

Векторная модель Е.Н. Соколова: Формализм и возможные обобщения

Фундаментальная *Векторная модель* Евгения Николаевича Соколова с соавторами была разработана более сорока лет назад, но заложенные в ней принципы до сих пор широко используются при моделировании процессов в нейронных системах и психофизике. Имеется огромное количество работ, авторы которых используют терминологию, в той или иной степени близкую применяемой Е.Н. Соколовым и его последователями. Между тем, логика дальнейшего развития моделей обработки сенсорной информации требует уточнения и более строгого определения ряда фундаментальных для этой теории понятий. При этом основные монографии (Фомин и др., 1979) и (Соколов и др., 1989), в которых сформулированы формальные допущения теории, давно стали библиографической редкостью.

К сожалению, формальная часть теории изложена в работах школы Е.Н. Соколова очень кратко, без детального математического обоснования. Вероятно, это можно объяснить безвременной кончиной в 1975 году одного из создателей теории, известного математика С.В. Фомина. Как нам представляется, аккуратное аксиоматическое изложение *Векторной модели* необходимо как для уточнения основных понятий, так и для дальнейших разработок и развития теории. В данной работе мы попытаемся определить систему базовых постулатов *Векторной модели*, и дать ее формальное изложение и дальнейшее обобщение, используя аппарат функционального (Колмогоров и др., 2004) и выпуклого (Рокафеллар, 1973) анализа. При этом изложение существенно опирается на результаты одного из авторов (Дубровский, 2009).

Используемая Е.Н. Соколовым с соавторами система обозначений во многом отличается от общепринятой. Тем не менее, мы также будем ей пользоваться, чтобы сохранить преемственность изложения. При этом наряду с редко употребляемым термином “модуль вектора $|\cdot|$ ” будет применяться понятие “норма вектора $\|\cdot\|$ ”.

Итак, попытаемся сформулировать в явном виде те основные предположения, на которых основана *Векторная модель*.

|| “Основным элементом сети является формальный нейрон.” (Соколов и др., 1989; ||
|| с. 59) ||

По сути дела, используется формальный нейрон МакКаллока-Питтса, но без пороговой функции активации, т.е. простой линейный сумматор. Это означает, что если на вход нейрона подается взвешенная сумма входных воздействий, то сигнал на выходе нейрона

также является взвешенной суммой (с теми же весовыми коэффициентами) реакций нейрона на каждое из этих воздействий по отдельности.

“Детектор производит суммирование попарных произведений каждого входного сигнала на соответствующий коэффициент синаптической передачи

$$d_j = c_{j1}f_1 + \dots + c_{ji}f_i + \dots + c_{jn}f_n = \sum_{i=1}^n c_{ji}f_i$$

где d_j – величина возбуждения j -го детектора. Набор приходящих возбуждений образует вектор возбуждения $\mathbf{F} = \{f_1 \dots f_i \dots f_n\}$. Набор коэффициентов связи образует вектор связи $\mathbf{C}_j = \{c_{j1}, \dots, c_{ji}, \dots, c_{jn}\}$.” (Фомин и др., 1979; с. 12-13)

“Набор чисел $\{f_i\}$ можно рассматривать как компоненты вектора возбуждения $\mathbf{F}$.” (Соколов и др., 1989; с. 60)

На самом деле, за этим простым утверждением скрывается ряд очень сильных предположений. Называя набор чисел вектором, мы автоматически считаем его элементом линейного векторного пространства. Для векторов должны быть определены операции сложения и умножения на число, удовлетворяющие стандартным аксиомам (Колмогоров и др., 2004). Мы не предполагаем, что читатель знаком с литературой по функциональному анализу, поэтому здесь и далее приводим кратко все необходимые определения.

Непустое множество $\mathbb{L}$, содержащее элементы $\mathbf{x}, \mathbf{y}, \mathbf{z}, \dots$ называется линейным пространством, если удовлетворяются следующие аксиомы:

Для любых двух элементов $\mathbf{x}, \mathbf{y} \in \mathbb{L}$ однозначно определен третий элемент $\mathbf{z} \in \mathbb{L}$, называемый их суммой, и обозначаемый $\mathbf{x} + \mathbf{y}$, причем

1. $\mathbf{x} + \mathbf{y} = \mathbf{y} + \mathbf{x}$
2. $\mathbf{x} + (\mathbf{y} + \mathbf{z}) = (\mathbf{x} + \mathbf{y}) + \mathbf{z}$
3. В $\mathbb{L}$ существует нулевой элемент $\mathbf{0}$ такой, что $\mathbf{0} + \mathbf{x} = \mathbf{x}$ для всех $\mathbf{x} \in \mathbb{L}$
4. Для каждого $\mathbf{x} \in \mathbb{L}$ существует противоположный элемент $-\mathbf{x} \in \mathbb{L}$ такой, что $\mathbf{x} + (-\mathbf{x}) = \mathbf{0}$

Для любых чисел α, β определено произведение вектора на число $\alpha\mathbf{x} \in \mathbb{L}$, причем

5. $\alpha(\beta\mathbf{x}) = (\alpha\beta)\mathbf{x}$
6. $1 \cdot \mathbf{x} = \mathbf{x}$
7. $(\alpha + \beta)\mathbf{x} = \alpha\mathbf{x} + \beta\mathbf{x}$
8. $\alpha(\mathbf{x} + \mathbf{y}) = \alpha\mathbf{x} + \alpha\mathbf{y}$

В Векторной модели используется так называемое арифметическое пространство $\mathbb{R}^n$, элементы которого (т.е. вектора) являются наборами n чисел: $\mathbf{F} = \{f_i\}_{i=1}^n$, $\mathbf{C}_j = \{c_{ji}\}_{i=1}^n$ ($j = 1, \dots, m$). При этом операция сложения векторов определяется как покомпонентное

сложение наборов чисел, а умножение вектора на число – как умножение каждой его компоненты на это число. Легко проверить, что все аксиомы при этом выполняются.

В дальнейшем для придания выражениям более компактной формы будет использоваться матричная система записи. При этом вектора $\mathbf{F}, \mathbf{C}_j, \dots$ по умолчанию считаются вектор-столбцами. Вектор-строки получаются путем операции транспонирования: $\mathbf{F}^T, \mathbf{C}_j^T, \dots$

В соответствии с логикой теории измерений необходимо потребовать, чтобы все возможные в линейном пространстве операции над наборами чисел, как векторами, имели содержательный смысл и для описываемых ими формальных нейронов. Только в этом случае можно утверждать, что, оперируя математическими объектами (векторами), мы всегда сможем интерпретировать результаты в терминах нейронной модели. И вот тут возникает серьезная проблема. Сложение векторов возбуждения соответствует суммированию входных воздействий, умножение вектора возбуждения на число – это его пропорциональное увеличение и уменьшение. Аналогичный смысл можно придать сложению и умножению на число векторов связи. Но, к примеру, сумма вектора возбуждения и вектора связи $\mathbf{F} + \mathbf{C}_j$ содержательного смысла не имеет – эти вектора описывают объекты разной физической природы. Таким образом, считая эти вектора элементами одного и того же пространства, мы неявно наделяем модель некоторым формальным свойством, которое отсутствует у нейронов. К этой фундаментальной проблеме мы еще вернемся в дальнейшем.

“Компоненты – это числа, равные длине проекции вектора $\mathbf{F}$ на n координатных осей” (Соколов и др., 1989; с. 60).

Здесь также содержится неявное допущение. Рассматривая наборы чисел как вектора в линейном пространстве, мы ни в коей мере не связываем их с какой-либо конкретной системой координат. Чтобы говорить о координатных осях, нужно сначала ввести понятие базиса, как набора линейно-независимых векторов с максимально возможным числом элементов. Это число называется размерностью пространства, причем, вообще говоря, пространство может быть и бесконечномерным. Для множества всех наборов по n чисел количество базисных векторов также равно n , т.е. мы имеем дело с n -мерным пространством.

Каноническим базисом называется набор векторов

$$\mathbf{e}_1 = \begin{bmatrix} 1 \\ 0 \\ \dots \\ 0 \\ \dots \\ 0 \end{bmatrix}; \quad \mathbf{e}_2 = \begin{bmatrix} 0 \\ 1 \\ \dots \\ 0 \\ \dots \\ 0 \end{bmatrix}; \quad \dots \quad \mathbf{e}_i = \begin{bmatrix} 0 \\ 0 \\ \dots \\ i \\ \dots \\ 0 \end{bmatrix}; \quad \dots \quad \mathbf{e}_n = \begin{bmatrix} 0 \\ 0 \\ \dots \\ 0 \\ \dots \\ n \end{bmatrix};$$

такой, что компоненты вектора совпадают с коэффициентами его разложения по этому базису. Например, $\mathbf{F} = \{f_i\}_{i=1}^n = f_1\mathbf{e}_1 + f_2\mathbf{e}_2 + \dots + f_i\mathbf{e}_i + \dots + f_n\mathbf{e}_n$. Но, в то же время, можно выбрать другие базисы, для которых это условие не выполняется. Таким образом, принимая это условие, мы неявно фиксируем определенный базис.

“Реакция на выходе детектора возбуждения равна скалярному произведению вектора возбуждения на вектор связи $d_j = (\mathbf{F}, \mathbf{C}_j)$.” (Фомин и др., 1979; с. 13)

За этим утверждением также скрывается цепочка допущений. Дело в том, что в линейном пространстве совершенно не обязательно должно быть определено скалярное произведение. Это – дополнительная конструкция, которую можно задавать разными способами (Колмогоров и др., 2004).

Скалярное произведение – это функция, ставящая в соответствие каждой паре векторов $\mathbf{x}, \mathbf{y} \in \mathbb{L}$ некоторое число $(\mathbf{x}, \mathbf{y})$, и удовлетворяющая следующим условиям:

1. $(\mathbf{x}, \mathbf{y}) = (\mathbf{y}, \mathbf{x})$
2. $(\mathbf{x}_1 + \mathbf{x}_2, \mathbf{y}) = (\mathbf{x}_1, \mathbf{y}) + (\mathbf{x}_2, \mathbf{y})$
3. $(\alpha\mathbf{x}, \mathbf{y}) = \alpha(\mathbf{x}, \mathbf{y})$
4. $(\mathbf{x}, \mathbf{x}) \geq 0$, причем $(\mathbf{x}, \mathbf{x}) = 0$ только если $\mathbf{x} = \mathbf{0}$

В -мерном арифметическом пространстве скалярное произведение задается в виде

$$(\mathbf{x}, \mathbf{y}) = \mathbf{x}^T \mathbf{A} \mathbf{y} = \sum_{i=1}^n \sum_{j=1}^n a_{ij} x_i y_j$$

причем матрица $\mathbf{A}$ с элементами a_{ij} должна быть такой, чтобы аксиомы выполнялись (симметричной положительно определенной). В принципе, таких матриц существует бесконечно много. В *Векторной модели* считается, что скалярное произведение имеет вид

$$(\mathbf{x}, \mathbf{y}) = \mathbf{x}^T \mathbf{y} = \sum_{i=1}^n x_i y_i$$

что согласуется с моделью формального нейрона. При этом $\mathbf{A}$ – единичная диагональная матрица.

Таким образом, мы наделили наше пространство еще одной конструкцией – скалярным произведением специального вида.

“При изменении сигнала на входе реакция нейрона меняется, достигая максимума тогда, когда вектор возбуждения коллинеарен вектору связи.” (Фомин и др., 1979; с. 13)

Здесь также используется неявное допущение: предполагается, что на вход системы подаются нормированные сигналы. Для этого нужно дать определение нормы, т.е. определить, что считается длиной вектора.

Вообще говоря, понятие нормы вектора в линейном пространстве вводится независимо от скалярного произведения (Колмогоров и др., 2004). Каждому вектору $\mathbf{x} \in \mathbb{L}$ ставится в соответствие некоторое число $\|\mathbf{x}\|$ – его норма, причем должны выполняться четыре аксиомы:

1. $\|\mathbf{x}\| \geq 0$
2. $\|\mathbf{x}\| = 0$ только при $\mathbf{x} = 0$
3. $\|\alpha\mathbf{x}\| = |\alpha|\|\mathbf{x}\|$ для любого $\mathbf{x} \in \mathbb{L}$ и любого коэффициента α (однородность)
4. $\|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$ для любых $\mathbf{x}, \mathbf{y} \in \mathbb{L}$ (неравенство треугольника)

Норму можно определить множеством различных способов, но если задано скалярное произведение, то естественно считать, что квадрат длины вектора равен скалярному произведению вектора на самого себя: $\|\mathbf{x}\| = \sqrt{(\mathbf{x}, \mathbf{x})}$ (евклидова норма). Однако далее мы будем рассматривать и другие нормы, не согласованные со скалярным произведением.

Скалярному произведению

$$(\mathbf{x}, \mathbf{x}) = \mathbf{x}^T \mathbf{x} = \sum_{i=1}^n x_i x_i$$

соответствует квадратичная норма

$$\|\mathbf{x}\| = \left(\sum_{i=1}^n x_i^2 \right)^{\frac{1}{2}}$$

Чтобы доказать, что это действительно норма, нужно проверить выполнение аксиом. Единственная трудность здесь связана с доказательством неравенства треугольника. Оно следует из неравенства Коши-Буняковского $(\mathbf{x}, \mathbf{y})^2 \leq (\mathbf{x}, \mathbf{x})(\mathbf{y}, \mathbf{y})$, которое, в свою очередь, выводится из аксиом скалярного произведения (Колмогоров и др., 2004). Это неравенство превращается в равенство в том случае, когда вектора коллинеарны: $\alpha\mathbf{x} = \mathbf{y}$.

“Если длина вектора связей $\mathbf{C}_j$ для всех нейронов-детекторов одинакова, т.е. $|\mathbf{C}_j| = \text{const}$, то величина реакции нейрона-детектора при неизменном стимуле будет зависеть только от угла $\widehat{\mathbf{C}_j \mathbf{F}}$.” (Соколов и др., 1989; с. 60)

Если норма согласована со скалярным произведением, то неравенство Коши-Буняковского можно записать в виде

$$|(\mathbf{x}, \mathbf{y})| \leq \sqrt{(\mathbf{x}, \mathbf{x})} \sqrt{(\mathbf{y}, \mathbf{y})} = \|\mathbf{x}\| \|\mathbf{y}\|$$

или

$$-1 \leq \frac{(\mathbf{x}, \mathbf{y})}{\|\mathbf{x}\| \|\mathbf{y}\|} \leq 1$$

Отсюда следует, что угол между векторами может быть корректно определен выражением

$$\cos \theta = \frac{(x, y)}{\|x\| \|y\|}$$

причем $\cos \theta = 1$, и $\theta = 0$, если $\alpha x = y$.

“Кодирование сигнала номером канала – это такое кодирование, когда определенному значению сигнала соответствует один максимально возбужденный элемент из целого набора.” (Соколов и др., 1989; с. 59)

Сформулируем в общем виде экстремальную задачу определения оптимального детектора C_j для заданного вектора возбуждения F . Необходимо найти такой детектор $C_{j_{max}}$, чтобы его отклик был максимальным:

$$C_{j_{max}} = \arg \max_{C_j \in \mathcal{C}} (C_j, F)$$

Условие $C_j \in \mathcal{C}$ означает, что каким-либо образом задано множество $\mathcal{C}$ допустимых векторов связи.

В *Векторной модели* предполагается, что $\mathcal{C}$ – множество всех векторов связи ограниченной длины: $\mathcal{C} = \{C_j \mid \|C_j\| \leq \lambda\}$, где λ – произвольная константа. При пропорциональном увеличении или уменьшении всех векторов $\{C_j\}$ номер детектора с максимальным откликом не изменится. Поэтому без ограничения общности будем считать, что $\lambda = 1$.

Мы ищем детектор, дающий максимальный отклик, т.е. с наибольшим возможным значением (C_j, F) . Решением могут являться только такие вектора связи, для которых $\|C_j\| = 1$. Вектора, для которых $\|C_j\| < 1$, дают заведомо меньший отклик, и их можно было бы вообще исключить из рассмотрения. Но для дальнейших обобщений удобнее считать, что $\mathcal{C} = \{C_j \mid \|C_j\| \leq 1\}$.

В силу линейности скалярного произведения, значение индекса j_{max} , соответствующее детектору с максимальным откликом, не изменится, если заменить C_j на αC_j , т.е. искать решение задачи

$$C_{j_{max}} = \arg \max_{\alpha C_j \in \mathcal{C}} (\alpha C_j, F)$$

Если выбрать $\alpha = \frac{1}{\|C_j\|}$ ($C_j \neq 0$), то $\left\| \frac{C_j}{\|C_j\|} \right\| = 1$, и условие $\frac{C_j}{\|C_j\|} \in \mathcal{C}$ выполняется автоматически для любых $C_j \neq 0$. Итак, задачу поиска оптимального вектора связи в *Векторной модели* можно окончательно сформулировать следующим образом:

$$C_{j_{max}} = \arg \max_{C_j \neq 0} \left(\frac{C_j}{\|C_j\|}, F \right) = \arg \max_{C_j \neq 0} \frac{(C_j, F)}{\|C_j\|} = \arg \max_{C_j \neq 0} \|F\| \cos(\widehat{C_j F}) = \arg \max_{C_j \neq 0} \cos(\widehat{C_j F})$$

В силу неравенства Коши-Буняковского, $\frac{\langle \mathbf{C}_j, \mathbf{F} \rangle}{\|\mathbf{C}_j\|} \leq \|\mathbf{F}\|$. Зависящее от $\mathbf{C}_j$ выражение в левой части этого неравенства ограничено некоторой константой $\|\mathbf{F}\|$. Его максимум достигается, когда неравенство при условии $\alpha \mathbf{F} = \mathbf{C}_{j_{max}}$ превращается в равенство, и $\cos(\widehat{\mathbf{C}_j, \mathbf{F}}) = 1$. Это обосновывает условие коллинеарности векторов возбуждения и связи для оптимального детектора (Фомин и др., 1979).

Перейдем теперь к вопросам, связанным с обобщением *Векторной модели*. Как было отмечено выше, вектора возбуждения $\mathbf{F}$ и вектора связи $\mathbf{C}_j$ не должны быть элементами одного и того же арифметического пространства, т.к. операция их сложения не имеет смысла. Вместо этого будем считать, что эти вектора принадлежат двум n -мерным арифметическим пространствам: $\mathbf{F} \in \mathbb{F}$ и $\mathbf{C}_j \in \mathbb{C}$. Скалярные произведения в этих пространствах не определяются, так как им нельзя предать содержательный смысл в терминах обобщенной векторной модели. Соответственно, мы не имеем права говорить об углах между векторами в каждом из пространств.

Тем не менее, на декартовом произведении пространств $\mathbb{C} \times \mathbb{F}$ можно определить билинейную форму $\langle \mathbf{C}_j, \mathbf{F} \rangle$ (см. (Дубровский, 2009) и дальнейшие ссылки в этой работе).

Билинейная форма – это функция, которая ставит в соответствие каждой упорядоченной паре векторов $\mathbf{F} \in \mathbb{F}$ и $\mathbf{C}_j \in \mathbb{C}$ число $\langle \mathbf{C}_j, \mathbf{F} \rangle$. При этом для любых векторов и произвольных коэффициентов α, β должны выполняться следующие аксиомы:

1. $\langle \alpha \mathbf{C}_j + \beta \mathbf{C}_k, \mathbf{F} \rangle = \alpha \langle \mathbf{C}_j, \mathbf{F} \rangle + \beta \langle \mathbf{C}_k, \mathbf{F} \rangle$
2. $\langle \mathbf{C}_j, \alpha \mathbf{F}_x + \beta \mathbf{F}_y \rangle = \alpha \langle \mathbf{C}_j, \mathbf{F}_x \rangle + \beta \langle \mathbf{C}_j, \mathbf{F}_y \rangle$
3. Если для любых $\mathbf{C}_j \in \mathbb{C}$ выполняется условие $\langle \mathbf{C}_j, \mathbf{F} \rangle = 0$, то $\mathbf{F} = \mathbf{0}$
4. Если для любых $\mathbf{F} \in \mathbb{F}$ выполняется условие $\langle \mathbf{C}_j, \mathbf{F} \rangle = 0$, то $\mathbf{C}_j = \mathbf{0}$

Первые две аксиомы являются аксиомами линейности, а две оставшиеся – аксиомами отделимости. Заданная таким образом билинейная форма приводит пространства $\mathbb{C}$ и $\mathbb{F}$ в двойственность.

Непосредственно проверяется, что выражение

$$\langle \mathbf{C}_j, \mathbf{F} \rangle = \mathbf{C}_j^T \mathbf{F} = \sum_{i=1}^n c_{ji} f_i$$

описывающее отклик j -го детектора на вектор возбуждения $\mathbf{F}$, действительно удовлетворяет этим аксиомам, и, следовательно, является билинейной формой.

Главное отличие обобщенной модели от *Векторной модели* Е.Н. Соколова заключается в том, что здесь квадратичная норма не играет какой-то специальной роли, т.к. не используется конструкция скалярного произведения. Вместо этого в каждом из пространств

задается своя норма, соответственно $\|\mathbf{C}_j\|_{\mathbb{C}}$ и $\|\mathbf{F}\|_{\mathbb{F}}$. Эти нормы должны быть согласованы в том смысле, чтобы выполнялся аналог неравенства Коши-Буняковского:

$$|\langle \mathbf{C}_j, \mathbf{F} \rangle| \leq \|\mathbf{C}_j\|_{\mathbb{C}} \cdot \|\mathbf{F}\|_{\mathbb{F}}$$

Если задана одна из норм, $\|\mathbf{C}_j\|_{\mathbb{C}}$, то двойственную ей можно определить следующим образом:

$$\|\mathbf{F}\|_{\mathbb{F}} = \sup_{\|\mathbf{C}_j\|_{\mathbb{C}} \leq 1} |\langle \mathbf{C}_j, \mathbf{F} \rangle|$$

или, что эквивалентно,

$$\|\mathbf{F}\|_{\mathbb{F}} = \sup_{\mathbf{C}_j \neq 0} \frac{|\langle \mathbf{C}_j, \mathbf{F} \rangle|}{\|\mathbf{C}_j\|_{\mathbb{C}}}$$

Вообще говоря, здесь нужно доказать, что это выражение действительно определяет норму $\|\mathbf{F}\|_{\mathbb{F}}$, т.е. выполняются соответствующие аксиомы. Первые три аксиомы нормы с очевидностью следуют из определения и свойств билинейной формы. Неравенство треугольника легко доказывается (учитывая, что сумма модулей не меньше модуля суммы):

$$\|\mathbf{F}_1\|_{\mathbb{F}} + \|\mathbf{F}_2\|_{\mathbb{F}} = \sup_{\mathbf{C}_j \neq 0} \frac{|\langle \mathbf{C}_j, \mathbf{F}_1 \rangle|}{\|\mathbf{C}_j\|_{\mathbb{C}}} + \sup_{\mathbf{C}_j \neq 0} \frac{|\langle \mathbf{C}_j, \mathbf{F}_2 \rangle|}{\|\mathbf{C}_j\|_{\mathbb{C}}} \geq \sup_{\mathbf{C}_j \neq 0} \frac{|\langle \mathbf{C}_j, \mathbf{F}_1 + \mathbf{F}_2 \rangle|}{\|\mathbf{C}_j\|_{\mathbb{C}}} = \|\mathbf{F}_1 + \mathbf{F}_2\|_{\mathbb{F}}$$

Аналогично, если определена норма $\|\mathbf{F}\|_{\mathbb{F}}$, то

$$\|\mathbf{C}_j\|_{\mathbb{C}} = \sup_{\mathbf{F} \neq 0} \frac{|\langle \mathbf{C}_j, \mathbf{F} \rangle|}{\|\mathbf{F}\|_{\mathbb{F}}}$$

Нам придется отказаться от утверждения о том, что вектор возбуждения и соответствующий ему вектор связи оптимального детектора коллинеарны. Эти вектора теперь принадлежат разным подпространствам. Но, как и ранее, пользуясь тем, что

$$-1 \leq \frac{\langle \mathbf{C}_j, \mathbf{F} \rangle}{\|\mathbf{C}_j\|_{\mathbb{C}} \|\mathbf{F}\|_{\mathbb{F}}} \leq 1$$

можно определить обобщенный угол между векторами из двух разных пространств

$$\cos \theta = \cos(\widehat{\mathbf{C}_j \mathbf{F}}) = \frac{\langle \mathbf{C}_j, \mathbf{F} \rangle}{\|\mathbf{C}_j\|_{\mathbb{C}} \|\mathbf{F}\|_{\mathbb{F}}}$$

причем $\cos \theta = 1$ и $\theta = 0$, если неравенство превращается в равенство.

Будем считать, что задано некоторое ограничение на длину вектора связи, и определено множество $\mathcal{C} = \{\mathbf{C}_j \in \mathbb{C} \mid \|\mathbf{C}_j\|_{\mathbb{C}} \leq 1\}$ допустимых векторов связи. Как и ранее, нас интересует задача нахождения оптимального детектора

$$\mathbf{C}_{j_{max}} = \arg \max_{\mathbf{C}_j \in \mathcal{C}} \langle \mathbf{C}_j, \mathbf{F} \rangle = \arg \max_{\mathbf{C}_j \neq 0} \frac{\langle \mathbf{C}_j, \mathbf{F} \rangle}{\|\mathbf{C}_j\|_{\mathbb{C}}} = \arg \max_{\mathbf{C}_j \neq 0} \cos(\widehat{\mathbf{C}_j \mathbf{F}})$$

Детектор, вектор связи которого $\mathbf{C}_{j_{max}}$ является решением этой задачи, будем называть согласованным детектором для вектора возбуждения $\mathbf{F}$.

Задаваясь различными классами ограничений на вектора связи, можно подбирать параметры модели и определять вектора связи оптимальных детекторов. При этом для вычисления двойственных норм удобно использовать известные неравенства. Пары векторов $\mathbf{F} \in \mathbb{F}$ и $\mathbf{C}_j \in \mathbb{C}$, для которых неравенство превращается в равенство, как раз и позволяют найти оптимальный вектор связи для заданного вектора возбуждения. Отметим, что одновременно решается и симметричная задача: для определенного детектора с вектором связи $\mathbf{C}_j$ найти оптимальный вектор возбуждения из заданного набора $\mathcal{F} = \{\mathbf{F} \in \mathbb{F} \mid \|\mathbf{F}\|_{\mathbb{F}} \leq 1\}$:

$$\mathbf{F}_{max} = \arg \max_{\mathbf{F} \in \mathcal{F}} \langle \mathbf{C}_j, \mathbf{F} \rangle = \arg \max_{\mathbf{F} \neq 0} \frac{\langle \mathbf{C}_j, \mathbf{F} \rangle}{\|\mathbf{F}\|_{\mathbb{F}}} = \arg \max_{\mathbf{F} \neq 0} \cos(\widehat{\mathbf{C}_j \mathbf{F}})$$

Таким образом, можно говорить о паре согласованных векторов $\mathbf{C}_j$ и $\mathbf{F}$.

Приведем несколько примеров.

1. Неравенство Коши-Буняковского (в матричной записи) $|\mathbf{C}_j^T \mathbf{F}|^2 \leq \mathbf{C}_j^T \mathbf{C}_j \cdot \mathbf{F}^T \mathbf{F}$.

Позволяет получить пару двойственных квадратичных норм:

$$\|\mathbf{F}\|_{\mathbb{F}} = \sqrt{\mathbf{F}^T \mathbf{F}} = \sqrt{\sum_{i=1}^n f_i^2} \quad \text{и} \quad \|\mathbf{C}_j\|_{\mathbb{C}} = \sqrt{\mathbf{C}_j^T \mathbf{C}_j} = \sqrt{\sum_{i=1}^n c_{ji}^2}$$

Равенство достигается при $\alpha \mathbf{F} = \mathbf{C}_j$.

В этом случае обобщенная модель тождественна *Векторной модели*.

2. Обобщенное векторное неравенство Коши-Буняковского.

$|\mathbf{C}_j^T \mathbf{F}|^2 \leq \mathbf{C}_j^T \mathbf{A}^{-1} \mathbf{C}_j \cdot \mathbf{F}^T \mathbf{A} \mathbf{F}$, где $\mathbf{A}$ – симметричная положительно определенная матрица.

Равенство достигается при $\alpha \mathbf{A} \mathbf{F} = \mathbf{C}_j$.

Двойственные нормы:

$$\|\mathbf{F}\|_{\mathbb{F}} = \sqrt{\mathbf{F}^T \mathbf{A} \mathbf{F}} \quad \text{и} \quad \|\mathbf{C}_j\|_{\mathbb{C}} = \sqrt{\mathbf{C}_j^T \mathbf{A}^{-1} \mathbf{C}_j}$$

3. Неравенство Гельдера

$$\sum_{i=1}^n c_{ji} f_i \leq \left(\sum_{i=1}^n |c_{ji}|^p \right)^{\frac{1}{p}} \left(\sum_{i=1}^n |f_i|^q \right)^{\frac{1}{q}}$$

где $p \geq 1$, $q \geq 1$ и $\frac{1}{p} + \frac{1}{q} = 1$. Отсюда $\frac{q}{p} = q - 1$.

Равенство достигается при $\alpha |f_i|^q = |c_{ji}|^p$ или $\alpha |f_i|^{q-1} \text{sign}(f_i) = c_{ji}$.

Неравенство Гельдера позволяет получить пару двойственных норм:

$$\|\mathbf{F}\|_{\mathbb{F}} = \left(\sum_{i=1}^n |f_i|^q \right)^{\frac{1}{q}} \quad \text{и} \quad \|\mathbf{C}_j\|_{\mathbb{C}} = \left(\sum_{i=1}^n |c_{ji}|^p \right)^{\frac{1}{p}}$$

Для того, чтобы доказать, что это действительно нормы, нужно воспользоваться неравенством Минковского (Колмогоров и др., 2004), которое, в свою очередь, следует из неравенства Гельдера.

4. Рассмотрим важный частный случай: $p = 1, q = \infty$. При этом

$$\|\mathbf{F}\|_{\mathbb{F}} = \left(\sum_{i=1}^n |f_i|^q \right)^{\frac{1}{q}} \xrightarrow{q \rightarrow \infty} \max_i |f_i| \quad \text{и} \quad \|\mathbf{C}_j\|_{\mathbb{C}} = \sum_{i=1}^n |c_{ji}|$$

5. Случаю $p = \infty, q = 1$ соответствует пара норм

$$\|\mathbf{F}\|_{\mathbb{F}} = \sum_{i=1}^n |f_i| \quad \text{и} \quad \|\mathbf{C}_j\|_{\mathbb{C}} = \max_i |c_{ji}|$$

Для интерпретации результатов психофизических экспериментов с помощью *Векторной модели* и ее обобщений нужно дополнительно сделать какое-то предположение о том, как на основании откликов набора детекторов принимается решение о наличии стимула. Простейшее пиковое решающее правило предполагает, что предъявленный стимул обнаруживается, если реакция хотя бы одного детектора превысит пороговое значение. Путем выбора соответствующей нормировки векторов связи величину порога можно сделать равной единице.

Введем понятие подпорогового множества как множества всех векторов $\mathbf{F}$, которые не обнаруживаются анализатором вообще или обнаруживаются на пороговом уровне. Для каждого такого вектора должно выполняться условие $\langle \mathbf{C}_j, \mathbf{F} \rangle \leq 1$ для всех $\mathbf{C}_j \in \mathcal{C}$, или, что эквивалентно,

$$\sup_{\mathbf{C}_j \in \mathcal{C}} \langle \mathbf{C}_j, \mathbf{F} \rangle \leq 1$$

Но $\mathbf{C}_j \in \mathcal{C}$, если $\|\mathbf{C}_j\|_{\mathbb{C}} \leq 1$. Кроме того, ранее было показано, что

$$\|\mathbf{F}\|_{\mathbb{F}} = \sup_{\|\mathbf{C}_j\|_{\mathbb{C}} \leq 1} |\langle \mathbf{C}_j, \mathbf{F} \rangle| = \sup_{\mathbf{C}_j \neq \mathbf{0}} \frac{|\langle \mathbf{C}_j, \mathbf{F} \rangle|}{\|\mathbf{C}_j\|_{\mathbb{C}}}$$

Отсюда следует, что вектор $\mathbf{F}$ не обнаруживается, если $\|\mathbf{F}\|_{\mathbb{F}} \leq 1$, т.е. $\mathcal{F} = \{\mathbf{F} \in \mathbb{F} \mid \|\mathbf{F}\|_{\mathbb{F}} \leq 1\}$ – подпороговое множество.

Итак, мы пришли к очень важному результату. Задав норму $\|\mathbf{C}_j\|_{\mathbb{C}}$, и определив двойственную норму $\|\mathbf{F}\|_{\mathbb{F}}$ с помощью описанной выше техники, можно для каждого вектора $\mathbf{F}$ предсказать теоретически, является ли он подпороговым, и проверить это экспериментально. Естественно задаться вопросом: всегда ли можно подобрать норму $\|\mathbf{C}_j\|_{\mathbb{C}}$ так, чтобы форма подпорогового множества соответствовала экспериментальным данным?

Для ответа на этот вопрос необходимо вначале указать, какие ограничения накладывают аксиомы нормы на форму подпорогового множества, задаваемого этой нормой.

1. $\mathcal{F}$ содержит нулевой вектор. Это условие должно выполняться автоматически, т.к. нулевой вектор является подпороговым по определению.
2. Из свойств нормы следует, что $\|\mathbf{F}\|_{\mathbb{F}} = \|-\mathbf{F}\|_{\mathbb{F}}$, т.е. $\mathcal{F}$ - центрально-симметричное множество.
3. Множество $\mathcal{F}$ выпукло. Это означает, что если оно содержит любые два вектора $\mathbf{F}_1, \mathbf{F}_2 \in \mathcal{F}$, то ему принадлежит и весь соединяющий их отрезок: $\alpha\mathbf{F}_1 + (1 - \alpha)\mathbf{F}_2 \in \mathcal{F}$ (где $0 \leq \alpha \leq 1$). Действительно, пусть $\|\mathbf{F}_1\|_{\mathbb{F}} \leq 1$ и $\|\mathbf{F}_2\|_{\mathbb{F}} \leq 1$. Тогда

$$\|\alpha\mathbf{F}_1 + (1 - \alpha)\mathbf{F}_2\|_{\mathbb{F}} \leq \alpha\|\mathbf{F}_1\|_{\mathbb{F}} + (1 - \alpha)\|\mathbf{F}_2\|_{\mathbb{F}} \leq \alpha + (1 - \alpha) = 1$$

что и доказывает утверждение.

Эти условия являются необходимыми. Если полученное в эксперименте подпороговое множество не является центрально-симметричным или выпуклым, то невозможно построить для его описания векторную модель с какой-либо нормой $\|\mathbf{C}_j\|_{\mathbb{C}}$. Далее будет показано, как можно по множеству, удовлетворяющую этим условиям, построить соответствующую норму. Таким образом, эти условия являются и достаточными.

Вообще говоря, требования к модели, связанные с выполнением аксиом нормы, избыточны, и их можно ослабить. Введение в модель норм $\|\mathbf{F}\|_{\mathbb{F}}$ и $\|\mathbf{C}_j\|_{\mathbb{C}}$ является на самом деле техническим приемом, облегчающим расчеты. Этот прием заимствован из теории сигналов – там квадратичная норма имеет естественный физический смысл энергии. Однако в рассматриваемой модели квадратичная (энергетическая) норма не обладает какими-то специфическими преимуществами.

Общая формулировка задач моделирования не требует использования норм. Это можно сделать непосредственно в терминах множеств $\mathcal{F}$ и $\mathcal{C}$:

1. Из каких-либо соображений выбирается набор детекторов с заданными векторами связи $\mathbf{C}_j \in \mathcal{C}$. Нужно найти подпороговое множество $\mathcal{F}$.
2. Экспериментально получено подпороговое множество $\mathcal{F}$. Необходимо найти соответствующий ему набор детекторов $\mathcal{C}$.

В обоих случаях нужно определить, какой из детекторов будет оптимальным образом реагировать на каждый стимул.

Решение первой задачи можно выписать в явном виде, поставив в соответствие множеству $\mathcal{C}$ его поляр $\mathcal{C}^\circ$ (Рокафеллар, 1973):

$$\mathcal{F} = \mathcal{C}^\circ = \{\mathbf{F} \in \mathbb{F} \mid \sup_{\mathbf{C}_j \in \mathcal{C}} \langle \mathbf{C}_j, \mathbf{F} \rangle \leq 1\}$$

Поляра всегда содержит нулевой вектор, и является выпуклым множеством (Рокафеллар, 1973). Таким образом, выпуклость подпорогового множества является фундаментальным свойством всех моделей рассматриваемого типа. Любая модель, использующая линейные детекторы и пиковое решающее правило требует, чтобы подпороговое множество было выпуклым.

Отметим, что поляра $\mathcal{F} = \mathcal{C}^\circ$ не изменится, если к множеству $\mathcal{C}$ добавить все возможные выпуклые комбинации его элементов, т.е. вектора вида

$$\mathbf{C} = \sum_j \alpha_j \mathbf{C}_j \quad (0 \leq \alpha_j \leq 1; \sum_j \alpha_j = 1)$$

Полученное множество называется выпуклым замыканием $\mathcal{C}$ и обозначается $\text{conv}(\mathcal{C})$. Таким образом, $\mathcal{C}^\circ = (\text{conv}(\mathcal{C}))^\circ$. В терминах обобщенной векторной модели это означает, что детектор, вектор связи которого является выпуклой комбинацией векторов связи $\mathbf{C}_j$ других детекторов, не оказывает влияния на обнаружение стимула всей системой (т.е., на подпороговое множество $\mathcal{F}$). Его отклик никогда не будет превышать отклики других детекторов:

$$\langle \mathbf{C}, \mathbf{F} \rangle = \left\langle \sum_j \alpha_j \mathbf{C}_j, \mathbf{F} \right\rangle = \sum_j \alpha_j \langle \mathbf{C}_j, \mathbf{F} \rangle \leq \sup_{\mathbf{C}_j \in \mathcal{C}} \langle \mathbf{C}_j, \mathbf{F} \rangle \cdot \sum_j \alpha_j = \sup_{\mathbf{C}_j \in \mathcal{C}} \langle \mathbf{C}_j, \mathbf{F} \rangle$$

Решение второй задачи также находится в форме поляры:

$$\mathcal{C} = \mathcal{F}^\circ = \{ \mathbf{C}_j \in \mathbb{C} \mid \sup_{\mathbf{F} \in \mathcal{F}} \langle \mathbf{C}_j, \mathbf{F} \rangle \leq 1 \}$$

Полученное множество $\mathcal{C}$ является выпуклым замыканием реального набора векторов связи детекторов, т.е. включает все их выпуклые комбинации. Нахождение такого минимального подмножества векторов связи, чтобы все остальные элементы $\mathcal{C}$ были их выпуклыми комбинациями, представляет отдельную задачу (Дубровский, 2009).

Итак, мы видим, что свойство выпуклости подпорогового множества является фундаментальным для моделей данного класса. Однако известно множество экспериментов, в которых свойство центральной симметрии подпорогового множества нарушается. От этого свойства можно отказаться, если вместо норм использовать в обобщенной векторной модели калибровочные функции (Рокафеллар, 1973).

Будем называть функцию $\mathbf{k}[\mathbf{x}]$ калибровочной функцией, если она удовлетворяет следующим условиям:

1. $\mathbf{k}[\mathbf{x}] \geq 0$
2. $\mathbf{k}[\mathbf{0}] = 0$
3. $\mathbf{k}[\alpha \mathbf{x}] = \alpha \mathbf{k}[\mathbf{x}]$ для любого $\mathbf{x} \in \mathbb{L}$ и любого коэффициента $\alpha > 0$ (положительная однородность)

4. $k[x + y] \leq k[x] + k[y]$ для любых $x, y \in \mathbb{L}$ (субаддитивность)

С помощью калибровочных функций (аналогично тому, как это делалось с помощью норм) задаются выпуклые множества $\mathcal{F}$ и $\mathcal{C}$:

$$\mathcal{F} = \{\mathbf{F} \in \mathbb{F} \mid k[\mathbf{F}] \leq 1\} \text{ и } \mathcal{C} = \{\mathbf{C}_j \in \mathbb{C} \mid k^\circ[\mathbf{C}_j]\}$$

При этом, как и для норм, если выполняется неравенство

$$\langle \mathbf{C}_j, \mathbf{F} \rangle \leq k^\circ[\mathbf{C}_j] \cdot k[\mathbf{F}]$$

то $\mathcal{C} = \mathcal{F}^\circ$, $\mathcal{F} = \mathcal{C}^\circ$, и, зная одну калибровочную функцию, можно вычислить вторую:

$$k[\mathbf{F}] = \sup_{\mathbf{C}_j \neq \mathbf{0}} \frac{\langle \mathbf{C}_j, \mathbf{F} \rangle}{k^\circ[\mathbf{C}_j]} \text{ или } k^\circ[\mathbf{C}_j] = \sup_{\mathbf{F} \neq \mathbf{0}} \frac{\langle \mathbf{C}_j, \mathbf{F} \rangle}{k[\mathbf{F}]}$$

При теоретической разработке моделей калибровочные функции обычно задаются аналитически. Но если подпороговое множество $\mathcal{F}$ определяется экспериментально, то в качестве калибровочной функции используется функционал Минковского (Рокафеллар, 1973):

$$k[\mathbf{F}] = \mu(\mathbf{F} \mid \mathcal{F}) = \begin{cases} 0, & \text{если } \mathbf{F} = \mathbf{0} \\ \inf\{\lambda > 0 \mid \lambda^{-1}\mathbf{F} \in \mathcal{F}\}, & \text{если } \mathbf{F} \neq \mathbf{0} \end{cases}$$

Если $\mathcal{F}$ - выпукло, то этот функционал удовлетворяет всем аксиомам калибровочной функции. Фактически построение функционала Минковского повторяет логику хорошо известного в психофизике метода установки: необходимо подобрать такое минимальное значения коэффициента ослабления интенсивности стимула λ , чтобы стимул $\frac{1}{\lambda}\mathbf{F}$ оказался пороговым.

На этом мы завершаем краткий формальный обзор *Векторной модели* и ее возможных обобщений. Дальнейшие пути развития теории намечены в работе (Дубровский, 2009). Итак, можно сделать следующие выводы:

- В *Векторной модели* предложен (часто в неявной форме) базовый набор постулатов, позволяющий описывать сенсорный анализатор в терминах некоторой задачи оптимизации.
- Основные предположения, заложенные в *Векторной модели*, допускают ряд естественных обобщений. Это приводит к построению целого класса моделей сенсорных анализаторов. Исходная *Векторная модель* является одним из частных случаев этой теории.
- Оказывается, что формализм выпуклого анализа может служить адекватным математическим языком для описания моделей этого класса.
- Обобщенные модели допускают непосредственную проверку в психофизическом эксперименте. При этом выпуклость подпороговых множеств может служить критерием применимости разработанной теории.

Литература

1. Фомин С. В., Соколов Е. Н., Вайткавичус Г. Г. Искусственные органы чувств. - М.: Наука, 1979. – 180 с.
2. Соколов Е. Н., Вайткавичус Г. Г. НЕЙРОИНТЕЛЛЕКТ от нейрона к нейрокомпьютеру - М.: Наука, 1989. – 240 с.
3. Колмогоров А. Н., Фомин С. В. Элементы теории функций и функционального анализа. - М., ФИЗМАТЛИТ, 2004. - 572 с.
4. Рокафеллар Р. Выпуклый анализ. - М.: Мир, 1973. - 471 с.
5. Дубровский В. Е. Геометрический подход к задаче сенсорного различения // Современная психофизика - под ред. В. А. Барабанщикова - М.: изд-во “ Институт психологии РАН”, 2009. - с. 110 - 144.