Семинар 5

Билинейные формы

Пусть V – векторное пространство, можно думать для простоты, что $V = \mathbb{R}^n$. Тогда *билинейная форма* на V – это отображение $\beta\colon V \times V \to \mathbb{R}$ такое,[1] что

1. $\beta(v_1 + v_2, u) = \beta(v_1, u) + \beta(v_2, u)$ для всех $v_1, v_2, u \in V$.
2. $\beta(\lambda v, u) = \lambda\beta(v, u)$ для всех $v, u \in V$ и $\lambda \in \mathbb{R}$.
3. $\beta(v, u_1 + u_2) = \beta(v, u_1) + \beta(v, u_2)$ для всех $v, u_1, u_2 \in V$.
4. $\beta(v, \lambda u) = \lambda\beta(v, u)$ для всех $v, u \in V$ и $\lambda \in \mathbb{R}$.

Думать про эту процедуру надо так: у нас даны два вектора v и u из V, мы их «перемножаем» и получаем число $\beta(v, u) \in \mathbb{R}$. Самый важный пример билинейной формы – стандартное скалярное произведение: пусть даны два вектора $x, y \in \mathbb{R}^n$, зададим тогда $\beta(x, y) = x^t y = \sum_{i=1}^n x_i y_i$.

Основной план взаимодействия с билинейными формами такой. Среди всех билинейных форм мы выделим «хорошие» и назовем их скалярными произведениями. Эти товарищи будут иметь хороший геометрический смысл, с помощью которого мы определим *движения* в векторных пространствах. Но нашей конечной целью будет изучение самих движений с помощью скалярных произведений.

Как задавать билинейные формы

Пусть V – векторное пространство с базисом $e_1, \ldots, e_n$ и β – билинейная форма на V. Тогда определим числа $b_{ij} = \beta(e_i, e_j)$ – произведения базисных векторов, и составим из них матрицу $B \in \operatorname{M}_n(\mathbb{R})$. Тогда для любых векторов $v = x_1 e_1 + \ldots + x_n e_n$ и $u = y_1 e_1 + \ldots + y_n e_n$ имеем

$$\beta(v, u) = \sum_{ij} x_i y_j \beta(e_i, e_j) = \begin{pmatrix} x_1 & \ldots & x_n \end{pmatrix} B \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$$

В частности, если $V = \mathbb{R}^n$ и $x, y \in \mathbb{R}^n$, а e_i – стандартный базис. То получаем $\beta(x, y) = x^t B y$.

Таким образом и линейные операторы и билинейные формы задаются матрицами. Основная разница между ними – как эта самая матрица меняется при замене базиса. Для операторов ответы мы знаем, для билинейных форм мы сейчас займемся данным вопросом.

Смена базиса

Пусть в векторном пространстве V заданы два базиса $e_1, \ldots, e_n$ и $f_1, \ldots, f_n$ с матрицей перехода $C \in \operatorname{M}_n(\mathbb{R})$, т.е. $(f_1, \ldots, f_n) = (e_1, \ldots, e_n)C$. Пусть нам даны два вектора v и u в V. Тогда их можно разложить по базисным векторам следующим образом

$$v = x_1 e_1 + \ldots + x_n e_n = \begin{pmatrix} e_1 & \ldots & e_n \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}, \quad u = y_1 e_1 + \ldots + y_n e_n = \begin{pmatrix} e_1 & \ldots & e_n \end{pmatrix} \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$$

$$v = x'_1 f_1 + \ldots + x'_n f_n = \begin{pmatrix} f_1 & \ldots & f_n \end{pmatrix} \begin{pmatrix} x'_1 \\ \vdots \\ x'_n \end{pmatrix}, \quad u = y'_1 f_1 + \ldots + y'_n f_n = \begin{pmatrix} f_1 & \ldots & f_n \end{pmatrix} \begin{pmatrix} y'_1 \\ \vdots \\ y'_n \end{pmatrix}$$

Благодаря матрице перехода C, мы знаем, что

$$x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = C \begin{pmatrix} x'_1 \\ \vdots \\ x'_n \end{pmatrix} = Cx' \text{ и } y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = C \begin{pmatrix} y'_1 \\ \vdots \\ y'_n \end{pmatrix} = Cy'$$

[1] На самом деле можно рассматривать отображения $\beta\colon V \times U \to \mathbb{R}$, то есть можно перемножать вектора из разных пространств, но мы этого делать не будем.

Тогда в базисе e_i форма записывается в виде $\beta(v,u) = x^t By$, а в базисе f_i в виде $\beta(v,u) = (x')^t B' y'$. Но это одно и то же число посчитанное в разных базисах. Значит

$$(x')^t B' y' = x^t By = (Cx')^t BCy' = (x')^t C^T BCy'$$

для всех $x', y' \in \mathbb{R}^n$. Значит $B' = C^t BC$.[2]

Симметричность и кососимметричность

Форма $\beta\colon V \times V \to \mathbb{R}$ называется *симметричной*, если $\beta(v,u) = \beta(u,v)$ для всех $v, u \in V$. Она называется *кососимметричной*, если $\beta(v,u) = -\beta(u,v)$.

Если в координатах $\beta(x,y) = x^t By$, то $\beta(y,x) = y^t Bx$. Так как выражение $y^t Bx$ является числом, то оно не меняется при транспонировании, то есть $y^t Bx = (y^t Bx)^t = x^t B^t y$. Значит симметричность означает $x^t By = x^t B^t y$ для всех $x, y \in \mathbb{R}^n$. А это равносильно тому, что $B = B^t$. Такая матрица B называется *симметричной*. Аналогично, форма кососимметрична, тогда и только тогда, когда $B^t = -B$. В этом случае матрица B называется *кососимметричной*.

Характеристики билинейных форм

Как и выше $\beta\colon V \times V \to \mathbb{R}$ – билинейная форма. И пусть в некотором базисе она задана в виде $\beta(x,y) = x^t By$ для некоторой матрицы $B \in \mathrm{M}_n(\mathbb{R})$. Посмотрим какие характеристики матрицы B не зависят от выбора базиса.

1. Ранг матрицы B не меняется при замене $B \mapsto C^t BC$, где $C \in \mathrm{M}_n(\mathbb{R})$ – невырожденная матрица.

2. Знак определителя B не меняется при замене $B \mapsto C^t BC$, где $C \in \mathrm{M}_n(\mathbb{R})$ – невырожденная матрица. Но сам определитель меняется на $\det(C)^2$. Потому можно лишь говорить о ситуации определитель меньше нуля, больше нуля или равен нулю.

3. Обратим внимание, что невырожденность матрицы B не меняется при замене $B \mapsto C^t BC$, где $C \in \mathrm{M}_n(\mathbb{R})$ – невырожденная матрица.

4. След матрицы B вообще говоря может стать каким угодно при замене $B \mapsto C^t BC$. Потому он не несет никакой информации.

5. Симметричность и кососимметричность матрицы B не зависят от замены $B \mapsto C^t BC$.

Ядра и ортогональные дополнения

Как и выше $\beta\colon V \times V \to \mathbb{R}$ – билинейная форма. Множества ${}^\perp V = \{v \in V \mid \beta(v,V) = 0\}$ и $V^\perp = \{v \in V \mid \beta(V,v) = 0\}$ называются *левым* и *правым ядрами* формы β. Эти подмножества являются подпространствами в V. Если в координатах форма задана $\beta(x,y) = x^t By$, то ${}^\perp\mathbb{R}^n = \{x \in \mathbb{R}^n \mid x^t B = 0\}$ и $(\mathbb{R}^n)^\perp = \{y \in \mathbb{R}^n \mid By = 0\}$. В частности отсюда видно, что ядра имеют одинаковую размерность равную $n - \operatorname{rk} B$. Если форма симметричная или кососимметричная, то нет разницы между правыми и левыми ядрами. Ядра – это неинтересная часть пространства, которая «ортогональна» всему относительно этой формы.

Более обще, пусть $U \subseteq V$ – подпространство в V. Тогда его *левым ортогональным дополнением* является подпространство ${}^\perp U = \{v \in V \mid \beta(v,U) = 0\}$. Аналогично, *правое ортогональное дополнение* это $U^\perp = \{v \in V \mid \beta(U,v) = 0\}$. Если в координатах форма задана $\beta(x,y) = x^t By$ и $U = \langle u_1, \ldots, u_k \rangle$. Пусть D – матрица составленная из столбцов u_i. Тогда ${}^\perp U = \{x \in \mathbb{R}^n \mid x^t BD = 0\}$ и $U^\perp = \{y \in \mathbb{R}^n \mid D^t By = 0\}$. Обычно ортогональные дополнения и ядра интересны в случае симметрических или кососимметрических форм, так как в этом случае левые и правые ортогональные дополнения равны между собой.

Двойственность для подпространств

Напомню, что для двух подпространств $U, W \subseteq V$ определены их сумма $U + W = \{u + w \mid u \in U, w \in W\}$ и пересечение $U \cap W$. Сумма – это наименьшее подпространство, которое содержит U и W одновременно

[2] Напомним, что матрица A линейного оператора $\phi\colon V \to V$ меняется по правилу $A' = C^{-1}AC$.

(объединение не является подпространством вообще говоря). А пересечение – это наибольшее подпространство, которое лежит и в том и в том. Про них надо думать, как бы как про НОК и НОД подпространств, соответственно.

В общем случае нет хорошей связи между подпространством и его ортогональным (левым или правым) дополнением. Например, если билинейная форма нулевая, то есть $\beta\colon V \times V \to \mathbb{R}$ все переводит в ноль (это соответствует нулевой матрице в любом базисе), то любые два подпространства ортогональны, и ортогональным дополнением к чему угодно будет все пространство V. Однако для хороших билинейных форм можно доказать теорему о двойственности на подпространствах.

Утверждение. *Пусть $\beta\colon V \times V \to \mathbb{R}$ – невырожденная билинейная форма (то есть ее матрица не вырождена). Тогда:*

1. *Для любого подпространства $W \subseteq V$ выполнено*
$$\dim W^{\perp} + \dim W = \dim V$$
2. *Для любого подпространства $W \subseteq V$ выполнено ${}^{\perp}(W^{\perp}) = W$.*
3. *Для любых подпространств $W \subseteq E \subseteq V$ верно, что $W^{\perp} \supseteq E^{\perp}$. Причем $W = E$ тогда и только тогда, когда $W^{\perp} = E^{\perp}$.*
4. *Для любых подпространств $W, E \subseteq V$ выполнено равенство*
$$(W + E)^{\perp} = W^{\perp} \cap E^{\perp}$$
5. *Для любых подпространств $W, E \subseteq V$ выполнено равенство*
$$(W \cap E)^{\perp} = W^{\perp} + E^{\perp}$$

Аналогично выполнены все свойства для подпространств $W \subseteq U$ и их левых ортогональных дополнений ${}^{\perp}W$.

Таким образом, мы как бы переворачиваем подпространства вверх ногами под действием операции взятия ортогонального дополнения. Операции взятия левого и правого ортогонального дополнения становятся взаимнообратными. При этом эти операции большие подпространства переводят в маленькие, обращают вложения и меняют НОК и НОД подпространств местами.

Симметричные формы

Утверждение. *Пусть $\beta\colon \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ – симметрическая билинейная форма заданная $\beta(x, y) = x^t B y$. Тогда существует такой базис, что матрица B диагональная и на диагонали стоят либо 1, либо -1, либо 0, т.е. блочно имеет вид $B' = \begin{pmatrix} E & & \\ & -E & \\ & & 0 \end{pmatrix}$. При этом количество единиц и минус единиц на диагонали не зависит от базиса.*

На это утверждение еще можно смотреть так. Для любой симметрической матрицы $B \in \mathrm{M}_n(\mathbb{R})$ можно найти такую невырожденную матрицу $C \in \mathrm{M}_n(\mathbb{R})$, что матрица $C^t B C$ имеет описанный диагональный вид.

Суммарно количество единиц и минус единиц дает ранг матрицы B, то есть ранг билинейной формы. Тот факт, что количество единиц и минус единиц является инвариантом формы надо понимать так: у нас ранг как бы складывается из положительной и отрицательной части и размеры этих частей определены однозначно.

Количество единиц $\#1$ в таком виде называется положительным индексом инерции формы, количество минус единиц $\# -1$ – отрицательным индексом, а количество нулей $\#0$ – нулевым индексом. Вместе набор чисел $(\#1, \# -1, \#0)$ называется сигнатурой формы.

Определение сигнатуры формы

Для определения сигнатуры билинейной формы можно воспользоваться разными методами. Самый простой – Симметричный Гаусс. Мы диагонализуем форму и считаем количество положительных, отрицательных и нулевых элементов на диагонали. Этот метод работает всегда.

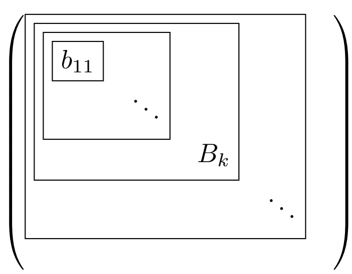
Симметрический Гаусс Теперь, когда мы знаем, что симметрические билинейные формы диагонализуются в каком-то базисе, хорошо было бы иметь какой-нибудь (ну хотя бы плохонький) алгоритм, приводящий форму к диагональному виду, если она задана в каком-то случайном базисе. Пусть, скажем, нам задана билинейная форма $\beta\colon F^n \times F^n \to F$ по правилу $(x, y) \mapsto x^t By$, где $B \in \mathrm{M}_n(F)$ – некоторая симметричная матрица. Тогда в новом базисе матрица будет иметь вид C^tBC, где $C \in \mathrm{M}_n(F)$ – некоторая невырожденная матрица.[3] Любая невырожденная матрица C раскладывается в произведение элементарных матриц. С другой стороны, если C – матрица элементарного преобразования, то $B \mapsto C^tBC$ – это выполнение одного и того же преобразования и над строками и над столбцами (не важно в каком порядке, так как произведение матриц ассоциативно). То есть у нас есть следующий запас операций:

- Прибавляем i-ю строку умноженную на λ к j-ой строке, потом прибавляем i-ый столбец умноженный на λ к j-ому столбцу.
- Меняем местами i и j строки, после чего меняем местами i и j столбцы.
- Умножаем на ненулевое λ i-ю строку, потом умножаем на λ i-ый столбец.

Таким образом предыдущая теорема гласит, что выполняя подобные симметричные элементарные преобразования над симметрической матрицей, мы обязательно приведем ее к диагональному виду. Если при этом надо восстановить матрицу C, то рассматриваем $(B|E)$ и делаем симметричные элементарные преобразования над ней в том смысле, что преобразования над строками выполняются над всей матрицей, а преобразования над столбцами только над часть, где лежит B. Тогда матрица приведется к виду $(B'|C^t)$.

Метод Якоби Также для определения сигнатуры формы используется метод Якоби. Этот метод работает почти всегда и я поясню, что это значит и что делать, когда он не работает. Но прежде всего я хочу обратить внимание, что у него есть ограничения на входные данные. Матрица B обязательно должна быть невырождена. Это в частности означает, что метод работает только для форм у которых в сигнатуре только единицы и минус единицы и совсем нет нулей.

Пусть $B \in \mathrm{M}_n(\mathbb{R})$ – симметричная невырожденная матрица и $\beta(x, y) = x^tBy$. Выделим в матрице B верхние левые блоки:

$B =$ 

То есть B_k – подматрица состоящая из первых k строк и столбцов. Теперь определим числа $\Delta_k = \det(B_k)$, которые называются угловыми минорами. Если так получилось, что все числа Δ_k НЕ равны нулю[4], то мы строим последовательность

$$\Delta_1, \frac{\Delta_2}{\Delta_1}, \frac{\Delta_3}{\Delta_2}, \ldots, \frac{\Delta_n}{\Delta_{n-1}}$$

Тогда положительный индекс инерции для B равен количеству положительных чисел в этой последовательности, а отрицательный индекс инерции равен количеству отрицательных чисел в этой последовательности.

Что делать, если встретились нули По-хорошему надо пользоваться симметричным Гауссом. Но если все таки хочется воспользоваться именно методом Якоби, то надо чуть-чуть пошевелить матрицу B правильным образом. Мы можем сгенерировать случайную матрицу C. Она с вероятностью один будет невырожденной. Потом надо рассмотреть матрицу $B' = C^tBC$ и применить метод Якоби к матрице B' вместо B. Сделаю одно замечание по организации вычислений. В этом методе надо генерировать случайную матрицу C и НЕ проверять ее на невырожденность. Вместо этого, надо сразу применить метод Якоби к матрице B'. Если все

[3] На самом деле C – матрица перехода из старого в новый базис.

[4] Матрицу B с таким условием можно разложить в виде $B = LU$, где L – нижнетреугольная матрица, U – верхнетреугольная матрица с 1 на диагонали. Это называется LU разложением. Для этого надо применить Гаусса к B вычитая из более верхних строчек более низкие. Тогда мы приведем B к верхнетреугольному виду. Отсюда можно вытащить LU разложение стандартным рассуждением.

Δ_k оказались не нулевые, то нам повезло и метод и так сработал (матрица C в этом случае автоматически окажется невырожденной). А если не повезло, то нам все равно надо будет генерировать новую матрицу C и не важно какой она была.

Продвинутый метод определения сигнатуры Пусть $B \in \mathrm{M}_n(\mathbb{R})$ – симметричная матрица и $\beta(x, y) = x^t By$. Тогда найдем спектр матрицы B с кратностями. В случае симметрической матрицы окажется, что спектр будет обязательно вещественным. Тогда количество положительных чисел в спектре с учетом кратности равно положительному индексу инерции, количество отрицательных чисел в спектре с кратностью равно отрицательному индексу инерции, а количество нулей – нулевому индексу. Надо понимать, что сам спектр не является корректно определенной величиной для билинейной формы, он может измениться кардинальной при смене базиса, но знаки собственных значений, оказывается, не изменятся.

Последнее свойство дает способ оценить количество собственных значений у матрицы на отрезке. Например у матрицы $A - \lambda E$ спектр состоит из $\lambda_i - \lambda$, где λ_i – собственные значения A. Тогда количество положительных собственных значений у $A - \lambda E$ можно определить симметричным Гауссом, и оно совпадает с количеством собственных значений A строго больших λ. Подбирая λ и считая количество положительных или отрицательных корней, мы можем определить количество собственных значений на заданном отрезке.

Квадратичные формы

Если нам дана какая-то билинейная форма $\beta\colon V \times V \to \mathbb{R}$, то отображение $Q\colon V \to \mathbb{R}$ вида $Q(x) = \beta(x, x)$ называется квадратичной формой. Если векторное пространство $V = \mathbb{R}^n$, то билинейная форма превращается в $\beta(x, y) = x^t By$, а соответствующая квадратичная форма в $Q(x) = x^t Bx$. Если расписать явно последнее выражение, то мы получим

$$Q(x) = \beta(x, x) = x^t Bx = \sum_{ij} b_{ij} x_i x_j = \sum_i b_{ii} x_i^2 + \sum_{i<j} (b_{ij} + b_{ji}) x_i x_j$$

Обратите внимание, что в отличие от билинейной формы, квадратичная форма не однозначно задается матрицей B. Действительно,

$$Q(x_1, x_2) = \begin{pmatrix} x_1 & x_2 \end{pmatrix} \begin{pmatrix} 0 & 2 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} x_1 & x_2 \end{pmatrix} \begin{pmatrix} 0 & 0 \\ 2 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} x_1 & x_2 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = 2x_1 x_2$$

За счет этого эффекта, при переходе к квадратичным формам от билинейных, мы теряем часть информации. Однако, квадратичная форма однозначно задается симметрической матрицей B, то есть матрицей B с условием $B^t = B$. В примере выше – это последний случай.

Для полноты картины добавлю, что в случае симметричной матрицы B или что то же самое симметричной билинейной формы β, мы можем вернуться от квадратичной формы к билинейной с помощью так называемой поляризационной формулы, а именно

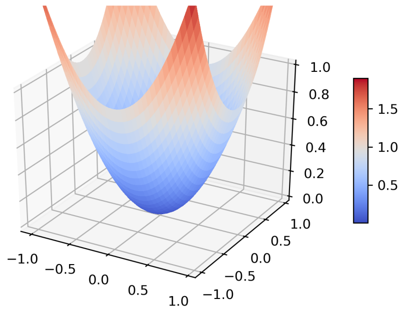
$$\beta(x, y) = \frac{Q(x + y) - Q(x) - Q(y)}{2}$$

Идейно это означает, что изучать симметричные билинейные формы – это то же самое, что изучать квадратичные формы. Но у квадратичных форм есть красивый геометрический смысл. Его мы и обсудим далее.

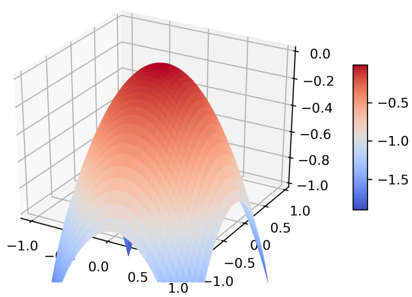
Графики квадратичных форм

Пусть $V = \mathbb{R}^2$. Тогда квадратичная форма $Q(x, y)$ задает функцию от двух переменных, а именно $z = Q(x, y)$. Давайте нарисуем ее графики в некоторых частных случаях.

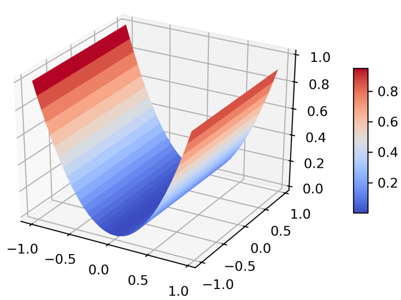
1. $z = x^2 + y^2$. Начало координат – точка минимума. Матричная запись $z = Q(x, y) = \begin{pmatrix} x \\ y \end{pmatrix}^t \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}$



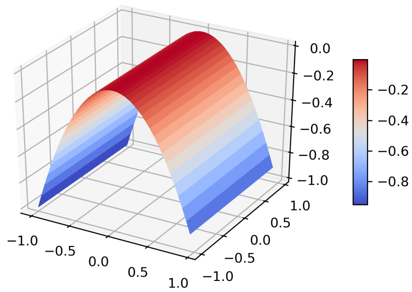
2. $z = -x^2 - y^2$. Начало координат – точка максимума. Матричная запись $z = Q(x,y) = \begin{pmatrix} x \\ y \end{pmatrix}^t \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}$



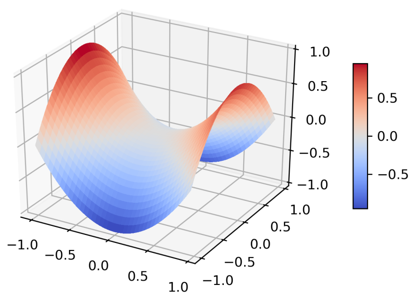
3. $z = x^2$. Минимум достигается на прямой $x = 0$. Матричная запись $z = Q(x,y) = \begin{pmatrix} x \\ y \end{pmatrix}^t \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}$



4. $z = -x^2$. Максимум достигается на прямой $x = 0$. Матричная запись $z = Q(x,y) = \begin{pmatrix} x \\ y \end{pmatrix}^t \begin{pmatrix} -1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}$



5. $z = x^2 - y^2$. Начало координат – седловая точка. Матричная запись $z = Q(x, y) = \begin{pmatrix} x \\ y \end{pmatrix}^t \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}$



Обратите внимание, что поведение графика зависит от знаков чисел на диагонали матрицы два на два. В общем случае поведение графика зависит от сигнатуры формы.

Классификация билинейных и квадратичных форм Ниже я все определения отразил в единой табличке. В ней подразумевается, что билинейная форма задана на пространстве размерности n.

Термин	**Обозначения**	**Условие**	**Индексы**
Положительная	$\beta > 0$ или $Q > 0$	$\forall x \neq 0 \Rightarrow Q(x) > 0$	$\#1 = n$
Отрицательная	$\beta < 0$ или $Q < 0$	$\forall x \neq 0 \Rightarrow Q(x) < 0$	$\# - 1 = n$
Неотрицательная	$\beta \geqslant 0$ или $Q \geqslant 0$	$\forall x \Rightarrow Q(x) \geqslant 0$	$\# - 1 = 0$
Неположительная	$\beta \leqslant 0$ или $Q \leqslant 0$	$\forall x \Rightarrow Q(x) \leqslant 0$	$\#1 = 0$
Неопределенная		$\exists x, y \Rightarrow Q(x) > 0$ и $Q(y) < 0$	$\#1 > 0$ и $\# - 1 > 0$

Скалярные произведения

Билинейная форма $\beta\colon V \times V \to \mathbb{R}$ называется *скалярным произведением*, если она

1. симметрична $\beta(v, u) = \beta(u, v)$.
2. *положительно определена*, т.е. для любого ненулевого вектора $v \in V$ имеем $\beta(v, v) > 0$.

В этом случае пишут (v, u) вместо $\beta(v, u)$. Самый важный пример – стандартное скалярное произведение: $(x, y) = x^t y$, где $x, y \in \mathbb{R}^n$. Векторное пространство, в котором зафиксировано какое-либо скалярное произведение называется *Евклидовым пространством*.

По определению скалярного произведения у него в сигнатуре присутствуют только единицы, а минус единиц и нулей нет. В частности это означает, что матрица скалярного произведения всегда невырождена. Кроме того это еще означает, что для любого скалярного произведения существует такой базис, что в нем матрица B становится единичной матрицей. По-другому, на этот факт можно смотреть так: какие-бы два евклидовых пространства одинаковой размерности вы ни взяли бы, они оказываются одинаковыми (формально изоморфными).

Экзотические скалярные произведения

Ради интереса вот два примера любопытных скалярных произведений в неожиданных ситуациях.

1. Пусть в качестве векторного пространства у нас будет пространство матриц $V = \mathrm{M}_{m\,n}(\mathbb{R})$ не обязательно квадратных. Пусть $A, B \in \mathrm{M}_{m\,n}(\mathbb{R})$, зададим скалярное произведение следующим образом $(A, B) = \operatorname{tr}(A^t B)$.[5] Можно руками проверить, что $(A, B) = (B, A)$ и что $(A, A) = \sum_{ij} a_{ij}^2 > 0$, если $A \neq 0$. А потому эта штука удовлетворяет свойствам скалярного произведения. Значит можно в пространстве матриц мерить длины матриц и углы между ними. Длина матрицы в этом случае будет $|A|_F = \sqrt{\operatorname{tr}(A^t A)}$ и называется «нормой Фробениуса» матрицы A.

2. Пусть теперь в качестве векторного пространства у нас будет множество всех непрерывных функций на отрезке $[0, 1]$, то есть $V = C[0, 1] = \{f\colon [0, 1] \to \mathbb{R} \mid f \text{ непрерывна}\}$. Тогда для двух функций $f, g \in C[0, 1]$ определим скалярное произведение следующим образом $(f, g) = \int_0^1 f(x)g(x)\,dx$. Получается, что теперь можно мерить длины функций и углы между функциями. Например, длина функции f будет $\sqrt{\int_0^1 f(x)^2\,dx}$.

Углы и расстояния

Пусть V – евклидово пространство. Тогда *длина* вектора v это $|v| = \sqrt{(v, v)}$. Если $v, u \in V$ – два вектора, то определим *угол* $\alpha_{v,u}$ между этими векторами из равенства $\cos \alpha_{v,u} = \frac{(v,u)}{|v||u|}$.

Два вектора v и u называются *ортогональными*, если $(v, u) = 0$, т.е. угол между векторами 90°. Базис $e_1, \ldots, e_n$ называется *ортогональным*, если любая пара векторов из базиса ортогональна, т.е. $(e_i, e_j) = 0$ при $i \neq j$. Базис называется *ортонормированным*, если он ортогонален и все вектора имеют длину 1, т.е. $(e_i, e_j) = 0$ при $i \neq j$ и $(e_i, e_i) = 1$. По определению матрицы билинейной формы $b_{ij} = (e_i, e_j)$, а значит в ортонормированном базисе скалярное произведение имеет вид $(x, y) = x^t y$.

Для любого множества векторов $S \subseteq \mathbb{R}^n$ мы можем определить ортогональное дополнение

$$S^\perp = \{v \in V \mod (v, s) = 0 \text{ для любого } s \in S\}$$

То есть это все возможные вектора, которые ортогональны всем векторам из S. Обратите внимание, что S может быть любым множеством, например, может состоять из одного вектора, но при этом $S^\perp$ всегда будет векторным подпространством в $\mathbb{R}^n$.

Расстоянием между двумя векторами v и u пространства $\mathbb{R}^n$ называется $\rho(v, u) = |v - u|$. Если мы хотим найти расстояние между двумя подмножествами, например, $X, Y \subseteq V$, то по определению расстояние между ними – это наименьшее расстояние между всеми парами точек из них, то есть

$$\rho(X, Y) = \inf_{\substack{x \in X\\ y \in Y}} \rho(x, y)$$

В частности можно говорить о расстоянии от вектора до подпространства. Напомню, что если $U \subseteq \mathbb{R}^n$ – некоторое подпространство и $U^\perp$ – его ортогональное дополнение, то любой вектор однозначно раскладывается в сумму вектора из U и вектора из $U^\perp$, то есть любой $v \in V$ имеет вид $v = u + w$, где $u \in U$ и $w \in U^\perp$. Тогда вектор u называется ортогональной проекцией v на U, а вектор w называется ортогональной составляющей v относительно U.

Утверждение. *Пусть $V = \mathbb{R}^n$ – евклидово пространство, $U \subseteq V$ – подпространство, $v \in V$ – некоторый вектор. Тогда расстояние от v до U равно длине ортогональной составляющей v относительно U.*

Пример Пусть $V = \mathbb{R}^n$ и задано стандартное скалярное произведение $(x, y) = x^t y$. Пусть U – подпространство заданное системой следующего вида: $U = \{y \in \mathbb{R}^n \mid (a_1, \ldots, a_n)y = 0\}$ (мы считаем, что хотя бы одно из чисел $a_i \neq 0$). Если положим $v = (a_1, \ldots, a_n)^t$. То $U = \langle v\rangle^\perp$ по определению. То есть оно задается в виде $\{y \in \mathbb{R}^n \mid (y, v) = 0\}$ для некоторого фиксированного ненулевого вектора v. Заметим, что это подпространство делит все пространство на два класса:

1. Положительные векторы $\{y \in \mathbb{R}^n \mid (y, v) > 0\}$.

[5] Обратите внимание, что матрица $A^t B$ будет квадратной размера n на n, а потому для нее корректно определено понятие следа.

2. Отрицательные векторы $\{y \in \mathbb{R}^n \mid (y, v) < 0\}$.

Между этими классами как раз и расположено подпространство $\langle v\rangle^\perp$.

Пусть $w \in V$ – произвольный вектор. Давайте поймем как найти расстояние от w до подпространства U. Вспомним, что $w = \alpha v + u$, где $u \in U$, а вектор αv будет ортогональной составляющей v относительно U. Значит его длина и будет расстоянием. Надо лишь найти неизвестное α. Для этого умножим равенство $w = \alpha v + u$ скалярно на v и получим $(w, v) = \alpha(v, v)$. Откуда $\alpha = \frac{(w,v)}{(v,v)}$. Значит

$$\rho(w, U) = \left|\frac{(w, v)}{(v, v)} v\right| = \frac{|(w, v)|}{|v|} = |(w, v/|v|)|$$

Если вектор v имел единичную длину, то формула упрощается так

$$\rho(w, \langle v\rangle^\perp) = |(w, v)| = |w^t v|$$[6]

Двойственность для подпространств

Если у нас задано скалярное произведение, то оно является невырожденной и симметричной билинейной формой. А значит можно применить утверждение о двойственности для подпространств. При этом ситуация сильно упрощается, так как нет разницы между левыми и правыми ортогональными дополнениями. Давайте я явно проговорю, что получится в этом случае.

Утверждение. *Пусть $V = \mathbb{R}^n$ – евклидово пространство. Тогда:*

1. *Для любого подпространства $W \subseteq V$ выполнено*
$$\dim W^\perp + \dim W = \dim V$$
2. *Для любого подпространства $W \subseteq V$, $V = W \oplus W^\perp$.*
3. *Для любого подпространства $W \subseteq V$ выполнено $(W^\perp)^\perp = W$.*
4. *Для любых подпространств $W \subseteq E \subseteq V$ верно, что $W^\perp \supseteq E^\perp$. Причем $W = E$ тогда и только тогда, когда $W^\perp = E^\perp$.*
5. *Для любых подпространств $W, E \subseteq V$ выполнено равенство*
$$(W + E)^\perp = W^\perp \cap E^\perp$$
6. *Для любых подпространств $W, E \subseteq V$ выполнено равенство*
$$(W \cap E)^\perp = W^\perp + E^\perp$$

Отртогонализация Грама-Шмидта

Дано Множество векторов $v_1, \ldots, v_k \in \mathbb{R}^n$.

Задача Найти множество $u_1, \ldots, u_s$ такое, что u_i попарно ортогональны и $\langle v_1, \ldots, v_k\rangle = \langle u_1, \ldots, u_s\rangle$.

Алгоритм

1. Берем первый ненулевой вектор среди v_i. Пусть это будет v_1. Тогда полагаем $u_1 = v_1$.
2. Рассмотрим $v_2 - \frac{(v_2,u_1)}{(u_1,u_1)} u_1$. Если этот вектор не ноль, то обозначим его за u_2. Если ноль, то выкинем v_2 и перенумеруем вектора так, что v_3 теперь будет вектором v_2. Повторяем этот шаг до тех пор, пока не найдем u_2 или пока не закончатся вектора v_i.
3. Рассмотрим $v_3 - \frac{(v_3,u_1)}{(u_1,u_1)} u_1 - \frac{(v_3,u_2)}{(u_2,u_2)} u_2$. Если он не ноль, то обозначим его за u_3. Иначе как и в предыдущем пункте переходим к следующему вектору и повторяем этот шаг.
4. Для поиска u_i надо рассмотреть вектор $v_i - \frac{(v_i,u_1)}{(u_1,u_1)} u_1 - \ldots - \frac{(v_i,u_{i-1})}{(u_{i-1},u_{i-1})} u_{i-1}$. Аналогично предыдущему пункту, если этот вектор не ноль, то это u_i. Если ноль, то рассматриваем следующий v_{i+1} вместо него и повторяем этот шаг.

[6] Последнее равенство в силу того, что у нас стандартное скалярное произведение.

Пример Пусть у нас заданы векторы

$$v_1 = \begin{pmatrix}1\\1\\1\\1\end{pmatrix}, \quad v_2 = \begin{pmatrix}3\\3\\1\\1\end{pmatrix}, \quad v_3 = \begin{pmatrix}2\\2\\1\\1\end{pmatrix} \text{ и } v_4 = \begin{pmatrix}2\\0\\1\\-1\end{pmatrix} \in \mathbb{R}^4$$

Первый вектор не ноль, значит $u_1 = v_1$. Теперь рассмотрим

$$v_2 - \frac{(v_2, u_1)}{(u_1, u_1)} u_1 = \begin{pmatrix}3\\3\\1\\1\end{pmatrix} - \frac{3+3+1+1}{1+1+1+1}\begin{pmatrix}1\\1\\1\\1\end{pmatrix} = \begin{pmatrix}1\\1\\-1\\-1\end{pmatrix}$$

Значит $u_2 = v_2$. Теперь рассмотрим

$$v_3 - \frac{(v_3, u_1)}{(u_1, u_1)} u_1 - \frac{(v_3, u_2)}{(u_2, u_2)} u_2 = \begin{pmatrix}2\\2\\1\\1\end{pmatrix} - \frac{2+2+1+1}{1+1+1+1}\begin{pmatrix}1\\1\\1\\1\end{pmatrix} - \frac{2+2-1-1}{1+1+1+1}\begin{pmatrix}1\\1\\-1\\-1\end{pmatrix} = 0$$

Значит забываем про v_3 и переходим к следующему вектору.

$$v_4 - \frac{(v_4, u_1)}{(u_1, u_1)} u_1 - \frac{(v_4, u_2)}{(u_2, u_2)} u_2 = \begin{pmatrix}2\\0\\1\\-1\end{pmatrix} - \frac{2+1-1}{1+1+1+1}\begin{pmatrix}1\\1\\1\\1\end{pmatrix} - \frac{2-1+1}{1+1+1+1}\begin{pmatrix}1\\1\\-1\\-1\end{pmatrix} = \begin{pmatrix}1\\-1\\1\\-1\end{pmatrix}$$

Таким образом ответ

$$u_1 = \begin{pmatrix}1\\1\\1\\1\end{pmatrix}, \quad u_2 = \begin{pmatrix}1\\1\\-1\\-1\end{pmatrix} \text{ и } u_3 = \begin{pmatrix}1\\-1\\1\\-1\end{pmatrix}$$

Ортогональные матрицы

Утверждение. *Для матрицы $A \in \mathrm{M}_n(\mathbb{R})$ следующие условия эквивалентны:*

1. $A^t A = E$.
2. $A A^t = E$.
3. $A^t = A^{-1}$.

Матрица обладающая одним из этих эквивалентных условий называется *ортогональная*. Таким образом в ортонормированном базисе движение задается ортогональной матрицей.

QR-разложение Пусть у нас дан набор векторов $v_1, \ldots, v_k \in \mathbb{R}^n$ и задано стандартное скалярное произведение $(x, y) = x^t y$. Давайте для начала для простоты будем считать, что все векторы линейно независимы. Проведем для них процесс ортогонализации Грама-Шмидта и получим векторы $v_1', \ldots, v_k'$. Если в формулах для Грама-Шмидта перенести все старые векторы влево, а новые вправо, то получится следующий набор равенств:

$$\begin{pmatrix} v_1 & \ldots & v_k \end{pmatrix} = \begin{pmatrix} v_1' & \ldots & v_k' \end{pmatrix} \begin{pmatrix} 1 & \frac{(v_2, v_1')}{(v_1', v_1')} & \frac{(v_3, v_1')}{(v_1', v_1')} & \ldots & \frac{(v_k, v_1')}{(v_1', v_1')} \\ & 1 & \frac{(v_3, v_2')}{(v_2', v_2')} & \ldots & \frac{(v_k, v_2')}{(v_2', v_2')} \\ & & 1 & & \\ & & & \ddots & \vdots \\ & & & & 1 \end{pmatrix}$$

При этом векторы $v_1', \dots, v_k'$ являются ортогональными. В силу предположенной линейной независимости исходных векторов новые векторы тоже будут линейно независимыми, а в частности не нулевыми. Если теперь поделить каждый из них на его длину, то получится

$$\begin{pmatrix} v_1 & \dots & v_k \end{pmatrix} = \begin{pmatrix} \frac{v_1'}{|v_1'|} & \dots & \frac{v_k'}{|v_k'|} \end{pmatrix} \begin{pmatrix} |v_1'| & \frac{(v_2, v_1')}{|v_1'|} & \frac{(v_3, v_1')}{|v_1'|} & \dots & \frac{(v_k, v_1')}{|v_1'|} \\ & |v_2'| & \frac{(v_3, v_2')}{|v_2'|} & \dots & \frac{(v_k, v_2')}{|v_2'|} \\ & & |v_3'| & & \\ & & & \ddots & \vdots \\ & & & & |v_k'| \end{pmatrix}$$

Пусть теперь $A = (v_1| \dots |v_k) \in \mathrm{M}_{n\,k}(\mathbb{R})$ – матрица составленная из исходных векторов, а $Q' = \left(\frac{v_1'}{|v_1'|}|\dots|\frac{v_k'}{|v_k'|}\right)$, а $R' \in \mathrm{M}_k(\mathbb{R})$ – матрица из коэффициентов справа в верхнем равенстве. Тогда верхнее равенство записывается так $A = Q'R'$.Мы можем добавить столбцы в матрицу Q' в конце так, чтобы получилась ортогональная матрица Q, тогда $A = Q(R'|0)$. Таким образом A представлена в виде произведения ортогональной матрицы Q и верхнетреугольной матрицы $R = (R'|0)$. Такое разложение называется QR разложением.

Если же у исходной матрицы A столбцы оказались линейно зависимыми, то в результате нашего алгоритма, Q' может содержать нулевые столбцы, но все они будут ортогональны. В этом случае поступают так. Для каждого нулевого столбца Q' зануляют соответствующую строку в матрице R'. После этого все равно, что стоит в соответствующем столбце Q'. В этом случае в место нулевых столбцов вставляют недостающие векторы до ортонормированного базиса. А последний шаг перехода от Q' и R' к Q и R абсолютно такой же.

Линейные многообразия

По-простому, подпространства – это «линейные поверхности» проходящие через начало координат. А если мы их хотим сдвинуть, то такие штуки называются линейными многообразиями. А именно, пусть $V = \mathbb{R}^n$, $v \in V$ – некоторый вектор и $U \subseteq V$ – подпространство. Тогда множество $v+U$ называется линейным многообразием.

Линейные многообразия в $\mathbb{R}^n$

Любое линейное многообразие в $\mathbb{R}^n$ можно задать двумя способами:

1. $L = v + \langle v_1, \dots, v_k \rangle$, где $v_1, \dots, v_k$ можно выбрать линейно независимыми.
2. $L = \{y \in \mathbb{R}^n \mid Ay = b\}$, где $A \in \mathrm{M}_{m\,n}(\mathbb{R})$, $b \in \mathbb{R}^m$. Кроме того, строки матрицы A можно выбрать линейно независимыми.

Таким образом линейные многообразия возникают, когда мы решаем неоднородные линейные системы, а по своей сути они всего лишь сдвиги подпространств на какой-то вектор.

Примеры

1. Прямая. Пусть $p, v \in V$ – некоторые векторы, причем $v \neq 0$. Тогда $p + \langle v \rangle$ – это сдвиг одномерного подпространства. Такое подпространство называется прямой. Явно на такой прямой лежат вектора вида $p + \lambda v$, где $\lambda \in \mathbb{R}$.

2. Гиперповерхность. Пусть $p \in V$ и $U \subseteq V$ подпространство размерности на 1 меньше, чем V. Тогда $p+U$ называется гиперповерхностью. Как было показано выше, $U = \langle v \rangle^\perp$ для некоторого ненулевого вектора v. В этом случае $p+U = \{y \in V \mid (y, v) = (p, v)\}$. Кроме этого мы можем разделить все пространство на два класса:

   (a) Положительные векторы $\{y \in V \mid (y, v) > (p, v)\}$.
   (b) Отрицательные векторы $\{y \in V \mid (y, v) < (p, v)\}$.

   Гиперповерхность является простейшим примером линейного классификатора. Если точки находятся по одну сторону от нее, то они относятся к одному класса, а по другую – к другому.

   Пусть $w \in V$ – произвольный вектор. Как посчитать расстояние от w до $p + U$. Для этого надо сдвинуть начало координат в точку p и посчитать расстояние до подпространства U. Из вычислений ранее получаем

   $$\rho(w, p+U) = \rho(w-p, U) = \frac{|(w-p, v)|}{|v|}$$