Информатика

Петров Игорь Борисович
Доктор физико-математических наук, профессор, заведующий кафедрой информатики Московского физико-технического института (МФТИ).

Можно ли провести прямую через три точки?

Делая первые шаги по освоению методов обработки результатов, начинающие физики-экспериментаторы встречаются с такой задачей: провести наиболее подходящую прямую через набор точек на плоскости. Есть простое решение: взять линейку и провести прямую на глаз, так, чтобы данные точки находились близко от этой прямой. Но первое правило исследователя – всегда чётко понимать, что делается, иметь представление о точности результатов и обосновывать методы с помощью математических моделей.

Что скрывается за словосочетанием» «*наиболее подходящая прямая*»? Давайте в этом разберёмся.

Можно ли провести прямую через три точки?

Иногда можно, если очень сильно повезёт. А в большинстве случаев случайно взятые точки на координатной плоскости не лежат на одной прямой – это понимает любой школьник.

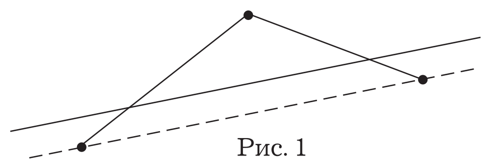

Рис. 1

Однако как в экспериментальной, так и в вычислительной науках подобная задача может встретиться. Более того, возможна и более проблемная постановка задачи о проведении прямой через N точек, где N намного больше трёх, что, разумеется, также невозможно, если все эти точки не лежат на одной прямой. Откуда же взяться столь странным задачам?

Предположим, что Вы, проведя некоторый эксперимент (лабораторный или вычислительный), получили N экспериментальных точек. Из теории Вы знаете, что они должны лежать на одной прямой (линейная зависимость функции от аргумента). В линейной зависимости вида

$$y = ax + b \qquad (1)$$

Вам необходимо вычислить два коэффициента a и b, определяющие эту линейную функцию. Однако экспериментальные точки лежат не на одной прямой, а *почти все* на одной прямой. «Что же делать?» – это вполне резонный вопрос, который Вы зададите.

Действительно, положим, что в плоскости $\{x,y\}$, где y – функция, x – аргумент, наши три точки имеют соответственно координаты

$$(x_1;y_1),\ (x_2;y_2),\ (x_3;y_3).$$

В таком случае из (1) следует, что справедливы три уравнения:

$$\begin{aligned} y_1 &= ax_1 + b, \\ y_2 &= ax_2 + b, \\ y_3 &= ax_3 + b. \end{aligned} \qquad (2)$$

Но если одно из приведённых уравнений не является следствием любого из оставшихся двух, что может быть лишь в случае, если все три точки лежат на одной прямой, то система (2) несовместна. В экспериментальной практике такие ситуации встречаются часто. Разумеется, Вы можете предложить «выбросить» одну точку и успешно провести прямую через две оставшиеся точки. Но при этом Вы потеряете часть экспериментальной информации. К тому же таких экспериментальных «точек» может быть не две, не три, а значительно больше. Допустим, N. Так не выбрасывать же $(N–2)$ точки, оставляя лишь две понравившиеся и теряя при этом почти всю информацию, полученную в эксперименте, который, кстати сказать, может оказаться весьма дорогостоящим.

Поясним цену некоторых экспериментов на примере, далёком и от математики, и от физики.

Во всей человеческой цивилизации нет более таинственного явления, чем смерть. Но, тем не менее, современные врачи спасают многих людей от смерти. Можно сказать, что из безнадёжных ситуаций людей выволакивают. А почему? Да потому, прежде всего, что многое знают о том, как смерть протекает, и представляют себе, когда этот процесс можно прервать. А как узнали? Тысячи экспериментов провели. Истории хорошо известны два из них. Первый эксперимент устроил сам над собой древнегреческий философ Сократ. В древних Афинах когда-то его приговорили к смерти: он должен был выпить чашу с ядом. При вынесении приговора судьи разделились почти пополам с маленьким перевесом в сторону казни… Сократ, известный своим остроумием и красноречием, мог бы легко переубедить судей и остаться в живых. Но в тот момент им владела совсем другая идея – он доказывал, что философ не может бояться смерти. Когда ему поднесли чашу, он спокойно выпил её, а потом рассказывал ученикам, которые пришли к нему в тюрьму, о своих предсмертных ощущениях.

Второй эксперимент устроил также над собой академик Павлов. В первой трети XX века он был уже очень пожилым человеком. Почувствовав, что умирает, созвал своих учеников и до самого конца, сколь хватило ему сил, диктовал им свои ощущения. А те плакали и записывали. Вот так были получены только две из экспериментальных «точек». Каждую «точку» знаний человек часто добывал ценой собственной жизни.

Если нельзя пренебрегать экспериментальными «точками», то, может быть, есть смысл изменить постановку задачи?

Это хорошая мысль. Вы, возможно, уже догадались, что прямую имеет смысл проводить не через три точки, что невозможно, а наиболее близко к ним, конечно же, предварительно оценив это новое понятие «наиболее близко».

С точки зрения специалиста по решению систем линейных алгебраических уравнений, мы получили переопределённую систему уравнений, в которой число уравнений превышает число неизвестных. Решения такой системы не существует.

Но вернёмся к интересующему нас понятию «наиболее близко». Методов его трактовки может быть несколько. Мы приведём идею проверенного и хорошо зарекомендовавшего себя метода. Он заключается в следующем: «Да, мы не можем решить систему (2), но мы можем построить некую функцию, которая способна количественно измерить «близость» интересующей нас прямой к трём известным точкам». В качестве такой функции резонно выбрать следующую:

$$\begin{aligned}\varphi(a,b)=(ax_1+b-y_1)^2+\\+(ax_2+b-y_2)^2+\\+(ax_3+b-y_3)^2.\end{aligned} \tag{3}$$

Почему «резонно»? Часто за такими, безобидными на первый взгляд словами, прячется нежелание автора утомлять читателя длинными выкладками. Объясним наши резоны аналогией с формулой квадрата расстояния между двумя точками $P\ (p_1;p_2)$ и $C\ (c_1;c_2)$:

$$PC^2=(p_1-c_1)^2+(p_2-c_2)^2.$$

Чем ближе друг к другу располагаются точки P и C, тем меньше отличаются от нуля квадраты разностей (p_1-c_1) и (p_2-c_2) и тем меньше становится квадрат длины отрезка PC. Видимо поэтому метод, основанный на формуле (3), называют методом наименьших квадратов. В самом деле, интуитивно ясно, что любой вопрос о *«близости»*, о расстоянии между некоторой точкой и прямой сводится к вопросу о длине отрезка, один конец которого бегает по прямой, а другой располагается в интересующей нас точке.

Очевидно, что при этом уравнения неопределённой системы (2) не будут выполняться, иначе $\varphi\ (a,b) = 0$. Однако очевидно и другое: чем ближе функция φ к нулю, тем ближе к выполнению все три равенства системы (2). При этом, конечно же, вместо этих трёх равенств мы получим три других:

$$\begin{aligned}ax_1+b-y_1=r_1,\\ax_2+b-y_2=r_2,\\ax_3+b-y_3=r_3,\end{aligned} \tag{4}$$

в которых величины r_1, r_2, r_3 называются невязками. Тогда функцию φ можно переписать в виде:

$$\varphi(a,b)=r_1^2+r_2^2+r_3^2. \tag{5}$$

В таком случае, если найти такие значения коэффициентов a и b нашей прямой, при которых функция (5) достигает своего минимума (не максимума, поскольку $\varphi > 0$), то эти коэффициенты будут искомыми.

Для школьников, знакомых с дифференцированием, скажем, что функция двух переменных $\varphi(a,b)$ достигает минимума, когда её производные по a и по b равны нулю.

Если обратиться к геометрической интерпретации более простого случая, т.е. случая, когда рассматривается функция одного переменного $f(z)$, то точкой её минимума будет точка M с координатами $\left(z_{\min};\ f_{\min}\right)$ (рис. 2).

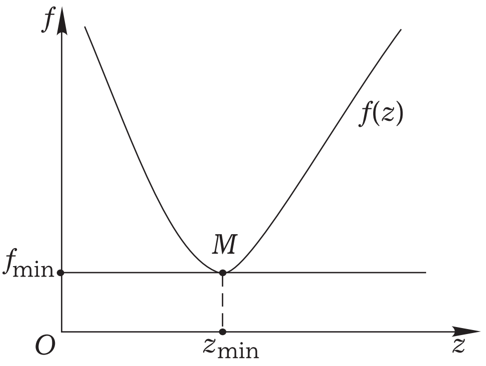

Рис. 2

Условия минимума функции φ, которые Вы можете получить, продифференцировав её сначала по a, потом по b и приравняв обе производные к нулю, дают систему из двух линейных уравнений, решив которую, Вы определите параметры a и b:

$$\begin{cases} a(x_1^2+x_2^2+x_3^2)+b(x_1+x_2+x_3)= \\ =x_1y_1+x_2y_2+x_3y_3, \\ a(x_1+x_2+x_3)+3b=y_1+y_2+y_3. \end{cases} \qquad (6)$$

В том случае, если нужно провести прямую, наиболее близкую (в смысле метода наименьших квадратов) через N точек, система будет иметь такой вид:

$$\begin{cases} a(x_1^2+x_2^2+\ldots+x_N^2)+ \\ +b(x_1+x_2+x_3+\ldots+x_N)= \\ =x_1y_1+x_2y_2+\ldots+x_Ny_N, \\ a(x_1+x_2+\ldots+x_N)+Nb= \\ =y_1+y_2+\ldots+y_N. \end{cases} \qquad (7)$$

В этом случае Вы решаете задачу уже для N точек (рис. 3).

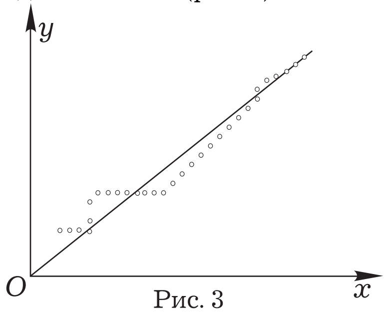

Рис. 3

Если же ваша экспериментальная зависимость не является линейной в соответствии с теорией, а, например, оказывается квадратичной, то в таком случае соответствующая функция будет записана в виде:

$$y=ax^2+bx+c.$$

Например, парабола $f=z^2$, как Вы хорошо знаете, достигает своего минимума в точке $z = 0$. А теперь повращайте эту параболу вокруг вертикальной оси и Вы получите трёхмерную фигуру, называемую параболоидом вращения. Она также имеет минимум. Но в этом случае функция f зависит уже не от одной, а от двух переменных z_1 и z_2 (рис. 4).

В заключение отметим, что теория может диктовать и ещё более сложную зависимость для экспери-

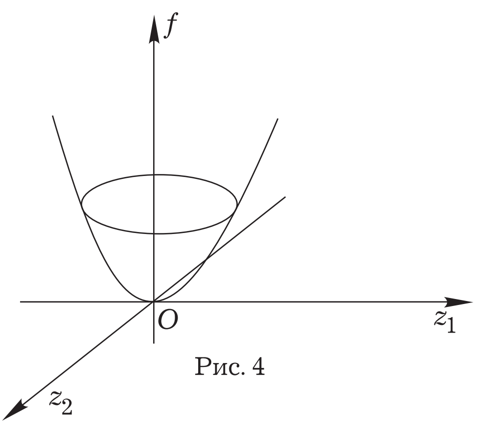

Рис. 4

ментальных «точек». Например, эта зависимость может иметь вид многочлена N-го порядка

$$y=a_Nx^N+a_{N-1}x^{N-1}+\ldots+a_1x+a_0.$$

В таком случае надо соблюдать осторожность при использовании метода наименьших квадратов, поскольку уже при $N > 7$ численное решение получившейся системы линейных уравнений может привести к заметным вычислительным ошибкам.