appropos

Михеев С.Е.

October 2020

Оглавление

Глава 1.
Линейная интерполяция и аппроксимация

§1.1. ПОЛУЛИНЕЙНАЯ АППРОКСИМАЦИЯ

Задача полулинейной аппроксимации (ПА). Пусть в n-мерном векторном пространстве над числовым полем P значения скалярной функции $g : P^n \to P$ известны в узлах $x^1, ..., x^s$. Имеется ω — k–мерная функция-столбец $\omega : P^n \to P^k$. Требуется найти строку $\bar{g}$, $\bar{g}^{\mathrm{T}} \in P^k$, такую, что аппроксиматор (аппроксимирующий агрегат) $\bar{g}\omega(x)$, в некотором указанном смысле наилучшим образом приближает функцию g. Поле P может быть как вещественным, так и комплексным. Величина k называется размерностью аппроксиматора. ■

Отметим, что "полу" в названии этой задачи обязано возможной нелинейности аппроксиматора по x, а "линейность" – линейному вхождению параметров $\bar{g}$ в аппроксиматор. Существуют и нелинейные аппроксиматоры общего вида $G(\bar{g}, x)$. Они значительно сложнее в работе, но обладают другими хорошими качествами, и применяются в основном в одномерных пространствах.

Очень близка к задаче ПА, а после полной формализации и не отличима от нее, задача, рождающаяся существенно иначе. Функция $g(x)$ может быть известна в виде некоторого *закона* с неизвестными параметрами: $g(x) = G(\bar{g}, x)$, которые требуется найти в результате эксперимента. Закон этот, однако, может лишь приближенно соответствовать исследуемой функции, да и сами измерения ее значений $\{y_i\}$ в узлах $\{x_i\}$ могут содержать погрешности. Откуда появляется проблема поиска наилучшего в том или ином смысле набора параметров $\bar{g}$.

В этом разделе мы отойдем от вопросов погрешностей измерений функции g и качества ее аппроксимации в точках, отличных от узлов. То есть все что остается от функции g — это ее значения $y_1, ..., y_s$ в узлах $x_1, ..., x_s$.

Здесь будут рассмотрены только линейные по $\bar{g}$ аппроксиматоры G, т.е. те, что предложены в задаче ПА: $G(\bar{g}, x) = \bar{g}\omega(x)$. Для них одним из самых удобных для использования критериев качества является квадратичное отклонение $Q(\bar{g}) := \sum \left| y_i - \bar{g}\omega(x^i) \right|^2$: чем он меньше, тем лучше.

З а д а ч а полулинейной аппроксимации по методу наименьших квадратов (ПАНК). Пусть набор чисел $y_1, ..., y_s$ из числового поля P сопоставлен узлам $x^1, ..., x^s$ из n-мерного векторного пространства над полем P. Имеется ω — k–мерная функция-столбец $\omega : P^n \to P^k$. Требуется найти строку $\bar{g}$, $\bar{g}^{\text{т}} \in P^k$, которая является минимайзером квадратичной программы

$$Q(\bar{g}) := \sum \left| y_i - \bar{g}\omega(x^i) \right|^2 \longrightarrow \min_{g_1, \cdots, g_k} . \tag{1.1}$$

Поле P может быть как вещественным, так и комплексным. ■

Отметим, что как в ПА, так и в ПАНК не возбраняется совпадение некоторых узлов. это интерпретируется как многократное измерение с погрешностью значения функции при одном значении переменной.

Будем далее обозначать через * операцию сопряжения матриц, т.е. суперпозицию операций покомпонентного комплексного сопряжения с последующим транспонированием; в вещественном случае * будет означать простое транспонирование, а в скалярном – операцию комплексного сопряжения.

Теорема 1.1. *Множество конечных решений задачи ПАНК не пусто, т.е. у Q из* (1.1) *существует конечный минимайзер.*

Д о к а з а т е л ь с т в о. Преобразуем целевую функцию из (1.1)

$$Q(\bar{g}) := \sum \left(y_i - \bar{g}\omega(x^i)\right)\left(y_i - \bar{g}\omega(x^i)\right)^* =$$

$$= \sum |y_i|^2 - \sum y_i \omega^*(x^i)\bar{g}^* - \bar{g}\sum \omega(x^i)y_i^* + \bar{g}\left(\sum \omega(x^i)\omega^*(x^i)\right)\bar{g}^*. \tag{1.2}$$

Приведем эрмитову форму $E = \bar{g}\left(\sum \omega(x^i)\omega^*(x^i)\right)\bar{g}^*$ в разложении (1.2) неособым преобразованием $\bar{g} = hS$ к нормальному виду. В силу неотрицательности целевой функции и закона инерции эрмитовых форм имеем

$$Q(\bar{g}) = Q_h(h) := h_1 h_1^* + ... + h_r h_r^* + ha + a^* h^* + c_0, \tag{1.3}$$

$$r = \text{rang}\,(E), \qquad c_0 = \text{const} \in R^1, \qquad a = \text{const} \in P^n.$$

Пусть e^k — единичная строка с единицей на k-м месте; $\Re$ — функция, выдающая действительную часть своего комплексного аргумента; $\Im$ — функция, выдающая мнимую часть своего комплексного аргумента.

Полагая $h := te^k \Re a_k$, $k = \overline{r+1, n}$ и устремляя t к $-\infty$, из неотрицательности $Q_h(h) \equiv 2t(\Re a_k)^2 + c_0$ для всех t имеем

$$\Re a_{r+1} = ... = \Re a_n = 0.$$

Полагая $h := tie^k \Im a_k$, $k = \overline{r+1, n}$ и устремляя t к $+\infty$, из неотрицательности $Q_h(h) \equiv -2t\Im {a_k}^2 + c_0$ для всех t имеем $\Im a_{r+1} = ... = \Im a_n = 0$.

Выделяя полные квадраты в Q_h, получаем

$$Q_h(h) = (h_1 + a_1)(h_1 + a_1)^* + ... + (h_r + a_r)(h_r + a_1)^* + c_1.$$

Существование конечных минимума и его минимайзера $h^{\min}$ у Q_h очевидно. (Например такой: $h^{\min} = (-a_1, ..., -a_r, 0, ..., 0)$.) Ясно также, что $\bar{g}^{\min} := h^{\min} S$ будет конечным минимайзером у Q. ■

Теорема 1.2. *Множество минимайзеров квадратичной программы* (1.1) *совпадает с множеством решений системы линейных уравнений относительно* $\bar{g}$:

$$\bar{g}\Omega = \sum_1^s y_k \omega^*(x^k), \qquad \Omega := \sum_1^s \omega(x^k)\, \omega^*(x^k). \tag{1.4}$$

Доказательство. 1) <u>Решение системы есть минимайзер.</u> Возьмем некоторую строку e, число t и исследуем значение целевой функции $Q(\bar{g} + te)$, используя представление (1.2).

$$Q(\bar{g} + te) = Q(\bar{g}) - \bar{t} \sum \left(y_k - \bar{g}\omega(x^k)\right) \omega^*(x^k) e^* -$$

$$-te \sum \omega(x^k) \left(y_k^* - \omega^*(x^k)\bar{g}^*\right) + |t|^2 e\Omega e^* =$$

$$= Q(\bar{g}) + |t|^2 e\Omega e^* - \bar{t} \left(\sum y_k \omega^*(x^k) - \bar{g}\Omega\right) e^* - te \left(\sum \omega(x^k) y_k^* - \Omega \bar{g}^*\right). \tag{1.5}$$

Если строка $\bar{g}$ есть решение системы (1.4), то, в силу эрмитовости Ω, эта строка также решение системы

$$0 = \sum \omega(x^k) y_k^* - \Omega \bar{g}^* =: \ W^*.$$

Поэтому, если $\bar{g}$ является решением системы (1.4), то, в силу очевидной положительной полуопределенности матрицы Ω, верно $Q(\bar{g} + te) \geq Q(\bar{g})$ $\forall e$, $\forall t$. Следовательно, $\bar{g}$ — минимайзер для Q.

2) Обратно. Пусть $\bar g$ — минимайзер. Положим $e = e^k$, где e^k — единичная строка с единицей на k-м месте. Для малых по модулю вещественных t согласно разложению (1.5) знак приращения $Q(\bar g + te) - Q(\bar g)$ совпадает со знаком линейной части, если она отлична от нуля, т.е. со знаком величины $-2\Re W_k$, W_k — k-я компонента столбца W. Полагая $\operatorname{sgn} t = \operatorname{sgn} \Re W_k$, приходим к противоречию с тем, что $\bar g$ — минимайзер (при условии $\Re W_k \neq 0$). Перебрав все единичные строки, получаем $\Re W = 0$.

Пусть теперь $e = ie^k$. Тогда линейная часть разложения (1.5) есть $-2t\Im W_k$. И аналогично $\Im W = 0$. Т.е. $W = 0$. ∎

Следствие 1.1. Поскольку по теореме 1.1 *конечный минимайзер задачи ЛАНК всегда существует и в силу второй половины доказательства теоремы* 1.2 *он является решением системы* (1.4), *решение системы* (1.4) *всегда существует.*

В описанниях задач ПА и ПАНК вектор-функция ω была общего вида. Поэтому пришлось оставить без внимания количество и расположение узлов, а единственность решения задачи ЛА и ЛАНК зависит от вида набора узлов. Конкретизация ω позволяет наложить требование на узлы, обеспечивающее единственность задачи ЛАНК.

§1.2. ЛИНЕЙНАЯ АППРОКСИМАЦИЯ

Определение 1.1. *Векторы* $x^0, x^1, ..., x^n$ *находятся в общем положении, если векторы* $x^1 - x^0, ..., x^n - x^0$ *линейно независимы.*

Пусть в произвольном базисе набору векторов $x^0, x^1, ..., x^n$ сопоставлен набор координатных столбцов $\mathrm{x}^0, \mathrm{x}^1, ..., \mathrm{x}^n$. Тогда общее положение векторов набора эквивалентно невырожденности матрицы $\Xi := (\mathrm{x}^1 - \mathrm{x}^0 \dots \mathrm{x}^n - \mathrm{x}^0)$. Положим $\check x := \begin{pmatrix} 1 \\ \mathrm{x} \end{pmatrix}$. Элементарно доказывается, что невырожденность матрицы Ξ эквивалентна невырожденности матрицы $(\check x^0 \dots \check x^n)$.

Отсюда, очевидно, вытекает инвариантность определения 1.1 относительно замены "вычитаемого" вектора x_0 на любой другой из набора.

Задача линейной аппроксимации по методу наименьших квадратов (ЛАНК). [1] Пусть значения скалярной функ-

[1] Название задачи обязано линейности аппроксимирующего агрегата $\widetilde g$ по параметрам $g_0, ..., g_n$ и по переменной x.

ции $g(x)$ от n аргументов известны в узлах $x^1, ..., x^s$ и среди них есть $n+1$, находящихся в общем положении [2]. Требуется найти параметры $g_0, ..., g_n$ линейной функции 1-го рода $\widetilde{g}(x) = g_0 + g_1 x_1 + ... + g_n x_n$, приближающую g в этих узлах наилучшим образом в смысле минимизации невязки:

$$\sum \left| g(x^i) - \widetilde{g}(x^i) \right|^2 \longrightarrow \min_{g_0, \dots, g_n} . \tag{1.6}$$

При обозначениях $y_i = g(x^i)$, $i = \overline{1,s}$; $\bar{g} = (g_0\ g_1\ ...\ g_n)$; $\check{x}^i = (1\ x_1^i\ ...\ x_n^i)^{\mathrm{T}}$ квадратичная программа (1.6) принимает вид

$$R(\bar{g}) := \sum_{i=1}^{s} \left| y_i - g_0 - \sum_{j=1}^{n} g_j x_j^i \right|^2 \equiv \sum_i \left| y_i - \bar{g}\check{x}^i \right|^2 \longrightarrow \min_{\bar{g}} . \tag{1.7}$$

■

Упрощением предыдущей теоремы для задачи ЛАНК будет

Теорема 1.3. *Множество решений задачи* (1.7) *(квадратичной программы) совпадает с множеством решений системы*

$$\begin{cases} 0 = \displaystyle\sum_{i=1}^{s} y_i - s g_0 - \sum_{p=1}^{n} \left(g_p \sum_{i=1}^{s} x_p^i \right) = \sum_{i=1}^{s} y_i - \sum_{i=1}^{s} \bar{g}\check{x}^i, \\ 0 = \displaystyle\sum_{i=1}^{s} \left(y_i - \bar{g}\check{x}^i \right) \bar{x}_j^i = \sum_i y_i \bar{x}_j^i - \bar{g} \left(\sum_i \check{x}^i \bar{x}_j^i \right), \quad j = \overline{1,n}, \end{cases} \tag{1.8}$$

или в матричном виде системы

$$\bar{g}X = \sum_{i=1}^{s} y_i (\check{x}^i)^*, \qquad X := \sum_{i=1}^{s} \check{x}^i (\check{x}^i)^* . \tag{1.9}$$

■

Практически ничего нового не появляется при аппроксимации вектор-функции вместо скалярной.

Задача ЛАНК для вектор-функции. Пусть значения отображения $g : P^n \to P^m$ известны в узлах $x^1, ..., x^s$ и среди них есть $n+1$, находящихся в общем положении. Требуется найти параметры $\bar{g} = (g^0 ... g^n)$ аффинной функции 1-го рода

[2] Это требование позволит в дальнейшем доказать единственность решения задачи ЛАНК.

$\widetilde{g}(x) = g^0 + g^1 x_1 + ... + g^n x_n = \bar{g}\check{x}$, приближающую g в этих узлах наилучшим образом в смысле минимизации невязки:

$$\sum \left(g(x^i) - \widetilde{g}(x^i)\right)^* \left(g(x^i) - \widetilde{g}(x^i)\right) \longrightarrow \min_{g^0,...,g^n}, \tag{1.10}$$

т.е. по методу наименьших квадратов. Здесь $g^0, ..., g^n$ — m-мерные столбцы. ■

Сохраняя обозначение $\bar{g}$ для матрицы из этих столбцов, имеем $\widetilde{g}(x) = \bar{g}\check{x}$ и в вещественном случае тогда, вплоть до обозначений, справедлива теорема 1.3 в применении к задаче (1.10), т.е. решение задачи (1.10) есть решение матричного уравнения (1.9) относительно матрицы $\bar{g}$. Это можно увидеть не только из доказательства теоремы 1.3, но и заметив, что программа (1.10) расщепляется на m независимых программ с минимизацией по $\bar{g}_j$ — строкам матрицы $\bar{g}$:

$$\sum \left(g(x^i) - \widetilde{g}(x^i)\right)^* \left(g(x^i) - \widetilde{g}(x^i)\right) = \sum_i \sum_j \left|g_j(x^i) - \bar{g}_j \check{x}^i\right|^2 =$$

$$= \sum_j \sum_i \left|g_j(x^i) - \bar{g}_j \check{x}^i\right|^2 .$$

Помимо очевидной симметричности матрицы X и эрмитовости матрицы Ω, оказывается, они обладают еще одним хорошим свойством: положительной полуопределенностью (ППО), а при удачном выборе узлов и положительной определенностью (ПО). Последняя гарантирует единственность, уже существующего в силу теорем 1.1, 1.2 и 1.3, решения систем (1.4), (1.9). Это позволяет применять широкий круг методов поиска решения систем (1.4), (1.9), а в методе Гаусса не будут появляться нули на главной диагонали на каком либо шаге.

В выяснении ПО матрицы X используем следующий результат.

Теорема 1.4. *Общего положения векторов* $x^1, ..., x^{n+1}$ *векторного пространства* P^n *, необходимо и достаточно, чтобы для всякого набора векторов* $\{y^1, ..., y^{n+1}\}$ *из векторного пространства* P^m *существовала единственная афинная функция 1-го рода* $\widetilde{g}(x) = g^0 + g^1 x_1 + ... + g^n x_n$, *такая, что* $g^0, ..., g^n \in P^m$ *и* [3)]

$$\widetilde{g}(x^i) = y^i, \qquad i = \overline{1, n+1}. \tag{1.11}$$

[3)] Ортега в теореме 7.2.3 из [3] формулирует этот результат для $n = m$, и, когда набор $y^1, ..., y^{n+1}$ в общем положении, утверждает, что система $g^1, ..., g^n$ линейно независима. Следует отметить, что в русский перевод [3] формулировки этой теоремы вкрались серьезные ошибки. Но не в доказательство.

Действительно. В матричном виде условие (1.11) есть

$$\bar{g}\left(\check{x}^1 \dots \check{x}^{n+1}\right) = \left(y^1 \dots y^{n+1}\right), \tag{1.12}$$

где через $\bar{g}$ обозначена матрица $(g^0\ g^1 \dots g^n)$. Для существования единственного решения системы (1.12) относительно $\bar{g}$ при любом наборе $y^1, \dots, y^{n+1}$, как известно, необходимо и достаточно невырожденности матрицы $(\check{x}^1 \dots \check{x}^{n+1})$, т.е. общего (см. замечание 1) положения векторов $x^1, \dots, x^{n+1}$. ■

Если число узлов равно $n+1$ и они находятся в общем положении, то матрица X из (1.9) положительно определена. Действительно. Из теоремы 1.4 следует существование единственного (зависимого от y) набора $\bar{g} := (g^0,\ g^1, \dots, g^n)$, который дает ноль невязке (1.6). Так как эта невязка неотрицательна, то $\bar{g}$ — ее минимайзер и он единствен. Поэтому согласно теореме 1.3 множество решений системы (1.9), $s = n+1$ состоит из единственного $\bar{g}$, что эквивалентно невырожденности X. А по построению X положительно полуопределена. Поэтому и положительно определена.

А могут ли дополнительные узлы "испортить" матрицу X? На этот вопрос отвечает теорема 1.5 (см. далее). Для ее формулировки потребуется

Задача полулинейной интерполяции (ПИ). Пусть интерполятор (интерполяционный агрегат) имеет вид $\bar{g}\omega(x)$, где $\bar{g}$ — k–мерная строка, $x \in P^n$, $\omega(x)$ — k–мерная функция-столбец, элементы которой зависят от x. Требуется для заданного набора чисел y_i и узлов x^i, $i \in \mathcal{M}$, найти параметры $\bar{g}$, которые обеспечат

$$y_i = \bar{g}\omega(x^i) \qquad \forall i \in \mathcal{M}. \quad ■ \tag{1.13}$$

Отметим, что задача ПИ хорошо поставлена, когда s – мощность множества $\mathcal{M}$ – равна k: количества выходных данных и входных равны.

Если $\omega(x) = \check{x}$, то:

1) поскольку $r = n+1$, этому соответствует $s = n+1$ (о существовании и единственности решения задачи ПИ см. т. 1.4);

2) когда $s < n+1$ и решение задачи ПИ существует, то оно не единственно;

3) когда $\operatorname{rang}\{\check{x}^i\}_{i \in M} < s$ (а это всегда при $s > n+1$), то задача ЛИ не для всякого набора $\{y_i\}_{i \in M}$ имеет решение, в частности, может ни для какого.

Лемма 1.1. *Если H является ППО матрицей (эрмитовой), то квадратичная форма $K(x) := x^*Hx$ выпукла, а если H — ПО матрица, то K — строго выпукла. Оба утверждения верны и для вещественного случая.*

Доказательство. Возьмем два вектор-столбца x, y и рассмотрим отрезок в параметрическом виде, их соединяющий: $x + t(y-x),\ \ t \in [0,1]$. Для строгой выпуклости квадратичной формы необходимо и достаточно, чтобы

$$(\forall x \neq y)(\forall t \in (0,1))\ \ 0 > K(x+t(y-x)) - K(x) - t[K(y)-K(x)] =: P_2(t).$$

Раскрывая скобки, получаем

$$P_2(t) = t^2(y-x)^*H(y-x) + t[x^*H(y-x) + (y-x)^*Hx - x^*Hx - y^*Hy].$$

По построению многочлен второго порядка P_2 имеет два корня: 0 и 1. Из ПО матрицы H следует положительность коэффициента перед t^2. Что влечет $0 > P_2(t)\ \ \forall t \in (0,1)$, т.е. K — строго выпукла. Если H — ППО, то возможно, что $(y-x)^*H(y-x) = 0$, и тогда многочлен P_2 — тождественный нуль, как многочлен первой степени, имеющий два корня: 0 и 1.

Теорема 1.5. *Если среди узлов $x^1, ..., x^s \in P^n$ есть подмножество с индексами из $\mathcal{M}$, такое, что задача ПИ на этом подмножестве имеет единственное решение для какого-нибудь набора $\{y_i\}_{i\in\mathcal{M}} \subset R^1$, то:*
1) *матрица эрмитовой квадратичной формы, исчерпывающей квадратичность функции Q из* (1.1) *и* (1.2) :

$$Q(\bar{g}) = \sum_1^s \left(y_i - \bar{g}\omega(x^i)\right)\left(y_i - \bar{g}\omega(x^i)\right)^*,$$

т.е. матрица $\Omega := \displaystyle\sum_1^s \omega(x^i)\omega^*(x^i)$, *положительно определена;*
2) *функция Q строго выпукла, ее минимайзер существует, конечен и единствен.*

Доказательство. **1)** Положим, не умаляя общности, $\mathcal{M} = \{1, ..., m\}$ и разобъем Q на два слагаемых:

$$Q_1 := \sum_1^m \left(y_i - \bar{g}\omega(x^i)\right)\left(y_i^* - \omega^*(x^i)\bar{g}^*\right), \qquad Q_2 := \sum_{m+1}^s \left|y_i - \bar{g}\omega(x^i)\right|^2.$$

Покажем сначала, что матрица $\Omega_1 := \sum\limits_1^m \omega(x^i)\omega^*(x^i)$ положительно определена. Положительная полуопределенность матрицы Ω_1 очевидна:

$$(\forall \bar{g}) \quad \bar{g}\Omega_1\bar{g}^* = \sum_1^m \left|\bar{g}\omega(x^i)\right|^2 \geq 0. \tag{1.14}$$

Пусть она не положительно определена. Тогда существует строка $e \neq 0$, такая, что $e\Omega_1 e^* = 0$.

Согласно условию этой теоремы существует единственная строка $\bar{g}^0$, решающая задачу ПИ по узлам с индексами из $\mathcal{M}$. Поэтому она же, очевидно, будет единственным минимайзером для Q_1, доставляющим Q_1 нулевое значение. Пусть вещественное t отлично от нуля, тогда в силу того, что Q_1 не отрицателен и обнуляется только на $\bar{g}^0$, справедливо:

$$0 < Q_1(\bar{g}^0 + te) = Q_1(\bar{g}^0) - t\sum_1^m \left(y_i - \bar{g}^0\omega(x^i)\right)\omega^*(x^i)e^* -$$

$$-t\sum_1^m e\omega(x^i)\left(y_i^* - \omega^*(x^i)(\bar{g}^0)^*\right) + t^2\sum_1^m e\omega(x^i)\omega^*(x^i)e^* =$$

$$= t\Big(\bar{g}^0\Omega_1 - \sum_1^m \omega(x^i)y_i^*\Big) + te\Big(\Omega_1(\bar{g}^0)^* - \sum_1^m y_i\omega(x^i)^*\Big) + t^2 e\Omega_1 e^* =$$

$$\overset{\text{т.1.2}}{=} t^2 e\Omega_1 e^* = 0. \tag{1.15}$$

Источником противоречия в цепи (1.15) является допущение не ПО матрицы Ω_1.

Из положительной полуопределенности Ω_2 и положительной определенности Ω_1 следует

$$\bar{g}(\Omega_1 + \Omega_2)\bar{g}^* \ \geq \ \bar{g}\Omega_1\bar{g}^* \ > \ 0,$$

т.е. Ω, равная $\Omega_1 + \Omega_2$, положительно определена.

2) Согласно предыдущей лемме, из ПО матрицы Ω следует строгая выпуклость квадратичной формы $\bar{g}\Omega\bar{g}^*$. Выпуклая линейная часть функции Q в сумме со строго выпуклой квадратичной формой, очевидно, строго выпукла. Легко заметить, что в силу положительной определенности Ω при $\|\bar{g}\| \longrightarrow \infty$ имеет место $Q(\bar{g}) \longrightarrow +\infty$. Отсюда вытекает существование конечного минимайзера функции Q и его единственность.

Теорема 1.6. *Если в наборе узлов* $x^1, ..., x^s$ *есть подмножество узлов, стоящих в общем положении, то матрица* X *при переменных в системе* (1.9) *положительно определена.*

Доказательство. Положив $\omega(x) = (1\ x_1\ ...\ x_n)^{\mathrm{T}}$ и $\bar{g} = (g_0\ ...\ g_n)$, видим, что Ω из предыдущей теоремы совпадает с X из (1.9). Применяем две предшествующие теоремы. ∎

§1.3. ЛИНЕЙНАЯ ИНТЕРПОЛЯЦИЯ

1.3.1. Место линейной интерполяции в общей задаче интерполирования

Напомним, что такое линейная интерполяция скалярной функции $f : P \to P$, где P – числовое поле. Дан интерполяционный агрегат (интерполятор)

$$A(x, p) := pw(x), \tag{1.16}$$

где $p = (p_1, ..., p_m)$, $w(x) = \begin{pmatrix} w_1(x) \\ \vdots \\ w_n(x) \end{pmatrix}$. Функции $w_1(x), ..., w_n(x)$ не обязаны быть линейными. Требуется подобрать параметры p такими, чтобы в узлах $x_0, ..., x_n$ было $A(x_i, p) = f(x_i)$; исследовать качество аппроксимации $A(x, p) \approx f(x)$ для x, отличного от узлов. Аналогично ставится задача интерполяции в многомерном случае. Пусть имеется

$$f : (\Omega \subset P^n) \to P \tag{1.17}$$

и $A(x, p)$ – интерполятор: A:$(\Omega \times \hat{\Omega} \subset P^n \times P^m) \to P$. Здесь P – числовое поле, $(x, p) \in \Omega \times \hat{\Omega}$, т.е. интерполяционный агрегат по первому аргументу задан в области, которая содержит область задания исходной функции. Число параметров, то есть размерность набора $p \in \hat{\Omega} \subset P^m$, в хорошем случае равно числу узлов, но, вообще говоря, это не обязательно, в общем случае $n \neq m$.

Укороченная (или абстрактная) интерполяционная задача. *Дан некоторый набор узлов и соответствующие им числа, то есть дана интерполяционная таблица*

x	$x^0 ... x^m$	(1.18)
y	$y_0 ... y_m$	

Дан интерполятор $A(p,x)$. Надо подобрать набор параметров p таким образом, чтобы

$$A(x^i,p) = y_i, \quad i = \overline{0,m}. \tag{1.19}$$

Полная интерполяционная задача. *Даны узлы и значения функции f в них (то есть задана интерполяционная таблица* (1.18) *и известно происхождение $y_0, ..., y_m$):*

$$y_i = f(x^i), \quad i = \overline{0,m}. \tag{1.20}$$

Дан интерполятор $A(p,x)$ и известно, что функции f принадлежат некоторому классу K. Надо:

1) *подобрать набор параметров p таким образом, чтобы выполнялось* (1.19) *то есть решить укороченную интерполяционную задачу*);

2) *оценить остаточный член* $(|r(x)| \leq ?)$ *либо для этой f, либо для класса K.*

$$r(x) := A(x,p) - f(x). \tag{1.21}$$

Дополнительная интерполяционная задача. *Определить каким должен быть набор узлов, чтобы минимизировать погрешность приближения на заданном множестве.*

Рассмотрим многомерную интерполяцию на примере полиноминального интерполянтора. Интерполяция полиномом нулевой степени тривиальна – она всегда единственна и решение всегда существует только тогда, когда имеется один узел.

Пусть речь идет об интерполяции полиномом первой степени. Если узлов больше, чем два, то решение существует не всегда.

1.3.2. Построение интерполянта

В этом случае

$$A(x,a,p) = a + p^{\mathrm{T}} x, \tag{1.22}$$

где $a \in R^1$ – число, $p \in R^n$. Количество параметров $n+1$. Возьмем $n+1$ узел, то есть $n = m$. Выписываем (1.19) для интерполянта (1.22). Получаем систему линейных алгебраических уравнений относительно неизвестных a и p:

$$\left.\begin{array}{c} a + p^{\mathrm{T}} x^0 = y_0 \\ \ldots\ldots \\ a + p^{\mathrm{T}} x^{(m)} = y_m \end{array}\right\}. \tag{1.23}$$

Перепишем систему в матричном виде ($n = m$):

$$\begin{pmatrix} 1 & x_1^0 & \cdots & x_n^0 \\ 1 & x_1^1 & \cdots & x_n^1 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_1^n & \cdots & x_n^n \end{pmatrix} \begin{pmatrix} a \\ p_1 \\ \vdots \\ p_n \end{pmatrix} = \begin{pmatrix} y_0 \\ y_1 \\ \vdots \\ y_n \end{pmatrix}. \tag{1.24}$$

Нижний индекс – номер компоненты, верхний – номер узла.

Иногда систему (1.21) записывают с помощью обобщенного столбца $\hat{x}$:

$$\begin{pmatrix} (\hat{x}^0)^{\text{т}} \\ \vdots \\ (\hat{x}^n)^{\text{т}} \end{pmatrix} \begin{pmatrix} a \\ p \end{pmatrix} = Y, \tag{1.25}$$

где $\hat{x}^i = \begin{pmatrix} 1 \\ x^i \end{pmatrix}$, $Y = \begin{pmatrix} y_0 & \ldots & y_n \end{pmatrix}^{\text{т}}$. Обозначая $X = \begin{pmatrix} (\hat{x}^0)^{\text{т}} \\ \vdots \\ (\hat{x}^n)^{\text{т}} \end{pmatrix}$, получаем

$$X \begin{pmatrix} a \\ p \end{pmatrix} = Y. \tag{1.26}$$

Система (1.26) имеет единственное решение при любом Y тогда и только тогда, когда ее определитель неравен нулю.

$$|X| \neq 0 \tag{1.27}$$

Отметим, что если количество узлов меньше размерности $n+1$ (то есть $m < n$), то, когда решение систем (1.23) $\Leftrightarrow$ (1.24) существует, оно не единственно. А когда количество узлов больше $n+1$ (то есть $m > n$), то у (1.23) $\Leftrightarrow$ (1.24) либо для всех, либо для некоторых наборов $(y_0, ..., y_m)$ решение отсутствует.

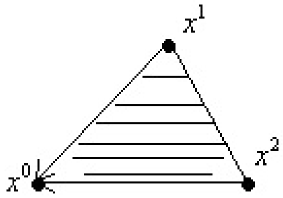
Определение 1.2. *Про набор точек (векторов) аффинного пространства говорят, что они (точки) находятся в общем положении, если они образуют симплекс.*

Определение 1.3. *Набор векторов* $x_0, ..., x_m$ *из* m*–мерного векторного пространства* V *называется симплексом, если для любого вектора из этого набора разности*

$x^0 - x^i, \quad i = 0, ..., m$ *образуют базис этого пространства* V. (*Тогда выпуклая оболочка этих векторов имеет ненулевой объем в данном пространстве.*)

Очевидно, что симплекс в пространстве R^n состоит из $m+1$ элемента. Отметим также, что требования "любое подмножество из m элементов образует базис" недостаточно для симплекса.

Пример. Симплекс в двумерном пространстве



$x^0 - x^1$ и $x^0 - x^2$ – линейно независимые.

Площадь линейной оболочки (треугольника, образованного $x^0 - x^1$ и $x^0 - x^2$) не равна нулю.

Теорема 1.7. *Определитель матрицы,составленный из координатных столбцов базисных векторов, отличен от нуля в каком бы ином базисе не были получены эти координатные столбцы. И обратно, если определитель матрицы, составленный из координатных столбцов базисных векторов, отличен от нуля, то эти вектора образуют базис.* (Доказывается в курсе алгебры.)

Покажем, что (1.27) эквивалентно общему положению узлов. По построению

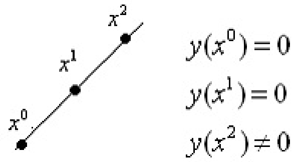
$$|X| = \begin{vmatrix} 1 & x_1^0 & \cdots & x_n^0 \\ \cdots & \cdots & \cdots & \cdots \\ 1 & x_1^n & \cdots & x_n^n \end{vmatrix}.$$

Каждую предыдущую строку, начиная с предпоследней, умножим на x_1 и вычтем из последующих и разложим определитель по первому столбцу.

$$|X| = \begin{vmatrix} 1 & x_1^0 & \cdots & x_n^0 \\ 0 & x_1^1 - x_1^0 & \cdots & x_n^1 - x_n^0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & x_1^n - x_1^{n-1} & \cdots & x_n^n - x_n^{n-1} \end{vmatrix} = \begin{vmatrix} x_1^1 - x_1^0 & \cdots & x_n^1 - x_n^0 \\ \cdots & \cdots & \cdots \\ x_1^n - x_1^{n-1} & \cdots & x_n^n - x_n^{n-1} \end{vmatrix}.$$

Согласно теореме 1.7 последний определитель отличен от 0, тогда и только тогда, когда узлы находятся в общем положении.

Рассмотрим для $n = 2$ случай, когда узлы не образуют симплекса. Мы хотим провести плоскость в трехмерном пространстве (третье измерение перпендикулярно чертежу) через три точки $(x^0, y(x^0))$, $(x^1, y(x^1))$, $(x^2, y(x^2))$.



Если все $y(x^i) = 0$, то мы можем провести бесконечно много плоскостей, содержащих эти три точки.

Для этого примера геометрически плоскость существует перпендикулярно чертежу. Но ей не соответствует линейная функция, определенная на R^2.

1.3.3. Погрешность линейной интерполяции в многомерном пространстве

Пусть известны значения функции в узлах $x^0, ..., x^n$. Возьмем интерполяционный агрегат (1.22). Задача интерполяции с этим агрегатом сводится к решению системы (1.23) при $m = n$ или (1.24) $\Leftrightarrow$ (1.25) $\Leftrightarrow$ (1.26).

Пусть f дифференцируема по Фреше в некоторой точке $\hat{x}$. То есть верна формула

$$f(x^i) = f(\hat{x}) + \nabla f(\hat{x})(x^i - \hat{x}) + o\left(\left\|x^i - \hat{x}\right\|, x^i, \hat{x}\right), \quad i = \overline{0, n}. \quad (1.28)$$

Здесь $\left|o\left(\Delta, x^i, \hat{x}\right)\right| / \Delta \xrightarrow[\Delta \to 0]{} 0$, градиент ∇ трактуется как строка.

Вычтем из каждого последующего уравнения системы (1.23) первое и преобразуем согласно (1.19), (1.20), (1.22) и (1.28)

$$p(x^i - x^0) = y_i - y_0 =$$

$$= \nabla f(\hat{x})(x^i - x^0) + o\left(\left\|x^i - \hat{x}\right\|, x^i, \hat{x}\right) - o\left(\left\|x^0 - \hat{x}\right\|, x^0, \hat{x}\right), \quad i = \overline{1, n}.$$

Преобразуем эту систему к матричному виду.

$$\left(p - \nabla f(\hat{x})\right) V = \hat{o}, \quad (1.29)$$

где $\hat{o} =$

$$= \left(o\left(\left\|x^1 - \hat{x}\right\| x^1, \hat{x}\right) - o\left(\left\|x^0 - \hat{x}\right\| x^0, \hat{x}\right), ..., o\left(\left\|x^n - \hat{x}\right\| x^n, \hat{x}\right) - o\left(\left\|x^0 - \hat{x}\right\| x^0, \hat{x}\right)\right)$$

и $V = (x^1 - x^0, ..., x^n - x^0)$.

Пусть $q^1, ..., q^n$ – линейно независимые векторы. Рассмотрим семейство узлов

$$\left\{x^0, x^0 + \mu q^1, ..., x^0 + \mu q^n\right\}_{\mu>0}. \tag{1.30}$$

Положим $x^i = x^0 + \mu q^i$, $i = \overline{1, n}$, $\mathrm{V}_1 := (q^1, ..., q^n)$, $\mathrm{V}_\mu := (\mu q^1, ..., \mu q^n) = \mu \mathrm{V}_1$. Очевидно, что

$$(\mathrm{V}_\mu)^{-1} = \frac{1}{\mu} \cdot \mathrm{V}_1^{-1}. \tag{1.31}$$

Если положить $\widehat{x} = x^0$, то вид $\hat{o}$ упрощается:

$$\hat{o}{=}o^0 := \left(o(\left\|x^1 - x^0\right\|, x^1, x^0), ..., o(\left\|x^n - x^0\right\|, x^n, x^0)\right) =$$

$$= \left(o(\mu\left\|q_1\right\|, x^1, x^0), ..., o(\mu\left\|q_n\right\|, x^n, x^0)\right) =: (o_1^0, ..., o_n^0).$$

И тогда для симплекса (1.30) система (1.29) такова:

$$\left(p - \nabla f(x^0)\right) V_\mu = o^0 \iff$$

$$p - \nabla f(x^0) = o^0 V_\mu^{-1} {=} \left(\frac{o_1^0}{\mu}, ..., \frac{o_n^0}{\mu}\right) V_1^{-1}. \tag{1.32}$$

Введем обозначение $M := \max\|q^i\|$. Тогда

$$\frac{\left|o_i^0\right|}{\mu} = \frac{\left|o\left(\mu\left\|q^i\right\|, x^i(\mu), x^0\right)\right| ||q_i||}{\mu ||q_i||} \leq$$

$$\leq \frac{\left|o\left(\mu\left\|q^i\right\|, x^i(\mu), x^0\right)\right| M}{\mu\left\|q^i\right\|} \underset{\mu\to 0}{\longrightarrow} 0, \quad i = \overline{1, n}. \tag{1.33}$$

Из (1.32) и (1.33) следует $p \underset{\mu\to 0}{\longrightarrow} \nabla f(x^0)$, то есть параметр p из задачи линейной интерполяции стремится к градиенту fв точке x^0 при стягивании симплекса (1.29) в эту точку. Таким образом получается, что из решений задачи интерполяции можно извлечь приближение к градиенту в точке x^0. Погрешность приближений к исходной функции извлекается из формул (1.32), (1.33). Найдем погрешность интерполятора в некоторой точке x.

$$r(x) := a + p^{\mathrm{T}}x - f(x) = a + p^{\mathrm{T}}x - (a + p^{\mathrm{T}}x^0) + (a + p^{\mathrm{T}}x^0 - f(x^0)) + f(x^0) - f(x) =$$

$$= p^{\mathrm{T}}(x - x^0) + \nabla f(x^0)(x - x^0) + o(\|x - x^0\|, x, x^0) =$$

$$= o^0 V_\mu^{-1}(x - x^0) + o(\|x - x^0\|, x, x^0).$$

1.3.4. Линейная интерполяция вектор-функций

Пусть

$$f : (\Omega \subset P^n) \to P^r. \tag{1.34}$$

В качестве интерполятора берется

$$A(x, a, p) := a + p^{\mathrm{T}} x, \tag{1.35}$$

где $a \in P^r$, $p = (p_1, ..., p_n)$, $\{p_i\}_1^r \subset P^n$, т.е. p – матрица размерности $(r \times n)$.

Ставится задача найти p, такую, что

$$f(x^i) = A(x^i, a, p), \quad i = 1, ..., n. \tag{1.36}$$

Нетрудно заметить, что задача (1.36) разбивается на r независимых подзадач

$$f_j(x^i) = a_j + (p_{j1} ... p_{jn}) x_i, \quad j = 1, ..., r, \tag{1.37}$$

где f_j – j-я компонента f, a_j – j-й элемент столбца a, p_{jk} – k-й элемент j-го столбца матрицы p.

К каждой такой задаче применимы результаты предыдущего пункта.

§1.4. ИНТЕРПОЛЯЦИЯ ВЕЩЕСТВЕННОЗНАЧНОЙ ФУНКЦИИ, ЗАДАННОЙ В КОМПЛЕКСНОМ ВЕКТОРНОМ ПРОСТРАНСТВЕ

Пусть

$$f : C^n \longrightarrow R \tag{1.38}$$

В качестве интерполятора рассмотрим

$$G(x, h) = h_0 + hx + x^* h^*, \tag{1.39}$$

где $h_0 \in R$, R – множество вещественных чисел, $h = (h_1 ... h_n)$, $\{h_i\}_1^n \in C$. (О том, насколько велика общность представления (1.39) см. далее теорему 2.1.) Нетрудно заметить, что интерполятор взаимно однозначно представим в виде

$$G(x, h) = v_0 + v \begin{pmatrix} \Re x \\ \Im x \end{pmatrix}, \tag{1.40}$$

где $v^{\mathrm{T}} \in R^{2n}$, $v_0 = h_0$, очевидно, что $v = 2(\Re h, -\Im h)$.

Таким образом, интерпретируя множество комплексных чисел как R^2, и, соответственно, C^n как R^{2n}, задача интерполяции в комплексном векторном пространстве сводится к уже рассмотренной ранее интерполяции в R^{2n}.

Комплексную специфику можно использовать, когда есть возможность выбора узлов для интерполяции. Судя по (1.40), нужно получить 2n+1 вещественных числа. Т.е. измерить функцию f в 2n+1 различных узлах.

Отметим несколько парадоксальный факт. Если бы нужно было найти линейный интерполянт вида $\Gamma(x,\widehat{h}) = \gamma + \widehat{h}x$ функции $\varphi : C^n \to C$, то хватило бы измерений функции φ только в $n+1$ узле, причем интерполянту Γ для описания требуется даже на одно вещественное число больше информации, чем для описания G (дополненение за счет комплексности γ).

Легче всего вычисляются параметры в (1.40) если придавать узлам $(x^0,...,x^{2n})$ значения $(0,e_1,...,e_n,\mathrm{i}e_1,...,\mathrm{i}e_n)$. Тогда

$$v_0 = f(0), \quad v_i = f(x^i) - v_0, \quad i = 1,...,2n.$$

Что в переходе к параметрам h_0, h дает

$$h_0 = f(0), \quad \Re h_j = (f(e_j) - h_0)/2, \quad \Im h_j = -\mathrm{i}(f(\mathrm{i}e_j) - h_0)/2, \quad j = 1,...,n.$$

То есть

$$h_j = (f(e_j) - h_0)/2 - \mathrm{i}(f(\mathrm{i}e_j) - h_0)/2, \quad j = 1,...,n.$$

Глава 2. Квадратичная интерполяция и аппроксимация

Каждый шаг метода квадратичной оптимизации состоит из двух ступеней: получения коэффициентов аппроксимирующей квадратичной программы и ее решения. Рассмотрим их по отдельности.

§2.1. РАЗЛИЧНЫЕ ПОДХОДЫ К ПОСТРОЕНИЮ КВАДРАТИЧНЫХ АППРОКСИМАЦИЙ

Существуют различные способы вычисления коэффициентов квадратичного приближения функции многих переменных.

Аналитический – это приближение функции тремя первыми слагаемыми ее ряда Тэйлора

$$f(x) \approx \tilde{f}(x, x^0) := f(x^0) + \nabla f(x^0)(x - x^0) + \frac{1}{2}(x - x^0)^{\mathrm{T}} H_f(x^0)(x - x^0),$$

где ∇f — градиент функции, трактуемый как строка, H_f – матрица Гесса функции f, т.е. матрица из вторых производных этой функции. Использование этого способа подразумевает возможность получить расчетные аналитические формулы для градиента и матрицы Гесса, а также наличие оценки в некоторой окрестности x^0 погрешности приближения вида $|f(x) - \tilde{f}(x, x^0)| < E(x, x^0)$. Где E некоторая функция, бесконечно малая относительно скалярного квадрата разности аргументов: $x \to x^0 \implies E(x, x^0)/(x - x^0)^2 \to 0$.

Под *косвенно аналитическим* способом будем понимать такой, в котором градиент и матрица Гесса в узле x^0 вычисляются приближенно тем или иным способом. При использовании *разностных схем* они обычно определяются по формулам центральных разностей:

$$\frac{\partial f(x)}{\partial x_i} \approx \Delta_f(x, h_i, e_i) := \frac{f(x + h_i e_i) - f(x - h_i e_i)}{2h_i}, \qquad i = \overline{1, n},$$

$$\frac{\partial^2 f(x)}{\partial x_i \partial x_j} \approx \frac{\Delta_f(x + h_j e_j, h_i, e_i) - \Delta_f(x - h_j e_j, h_i, e_i)}{2h_j}, \qquad i, j = \overline{1, n},$$

здесь e_i — единичный вектор; $h_i, h_j > 0$ — приращения по i-й и j-й координатам соответственно.

Разностные схемы могут применяться для получения коэффициентов квадратичных аппроксимаций нелинейных программ произвольного происхождения. аналитический способ и использование разностных схем имеет свои преимущества и свои недостатки. Кроме того, они имеют один общий неустранимый недостаток. Даже если априори целевая функция выпукла, это еще не гарантия того, что вычислительные погрешности не породят невыпуклые квадратичные аппроксимциии. Появление последних крайне отрицательно скажется на сходимости всего процесса. Устранение такого дефекта возможно, когда порождение прграмм-аппроксимаций является решением задачи аппроксимации функции по узлам и ее значениям, и эта задача обременена дополнительным условием выпуклости результата.

§**2.2. КВАДРАТИЧНЫЕ АГРЕГАТЫ**

Перейдем к квадратичным аппроксиматорам. Если в вещественном случае квадратичность понимается однозначно, то в комплексном у нее возможны весьма различные трактовки.

Пусть P — вещественное или комплексное поле; $h_{00} \in P$; h_0 — n-мерная строка над полем P; H — симметричная $[n \times n]$-матрица над полем P. С использованием этих обозначений произвольная вещественная $(P = R)$ квадратичная функция и произвольная квадратичная функция 1-го рода $(P = C)$ представимы в виде

$$\widetilde{f}(\check{H}, x) = \check{x}^{\mathrm{T}} \check{H} \check{x}, \qquad \check{x} = \begin{pmatrix} 1 \\ x \end{pmatrix}, \qquad \check{H} = \begin{pmatrix} h_{00} & h_0^{\mathrm{T}} \\ h_0 & H \end{pmatrix}. \qquad (2.1)$$

Назовем агрегаты вида (2.1) РР-агрегатами. Далее для задач интерполяции и аппроксимации данным от измерений комплексно-значных функций на комплексных узлах квадратичными агрегатами не 1-го рода будет предлагаться переход к соответствующей вещественной интерпретации. Такой же переход будет предлагаться и для вещественных данных на комплексных узлах, кроме эрмитовых аппроксимирующих агрегатов (Э-агрегатов)

$$\widetilde{f}(\check{H}, x) = \check{x}^{*} \check{H} \check{x}, \qquad \check{x} = \begin{pmatrix} 1 \\ x \end{pmatrix}, \quad \check{H} = \begin{pmatrix} h_{00} & h_0 \\ h_0^{*} & H \end{pmatrix}, \quad h_{00} \in R, \quad H = H^{*}, \qquad (2.2)$$

для которых возможно менее трудоемкое решение исходных задач.

Определение 2.1. *Задачу ПИ квадратичным интерполятором вида* (2.1) *или* (2.2) *назовем задачей линейно-квадратичной интерполяции (ЛКИ)* [1].

Задача линейно-квадратичной аппроксимации с критерием $\widetilde{Q}$ (ЛКА($\widetilde{Q}$)). Пусть даны значения $y_1, ..., y_s$ скалярной функции $f(x)$, в узлах $x^1, ..., x^s \in P^n$, и среди них имеется подмножество узлов такое, что существует единственная квадратичная функция, решающая задачу ЛКИ по этому подмножеству. (См. вариант *подходящего выбора узлов* далее, формулы (2.7), (2.11) и дополнительно для Э-агрегата формулы (2.7), (2.11).) Требуется построить квадратичную функцию $\widetilde{f}(x)$ вида (2.1), либо (2.2), наилучшим образом в смысле минимизации некоторого критерия $\widetilde{Q}$ приближающую f. ■

Сколь благоприятен выбор

$$Q(\check{H}) := \sum_{1}^{s} \left| y_i - \widetilde{f}(\check{H}, \check{x}^i) \right|^2 . \tag{2.3}$$

в качестве критерия $\widetilde{Q}$ для задачи ЛКА($\widetilde{Q}$)?

Лемма 2.1. *Если среди узлов аппроксимации имеется подмножество, на котором задача ЛКИ имеет единственное решение, то функция Q, определенная в* (2.3)*, строго выпукла.*

Доказательство. Когда $\widetilde{f}$ из (2.1) расположим все наддиагональные и диагональные элементы матрицы $\check{H}$, друг за другом в виде строки $\bar{g}$, и все наддиагональные и диагональные элементы матрицы $\check{x}\check{x}^{\mathrm{T}}$ друг под другом в виде столбца $w(\check{x})$ в том же порядке. Тогда $\check{x}^{\mathrm{T}}\check{H}\check{x} = \bar{g}\omega(x)$. Применим теорему 1.5.

В случае (2.2) существует такая вещественная симметричная матрица $\check{V}$, что

$$\widetilde{f}(\check{H}, x) = (1, \Re x^{\mathrm{T}}, \Im x^{\mathrm{T}})\, \check{V}\, (1, \Re x^{\mathrm{T}}, \Im x^{\mathrm{T}})^{\mathrm{T}} .$$

Воспользуемся предыдущими рассуждениями. ■

---

[1] Т.е. линейной по параметрам интерполяционного агрегата, являющегося квадратичной функцией.

Если считать, что операция сопряжения (*) матрицы над вещественным полем есть ее обычное транспонирование, то вещественный РР-агрегат из (2.1) и Э-агрегат из (2.2) описываются единообразно:

$$\left.\begin{array}{c} \widetilde{f}(x) = h_{00} + h_0 x + x^* h_0^* + x^* H x = \check{x}^* \check{H} \check{x}, \\ \check{x} = \begin{pmatrix} 1 \\ x \end{pmatrix}, \qquad \check{H} = \begin{pmatrix} h_{00} & h_0 \\ h_0^* & H \end{pmatrix}, \qquad h_{00} \in R, \qquad H = H^*. \end{array}\right\} \tag{2.4}$$

Действительно. Упростив операцию "звездочка" в (2.4) до операции транспонирования, получим тоже, что и выполнив поблочное перемножение в (2.1):

$$\widetilde{f}(x) = h_{00} + h_0 x + x^{\mathrm{T}} h_0^{\mathrm{T}} + x^{\mathrm{T}} H x = h_{00} + 2h_0 x + x^{\mathrm{T}} H x \tag{2.5}$$

Насколько общим является представление вещественной квадратичной функции в виде (2.4)?

В вещественном случае очевидно, что этот вид — самый общий.

Для векторных пространств над комплексным полем потребуется более четко определиться с понятиями. Трактуя C^n как R^{2n}, в качестве наиболее общего целесообразно принять

Определение 2.2. *Назовем вещественной квадратичной функцией от комплексного аргумента* $x \in C^n$ *функцию вида*

$$\widetilde{f}(x) = v_{00} + 2v_0 \begin{pmatrix} \Re x \\ \Im x \end{pmatrix} + (\Re x^T \ \Im x^T)\, V \begin{pmatrix} \Re x \\ \Im x \end{pmatrix} = (1 \ \Re x^T \ \Im x^T)\, \check{V} \begin{pmatrix} 1 \\ \Re x \\ \Im x \end{pmatrix}, \tag{2.6}$$

здесь $\Re$ *и* $\Im$ *операции выделения вещественной и мнимой части,* v_{00} *— вещественное число,* v_0 *— вещественная строка длинной* $2n$, V *— размером* $[2n \times 2n]$ *вещественная симметричная матрица* $\check{V} = \begin{pmatrix} v_{00} & v_0 \\ v_0^* & V \end{pmatrix}$. *Линейную часть функции* (2.6) *назовем вещественной линейной функцией комплексного аргумента* $x \in C^n$.

Агрегат (2.6) зависит от $(2n+1)(n+1)$ вещественных параметров. Вещественная часть агрегата (2.4) зависит от $(n+1)(n+2)/2$ параметров, в его мнимой части — $n(n+1)/2$ параметра, т.е. агрегат (2.4) описывается в общей сложности $(n+1)^2$ вещественными параметрами. Следовательно, (2.4) беднее (2.6) на $n(n+1)$ вещественных параметра.

Закрывает вопрос об общности в комплексном случае

Теорема 2.1. *Агрегат* (2.4) *является частным случаем агрегата* (2.6).

Доказательство. Пусть имеется агрегат (2.4). Положим $v_{00} = h_{00}$. Сделаем преобразования:

$$x^* h_0^* + h_0 x = \Re x^{\mathrm{T}} \Re h_0^{\mathrm{T}} - \Im x^{\mathrm{T}} \Im h^{\mathrm{T}} + \Re h_0 \Re x - \Im h_0 \Im x = 2\Re h_0 \Re x - 2\Im h_0 \Im x.$$

Положим $v_0 = 2\,(\Re h_0,\ -\Im h_0)$. Тогда

$$h_{00} + h_0 x + x^* h_0^* = v_{00} + v_0 \begin{pmatrix} \Re x \\ \Im x \end{pmatrix},$$

и, таким образом, линейная часть агрегата (2.4) представлена в виде линейной части агрегата (2.6). Верно и обратное: произвольной линейной части агрегата (2.6) соответствует линейная часть того вида, который использован в агрегате (2.4). Таким образом, линейная часть агрегата (2.4) является линейной вещественной функцией общего вида.

Установим связь между квадратичными частями.

$$x^* H x = \Re x^{\mathrm{T}} H \Re x + \Im x^{\mathrm{T}} H \Im x + \mathrm{i} \Re x^{\mathrm{T}} H \Im x - \mathrm{i} \Im x^{\mathrm{T}} H \Re x =$$

$$= \Re x^{\mathrm{T}} \Re H \Re x + \Im x^{\mathrm{T}} \Re H \Im x - \Re x^{\mathrm{T}} \Im H \Im x + \Im x^{\mathrm{T}} \Im H \Re x =$$

$$= (\Re x^{\mathrm{T}}\ \Im x^{\mathrm{T}}) \begin{pmatrix} \Re H & -\Im H \\ \Im H & \Re H \end{pmatrix} \begin{pmatrix} \Re x \\ \Im x \end{pmatrix}.$$

В этой цепи во втором переходе использованы кососимметричность $\Im H$ и симметрчность $\Re H$, в силу которых $y^{\mathrm{T}} \Im H y = 0\ \forall y$ и $y^{\mathrm{T}} \Re H z = z^{\mathrm{T}} \Re H y\ \forall y, z$.

Таким образом, агрегат (2.4) представлен в виде агрегата (2.6), но полученная в этом представлении симметричная блочная матрица не является симметричной матрицей общего вида, ибо по диагонали стоят совпадающие подматрицы $\Re H$ (минус $n(n+1)/2$ параметров), а внедиагональная подматрица в силу кососимметричности $(-\Im H = \Im H^{\mathrm{T}})$ не является произвольной матрицей как соответствующая ей подматрица в V из (2.6). (В $\Im H$ всего лишь $n(n-1)/2$ независимых параметров) ■

Таким образом, общий случай вещественнозначного квадратичного интерполяционного агрегата от комплексного аргумента из пространства C^n соответствует вещественному квадратичному агрегату (2.6), заданному в R^{2n}. Тогда для построения (2.6) используются те же алгоритмы, что и для вещественного случая. В частном случае, когда (2.6) представим в виде (2.4) будут предложены вычислительно более экономные алгоритмы.

Замечание 1. Когда все узлы исходной задачи лежат в вещественном пространстве, допущение комплексности матриицы $\check{H}$ в интерполяционном агрегате (2.4) не дает возможности уменьшить критерий Q. В самом деле, мнимая компонента матрицы $\check{H}$, как легко проверить, не сказывается на значении $\tilde{f}(x)$, если x вещественен. ■

§2.3. ВЫБОР УЗЛОВ ИНТЕРПОЛЯЦИИ

Введенные выше задачи ЛКИ и ЛКА могут являться частями более общих проблем, в которых перед определением параметров аппроксимирующего агрегата предстоит еще выбрать узлы и вычислить в них значения исходной функции. Весьма часто из каких-то внешних соображений назначается некоторый начальный узел x^0, который должен быть включен в набор, постулируются ограничения сверху и иногда снизу на удаленность прочих узлов от x^0 и друг от друга, и задается предельный уровень сплющенности набора.

Рассмотрим по отдельности РР-агрегаты и Э-агрегаты.

РР-агрегат (2.1) или (2.5) зависит от $N = (n+1)(n+2)/2$ параметров (вещественных, либо комплексных). Следовательно, единственность решения задачи интерполяции агрегатом (2.4) на s узлах может быть обеспечена только при $s \geq N$. Понятно, однако, что этого не достаточно. Общее условие, которому должно подчиняться расположение узлов для гарантии единственности, громоздко аналитически, и из него затруднительно извлечь общие рекомендации по их выбору. Можно, однако, предложить конкретный вариант выбора $s = N$ узлов [10], обеспечивающих и единственность и низкую чувствительность решения задач ЛКИ к погрешностям в измерениях значений $y_1, ..., y_s$, и который выделяется удобством вычисления параметров агрегата. [2)]

Итак, узлы для построения (2.5).

1. Начальный узел x^0 получим из внешней задачи.

2. Также, исходя из внешней задачи, вычислим параметр $\varepsilon > 0$, характеризующий удаленность узлов друг от друга. Им может быть, например, половина радиуса минимального шара с центром

[2)] Впрочем следует отметить, что любой вариант выбора узлов, гарантирующий единственность, при идеальных вычислениях приведет к одинаковым параметрам агрегата (2.4). В частности, всегда $h_{00} = f(0)$.

в x^0, содержащего область, в которой намечено получать значения аппроксимации. Если измерение значений аппроксимируемой функции происходят с ошибкой, то параметр ε должен существенно превосходить оценку этой ошибки, деленную на оценку нормы градиента.

3. Назначим

$$\left.\begin{aligned} x^k\ &= x^0 + \varepsilon e^k \\ x^{n+k} &= x^0 - \varepsilon e^k \end{aligned}\right\} \qquad k=\overline{1,n}. \tag{2.7}$$

Здесь e^k, $k=\overline{1,n}$ — единичные столбцы.

Преобразуем представление (2.5):

$$\widetilde{f}(x) = \widetilde{f}(x \pm x^0) =$$

$$= \widetilde{f}(x^0) + 2\left((x^0)^{\mathrm{T}} H + h_0\right)(x - x^0) + (x - x^0)^{\mathrm{T}} H (x - x^0). \tag{2.8}$$

При подстановке в (2.8) вместо x столбцов x^i из (2.7) и обозначениях

$$w_k = 2\left((x^0)^{\mathrm{T}} H + h_0\right) e^k, \qquad k=\overline{1,n}, \tag{2.9}$$

видно, что для $k=\overline{1,n}$ решение задачи ЛКИ должно удовлетворять системе линейных уравнений относительно w_k и диагональных элементов h_{kk}:

$$\left.\begin{aligned} y_0 + \varepsilon\, w_k + \varepsilon^2 h_{kk} = y_k, \\ y_0 - \varepsilon\, w_k + \varepsilon^2 h_{kk} = y_{n+k}. \end{aligned}\right\}. \tag{2.10}$$

Определитель матрицы при неизвестных h_{kk}, w_k в системе (2.10) равен $2\varepsilon^3 \neq 0$, следовательно, она их однозначно определяет.

Отметим, что интерполятор вида

$$h_{00} + 2h_0(x - x^0) + (x - x^0)^{\mathrm{T}} H (x - x^0)$$

применяется не реже, чем (2.5). В этом случае $h_{00} = f(x^0)$ и $h_0 = w/2$.

4. Для определения недиагональных элементов матрицы H назначим следующие узлы:

$$x^{2n+(2n-k)(k-1)/2+j-k} = x^0 + \varepsilon e^k + \varepsilon e^j, \quad k=\overline{1,n-1}, \quad j=\overline{k+1,n}. \tag{2.11}$$

На таких узлах [3] условия совпадения значений агрегата и исходной функции имеют вид

$$y_0 + \varepsilon\, w_k + \varepsilon\, w_j + \varepsilon^2 h_{kk} + \varepsilon^2 h_{jj} + 2\varepsilon^2 h_{kj} = y_{(k+1)n-k(k+1)/2+j}.$$

Отсюда [4] однозначно определяются недиагональные элементы матрицы H. Потом однозначно из формул (2.9) вычисляется строка h_0, затем h_{00} из условия $f(x^0) = \widetilde{f}(x^0)$:

$$h_0 = \frac{1}{2}(w_1, ..., w_n) - (x^0)^{\mathrm{T}} H, \qquad h_{00} = y_0 - 2h_0 x^0 - (x^0)^{\mathrm{T}} H x^0. \qquad \blacksquare$$

Теорема 2.2. *Наличие в составе узлов подмножества вида* (2.7) *и* (2.11) *гарантирует существование и единственность решения задачи ЛКА*(Q) (Q *из* (2.3)).

Д е й с т в и т е л ь н о. Положив

$$\omega(x) = (1 \;\; x_1 \; ... \; x_n \;\; x_1^2 \; ... \; x_1 x_n \;\; x_2^2 \; ... \; x_2 x_n \; ... \; x_3^2 \; ... \; x_n^2)^{\mathrm{T}},$$

$$\bar{g} = (h_{00} \;\; h_0 \;\; h_{11} \;\; 2h_{12} \; ... \; 2h_{1n} \;\; h_{22} \; ... \; 2h_{2n} \; ... \; h_{33} \; ... \; h_{nn}),$$

видим, что

$$Q_1(g) \equiv \sum_1^s \left|y_k - \bar{g}\omega(x^k)\right|^2 = \sum_1^s \left|y_k - h_{00} - h_0 x^k - (x^k)^{\mathrm{T}} H x^k\right|^2 =$$
$$= \sum_1^s \left|y_k - (\check{x}^k)^{\mathrm{T}} \check{H} \check{x}^k\right|^2 \equiv Q(\check{H}). \tag{2.12}$$

Следовательно, предлагаемый набор узлов обеспечивает существование и единственность решения задачи ЛКИ агрегатом Q_1. Поэтому в силу теоремы 1.5 минимайзер агрегата Q_1 существует, конечен и единствен. Что верно в силу (2.12) и ддя агрегата Q. $\blacksquare$

Включение произвольных дополнительных узлов в совокупность (2.7), (2.11) может повысить чувствительность решения задачи ЛКА(Q) к погрешностям входных данных. Так, например, происходит когда эти узлы располагаются в линейном многообразии,

---

[3] Отметим, что относительно сложный вариант отображения $i(\cdot,\cdot)$ пар индексов (k, j) во множество индексов $\{2n+1, N\}$, приведенный в (2.11) не принципиален. Сгодится любое взаимно однозначное отображение.

[4] Отметим, что здесь немного иначе записанный индекс у y совпадает с индексом у x в (2.11). И в общем случае отображения пар индексов (k, j) во множество индексов $\{2n+1, N\}$ в обеих формулах должны совпадать.

проходящем через x^0. Напротив, если они выбирались равномерно распределенными в шаре или на сфере с центром в x^0 радиусом ε, то чувствительность решения БЗ к погрешностям в численных экспериментах понижалась. А если исходная функция была квадратичной, то вычисляемые параметры аппроксимации стремились к параметрам исходной функции. То же происходило и при добавлении новых узлов сериями вида (2.7) + (2.11) с различными ε.

В пересчете на вещественные параметры Э-агрегат (2.2) беднее РР-агрегата над комплексным полем: $(n+1)(n+2)/2+n(n+1)/2 = (n+1)^2$ параметров против $(n+1)(n+2)$. Однако для определения параметров Э-агрегата приходиться вводить дополнительные узлы. Это объясняется тем, что каждый узел дает лишь одно вещественное число Э-агрегату и в два раза больше информации в виде комплексного числа РР-агрегату. Для построения вещественного квадратичного агрегата общего вида (2.6) потребуется еще большее число узлов, совпадающее, как и в случае Э-агрегатов с числом параметров. Для РР-агрегатов, Э-агрегатов и агрегатов вида (2.6) количества узлов, соответственно, таковы

$$(n+1)(n+2)/2 < (n+1)^2 < (2n+1)(n+1).$$

Итак, первые N узлов для Э-агрегатов назначаются также, как для вещественного случая (см. пп. 1–4, формулы (2.7), (2.11)). В прочие узлы вводится комплексное отклонение от начального. Пусть $\mathfrak{i}$ — мнимая единица.

5. Полагая, как и ранее, e^k единичным столбцом, назначим

$$x^{N+k-1} = x^0 + \mathfrak{i}\varepsilon e^k, \quad k = \overline{1, n}. \tag{2.13}$$

6. Назначим

$$x^{N+(2n-k+1)k/2+j-k-1} = x^0 + \varepsilon e^k + \mathfrak{i}\varepsilon e^j, \quad k = \overline{1, n-1}, \quad j = \overline{k+1, n}. \tag{2.14}$$

Покажем, что узлы (2.7)–(2.14) обеспечивают единственность решения задачи интерполяции агрегатом (2.2). Для этого опишем схему вычислений параметров этого агрегата, в которой единственность результата очевидна.

1) Подставляем узлы $x^k = x^0 + \varepsilon e^k$, $k = \overline{1, n}$ в агрегат (2.2) и выражаем коэффициент w_k при первой степени ε через параметры агрегата.

$$\widetilde{f}(\check{H}, x^k) = \widetilde{f}(\check{H}, x^0) + \left[\left((x^0)^* H + h_0\right) e^k + (e^k)^{\mathrm{T}} \left(H x^0 + h_0^*\right)\right] \varepsilon + (e^k)^{\mathrm{T}} H e^k \varepsilon^2, \tag{2.15}$$

$$w_k = \left((\bar{x}^0)^{\text{т}} H + (x^0)^{\text{т}} H^{\text{т}} + 2\Re h_0\right) e^k. \qquad (2.16)$$

Используем $\bar{H} = H^{\text{т}}$:

$$w_k = 2\left((\Re x^0)^{\text{т}} \Re H + (\Im x^0)^{\text{т}} \Im H + \Re h_0\right) e^k. \qquad (2.17)$$

Используя условие совпадения значений интерполянта и исходной функции на узлах, получаем

$$y_{k+n} = y_0 + w_k \varepsilon + e^k H e^k \varepsilon^2. \qquad (2.18)$$

Отметим, что при подстановке узлов $x^{n+k} = x^0 - \varepsilon e^k, \quad k = \overline{1,n}$ коэффициенты при первой степени ε будут равны аналогичным для предыдущей серии, взятым с обратным знаком, т.е.

$$y_{k+n} = y_0 - w_k \varepsilon + e^k H e^k \varepsilon^2. \qquad (2.19)$$

2) Связи (2.18) и (2.19) образуют невырожденную линейную систему относительно w_k и h_{kk}, совпадающую вплоть до обозначений с системой (2.10):

$$\left.\begin{array}{rcl} \varepsilon^2 h_{kk} + \varepsilon w_k &=& y_k - y_0 \\ \varepsilon^2 h_{kk} - \varepsilon w_k &=& y_{n+k} - y_0 \end{array}\right\} \qquad k = \overline{1,n}. \qquad (2.20)$$

Отсюда получаем диагональ матрицы H и набор вспомогательных параметров $w_1, ..., w_n$.

3) На узлах (2.11) получим уравнения для определения вещественных частей недиагональных элементов матрицы H:

$$\left.\begin{array}{c} 2\varepsilon^2 \Re h_{kj} + \varepsilon^2 h_{kk} + \varepsilon^2 h_{jj} + \varepsilon w_k + \varepsilon w_j = y_{2n+(2n-k)(k-1)/2+j-k} - y_0, \\ k = 1, ..., n-1, \quad j = k+1, ..., n. \end{array}\right\} \qquad (2.21)$$

4) Подставим узлы x^{N+k-1} серии (2.13) в (2.2) и выразим коэффициенты u_k при первых степенях ε через параметры агрегата:

$$u_k = \left((x^0)^* H + h_0\right) e^k \mathfrak{i} - (e^k)^{\text{т}} \left(H x^0 + h_0^*\right) \mathfrak{i} = \left[(\bar{x}^0)^{\text{т}} H \mathfrak{i} - (x^0)^{\text{т}} H^{\text{т}} \mathfrak{i} - 2\Im h_0\right] e^k =$$
$$= 2\left[\left(\Im x^0\right)^{\text{т}} \Re H - \left(\Re x^0\right)^{\text{т}} \Im H - \Im h_0\right] e^k. \qquad (2.22)$$

Значение параметра u_k определяются из линейного уравнения:

$$\varepsilon u_k + \varepsilon^2 h_{kk} = y_{N+k} - y_0. \qquad (2.23)$$

5) Для определения $\Im H$ используютя узлы (2.14):

$$\left.\begin{array}{c}\varepsilon^2(h_{kk}+h_{jj}-2\Im h_{kj})+\varepsilon w_k+\varepsilon u_j=y_{N+(2n-k+1)k/2+j-k-1}-y_0,\\ k=1,...,n-1,\ j=k+1,...,n.\end{array}\right\} \tag{2.24}$$

6) Для определения $\Re h_0$ используется (2.17), и (2.22) — для определения $\Im h_0$.

7) В заключение вычисляется постоянная составляющая аппроксимации:

$$h_{00}=y_0-h_0x^0-(x^0)^*Hx^0. \tag{2.25}$$

Итак, схема вычислений параметров Э-агрегата (2.2) при интерполяции по узлам (2.7), (2.11), (2.13), (2.14) такова: решаются линейные уравнения в порядке (2.20), (2.21), (2.23), (2.24), (2.17), (2.22), затем применяется (2.25).

§2.4. ВЫПУКЛОСТЬ ЛИНЕЙНО-КВАДРАТИЧНОГО АГРЕГАТА

Требование выпуклости принципиально, когда известно, что исходная функция выпукла. Невыпуклость аппроксимации тогда однозначно свидетельствует о ее низком качестве. Бывают и иные причины добиваться выпуклости аппроксимации.

Задача выпуклой линейно-квадратичной аппроксимации с критерием $\widetilde{Q}$ (ВКА($\widetilde{Q}$)). Имеются входные данные для задачи ЛКА($\widetilde{Q}$): узлы $x_1,...,x_s$ и измеренные в них величины $y_1,...,y_s$. Требуется найти минимайзер программы

$$\widetilde{Q}(\check{H}) \longrightarrow \min_{H\in\bar{K}}. \quad ■ \tag{2.26}$$

Задача ВКА(Q_1) весьма трудна для численного решения из-за сложности аналитического представления условия $H\in\bar{K}$, что хорошо будет видно в последующих двух параграфах.

Интенсивное применение решения задачи ВКА(Q)

$$Q(\check{H}):=\sum_1^s\left(y_i-(\check{x}^i)^*\check{H}\check{x}^i\right)^2, \tag{2.27}$$

как части общего алгоритма находит, например, в поиске экстремума нелинейной функции методом параболоидов [3] и методом последовательной оптимизации [12], где на итерациях строится квадратичная аппроксимация по значениям исходной целевой функции в выбранных узлах, после чего решаются квадратичные программы. Задача ВКА(Q_1) сама является нелинейной программой, для ее решения, помимо приведенных в предыдущих параграфах методов, известны и другие итеративные методы, но все с большой трудоемкостью [9], [10].

Как известно, выпуклость вещественных РР-агрегатов из (2.2) эквивалентна положительной полуопределенности (ППО) матрицы H.

Установим связь между ППО эрмитовой матрицы и выпуклостью агрегата (2.4).

Теорема 2.3. *Выпуклость агрегата* (2.4) *равносильна ППО матрицы* H, *а строгая выпуклость — ПО матрицы* H.

Д о к а з а т е л ь с т в о. Линейная часть агрегата (2.4) очевидно выпукла:

$$h_0(tx+(1-t)y)+(tx+(1-t)y)^*h_0^* = t(h_0x+x^*h_0^*)+(1-t)(h_0y+y^*h_0^*).$$

Выпуклость квадратичной части для ППО матрицы H и строгая выпуклость для ПО матрицы H следует из леммы 1.1. Как известно, сумма двух выпуклых функций выпукла, а выпуклой и строго выпуклой — строго выпукла. ■

Эта теорема позволяет условие выпуклости агрегата (2.4) заменить на условие ППО матрицы H.

Как известно, положительно полуопределенные матрицы образует замкнутый выпуклый конус $\bar{K}$. Положительно определенные матрицы тоже образуют выпуклый конус K, но открытый в отличие от $\bar{K}$. Этим оправдывается требование в задаче ВКА именно простой выпуклости, а не строгой.

§2.5. СКАЛЯРНОЕ ПРОИЗВЕДЕНИЕ МАТРИЦ И УПРОЩЕННЫЕ КРИТЕРИИ ЗНАКООПРЕДЕЛЕННОСТИ

Критерий Сильвестра для ПО матриц состоит из n неравенств с полиномами от элементов матрицы со степенями от 1 до n включительно. Для ППО матриц в критерии уже 2^n неравенств. При больших n использование этого критерия для какого либо метода спуска решения задачи ВКА становится очень трудоемким. В этом параграфе предлагаются упрощенные критерии ПО и ППО матриц, которые позволяют вычислить субантиградиент целевой функции

(2.3), в направлении которого допустимо смещение текущей матрицы, с сохранением ее положительной полуопределенности (ППО) или положительной определенности (ПО) [8], [9]. И, тем самым, производить поиск приближенного решения задачи ВКА.

Для описания критериев, а также для альтернативного подхода к решению задачи ВКА (см. следующий параграф), потребуются некоторые предварительные результаты относительно матриц квадратичных форм. Приведем их.

Не отступая от традиций, под сопряженной к матрице A матрицей A^* здесь будем понимать комплексно сопряженную и транспонированную матрицу: $A^* := (\bar{A})^{\mathrm{T}}$. Таким образом, если матрица A над вещественным полем, то $A^* = A^{\mathrm{T}}$.

Множество матриц одной размерности над полем P с классической операцией умножения на число из поля $P' \in P$ образует векторное пространство. Его можно обогатить до унитарного пространства, введя скалярное произведение $A \cdot B$, двух матриц $A = \|a_{i,j}\|$, $B = \|b_{i,j}\|$ формулой

$$A \cdot B \ := \ \sum_{i,j} \bar{a}_{ij} b_{ij}, \tag{2.28}$$

где “ $^-$ ” — операция комплексного сопряжения.

Нетрудно заметить, что согласно определению (2.28) выполняется

$$B^* \cdot A^* = A \cdot B \tag{2.29}$$

С помощью скалярного произведения введем шуровскую норму:

$$\|D\|_{E} := \sqrt{D \cdot D} = \sqrt{\sum |d_{ij}|^2}.$$

С сответствующей заменой терминов “унитарных матриц и преобразований”, на “ортогональные матрицы и преобразования” все приводимые далее результаты для эрмитовых матриц верны также и для вещественных симметричных матриц. Специально это оговариваться не будет.

Теорема 2.4. *Множество положительно определенных матриц есть выпуклый открытый конус K, а множество положительно полуопределенных матриц есть выпуклый замкнутый конус $\widetilde{K}$, который является замыканием конуса положительно определенных матриц: $\widetilde{K} = \bar{K}$, кроме того, $\partial\bar{K} = \partial K$.*

Д о к а з а т е л ь с т в о. Очевидно, что

$$A \in \bar{K} \Longrightarrow (\forall\alpha \geq 0)\ \alpha A \in \bar{K} \quad \text{и} \quad A \in K \Longrightarrow (\forall\alpha > 0)\ \alpha A \in K.$$

Поэтому оба множества $\widetilde{K}$ и K — конусы.

Выпуклость K проверяется элементарно. В самом деле. Пусть H_1, H_2 положительно определены, т.е. принадлежат K. Рассмотрим "отрезок соединяющий эти два элемента:

$$H(t) = tH_1 + (1-t)H_2, \qquad t \in (0,1).$$

Для всех отличных от нуля $x \in {}^n$ имеет место

$$x^*H(t)x = tx^*H_1x + (1-t)x^*H_2x > 0,$$

откуда $H(t) \in K$, что соответствует выпуклости.

Пусть матрица $A \in K$. Тогда ее угловые главные миноры положительны: $M_{J(j),J(j)}(A) > 0$, $J(j) := \{1,...,j\}$, $j = 1,..,n$. Следовательно, в силу непрерывной зависимости определителя от элеметов матрицы существует положительное ε, такое, что $\|H\|_E < \varepsilon \Longrightarrow (\forall j)\, M_{J(j),J(j)}(A+H) > 0$. А это является критерием ПО матрицы $A+H$, отсюда — открытость конуса K.

Пусть последовательность $\{A_k\} \subset K$ сходится к некоторой матрице A, т.е. $\|A_k - A\|_E \to 0$. Тогда все главные миноры $M_{J,J}(A_k)$, будучи положительными согласно критерию Сильвестра, также стремятся к главным минорам $M_{J,J}(A)$ для всех $J \subset \{1,...,n\}$. Поэтому $M_{J,J}(A) \geq 0$, что в силу критерия ППО матриц означает $A \in \widetilde{K}$, т.е. $\bar{K} \subseteq \widetilde{K}$.

Покажем обратное включение. Пусть $A \in \partial\bar{K}$. Рассмотрим произвольное $\varepsilon > 0$. Матрица $A + \varepsilon I$ положительно определена, как матрица положительно определенной формы. (Следствие того, что εI — положительно определена, и см. выше доказательство выпуклости K.) Таким образом, в любой окрестности A есть матрицы из $\bar{K}$. Поэтому $A \in \bar{K}$ и $\widetilde{K} \subseteq \bar{K}$.

Оказывается, в ∂K нет внутренних точек из конуса $\bar{K}$.

В самом деле, пусть $A \in \partial K$, тогда найдется унитарное преобразование переменных $y = Ux$, $U^*U = I$, такое, что $U^*AU = \operatorname{diag}\operatorname{Eig} A =: B$ и среди собственных чисел матрицы A, так как K открыто, должно быть хотя бы одно не положительное, например, λ_1:

$$x^*Ax = \lambda_1 y_1^2 + \lambda_2 y_2^2 + ... + \lambda_n y_n^2.$$

Рассмотрим матрицу Δ, которая отлична от нулевой только элементом $\delta_{11} = 1$. Но тогда матрицы вида $D_\varepsilon = B + \varepsilon\Delta$ не положительно полуопределены для любого $\varepsilon > 0$, ибо $e_1^{\mathrm{T}} D_\varepsilon e_1 < 0$. Перейдя обратно в переменные x, обнаружим, что матрица $A_\varepsilon := A + \varepsilon U D_\varepsilon U^*$ не имеет ППО для любого $\varepsilon > 0$. Т.е. в любой окрестности матрицы A найдется не положительно полуопределенная матрица. Таким образом A не может быть внутренней для $\bar{K}$.

Как известно, $\partial\bar{K} \subset \partial K \subset \bar{K}$. Но как только что показано, $\partial K \bigcap (\bar{K} \setminus \partial\bar{K}) = \emptyset$. Следовательно, $\partial\bar{K} = \partial K$. ∎

Теорема 2.5. *Скалярное произведение квадратных матриц инвариантно относительно их одновременного умножения слева, либо справа на унитарную матрицу U, а также ортогонального преобразования, т.е. если $UU^* = I$,*[5] *то для произвольных квадратных матриц A, B верно*

$$(AU) \cdot (BU) = (UA) \cdot (UB) = (U^*AU) \cdot (U^*BU) = A \cdot B$$

(I — единичная матрица того же порядка, что и U).

Доказательство. Пусть операция, обозначаемая далее символом $\widehat{\ }$, ставит в соответствие произвольной матрице столбец, состоящий из столбцов этой матрицы, размещенных последовательно в порядке возрастания их номеров в матрице. Обратную операцию к операции $\widehat{\ }$ обозначим также, различение по контексту. Таким образом, $\widehat{\widehat{A}} = A$ имеет смысл и верно, как для A – матрицы, так для A – столбца.

Очевидно, что для любых матриц A, B верно $A \cdot B = \widehat{A}^* \widehat{B}$. Т.е. операция $\widehat{\ }$ устанавливает изоморфизм между унитарным пространством $\mathbb{U}_m$ матриц из $C^{n \times n}$ со скалярным произведением из определения (2.28) и унитарным пространством $\mathbb{U}_c$ столбцов из C^{n^2} с самым простым скалярным призведением.

Пополним пространство $\mathbb{U}_m$ еще одним законом внешней композиции — умножением слева на квадратную матрицу над комплексным полем и соответствующей размерности. Т.е. тоже[6] из $C^{n \times n}$.

---

[5] Из унитарности U следует унитарность U^*, ибо $UU^* = I$ означает, что U^* — правая обратная матрица к U, а в курсе алгебры доказывается, что правая и левая обратные матрицы совпадают, ззначит $U^*U = I$.

[6] Построения можно развивать и когда $\mathbb{U}_m$ — множество прямоугольных матриц из $C^{m \times n}$. Тогда умножать слева нужно на квадратные матрицы из $C^{m \times m}$. Нам, однако, это развитие не потребуется.

Для получения изоморфной алгебраической структуры из пространства $\mathbb{U}_c$ умножению произвольной матрицы X слева на матрицу U следует сопоставить умножение столбца $\widehat{X}$ слева на матрицу

$$\widetilde{U} := \begin{pmatrix} U & 0 & 0 & \dots & 0 \\ 0 & U & 0 & \dots & 0 \\ \dots & & & \dots & \\ 0 & 0 & 0 & \dots & U \end{pmatrix}.$$

Тогда $\widehat{UX} = \widetilde{U}\widehat{X}$.

Нетрудно заметить, что матрица $\widetilde{U}$ унитарна, если унитарна матрица U. Поэтому

$$(UA)\cdot(UB) = (\widehat{UA})^*\widehat{UB} = (\widetilde{U}\widehat{A})^*\widetilde{U}\widehat{B} = \widehat{A}^*\widetilde{U}^*\widetilde{U}\widehat{B} = \widehat{A}^*\widehat{B} \;=\; A\cdot B, \tag{2.30}$$

$$(AU)\cdot(BU) \overset{(2.29)}{=} (BU)^*\cdot(AU)^* = (U^*B^*)\cdot(U^*A^*) \overset{(2.30)}{=} B^*\cdot A^* \overset{(2.29)}{=} A\cdot B.$$

Отсюда, очевидно, вытекает

$$(U^*AU)\cdot(U^*BU) = (U^*A)\cdot(U^*B) \overset{(2.30)}{=} A\cdot B. \qquad \blacksquare$$

Неплохой иллюстрацией к ПО и ППО матриц является понятие угла между матрицами, который можно задать с помощью скалярного произведения, введенного в определении 2.28.

Определение 2.3. *Угол γ между эрмитовыми матрицами $A, B \neq 0$ одних размеров — это число из $[0,\pi]$, удовлетворяющее соотношению*

$$\cos\gamma := \frac{A\cdot B}{\|A\|_E\|B\|_E}. \tag{2.31}$$

Но всегда ли такое γ существует?

Несложно заметить, что скалярное произведение эрмитовых матриц вещественно: $A = A^* \;\wedge\; B = B^* \;\Longrightarrow$

$$\mathfrak{Im}\,(A\cdot B) = -\mathfrak{Im}\,(A^*\cdot B^*) = -\mathfrak{Im}\,(A\cdot B) \;\Longrightarrow\; \mathfrak{Im}\,(A\cdot B) = 0.$$

С другой стороны, имеет место $|A\cdot B| \le \|A\|_E\|B\|_E$ — неравенство Коши-Буняковского, поэтому угол γ из (2.31) вещественен, чем и показана корректность определения 2.3.

Теорема 2.6. *Матрицы из $\bar{K}$ — конуса положительно полуопределенных матриц — образуют друг с другом только углы χ не более 90^0, если же некоторая матрица не принадлежит $\bar{K}$, то существует другая матрица из K, образующая с первой угол $\gamma > 90°$. Иными словами,*

$$B \in \bar{K} \implies (\forall A \in \bar{K}) \;\; A \cdot B \geq 0. \tag{2.32}$$

$$(\forall A \in K) \; A \cdot B \geq 0 \implies B \in \bar{K}. \tag{2.33}$$

Доказательство. Как известно, всегда можно привести симметричную матрицу B к диагональному виду $B_2 := \mathrm{diag}(\mathrm{Eig}\; B)$ унитарным преобразованием: $B_2 = U^*BU \;\; \wedge \;\; U^*U = I$.

Докажем сначала, что $A, B \in \bar{K} \Longrightarrow \chi \leq 90°$, т.е. (2.32).

Пусть $B \in \bar{K}$, тогда $\mathrm{Eig}\; B = \mathrm{Eig}\; B_2 \geq 0$. Любая матрица A из конуса $\bar{K}$ после того же преобразования останется положительно полуопределенной, иными словами $U^*AU =: A_2 \in \bar{K}$. В силу критерия Сильвестра ее диагональные элементы неотрицательны. Поэтому $A_2 \cdot B_2 \geq 0$. Но теорема 2.5 влечет $A_2 \cdot B_2 = A \cdot B$, где A — любая положительно полуопределенная. Следовательно, доказано, что $\chi \leq 90^0$.

Для доказательства импликации (2.33) проверим справедливость ее дополнения (которое прямо соответствует словесной формулировке теоремы):

$$B \notin \bar{K} \implies (\exists A \in K) \; A \cdot B < 0. \tag{2.34}$$

Отметим, что $B \notin \bar{K} \iff (\exists \lambda_i \in \mathrm{Eig}\; B) \; \lambda_i < 0$. Преобразуем ортогональным преобразованием U не являющуюся положительно полуопределенной матрицу B к диагональному виду $B_2 = U^*BU$. Пусть i-й элемент λ в B_2 отрицателен. Назначим матрицу $A_2(\delta)$ отличной от δI только i-м диагональным элементом, равным $1+\delta$. Очевидно, что для всех $\delta > 0$ матрица $A_2(\delta)$ положительно определена. С другой стороны, $A_2(\delta) \cdot B_2 = \lambda + \delta \,\mathrm{Sp}\, B_2$, что отрицательно при достаточно малых $\delta > 0$, тогда полагая $A = UA_2(\delta)U^*$ и используя теорему 2.5, получаем (2.34), эквивалентное (2.33), т.е. $\gamma > 90°$.

Теорема 2.7. *Матрицы из K — конуса положительно определенных матриц — образуют с матрицами из множества $\bar{K} \setminus \{0\}$ только углы χ, меньшие $90°$. Если же имеется матрица не из K, то в $\bar{K} \setminus \{0\}$ найдется матрица, образующая с первой угол $\gamma \geq 90^0$. Иными словами, справедливо:*

$$B \in K \iff \big(\forall A \in \bar{K} \setminus \{0\}\big) \quad A \cdot B > 0, \tag{2.35}$$

Доказательство. Приведем матрицу B к диагональному виду $B_2 := \mathrm{diag}(\mathrm{Eig}\, B)$ унитарным преобразованием:

$$B_2 = U^* B U \quad \wedge \quad U^* U = I.$$

Пусть A_2 и A — некоторые матрицы, связанные соотношением $A_2 = U^* A U$.

“ $\Longrightarrow$”. Если $B \in K$, то все элементы диагонали матрицы B_2 положительны. Если A из $\bar{K} \setminus \{0\}$, то A_2 также из $\bar{K} \setminus \{0\}$. На диагонали A_2 могут, в силу критерия положительной полуопределенности стоять только неотрицательные числа. Более того, одновременно они не могут все быть нулями. Действительно, поскольку $A_2 \neq 0$, в случае нулевой диагонали должна существовать пара недиагональных элементов, например $a_{i,j}^{(2)}$, $a_{j,i}^{(2)}$, отличных от нуля. Но минор второго порядка, включающий эту пару отрицателен, что противоречит критерию положительной полуопределенности. Следовательно, $A_2 \cdot B_2 > 0$, и, по теореме 2.5, $A \cdot B > 0$.

“ $\Longleftarrow$”. Пусть матрица A_2 отличается от нулевой только i-м диагональным элементом, равным 1. Нетрудно видеть, что A_2 из $\bar{K} \setminus \{0\}$, следовательно, A также из $\bar{K} \setminus \{0\}$. Поэтому

$$0 < A \cdot B \overset{\text{т.2.5}}{=} A_2 \cdot B_2 = \lambda_i.$$

Пробегая индексом i от 1 до n, выясняем, что все λ_i положительны, т.е. $B \in K$. ■

Критерий Сильвестра полностью отвечает на вопросы “когда?” и “только когда?” матрица положительно определена. Однако его некоторая громоздкость и низкая “аналитичность” затрудняют применение в теоретических исследованиях. На основе изложенного можно получить более простой достаточное (но не необходимое) и необходимое (но не достаточное) условия положительной полуопределенности (ППО) и положительной определенности (ПО). Эти критерии зауживают область необходимости и запрашивают больше для гарантий ПО и ППО, зато вычислительно очень просты и обладают хорошей “аналитичностью”.

Теорема 2.8. *Для ППО матрицы A необходимо, чтобы*

$$\mathrm{Sp}\, A \geq \|A\|_E, \tag{2.36}$$

и достаточно, чтобы

$$\operatorname{Sp} A \geq \sqrt{n-1}\, \|A\|_{E}. \tag{2.37}$$

Условие вида (2.36) *не улучшаемо в том смысле, что существует положительно полуопределенная матрица* A*, удовлетворяющая*

$$\operatorname{Sp} A = \|A\|_{E}. \tag{2.38}$$

Константа $\sqrt{n-1}$ *в условии* (2.37) *не улучшаема: для всякого* $\delta > 0$ *найдется знаконеопределенная матрица* B*, такая, что*

$$\operatorname{Sp} B > \left(\sqrt{n-1} - \delta\right) \|B\|_{E}. \tag{2.39}$$

Доказательство. Как известно [13], шуровская норма и след матрицы инвариантны относительно унитарного преобразования:

$$U^*U \implies \operatorname{Sp} U^*AU = \operatorname{Sp} A, \quad \|U^*AU\|_{E} = \|A\|_{E}.$$

Таким образом, можно ограничиться в доказательстве рассмотрением только диагональных матриц: $A = \operatorname{diag}(a_1, ..., a_2)$.

Для таких матриц квадрат условия (2.36) имеет вид

$$\sum a_i^2 + 2\sum_{i>j} a_i a_j \geq \sum a_i^2,$$

что, очевидно справедливо при ППО матрицы A. Необходимое условие доказано.

Матрица $A = \operatorname{diag}(1, 0, ..., 0)$ положительно полуопределена и удовлетворяет условию (2.38).

Достаточное условие. Для диагональной матрицы условие (2.37) принимает вид

$$\sum a_i \geq \sqrt{n-1}\sqrt{\sum a_i^2}. \tag{2.40}$$

Как известно, $\|x\|_1 \leq \sqrt{n}\|x\|_2$. Поэтому

$$(\exists j)\ a_j < 0 \implies \sum a_i < \sum_{i\neq j} a_i \leq \sqrt{n-1}\sqrt{\sum_{i\neq j} a_i^2} < \sqrt{n-1}\|A\|_{E}. \tag{2.41}$$

Дополнение к этому высказыванию имеет вид

$$\sum a_i \geq \sqrt{n-1}\|A\|_{E} \implies (\forall i)\ a_i \geq 0.$$

Левая часть последнего высказывания есть (2.40), а правая означает ППО матрицы A.

Для проверки утверждения (2.39) введем матрицу $B(t) = \mathrm{diag}(-t, w, ..., w)$, где $w = \sqrt{1 - t^2/(n-1)}$. Она вещественна и знаконеопределена, когда $t \in (0,1]$ и $n \geq 2$, и удовлетворяет соотношениям

$$\frac{\mathrm{Sp}\, B}{\|B\|_{\scriptscriptstyle E}} = \frac{w(n-1) - t}{\sqrt{n-1}} = w\sqrt{n-1} - \frac{t}{\sqrt{n-1}}. \tag{2.42}$$

Назначим $\varepsilon = \min(3/(2\sqrt{n-1}), \delta)$. Тогда больший корень t_2 уравнения $\varepsilon\sqrt{n-1} = t^2/2 + t$ окажется в $(0,1]$, что обеспечит вещественность и знаконеопределенность матрицы $B(t_2)$. Отметим, что $w > 1 - t^2/(2n-2)$, и, полагая $t = t_2$ в (2.42), имеем:

$$\frac{\mathrm{Sp}\, B}{\|B\|_{\scriptscriptstyle E}} > \sqrt{n-1} - \frac{t_2^2}{2\sqrt{n-1}} - \frac{t_2}{\sqrt{n-1}} = \sqrt{n-1} - \varepsilon \geq \sqrt{n-1} - \delta,$$

что доказывает истинность утверждения (2.39).

Замечание 1. Аналогичный результат имеет место для отрицательно полуопределенных матриц. Тогда надо в соотношениях (2.36) — (2.39) *поставить минус перед знаком следа матрицы.*

Теорема 2.9. *Для ПО матрицы A необходимо, чтобы*

$$\mathrm{Sp}\, A > \|A\|_{\scriptscriptstyle E}, \tag{2.43}$$

и достаточно, чтобы

$$\mathrm{Sp}\, A > \sqrt{n-1}\, \|A\|_{\scriptscriptstyle E}. \tag{2.44}$$

Постоянный множитель (единица) в правой части условия (2.43) *и константа* $\sqrt{n-1}$ *в* (2.44) *точны: для любого* $\delta > 0$ *найдется положительно определенная матрица A, такая, что*

$$\mathrm{Sp}\, A \leq (1+\delta)\, \|A\|_{\scriptscriptstyle E}, \tag{2.45}$$

и существует положительно полуопределенная матрица B, такая, что

$$\mathrm{Sp}\, B = \sqrt{n-1}\, \|B\|_{\scriptscriptstyle E}. \tag{2.46}$$

Доказательство необходимости (2.43) и достаточности (2.44) близко к доказательству необходимости (2.36) и достаточности (2.37) соответственно.

Матрица $B = \mathrm{diag}(0, 1, ..., 1)$ положительно полуопределена и удовлетворяет утверждению (2.46).

Для доказательства точности (2.43) возьмем некоторую ППО матрицу A, для которой имеет место утверждение (2.38). Рассмотрим матрицу $A_t = A + It\|A\|_E$. Она положительно определена для всех положительных t. Найдем такое t, что A_t будет удовлетворять соотношению (2.45). Имеем:

$$\mathrm{Sp}\, A_t = \mathrm{Sp}\, A + nt\, \|A\|_E = \|A\|_E(1 + nt),$$

$$\|A_t\|_E \geq \|A\|_E - \sqrt{n}\, t\|A\|_E = \|A\|_E(1 - \sqrt{n}\, t).$$

Отсюда

$$\mathrm{Sp}\, A_t \leq \frac{1 + nt}{1 - \sqrt{n}t}\|A_t\|_E.$$

Следовательно для справедливости утверждения (2.45) достаточно равенства

$$1 + n\, t = (1 - \sqrt{n}t)(1 + \delta).$$

Разрешая его относительно t, имеем: $t = \delta/(n + \sqrt{n}(1 + \delta))$, что всегда положительно. Таким образом, A_t является ПО матрицей и удовлетворяет утверждению (2.45).

Замечание 2. *Аналогичный результат имеет место для отрицательно определенных матриц. Тогда надо в* (2.43) – (2.46) *поставить минус перед* Sp.

Следствие 2.1. *При* $n = 2$ *условие* (2.36) *необходимо и достаточно для положительной полуопределенности, а условие* (2.43) *необходимо и достаточно для положительной определенности.* ■

Теоремы 2.8 и 2.9 можно, пользуясь определением 2.28, сформулировать в тригонометрической форме следующим образом:

Теорема 2.10. *Для ППО матрицы A необходимо, чтобы она находилась в замкнутом конусе необходимости*:

$$\cos\,(\widehat{I\ A}) \geq 1/\sqrt{n} \tag{2.47}$$

и достаточно, чтобы она находилась в замкнутом конусе достаточности:

$$\cos(I \widehat{\ } A) \geq \sqrt{1 - 1/n}; \tag{2.48}$$

правые части соотношений (2.47) *и* (2.48) *не улучшаемы*: *первую нельзя увеличить, вторую нельзя уменьшить.*

Действительно,

$$\cos(I \widehat{\ } A) \equiv \frac{I \cdot A}{\|I\|_{E} \|A\|_{E}} = \frac{\operatorname{Sp} A}{\sqrt{n}\|A\|_{E}}, \tag{2.49}$$

что совместно с соотношениями (2.36) и (2.37) дает (2.47) и (2.48).

Теорема 2.11. *Для ПО матрицы A необходимо, чтобы она находилась в открытом конусе необходимости*:

$$\cos(I \widehat{\ } A) > 1/\sqrt{n}, \tag{2.50}$$

и достаточно, чтобы она находилась в открытом конусе достаточности:

$$\cos(I \widehat{\ } A) > \sqrt{1 - 1/n}; \tag{2.51}$$

правые части в (2.50) *и* (2.51) *нельзя улучшить.* ∎

Между углами достаточности и необходимости ППО есть связь.

$$\left.\begin{aligned} \varphi_{H} &= \arccos 1/\sqrt{n} \\ \varphi_{D} &= \arccos \sqrt{1 - 1/n} \end{aligned}\right\} \implies \varphi_{D} < \varphi_{H}.$$

Отметим, что при $n = 2$ будет $\varphi_{D} = \varphi_{H} = 45^\circ$, то есть в пространстве размерности два упрощенные критерии необходимости и достаточности совпадают.

Имеется любопытная аналогия между упрощенными критериями и глобусом. Сопоставим единичную матрицу I земной оси. Северный полярный круг – граница упрощенного критерия достаточности ППО; Тропик Рака – граница упрощенного критерия необходимости ППО; Южный полярный круг и тропик Козерога – упрощенным критериям ОПО; Арктика – ПО; Антарктика – ОО.

Рисунок разумеется весьма условен. Градусы, определяемые согласно формулам (2.48) и (2.49), ни для какого n не совпадут с градусом Северного полярного круга ($66^\circ 33' 44''$) и тропиком Рака ($23^\circ 26' 16''$). Кроме того, число параметров симметричной матрицы равно $n(n-1)/2$. При $n = 2$ — три параметра, но это вырожденный случай. При $n = 3$ — шесть параметров, отображаются вначале на пятимерную сферу шестимерного пространства, которая проектируется в двумерную плоскость рисунка. В географической интерпретации двумерная сфера трехмерного пространства проектируется в двумерную плоскость рисунка. В обоих случаях по каким бы правилам ни шло проектирование происходит утрата некоторой информации.

§2.6. АЛЬТЕРНАТИВНЫЙ ПОДХОД

к решению задачи ЛКА($\widetilde{Q}$), обремененной требованием выпуклости аппроксимации использует решение задачи о собственных числах и собственных векторах [8], [9], [11]. Задача построения выпуклой квадратичной аппроксимации ВКА($\widetilde{Q}$) вещественнозначной функции, заданной в некоторых узлах, решается в два этапа. Вначале находится безусловный квадратичный аппроксиматор (БКА), наилучший согласно некоторому критерию качества $\widetilde{Q}$, т.е. решение задачи ЛКА($\widetilde{Q}$).[7] В качестве критерия качества для ВКА берет-

[7] Формулировку этой задачи см. в §2.2.

ся уклонение ее параметров от параметров БКА. На втором этапе конечным алгоритмом находятся лучшие по этому критерию параметры ВКА.

2.6.1. Выбор критерия качества

Надо признать, что при выборе критерия качества чаще руководствуются не каким то общим пониманием адекватности аппроксимации реалиям некоей исходной задачи, но удобством использования критерия. Таков, например, критерий минимальности сумм квадратов отклонений значений $\widetilde{f}$ из (2.1) или (2.1) от наблюдаемых величин $y_1, ..., y_s$ в узлах:

$$Q_1(\check{H}) = \sum_{i=1}^{s} \left| \widetilde{f}(\check{H}, x^i) - y_i \right|^d, \quad d = 2. \tag{2.52}$$

Его применение — суть метода наименьших квадратов.

Однако, если отвлечься от легкости использования этого критерия, ничуть не хуже будут критерии с произвольной степенью $d \in (0, +\infty)$, или же критерии, использующие отклонения точек (n+1)-мерного пространства $\bar{y}_i = (y_i, x^i)$ от графика Г агрегата:

$$Q_2(\check{H}) = \sum_{i=1}^{s} \rho^d \left(\Gamma(\check{H}), \bar{y}_i \right), \quad d > 0,$$

ρ — некоторая метрика в (n+1)-мерном пространстве.

Еще более отличные от Q_1 критерии качества можно предложить, когда задача аппроксимации обременена дополнительным условием выпуклости вещественного PP-агрегата или Э-агрегата. Согласно теореме 2.3 эта выпукость соответствует ППО матрицы H, т.е. $H \in \bar{K}$, где $\bar{K}$ — конус положительно полуопределенных матриц.

Взглянем на выбор критерия с другой стороны.

Пусть $\check{H}^{\text{БЗ}}$ есть решение задачи ЛКА($\widetilde{Q}$) (без условия выпуклости). Рассмотрим критерий

$$Q_3(\check{H}) = \rho(\check{H}, \check{H}^{\text{БЗ}}). \tag{2.53}$$

Его минимизация при условии $H \in \bar{K}$ дает наименее уклоняющийся элемент от решения задачи БКА($\widetilde{Q}$). Соответствующую задачу обозначим так: ВКА($\widetilde{Q}, Q_3$).

Если критерии Q_1, Q_2 понимать как *невязки*, то Q_3 можно было бы назвать аналогом *абсолютной погрешности*. Если не известно, что лучше: малая невязка или малая абсолютная погрешность, то есть смысл минимизировать то, что легче.

Решение задачи ВКА$(\widetilde{Q}, Q_3)$ при удачном выборе метрики ρ значительно легче, чем задачи ВКА$(\widetilde{Q})$.

Введем здесь метрику на множестве матриц одних размеров естественным образом.

Определение 2.4. *Расстояние между двумя матрицами есть*

$$\rho(X, Y) := \|X - Y\|_{E} = \sqrt{(X - Y) \cdot (X - Y)}. \qquad \blacksquare$$

Тогда Q_3 из (2.53) превращается в критерий

$$Q_{E}(\check{H}) = \|\check{H} - \check{H}^{\text{БЗ}}\|_{E}.$$

Задача ВКА$(Q_1,\ Q_{E})$ на порядок проще решается конечным алгоритмом. Для его обоснования потребуются еще 4 теоремы.

2.6.2. Обоснование алгоритма

Теорема 2.12. *Введенное в опр.* 2.4 *расстояние между матрицами инвариантно относительно пары унитарных преобразований над ними*:

$$\rho(U^*XU, U^*YU) = \rho(X, Y) \qquad \forall X, Y.$$

Доказательство. Известно [13] стр. 66, что шуровская норма инвариантна относительно унитарного преобразования, поэтому из задания здесь метрики через шуровскую норму следует

$$(\forall X, Y) \qquad \rho(U^*XU, U^*YU) = \|U^*(XU - YU)\|_{E} \overset{[13]}{=} \|XU - YU\|_{E} =$$

$$= \|U^*(X^* - Y^*)\|_{E} \overset{[13]}{=} \|X^* - Y^*\|_{E} = \|X - Y\|_{E} = \rho(X, Y).$$

Определение 2.5. *Расстояние* $\rho(K, X)$ *между вектором (матрицей)* X *и множеством* K *есть* $\inf\limits_{Y \in K} \rho(Y, X)$.

Теорема 2.13. *Если K — конус ППО или конус положительно определенных матриц, то расстояние $\rho(K, A)$ между некоторой матрицей A и K инвариантно относительно пары унитарных преобразований U^*, U над матрицей A:*

$$\rho(K, A) = \rho(K, U^*AU).$$

Доказательство. Так как унитарное преобразование не меняет $\operatorname{Eig} A$, то

$$(X \in K \Rightarrow U^*XU \in K) \implies U^*KU \subseteq K. \tag{2.54}$$

Аналогично $UKU^* \subseteq K$. Умножая последнее слева на U^* и справа на U, имеем $K \subseteq U^*KU$. Совместно с (2.54) это дает $K = U^*KU$. Поэтому согласно предыдущей теореме

$$\rho(K, U^*AU) = \inf_{Y \in K} \rho(Y, U^*AU) = \inf_{Y \in U^*KU} \rho(Y, U^*AU) =$$

$$= \inf_{Z \in K} \rho(U^*ZU, U^*AU) = \rho(K, A). \qquad \blacksquare$$

Теорема 2.14. *Ближайший к диагональной матрице D элемент из конуса $\bar{K}$ положительно полуопределенных матриц существует, единствен и равен диагональной матрице D^+, получаемой из D заменой всех отрицательных элементов на нули. Иными словами, если $D = \operatorname{diag}_i(d_{ii})$, то*

$$\rho^2(K, D) = \rho^2(D^+, D) \equiv \sum_1^n (d_{ii}^+ - d_{ii})^2,$$

где $D^+ = \operatorname{diag}_i(d_{ii}^+)$, $d_{ii}^+ = d_{ii}(1 + \operatorname{sgn} d_{ii})/2$.

Доказательство. Так как $\bar{K}$ замкнут, то существует матрица $X \in \bar{K}$, которая реализует расстояние между D и $\bar{K}$. Рассмотрим диагональную подматрицу $X^d = \operatorname{diag}_i(x_{ii})$ матрицы X. Согласно необходимой части известного критерия [4], ППО матрицы X влечет $x_{ii} \geq 0$, $i = \overline{1, n}$. Согласно достаточной части того же критерия, неотрицательность элементов диагональной матрицы влечет ее ППО, следовательно $X^d \in \bar{K}$. Но

$$\|X - D\|_E^2 \equiv \sum_1^n (x_{ii} - d_{ii})^2 + \sum_{i \neq j} |x_{ij}|^2 \geq \|X^d - D\|_E^2.$$

Поскольку X является минимально удаленной от D матрицей с ППО, а X^d тоже с ППО, то здесь возможно лишь равенство. Откуда следует, что все недиагональные элементы матрицы X равны нулю, т.е. $X = X^d$. Таким образом, ближайшая к D матрица с ППО находится среди диагональных.

Поиск решения квадратичной программы

$$\sum_{1}^{n}(x_{ii} - d_{ii})^2 \longrightarrow \min_{x_{ii}\geq 0,\ i=\overline{1,n}}$$

незатруднителен. Минимум, очевидно, существует, единствен и достигается когда $x_{ii} = d_{ii}^{+}$, $i = \overline{1,n}$, где $d_{ii}^{+} = d_{ii}(1 + \operatorname{sgn} d_{ii})/2$. ∎

Первый шаг нижеследующего алгоритма отсылает к решению задачи ЛКА($\widetilde{Q}$) без условия выпуклости. Для произвольного критерия $\widetilde{Q}$ эта задача может и не иметь конечного решения. Однако, для наиболее часто используемого критерия Q из (2.27) имеет место

Теорема 2.15. *Задача ЛКА($\widetilde{Q}$) независимо от выбора узлов всегда имеет хотя бы одно конечное решение.*

Это частный случай теоремы 1.1. ∎

Элемент h_{00}, строка h_0 и столбец h_0^* в матрице $\check{H}$, очевидно, никак не сказываются на выпуклости аппроксимации $\widetilde{f}$ из (2.27). Поэтому если $\check{H}$ есть решение задачи ВКА($\widetilde{Q}$, Q_E) и $\check{H}^{\text{БЗ}}$ — решение задачи ЛКА($\widetilde{Q}$), то $h_{00} = h_{00}^{\text{БЗ}}$ и $h_0 = h_0^{\text{БЗ}}$.

Таким образом, после нахождения $\check{H}^{\text{БЗ}}$ поиск решения задачи ВКА($\widetilde{Q}$, Q_E) сводится к поиску матрицы $[n \times n]$ из $\bar{K}$, ближайшей относительно введенной метрики ρ к $H^{\text{БЗ}}$.

2.6.3. Алгоритм

Шаг 1. Найти любым способом решение $\check{H}^{\text{БЗ}}$ безусловной задачи ЛКА($\widetilde{Q}$).

Шаг 2. Преобразовать унитарным преобразованием U подматрицу $H^{\text{БЗ}}$ матрицы $\check{H}^{\text{БЗ}}$ (см. (2.4)) к диагональному виду: $D = U^*HU$.

Шаг 3. Обнулить все отрицательные элементы в D, т. е. получить матрицу D^+.

Ш а г 4. Найти $H^+ = UD^+U^*$.

Ш а г 5. Заменить в $\check{H}^{\text{БЗ}}$ подматрицу $H^{\text{БЗ}}$ на H^+. То, что получится, и есть решение задачи ВКА$(\widetilde{Q}, Q_E)$.

З а м е ч а н и е 1. Если все узлы вещественны, то подматрица $H^{БЗ}$ — вещественна, симметрична и к диагональному виду она приводится ортогональным преобразованием: $D = C^T HC$ (шаг 2), и на шаге 3: $H^+ = CD^+C^T$.

2.6.4. Единственность и узлы

Лемма 2.2. *В евклидовом пространстве с естественной метрикой $\rho(a, b) := \sqrt{(a - b) \cdot (a - b)}$ расстояние между элементом x и выпуклым замкнутым множеством $\bar{K}$ реализуется на единственном элементе $y \in \bar{K}$.*

Д о к а з а т е л ь с т в о. Если $x \in \bar{K}$, то утверждение леммы очевидно. Пусть

$$x \notin \bar{K}, \quad y, z \in \bar{K}, \quad y \neq z, \quad \|y - x\|_E = \|z - x\|_E = \rho(\bar{K}, x). \quad (2.55)$$

1) Если $y - x = \tau(z - x)$ и τ вещественно, то $|\tau| = 1$, так как $\|y - x\|_E = \|z - x\|_E$. Если $\tau = 1$, то $y = x$, что противоречит допущению (2.55). Но при $\tau = -1$ будет $x = (y + z)/2 =: w$. Но в силу выпуклости $w \in \bar{K}$, что противоречит предположению $x \notin \bar{K}$ в (2.55). Следовательно, векторы $y - x$ и $z - x$ не коллинеарны.
2) Отсутствие коллинеарности дает “<"в следующей цепи

$$\rho(w, x) = \frac{1}{2}\|y - x + z - x\|_E < \frac{1}{2}\left(\rho(y, x) + \rho(z, x)\right) = \rho(\bar{K}, x),$$

что невозможно по определению расстояния до множества.

Теорема 2.16. *Одному конечному решению задачи ЛКА$(\widetilde{Q})$ соответствует одно и только одно решение задачи ВКА$(\widetilde{Q}, Q_E)$.*

Д о к а з а т е л ь с т в о. Если конечное решение задачи ЛКА$(\widetilde{Q})$ существует, то алгоритм из 2.7.3 всегда реализуем и выдает, согласно предыдущим теоремам, одно из решений задачи ВКА$(\widetilde{Q}, Q_E)$.

Покажем, что других решений нет. Допустим противное, есть два решения задачи ВКА$(\widetilde{Q}, Q_E)$: $\check{H}$ и $\check{G}$. Как уже говорилось в п.2.7.2, они могут отличаться только своими подматрицами H и

G размером $[n \times n]$, стоящих в их правых нижних углах. Пусть задача ЛКА$(\widetilde{Q})$ имеет решение $\check{H}^{\text{бз}}$ с подматрицей $H^{\text{бз}}$. Тогда H и G суть проекции $H^{\text{бз}}$ на выпуклый конус $\bar{K}$, т.е. матрицы, которые реализуют $\rho(\bar{K}, H^{\text{бз}})$. Но, в силу предыдущей леммы, проекция на выпуклое множество единственна, следовательно, $H = G$ и $\check{H} = \check{G}$. ■

Сама задача ЛКА$(\widetilde{Q})$ при неудачном выборе узлов может иметь бесконечное множество решений, что, как правило, является неудовлетворительным. Следовательно, число узлов s должно быть не менее числа параметров $N = (n+1)(n+2)/2$ в аппроксимирующем агрегате. Понятно, однако, что этого не достаточно. Общее условие, которому должно подчиняться расположение узлов, громоздко аналитически и из него затруднительно извлечь общие рекомендации по выбору узлов. Можно, однако, предложить конкретные варианты выбора [10], обеспечивающие и единственность и низкую чувствительность решения задачи ЛКА$(\widetilde{Q})$ к погрешностям в значениях $y_1, ..., y_s$.

1. Случай $s = N$. См. формулы (2.7), (2.11) и для Э-агрегатов дополнительно (2.13), (2.14).

Отметим, что если в значениях аппроксимируемой функции есть ошибки, то параметр ε в указанных формулах должен существенно превосходить оценку этих ошибок.

2. Случай $s > N$.
Добавление произвольных дополнительных узлов к совокупности (2.7), (2.11) и для Э-агрегата еще (2.13), (2.14) может повысить чувствительность решения задачи ЛКА$(\widetilde{Q})$ к погрешностям входных данных. Так, например, происходит когда эти узлы располагаются в линейном многообразии проходящем через x^0. Напротив, если они выбирались равномерно распределенными в шаре или на сфере с центром в x^0 радиусом ε, то чувствительность решения ЛКА$(\widetilde{Q})$ к погрешностям в численных экспериментах понижалась. А если исходная функция была квадратичной, то вычисляемые параметры аппроксимации стремились к параметрам исходной функции. То же происходило и при добавлении новых узлов сериями вида (2.7), (2.11), (2.13), (2.14) с различными ε.

Библиография

[1] **Бахвалов Н. С., Жидков Н. П., Кобельков Г. Н.** Численные методы. – М., 1987. 598 с.

[2] **Демидович Б.П, Марон И.А.** основы вычислительной математики.-- М., 1960. 659 с.

[3] **Ортега Дж., Рейнболдт В.** Итерационные методы решения нелинейных систем уравнений со многими неизвестными. – М., 1975. 558 с.

[4] **Гантмахер Ф. Р.** Теория матриц. М., 1988. 552 с.

[5] **Кирин Н. Е., Сеисов Ю. Б.** Оптимизация процессов в управляемых системах. – Ашхабад, 1991. 264 с.

[6] **Котельников Е. А.** Невыпуклая квадратичная оптимизация на параллелепипеде // Сибирский журнал вычислительной математики. 11(2008), 1(весна), с.69-81.

[7] **Кюнци Г., Крелле В.** Нелинейное программирование. – М., 1965. 304 с.

[8] **Михеев С. Е., Никольский А. М.** Метод спуска в задачах аппроксимации // Процессы управления и устойчивость: Труды XXX научн. конф. – СПб., 1999. С. 119-125.

[9] **Михеев С. Е.** Выпуклая квадратичная интерполяция // Вестник С.-Петербург. ун-та. 1999. Сер. 1, вып. 2 (т 8). С. 42-48.

[10] **Михеев С. Е.** Нелинейные методы в оптимизации. – СПб.: изд. СПбГУ, 2001. 276 с.

[11] **Михеев С. Е.** Выпуклая квадратичная аппроксимация // Вычислительные технологии. – Новосибирск, 2004. Т.9, т4. С. 66–76.

[12] **Пономарев В. М., Горичев Ю. В., Городецкий В. И.** Последовательная оптимизация нелинейного закона управления // Нелинейная оптимизация систем автоматического управления / Под ред. В. М. Пономарева. – М., 1970. С. 15-19.

[13] **Уилкинсон Дж.Х.** Алгебраическая проблема собственных значений. М.: Наука, 1970. 564 с.

[14] **Beale E. M. L.** Cycling in the dual simplex algorithm, Nav. Res. Logist. Quart. 2 (1955), p.269-275.