ называется ее сингулярным разложением.

Заметим, что отсюда легко получить укороченную версию этого разожения.

Утверждение Всякую матрицу $A \in M_{m\times n}(\mathbb{C})$ можно представить в виде $U\Lambda V^*$, где $\Lambda \in M_r(\mathbb{C})$ – диагональная матрица из положительных чисел, $U \in M_{m\times r}(\mathbb{C}), V \in M_{n\times r}(\mathbb{C})$, $U^*U = V^*V = I$

Доказательство

Воспользуемся сингулярным разложением $A = U\Lambda V^*$. Обозначим через λ_{ij} элементы матрицы Λ.
Пусть $\lambda_{i_1i_1}, \lambda_{i_2i_2}, \ldots, \lambda_{i_ri_r}$ – все ненулевые элементы, причем $i_1 < i_2 < \ldots < i_r$. Обозначим через U' матрицу размера $m \times r$, получающуюся из U выкидыванием всех столбцов, кроме столбцов с индексами i_k. Аналогично обозначим через V' матрицу размера $n \times r$, получающуюся из V выкидыванием всех столбцов, кроме столбцов с индексами i_k. Положим $\Lambda' = \operatorname{diag}(\lambda_{i_1i_1}, \lambda_{i_2i_2}, \ldots, \lambda_{i_ri_r})$. Тогда очевидно, что $(U')^*U' = (V')^*V' = I$. Осталось заметить, что $U\Lambda V^* = U'\Lambda'(V')^*$. ■

Определение Разложение из последнего утверждения назовем усеченным сингулярным разложением.

Теперь мы можем определить важнейшее понятие псевдообратной матрицы.

Определение Пусть $A \in M_{m\times n}(\mathbb{C})$. Псевдообратной к ней называется матрица $A^+ \in M_{n\times m}(\mathbb{C})$, такая что выполняются следующие соотношения:
1. $AA^+A = A$;
2. $A^+AA^+ = A^+$;
3. $(AA^+)^* = AA^+$;
4. $(A^+A)^* = A^+A$

Отметим, что из пункта 2 ясно, что псевдообратная матрица к нулевой – нулевая.

Теорема (**Мур – Пенроуз – Тихонов**) Обратная матрица к любой существует и единственна. При этом, если $A \in M_{m\times n}(\mathbb{C})$, $A \neq 0$, то следующие условия эквивалентны:
1. A^+ – псевдообратная к A;
2. $A^+ = \lim\limits_{\delta\to 0} A^*(AA^* + \delta^2 I)^{-1} = \lim\limits_{\delta\to 0}(A^*A + \delta^2 I)^{-1}A^*$;
3. Если $A = U\Lambda V^*$ – усеченное сингулярное представление, то $A^+ = V\Lambda^{-1}U^*$

Псевдообратные матрицы замечательны тем, что с помощью них можно записать формулу решения системы линейных уравнений.

Теорема Система линейных уравнений $Ax = b$, $A \in M_{p\times q}(\mathbb{C})$, $b \in \mathbb{C}^p$ совместна тогда и только тогда, когда $AA^+b = b$. В этом случае ее общее решение имеет вид $\{A^+b + (I - A^+A)z \mid z \in \mathbb{C}^q\}$.

Доказательство

Если выполнено $AA^+b = b$, то конечно же есть решение $x = A^+b$. Пусть теперь выполнено $Ax = b$.
Тогда $AA^+b = AA^+Ax = Ax = b$. Первая часть доказана. Теперь выведем формулу общего решения.
Покажем, что $x = A^+b + (I - A^+A)z$ является решением:

$$Ax = A(A^+b + (I - A^+A)z) = AA^+b + Az - AA^+Az = b + Az - Az = b$$

Покажем теперь, что произвольное решение имеет такой вид:

$$x = A^+Ax + (I - A^+A)x = A^+b + (I - A^+A)x$$

■

Из этой теоремы сразу следует формула для решения системы нормальных уравнений. Отметим, что нам достаточно найти любое решение.

Теорема Одно из решений системы $R_{\xi\eta} = XR_\eta$ задается формулой $X = R_{\xi\eta}R_\eta^+$

Доказательство

Сопряжем уравнение, тогда получим $R_\eta X^* = R_{\eta\xi}$. Это можно интерпретировать, как набор систем линейных уравнений по каждому столбцу матрицы X^*. Отсюда, пользуясь теоремой находим $X^* = R_\eta^+R_{\eta\xi} + (I - R_\eta^+R_\eta)Z$, где Z – произвольная матрица. В силу того, что нас интересует одно из решений, положим $Z = 0$:

$$X^* = R_\eta^+R_{\eta\xi} \implies X = R_{\xi\eta}(R_\eta^+)^*$$

Осталось показать, что (R_η^+) является эрмитовой матрицей. В самом деле, пусть $R_\eta = U\Lambda V^*$ – усеченное сингулярное разложение. Тогда:

$$R_\eta^+ = V\Lambda^{-1}U^* \implies (R_\eta^+)^* = U\Lambda^{-1}V^* = (R_\eta^*)^+ = R_\eta^+$$

■

7 Ортогонализация

Предположим, что у нас имеется имеется набор центрированных случайных величин $\eta_1, \eta_2, \dots, \eta_N : \Omega \to \mathbb{C}^p$. Обозначим $\eta = \eta_1 \oplus \eta_2 \oplus \dots \eta_N$. Мы хотим найти некоторое невырожденное преобразование $e = A\eta$, чтобы матрица R_e была устроена максимально просто, например, чтобы она была блочно-диагональной. Оказывается, что так можно сделать.

Теорема Пусть $\eta_1, \eta_2, \dots, \eta_N : \Omega \to \mathbb{C}^p$ – набор центрированных случайных величин. Обозначим $e_1 = \eta_1$, $e_i = \eta_i - S(\eta_i, \eta_1 \oplus \eta_2 \oplus \dots \oplus \eta_{i-1})$. Тогда верно следующее:
1. e_i, e_j не коррелированы при различных i, j;
2. Для всякого номера k имеем $span\{e_i\}_{i=1}^k = span\{\eta_i\}_{i=1}^k$;
3. $S(\eta_i, \eta_1 \oplus \eta_2 \oplus \dots \oplus \eta_{i-1}) = \sum\limits_{j=1}^{i-1} R_{\eta_i e_j} R_{e_j}^+ e_j$

Доказательство
Второе свойство очевидно. Докажем первое. Пусть $i < j$. Тогда:

$$\langle e_j, e_i \rangle = \langle \eta_j - S(\eta_j, \eta_1 \oplus \eta_2 \oplus \dots \oplus \eta_{j-1}), e_i \rangle = \langle \eta_j - S(\eta_j, \eta_1 \oplus \eta_2 \oplus \dots \oplus \eta_{j-1}), K(\eta_1 \oplus \eta_2 \oplus \dots \oplus \eta_{j-1}) \rangle = 0$$

Все сразу получается из свойства ортогональной проекции. Третье же следует из того, что в силу второго свойства $S(\eta_i, \eta_1 \oplus \eta_2 \oplus \dots \oplus \eta_{i-1}) = S(\eta_i, e_1 \oplus e_2 \oplus \dots \oplus e_{i-1})$. А теперь в силу некореллированности воспользуемся свойством решения: $S(\eta_i, e_1 \oplus e_2 \oplus \dots \oplus e_{i-1}) = \sum\limits_{j=1}^{i-1} S(\eta_i, e_j)$. Итого получаем:

$$S(\eta_i, \eta_1 \oplus \eta_2 \oplus \dots \oplus \eta_{i-1}) = S(\eta_i, e_1 \oplus e_2 \oplus \dots \oplus e_{i-1}) = \sum_{j=1}^{i-1} S(\eta_i, e_j) = \sum_{j=1}^{i-1} R_{\eta_i e_j} R_{e_j}^+ e_j$$

■

В соотвествии с теоремой мы уже знаем, что $e_i = \eta_i - \sum\limits_{j=1}^{i-1} R_{\eta_i e_j} R_{e_j}^+ e_j$. Можно записать это и в виде $\eta_i = e_i + \sum\limits_{j=1}^{i-1} R_{\eta_i e_j} R_{e_j}^+ e_j$
При этом матрица L в представлении $\eta = Le$ имеет вид:

$$L = \begin{pmatrix} I_p & 0 & 0 & 0 & \dots & 0 \\ R_{\eta_2 e_1} R_{e_1}^+ & I_p & 0 & 0 & \dots & 0 \\ R_{\eta_3 e_1} R_{e_1}^+ & R_{\eta_3 e_2} R_{e_2}^+ & I_p & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ R_{\eta_N e_1} R_{e_1}^+ & R_{\eta_N e_2} R_{e_2}^+ & R_{\eta_N e_3} R_{e_3}^+ & R_{\eta_N e_4} R_{e_4}^+ & \dots & I_p \end{pmatrix}$$

Тогда в представлении $e = A\eta$ в качестве A можно взять L^{-1}.

8 Случай ненулевых средних

С самого начала мы рассмтривали оценки ξ по η, где ξ, η имеют нулевые средние. Теперь рассмотрим общий случай. Интуиция подсказывает нам, что логично свести эту задачу к уже известной, оценивая $\xi - \mathbb{E}\xi$ по $\eta - \mathbb{E}\eta$. Поставим теперь задачу строго математически.

Определение Рассмотрим два случайных вектора с компонентами из $L^2_{\mathbb{C}}(\mathbb{P})$ $\xi : \Omega \to \mathbb{C}^n, \eta : \Omega \to \mathbb{C}^p$.
Рассмотрим задачу минимизации матрицы квадратичной ошибки:

$$\|\xi - K\eta - l\|^2 := \begin{pmatrix} \mathbb{E}|\xi_1 - (K\eta)_1 - l_1|^2 & \dots & \mathbb{E}[(\xi_1 - (K\eta)_1 - l_1)\overline{(\xi_n - (K\eta)_n - l_n)}] \\ \dots & \dots & \dots \\ \mathbb{E}[(\xi_n - (K\eta)_n - l_n)\overline{(\xi_1 - (K\eta)_1 - l_1)}] & \dots & \mathbb{E}|\xi_n - (K\eta)_n - l_n|^2 \end{pmatrix} \to \min_{l \in \mathbb{C}^n, K \in M_{n \times p}}$$

Минимизация понимается тут как и раньше, в смысле порядка на эрмитовых матрицах. Отметим, что теперь уже мы работаем не с ковариационной матрицей, поскольку среднее не равно нулю.

Определение Введем обозначение. Для случайного вектора ξ обозначим $\xi^\circ = \xi - \mathbb{E}\xi$.
Если ξ, η – случайные векторы, то $\langle \xi, \eta \rangle$ – это матрица с элементами $\mathbb{E}[\xi_i \overline{\eta_j}]$ на месте (i, j).

Теорема Эта задача всегда имеет единственное решение $\hat{\xi} = K\eta + l$

Доказательство
Верно тождество: $\|\xi - K\eta - l\|^2 = \|\xi^\circ + \mathbb{E}\xi - K\eta^\circ - K\mathbb{E}\eta - l\|^2 = \|\xi^\circ - K\eta^\circ\|^2 + \|\mathbb{E}\xi - K\mathbb{E}\eta - l\|^2$, поскольку $\langle \xi^\circ - K\eta^\circ, \mathbb{E}\xi - K\mathbb{E}\eta - l\rangle = 0$. Пусть $K_0\eta^\circ = S(\xi^\circ, \eta^\circ)$ – решение центрированной задачи. Для нее мы уже знаем, что $\|\xi^\circ - K\eta^\circ\|^2 = \|\xi^\circ - K_0\eta^\circ\|^2 + \|(K - K_0)\eta^\circ\|^2$ по свойству ортогональной проекции. Отсюда получим:

$$\|\xi - K\eta - l\|^2 = \|\xi^\circ - K_0\eta^\circ\|^2 + \|(K - K_0)\eta^\circ\|^2 + \|\mathbb{E}\xi - K\mathbb{E}\eta - l\|^2$$

Значит, оптимальное решение будет достигаться при $K = K_0,\ l = \mathbb{E}\xi - K_0\mathbb{E}\eta$. Существование мы доказали. Из этой же формулы выше вытекает и единственность, по аналогии с тем, как мы делали для центрированной задачи. ■

9 Модель Калмана

Все величины, которые будут встречаться далее считаем из $L^2(\mathbb{P})$.

Определение Последовательность случайных векторов одной размерности $\{\xi_k\}_k$ назовем белым шумом, если $\mathbb{E}\xi_k = 0$ и все они попарно некоррелированы.

Определение Считаем, что $\{\xi_i\}_{i=0}^T$ – рекуррентная последовательность случайных векторов, заданная уравнением $\xi_{i+1} = F_i\xi_i + G_i u_i$, где F_i, G_i – заданные комплексные матрицы, $\{u_i\}$ – процесс белого шума. При этом имеется еще процесс измерений $\{\eta_i\}_{i=0}^T$, определенный формулой $\eta_i = H_i\xi_i + v_i$, где H_i – заданные комплексные матрицы, $\{v_i\}_{i=0}^T$ – процесс белого шума. Потребуем также, что при $j \neq i$ шумы u_i, v_j не коррелированы между собой. Изначальное состояние ξ_0 предполагается с нулевым средним и ковариационной матрицей Π_0. При этом мы считаем, что оно не коррелировано со всеми шумами. Итого, основные формулы в модели выглядят вот так:

$$\begin{cases} \xi_{i+1} = F_i\xi_i + G_i u_i \\ \eta_i = H_i\xi_i + v_i \end{cases}$$

Определение Введем стандартные обозначения в этой модели. Пусть $\dim\xi_i = n, \dim\eta_i = \dim v_i = p, \dim u_i = m$. Пусть Π_i – ковариационная матрица ξ_i, Q_i – ковариационная матрица u_i, R_i – ковариационная матрица v_i, S_i – матрица попарных ковариаций u_i, v_i.

Определение Задача линейного оценивания ξ_i по вектору $\eta_0 \oplus \eta_1 \oplus \ldots \oplus \eta_{i-1}$ называется задачей прогнозирования, а задача линейного оценивания ξ_i по вектору $\eta_0 \oplus \eta_1 \oplus \ldots \oplus \eta_i$ называется задачей фильтрации. Их решения обозначим через $\hat{\xi}_{i|i-1}$ и $\hat{\xi}_{i|i}$ соответсвенно.

Определение Часто мы будем проецировать величины именно на $\eta_0 \oplus \eta_1 \oplus \ldots \oplus \eta_{i-1}$. Поэтому для всякой центрированной величины ξ введем обозначение $\hat{\xi}_{|i-1} := S(\xi, \eta_0 \oplus \eta_1 \oplus \ldots \oplus \eta_{i-1})$

Лемма Из предположений модели Калмана вытекает, что ξ_i не коррелированно с u_j и v_j при $j \geqslant i$, а также η_i не коррелировано с u_j и v_j при $j > i$.

Доказательство
Докажем по индукции для ξ_i. При $i = 0$ это следует из требований модели, где мы явно указываем что начальное состояние не коррелировано со всеми шумами. Пусть теперь это верно для всех $k \leqslant i$. Возьмем u_j, v_j, где $j \geqslant i + 1$. Тогда из первого уравнения ξ_{i+1} линейно выражается через ξ_i, u_i, но каждое из них по предположению индукции не коррелировано с u_j, v_j. Отсюда же из второго уравнения сразу следует утверждения для η_i. ■

Утверждение Верна рекуррентная формула: $\Pi_{i+1} = F_i\Pi_i F_i^* + G_i Q_i G_i^*$.

Доказательство

$$\Pi_{i+1} = \mathbb{E}[\xi_{i+1}\xi_{i+1}^*] = \mathbb{E}[(F_i\xi_i + G_i u_i)(F_i\xi_i + G_i u_i)^*] = \mathbb{E}[F_i\xi_i\xi_i^* F_i^*] + \mathbb{E}[F_i\xi_i u_i^* G_i^*] + \mathbb{E}[G_i u_i\xi_i^* F_i^*] + \mathbb{E}[G_i u_i u_i^* G_i^*].$$

Осталось заметить, что для каждого i состояние ξ_i не коррелировано с u_i по лемме и применить линейность ожидания:

$$\Pi_{i+1} = \mathbb{E}[\xi_{i+1}\xi_{i+1}^*] = \mathbb{E}[F_i\xi_i\xi_i^* F_i^*] + \mathbb{E}[G_i u_i u_i^* G_i^*] = F_i\Pi_i F_i^* + G_i Q_i G_i^*$$

■

Применим к $\{\eta_i\}_i$ процесс ортогонализации и перейдем к последовательности $\{e_i\}_i$. Напомним, что $e_i = \eta_i - \hat{\eta}_{i|i-1}$.

Теорема (**Рекуррентные формулы для задачи прогнозирования**)
Верна следующая формула: $\hat{\xi}_{i+1|i} = F_{pi}\hat{\xi}_{i|i-1} + K_{pi}\eta_i,\ F_{pi} = (F_i - R_{\xi_{i+1}e_i}R^+_{e_i}H_i),\ K_{pi} = R_{\xi_{i+1}e_i}R^+_{e_i}$ с начальными условием $\hat{\xi}_{0|-1} = 0$.

Доказательство
Используя ортогональность e_i напишем тождество:

$$\hat{\xi}_{i+1|i} = S(\xi_{i+1}, \eta_0 \oplus \eta_1 \oplus \ldots \oplus \eta_i) = S(\xi_{i+1}, e_0 \oplus e_1 \oplus \ldots \oplus e_i) = S(\xi_{i+1}, e_0 \oplus e_1 \oplus \ldots \oplus e_{i-1}) + S(\xi_{i+1}, e_i) = \hat{\xi}_{i+1|i-1} + R_{\xi_{i+1}e_i}R^+_{e_i}e_i$$

Разберемся теперь с каждым слагаемым. В силу линейности оценок и рекуррентной формулы для ξ_i получим:

$$\hat{\xi}_{i+1|i-1} = (\widehat{F_i\xi_i + G_i u_i})_{|i-1} = F_i\hat{\xi}_{i|i-1} + G_i\hat{u}_{i|i-1}$$

По лемме u_i не коррелирован с вектором $\eta_1 \oplus \ldots \oplus \eta_{i-1}$, а поэтому $\hat{u}_{i|i-1} = 0 \implies \hat{\xi}_{i+1|i-1} = F_i\hat{\xi}_{i|i-1}$. Теперь проделаем похожий трюк с e_i:

$$e_i = \eta_i - \hat{\eta}_{i|i-1} = \eta_i - (\widehat{H_i\xi_i + v_i})_{|i-1} = \eta_i - H_i\hat{\xi}_{i|i-1}$$

Итого получаем: $\hat{\xi}_{i+1|i} = F_i\hat{\xi}_{i|i-1} + R_{\xi_{i+1}e_i}R^+_{e_i}(\eta_i - H_i\hat{\xi}_{i|i-1}) = (F_i - R_{\xi_{i+1}e_i}R^+_{e_i}H_i)\hat{\xi}_{i|i-1} + R_{\xi_{i+1}e_i}R^+_{e_i}\eta_i$ ■

Теперь мы получим уравнение Рикатти, играющую ключевую роль во всей теории.

Определение Обозначим через Σ_i ковариационную матрицу $\hat{\xi}_{i|i-1}$, а через $\tilde{\xi}_i = \xi_i - \hat{\xi}_{i|i-1}$.
При этом обозначим через P_i ковариационную матрицу $\tilde{\xi}_i$.

Теорема (**Уравнение Рикатти**)
Выполнено следующее рекуррентное уравнение, называемое уравнением Рикатти: $P_{i+1} = F_iP_iF_i^* + G_iQ_iG_i^* - K_{pi}R_{e_i}K_{pi}^*$ с начальным условием $P_0 = \Pi_0$.

Доказательство
Заметим, что из интерпретации оптимальной оценки как ортогональной проекции вытекает, что $\hat{\xi}_{i|i-1}$ не коррелировано с $\tilde{\xi}_i$. Поэтому получим:

$$\Pi_i = \mathbb{E}[\xi_i\xi_i^*] = \mathbb{E}[(\tilde{\xi}_i + \hat{\xi}_{i|i-1})(\tilde{\xi}_i + \hat{\xi}^*_{i|i-1})] = \mathbb{E}[\tilde{\xi}_i\tilde{\xi}_i^*] + \mathbb{E}[\tilde{\xi}_i\hat{\xi}^*_{i|i-1}] + \mathbb{E}[\hat{\xi}_{i|i-1}\tilde{\xi}_i^*] + \mathbb{E}[\hat{\xi}_{i|i-1}\hat{\xi}^*_{i|i-1}] = P_i + \Sigma_i \implies P_i = \Pi_i - \Sigma_i$$

Теперь воспользуемся рекуррентными формулами для Π_i и Σ_i. Для Π_i мы получали формулу выше, а для Σ_i выведем ее сейчас. Из доказательства теоремы выше следует, что $\hat{\xi}_{i+1|i} = F_i\hat{\xi}_{i|i-1} + K_{pi}e_i$. При этом, e_i не коррелировано с $\hat{\xi}_{i|i-1}$ по построению e_i. Итого получим:

$$P_{i+1} = \Pi_{i+1} - \Sigma_{i+1} = F_i\Pi_iF_i^* + G_iQ_iG_i^* - F_i\Sigma_iF_i^* - K_{pi}R_{e_i}K_{pi}^* = F_iP_iF_i^* + G_iQ_iG_i^* - K_{pi}R_{e_i}K_{pi}^*$$

■

Теорема (**Вычисление матрицы K_{pi}**)
Верны следующие формулы: $R_{e_i} = H_iP_iH_i^* + R_i,\ K_{pi} = (F_iP_iH_i^* + G_iS_i)R^+_{e_i}$

Доказательство
Сначала вычислим R_{e_i}. Заметим, что $e_i = \eta_i - \hat{\eta}_{i|i-1} = H_i\xi_i + v_i - H_i\hat{\xi}_{i|i-1} + \hat{v}_{i|i-1}$. Мы уже показывали, что v_i оказывается не коррелированным с $\eta_k, k < i$, поэтому $\hat{v}_{i|i-1} = 0$. Итого получаем:

$$e_i = H_i\xi_i + v_i - H_i\hat{\xi}_{i|i-1} = H_i\tilde{\xi}_i + v_i \implies R_{e_i} = H_iP_iH_i^* + R_i$$

В последнем равенстве мы использовали еще некоррелированность $\tilde{\xi}$ и v_i. Теперь посчитаем K_{pi}:

$$K_{pi} = \mathbb{E}[\xi_{i+1}e_i^*]R^+_{e_i} = \mathbb{E}[F_i\xi_ie_i^* + G_iu_ie_i^*]R^+_{e_i} = (F_i\mathbb{E}[\xi_ie_i^*] + G_i\mathbb{E}[u_ie_i^*])R^+_{e_i}$$

$$\mathbb{E}[\xi_ie_i^*] = \mathbb{E}[(\hat{\xi}_{i|i-1} + \tilde{\xi}_i)(H_i\tilde{\xi}_i + v_i)^*] = P_iH_i^*$$

Последнюю формулу мы получили благодаря тому, что величина $\hat{\xi}_{i|i-1}$ не коррелирована ни с $\tilde{\xi}_i$, ни с v_i. Вычислим теперь по аналогии оставшееся слагаемое:

$$\mathbb{E}[u_ie_i^*] = \mathbb{E}[(u_i(H_i\tilde{\xi}_i + v_i)^*] = S_i$$

Из этих вычислений непосредственно вытекает наша формула. ■

Сформулируем все полученные результаты про модель Калмана в виде единой теоремы:

Теорема (**Рекуррентные формулы для задачи прогнозирования**)
Верна следующая формула: $\hat{\xi}_{i+1|i} = F_{pi}\hat{\xi}_{i|i-1} + K_{pi}\eta_i$ с начальным условием $\hat{\xi}_{0|-1} = 0$, где $K_{pi} = (F_iP_iH_i^* + G_iS_i)R_{e_i}^+$, $R_{e_i} = H_iP_iH_i^* + R_i$, $F_{pi} = F_i - K_{pi}H_i$, а матрица P_i находится из уравнения Рикатти $P_{i+1} = F_iP_iF_i^* + G_iQ_iG_i^* - K_{pi}R_{e_i}K_{pi}^*$ с начальным условием $P_0 = \Pi_0$.

10 Пример использования фильтра для простейшей модели

Рассмотрим простую одномерную вещественную модель:

$$\begin{cases} \xi_{i+1} = \xi_i \\ \eta_i = \xi_i + v_i \end{cases}$$

Будем считать, что шумы датчика имеют постоянную дисперсию $s^2 > 0$ и ξ_0 имеет дисперсию $\sigma^2 > 0$.
Найдем оптимальную оценку $\hat{\xi}_{i+1|i}$. Для этого запишем, как выглядят наши формулы в конкретном случае:

$$\begin{cases} R_{e_i} = P_i + s^2 \\ K_{pi} = \frac{P_i}{P_i+s^2} \\ P_{i+1} = P_i - K_{pi}R_{e_i}K_{pi} = P_i - \frac{P_i^2}{P_i+s^2} = \frac{P_is^2}{P_i+s^2} \\ P_0 = \sigma^2 \end{cases}$$

В данном случае рекуррентное соотношение из уравнения Рикатти легко разрешается, откуда находим $P_i = \frac{s^2\sigma^2}{i\sigma^2+s^2}$. Поэтому итоговая формула выглядит следующим образом:

$$\hat{\xi}_{i+1|i} = (1 - K_{pi})\hat{\xi}_{i|i-1} + K_{pi}\eta_i = \frac{s^2}{P_i + s^2}\hat{\xi}_{i|i-1} + \frac{P_i}{P_i + s^2}\eta_i,\ P_i = \frac{s^2\sigma^2}{i\sigma^2 + s^2}$$

Заметим, что в этом случае мы имеем сходимость в $L^2(\mathbb{P})$ наших измерений к истинному значению, поскольку

$$\|\xi_0 - \hat{\xi}_{i|i-1}\|^2 = P_i = \frac{s^2\sigma^2}{i\sigma^2 + s^2} \to 0$$

11 Фильтр Калмана в форме Шмидта

Утверждение (**Коррекция прогноза**)
Верна следующая формула: $\hat{\xi}_{i|i} = \hat{\xi}_{i|i-1} + K_{fi}e_i$, где $K_{fi} = P_iH_i^*R_{e_i}^+$. При этом ковариационная матрица ошибки $\xi_i - \hat{\xi}_{i|i}$ имеет вид $P_i - K_{fi}R_{e_i}K_{fi}^*$.

Доказательство
Напомним, что в теореме о вычислении матрицы K_{pi} было получено тождество $\mathbb{E}[\xi_ie_i^*] = P_iH_i^*$. Воспользовавшись этим соотношением, получим:

$$\hat{\xi}_{i|i} = \hat{\xi}_{i|i-1} + R_{\xi_ie_i}R_{e_i}^+e_i = \hat{\xi}_{i|i-1} + P_iH_i^*R_{e_i}^+e_i$$

Теперь оценим погрешность. Из свойств ортогональной проекции и формулы для $\hat{\xi}_{i|i}$ получим:

$$\mathbb{E}[(\xi_i - \hat{\xi}_{i|i})(\xi_i - \hat{\xi}_{i|i})^*] = \mathbb{E}[(\xi_i - \hat{\xi}_{i|i})\xi_i^*] - \mathbb{E}[(\xi_i - \hat{\xi}_{i|i})\hat{\xi}_{i|i}^*] = \mathbb{E}[(\xi_i - \hat{\xi}_{i|i})\xi_i^*] = \mathbb{E}[(\tilde{\xi}_i - K_{fi}e_i)\xi_i^*] = P_i - K_{fi}\mathbb{E}[e_i\xi_i^*]$$

Осталось вычислить только $\mathbb{E}[e_i\xi_i^*]$. Для этого воспользуемся уже известным представлением $e_i = H_i\tilde{\xi}_i + v_i$:

$$\mathbb{E}[e_i\xi_i^*] = \mathbb{E}[(H_i\tilde{\xi}_i + v_i)\xi_i^*] = H_iP_i$$

Итого имеем $\mathbb{E}[(\xi_i - \hat{\xi}_{i|i})(\xi_i - \hat{\xi}_{i|i})^*] = P_i - K_{fi}H_iP_i = P_i - P_iH_i^*R_{e_i}^+H_iP_i = P_i - P_iH_i^*R_{e_i}^+R_{e_i}R_{e_i}^+H_iP_i = P_i - K_{fi}R_{e_i}K_{fi}^*$ ■

Определение Ковариационную матрицу ошибки $\tilde{\xi}_{i|i} := \xi_i - \hat{\xi}_{i|i}$ обозначим через $P_{i|i}$

Утверждение (**Прогноз по полной оценке**)
Верна следующая формула: $\hat{\xi}_{i+1|i} = F_i\hat{\xi}_{i|i} + G_iS_iR_{e_i}^+e_i$. При этом матрица ошибки имеет следующий вид:

$$P_{i+1} = F_iP_{i|i}F_i^* + G_i(Q_i - S_iR_{e_i}^+S_i^*)G_i^* - F_iK_{fi}S_i^*G_i^* - G_iS_iK_{fi}^*F_i^*$$

Доказательство
Из линейности оценки ясно, что $\hat{\xi}_{i+1|i} = F_i\hat{\xi}_{i|i} + G_i\hat{u}_{i|i}$. В силу некоррелированности u_i с $e_k, k < i$ получим, что $G_i\hat{u}_{i|i} = G_iR_{u_ie_i}R^+_{e_i}e_i = G_iS_iR^+_{e_i}e_i$. Теперь выведем формулу для матрицы ошибки:

$$\mathbb{E}[\tilde{\xi}_{i+1}\tilde{\xi}^*_{i+1}] = \mathbb{E}[(\xi_{i+1}-F_i\hat{\xi}_{i|i}-G_iS_iR^+_{e_i}e_i)(\xi_{i+1}-F_i\hat{\xi}_{i|i}-G_iS_iR^+_{e_i}e_i)^*] = \mathbb{E}[(F_i\tilde{\xi}_{i|i}+G_iu_i-G_iS_iR^+_{e_i}e_i)(F_i\tilde{\xi}_{i|i}+G_iu_i-G_iS_iR^+_{e_i}e_i)^*]$$

Теперь будем раскрывать это выражение по частям:

$$\mathbb{E}[(F_i\tilde{\xi}_{i|i}(F_i\tilde{\xi}_{i|i} + G_iu_i - G_iS_iR^+_{e_i}e_i)^*] = F_iP_{i|i}F^*_i + F_i\mathbb{E}[(\xi_i - \hat{\xi}_{i|i})u^*_i]G^*_i - F_i\mathbb{E}[(\xi_i - \hat{\xi}_{i|i})e^*_i]R^+_{e_i}S^*_iG^*_i$$

$$\mathbb{E}[(\xi_i - \hat{\xi}_{i|i})u^*_i] = -\mathbb{E}[\hat{\xi}_{i|i}u^*_i] = -\mathbb{E}[(\hat{\xi}_{i|i-1} + K_{fi}e_i)u^*_i] = -K_{fi}\mathbb{E}[(H_i\tilde{\xi} + v_i)u^*_i] = -K_{fi}S^*_i$$

$$\mathbb{E}[(\xi_i - \hat{\xi}_{i|i})e^*_i] = \mathbb{E}[(\tilde{\xi}_i - K_{fi}e_i)e^*_i] = \mathbb{E}[\tilde{\xi}_i(H_i\tilde{\xi}_i + v_i)^*] - K_{fi}R_{e_i} = P_iH^*_i - K_{fi}R_{e_i}$$

Теперь заметим, что на самом деле слагаемое из последней строчки даст нулевой вклад:

$$F_i\mathbb{E}[(\xi_i - \hat{\xi}_{i|i})e^*_i]R^+_{e_i}S^*_iG^*_i = F_i(P_iH^*_i - K_{fi}R_{e_i})R^+_{e_i}S^*_iG^*_i = F_iP_iH^*_iR^+_{e_i}S^*_iG^*_i - F_iP_iH^*_iR^+_{e_i}R_{e_i}R^+_{e_i}S^*_iG^*_i = 0$$

Итого первая группа слагаемых имеет вид:

$$\mathbb{E}[(F_i\tilde{\xi}_{i|i}(F_i\tilde{\xi}_{i|i} + G_iu_i - G_iS_iR^+_{e_i}e_i)^*] = F_iP_{i|i}F^*_i - F_iK_{fi}S^*_iG^*_i$$

Аналогично считаются остальные слагаемые. ■

Благодаря этим утверждениям мы можем получить другой алгоритм. В классической версии фильтра Калмана мы действуем по следующей цепочке вычислений:

$$\hat{\xi}_{1|0} \to \hat{\xi}_{2|1} \to \hat{\xi}_{3|2} \to \hat{\xi}_{4|3} \to \dots$$

Фильтр Калмана в форме Шмидта, предполагает другой порядок вычислений:

$$\hat{\xi}_{1|0} \to \hat{\xi}_{1|1} \to \hat{\xi}_{2|1} \to \hat{\xi}_{2|2} \to \dots$$

12 Регуляризация по Тихонову

Рассмотрим модель Калмана с абсолютно точным датчиком, то есть при $v_i = 0$: $\begin{cases}\xi_{i+1} = F_i\xi_i + G_iu_i \\ \eta_i = H_i\xi_i\end{cases}$.

Тогда уравнение Рикатти примет вид: $P_{i+1} = F_iP_iF^*_i + G_iQ_iG^*_i - K_{pi}R_{e_i}K^*_{pi}$, где $K_{pi} = F_iP_iH^*_iR^+_{e_i}$, $R_{e_i} = H_iP_iH^*_i$. Мы хотели бы решить это уравнение приближенно, не используя псевдообратную матрицу. Это можно сделать с помощью регуляризации по Тихонову. Но сначала нам пригодится важная теорема:

Теорема Если $A \in M_n(\mathbb{C})$ – неотрицательно определенная матрица, то существует $B \in M_n(\mathbb{C})$, такая что $A = BB^*$.

Лемма Для всякой матрицы A справедливо равенство $A^+ = A^*(AA^*)^+$.

Доказательство
Для нулевой матрицы это очевидно. Рассмотрим теперь произвольную ненулевую матрицу A и напишем для нее усеченное сингулярное разложение $A = U\Lambda V^*$, где $U^*U = V^*V = I$. По определению тогда $A^+ = V\Lambda^{-1}U^*$. Но при этом:

$$A^*(AA^*)^+ = V\Lambda U^*(U\Lambda V^*V\Lambda U^*)^+ = V\Lambda U^*(U\Lambda^2U^*)^+ = V\Lambda U^*U\Lambda^{-2}U^* = V\Lambda^{-1}U^* = A^+$$

■

Теорема (**Регуляризация по Тихонову**)
Верны следующие утверждения:
1. $K_{pi}R_{e_i}K^*_{pi} = F_iP_iH^*_iR^+_{e_i}H_iP_iF^*_i$;
2. Для всякого $\delta > 0$ матрица $H_iP_iH^*_i + \delta^2I$ обратима;
3. $F_iP_iH^*_iR^+_{e_i}H_iP_iF^*_i = \lim\limits_{\delta\to0} F_iP_iH^*_i(H_iP_iH^*_i + \delta^2I)^{-1}H_iP_iF^*_i$.

Доказательство
Свойство 1 следует из определения псевдообратной матрицы. Свойство 2 следует из того, что $H_iP_iH^*_i$ неотрицательно определена, а δ^2I положительно определена. Осталось обосновать законность перехода к пределу. Воспользовавшись теоремой выше найдем матрицу A, такую что $AA^* = P_i$. Тогда получим:

$$\lim_{\delta\to0} P_iH^*_i(H_iP_iH^*_i + \delta^2I)^{-1} = \lim_{\delta\to0} AA^*H^*_i(H_iAA^*H^*_i + \delta^2I)^{-1} = \lim_{\delta\to0} A(H_iA)^*((H_iA)(H_iA)^* + \delta^2I)^{-1} = A(H_iA)^+$$

Здесь мы просто воспользовались определение псевдообратной матрицы. Теперь применим лемму:

$$A(H_iA)^+ = AA^*H^*_i(H_iAA^*H^*_i)^+ = P_iH^*_i(H_iP_iH^*_i)^+ = P_iH^*_iR^+_{e_i}$$

■

13 Фильтр для нецентрированных величин

Рассмотрим стандартную модель: $\begin{cases} \xi_{i+1} = F_i\xi_i + G_iu_i \\ \eta_i = H_i\xi_i + v_i \end{cases}$. Теперь будем считать, что ξ_0 имеет ненулевое среднее. Оценивать на каждом шаге будем аффинными оценками, подобно тому, как мы действовали в 8 части. Оптимальную аффинную оценку ξ по $\eta_0 \oplus \eta_1 \oplus \ldots \oplus \eta_i$ теперь будем обозначать $\hat{\xi}_{|i}$. Это не вызовет недоразумений, поскольку в случае центрированных величин наилучшая оценка будет аффинной. Это следует сразу же из прошлых результатов о формуле наилучшего аффинного прогноза. Оказывается, что формулы почти не изменятся, а именно верна следующая теорема:

Теорема (**Прогноз для нецентрированных величин**)
Верна формула: $\hat{\xi}_{i+1|i} = F_i\hat{\xi}_{i|i-1} + K_{pi}(\eta_i - H_i\hat{\xi}_{i|i-1})$ с начальным условием $\hat{\xi}_{0|-1} = \mathbb{E}\xi_0$.
Все матрицы тут опредляются для центрированных частей.

Доказательство
Итак, заметим что $\xi_{i+1} = F_i\xi_i + G_iu_i \implies \mathbb{E}\xi_{i+1} = F_i\mathbb{E}\xi_i$, а поэтому из равенства $\xi^\circ_{i+1} + \mathbb{E}\xi_{i+1} = F_i\xi^\circ_i + F_i\mathbb{E}\xi_i + G_iu_i$ вытекает $\xi^\circ_{i+1} = F_i\xi^\circ_i + G_iu_i$. Аналогично получается равенство $\eta^\circ_i = H_i\xi^\circ_i + v_i$. То есть центрированные величины подчиняются обычной модели Калмана: $\begin{cases} \xi^\circ_{i+1} = F_i\xi^\circ_i + G_iu_i \\ \eta^\circ_i = H_i\xi^\circ_i + v_i \end{cases}$. Тогда, в связи с пунктом 8, получим:

$$\hat{\xi}_{i+1|i} = S(\xi^\circ_{i+1}, \eta^\circ_0 \oplus \ldots \eta^\circ_i) + \mathbb{E}\xi_{i+1} = F_i\hat{\xi}^\circ_{i|i-1} + K_{pi}(\eta^\circ_i - H_i\hat{\xi}^\circ_{i|i-1}) + F_i\mathbb{E}\xi_i = F_i\xi_{i|i-1} + K_{pi}(\eta_i - H_i\mathbb{E}\xi_i - H_i\hat{\xi}_{i|i-1} + H_i\mathbb{E}\hat{\xi}_{i|i-1})$$

Осталось заметить, что в силу формулы аффинной проекции $H_i\mathbb{E}\xi_i = H_i\mathbb{E}\hat{\xi}_{i|i-1}$. Осталось только положить начальное условие. Мы хотим, чтобы $\hat{\xi}_{1|0} = F_0\xi_{0|-1} + K_{p0}(\eta_0 - H_0\xi_{0|-1})$. Подставляя $\xi_{0|-1} = \mathbb{E}\xi_0$ получим:

$$\hat{\xi}_{1|0} = F_0\mathbb{E}\xi_0 + K_{p0}(\eta_0 - H_0\mathbb{E}\xi_0) = \mathbb{E}\xi_1 + K_{p0}(\eta_0 - \mathbb{E}\eta_0) = \mathbb{E}\xi_1 + K_{p0}\eta^\circ$$

Эта формула согласуется с нашими вычислениями, а значит начальное значение найдено верно. ■

14 Дополнение Шура

Обсудим понятие дополнения Шура и получим с помощью него некоторые матричные равенства.

Определение Рассмотрим квадратную блочную матрицу $M = \begin{pmatrix} A & B \\ C & D \end{pmatrix}$, в которой подматрицы A и D – квадратные. Предположим, что A и D обратимы. Тогда $\Delta_A = D - CA^{-1}B$ назовем дополнением Шура блока A, а $\Delta_D = A - BD^{-1}C$ назовем дополнением Шура блока D.
Смысл этих определений в том, что в их терминах получаются удобные матричные разложения.

Теорема (**Блочный метод Гаусса**)
Рассмотрим квадратную блочную матрицу $M = \begin{pmatrix} A & B \\ C & D \end{pmatrix}$, в которой подматрицы A и D – квадратные. Предположим, что A и D обратимы. Тогда выполняются следующие соотношения:
1. $\begin{pmatrix} I & 0 \\ -CA^{-1} & I \end{pmatrix} \begin{pmatrix} A & B \\ C & D \end{pmatrix} \begin{pmatrix} I & -A^{-1}B \\ 0 & I \end{pmatrix} = \begin{pmatrix} A & 0 \\ 0 & \Delta_A \end{pmatrix}$;
2. $\begin{pmatrix} I & -BD^{-1} \\ 0 & I \end{pmatrix} \begin{pmatrix} A & B \\ C & D \end{pmatrix} \begin{pmatrix} I & 0 \\ -D^{-1}C & I \end{pmatrix} = \begin{pmatrix} \Delta_D & 0 \\ 0 & D \end{pmatrix}$;
3. $\begin{pmatrix} A & B \\ C & D \end{pmatrix} = \begin{pmatrix} I & 0 \\ -CA^{-1} & I \end{pmatrix} \begin{pmatrix} A & 0 \\ 0 & \Delta_A \end{pmatrix} \begin{pmatrix} I & A^{-1}B \\ 0 & I \end{pmatrix} = \begin{pmatrix} I & BD^{-1} \\ 0 & I \end{pmatrix} \begin{pmatrix} \Delta_D & 0 \\ 0 & D \end{pmatrix} \begin{pmatrix} I & 0 \\ D^{-1}C & I \end{pmatrix}$;
4. $\det \begin{pmatrix} A & B \\ C & D \end{pmatrix} = \det A \det \Delta_A = \det D \det \Delta_D$;
5. $\begin{pmatrix} A & B \\ C & D \end{pmatrix} = \begin{pmatrix} A \\ C \end{pmatrix} A^{-1} \begin{pmatrix} A & B \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ 0 & \Delta_A \end{pmatrix} = \begin{pmatrix} B \\ D \end{pmatrix} D^{-1} \begin{pmatrix} C & D \end{pmatrix} + \begin{pmatrix} \Delta_D & 0 \\ 0 & 0 \end{pmatrix}$

Доказательство
Все соотношения легко проверяются перемножением матриц. 4 следует из 1 и 2. Например, первое равенство следует из $\det \begin{pmatrix} I & 0 \\ -CA^{-1} & I \end{pmatrix} = \det \begin{pmatrix} I & -A^{-1}B \\ 0 & I \end{pmatrix} = 1$. Аналогично получается и второе равенство. ■

Теорема (**Формула обращения в терминах дополнений Шура**)
Предположим, что квадратная блочная матрица $M = \begin{pmatrix} A & B \\ C & D \end{pmatrix}$, с квадратными подматрицами A и D – обратима. Тогда верны следующие тождества:

$$\begin{pmatrix} I & -A^{-1}B \\ 0 & I \end{pmatrix} \begin{pmatrix} A^{-1} & 0 \\ 0 & \Delta_A^{-1} \end{pmatrix} \begin{pmatrix} I & 0 \\ CA^{-1} & I \end{pmatrix} = M^{-1} = \begin{pmatrix} I & 0 \\ -D^{-1}C & I \end{pmatrix} \begin{pmatrix} \Delta_D^{-1} & 0 \\ 0 & D^{-1} \end{pmatrix} \begin{pmatrix} I & -BD^{-1} \\ 0 & I \end{pmatrix}$$

$$\begin{pmatrix} A^{-1} + A^{-1}B\Delta_A^{-1}CA^{-1} & -A^1B\Delta_A^{-1} \\ -\Delta_A^{-1}CA^{-1} & \Delta_A^{-1} \end{pmatrix} = M^{-1} = \begin{pmatrix} \Delta_D^{-1} & -\Delta_D^{-1}BD^{-1} \\ -D^{-1}C\Delta_D^{-1} & D^{-1} + D^{-1}C\Delta_D^{-1}BD^{-1} \end{pmatrix}$$

Доказательство
Из невырожденности матрицы в силу пункта 4 следует, что дополнения Шура обратимы. Поэтмоу, нужные формулы сразу следуют из утверждения 3. ■

Теперь выведем важную лемму, которая нам вскоре понадобится.

Лемма (**Тождество Вудбери**) Пусть $A \in M_n(\mathbb{C}), C \in M_m(\mathbb{C}), B \in M_{n\times m}(\mathbb{C}), D \in M_{m\times n}(\mathbb{C})$ квадратные матрицы, причем A и C обратимы. Тогда:
1. Обратимость матрицы $A + BCD$ равносильна обратимости матрицы $C^{-1} + DA^{-1}B$;
2. Если матрицы из пункта 1 обратимы, то верна формула:

$$(A + BCD)^{-1} = A^{-1} - A^{-1}B(C^{-1} + DA^{-1}B)^{-1}DA^{-1}$$

Доказательство
Рассмотрим матрицу $\begin{pmatrix} A & B \\ -D & C^{-1} \end{pmatrix}$. В этой матрице имеем $\Delta_A = C^{-1} + DA^{-1}B$, $\Delta_{C^{-1}} = A + BCD$. Из пункта 4 теоремы о блочном методе Гаусса следует равносильность обратимостей. Теперь докажем формулу из пункта 2. Из формулы обращения в терминах дополнений Шура, приравнивая матрицы в верхнем левом углу получим:

$$\Delta_{C^{-1}}^{-1} = A^{-1} - A^{-1}B\Delta_A^{-1}DA^{-1}$$

Это и есть искомая формула. ■

15 Информационная форма фильтра Калмана

Для упрощения будем рассматривать стандартную модель Калмана с $R_i > 0$ и $S_i = 0$ для всех индексов i. Отметим, что в силу $R_{e_i} = H_iP_iH_i^* + R_i$ из наших предположений следует обратимость R_{e_i}. Идея информационной формы состоит в том, чтобы вычислять соотношения в обратных матрицах. Это имеет смысл, если они обладают лучшей вычислительной стабильностью, например, если исходные матрицы содержат большие числа. Обсудим коррекцию прогноза из формы Шмидта.

Утверждение Если P_i обратима, то $P_{i|i}$, являющаяся ковариационной матрицей $\xi_i - \hat{\xi}_{i|i}$, тоже является обратимой. При этом верны следующие соотношения:
1. $P_{i|i}^{-1} = P_i^{-1} + H_i^*R_i^{-1}H_i$;
2. $K_{fi} = P_{i|i}H_i^*R_i^{-1} = (P_i^{-1} + H_i^*R_i^{-1}H_i)^{-1}H_i^*R_i^{-1}$

Доказательство
Начнем с пункта 1. Воспользуемся тождеством Вудбери:

$$(P_i^{-1} + H_i^*R_i^{-1}H_i)^{-1} = P_i - P_iH_i^*(R_i + H_iP_iH_i^*)^{-1}H_iP_i = P_{i|i}$$

Мы воспользовались тем, что ранее вывели формулу $P_{i|i} = P_i - P_iH_i^*R_{e_i}^{-1}H_iP_i$, а также обратимостью R_{e_i}.
Для локазательства второй части снова применим лемму тождество Вудбери к K_{fi}. Получим:

$$K_{fi} = P_iH_i^*(H_iP_iH_i^* + R_i)^{-1} = P_iH_i^*(R_i^{-1} - R_i^{-1}H_i(P_i^{-1} + H_i^*R_i^{-1}H_i)^{-1}H_i^*R_i^{-1} = P_{i|i}H_i^*R_i^{-1}$$

■

Таким образом, вопрос о коррекции прогноза полностью решен, в силу того, что $\hat{\xi}_{i|i} = \hat{\xi}_{i|i-1} + K_{fi}e_i$, а формула для K_{fi} получена только что. Но нам было нужно условие обратимости P_i. Приведем одно достаточное условие.

Утверждение Если все F_i обратимы и Π_0 положительно определена, то все P_i тоже обратимы.

Доказательство

Докажем по индукции. То, что P_0 обратима нам дано. Предположим, что P_i обратима, покажем, что P_{i+1} тоже. Запишем уравнение Рикатти в наших предположениях (напомним, что мы работаем в предположении $S_i = 0, R_i > 0$). Получим: $P_{i+1} = F_iP_iF_i^* + G_iQ_iG_i^* - F_iP_iH_iR_{e_i}^{-1}H_iP_iF_i^*$. Теперь применим к нему тождество Вудбери:

$$P_{i+1} = G_iQ_iG_i^* + F_i(P_i + P_iH_i(H_iP_iH_i^* + R_i)^{-1}H_iP_i)F_i^* = G_iQ_iG_i^* + F_i(P_i^{-1} + H_i^*R_i^{-1}H_i)^{-1}F_i^*$$

Заметим, что матрица $G_iQ_iG_i^*$ неотрицательно определена, поскольку сама Q_i неотрицательно определена. Матрица P_i была положительно определена, поскольку она обратима. Покажем, что тогда и P_i^{-1} тоже положительно определена. В самом деле, для произвольного вектора $x \neq 0$ в силу биективности P_i найдется $z \neq 0$, такое что $x = Az$. Тогда получим:

$$x^*P_i^{-1}x = (P_iz)^*P_i^{-1}(P_iz) = z^*P_i^*P_i^{-1}P_iz = z^*P_iz > 0$$

По этой же причине матрица R_i^{-1} будет положительно определена, а тогда $H_i^*R_i^{-1}H_i$ неотрицательно определена. Осталось заметить, что из обратимости F_i сразу получается, что из положительной определенности A следует положительная определенность $F_iAF_i^*$. ∎

Теперь обсудим этап прогноза по полной оценке. В наших предположениях имеет место формула $\hat{\xi}_{i+1|i} = F_i\hat{\xi}_{i|i}$. Поэтому единмтвенная сложность – получение формулы для обратной матрицы погрешности.

Утверждение Если все F_i, Q_i обратимы и Π_0 положительно определена, то верна следующая формула

$$P_{i+1}^{-1} = (F_i^{-1})^*P_{i|i}^{-1}F_i^{-1} - (F_i^{-1})^*P_{i|i}^{-1}F_i^{-1}G_i\left(Q_i^{-1} + G_i^*(F_i^{-1})^*P_{i|i}^{-1}F_i^{-1}G_i\right)^{-1}G_i^*(F_i^{-1})^*P_{i|i}^{-1}F_i^{-1}$$

Доказательство

Это утверждение получается сразу применением тождества Вудбери к равенству $P_{i+1}^{-1} = (F_iP_{i|i}F_i^* + G_iQ_iG_i^*)^{-1}$, где $A = F_iP_{i|i}F_i^*, B = G_i, C = Q_i, D = G_i^*$ ∎

Отметим, что у нас уже есть формула для $P_{i|i}^{-1}$. Подставив ее в только что доказанное утверждение мы сможем получить рекуррентное уравнение для P_i^{-1}. Именно это мы и проделаем. Оно также называется уравнением Рикатти.

Теорема (**Обратное уравнение Рикатти**) Если все F_i, Q_i обратимы и Π_0 положительно определена, то выполняются следующие соотношения:

$$P_{i+1}^{-1} = (F_i^{-1})^*P_{i|i}^{-1}F_i^{-1} - (F_i^{-1})^*P_{i|i}^{-1}F_i^{-1}G_i\left(Q_i^{-1} + G_i^*(F_i^{-1})^*P_{i|i}^{-1}F_i^{-1}G_i\right)^{-1}G_i^*(F_i^{-1})^*P_{i|i}^{-1}F_i^{-1}$$

$$P_{i|i}^{-1} = P_i^{-1} + H_i^*R_i^{-1}H_i$$

16 Связь с задачей оптимизации

Рассмотрим задачу линейного оценивания. Будем оценивать вектор ξ по вектору $\eta = H\xi + v$. Предположим, что все величины вещественные, центрированные и их компоненты лежат в $L^2(\mathbb{P})$. Обозначим через P_0 – ковариационную матрицу ξ, через R – ковариационную матрицу v. Предположим, что эти ковариационные матрицы обратимы. Заметим, что эту задачу можно переписать в терминах модели Калмана:

$$\begin{cases} \xi_{i+1} = \xi_i \\ \eta_i = H_i\xi_i + v_i \end{cases}$$

Здесь $i = 0, v_0 = v, H_0 = H, \xi_0 = \xi, \eta_0 = \eta$.
Таким образом, мы свели задачу к нахождению $\hat{\xi}_{0|0}$. Мы уже знаем, что формула имеет вид

$$\hat{\xi}_{0|0} = (P_0^{-1} + H^TR^{-1}H)^{-1}H^TR^{-1}\eta$$

Рассмотрим теперь оптимизационную задачу:

$$f(x) = x^TP_0^{-1}x + (\eta - Hx)^TR^{-1}(\eta - Hx) \to \min_x$$

Эта функция выпукла, поэтому ее минимум ищется из равенства нулю производной. Отсюда находится, что

$$\hat{x} = (P_0^{-1} + H^T R^{-1} H)^{-1} H^T R^{-1} \eta$$

Итого оказывается, что соответсвующая детерминированная задача эквивалентна задачи оценивания.
Аналогичная ситуация происходит и в более общем случае.
Рассмотрим стандартную модель Калмана. При этом можно считать, что $\mathbb{E}\xi_0$ произвольное, а также F_i, Q_i, Π_0 обратимы.

$$\begin{cases} \xi_{i+1} = F_i \xi_i + G_i u_i \\ \eta_i = H_i \xi_i + v_i \end{cases}$$

Рассмотрим оптимизационную задачу:

$$\begin{cases} f(x_0, \dots, x_T, u_0, \dots, u_{T-1}) = (x_0 - \mathbb{E}\xi_0)^T \Pi_0^{-1} (x_0 - \mathbb{E}\xi_0) + \sum_{i=0}^{T} (\eta_i - H_i x_i)^T R_i^{-1} (\eta_i - H_i x_i) + \sum_{i=0}^{T-1} u_i^T Q_i^{-1} u_i \to \min \\ x_{i+1} = F_i x_i + G_i u_i, \ i = 0, 1, \dots, T-1 \end{cases}$$

Оказывается, что решение $\hat{x}_i$ этой задачи совпадает с $\hat{\xi}_{i|T}$.

17 LDU разложение

Определение Квадратную матрицу A назовем сильно регулярной, если все ее главные миноры не равны нулю. С помощью дополнений Шура можно доказать следующую теорему:

Теорема Пусть A – сильно регулярная матрица. Тогда ее можно представить единственным образом в виде $R = LDU$, где L – нижнетреугольная, U – верхнетреугольная матрица с единичными элементами на диагонали, D – диагональная обратимая матрица.

Отметим, что если A – сильно регулярная эрмитова матрица, то тогда в LDU разложении $U = L^*$. Это сразу же следует из единственности. Напомним критерий Сильвестра.

Теорема (**Критерий Сильвестра**) Эрмитова матрица A положительно определена тогда и только тогда, когда все ее главные миноры положительны.

Из всего сказанного следует, что положительно определенная матрица сильно регулярна, а значит ее можно представить в виде LDL^*. Оказывается, что и для вырожденных матриц можно придумать удобную факторизацию.

Теорема (**Факторизация неотрицательно определенной матрицы**)
Неотрицательно определенная матрица $A \in M_n(\mathbb{C})$ допускает единственное разложение в виде $A = MM^*$, где M – нижнетреугольная матрица, имеющая на главной диагонали ровно rkA ненулевых положительных элементов и $n - rkA$ полностью нулевых столбцов.

Стоит отметить, что доказательство этой теоремы – представляет собой конкретный алгоритм приведения матрицы к нужному виду. Мы не будем доказывать эту теорему в общем виде, рассмотрим ее на конкретном примере.

Пример Рассмотрим матрицу $P = \begin{pmatrix} 1 & 2 & 0 \\ 2 & 4 & 0 \\ 0 & 0 & 2 \end{pmatrix}$ Это, очевидно, вырожденная матрица. Покажем, что эта матрица неотрицаительно определена. Подействуем соответсвующей квадратичной формой на эту матрицу:

$$x^* P x = |x_1|^2 + 2\overline{x_1} x_2 + 2 x_1 \overline{x_2} + 4|x_2|^2 + 2|x_3|^2 = |x_1 + 2x_2|^2 + 2|x_3|^2 \geqslant 0$$

Приступим к факторизации матрицы. Обозначим $A = \begin{pmatrix} 1 \end{pmatrix}$, $B = \begin{pmatrix} 2 & 0 \end{pmatrix}$, $C = \begin{pmatrix} 2 \\ 0 \end{pmatrix}$, $D = \begin{pmatrix} 4 & 0 \\ 0 & 2 \end{pmatrix}$. Теперь наша матрица записывается в блочном виде $P = \begin{pmatrix} A & B \\ C & D \end{pmatrix}$. Воспользуемся теперь блочным методом Гаусса:

$$P = \begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix} \begin{pmatrix} 1 & 2 & 0 \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ 0 & \Delta_A \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix} \begin{pmatrix} 1 & 2 & 0 \end{pmatrix} + \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 2 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 2 & 0 & 0 \\ 0 & 0 & \sqrt{2} \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 2 & 0 & 0 \\ 0 & 0 & \sqrt{2} \end{pmatrix}^T$$

Аккуратное доказательство состоит в подсчете ненулевых элементов на диагонале, в проверке, что на каждом шаге дополнение Шура остается неотрицательно опредленной матрицей.

18 Формула Поттера

Рассмотрим модификацию, позволяющую улучшить численную реализацию алгоритма на этапе коррекции прогноза. Предположим, что все η_k одномерные и $R_k > 0$. Разложим матрицу $P_i = P_i^{1/2} P_i^{*/2}$ по теореме о факторизации, обозначая $P_i^{\frac{*}{2}} := (P_i^{\frac{1}{2}})^*$. Тогда по формуле для матрицы коррекции прогноза имеем:

$$P_{i|i} = P_i - P_i H_i^* R_{e_i}^{-1} H_i P_i = P_i^{1/2}(I - P_i^{*/2} H_i^* R_{e_i}^{-1} H_i P_i^{1/2}) P_i^{*/2} = P_i^{1/2}(I - R_{e_i}^{-1} P_i^{*/2} H_i^* (P_i^{*/2} H_i^*)^*) P_i^{*/2}$$

Здесь мы воспользовались тем, что $R_{e_i}^{-1}$ является числом, а поэтому верна коммутативность. Обозначим $A = P_i^{*/2} H_i^*$. Идея Поттера заключалась в том, чтобы представить в удобном виде матрицу $I - R_{e_i}^{-1} AA^*$. Подберем параметр γ, такой чтобы:

$$I - R_{e_i}^{-1} AA^* = (I - \gamma AA^*)^2 = I - 2\gamma AA^* + \gamma^2 A(A^*A)A^* = I - 2\gamma AA^* + \gamma^2 (A^*A)AA^* = I + (-2\gamma + \gamma^2 A^*A)AA^*$$

Поэтому параметр γ можно найти решая квадратное уравнение: $A^*A\gamma^2 - 2\gamma = -R_{e_i}^{-1}$. Из того, что $R_{e_i} = R_i + A^*A$ вытекает, что дискриминант у уравнения положительный и решение имеет вид $\gamma = \frac{1}{\sqrt{R_{e_i}}(\sqrt{R_{e_i}} \pm \sqrt{R_i})}$. Для численной устойчивости выбирают решение с плюсом. Отсюда получается итоговая формула Поттера:

$$P_{i|i}^{1/2} = P_i^{1/2} \left(I - \frac{AA^*}{\sqrt{R_{e_i}}(\sqrt{R_{e_i}} + \sqrt{R_i})} \right)$$

К сожалению, в размерности больше чем 1 эффективных аналогов этой формулы нет. Тем не менее, эта идея Поттера в дальнейшем и развилась в метод корня, широко применяемый в настоящее время.

19 Преобразования Хаусхолдера

Утверждение (**Вещественное преобразование Хаусхолдера**)
1. Пусть $z \in \mathbb{R}^k$ – вектор строка. Положим $g = z \pm \|z\| e_1$ со стандартной евклидовой нормой, где e_1 записан в строчку. Определим матрицу $\Theta = I - 2\frac{g^T g}{g g^T}$. Тогда эта матрица ортогональна и симметрична, а также $z\Theta = \mp \|z\| e_1$.
2. Пусть $z \in \mathbb{R}^k$ – вектор столбец. Положим $g = z \pm \|z\| e_1$ со стандартной евклидовой нормой, где e_1 записан в столбец. Определим матрицу $\Theta = I - 2\frac{g g^T}{g^T g}$. Тогда эта матрица ортогональна и симметрична, а также $\Theta z = \mp \|z\| e_1$.

Доказательство
В силу аналогичности этих утверждений докажем только первую часть. Симметричность матрицы проверяется легко:

$$\Theta^T = \left(I - 2\frac{g^T g}{g g^T} \right)^T = I - 2\frac{g^T g}{g g^T} = \Theta$$

Теперь проверим ортогональность:

$$\Theta \Theta^T = \Theta^2 = \Theta^T \Theta = \left(I - 2\frac{g^T g}{g g^T} \right)^2 = I - 4\frac{g^T g}{g g^T} + 4\left(\frac{g^T g}{g g^T} \right)^2 = I - 4\frac{g^T g}{g g^T} + 4\frac{g^T (g g^T) g}{(g g^T)^2} = I$$

Осталось проверить основное тождество. Сначала вычислим вспомогательные конструкции:

$$g g^T = (z \pm \|z\| e_1)(z \pm \|z\| e_1)^T = 2\|z\|^2 \pm 2\|z\| z_1, \ \ z g g^T = 2\|z\|^2 z \pm 2\|z\| z_1 z$$

$$z g^T g = z(z \pm \|z\| e_1)^T (z \pm \|z\| e_1) = \|z\|^2 z + \|z\|^2 z_1 e_1 \pm \|z\|^3 e_1 \pm \|z\| z_1 z$$

Теперь уже легко считается $z\Theta$:

$$z\Theta = \frac{z g g^T - 2 z g^T g}{g g^T} = \frac{2\|z\|^2 z \pm 2\|z\| z_1 z - 2\|z\|^2 z - 2\|z\|^2 z_1 e_1 \mp 2\|z\|^3 e_1 \mp 2\|z\| z_1 z}{2\|z\|^2 \pm 2\|z\| z_1} = -\frac{\|z\|^2 z_1 e_1 \pm \|z\|^3 e_1}{\|z\|^2 \pm \|z\| z_1} = \mp \|z\| e_1$$

■

Отметим, что на практике выбор знака становится важным. В выражении $g = z \pm \|z\| e_1$ знак выбирается так, чтобы $\|g\|$ была побольше. Это нужно, потому что $\|g\|$ находится в знаменателе, а значит вычисления будут точнее. Аналогичная теорема верна и для комплексного случая.

Утверждение (**Комплексное преобразование Хаусхолдера**)
1. Пусть $z \in \mathbb{C}^k$ – вектор строка. Положим $g = z \pm \|z\|e^{\arg z_1}e_1$ со стандартной евклидовой нормой, где e_1 записан в строчку. Определим матрицу $\Theta = I - 2\frac{g^*g}{gg^*}$. Тогда эта матрица унитарна и эрмитова, а также $z\Theta = \mp\|z\|e^{\arg z_1}e_1$.
2. Пусть $z \in \mathbb{C}^k$ – вектор столбец. Положим $g = z \pm \|z\|e^{\arg z_1}e_1$ со стандартной евклидовой нормой, где e_1 записан в столбец. Определим матрицу $\Theta = I - 2\frac{gg^*}{g^*g}$. Тогда эта матрица унитарна и эрмитова, а также $\Theta z = \mp\|z\|e^{\arg z_1}e_1$.

Доказательство
Доказательство полностью повторяет вещественное с заменой транспонирования на эрмитово сопряжение. ■

Эти преобразования помогают нам приводить матрицы к более простому виду. Рассмотрим на примере.

Пример Рассмотрим матрицу $P = \begin{pmatrix} 3 & 4 & 0 \\ -4 & 4 & 1 \\ 9 & 7 & 2 \end{pmatrix}$. Примерим к строчке $z_1 = (3\ \ 4\ \ 0)$ преобразование Хаусхолдера Θ_1, так чтобы $z_1\Theta_1 = \alpha_1 e_1$. Тогда матрица $P\Theta_1 = \begin{pmatrix} \alpha_1 & 0 & 0 \\ * & a & b \\ * & c & d \end{pmatrix}$. Далее возьмем строчку $z_2 = (a\ \ b)$ и для нее будем подбирать преобразование Θ_2, такое что $\Theta_2 z_2 = (\alpha_2\ \ 0)$. Тогда матрица $P\Theta_1(1 \oplus \Theta_2) = \begin{pmatrix} \alpha_1 & 0 & 0 \\ * & \alpha_2 & 0 \\ * & * & * \end{pmatrix}$, где $1 \oplus \Theta_2 = \begin{pmatrix} 1 & 0 \\ 0 & \Theta_2 \end{pmatrix}$.
Таким образом, с помощью преобразований Хаусхолдера можно приводить матрицу к треугольному виду.

20 Преобразования Гивенса

Напомним, что матрица поворота в $\mathbb{R}^2$ имеет вид $\Theta_\rho = \frac{1}{\sqrt{1+\rho^2}} \begin{pmatrix} 1 & -\rho \\ \rho & 1 \end{pmatrix}$ для некоторого числа $\rho \in \mathbb{R}$. При этом поворот является ортогональным преобразованием. Предположим, что $a \neq 0$, тогда взяв $\rho = \frac{b}{a}$ получим:

$$(a\ \ b)\Theta_\rho = \frac{|a|}{\sqrt{a^2+b^2}}(a\ \ b)\begin{pmatrix} 1 & -\frac{b}{a} \\ \frac{b}{a} & 1 \end{pmatrix} = \frac{|a|}{\sqrt{a^2+b^2}}\begin{pmatrix} a + \frac{b^2}{a} & 0 \end{pmatrix} = \begin{pmatrix} sgn(a)\sqrt{a^2+b^2} & 0 \end{pmatrix}$$

То есть с помощью правильно подобранного поворота мы можем привести вектор к более простому виду. Такие повороты и называются поворотами Гивенса. С их помощью, по аналогии с преобразованием Хаусхолдера можно приводить матрицы к треугольному виду. Рассмотрим пример.

Пример Рассмотрим матрицу $P = \begin{pmatrix} 3 & 4 & 0 \\ -4 & 4 & 1 \\ 9 & 7 & 2 \end{pmatrix}$. Примерим к строчке $z_1 = (3\ \ 4)$ поворот Гивенса Θ_{ρ_1}, так чтобы $z_1\Theta_{\rho_1} = (\alpha_1\ \ 0)$. Тогда матрица $P\begin{pmatrix} \Theta_{\rho_1} & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} \alpha_1 & 0 & 0 \\ * & a & b \\ * & * & * \end{pmatrix}$. Далее возьмем строчку $z_2 = (a\ \ b)$ и для нее будем подбирать преобразование Θ_{ρ_2}, такое что $\Theta_{\rho_2} z_2 = (\alpha_2\ \ 0)$. Тогда матрица $P\begin{pmatrix} \Theta_{\rho_1} & 0 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & \Theta_{\rho_2} \end{pmatrix} = \begin{pmatrix} \alpha_1 & 0 & 0 \\ * & \alpha_2 & 0 \\ * & * & * \end{pmatrix}$
Таким образом, с помощью вращений Гивенса можно приводить матрицу к треугольному виду.

В комплексном случае все делается полностью аналогично, но матрица поворота определяется иначе, а именно
$\Theta_\rho = \frac{1}{\sqrt{1+|\rho|^2}}\begin{pmatrix} 1 & -\rho \\ \rho^* & 1 \end{pmatrix}$. Для случайной матрицы метод Хаусхолдера работает быстрее, но если у матрицы изначально было много нулей, то метод Гивенса зачастую оказывается предпочтительнее.

21 Теорема об унитарной матрице

Основная идея метода корня, который будет скоро представлен кроется в теореме об унитарной матрице. Сформулируем ее и докажем.

Теорема (**Об унитарной матрице**)
Пусть имеются две матрицы $A, B \in M_{s\times k}(\mathbb{C})$, где $s \leqslant k$. Тогда выполнение равенства $AA^* = BB^*$ равносильно существованию унитарной мвтрицы Θ, такой что $A\Theta = B$.

Доказательство

В одну сторону это утверждение очевидно. В самом деле, если $A\Theta = B$, то $AA^* = A(\Theta\Theta^*)A^* = BB^*$. Докажем в обратную сторону. При помощи преобразований Гивесна и Хаусхолдера мы можем сделать матрицу A нижнетреугольной, то есть найти такую матрицу Θ_A, что $A\Theta_A = \begin{pmatrix} l_{11} & 0 & \dots & \dots & \dots & \dots & 0 \\ l_{21} & l_{22} & 0 & \dots & \dots & \dots & 0 \\ \dots & \dots & \dots & 0 & \dots & \dots & 0 \\ l_{s1} & \dots & \dots & l_{ss} & 0 & \dots & 0 \end{pmatrix} = \begin{pmatrix} L_A & 0 \end{pmatrix}$, где L_A – нижнетреугольная матрица. При этом, дополнительно применяя повороты, можно выбрать матрицу Θ_A так, чтобы $l_{ii} \geqslant 0$ и при этом из $l_{ii} = 0$ следовало бы $l_{ji} = 0$ при $j > i$. Аналогичную процедуру проделаем для другой матрицы: $B\Theta_B = \begin{pmatrix} L_B & 0 \end{pmatrix}$. Получим: $L_A L_A^* = A\Theta_A\Theta_A A^* = AA^* = BB^* = B\Theta_B\Theta_B^* B^* = L_B L_B^*$. Из теоремы о факторизации неотрицательно определенной матрицы получим, что $L_A = L_B \implies A\Theta_A = B\Theta_B \implies A\Theta_A\Theta_B^* = B$. ■

Вернемся к нашей модели: $\begin{cases} \xi_{i+1} = F_i\xi_i + G_i u_i \\ \eta_i = H_i\xi_i + v_i \end{cases}$. Будем считать, что $R_i > 0$ и $S_i \neq 0$.

Тогда рассмотрим уравнение для матрицы ошибки прогноза по полной оценке: $P_{i+1} = F_i P_{i|i} F_i^* + G_i Q_i G_i^*$. Матрицы $P_{i+1}, P_{i|i}, Q_i$ неотрицательно определены, поэтому их можно разложить. Для неотрицательно определенной матрицы A обозначим через $A^{1/2}$ нижнетругольную матрицу из теоремы о факторизации. Тогда получаем:

$$(P_{i+1}^{1/2} \;\; 0)(P_{i+1}^{1/2} \;\; 0)^* = P_{i+1}^{1/2}P_{i+1}^{*/2} = P_{i+1} = F_i P_{i|i} F_i^* + G_i Q_i G_i^* = (F_i P_{i|i}^{1/2} \;\; G_i Q_i^{1/2})(F_i P_{i|i}^{1/2} \;\; G_i Q_i^{1/2})^*$$

По теореме об унитарной матрице найдется унитарная матрица Θ, такая что $(F_i P_{i|i}^{1/2} \;\; G_i Q_i^{1/2})\Theta = (P_{i+1}^{1/2} \;\; 0)$.

Таким образом, делать прогноз по полной оценке очень легко. Мы запишем матрицу $(F_i P_{i|i}^{1/2} \;\; G_i Q_i^{1/2})$, а затем с помощью вращений Гивенса и преобразований Хаусхолдера приведем ее к нижнетругольному виду $(X \;\; 0)$. Если при этом добиваться того, чтобы у матрицы X на диагонали были положительные элементы, а остальные столбцы были полностью нулевыми, то по теореме о факторизации X совпадет с $P_{i+1}^{1/2}$. Это и есть метод корня.

Обсудим теперь этап коррекции. В наших предположениях $P_{i|i} = P_i - P_i H_i^* R_{e_i}^{-1} H_i P_i$. Заметим, что тогда выполнено следующее тождество:

$$\begin{pmatrix} R_i^{1/2} & H_i P_i^{1/2} \\ 0 & P_i^{1/2} \end{pmatrix} \begin{pmatrix} R_i^{1/2} & H_i P_i^{1/2} \\ 0 & P_i^{1/2} \end{pmatrix}^* = \begin{pmatrix} R_i + H_i P_i H_i^* & H_i P_i \\ P_i H_i^* & P_i \end{pmatrix} =$$

$$= \begin{pmatrix} R_{e_i} & H_i P_i \\ P_i H_i & P_{i|i} + P_i H_i^* R_{e_i}^{-1} H_i P_i \end{pmatrix} = \begin{pmatrix} R_{e_i}^{1/2} & 0 \\ P_i H_i R_{e_i}^{-*/2} & P_{i|i}^{1/2} \end{pmatrix} \begin{pmatrix} R_{e_i}^{1/2} & 0 \\ P_i H_i R_{e_i}^{-*/2} & P_{i|i}^{1/2} \end{pmatrix}^*$$

Значит, по теореме об унитарной матрице найдется унитарная Θ, такая что $\begin{pmatrix} R_{e_i}^{1/2} & 0 \\ P_i H_i R_{e_i}^{-*/2} & P_{i|i}^{1/2} \end{pmatrix} \Theta = \begin{pmatrix} R_i^{1/2} & H_i P_i^{1/2} \\ 0 & P_i^{1/2} \end{pmatrix}$.

Таким образом, мы можем действовать по следующему алгоритму. Рассмотрим матрицу $\begin{pmatrix} R_{e_i}^{1/2} & 0 \\ P_i H_i R_{e_i}^{-*/2} & P_{i|i}^{1/2} \end{pmatrix}$ и будем приводить ее к треугольному виду $\begin{pmatrix} X & 0 \\ Y & Z \end{pmatrix}$ таким образом, чтобы на диагонали были только положительный элементы, а остальные столбцы были полностью нулевыми. Тогда Z и будет искомой матрицей.

Теперь обсудим классический прогноз без формы Шмидта. Для этого нам каждый раз нужно решать уравнение Рикатти, которое в наших предположениях имеет вид $P_{i+1} = F_i P_i F_i^* + Q_i Q_i G_i^* - F_i P_i H_i^* R_{e_i}^{-1} H_i P_i F_i^*$. Обозначая $K_i = F_i P_i H_i^*$ это можно записать в виде

$$\begin{pmatrix} R_i^{1/2} & H_i P_i^{1/2} & 0 \\ 0 & F_i P_i^{1/2} & G_i Q_i^{1/2} \end{pmatrix} \begin{pmatrix} R_i^{1/2} & H_i P_i^{1/2} & 0 \\ 0 & F_i P_i^{1/2} & G_i Q_i^{1/2} \end{pmatrix}^* = \begin{pmatrix} R_{e_i}^{1/2} & 0 & 0 \\ K_i R_{e_i}^{-*/2} & P_{i+1}^{1/2} & 0 \end{pmatrix} \begin{pmatrix} R_{e_i}^{1/2} & 0 & 0 \\ K_i R_{e_i}^{-*/2} & P_{i+1}^{1/2} & 0 \end{pmatrix}^*$$

Отсюда следует, что существует унитарная матрица Θ, такая что $\begin{pmatrix} R_i^{1/2} & H_i P_i^{1/2} & 0 \\ 0 & F_i P_i^{1/2} & G_i Q_i^{1/2} \end{pmatrix} \Theta = \begin{pmatrix} R_{e_i}^{1/2} & 0 & 0 \\ K_i R_{e_i}^{-*/2} & P_{i+1}^{1/2} & 0 \end{pmatrix}$

Таким образом, задача снова сводится к триангулизации матрицы $\begin{pmatrix} R_i^{1/2} & H_i P_i^{1/2} & 0 \\ 0 & F_i P_i^{1/2} & G_i Q_i^{1/2} \end{pmatrix}$.

Оценка же делается по формуле $\hat{\xi}_{i+1} = F_i\hat{\xi}_i + K_{pi}(\eta_i - H_i\hat{\xi}_i) = F_i\hat{\xi}_i + K_i R_{e_i}^{-*/2} R_{e_i}^{-1/2}$