Integer Functions and Continued Fractions

Jeffrey O. Shallit
Department of Mathematics
Princeton University
Princeton, New Jersey
April 15, 1979

Acknowledgements

The author wishes to thank Dr. Bernard Dwork for his interest and help. Thanks go to Dr. Jerrold Tunnell for his assistance and suggestions. Finally, the author extends his sincere appreciation to Eugene McDonnell, without whom this thesis would never have been written.

This thesis is submitted towards the fulfillment of the requirements for the AB degree, Princeton University.

This represents my own work in accordance with University regulations.

Jeffrey O. Shallit

Jeffrey O. Shallit

Table of Contents

Introduction.

This thesis arose from considering the following question: How can the notion of continued fractions be extended to the complex plane?

This question has been considered several times by other writers. For example, see Hurwitz [1] and Schmidt [2].

All of the previous proposals suffer from various defects. In particular, Hurwitz's construction does not reduce to the case of ordinary continued fractions if purely real numbers are employed.

The complex continued fraction algorithm in this thesis is based on a construction of McDonnell [3] that extends the greatest integer function to the complex plane. Examination of the concept of integer function led to a useful definition which is presented in section I.1. Properties of the greatest integer function are given in section I.2. The notion of duality, which links the greatest integer function with the least integer function, is discussed in section I.3. In section I.4, I will examine the extension of integer functions to the complex plane. The chapter concludes with a discussion of the properties of McDonnell's function in section I.5.

Chapter II examines some of the theory of continued

fractions. Sections II.1 and II.2 examine the real continued fraction algorithm and the real greatest common divisor algorithm, respectively. Section II.3 examines the extension of continued fractions to the complex plane. In section II.4, the continued fractions for pure imaginary numbers will be discussed, and the complex gcd and linear Diophantine equation are examined in section II.5. In section II.6, I will discuss the convergence of infinite continued fractions. Finally, in section II.7 I will discuss some unsolved problems and conjectures.

Some comments about the notation used should be made. We use the symbol $\mathbb{Z}$ to mean the set of integers. The symbol $\mathbb{R}$ is the set of real numbers. $\mathbb{Q}$ is the set of real rational numbers. $\mathbb{C}$ is the set of complex numbers. We use $\mathbb{Z}[i]$ to mean the set of Gaussian or complex integers. Similarly, $\mathbb{Q}[i]$ is the set of complex rational numbers.

If A and B are sets, we write A - B for the set theoretic difference $\{x:x\in A,\ x\notin B\}$. If A is a set and x is a number, then by xA we mean the set $\{xa:a\in A\}$. Similarly, by x + A we mean the set $\{x+a:a\in A\}$.

We write a|b (read: a divides b) iff b/a = n, where $n\in\mathbb{Z}$ or $\mathbb{Z}[i]$, depending on the domain in question.

The symbol |A| is used in several ways. If A is a set,

then by |A| we mean the number of elements in the set A. If A is a real or complex number, then by |A| we understand the magnitude of A.

We write Re(z) and Im(z) for the real and imaginary parts of a complex number.

Numbers in square brackets refer to the references at the end of the thesis.

1. Real Integer Functions.

Definition I.1.1.

A (real) integer function is a map $f:\mathbb{R}\to\mathbb{Z}$ such that

(a) $f(x+n) = f(x) + n$ for all $x\in\mathbb{R}$, $n\in\mathbb{Z}$

(b) $|x-f(x)| < 1$ for all $x\in\mathbb{R}$.

This definition is a generalization of familiar functions such as the greatest integer function.

The question immediately arises: what points in $\mathbb{R}$ are fixed by f? Clearly, since f maps $\mathbb{R}$ to $\mathbb{Z}$, if $x\notin\mathbb{Z}$, $f(x)\neq x$. I claim the converse is true, that is, if $n\in\mathbb{Z}$, $f(n)=n$. For by part (a) of the definition, $f(0+n)=f(0)+n$. What is $f(0)$? By part (b) we see $|0-f(0)|=|f(0)|<1$. But $f(0)\in\mathbb{Z}$. Hence $f(0)=0$. Thus we see $f(n)=n$. These results are summed up in the following

Theorem I.1.1.

$f(x) = x$ iff $x\in\mathbb{Z}$.

This shows that integer functions are surjective.

Suppose now we know that x lies between two integers. What can we say about $f(x)$?

Chapter I: Integer Functions

Theorem I.1.2.

If $n \leq x < n+1$, then $f(x)=n$ or $f(x)=n+1$.

Proof.

Assume $f(x) \leq n-1$. Then, since $x \geq n$ by hypothesis, $x-f(x) \geq 1$ which contradicts part (b) of Definition I.1.1.

Now assume $f(x) \geq n+2$. Then, since $x<n+1$, $x-f(x)>1$ which also contradicts part (b) of the definition.

Hence $f(x)=n$ or $f(x)=n+1$, as was to be shown.

Theorem I.1.3.

The set of all real integer functions may be placed in 1-1 correspondence with the set of all functions $g:(0,1)\rightarrow\{0,1\}$.

In fact, any function $g:(0,1)\rightarrow\{0,1\}$ may be extended to a real integer function g'.

Proof.

Suppose we have a real integer function g. Since $g(a+n)=n+g(a)$ for any integer n, g is completely determined by its action on $[0,1)$. But $g(0)=0$ by Theorem I.1.1; hence g is determined by its action on $(0,1)$. Define g' as the restriction of g on $(0,1)$. By Theorem I.1.2, $g'(x)=0$ or $g'(x)=1$ for $0<x<1$.

Now suppose $g \neq f$. We wish to show $g' \neq f'$. But $g \neq f \Rightarrow$ there

is some x such that $g(x) \neq f(x)$. Evidently $x \notin \mathbb{Z}$ for by Theorem I.1.1, $g(x)=x$ and $f(x)=x$ if $x \in \mathbb{Z}$. Hence we may assume $x=n+x'$ where $0<x'<1$. Then $g(x)=g(n+x')=n+g(x')=n+g'(x')$; similarly, we see $f(x)=f(n+x')=n+f(x')=n+f'(x')$. Hence $g(x)-f(x)=g'(x)-f'(x)$. The left side is not 0 by hypothesis; hence $g(x') \neq f(x')$. Thus $g'(x') \neq f'(x')$.

To show the second part of the theorem, suppose we have some $g:(0,1)\rightarrow\{0,1\}$. Define g', the extension of g, as follows:

$g'(n) = n$ for $n \in \mathbb{Z}$.

If $n<x<n+1$, $g'(x) = n+g(x-n)$.

I claim g' is a real integer function. The translation property is obvious. Also,

$$|g'(x)-x| = |n+g(x-n)-x|$$
$$= |n+g(u)-(u+n)|, \quad u=x-n$$
$$= |g(u)-u| < 1, \quad \text{since } 0<u<1 \text{ and } g(u)=0 \text{ or } g(u)=1.$$

Another way to consider these functions is to consider the set $g^{-1}(n)$. Since $g^{-1}(n)=n+g^{-1}(0)$, it suffices to examine the kernel. The next theorem completely characterizes the kernels of all real integer functions.

Theorem I.1.4.

Let X be a set of real numbers. Then X is the kernel of

some real integer function iff

(a) If $x \in X$, $|x|<1$.

(b) $X \cup X+1 \supset [0,1]$

(c) $X \cap X+1 = \emptyset$

Proof.

Suppose $X = f^{-1}(0)$. Then $1+f^{-1}(0)=f^{-1}(1)$. Now clearly $f^{-1}(0) \cap f^{-1}(1) = \emptyset$. This proves (c). The definition of f implies (a). If $x \in [0,1]$, then $f(x)=0$ or $f(x)=1$ by Theorem I.1.2. Hence (b).

Now suppose (a)-(c) hold. Define f on [0,1) by

$f(a) = 0$ if $a \in X$

$f(a) = 1$ if $a \notin X$.

I claim that f', the extension of f as given in Theorem I.1.3, is a real integer function and that the kernel of f' is X.

Evidently $|f(a)-a|<1$ for $a \neq 0$. Since $0 \in X \cup X+1$, we know that either $0 \in X$ or $-1 \in X$. But by (a), if $x \in X$, $|x|<1$; hence $0 \in X$. Thus $|f(x)-x|<1$ for all $x \in [0,1)$. Thus f', the extension of f, is a real integer function.

Now we wish to show that X is the kernel of f'. Suppose $f'(a)=0$. Then either $a \in X \cap [0,1)$ or $a+1 \notin X \cap [0,1)$. In the first case, clearly $a \in X$. In the second case, by (a) and (b) of the hypothesis, we see that $a+1 \in X+1$, which implies that $a \in X$.

This shows $f'^{-1}(0) \subset X$. Now suppose $a \in X$. We must show $f'(a)=0$. Clearly if $a \in X \cap [0,1)$, then $f'(a)=0$ by definition of f'. Suppose $a \in X$ but $a \notin X \cap [0,1)$. Then $a \in (-1,0)$ by (a). Hence $a+1 \in (0,1)$. Then $a+1 \notin X+1 \cap (0,1)$. Hence $f'(a+1)=1$ and $f'(a)=0$.

Now we will use the definition of integer function to define a generalized "remainder" or "residue" relative to any integer function f.

Definition I.1.2.

$$\text{res}_f(a,b) = \begin{cases} b-a\cdot f(b/a), & a\neq 0 \\ b, & a=0 \end{cases}$$

for $a,b \in \mathbb{R}$.

Definition I.1.3.

$$\text{fr}_f(x) = \text{res}_f(1,x) = x-f(x) \quad (x \in \mathbb{R}).$$

The function fr_f is often called the fractional part.

Theorem I.1.5.

$$|\text{res}_f(a,b)| < |a| \text{ for } a,b \in \mathbb{R},\ a\neq 0.$$

Proof.

$$|b/a - f(b/a)| < 1$$

$$|a||b/a - f(b/a)| < |a|$$

$$|b - a\cdot f(b/a)| < |a|$$

$$|res_f(a,b)| < |a|.$$

Theorem I.1.5 is sometimes stated in the following form, often called the division theorem.

Theorem I.1.6.

Given $a,b \in \mathbb{R}$, there exists $q \in \mathbb{Z}$ and $r \in \mathbb{R}$ such that $|r|<|a|$ and $b = aq+r$. (McDonnell [3]).

Proof.

Put $q=f(b/a)$ where f is an integer function, and $r = res_f(a,b)$. Then $|r|<|a|$ by the preceeding theorem and $aq+r = a \cdot f(b/a)+b-a \cdot f(b/a) = b$.

Of course, as of yet we have not exhibited any functions f with the desired properties; this will be done in section I.2.

The function res defined above corresponds closely to the notion of congruence, as will be shown by the following theorem.

Theorem I.1.7.

$$res_f(a,b) = res_f(a,c) \text{ for } a \neq 0 \text{ iff } a|b-c.$$

Proof.

Assume $res_f(a,b)=res_f(a,c)$. Then $b-a \cdot f(b/a)=c-a \cdot f(c/a)$. Hence $b-c = a[f(c/a)-f(b/a)]$. But a divides the right side.

Hence $a|b-c$.

Now assume $a|b-c$. Then

$$\begin{aligned} res_f(a,b)-res_f(a,c) &= b-a\cdot f(b/a) - [c-a\cdot f(c/a)] \\ &= b-c + a[f(c/a)-f(b/a)]. \end{aligned}$$

But $a|b-c$; hence $b - c = ka$ for some k, and $b = ka+c$. Hence

$$\begin{aligned} res_f(a,b)-res_f(a,c) &= b-c + a[f(c/a)-f((ka+c)/a)] \\ &= b-c + a[f(c/a) - f(c/a) - k] \\ &= b-c - ak \\ &= 0. \end{aligned}$$

Theorem I.1.8.

$$res_f(a,c) = b\cdot res_f(a/b,\ c/b), \quad b\neq 0.$$

Proof.

$$\begin{aligned} res_f(a,c) &= c - a\cdot f(c/a) \\ &= b(c/b - a/b\cdot f(c/a)) \\ &= b(c/b - a/b\cdot f((c/b)/(a/b)) \\ &= b\cdot res_f(a/b,c/b). \end{aligned}$$

Since $res_f(a,b)$ is an integer if a and b are both integers, let us consider what the results of res_f are if a is fixed and b ranges over all integers.

Definition I.1.4.

$$csr_f(a) = \{z : z = res_f(a,b) \text{ for some } b\in\mathbb{Z}\}.$$

The set $csr_f(a)$ is called a complete system of residues for a (relative to the function f). A criterion for an integer x to be in the complete system of residues for a given integer a is the following

Theorem I.1.9.

$$x \in csr_f(a) \iff res_f(a,x) = x.$$

Proof.

$\Leftarrow$ is clear. Assume $x = res_f(a,b)$ for some $b \in \mathbb{Z}$. Thus

$$x = b - a \cdot f(b/a)$$

$$x - b = -a \cdot f(b/a)$$

Hence $a \mid x-b$. By Theorem I.1.7, $res_f(a,x) = res_f(a,b)$. But then $res(a,x) = x$, as desired.

The next theorem is based on a suggestion of McDonnell [13] and gives another characterization of the set $csr_f(a)$.

Theorem I.1.10.

If $x,a \in \mathbb{Z}$ and $a \neq 0$, then $x \in csr_f(a)$ iff $x \in a \cdot f^{-1}(0)$.

Proof.

Let $x \in csr_f(a)$. Then, by Theorem I.1.9, $x = res_f(a,x)$. But then $x = x - a \cdot f(x/a)$. Hence $a \cdot f(x/a) = 0$. But $a \neq 0$ by hypothesis. Hence $f(x/a) = 0$ and if $n = x/a$, then $x = an$ where $n \in f^{-1}(0)$.

Now let $x=as$ where $f(s)=0$. Hence $res_f(s)=s$. Hence by Theorem I.1.7, $res_f(a,as)=as$. Hence $res_f(a,x)=x$.

Theorem I.1.11.

$$|csr_f(a)| = |a|.$$

Proof.

Since $|res_f(a,b)| < |a|$ by Theorem I.1.5, it is clear that $res_f(a,b)$ can attain at most the $2|a|-1$ distinct values $1-|a|$, $2-|a|,\ldots-1,0,1,2,\ldots|a|-1$. We see that 0 is always in the complete system of residues. I claim that if $x<0$ is in the complete system, then $x+|a|$ is not, and if $x>0$ is in the complete system, then $x-|a|$ is not. This follows from the (readily verified) fact that $res_f(a,|a|-k) = res_f(a,-k)$. Also, the numbers $res_f(a,0)$, $res_f(a,1),\ldots res_f(a,|a|-1)$ are different by Theorem I.1.7. Hence there are exactly $|a|$ elements in the complete system for a.

The next theorem characterizes all complete systems of residues.

Theorem I.1.12.

There are exactly $2^{|a|-1}$ different complete systems of residues for any $a\neq 0 \in \mathbb{Z}$.

Proof.

0 is a member of any system of residues. If we arrange

the integers from $1-|a|$ to $|a|-1$ in two lines

$$1-|a|,\ 2-|a|,\ 3-|a|,\ \ldots,\ -3,\quad -2,\quad -1$$

$$1,\quad 2,\quad 3,\ \ldots,\ |a|-3,\ |a|-2,\ |a|-1$$

then we see from the previous theorem, if any number in the first line is in the complete system, the corresponding number in the second line is not. Hence there are at most $2^{|a|-1}$ subsets corresponding to this choice of residues.

Now we must show that, given any subset of residues chosen according to this system, there corresponds an integer function f.

Let $a_1, a_2, \ldots, a_n$ be the $|a|$ distinct residues. Let X be the set

$$\bigcup_k [(2a_k-1)/2|a|,\ (2a_k+1)/2|a|),$$

a union of half-open intervals. It is easy to see that by Theorem I.1.4, X is the kernel of an integer function f. In fact, the function f constructed in the proof of the theorem is the very one such that $csr_f(a)$ consists of the residues $a_1, a_2, \ldots, a_n$.

This concludes our discussion of the general properties of real integer functions. In the next section, we will discuss some of the properties of a particular integer function, the greatest integer or floor function.

2. The Floor Function.

Probably the most familiar integer function is the floor function. This function is variously known as the greatest integer, integer part, and entier function. See Knuth [4].

Definition I.2.1.

The floor function fl(x) is defined as

$$fl(x) = \sup\{k \in \mathbb{Z} \mid k \leq x\}$$

This definition shows why the floor function is often

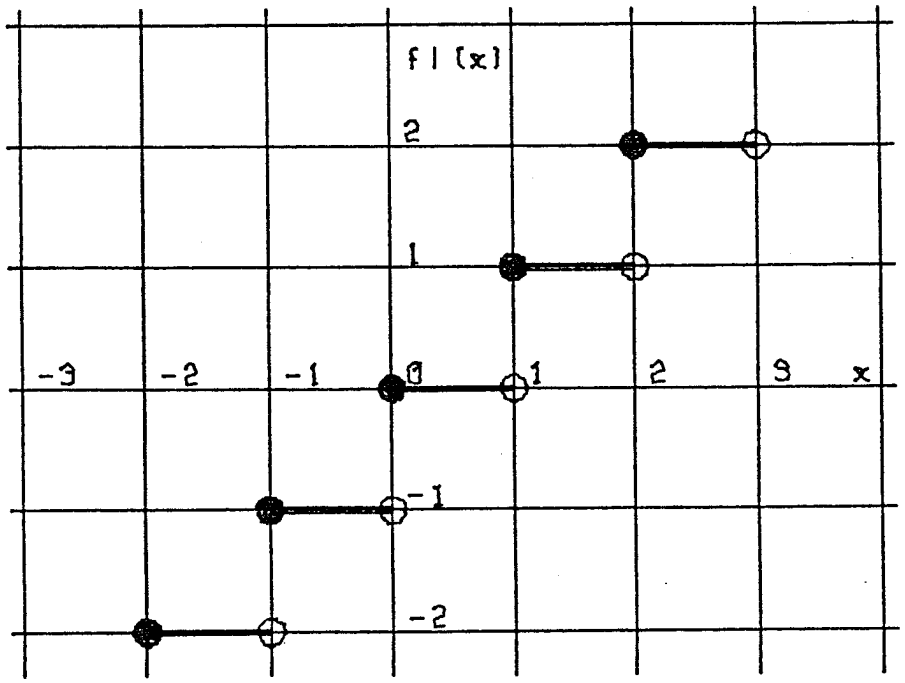

Figure 1: The Floor Function

called the "greatest integer" function. See Figure 1.

The fundamental property of the floor function is given in the next theorem, which is easily derived from the definition.

Theorem I.2.1.

$$0 \leq x - \mathrm{fl}(x) < 1.$$

The next theorem shows that, indeed, the floor function is an integer function.

Theorem I.2.2.

fl(x) is a real integer function.

Proof.

Evidently fl(x) maps $\mathbb{R} \to \mathbb{Z}$. We also have

$$\begin{aligned} \mathrm{fl}(x+n) &= \sup\{k \in \mathbb{Z} \mid k \leq x+n\} \\ &= \sup\{k \in \mathbb{Z} \mid k-n \leq x\} \\ &= \sup\{k'+n \in \mathbb{Z} \mid k' \leq x\} \\ &= n + \sup\{k' \in \mathbb{Z} \mid k' \leq x\} \\ &= n + \mathrm{fl}(x). \end{aligned}$$

Hence fl(x) satisfies part (a) of Definition I.1.1. It remains to show that fl satisfies part (b). But this is just Theorem I.2.1.

For the rest of this section, we will understand fr(x) and

res(a,b) to mean the functions of section I.1 with respect to the floor function. In particular we have the following useful theorem, which is a consequence of Theorem I.2.1 and the definition of fr(x).

Theorem I.2.3.

$$0 \leq fr(x) < 1.$$

The next two theorems were given as exercises in Knuth [4], and are easily proved.

Theorem I.2.4.

$$n \leq x < n+1 \text{ iff } fl(x) = n \text{ for } n \in \mathbb{Z},\ x \in \mathbb{R}.$$

Theorem I.2.5.

$$x-1 < n \leq x \text{ iff } n = fl(x) \text{ for } n \in \mathbb{Z},\ x \in \mathbb{R}.$$

Theorem I.2.6.

$$fl(x) + fl(y) = fl(x+y) \quad \text{iff } fr(x) + fr(y) < 1.$$

$$fl(x) + fl(y) = fl(x+y) - 1 \text{ iff } fr(x) + fr(y) \geq 1.$$

Hence $fl(x) + fl(y) \leq fl(x+y)$.

Proof.

(i) Suppose $0 \leq fr(x) + fr(y) < 1$. Then by Theorem I.2.1, $fl(x) \leq (x)$ and $fl(y) \leq (y)$. Hence $fl(x) + fl(y) \leq x+y$.

Also, $x = fl(x) + fr(x)$, $y = fl(y) + fr(y)$ by definition.

Hence

$$x + y = fl(x) + fl(y) + fr(x) + fr(y)$$
$$< fl(x) + fl(y) + 1.$$

Combining these, we get

$$fl(x) + fl(y) \leq x + y < fl(x) + fl(y) + 1.$$

Hence by Theorem I.2.4, $fl(x+y) = fl(x) + fl(y)$.

(ii) Now suppose $1 \leq fr(x) + fr(y) < 2$. Then an argument similar to that for (i) shows that $fl(x+y) = fl(x) + fl(y) + 1$.

To see the converses, assume $fl(x) + fl(y) = fl(x+y)$ but $1 \leq fr(x) + fr(y) < 2$. Then by part (ii), above, $fl(x+y) = fl(x) + fl(y) + 1$, contrary to assumption.

Similarly, assume $fl(x) + fl(y) = fl(x+y) + 1$ but $0 \leq fr(x) + fr(y) < 1$. Then by part (i), above, $fl(x+y) = fl(x) + fl(y)$, contrary to assumption.

Theorem I.2.7.

$fl_a(x) = fl(x+a)$, $0 \leq a < 1$, is a real integer function.

Proof.

$$fl_a(x+n) = fl(x+n+a) = fl(x+a) + n = fl_a(x) + n.$$

Hence fl_a satisfies part (a) of the definition. It is also easy to see that fl_a satisfies part (b).

In particular, taking a = 1/2, we get

Definition I.2.2.

$$\mathrm{mid}(x) = \mathrm{fl}_{1/2}(x) = \mathrm{fl}(x + 1/2).$$

mid(x) is usually interpreted as the "integer nearest x", with the proviso that if $x = n + 1/2$, then $\mathrm{mid}(x) = n + 1$. See Figure 2, below.

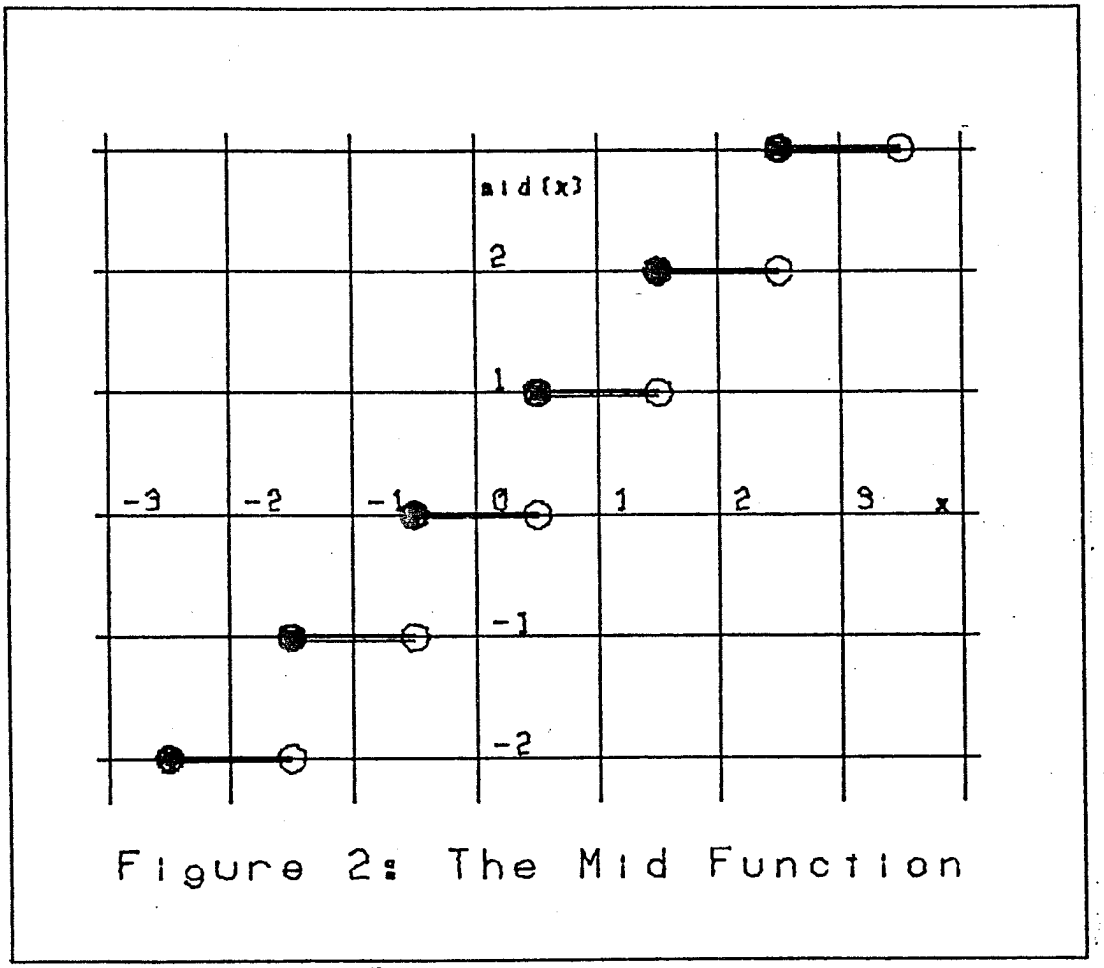

Figure 2: The Mid Function

Theorem I.2.8.

$\mathrm{rnd}(n,x) = (\mathrm{mid}(x \cdot 10^n))/10^n$ is precisely x rounded to n

decimal digits.

Examples.

$$\text{rnd}(2,3.14159) = 3.14$$

$$\text{rnd}(1,2.75) = 2.8$$

$$\text{rnd}(-2,34567) = 34600$$

$$\text{rnd}(3,-.2345) = -.234$$

Theorem I.2.9. (Knuth, [4]).

$$\text{fl}((m+x)/n) = \text{fl}((m+\text{fl}(x))/n) \text{ for } x \in \mathbb{R},\ m,n \in \mathbb{Z},\ n>0.$$

Proof.

$$\begin{aligned}(m+x)/n - 1 &= (m+x)/n - 1/n - (n-1)/n \\ &= (m+x-1)/n - (n-1)/n \\ &< (m+\text{fl}(x))/n - (n-1))/n \quad \text{(Theorem I.2.1)}\end{aligned}$$

I claim

$$(m+\text{fl}(x))/n - (n-1)/n \leq \text{fl}((m+\text{fl}(x))/n)$$

This is evidently true for $n = 1$. Assume $n \geq 2$. Now write $m+\text{fl}(x) = kn + c$ where $0 \leq c \leq n - 1$. The proof splits into two cases:

Case I. $0 \leq c \leq n - 2$.

Then

$$\begin{aligned}(m+\text{fl}(x))/n - (n-1)/n &= (kn+c-(n-1))/n \\ &= (k-1) + (c+1)/n \\ &< k\end{aligned}$$

$$= fl((kn+c)/n)$$

$$= fl((m+fl(x))/n)$$

Case II. $c = n - 1$.

Then

$$(m+fl(x))/n - (n-1)/n = (kn)/n = k = fl((m+fl(x)/n).$$

Thus we have

$$(m+x)/n - 1 < fl((fl(x)+m)/n) \leq (m+x)/n.$$

By Theorem I.2.5 we have the desired conclusion.

Remark.

If $m = 0$, we obtain the useful result

$$fl(x/n) = fl(fl(x)/n).$$

Theorem I.2.10. (Uspensky and Heaslet)

$$\sum_{K=0}^{n-1} fl(x + k/n) = fl(nx) \text{ for } x \in \mathbb{R},\ n>0 \in \mathbb{Z}.$$

Proof.

Let $j/n \leq fr(x) < (j+1)/n$ for some j, $0 \leq j < n$. Now for $0 \leq k \leq n-j-1$, $fl(x + k/n) = fl(x)$. For $n-j \leq k \leq n-1$, we have $fl(x + k/n) = fl(x) + 1$.

Hence

$$\sum_{k=0}^{n-1} fl(x + k/n) = \sum_{k=0}^{n-j-1} fl(x + k/n) + \sum_{k=n-j}^{n-1} fl(x+ k/n)$$
$$= (n-j)fl(x) + j(1+fl(x))$$
$$= nfl(x) + j$$

But $fl(x) + j/n \leq x < fl(x) + (j+1)/n$. By Theorem I.2.7, we see that

$$fl(nx) = nfl(x) + j.$$

Hence the desired conclusion.

The functional relation in the previous theorem is known as the replicative relation.

Theorem I.2.11.

$$fl(2x) + fl(2y) \geq fl(x) + fl(y) + fl(x+y)$$

Proof.

We can split the proof into four cases:

Case i: $fr(x) < 1/2$; $fr(y) < 1/2$

Case ii: $fr(x) < 1/2$; $fr(y) \geq 1/2$

Case iii: $fr(x) \geq 1/2$; $fr(y) < 1/2$

Case iv: $fr(x) \geq 1/2$; $fr(y) \geq 1/2$

Case i: We have $x = fl(x) + fr(x)$ and $2x = 2fl(x) + 2fr(x)$.

But $2fr(x) < 1$; hence $fl(2x) = 2fl(x)$. Similarly $fl(2y) = 2fl(y)$ and $fl(x+y) = fl(x) + fl(y)$ by Theorem I.2.6.

Hence $fl(2x) + fl(2y) = fl(x) + fl(y) + fl(x+y)$.

The remaining three cases can be disposed of with tedious reasoning similar to that in case i.

Many other properties of the floor function are given in Uspensky and Heaslet [5], Roberts [6], Knuth [7], Leveque [4], and Ryde [15].

3. Duality.

Another commonly used integer function is the so-called least integer or ceiling function. This function is intimately related to the floor function of the previous section.

Definition I.3.1.

$$ce(x) = \inf\{n \in \mathbb{Z} \mid n \geq x\}.$$

Theorem I.3.1. (Duality)

$$ce(x) = -fl(-x).$$

Proof.

$$\begin{aligned} fl(-x) &= \sup\{n \in \mathbb{Z} \mid n \leq -x\} \\ &= \sup\{n \in \mathbb{Z} \mid -n \geq -x\} \\ &= -\inf\{n \in \mathbb{Z} \mid n \geq x\} \\ &= -ce(x). \end{aligned}$$

We define the dual function of f, $\hat{f}$ as $\hat{f}(x) = -f(-x)$. It is easy to see that $\hat{\hat{f}}$ is the same as f. Then we have the following

Theorem I.3.2.

If f is an integer function, then so is its dual $\hat{f}$.

Proof.

(a) $\hat{f}(x+n) = -f(-x-n)$

$$= -\underline{f}(-x) + n$$

$$= \hat{f}(x) + n$$

(b) $$|x - \hat{f}(x)| = |x - (-f(-x))|$$

$$= |x + f(-x)|$$

$$= |f(u) - u| \qquad (u = -x)$$

$$< 1.$$

For more on the notion of duality, see Halpern [12].

From Theorem I.3.1 and Theorem I.3.2 we immediately find

Theorem I.3.3

$ce(x)$ is a real integer function and $0 \leq ce(x) - x < 1$.

Figure 3 shows a graph of the ceiling function.

The duality relationship between the floor and ceiling functions expressed in Theorem I.3.1 implies many theorems about the ceiling function which are easily derived.

Theorem I.3.4

(a) $ce(x) = n$ iff $n - 1 < x \leq n$.

(b) $ce(x) = n$ iff $x \leq n < x + 1$.

(c) $ce_a(x) = ce(x-a)$ is a real integer function for $0 \leq a < 1$.

(d) $ce((m+x)/n) = ce((m+ce(x))/n)$.

(e) $$\sum_{k=0}^{n-1} ce(x - k/n) = ce(nx) \text{ for } x \in \mathbb{R},\ n>0 \in \mathbb{Z}.$$

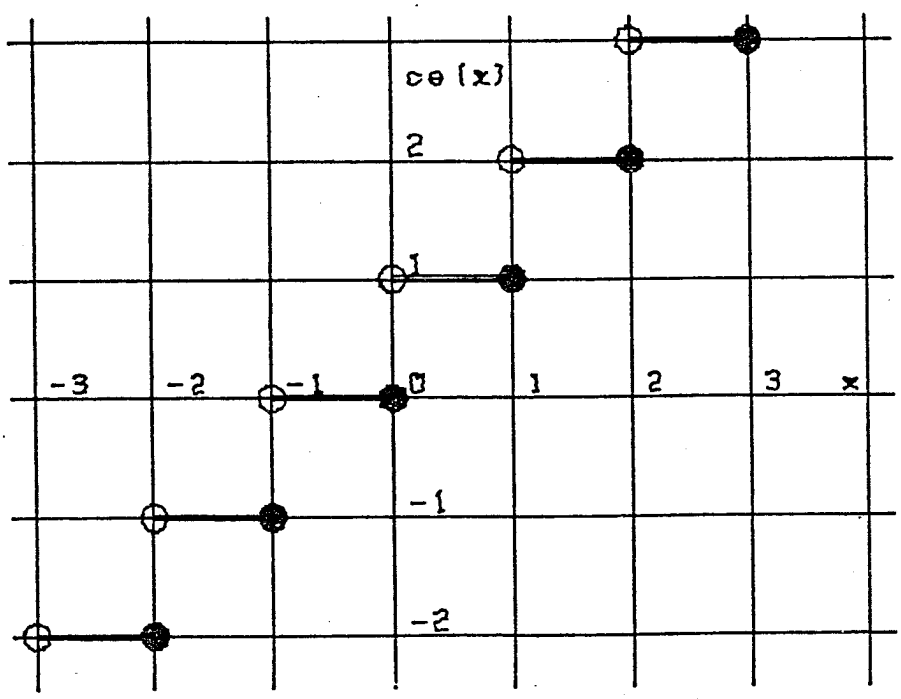

Figure 3: The Ceiling Function

(f) $ce(x) + ce(y) \geq ce(x+y)$

(g) $ce(2x) + ce(2y) \leq ce(x) + ce(y) + ce(x+y)$.

As an example, let us prove (d):

We have

$$fl((-m-x)/n) = fl((-m+fl(-x))/n) \text{ by Theorem I.2.9.}$$

$$fl(-(m+x)/n) = fl(-(m-fl(-x))/n)$$

$$-fl(-(m+x)/n) = -fl(-(m+ce(x))/n)$$

$$ce((m+x)/n) = ce((m+ce(x))/n).$$

Another useful inequality is the following, which relates the floor to the ceiling function.

Theorem I.3.5.

$$ce(x) - 1 \leq fl(x) \leq ce(x) \leq 1 + fl(x).$$

Proof.

If $x = n$, an integer, then $n = fl(n) = ce(n)$. Hence assume $x = fl(x) + fr(x)$, $fr(x) \neq 0$. Then $ce(x) = 1 + fl(x)$, from which the inequalities follow.

4. Complex Integer Functions.

At this point, we would like to extend our notion of integer functions to the complex plane. In analogy with Definition I.1.1, we make the following

Definition I.4.1.

A complex integer function is a map $f:\mathbb{C}\to\mathbb{Z}[i]$ such that

(a) $f(w+z) = f(w) + z$ for all $w\in\mathbb{C}$, $z\in\mathbb{Z}[i]$

(b) $|w - f(w)| < 1$ for all $w\in\mathbb{C}$.

It should be clear that the restriction of any complex integer function to the real line is a real integer function.

As in the case of real integer functions, we have the following

Theorem I.4.1.

If f is a complex integer function, then $f(z) = z$ iff z is a complex integer.

The proof is exactly as in the real case.

Theorem I.4.2.

If $m \leq \mathrm{Re}(z) < m + 1$ and $n \leq \mathrm{Im}(z) < n + 1$, then

$f(z) = m+ni$ or $f(z) = m+1 + ni$ or $f(z) = m + (n+1)i$

or $f(z) = (m+1) + (n+1)i$.

Again, the proof is a trivial consequence of part (b) of the definition.

Remark.

Unlike the real case, not every function $g:S\to\{0, 1, i, 1+i\}$ where $S = \{a+bi \mid 0<a<1, 0<b<1\}$, is an integer function. For example, if $g(x) = 1+i$ for all $x \in S$, then $g(.1) = 1+i$; so we find

$$|g(.1) - .1| = 1.3 > 1.$$

As in the real case, we can also characterize the kernel of a complex integer function abstractly.

Theorem I.4.3.

A set X is the kernel of a complex integer function iff

(a) If $x \in X$, then $|x| < 1$.

(b) Let $S = \{a+bi : 0 \leq a \leq 1, 0 \leq b \leq 1\}$

Then $X \cup X+1 \cup X+i \cup X+1+i \supset [0,1]$.

(c) The sets X, X+1, X+i, X+1+i are pairwise disjoint.

The proof uses the same reasoning as the proof of Theorem I.1.4, and is left to the reader.

The residue and fractional part functions, previously defined only for real numbers, have analogues in the complex plane. We make the following two definitions:

Definition I.4.2.

$$\mathrm{res}_f(w,z) = \begin{cases} z - w\cdot f(z/w), & w\neq 0. \\ z, & w=0. \end{cases}$$

for $w,z \in \mathbb{C}$.

Definition I.4.3.

$$\mathrm{fr}_f(w) = \mathrm{res}_f(1,w) = w - f(w),\ w \in \mathbb{C}.$$

The following four theorems are easily proved following the ideas for the proofs of their real analogues given in section I.1.

Theorem I.4.4.

$$|\mathrm{res}_f(w,z)| < |w| \text{ for } w,z \in \mathbb{C},\ w\neq 0.$$

Theorem I.4.5.

Given $w,\ z \in \mathbb{C}$, there exists $q \in \mathbb{Z}[i]$ and $r \in \mathbb{C}$ such that $|r| < |w|$ and $z = wq + r$.

Theorem I.4.6.

$$\mathrm{res}_f(w,z) = \mathrm{res}_f(w,x) \text{ for } w\neq 0 \text{ iff } w \mid z-x.$$

Theorem I.4.7.

$$\mathrm{res}_f(w,z) = x\ \mathrm{res}_f(w/x, z/x),\ x\neq 0.$$

We can also extend the idea of a complete system of residues to the complex plane:

Definition I.4.4.

$$cscr_f(a) = \{z : z = res_f(a,b) \text{ for some } b \in \mathbb{Z}[i]\}.$$

The set $cscr_f(a)$ is called a complete system of complex residues. As in the real case, we have the following theorems:

Theorem I.4.8.

$$z \in cscr_f(w) \iff res_f(w,z) = z.$$

Theorem I.4.9.

If $z, w \in \mathbb{Z}[i]$ and $w \neq 0$, then $z \in cscr_f(w)$ iff $z \in \bar{w} \cdot f^{-1}(0)$.

The previous two theorems can be proved using the same methods used in the real case.

Theorem I.4.10.

$$|cscr_f(w)| = |w|^2 \text{ for } w \neq 0 \in \mathbb{Z}[i].$$

Proof.

From Theorem I.4.6 and Theorem I.4.8 we see that $|cscr_f(w)|$ is the same as the number of distinct cosets in the quotient $\mathbb{Z}[i]/(w)\mathbb{Z}[i]$. We use the following argument due to G. Tunnell:

As an abelian group $\mathbb{Z}[i]$ is isomorphic to $\mathbb{Z} \oplus \mathbb{Z}$. The basis $\{1,i\}$ is used for $\mathbb{Z}[i]$. Let $w = a+bi$. The submodule $(a+bi)\mathbb{Z}[i]$ is a subgroup with basis $\{a+bi, -b+ai\}$. The matrix

relating the two bases is

$$M = \begin{bmatrix} a & b \\ -b & a \end{bmatrix}$$

The abelian group $\mathbb{Z}[i]/(a+bi)\mathbb{Z}[i]$ is isomorphic to

$$\mathbb{Z}/d_1\mathbb{Z} \oplus \mathbb{Z}/d_2\mathbb{Z} \oplus \ldots \mathbb{Z}/d_s\mathbb{Z}$$

with $d_1 d_2 \ldots d_s = \det M = a^2 + b^2$. The order of $\mathbb{Z}[i]/(a+bi)\mathbb{Z}[i]$ is therefore also $a^2 + b^2 = |w|^2$. See Jacobson [14].

It would be interesting to obtain a purely arithmetic proof of this theorem that did not use the theory of modules.

The first example of a complex integer function was given by Hurwitz [1]. His definition is equivalent to the following:

Definition I.4.2.

Let $z = x+iy$, $x,y \in \mathbb{R}$. Write $z' = fl(x) + ifl(y)$. Then

$$cmid(z) = \begin{cases} z' + 0 & \text{if } fr(x)<1/2,\ fr(y)<1/2 \\ z' + 1 & \text{if } fr(x)\geq 1/2,\ fr(y)<1/2 \\ z' + i & \text{if } fr(x)<1/2,\ fr(y)\geq 1/2 \\ z' + 1+i & \text{if } fr(x)\geq 1/2,\ fr(y)\geq 1/2 \end{cases}$$

Theorem I.4.11.

cmid is a complex integer function.

Proof.

The translation property is evident. Let us prove

$$|\text{cmid}(z) - z| < 1.$$

Case I. $\text{fr}(x)<1/2$, $\text{fr}(y)\geq 1/2$.

Then $\text{cmid}(z) = \text{cmid}(x+iy) = \text{fl}(x) + i\,\text{fl}(y)$.

$$\begin{aligned} |\text{cmid}(z)-z|^2 &= (x-\text{fl}(x))^2 + (y-\text{fl}(y))^2 \\ &= (\text{fr}(x))^2 + (\text{fr}(y))^2 \\ &< 1/4 + 1/4 < 1/2. \end{aligned}$$

Cases II, III, and IV are equally tedious and are left to the reader.

Figure 4 below is a representation of the cmid function.

Remark.

cmid is an extension of the mid function to the complex plane. It is easily seen that $\text{cmid}(x) = \text{mid}(x)$ if $x \in \mathbb{R}$. Also, $\text{cmid}(x)$ may be interpreted as the "complex integer nearest x" with the proviso that when there are two or four "nearest integers", x gets mapped to the number with the larger real or imaginary part. That is, $\text{cmid}(1/2) = 1$, $\text{cmid}(i/2) = i$, and $\text{cmid}((1+i)/2) = 1+i$.

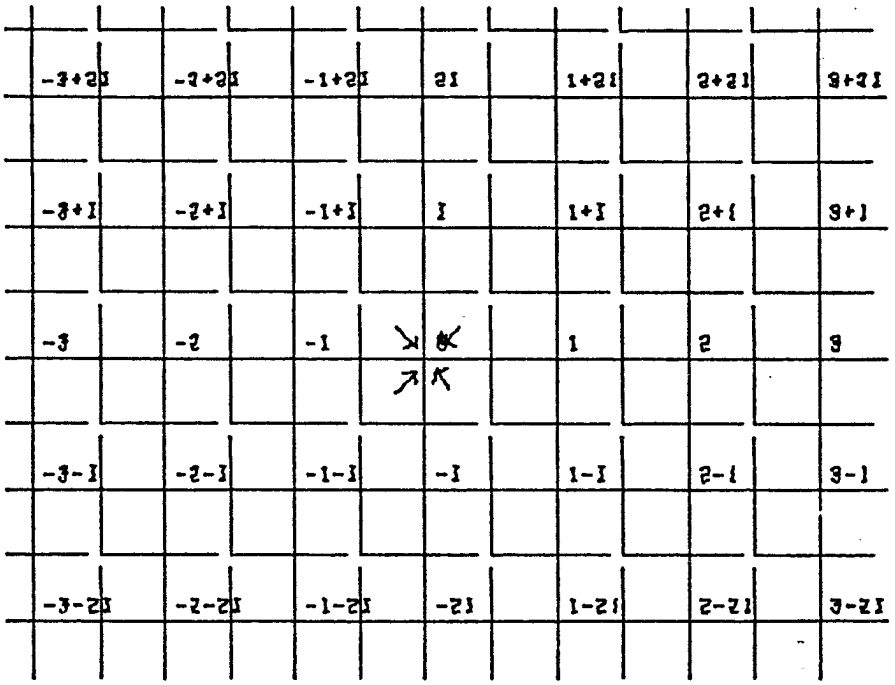

Figure 4: The cmld Function

5. The Complex Floor Function.

At this point, we would like to extend the floor function of section I.2 to the complex plane. One such extension was provided by McDonnell [3].

Definition I.5.1.

Let $z = x+iy$, $z' = fl(x) + ifl(y)$. Then

$$cfl(z) = \begin{cases} z' \text{ if } fr(x) + fr(y) < 1 \\ z' + 1 \text{ if } fr(x) + fr(y) \geq 1 \text{ and } fr(x) \geq fr(y) \\ z' + i \text{ if } fr(x) + fr(y) \geq 1 \text{ and } fr(x) < fr(y). \end{cases}$$

Theorem I.5.1.

$cfl(z)$ is a complex integer function.

Proof.

Again, the translation property should be obvious. The reader may convince himself that $|cfl(z) - z| < 1$ by noting that $cfl^{-1}(0)$ lies totally within the unit circle. See Figure 5 below.

Some of the nice identities for the real floor function do not extend directly in this generalization to the complex plane. For example, it is not true in general that

$$|cfl(z)+cfl(w)| \leq |cfl(w+z)|$$

or $$|cfl(w)| + |cfl(z)| \leq |cfl(w+z)|.$$

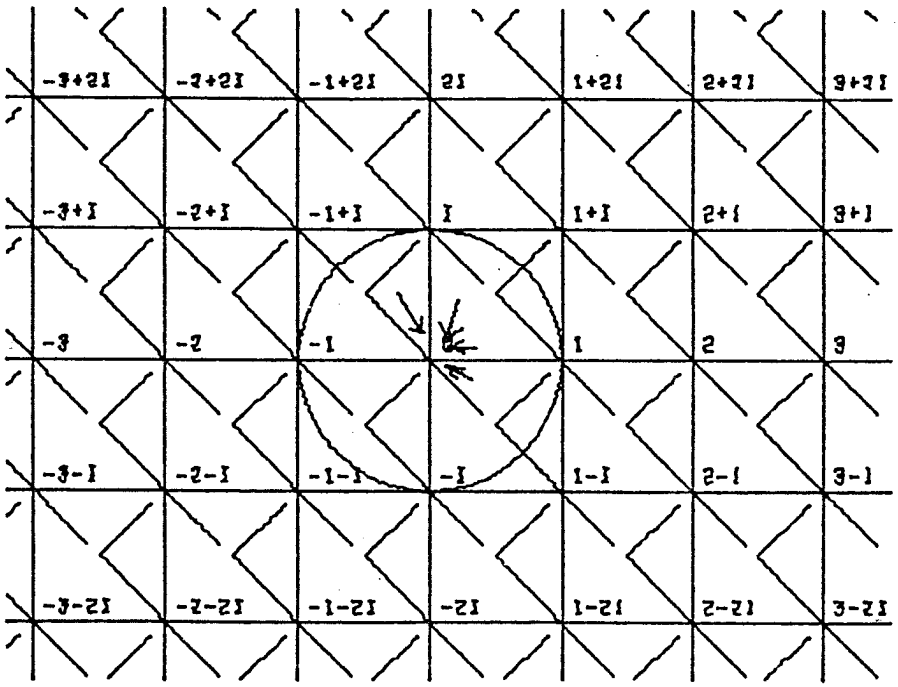

Figure 5: The cfl Function

As the reader may verify, a counter-example to both of these is w = 1.2 - 1.9i, z = 1.1 - 1.2i.

However, we can obtain a certain generalization of the real identities. These generalizations use the following function.

Definition I.5.2.

sum(z) = Re(z) + Im(z).

Theorem I.5.2.

sum(z) is a linear function of z, i. e.,

sum(w+z) = sum(w) + sum(z)

sum(cw) = c·sum(w) if c is real.

The proof is trivial and is left to the reader.

The following theorem relates the cfl and sum functions in an interesting way:

Theorem I.5.3.

sum(cfl(z)) = fl(sum(z)).

Proof.

The proof splits into three cases that correspond to the three parts of Definition I.5.1. We will show the first case, and leave the other two cases to the reader.

Case I. z=x+iy, fr(x) + fr(y) < 1.

Then cfl(z) = fl(x) + ifl(y). Hence sum(cfl(z)) = fl(x) + fl(y). On the other hand, sum(z) = x+y and by Theorem I.2.9, fl(sum(z)) = fl(x+y) = fl(x) + fl(y) since fr(x) + fr(y) < 1.

We have the following generalization of Theorem I.2.9.

Theorem I.5.4.

$$\text{sum}(\text{cfl}(w)+\text{cfl}(z)) \leq \text{sum}(\text{cfl}(w+z)) \text{ for } w,z \in \mathbb{C}.$$

Proof.

$$\begin{aligned}
\text{sum}(\text{cfl}(w)+\text{cfl}(z)) &= \text{sum}(\text{cfl}(w)) + \text{sum}(\text{cfl}(z)) \\
&= \text{fl}(\text{sum}(w)) + \text{fl}(\text{sum}(z)) \\
&\leq \text{fl}(\text{sum}(w) + \text{sum}(z)) \\
&= \text{fl}(\text{sum}(w+z)) \\
&= \text{sum}(\text{cfl}(w+z)).
\end{aligned}$$

Repeated use of the above technique proves the following three theorems.

Theorem I.5.5.

$$\text{sum}(\text{cfl}((z+m)/n)) = \text{sum}(\text{cfl}((\text{cfl}(z)+m)/n))$$

for all $z \in \mathbb{C}$, $m \in \mathbb{Z}[i]$, $n \in \mathbb{Z}$, $n>0$.

Theorem I.5.6.

$$\text{sum}\left(\sum_{k=0}^{n-1} \text{cfl}(z + k/n)\right) = \text{sum}(\text{cfl}(nz)) \text{ for } z \in \mathbb{C},\ n>0.$$

Theorem I.5.7.

$$\text{sum}(\text{cfl}(2x) + \text{cfl}(2y)) \geq \text{sum}(\text{cfl}(x) + \text{cfl}(y) + \text{cfl}(x+y)).$$

The following theorem relates the real and imaginary parts of cfl(z) and will be useful in Chapter II.

Theorem I.5.8.

If $Re(z) \geq Im(z)$, then $Re(cfl(z)) \geq Im(cfl(z))$.

Proof.

Let $z = x+iy$. The proof splits into two cases:

Case I. $x \geq y+1$. Then by Definition I.5.1, $cfl(z) = fl(x) + ifl(y)$ or $1+fl(x) + ifl(y)$ or $fl(x) + i(1+fl(y))$. In each instance, the desired conclusion follows.

Case II. $y \leq x < y+1$. Again, by definition I.5.1, either $cfl(z) = fl(x) + ifl(y)$ or $1+fl(x) + ifl(y)$ or $fl(x) + i(1+fl(y))$. In the first two cases, the conclusion follows. The last, $cfl(z) = fl(x) + i(1+fl(y))$, occurs only when $fr(x) + fr(y) \geq 1$ and $fr(x) < fr(y)$. Now $y \leq x < y+1 \Rightarrow fl(x) = fl(y)$ or $fl(x) = 1+fl(y)$. Assume $fl(x) = fl(y)$. Then $x = fl(x) + fr(x)$, $y = fl(y) + fr(y) = fl(x) + fr(y)$. Now $fr(x) < fr(y)$, so we find $x < y$, contrary to assumption. Thus we must have $fl(x) = 1+fl(y)$ and the conclusion follows.

We also have the following theorem which is easily verified from the definition of cfl.

Theorem I.5.9.

If a is real, then $cfl(ai) = i \cdot fl(a)$.

Now we will define the analogue of the ceiling function in the complex plane. We will use the idea of duality in the following definition.

Definition I.5.3.

$$cce(z) = -cfl(-z) \text{ for all } z \in \mathbb{C}.$$

By the analogue of Theorem I.3.2 in the complex plane, we immediately find

Theorem I.5.10.

cce(z) is a complex integer function.

Figure 6 illustrates the complex ceiling function.

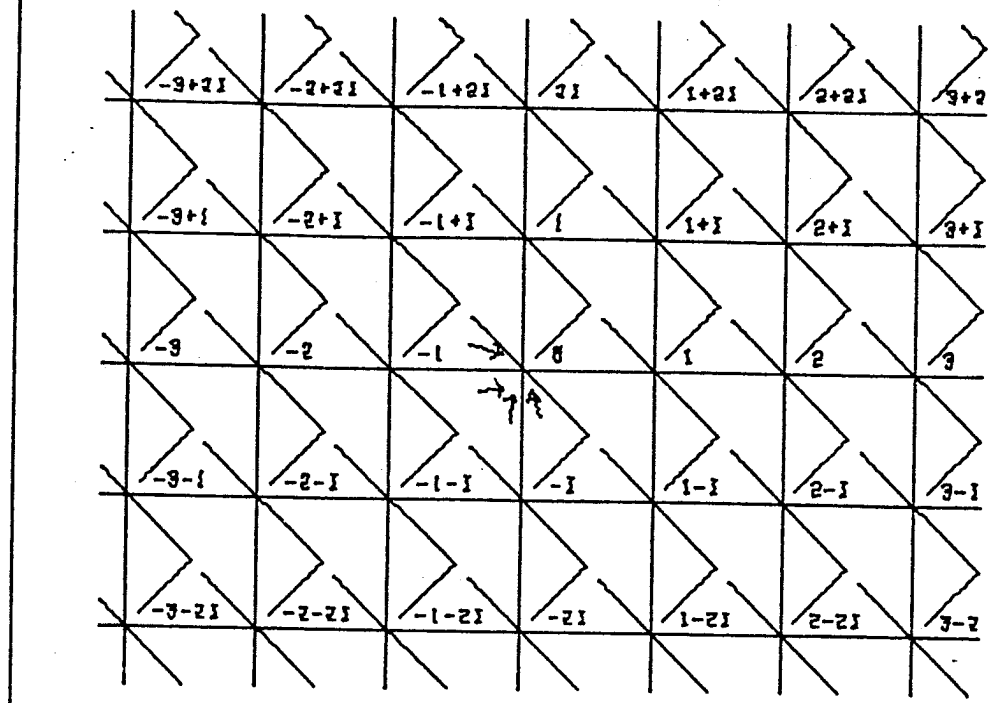

Figure 6: The cce Function