

Chapter 1

Prologue

1.1 An Example

A basic problem in mathematics is to make sure roots of a given function exist or do not exist. There are now numerous means for handling such problems. Yet the idea of transforming the root seeking problem into one that counts the number of tangent lines of a related function does not seem to have drawn much attention. In this book, we intend to explain in great detail how this idea can be realized for functions that involve a reasonable number of parameters.

To motivate what follows, let us consider the familiar real quadratic polynomial

$$f(\lambda|a, b) = \lambda^2 + a\lambda + b,$$

involving two real parameters a and b .

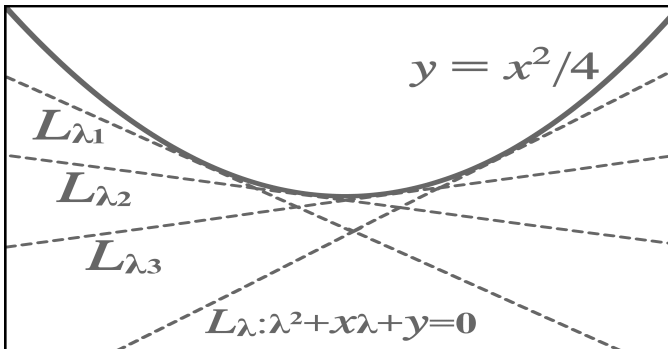


Fig. 1.1

Suppose we want to determine the necessary and sufficient condition satisfied by the coefficients a and b in order that all the roots of f are nonreal. This is an easy question since the roots of f are nonreal if, and only if, the discriminant $a^2 - 4b < 0$. We may however, approach this problem in a different manner. Let us collect all the desired pairs (a, b) into the plane set Ω .

Note first that,

$$\begin{aligned} & (c, d) \text{ is in the complement of } \Omega \\ \Rightarrow & \lambda^2 + c\lambda + d = 0 \text{ has a real root } \lambda_* \in \mathbf{R} \\ \Rightarrow & (c, d) \text{ lies on the straight line } L_{\lambda_*} : \lambda_*x + y = -\lambda_*^2. \end{aligned}$$

This motivates us to draw all possible straight lines defined by

$$L_\lambda : \lambda x + y = -\lambda^2, \lambda \in \mathbf{R}. \quad (1.1)$$

The resulting figure (see Figure 1.1) suggests the existence of a curve which ‘touches’ each of the possible straight lines at exactly one point. Such a curve is called the envelope of the family of straight lines $\{L_\lambda : \lambda \in \mathbf{R}\}$. Furthermore, this curve, as can be checked by computer experiments (and by analysis to be given later), coincides with the parabola

$$S : y = \frac{x^2}{4}, x \in \mathbf{R}.$$

Thus

$$(c, d) \text{ is in the complement of } \Omega \Rightarrow (c, d) \text{ lies on a tangent line of } S.$$

Furthermore, by reversing the above arguments, it is easily seen that the converse is also true! In other words,

$$\text{complement of } \Omega = \{(c, d) \mid \text{there is a tangent of } S \text{ passing through } (c, d)\},$$

or,

$$\Omega = \{(a, b) \mid \text{no tangents of } S \text{ pass through } (a, b)\}.$$

By inspecting Figure 1.1, it is easy to see that Ω is just the set of points lying above the parabola S , that is,

$$\Omega = \left\{ (a, b) \mid b > \frac{a^2}{4} \right\},$$

which is the correct answer!

The same method can be applied to determine the set of points (a, b) such that all roots of $f(\lambda|a, b)$ are nonpositive. We consider, instead of $\{L_\lambda \mid \lambda \in (-\infty, \infty)\}$ above, the family of straight lines $\{L_\lambda \mid \lambda \in (0, \infty)\}$. Then a unique point P_λ can be associated with each L_λ so that the totality of these points form a curve S and L_λ is the tangent line of S that passes through P_λ . See Figure 1.2. Furthermore, S is the graph of the function $y = x^2/4$ over the interval $(-\infty, 0)$ (instead of $(-\infty, \infty)$ as in the previous case). Again, it is clear from the Figure 1.2 that (a, b) is in the required region if, and only if, $a < 0$ and $b > a^2/4$, or, $a \geq 0$ and $b \geq 0$.

The same conclusion can easily be checked by manipulating the well known formula for the roots of $f(\lambda|a, b)$. Indeed, for each point (a, b) in the lower half plane, $f(0|a, b) = b < 0$. Since $\lim_{\lambda \rightarrow \infty, \lambda \in \mathbf{R}} f(\lambda|a, b) = +\infty$, we see that $f(\lambda|a, b)$ must have a positive root. Therefore, the lower half plane cannot be part of the

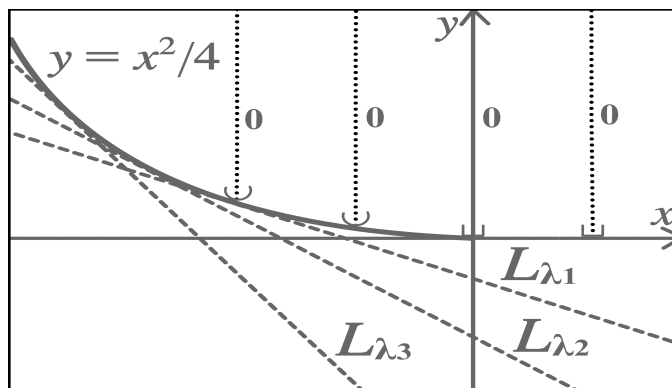


Fig. 1.2

required region. So, we consider the closed upper half plane. For each (a, b) which lies strictly above the parabola defined by $y = x^2/4$, since $a^2 - 4b < 0$, $f(\lambda|a, b)$ has two conjugate roots (which are definitely nonpositive). Suppose now (a, b) lies below the parabola and above the x -axis. Since $a^2 - 4b \geq 0$, $f(\lambda|a, b)$ has two real nonzero roots

$$\lambda_{\pm} = \frac{-x \pm \sqrt{a^2 - 4b}}{2}.$$

Furthermore, if $\lambda_+, \lambda_- \leq 0$, then

$$-x = \lambda_+ + \lambda_- \leq 0 \text{ and } y = \lambda_+ \lambda_- \geq 0,$$

and conversely, if $x \geq 0, y \geq 0$, then

$$\lambda_- \leq \lambda_+ = \frac{-a + \sqrt{a^2 - 4b}}{2} \leq \frac{-a + a}{2} = 0.$$

We have thus verified that the region that we are seeking is defined by

$$x < 0 \text{ and } y > x^2/4, \text{ or, } x \geq 0 \text{ and } y \geq 0.$$

In this book, we intend to make clear the various concepts involved in the above examples, to develop the corresponding mathematical tools, and to illustrate our idea by studying in great details two types of functions (to be collectively called quasi-polynomials) that are much studied in the theory of ordinary difference equations and ordinary differential equations.

1.2 Basic Definitions

Basic concepts from Calculus will be assumed in this book. For the sake of completeness, we will, however, briefly go through some of these concepts and their related information. We will also introduce here some common notations and conventions which will be used in this book.

First of all, sums and products of a set of numbers are common. However, empty sums or products may be encountered. In such cases, we will adopt the convention that an empty sum is taken to be zero, while an empty product will be taken as one.

The union of two sets A and B will be denoted by $A \cup B$ or $A + B$, their intersection by $A \cap B$ or AB , their difference by $A \setminus B$, and their Cartesian product by AB . The notations A^2, A^3, \dots , stand for the Cartesian products $A \times A, A \times A \times A, \dots$, respectively. It is also natural to set $A^1 = A$. The number of elements in a set Ω will be denoted by $|\Omega|$.

The set of real numbers will be denoted by \mathbf{R} , the set of complex numbers by \mathbf{C} , the set of integers by \mathbf{Z} , the set of nonnegative integers by \mathbf{N} and the set of all real m vectors by \mathbf{R}^m . The imaginary unit will be denoted by \mathbf{i} .

Therefore, the set of integers which are greater than or equal to $n \in \mathbf{R}$ is $\mathbf{Z}[n, \infty)$, the set of nonpositive complex numbers is $\mathbf{C} \setminus (0, \infty)$, the set of nonnegative complex numbers is $\mathbf{C} \setminus (-\infty, 0)$, the set of nonzero complex numbers is $\mathbf{C} \setminus \{0\}$, etc.

Real functions defined on intervals of \mathbf{R} will usually be denoted by f, g, h, F, G, H , etc. However, the identity function will be denoted by Υ , that is $\Upsilon(x) = x$ for all $x \in \mathbf{R}$; while the constant function will be denoted by Θ_γ , that is, $\Theta_\gamma(x) = \gamma$ for $x \in \mathbf{R}$. In particular, the null function is denoted by Θ_0 . Given an interval I in \mathbf{R} , the chi-function $\chi_I : I \rightarrow \mathbf{R}$ is defined by

$$\chi_I(x) = 1, \quad x \in I. \quad (1.2)$$

The restriction of a real function f defined over an interval J (which is not disjoint from I) will be written as $f\chi_I$, so that $f\chi_I$ is now defined on $I \cap J$ and

$$(f\chi_I)(x) = f(x), \quad x \in I \cap J.$$

Let f be a real function defined for points near $\alpha \in \mathbf{R}$. Recall that the derivative of f at α is $f'(\alpha) = \lim_{x \rightarrow \alpha} (f(x) - f(\alpha))/(x - \alpha)$, the right derivative of f at α is $f'_+(\alpha) = \lim_{x \rightarrow \alpha^+} (f(x) - f(\alpha))/(x - \alpha)$ and the left derivative of f at α is $f'_-(\alpha) = \lim_{x \rightarrow \alpha^-} (f(x) - f(\alpha))/(x - \alpha)$. For real functions of two or more independent variables, we may define partial derivatives in similar manners. In particular, let $g(x, y)$ be a real function of two variables, the partial derivative of g with respect to x at (α, β) is denoted by $g'_x(\alpha, \beta)$.

Let I be an interval which may be open or closed on either sides. As usual, $f : I \rightarrow \mathbf{R}$ is said to be continuous at an interior point ξ of I if $\lim_{x \rightarrow \xi} f(x) = f(\xi)$. If I is closed on the left (right) with the boundary point c (respectively d), then f is said to be continuous at c if $\lim_{x \rightarrow c^+} f(x) = f(c)$ (respectively $\lim_{x \rightarrow d^-} f(x) = f(d)$). f is said to be continuous on I if it is continuous at every point of I . The derivative of f at an interior point ξ of I is $f'(\xi)$. If I is closed on the left with the boundary point c , then the *derivative* of f at c is taken to be the right derivative of f at c , i.e.,

$$f'(c) = \lim_{x \rightarrow c^+} \frac{f(x) - f(c)}{x - c} = f'_+(c);$$

while if I is closed on the right with the boundary point d , then the *derivative* of f at d is

$$f'(d) = \lim_{x \rightarrow d^-} \frac{f(x) - f(d)}{x - d} = f'_-(c).$$

Therefore, if f is a real function defined on an interval I (e.g. $[c, d)$), then f is said to be differentiable if $f'(x)$ exists for every $x \in I$, and in such a case, the derived function f' is the function defined on I with $f'(x)$ as its value at $x \in I$. f is said to be *smooth* on I if its derived function f' is continuous on I . A smooth function g defined on an interval I is denoted by $g \in C^1(I)$.

In the sequel, we need sufficient conditions for a function to be smooth on an interval I .

Example 1.1. Suppose $g : I \rightarrow \mathbf{R}$, where $I = [c, d)$, is continuous. If $f : I \rightarrow \mathbf{R}$ satisfies $f'(x) = g(x)$ for $x \in (c, d)$, then $f \in C^1(I)$. Indeed, it suffices to show that f' is continuous at c . This can be seen by using the mean value theorem:

$$\lim_{x \rightarrow c^+} \frac{f(x) - f(c)}{x - c} = \lim_{x \rightarrow c^+} q(\zeta_x)$$

where ζ_x is a number in $(0, x)$. Since $\lim_{x \rightarrow c^+} q(\zeta_x) = q(c)$, we see that $f'_+(c)$ exists, is equal to $q(c)$, and

$$\lim_{x \rightarrow c^+} f'(x) = \lim_{x \rightarrow c^+} q(x) = q(c) = f'_+(c).$$

In other words, f' is continuous at c .

Given a real function g of a real variable, the following limits (which may, or may not exist)

$$\begin{aligned} & \lim_{x \rightarrow c^+} g(x), \quad \lim_{x \rightarrow c^-} g(x), \quad \lim_{x \rightarrow c^+} g'(x), \quad \lim_{x \rightarrow c^-} g'(x), \\ & \lim_{x \rightarrow -\infty} g(x), \quad \lim_{x \rightarrow +\infty} g(x), \quad \lim_{x \rightarrow -\infty} g'(x), \quad \lim_{x \rightarrow +\infty} g'(x) \end{aligned}$$

will be needed quite extensively for expressing various facts. For this reason, we will employ the following notations

$$\begin{aligned} g(c^+) &= \lim_{x \rightarrow c^+} g(x), & g(c^-) &= \lim_{x \rightarrow c^-} g(x), \\ g'(c^+) &= \lim_{x \rightarrow c^+} g'(x), & g'(c^-) &= \lim_{x \rightarrow c^-} g'(x), \\ g(-\infty) &= \lim_{x \rightarrow -\infty} g(x), & g(+\infty) &= \lim_{x \rightarrow +\infty} g(x), \\ g'(-\infty) &= \lim_{x \rightarrow -\infty} g'(x), & g'(+\infty) &= \lim_{x \rightarrow +\infty} g'(x). \end{aligned}$$

The *graph* of a real function f defined on a set J of real numbers is the set

$$\{(x, y) \in \mathbf{R}^2 \mid y = f(x), x \in J\}.$$

For the sake of convenience, we will use the same notation to indicate a (real) function of a real variable and its graph. Therefore, in the sequel, we will meet

statements such as ‘the set S is also the graph of a function $y = S(x)$ defined on the interval $I \dots$ ’.

Now let g be a function defined on an interval I . If the derivative $g'(\lambda)$ exists, then the *tangent line* of the graph g through the point $(\lambda, g(\lambda))$ is taken to mean the graph of the function $L_{g|\lambda}$ defined by

$$L_{g|\lambda}(x) = g'(\lambda)(x - \lambda) + g(\lambda), \quad x \in \mathbf{R}. \quad (1.3)$$

The so called ‘vertical tangents’ in some of the elementary analysis text books will not be regarded as tangent lines of our functions.

We say that a point (α, β) in the plane is strictly above (above, strictly below, below) the graph of a function g if α belongs to the domain of g and $g(\alpha) < \beta$ (respectively $g(\alpha) \leq \beta$, $g(\alpha) > \beta$ and $g(\alpha) \geq \beta$). The corresponding notations¹ are $(\alpha, \beta) \in \vee(g)$, $(\alpha, \beta) \in \bar{\vee}(g)$, $(\alpha, \beta) \in \wedge(g)$ and $(\alpha, \beta) \in \underline{\wedge}(g)$.

Example 1.2. Let $g(x) = x^2/4$ for $x \in \mathbf{R}$. Then $(a, b) \in \vee(g)$ if, and only if, $b > a^2/4$, while $(a, b) \in \bar{\vee}(g)$ if, and only if, $b \geq a^2/4$.

We also need to handle the ‘ordering’ relations between points and several graphs in the plane. Suppose we now have two real functions g_1 and g_2 defined on real subsets I_1 and I_2 respectively. We say that $(\alpha, \beta) \in \vee(g_1) \oplus \vee(g_2)$ if $\alpha \in I_1 \cap I_2$ and $\beta > g_1(\alpha)$ and $\beta > g_2(\alpha)$, or, $\alpha \in I_1 \setminus I_2$ and $\beta > g_1(\alpha)$, or, $\alpha \in I_2 \setminus I_1$ and $\beta > g_2(\alpha)$. The notations $(\alpha, \beta) \in \bar{\vee}(g_1) \oplus \vee(g_2)$, $(\alpha, \beta) \in \bar{\vee}(g_1) \oplus \wedge(g_2)$, etc. are similarly defined.

If we now have n real functions g_1, \dots, g_n defined on intervals I_1, \dots, I_n respectively, we write $(\alpha, \beta) \in \vee(g_1) \oplus \vee(g_2) \oplus \dots \oplus \vee(g_n)$ if $\alpha \in I_1 \cup I_2 \cup \dots \cup I_n$, and if

$$\alpha \in I_{i_1} \cap I_{i_2} \cap \dots \cap I_{i_m} \Rightarrow \beta > g_{i_1}(\alpha), \beta > g_{i_2}(\alpha), \dots, \beta > g_{i_m}(\alpha), \quad i_1, \dots, i_m \in \{1, \dots, n\}.$$

The notations $(\alpha, \beta) \in \bar{\vee}(g_1) \oplus \bar{\vee}(g_2) \oplus \dots \oplus \bar{\vee}(g_n)$, etc. are similarly defined.

Example 1.3. Let $f(x) = x^2/4$ for $x < 0$ and $g(x) = x^2/4$ for $x \in (-\infty, 1]$. Then

$$(a, b) \in \vee(f) \oplus \bar{\vee}(\Theta_0) \text{ if, and only if, } a < 0 \text{ and } b > a^2/4, \text{ or, } a \geq 0 \text{ and } b \geq 0,$$

and

$$(a, b) \in \vee(g) \oplus \bar{\vee}(\Upsilon/4) \text{ if, and only if, } a < 1 \text{ and } b > a^2/4, \text{ or, } a \geq 1 \text{ and } b \geq a/4.$$

Example 1.4. Let G_1, G_2 and G_3 be respectively the functions (see Figure 1.3)

$$G_1(x) = 2 \exp(1 - x) - 3, \quad x \in (-\infty, 1),$$

$$G_2(x) = 2 \exp(1 + x) - 3, \quad x \in (-1, \infty),$$

and

$$G_3(x) = -x^2, \quad x \in [-1, 1].$$

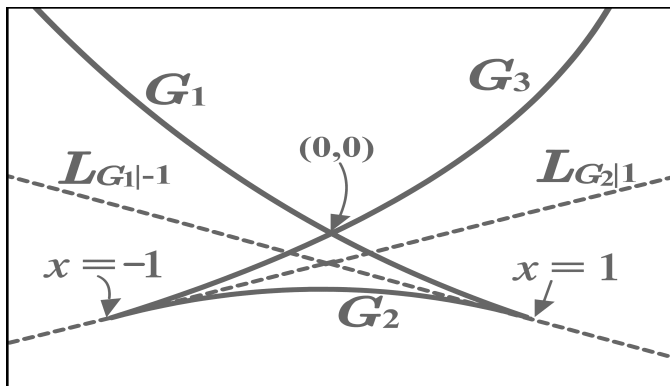


Fig. 1.3

Then

$$\{(x, y) \in \mathbf{R}^2 : y > \max \{G_1(x), G_2(x)\}\} = \vee(G_1) \oplus \vee(G_2) = \vee(G_1) \oplus \vee(G_2) \oplus \vee(G_3).$$

Note that $\vee(G_1) \oplus \vee(G_2)$ is a better description of the set on the left hand side than $\vee(G_1) \oplus \vee(G_2) \oplus \vee(G_3)$ since less work is needed to figure out its precise content. An even better one is as follows. We first note by solving $G_1(x) = G_2(x)$ that the x -coordinate of the point of intersection of the two graphs G_1 and G_2 is 0. Then $\vee(G_1) \oplus \vee(G_2)$ is equal to $\vee(G_1 \chi_{(-\infty, 0)}) \oplus \vee(G_2 \chi_{[0, \infty)})$. Again, the latter description is preferred since it is ‘precise’ in the sense that the domains of $G_1 \chi_{(-\infty, 0)}$ and $G_2 \chi_{[0, \infty)}$ have empty intersection. In spite of the precision obtained, however, cumbersome calculations (such as finding the point of intersection of the graphs G_1 and G_2) may be needed which are irrelevant to the illustration of the principles involved. Therefore in the sequel, we may use less precise formulas such as $\vee(G_1) \oplus \vee(G_2)$.

¹ $\nabla(g)$ is the epigraph of g . Such a term, however, is not needed in the sequel.