Theory of R-functions and Applications: A Primer*

Vadim Shapiro**

91-1219
July 1991

Department of Computer Science
Cornell University
Ithaca, NY 14853-7501

*This document is also available as CPA Technical Report CPA88-3, Cornell Programmable Automation, Sibley School of Mechanical Engineering, Ithaca NY 14853. It was prepared while the author was a member of the CPS project, and was supported by the General Motors Corporation (through a Corporate Fellowship) and by the National Science Foundation under grant MIP-8719196. This report was revised with the support of The Advanced Research Projects Agency of the Department of Defense under Office of Naval Research Contract N00014-88-K-0591.

**On leave from the Computer Science Department, General Motors Research Laboratories, Warren, MI 48090, shapiro@gmr.com. Current address: Department of Computer Science, Cornell University, Ithaca, NY 14853, shapiro@cornell.edu.

Theory of R-functions and Applications: A Primer *

Vadim Shapiro†

November 1988
(Revised June 1991)

*This document is also available as CPA Technical Report CPA88-3, Cornell Programmable Automation, Sibley School of Mechanical Engineering, Ithaca, NY 14853. It was prepared while the author was a member the CPA project, and was supported by the General Motors Corporation (through a Corporate Fellowship) and by the National Science Foundation under grant MIP-8719196. This report was revised with the support of The Advanced Research Projects Agency of the Department of Defense under Office of Naval Research Contract N00014-88-K-0591.

†On leave from the Computer Science Department, General Motors Research Laboratories, Warren, MI 48090, shapiro@gmr.com. Current address: Department of Computer Science, Cornell University, Ithaca, NY 14853, shapiro@cs.cornell.edu.

Foreword

An R-function is real-valued function characterized by some property that is completely determined by the corresponding property of its arguments, e.g., the sign of some real functions is completely determined by the sign of their arguments. More generally, such a property could be determined by some partition of the real axis. If the axis is partitioned into k subsets, each R-function corresponds to a companion function of k-valued logic. This relationship allows one to represent a logical predicate of n variable by a real-valued function of n arguments. The latter can be evaluated, differentiated, and possesses many other interesting properties. V. L. Rvachev first suggested R-functions in 1963. Since then, he and his colleagues have significantly developed the theory and found many applications. Their work is described in a numerous books and articles, unfortunately mostly in Russian. A complete list of references through 1987 can be found in [Shi88].

An important application of R-functions is in the description of geometric objects. Any object defined by a predicate on "primitive" geometric regions (e.g. regions defined by a system of inequalities) can now be represented by a *single* inequality, or equation. Furthermore, these real-valued functions can be constructed so that they have certain useful logic and differential properties. Application of theory of R-functions could have a profound effect on many problems where geometric information can be accounted for analytically. For example, according to [Rva82], R-functions have found applications in many unexpected areas, such as study of stability of motion, medical diagnostics, and chemical engineering, in addition to those described in this report.

This primer summarizes some basic results from the theory of R-functions and describes (rather superficially) some of the applications studied in the Soviet Union. As far as I know, this is the first such introduction to R-functions in English. Its main prupose is to stimulate interest in R-functions in the research community; it is not intended to serve as a comprehensive reference. While this document contains no original results, absorbing, translating, interpreting, and condensing the contents of the references did require a substantial judgement on my part. The original sources offer a wealth of additional material that was omitted for the sake of simplicity and coherence of this document. Thus, I also accept the responsibility for all mistakes, misinterpretations, and omissions in this report. While the main applications of R-functions have been in the description of geometric objects, the developed theory does not seem to rely on many known results in combinatorial, algebraic, and computational geometry and topology. In an effort to preserve the spirit of the original work, I resisted making "improvements" in the presentation.

My current interest in R-functions is stimulated by their obvious relationship to geometric and solid modeling. It remains to be seen whether problems in geometric modeling can be formulated and solved using this theory. Topics of interest include automatic construction of real functions with desired properties, relationship to regularized Boolean operations on solid objects, interference detection, robust computations, and applications in design and analysis of solids.

Lastly, I would like to thank Vadim Komkov who pointed me to the 1967 book by Rvachev [Rva67], and Rick Palmer who suggested a number of improvements in this primer.

Contents

1 Theory of R-functions

1.1 R-functions.

Some real-valued functions of real variables have the property that their signs are completely determined by the signs of their arguments and are independent of the magnitude of the arguments. For example, the function $W_1 = xyz$ is negative if the number of its negative arguments is odd and positive otherwise. A similar property is possessed by functions

$$
\begin{aligned}
W_2 &= x + y + \sqrt{xy + x^2 + y^2}, \\
W_3 &= 2 + x^2 + y^2 + z^2, \\
W_4 &= x + y + z - \sqrt{x^2 + y^2} - \sqrt{x^2 + z^2} - \sqrt{y^2 + z^2} + \sqrt{x^2 + y^2 + z^2}, \\
W_5 &= xy + z + |z - yx|,
\end{aligned}
$$

and so on. Table 1.1 shows how the signs of these functions depend on the signs of their arguments.

For comparison, here are some functions whose sign depends not only on the sign of the arguments but also on their magnitude:

$$
\begin{aligned}
W_6 &= xyz + 1, \\
W_7 &= \sin xy, \\
W_8 &= x + y + z - \sqrt{x^2 + y^2},
\end{aligned}
$$

and so on. Specifying distribution of signs for arguments of functions $W_1 - W_5$ completely determines the corresponding sign distribution of the functions; functions W_6, W_7 and W_8 do not behave in this fashion.

Besides the partition of real numbers into positive and negative, there are infinitely many choices for partitions. For example, one can partition real numbers into rational and irrational numbers, or, say, into all real numbers in interval $[0, 1]$, and the the rest of the real numbers. It is possible to introduce several or even infinitely many gradations when subdividing the set of real numbers. In general, any such partition Γ of the set of real numbers (based on some criterion) also determines a set $R(\Gamma)$ of those real functions that in some sense "inherit" the partition criterion (sign, rationality, membership in $[0, 1]$, etc.). Such functions will be called **R-functions**.

In this document, we will only consider R-functions defined by the decomposition of the real axis into positive and negative numbers, namely partition $\Gamma_2^0 = \{(-\infty, 0]$ and $[0, +\infty)\}$. Including zero in both intervals may seem strange, and we will return to this issue in section 1.6. For the time being, we assume that zero is always signed: $+0$, or -0; this allows us to determine whether it belongs to the set of positive or negative numbers. Consider a function $S_2(x)$ defined on the real axis as follows:

$$
S_2(x) = \begin{cases} 0 & \text{if } x \leq -0, \\ 1 & \text{if } x \geq +0. \end{cases}
$$

x	y	z	W_1	W_2	W_3	W_4	W_5
$-$	$-$	$-$	$-$	$-$	$+$	$-$	$+$
$-$	$-$	$+$	$+$	$-$	$+$	$-$	$+$
$-$	$+$	$-$	$+$	$+$	$+$	$-$	$-$
$-$	$+$	$+$	$-$	$+$	$+$	$-$	$+$
$+$	$-$	$-$	$+$	$+$	$+$	$-$	$-$
$+$	$-$	$+$	$-$	$+$	$+$	$-$	$+$
$+$	$+$	$-$	$-$	$+$	$+$	$-$	$+$
$+$	$+$	$+$	$+$	$+$	$+$	$+$	$+$

Table 1: The signs of real functions W_1 - W_5 depend only on the signs of their arguments x, y, and z

The foregoing concept of R-function implies that a function $y = f(x_1, \dots, x_n)$ is an R-function if there exists a Boolean[1] function $Y = F(X_1, \dots, X_n)$ such that the following equality is satisfied:

$$S_2[(f(x_1, x_2, \dots, x_n)] = F[S_2(x_1), S_2(x_2), \dots, S_2(x_n)].$$

We will refer to this Boolean function as a *companion* function of a given R-function. Informally, a real function is an R-function if it can change its property (sign) only when some of its arguments change the same property (sign).

It further follows that to every Boolean function there correspond an infinite number of R-functions. For example, the Boolean companion function for R-function $w_1 = xy$ is $X \Leftrightarrow Y$ (X is equivalent to Y). Just check to see that

$$S_2(xy) = S_2(x) \Leftrightarrow S_2(y).$$

But logical equivalence is also a Boolean companion function for R-functions like

$$\begin{aligned} w_2 &= xy(1 + x^2 + y^2), \\ w_3 &= (1 - 2^{-x})(3^y - 1), \end{aligned}$$

and so on. The set of all R-functions that have the same companion Boolean function is called a *branch* of the set of R-functions. Since the number of distinct Boolean functions of n arguments is 2^{2^n}, there are also 2^{2^n} distinct branches of R-functions of n arguments.

[1]In this document, by Boolean functions we mean functions of the Boolean algebra of logic with truth value 1 and false value 0. Such functions can be defined using logical operations $\wedge$, $\vee$, and $\neg$ on n logical variables.

1.2 Branches and systems of R-functions.

The following is only a partial list of some general properties of R-functions. Complete proofs, as well as many additional properties, can be found in the references, notably in [Rva67] and [Rva82].

1. The set of R-functions is closed under composition. In other words, any function obtained by composition of R-functions is also an R-function.

2. If a continuous function $y = f(x_1, \ldots, x_n)$ has zeros only on coordinate hyperplanes (i.e. $y = 0$ implies that one or more $x_j = 0, j = 1, 2, \ldots, n$), then y is an R-function;

3. Product of R-functions is another R-function (because Boolean companion of the binary product is logical equivalence). If R-function $y = f(x_1, \ldots, x_n)$ belongs to some branch, and $g(x_1, \ldots, x_n) > 0$ is an arbitrary function, then the function fg also belongs to the same branch;

4. If f_1 and f_2 are R-functions from the same branch, then the sum f_1+f_2 is an R-function belonging to the same branch;

5. If f is an R-function whose companion Boolean function is F, and C is some constant, then Cf is also an R-function. The Boolean companion function of Cf is F if $C > 0$, or $\neg F$ if $C < 0$.

6. If $f(x_1, \ldots, x_n)$ is an R-function whose Boolean companion function is $F(X_1, \ldots, X_n)$ and f can be integrated by x_i, then function

$$\phi(x_1, \ldots, x_n) \equiv \int_0^{x_i} f(x_1, \ldots, x_n) dx_i$$

is an R-function whose Boolean companion function is

$$\Phi(X_1, \ldots, X_n) \equiv [X_i \Leftrightarrow F(X_1, \ldots, X_n)].$$

7. The following functions are R-functions (followed by the Boolean companion function in parethesis):

$$\begin{array}{rll} C & \equiv \text{const}; & \text{(logical 1)} \\ \overline{x} & \equiv -x; & \text{(logical negation } \neg) \\ x_1 \wedge_1 x_2 & \equiv \min(x_1, x_2); & \text{(logical conjunction } \wedge) \\ x_1 \vee_1 x_2 & \equiv \max(x_1, x_2); & \text{(logical disjunction } \vee) \end{array}$$

The set of R-functions is infinite. However, for applications, it is not necessary to know all R-functions; we only need to be able to construct R-functions that belong to any specified branch. This leads to the notion of *sufficiently complete* systems of R-functions.

Theorem 1 ([Rva67]) *Let H be some system of R-functions, and G be the corresponding system of companion Boolean functions. The system H is sufficiently complete, if the system G is complete.*

The criteria for completeness of the system of Boolean functions are well understood [Rva82, page 53]. For example, take $G = \{0, \neg X, X_1 \wedge X_2\}$. It is well known that all logic functions can be constructed using just conjunction and negation; in other words, G is complete. Let us define R-negation and R-conjunction as

$$\begin{aligned} \overline{x} \quad &\equiv -x; \\ x_1 \wedge_0 x_2 \quad &\equiv x_1 + x_2 - \sqrt{x_1^2 + x_2^2}. \end{aligned}$$

Theorem 1 states that an R-function from any branch can be defined using composition of just these two functions.

Theorem 2 ([Rva74, page 62]) *Every branch of the set of R-functions contains at least one continuous R-function.*

Since the R-conjunction and the R-negation as defined above are both continuous functions, this result follows directly from Theorem 1. It is worth noting that, if the real axis were subdivided into two intervals $(-\infty, 0)$ and $[0, +\infty)$ (i.e. 0 is considered to be a positive number), the Theorem 2 would not be true. To see this, observe that any R-negation $f(x)$ (i.e. R-function whose companion Boolean function is logical negation) would have to be discontinuous at $x = 0$ [Rva74, page 58].

1.3 Sufficiently complete systems of R-functions.

Following Theorem 1, it is fairly easy to come up with any number of sufficiently complete systems of R-functions. In this section we compare several such systems. Our main considerations include simplicity, continuity, and differential properties, as well as convenience of use. As the first step, let us decide on the basic system of Boolean companion functions that is complete. It is customary to choose such a system as

$$G = \{0;\ \neg X;\ X_1 \wedge X_2;\ X_1 \vee X_2\},$$

even though only one of conjunction and disjunction is required to form a complete system.

Using property 7 in the previous section, we can see that

$$R_1(\Gamma) = \{-1;\ -x;\ \min(x_1, x_2);\ \max(x_1, x_2)\}$$

is the corresponding sufficiently complete system of R-functions (subscript 1 will become clear below). Furthermore, since

$$\min(x_1, x_2) + \max(x_1, x_2) = x_1 + x_2, \quad \text{and} \quad \min(x_1, x_2)\max(x_1, x_2) = x_1 x_2,$$

$\min(x_1, x_2)$ and $\max(x_1, x_2)$ are the smallest and the largest root respectively of the equation

$$z^2 - (x_1 + x_2)z + x_1 x_2 = 0.$$

Solving this equation, we get

$$\min(x_1, x_2) \equiv \frac{1}{2}[x_1 + x_2 - \sqrt{(x_1 - x_2)^2}];$$

$$\max(x_1, x_2) \equiv \frac{1}{2}[x_1 + x_2 + \sqrt{(x_1 - x_2)^2}].$$

The operation of $\sqrt{x^2}$ can be replaced with $|x|$ which is convenient for computational purposes. Because $|x_1 - x_2|$ is not differentiable along the lines $x_1 = x_2$, system $R_1(\Gamma)$ is not usable in many applications where differential properties of R-functions are important. For this and other reasons that will become clear later, we want to generalize the above system by finding alternative R-disjunction and R-conjunction. In particular, we obtain a system $R_\alpha(\Gamma)$ as follows:

$$x_1 \wedge_\alpha x_2 \equiv \frac{1}{1+\alpha}(x_1 + x_2 - \sqrt{x_1^2 + x_2^2 - 2\alpha x_1 x_2});$$

$$x_1 \vee_\alpha x_2 \equiv \frac{1}{1+\alpha}(x_1 + x_2 + \sqrt{x_1^2 + x_2^2 - 2\alpha x_1 x_2}),$$

where $\alpha(x_1, x_2)$ is an arbitrary *symmetric* function such that $-1 < \alpha(x_1, x_2) \leq 1$.

Note that the multiplier $1/(1 + \alpha)$ is always positive and thus does not affect which branch an R-function belongs to. It is needed for validity of certain differential properties of functions in $R_\alpha(\Gamma)$ in section 1.5. The precise value of α often may not matter, and it will be set to constant. For example, $\alpha = 1$ would yield the system $R_1(\Gamma)$ defined above. Similarly, $\alpha = 0$ results in the system $R_0(\Gamma)$ which uses:

$$x_1 \wedge_0 x_2 \equiv x_1 + x_2 - \sqrt{x_1^2 + x_2^2};$$

$$x_1 \vee_0 x_2 \equiv x_1 + x_2 + \sqrt{x_1^2 + x_2^2}.$$

This system may sometimes be advantageous because the defined R-functions are not differentiable only at the point $(0,0)$.

Let us also introduce a system $R_\alpha^m(\Gamma)$ using:

$$x_1 \wedge_\alpha^m x_2 \equiv (x_1 \wedge_\alpha x_2)(x_1^2 + x_2^2)^{\frac{m}{2}};$$

$$x_1 \vee_\alpha^m x_2 \equiv (x_1 \vee_\alpha x_2)(x_1^2 + x_2^2)^{\frac{m}{2}}.$$

These R-functions are analytic everywhere except at the origin, where they are at least m times differentiable (i.e. they are in C^m).

One may wonder whether the above R-functions are as simple as possible. For example, can we find a sufficiently complete system of R-functions among polynomials? The answer is no. It was shown in [Rva67] that addition and mupliplication alone are not enough to define a sufficiently complete system of R- functions. On the other hand, a sufficiently complete system does not have to use $\sqrt{\ }$. For example,

$$x_1 \wedge_C^m x_2 \equiv \left(\frac{x_1 + x_2}{2}\right)^m \operatorname{sign}(x_1 + x_2)^{m+1} - \left(\frac{x_1 - x_2}{2}\right)^m \operatorname{sign}(x_1 - x_2)^m;$$

$$x_1 \vee_C^m x_2 \equiv \left(\frac{x_1 + x_2}{2}\right)^m \operatorname{sign}(x_1 + x_2)^{m+1} + \left(\frac{x_1 - x_2}{2}\right)^m \operatorname{sign}(x_1 - x_2)^m$$

are also legitimate choices. Other sufficiently complete systems of R-functions can be found in [Rva82]. Given any of the above sufficiently complete systems of R-functions, we can construct R-functions for any companion Boolean function.

Example 1.1 Consider an R_0-function corresponding to the Boolean function $Y = (\neg X_1 \wedge X_2) \vee (X_1 \wedge \neg X_2)$:

$$\begin{aligned}
y &\equiv (\overline{x}_1 \wedge_0 x_2) \vee_0 (x_1 \wedge_0 \overline{x}_2) \\
&\equiv -x_1 + x_2 - \sqrt{x_1^2 + x_2^2} + x_1 - x_2 - \sqrt{x_1^2 + x_2^2} \\
&+ \left[(-x_1 + x_2 - \sqrt{x_1^2 + x_2^2})^2 + (x_1 - x_2 - \sqrt{x_1^2 + x_2^2})^2\right]^{\frac{1}{2}}.
\end{aligned}$$

After simplification we get

$$y = -\frac{2x_1x_2}{\sqrt{x_1^2 + x_2^2 - x_1x_2} + \sqrt{x_1^2 + x_2^2}}.$$

If we happen know that either x_1 or x_2 in Example 1.1 are not zero we can further drop the positive multiplier to obtain $y = -x_1x_2$, an R-function that belongs to the same branch. Naturally, in many cases, other properties of R-functions are also important and such a simplification cannot be used. Other simplification techniques are also known, for example for n-ary R-disjunctions and R-conjunctions. But, more generally, optimization of R-functions according to some criteria remains a challenging open problem [Rva82, page 127].

1.4 Logic properties of elementary R-functions.

R-functions of the sufficiently complete systems possess a number of properties that are similar to the properties of their companion Boolean functions. Specifically, for any system $R_\alpha(\Gamma)$ as defined, we have:

1. $\overline{\overline{x}} \equiv x$;
2. $x_1 \wedge_\alpha x_2 \equiv x_2 \wedge_\alpha x_1$;
3. $x_1 \vee_\alpha x_2 \equiv x_2 \vee_\alpha x_1$;
4. $\overline{x_1 \wedge_\alpha x_2} \equiv \overline{x}_1 \vee_\alpha \overline{x}_2$;
5. $\overline{x_1 \vee_\alpha x_2} \equiv \overline{x}_1 \wedge_\alpha \overline{x}_2$;

6. $(x_1 \wedge_\alpha x_2) + (x_1 \vee_\alpha x_2) \equiv \frac{2}{1+\alpha}(x_1 + x_2);$

7. $(x_1 \wedge_\alpha x_2)(x_1 \vee_\alpha x_2) \equiv \frac{2}{1+\alpha} x_1 x_2;$

8. $x_1 \wedge_\alpha x_2 = 0 \quad \Longleftrightarrow \quad x_1 = 0, x_2 \geq 0, \text{ } or \text{ } x_2 = 0, x_1 \geq 0;$

9. $x_1 \vee_\alpha x_2 = 0 \quad \Longleftrightarrow \quad x_1 = 0, x_2 \leq 0, \text{ } or \text{ } x_2 = 0, x_1 \leq 0.$

It is easy to show that the above properties are also valid for R-functions in system $R_\alpha^m(\Gamma)$ (properties 6 and 7 need to be changed appropriately). If $\alpha = 1$, the functions $x_1 \wedge_1 x_2$ and $x_1 \vee_1 x_2$ possess the following properties in addition to the above.

1. $x \wedge_1 x \equiv x;$

2. $x \vee_1 x \equiv x;$

3. $x \wedge_1 \overline{x} \equiv -|x|;$

4. $x \vee_1 \overline{x} \equiv |x|;$

5. $x_1 \wedge_1 (x_2 \wedge_1 x_3) \equiv (x_1 \wedge_1 x_2) \wedge_1 x_3;$

6. $x_1 \vee_1 (x_2 \vee_1 x_3) \equiv (x_1 \vee_1 x_2) \vee_1 x_3;$

7. $x_1 \wedge_1 (x_2 \vee_1 x_3) \equiv (x_1 \wedge_1 x_2) \vee_1 (x_1 \wedge_1 x_3);$

8. $x_1 \vee_1 (x_2 \wedge_1 x_3) \equiv (x_1 \vee_1 x_2) \wedge_1 (x_1 \vee_1 x_3);$

9. $(x_1 \wedge_1 x_2) \vee_1 x_1 \equiv x_1;$

10. $(x_1 \vee_1 x_2) \wedge_1 x_1 \equiv x_1;$

Because of these properties and other computational considerations, R-functions in the system $R_1(\Gamma)$ are more convenient to use when the constructed R-functions do not need to be differentiable.

1.5 Differential properties of the elementary R-functions.

Using an appropriate system of R-functions from $R_\alpha^m(\Gamma)$, it is possible to construct a composite R-function from any branch such that it is m times continuously differentiable everywhere. While R-functions are interesting in their own right, most applications require that arguments of R-functions are some other real functions (not necessarily R-functions.)

Suppose we are given an R-function $y = f(x_1, x_2, \ldots, x_n)$ and we substitute for its arguments some real functions $\varphi_1, \varphi_2, \ldots, \varphi_n$. The obtained function $y = f(\varphi_1, \varphi_2, \ldots, \varphi_n)$ is not, generally speaking, an R-function. Nevertheless, by appropriately choosing a system of R-functions as a basis, certain differential properties of the composite function $f(\varphi_1, \varphi_2, \ldots, \varphi_n)$ can be guaranteed.

Let $f(x_1, x_2, \ldots, x_n)$ be some R-function which is a composite of functions $x_1 \wedge_\alpha x_2$, $x_1 \vee_\alpha x_2$, and $\overline{x}$. Suppose that the argument x_1 appears in expression $f(x_1, x_2, \ldots, x_n)$ only once. Let us fix the values of all other arguments in f. Then, in order to compute the value of $f(x_1)$, it is necessary to perform a number of operations, possibly including several R-negations of x_1, or of sub-expressions containing x_1. The number of such R-negations will be called the *inversion degree* of the argument x_1. For example, in $y = \overline{\overline{x_1 \wedge_\alpha \overline{(x_2 \vee_\alpha \overline{x}_3)}}}$ the inversion degree of x_1 is 1, the inversion degree of x_2 is 2, and the inversion degree of x_3 is 3.

Theorem 3 *Let R-function $f(x_1, x_2, \ldots, x_n)$ be a composite of the R-functions $x_1 \wedge_\alpha x_2$, $x_1 \vee_\alpha x_2$, and $\overline{x}$ such that argument x_1 appears in f only once and has the inversion degree m. Suppose functions $\varphi_1, \varphi_2, \ldots, \varphi_n$, and f are continuously differentiable and satisfy the following conditions at point M_0*

$$\varphi_1(M_0) = 0; \quad \varphi_i(M_0) \neq 0, \ i = 2, 3, \ldots, n;$$

$$f(\varphi_1, \varphi_2, \ldots, \varphi_n)\Big|_{M_0} = 0.$$

Then, for any vector direction l, the following equality holds

$$\left[\frac{\partial}{\partial l} f(\varphi_1, \varphi_2, \ldots, \varphi_n)\right]\Bigg|_{M_0} = (-1)^m \left(\frac{\partial \varphi_1}{\partial l}\right)\Bigg|_{M_0}.$$

In particular, suppose that $\varphi_1, \varphi_2 \in C^1$ are real functions. Then the following special cases of the Theorem 3 are easier to visualize.

- If $\varphi_1(M_0) = 0$, and $\varphi_2(M_0) > 0$, then, for any vector direction l,

$$\left[\frac{\partial}{\partial l}(\varphi_1 \wedge_\alpha \varphi_2)\right]\Bigg|_{M_0} = \left(\frac{\partial \varphi_1}{\partial l}\right)\Bigg|_{M_0}.$$

- Similarly, if $\varphi_1(M_0) = 0$, and $\varphi_2(M_0) < 0$, then, for any vector direction l,

$$\left[\frac{\partial}{\partial l}(\varphi_1 \vee_\alpha \varphi_2)\right]\Bigg|_{M_0} = \left(\frac{\partial \varphi_1}{\partial l}\right)\Bigg|_{M_0}.$$

Further generalizations [Rva82] and additional results concerning behavior of second-order derivatives of some R-functions [Rva67] have also been proved, but are not presented here.

1.6 Other partitions of the real axis.

We have chosen the sign of a real number as the criterion for partition of real numbers, but it is not obvious what the "correct" partition Γ of the real axis is. Specifically, consider three such partitions:

$$\Gamma_2 = \{(-\infty, 0), \ [0, +\infty)\};$$

$$\begin{aligned}\Gamma_2^0 &= \{(-\infty, 0],\ [0, +\infty)\};\\ \Gamma_3 &= \{(-\infty, 0),\ 0,\ (0, +\infty)\}.\end{aligned}$$

All three seem to be suitable, since all distinguish between the positive and the negative real numbers. But they are different in handling of zero. Does it matter? Turns out that it does, because each partition determines the set of associated R-functions $R(\Gamma)$ and their properties.

The sets $R(\Gamma_2)$, $R(\Gamma_2^0)$ and $R(\Gamma_3)$ intersect [Rva74, page 57]. For example, functions $y_1 = x_1 + x_2 - |x_1 - x_2|$ and $y_2 = l^{x_1} + l^{x_2} + l^{x_3}$ belong to each of these sets, i.e. y_1 and y_2 are R-functions for each of the above partitions. At the same time, function $y = x_1 x_2$ is in $R(\Gamma_3)$ and $R(\Gamma_2^0)$ but is not in $R(\Gamma_2)$, and function $y = x_1^2 x_2^2 (1 - x_1)^2$ is in $R(\Gamma_2)$ but is not in $R(\Gamma_3)$, and so on.

It can also be shown that some branches of both $R(\Gamma_2)$ and $R(\Gamma_3)$ do not contain any continuous R-functions. This will turn out to be an important consideration for applications described here, and is one of the main reason for choosing Γ_2^0 as the "standard" partition.

One may wonder why we would bother with Γ_3 to begin with. Note that Γ_3 partitions the real axis into three intervals, not two. This forces one to use 3-valued logic for the companion functions as opposed to the Boolean functions. (In fact, this notion is generalized to a function of k-valued logic in [Rva82]. Essentially, such a partition allows one to specifically distinguish zero value from all other values, which is not possible with either Γ_2 or Γ_2^0 and may be important for some applications. On the other hand, $R(\Gamma_3)$ contains some R-functions with "undesirable" properties and 3-valued logic brings complications of its own [Rva82]. To make a long story short, it turns out that all *continuous* functions in $R(\Gamma_3)$ are also in $R(\Gamma_2^0)$. Thus, we rely on Boolean algebra and use only R_2^0-functions, but occasionally treat them as R_3-functions. This allows us to use 3-valued logic in order to identify and to *rule out* any situations where zero values may cause anomalies or ambiguities.

These and related issues are formally studied in [Rva82], [RR79], and [Rva74]. On the other hand, only Γ_2^0 was used in [Rva67], while Γ_2 was employed in [RKSU73] and [RS76] which are more concerned with applications.

2 The inverse problem of analytic geometry

2.1 Problem formulation

In his 1637 treatise "Geometry" Descartes suggested the method of coordinates to establish the correspondence between geometric objects (points, lines, bodies, etc.) and analytical objects (numbers, equations, inequalities, etc.). Since then, for various historical reasons, much attention has been paid to the so-called *direct* problem of analytic geometry: the study of curves and surfaces specified by given equations. This also led to the discovery and development of differential and integral calculus.

The *inverse* problem, i.e., the description of specified geometric objects with equations, was also posed and solved for a small set of objects of a very simple form: straight line, circle, conic sections, second-degree surfaces, and some others. Such objects are typically described by an equation $f = 0$, where f is a polynomial in two or three variables. Polynomials is the simplest class of functions that are built using just constants, addition, and multiplication. Yet significant difficulties are encountered when one tries to study or solve equations with functions from even this simplest class. This, seemingly, leaves little hope that the inverse problem can be solved for more general objects. However, it turns out that by introducing certain additional operations it is possible to significantly increase the descriptive power of the analytical method. Below, we describe how to construct an equation for geometric objects of practically arbitrary shape, thus solving the inverse problem of analytic geometry, at least in principle.

2.2 Functions for object description

The set of all points in space E^n where a function $y = f(x_1, \dots, x_n)$ takes on zero values will be called a *drawing* described by the equation $f(x_1, \dots, x_n) = 0$. Similarly, the set of all points in the Euclidean space E^n where a function $f(x_1, \dots, x_n)$ takes on non-negative values is called a *region* described by the inequality $f(x_1, \dots, x_n) \geq 0$. Clearly, these definitions are somewhat arbitrary. We can similarly talk about geometric objects described by $f > 0$, $f < 0$, $f \leq 0$, $f \neq 0$, and so on [Rva74].

If we restrict our choices to the functions f from the set C^0 of all continuous functions in E^n, the corresponding set of drawings will include all closed sets in E^n [Rva74, page 33]. Many of such closed sets are not interesting for our applications, e.g. it may be difficult to draw an arbitrary infinite set of points containing all its limit points. Thus the space of all continuous functions is too big, and so are the spaces C^n (n times continuously differentiable functions) and C^∞.

On the other hand, the set C^A of real functions that are analytic everywhere is too small. For example, it does not contain any functions describing a rectangle [Rva82, page 144]. The same limitation applies to the set of rational functions that describe the so-called *algebraic* drawings of curves and surfaces, e.g. straight line, circle, plane, hyperbola, quadric surfaces, etc.

Since the algebraic drawings include many familiar simple geometric objects, and the corresponding polynomial functions have been extensively studied, it make sense to think of them as "building blocks", or *primitives*. If we allow these primitives to be combined using the logical operations ("and" $\wedge$, "or" $\vee$, "not" $\neg$, etc.), most geometric objects of practical interest can be described. Such sets are called *semi-algebraic sets*, and their drawings are called semi-algebraic drawings. Among others, the semi-algebraic sets include unions of semi-algebraic drawings and any sets whose points satisfy systems of algebraic equations and inequalities. At the same time, there are semi-algebraic sets that are *not* algebraic; such sets cannot be described by (a system of) equations. In other words, in terminology of [Rva82], the system of algebraic functions is not *algorithmically complete.*

We could expand the set of primitives to include all *elementary* sets defined by any arithmetic, logarithmic, and trigonometric functions. Boolean combinations of such primitives result in *semi-elementary* sets and drawings respectively. It is easy to show that the system of elementary functions is algorithmically complete, i.e. any semi-elementary set is also an elementary set. Yet, despite their great usefulness and broad application, the semi-elementary sets suffer from a serious disadvantage: their drawings are not specified by equations of the form $f = 0$. As a result, using the geometric information analytically in various applications can be complicated and difficult. The problem is solved using R-functions; the following sections show how to construct appropriate equations and inequalities for arbitrary semi-elementary sets.

2.3 From Boolean expressions to real functions

Suppose we are given a description of a geometric object $D \subseteq E^n$

$$D = F[(\phi_1 \geq 0), \ldots, (\phi_m \geq 0)], \tag{1}$$

where real-function inequalities $\phi_i(x_1, \ldots, x_n) \geq 0$ define primitive geometric regions, and F is a set function constructed using standard set operations $\cap, \cup, -$ on these primitive regions. Alternatively, replacing the set operations by the corresponding logical functions $\wedge, \vee, \neg$, we can view F as a Boolean logic function. The equation 1 becomes a predicate equation

$$F[S_2(\phi_1), \ldots, S_2(\phi_m)] = 1, \tag{2}$$

which holds for all points $p \in D \subseteq E^n$.

We seek a single real-function inequality $f(x_1, \ldots, x_n) \geq 0$ that defines the composite object D. Note that Eqs. 1 and 2 may define an arbitrary subset of E^n, while any inequality $f \geq 0$ defines a *closed* subset of E^n. Hence, we restrict our attention to those Boolean functions that map closed sets into closed sets, called *closing* functions. It is well known that the class of all such functions is functionally closed. Then we have the following general result [Rva74, page 97]:

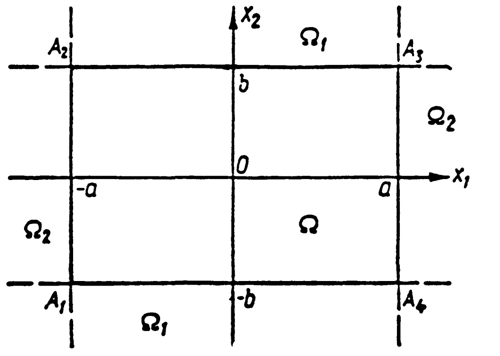


Figure 1: Rectangular region $A_1A_2A_3A_4$ is defined by the intersection of the horizontal and vertical strips

Theorem 4 *Let $F(X_1, \ldots, X_m)$ be a closing Boolean function that is the companion of a continuous R-function $f(x_1, \ldots, x_m)$. Then for any continuous real functions ϕ_i, $i = 1, 2, \ldots, m$, and a closed region $D \subset E^n$ given by Eq. 2, D can be defined by the inequality*

$$f(\phi_1, \ldots, \phi_m) \geq 0. \tag{3}$$

In other words, to obtain a real function defining the region D constructed from primitive regions $\phi_i \geq 0$, it suffices to construct an appropriate R-function and substitute for its arguments the real functions ϕ_i defining the primitive regions. To see that the theorem 4 is true observe that

$$S_2[f(\phi_1, \ldots \phi_m)] = F[S_2(\phi_1) \ldots, S_2(\phi_m)].$$

Therefore, as the predicate, Eq. 3 is true for any point $p \in D$ and is false for all points $p \notin D$. Thus, given a definition of any complex object as a Boolean combination of some primitives, a single real-function inequality can be constructed, using any sufficiently complete system of R-functions described in section 1.3.

Example 2.1 Suppose region Ω is a vertical strip $\omega_1 \equiv a^2 - x_1^2 \geq 0$, and Ω_2 is a horizontal strip $\omega_2 \equiv b^2 - x_2^2 \geq 0$ (Figure 1). Let us use R-conjunction $u_1 \wedge_0 u_2$. The inequality

$$(a^2 - x_1^2) \wedge_0 (b^2 - x_2^2) \geq 0$$

defines the rectangular region $A_1A_2A_3A_4$ formed by the intersection of the two strips. This inequality can be rewritten as

$$a^2 + b^2 - x_1^2 - x_2^2 - \sqrt{(a^2 - x_1^2)^2 + (b^2 - x_2^2)^2} \geq 0.$$

On the other hand, if we use the R-conjunction $u_1 \wedge_1 u_2$, the inequality defining the rectangular region becomes

$$a^2 + b^2 - x_1^2 - x_2^2 - |a^2 - b^2 - x_1^2 + x_2^2| \geq 0.$$

Similarly, we can obtain an inequality using m times continuously differentiable R-conjunction $u_1 \wedge_0^m u_2$.

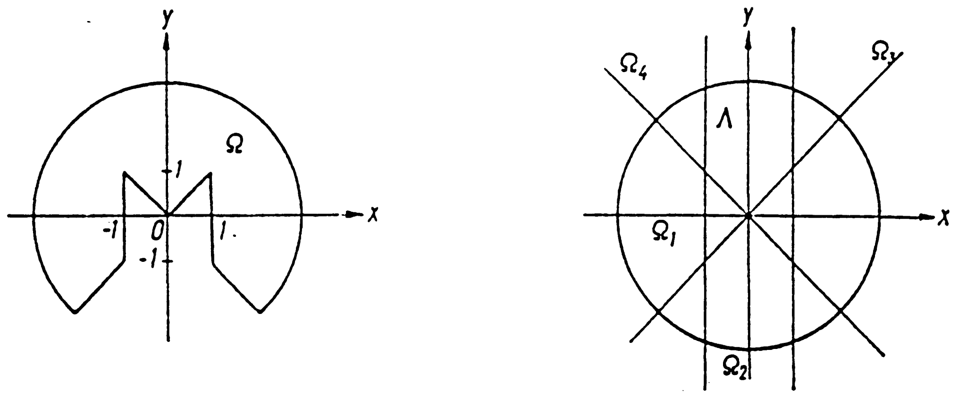


Figure 2: Region Ω can be defined by a Boolean combination of the primitive regions.

From now on we will use R-conjunctions and R-disjunctions as elementary functions, i.e. we will just write $u_1 \wedge_\alpha^m u_2$, $u_1 \vee_1 u_2$, etc. We know how to evaluate and differentiate these functions, and the notation becomes much simpler.

Example 2.2 Let us write an equation for the boundary of the region shown in Figure 2. First choose the primitive regions as:

$$\begin{aligned}
\Omega_1 &= (9 - x^2 - y^2 \geq 0) && \text{– disk of radius 3,}\\
\Omega_2 &= (x^2 - 1 \geq 0) && \text{– vertical strip,}\\
\Omega_3 &= (y - x \geq 0) && \text{– halfplane,}\\
\Omega_4 &= (y + x \geq 0) && \text{– halfplane.}
\end{aligned}$$

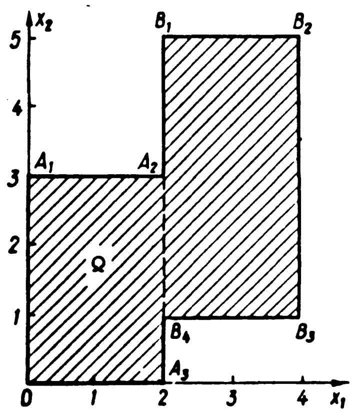
It is easy to check that the region Ω is defined by the Boolean predicate

$$[(\Omega_3 \wedge \Omega_4) \vee \Omega_2] \wedge (\Omega_3 \vee \Omega_4) \wedge \Omega_1.$$

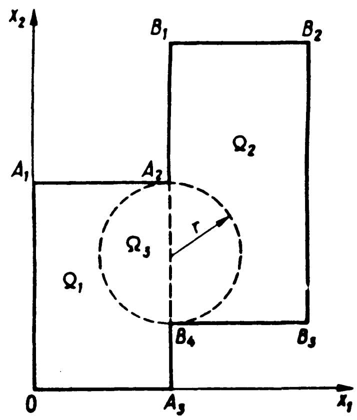
Substituting the expressions from the primitive definitions into appropriate R_0-functions, we get:

$$\Omega = \left\{ \left\{ [(y-x) \wedge_0 (y+x)] \vee_0 (x^2 - 1) \right\} \wedge_0 \left\{ [(y-x) \vee_0 (y+x)] \wedge_0 (9 - x^2 - y^2) \right\} \geq 0 \right\}.$$

Theorem 4 assumes that a region is specified as a Boolean function F of m primitive regions. Intersection of these primitives partition E^n into a number of regions, which is usually much smaller than 2^m possible Boolean truth assignments. In fact, the ratio of non-empty regions in the decomposition of E^n to 2^m rapidly goes to zero, as $m \to \infty$. Thus many Boolean functions define empty sets. In other words, Boolean predicates defining regions can be viewed as partially specified logic functions with a large number of "don't care" conditions. These "don't care" conditions can be often used to simplify Boolean and real-function expressions.



(a) Union of two rectangles leads to real functions with zeros in the interior of the combined object.



(b) Zero points can be covered by unioning an additional primitive region.

Figure 3: Eliminating interior zero points

2.4 Boundaries and interiors of regions

In the examples 2.1 and 2.2, replacing the inequality sign $\geq$ by equality $=$ results in an equation of the boundary of the defined regions. However, in general, this is not true. Just because we have found $f(x_1, \ldots, x_m) \geq 0$, it does not automatically imply that $f(x_1, \ldots, x_m) = 0$ is the *desired* equation of the drawing. For example, set S defined by $f(x_1, \ldots, x_m) = 0$ can contain points interior to S. This ambiguity is explained by the inclusion of zero in the set of positive numbers when we have partitioned the real axis. Such situations can be controlled (and, if necessary, eliminated) by the choice of the primitive real functions and disallowing certain 3-valued logic functions [Rva82, page 150].[2]

Figure 3(a) shows a more interesting example, where points satisfying $f(x, y) = 0$ lie in the interior of the *region* defined by $f(x, y) \geq 0$. Here, the drawings of the primitive regions (the two rectangles) define their respective boundaries, but the drawing of the combined region does not correspond to its boundary. Figure 3(b) shows that the interior zero points can be "covered" by unioning an additional primitive region. Similarly, a region defined by $f \geq 0$ may not contain any points satisfying $f > 0$, i.e. there are only boundary points.

Automatic detection of such situations is a difficult problem. In some cases, these "peculiarities" can be eliminated by constructing an alternative Boolean expression that uses the

[2]This requires a different partition Γ_3 of the real axis and transition to 3-valued logic, as described in section 1.6.

same primitives. But, more generally, additional primitive regions may be required.

The above procedures assume that for any desired geometric object S, a predicate (in Boolean or 3-valued logic) can be constructed using some geometric primitives for which the defining inequalities are known. Usually, this is not a problem. With the exception of the above "peculiarities", the boundary of an object S naturally suggests appropriate primitives. At other times, suitable approximations can be used. There are also special situations, where the procedure of obtaining a predicate equation (and hence the construction of the corresponding real function) can be completely automated. For example, an algorithm to construct a Boolean predicate for any two-dimensional linear polygon is given, using only halfplanes associated with the polygon's sides [RKSU73].[3]

Another common situation arises when only the boundary L of an object S need be defined by an equation $f(x_1, \ldots, x_m) = 0$, but the interior and exterior of S need not be distinguished. In this case, let L be a union of (boundary) elements $l_i, i = 1, \ldots, k$, and suppose we construct a predicate and real function equation $f_i = 0$ defining every element l_i. For example, an equation of a straight-line segment connecting points (x_1, y_1) and (x_2, y_2) is given by

$$\begin{aligned}
\phi(x,y) &= \Big\{[(x-x_1)(y_2-y_1)-(y-y_1)(x_2-x_1)]^2 + \\
&+ \Big[\frac{d^2}{4} - \Big(x - \frac{x_1+x_2}{2}\Big)^2 - \Big(y - \frac{y_1+y_2}{2}\Big)^2\Big]^2\Big\}^{1/2} - \\
&- \frac{d^2}{4} + \Big(x - \frac{x_1+x_2}{2}\Big)^2 + \Big(y - \frac{y_1+y_2}{2}\Big)^2]^2 = 0,
\end{aligned}$$

where d is the length of the line segment [Rva74, page 128]. Function ϕ is zero at all points lying on the segment and is strictly positive everywhere else. Similarly, one can write an equation for a circular arc, or for a planar face (polygon in E^3) of a polyhedron, but this task becomes difficult for more complicated boundaries. Union of elements l_i defines the desired boundary L, and there are several methods to derive an equation of L, given the primitive functions $f_i, i = 1, \ldots, k$. The easist method is to define L by $\prod_i^k f_i = 0$, but this function may not be attractive numerically; many other choices are possible, e.g. any R-conjunction of f_i's also defines L.

2.5 Equations of trimmed curves, surfaces, and regions

Sometimes it is desirable to derive a function such that $f(x_1, \ldots, x_n) = 0$ describes a segment of a curve, or some portion of a surface. More generally, we could be interested in some part of an $(n-1)$-dimensional hypersurface lying inside an n-dimensional region in E^n.

Suppose, for example, we need an equation of a segment of the curve $f(x,y) = 0$ lying inside the region $\varphi(x,y) \geq 0$ (Figure 4). The curve defined by $f = 0$ can be considered as an intersection of two regions $f \geq 0$ and $-f \geq 0$, and defined by an equality $-f^2 \geq 0$; then

[3] As far as I know, this reference precedes the recursive "Decreasing Convex Hull" algorithm that became popular in 1980s in the Western literature.

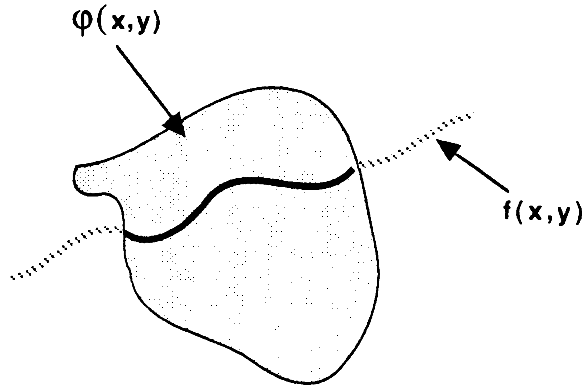


Figure 4: The intersection of the curve $f(x,y) = 0$ and a region $\varphi(x,y) \geq 0$ defines a segment of the curve.

the segment can be considered as an intersection of the curve and the region $\varphi \geq 0$. Thus we can write an equation for the curve segment as the intersection of the two regions:

$$(-f^2) \wedge_\alpha^m \varphi \geq 0. \tag{4}$$

But the left-hand side of Eq. 4 is identically zero for all points of the curve segment; hence we can we can replace sign $\geq$ by $=$ to get an equation of the curve segment. Similarly one derives an equation $f(x,\ldots,x_n) = 0$ defining a portion of a curve, surface, or hypersurface lying inside an n-dimensional space.

An n-dimensional region D is usually defined by an inequality $\varphi(x_1, x_2, \ldots, x_n) \geq 0$. In order to specify the same region D by an *equation,* consider it in $(n+1)$-dimensional space $\{z, x_1, x_2, \ldots, x_n\}$, as the intersection of the hyperplane $z = 0$ with the the cylinder $\varphi = 0$. Thus, for example,

$$z \wedge_0 \varphi = z + \varphi - \sqrt{z^2 + \varphi^2}.$$

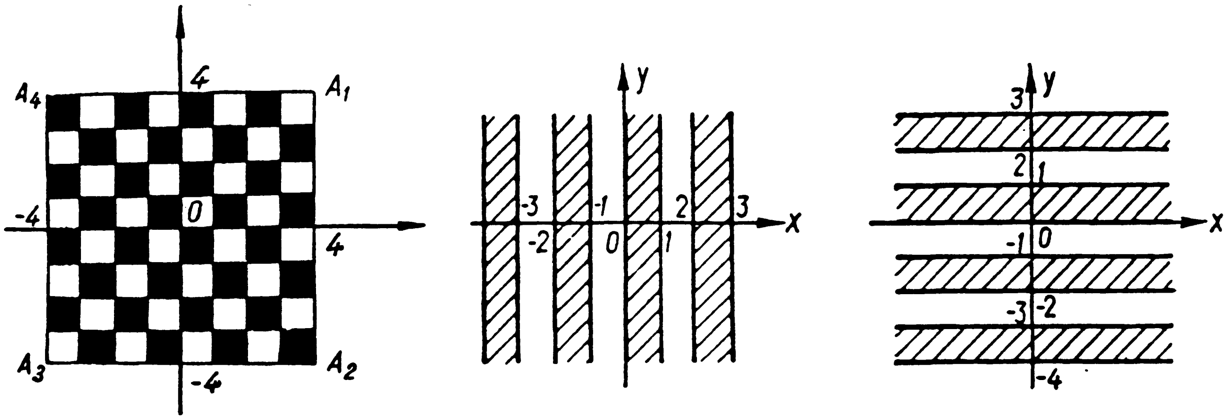
Setting $z = 0$ we get

$$\varphi - |\varphi| = 0 \tag{5}$$

as an equation for an arbitrary region D defined by $\varphi \geq 0$. Similar procedure can be used to obtain an m-times differentiable equation for D.

Example 2.3 Let us write an equation for a chess board (Figure 5a.) The equation should be satisfied everywhere inside the dark squares as well as at all the points of the boundary. Suppose the size of the squares is unity. For primitive regions we select

$$\begin{aligned} D_1 &= (sin\pi x \geq 0) \\ D_2 &= (sin\pi y \geq 0) \\ D_3 &= 32 - x^2 - y^2 - |x^2 - y^2| \geq 0 \end{aligned}$$



(a) (b) (c)

Figure 5: Constructing an equation for a chess board.

The region D_1 is a system of vertical strips (see Figure 5b) intersecting x-axis at segments $[2k, 2k+1]$, $(k = 0, \pm 1, \pm 2, \ldots)$. Similarly the region D_2 is a system of horizontal strips (Figure 5c). The region D_3 defines the square region $A_1\ A_2 A_3 A_4$ corresponding to the whole board. The region of interest to us can be defined as

$$[(D_1 \wedge D_2) \vee (\overline{D}_1 \wedge \overline{D}_2)] \wedge D_3.$$

Noticing that $(D_1 \wedge D_2) \vee (\overline{D}_1 \wedge \overline{D}_2) = D_1 \Leftrightarrow D_2$, the above Boolean predicate becomes

$$(D_1 \Leftrightarrow D_2) \wedge D_3.$$

Then the inequality

$$(\sin \pi x\ sin \pi y) \wedge_\alpha (32 - x^2 - y^2 - |x^2 - y^2|) \geq 0$$

defines the region D. Applying Eq. 5, we obtain the equation for the chess board

$$(\sin \pi x\ sin \pi y) \wedge_\alpha (32 - x^2 - y^2 - |x^2 - y^2|) - \left|(\sin \pi x \sin \pi y) \wedge_\alpha (32 - x^2 - y^2 - |x^2 - y^2|)\right| = 0.$$

2.6 Normal and normalized equations

There is an infinite number of equations that define a given drawing. Specific situations often require that these equations possess various additional properties. Suppose L is a drawing in E^d and $p \in E^d$ is an arbitrary point. Then the *distance* from p to the drawing L is defined as

$$\rho(p, L) \equiv \inf \rho(p, q), \quad q \in L \tag{6}$$

where $\rho(p,q)$ is the distance between the points p and q. If for every point $p \in E^d$ $f(p) = \rho(p,L)$, then $f(x_1,\ldots,x_n) = 0$ is called a *normal* equation, and function f is called the normal function of a drawing L.

Example 2.4 A normal equation of a line in E^2 is given by

$$|x\cos\alpha + y\sin\alpha - D| = 0,$$

where D is the distance to the line from the origin, and α is the angle between the x-axis and the direction in which D is measured. The distance from any point $p = (x,y)$ to the line is given by $\rho = |x\cos\alpha + y\sin\alpha - D|$. Similarly, a normal equation of a plane in E^3 is given by

$$|x\cos\alpha + y\cos\beta + z\cos\beta - D| = 0,$$

and a normal equation of a circle in E^2 with radius R and centered at a point (a,b) is

$$|\sqrt{(x-a)^2 + (y-b)^2} - R| = 0.$$

If f_1 and f_2 are normal functions for drawings L_1 and L_2 respectively, then

$$f_1 \wedge_1 f_2 \equiv \min(f_1, f_2) \tag{7}$$

is the normal function for the drawing $L = L_1 \cup L_2$. Normal equations for simple objects such as line segments and circular arcs E^2 and polygons positioned in E^3 can be constructed using appropriate R-functions. These simple objects can be viewed as standard boundary elements that can be "glued" together to form a boundary of more complex objects, whose normal equation is obtained using Eq. 7.

Normal functions are important in many applications, notably for constructing admissible approximations to solutions of boundary value problems (see section 3.3). There are at least two reasons preventing their use:

- Determining normal equations for more general objects is a difficult problem. For example, this problem reduces to solving a fourth order algebraic equation for an ellipse, and transcendental equations for logarithmic and sinusoidal curves;
- A normal function for a drawing L may have discontinuous partial derivatives at points that are equidistant from L, which makes it unsuitable for many applications.

Thus, we are led to the notion of *normalized* function (and equation) for a drawing, which is, in some sense, an approximation of a normal function that can be constructed for complex objects and has desired differentiable properties. Specifically, let L be a drawing and let ν be a normal vector at some regular[4] point $p \in L$. Then equation $w(x) = 0$ of L is called *normalized to m-th order* if function $w(x)$ satisfies

$$w(x) = 0|_L, \quad \frac{\partial w}{\partial \nu}\Big|_L = 1, \quad \frac{\partial^k w}{\partial \nu^k}\Big|_L = 0 \ (k = 2,3\ldots,m) \tag{8}$$

[4]A point $p \in L$ is regular if there is a unique normal to L at p.

at all regular points of L. Note that the normal function of L is normalized to an arbitrary order. The conditions 8 mean that function w behaves approximately as the distance function ρ defined in Eq. 6 along the normal ν in the neighborhood of L.

There are several methods of the first order normalization. Suppose a function $w \in C^m$, $w = 0$, and $\frac{\partial w}{\partial \nu} > 0$ on L. Then the function

$$w' \equiv w(w^2 + \|\nabla w\|^2)^{-\frac{1}{2}} \in C^{m-1} \tag{9}$$

is normalized to the first order, i.e. $\frac{\partial w'}{\partial \nu} = 1$. In practice, this method of normalization may be difficult to use for complicated functions. Fortunately, many R-functions preserve normalization properties. For example, let $w_1 \geq 0$ and $w_2 \geq 0$ be inequalities defining regions Ω_1 and Ω_2 respectively such that w_1 and w_2 are normalized to the first order. Then it can be shown that the functions $w_1 \wedge_\alpha w_2$ and $w_1 \vee_\alpha w_2$ are also normalized to the first order.

A number of techniques for normalization to the m-th order are also known [Rva82].

3 Applications

3.1 Problems of mathematical programming.

A typical mathematical programming problem is that of finding a point $x^0 = (x_1^0, \ldots, x_n^0)$ in some n-dimensional region $\Omega \in E^n$ where some function $f(x) = f(x_1, \ldots, x_n)$ reaches a maximum or a minimum value. The region Ω is usually specified as a union of N systems of inequalities

$$\sigma_{ki}(x_1, \ldots, x_n) \geq 0, \quad i = 1, \ldots, M,$$

where each $\sigma_{ki} \geq 0$ defines a region Σ_{ik}, and $k = 1, \ldots, N$. In other words, the region of interest is given as

$$\Omega = \bigcup_k (\bigcap_i \Sigma_{ki}), \quad i = 1, \ldots, M, \quad k = 1, \ldots, N. \tag{10}$$

Using R-functions, Ω can be defined by a single inequality

$$w(x) \equiv \overset{*}{\bigvee_k} (\overset{*}{\bigwedge_i} \sigma_{ki}) \geq 0`, \quad i = 1, \ldots, M, \quad k = 1, \ldots, N, \tag{11}$$

where $\wedge^*, \vee^*$ are any R-conjunctions and R-disjunctions respectively. Thus, any such problem of mathematical programming can be reformulated as follows: find a point x^0 in a region Ω defined by a single inequality $w(x) \geq 0$, such that $f(x)$ is maximum (or minimum).

In practice, it is often possible to estimate that the desired value of the goal function $f(x)$ is $z_0 \in [z_{\min}, z_{\max}]$. In that case, R-functions can be used to transform the original problem into a sequence of the unconstrained optimization problems. Let us define a region $\Sigma_0(z)$ defined by inequality

$$z - f(x) \geq 0, \quad z_{\min} \leq z \leq z_{\max}, \tag{12}$$

for every fixed value of z, and consider a region $Q(z) \equiv \Omega \cap \Sigma_0(z)$ defined by an inequality

$$q(x, z) \equiv w(x) \wedge_\alpha [z - f(x)] \geq 0. \tag{13}$$

If $z > z_0$, then the region $Q(z)$ has interior points, and if $z < z_0$, then $Q(z)$ is an empty set $\emptyset$. Assuming that $Q(z)$ is bounded, and w, f are continuous functions, it follows that the maximum value of the goal function $f(x)$ is achieved when $q_0(z) = \max_x q(x, z) = 0$. Note that $q_0(z)$ is monotone in z. Thus, the original problem is transformed into a sequence of the optimization problems (each problem is defined by a fixed value of $z \in [z_{\min}, z_{\max}]$) of the function $q(x, z)$ without any constraints on x. Additional details can be found in [Rva67, Rva82].

3.2 Optimal placement of geometric objects.

The following material was first discussed in [Rva67], further developed in [Sto75, SP78, SS80], and is summarized in [Rva82].

Many practical problems belong to the class of problems in optimal placement of geometric objects. Among them: maximizing the number of objects of similar shape that can be cut out from a piece of material,[5] minimizing the quantity of material necessary to produce certain shapes, various optimal packaging problems, and so on. Consider the following two-dimensional example of such a problem.

Given a planar piece of material that is fixed in the global coordinate system $\{x, y\}$, let Ω_1 be a rigid planar object that is fixed at some location and is defined by the inequality $\Omega_1 = [w_1(x, y) \geq 0]$, such that $w_1 > 0$ in the interior of Ω_1, $w_1 = 0$ for points of the boundary $\partial\Omega_1$, and $w_1 < 0$ outside of Ω_1. The object has three degrees of freedom in the plane: x_1, y_1 determine the position of Ω_1 with respect to the global coordinate system, and θ_1 defines its orientation. The same object Ω_1 in *general location* can be defined by

$$\Omega_1 = \{w_1[(x - x_1)\cos\theta_1 + (y - y_1)\sin\theta_1; \quad -(x - x_1)\sin\theta_1 + (y - y_1)\cos\theta_1] \geq 0\}. \qquad (14)$$

Consider now two such objects, Ω_1 and Ω_2, that are to be placed on a single sheet Ω. It is often important to define conditions under which the two objects do not overlap (i.e. the interiors of the two objects to not have any points in common). Suppose they do overlap. Then the region of their intersection $S = \Omega_1 \cap \Omega_2$ is defined by some function $\varphi = w_1 \wedge_\alpha^m w_2$, which is non-negative in S and is negative outside of S. It follows that the two objects do not overlap if and only if S is either empty or contains only boundary points. In other words, we can define the *non-overlap* condition as

$$-\max_{(x,y)} \varphi(x, y, x_1, y_1, \theta_1, x_2, y_2, \theta_2) \geq 0, \qquad (15)$$

where x_i, y_i, θ_i are position and orientation of Ω_i. If both Ω_1 and Ω_2 are convex it is not difficult to construct equations so that the function φ in the inequality (15) has a unique global maximum, which can be computed numerically starting from an arbitrary point. Otherwise, there may be multiple extrema, in which case the problem is much more difficult.

Given a number of objects, they can be "glued" together by taking their union (i.e. by applying R-disjunction). Because pairwise non-overlap conditions must be satisfied simultaneously, the global non-overlap condition (for all objects involved) can be written as an R-conjunction of all the non-overlap conditions. In particular, sufficiently good approximations of various geometric objects can be obtained as unions (covers) of circular balls. In this case, the problem simplifies significantly, because the non-overlap conditions can be written in a closed-form, and the search for the extrema is not required. Closed-form expressions for non-overlap conditions can be also written in other cases, for instance, for objects bounded by linear and circular arc segments.

Straightforward generalizations of the above methods allow us to formulate constraints for problems involving variable-form material and n deformable objects $\Omega_1, \ldots, \Omega_n$ whose shapes may be determined by additional parameters. Placement of n objects of similar shape often leads to geometric problems with high degree of periodicity and symmetry, which can be

[5] Efficient nesting of blanks to minimize the amount of sheet metal scrap is one important example of such problems.

taken into account during the construction of the corresponding real functions [Rva82]. R-functions can also be useful for defining objective functions whose extrema correspond to the best (in some sense) placement of the geometric objects, e.g. to minimize distances between objects, or the amount of "scrap" on the piece of material.

These techniques allow us to pose problems in optimal placement as problems of mathematical programming which then can be reformulated using the methods described above in section 3.1, and solved using various numerical algorithms.

3.3 Boundary value problems

Probably the most important area of application of R-functions to date is in constructing approximate solutions for boundary value problems. For example, [Rva67] describes application to problems of elastic torsion and contact, [RKSU73] and [RK87] are completely devoted to problems of plate bending and vibration, [RS76] deals mostly with problems of thermodynamics, and [Rva82] contains references to problems on electrostatics and magnetism. A family of software systems that generate solutions of physical field problems from high-level mathematical descriptions are described in [RMS86], [RM83], and [RK87].

3.3.1 Bundles of functions satisfying boundary conditions

A drawing in E^n can be specified by an infinite number of different equations. If $\partial\Omega = [w(x) = 0]$ is some drawing, and Φ is a bounded sign-preserving function, then $\partial\Omega = [w\Phi = 0]$. More generally, if $\Phi : E^n \rightarrow E^m$,

$$u = w\Phi, \tag{16}$$

defines the *bundle* of all possible functions equal to zero on $\partial\Omega$. Φ can be a vector or tensor function and is called the *undetermined component* of the bundle. The problem of constructing a bundle of functions equal to zero on the boundary of a given region is a special case of a more general problem of finding a bundle of functions possessing various properties at different points of some geometric object, such as coincide with some known functions, have specified normal derivatives, and so on.

Suppose that the boundary $\partial\Omega = \bigcup_i \partial\Omega_i$, where $\partial\Omega_i$ are possibly overlapping boundary pieces, and for each of them we have a boundary condition of the type

$$\varphi|_{\partial\Omega_i} = \varphi_i|_{\partial\Omega_i}. \tag{17}$$

Each $\partial\Omega_i$ can be defined by a real-valued function w_i such that $w_i = 0$ for all points of $\partial\Omega_i$ and $w_i > 0$ for all points outside of $\partial\Omega_i$. We need to construct a function φ such that it satisfies the boundary conditions (17) and is defined everywhere in Ω. In other words, we need to "glue" all these individual boundary conditions together. There are several ways of accomplishing this. For example,

$$\varphi = \frac{\varphi_1 w_1^{-1} + \cdots + \varphi_m w_m^{-1}}{w_1^{-1} + \cdots + w_m^{-1}} + w\Phi \tag{18}$$

is a bundle of functions that assume the desired values on each $\partial\Omega_i$ and is defined almost everywhere in Ω (except corner points).

Equation (18) is an example of the Langrange-Taylor-Hermite interpolation formulas where desired functions and their derivatives are specified not only at individual points, but also on curves, surfaces, or hypersurfaces. [Rva82] discusses other interpolation techniques resulting in bundles of functions satisfying various differential properties, as well as their relationship to the classical methods.

3.3.2 Structure of solutions for boundary value problems

The ability to construct bundles of functions that satisfy given boundary conditions is important for approximating solutions of boundary value problems. Typically, a boundary value problem can be reduced to determination of some function u such that in some region Ω it satisfies

$$Au = f, \tag{19}$$

with boundary conditions specified on the boundary $\partial\Omega = \bigcup_i \partial\Omega_i$ as

$$L_i u = \varphi_i, \text{ on } \partial\Omega_i \; i = 1, \ldots, m, \tag{20}$$

where A and L_i are some differential operators, f and φ_i are functions defined in Ω and on $\partial\Omega$. If a solution u can be found such that it satisfies all equations (19) and (20), then u is the exact solution of the boundary value problem.

There are very few practical cases for which the exact solutions are known to exist; instead, we usually seek a function that approximates the exact solution in some sense. Among variational or projective approximation methods, popular are the so-called direct methods which reduce a boundary value problem to the solution of a system of algebraic equations. All such methods approximate the solution of a boundary value problem with a function of the form

$$u_n = \sum_{k=1}^{n} c_k g_k + \varphi, \tag{21}$$

where function φ is chosen so that $u_n = \varphi$ on $\partial\Omega$ satisfies the appropriate boundary conditions of (20), functions g_k form a complete and linearly independent system of functions, and coefficients c_k are determined by solving the system of linear equations. The differences between various direct methods (e.g., Ritz/Galerkin, Finite-Element, Courant, etc.) are determined by (1) choice of the functions g_k, and (2) method of computing coefficients c_k. These differences are important because they determine the ease of constructing the approximate solution, as well as numerical stability and convergence of the solution method. We will see now that R-functions significantly simplify and unify the techniques available for construction of admissible approximations in Eq. (21).

Consider the simplest Dirichlet boundary value problem, with homogeneous boundary conditions: Eq. (20) is simply

$$u|_{\partial\Omega} = 0.$$

Let w be a real function satisfying

$$w > 0 \ \text{ in } \Omega, \qquad w = 0 \ \text{ on } \partial\Omega.$$

Then choose approximation of u from the bundle of functions defined in Eq. (16) as

$$u_n = w \sum_{k=1}^{n} c_k \Phi_k, \tag{22}$$

where undetermined functions $\Phi_1, \Phi_2, \ldots, \Phi_n$ can be chosen from the complete system of functions, such as from a sequence of polynomials[6] (algebraic, trigonometric, or other classical), or various functions with finite support (e.g. splines). If the boundary conditions are not homogeneous, but are specified by a system of Eqs. (17), the admissible approximation can be constructed using Eq. (18) in the form

$$u_n = \frac{\varphi_1 w_1^{-1} + \cdots + \varphi_m w_m^{-1}}{w_1^{-1} + \cdots + w_m^{-1}} + w \sum_{k=1}^{n} c_k \Phi_k.$$

Introduction of derivatives in the boundary conditions significantly complicates the structure of the solution. For instance, the Neumann boundary conditions are given as

$$\left.\frac{\partial u}{\partial \nu}\right|_{\partial\Omega} = \varphi,$$

where ν is a normal to $\partial\Omega$ (if it is defined). Generally speaking, it is not enough to "glue" the boundary conditions together as for the Dirichlet problem; we may also need to assure the existence of these derivatives in the domain Ω, may have to normalize the admissible approximation, and so on. Constructions for various boundary value problems (including mixed boundary conditions, tangential derivatives, fourth-order differential equations, and many others) have been studied in great detail in [Rva74] and [Rva82].

The above considerations lead to a notion of the *structure* of the solution of a boundary value problem. It is an expression

$$u = B(\Phi, w, \varphi),$$

which satisfies boundary conditions φ prescribed on the boundary defined by w, and where Φ is the undetermined component of the solution. The bundles of functions defined in Eqs. (16) and (18) are simple examples of such structures. Structures for many common boundary value problems have been constructed [Rva82, RM83]. More generally, the theory of R-functions allows us to construct solution structures for problems with intricate boundary conditions, and to account for known behavior near singularities. The concept of solution structure seems to unify the various direct methods for solving boundary value problems and make explicit their differences. For example, the choice of undetermined functions Φ is an important consideration affecting numerical stability and convergence of approximate solutions. These and many additional topics are studied in [Rva82].

[6] According to [RM83], such a construction was first suggested by Kantorovich [KK58].

References

[KK58] L. V. Kantorovich and V. I. Krylov. *Approximate Methods of Higher Analysis.* Interscience Publishers, 1958.

[RK87] V. L. Rvachev and L. V. Kurpa. *R-functions in Problems of Theory of Plates.* Naukova Dumka, 1987. In Russian.

[RKSU73] V. L. Rvachev, L. V. Kurpa, N. G. Sklepus, and L. A. Uchishvili. *Method of R-functions in Problems on Bending and Vibrations of Plates of Complex Shape.* Naukova Dumka, 1973. In Russian.

[RM83] V. L. Rvachev and G. P. Manko. *Automation of Programming for Boundary Value Problems.* Naukova Dumka, 1983. In Russian.

[RMS86] V. L. Rvachev, G. P. Manko, and A. N. Shevchenko. The R-function approach and software for the analysis of physical and mechanical fields. In J. P. Crestin and J. F. McWaters, editors, *Software for Discrete Manufacturing*, Paris, 1986. North-Holland.

[RR79] V. L. Rvachev and V. A. Rvachev. *Nonclassical Methods of Approximation Theory in Boundary Value Problems.* Naukova Dumka, 1979. In Russian.

[RS76] V. L. Rvachev and A. P. Slesarenko. *Logic Algebra and Integral Transforms in Boundary Value Problems.* Naukova Dumka, 1976. In Russian.

[Rva67] V. L. Rvachev. *Geometric Applications of Logic Algebra.* Naukova Dumka, 1967. In Russian.

[Rva74] V. L. Rvachev. *Methods of Logic Algebra in Mathematical Physics.* Naukova Dumka, 1974. In Russian.

[Rva82] V. L. Rvachev. *Theory of R-functions and Some Applications.* Naukova Dumka, 1982. In Russian.

[Shi88] A. K. Shidlovsky, editor. *Vladimir Logvinovich Rvachev.* Biobibliography of scientists of UkSSR. Naukova Dumka, 1988. In Russian.

[SP78] Yu. G. Stoian and A. A. Panasenko. *Periodic Placement of Geometric Objects.* Naukova Dumka, 1978. In Russian.

[SS80] Yu. G. Stoian and V. Z. Sokolovsky. *Solution of some multi-extremal problems by method of contracting domains.* Naukova Dumka, 1980. In Russian.

[Sto75] Yu. G. Stoian. *Placement of Geometric Objects.* Naukova Dumka, 1975. In Russian.