THE BASE CONVERSION THEOREM

DAVID W. MATULA

I. **Introduction and summary.** Utilizing a fixed number of digits and the power of a base to denote a real number is variously called a normalized, floating point, or scientific representation. Allowing negative as well as positive powers of the base in a radix representation and truncating all polynomials after a fixed number of terms will provide a characterization of such a space of "normalized numbers with a fixed number of digits." Conversion of normalized numbers from one base to another requires more than just a change in the representation of the number, since the real numbers in one such space will in general not be elements of another such space.

Mappings between different spaces of normalized numbers are provided by truncation conversion and rounding conversion. We seek the conditions under which these conversion mappings will be one-to-one and/or onto. Our major result is termed the Base Conversion Theorem: A rounding or truncation mapping of the space of normalized n digit numbers of the base β into the space of normalized m digit numbers of the base ν, where $\beta^i \neq \nu^j$ for any positive integers i, j, is one-to-one if and only if $\nu^{m-1} \geq \beta^n - 1$, and is onto if and only if $\beta^{n-1} \geq \nu^m - 1$. Thus if we call the fixed number of digits utilized in the representation the significance of the representation, then the Base Conversion Theorem implies that the set of real numbers with a given significance in one base cannot, in general, be converted into a set of numbers in a new base with uniform significance.

In a previous paper [1] we established notation, proved only a small part of the above theorem, and then concentrated on the implications of these results for digital computers. In this paper we repeat the necessary preliminaries, and then give a complete proof of the Base Conversion Theorem.

II. **Finite significance number spaces.**

DEFINITION. For the integers $\beta \geq 2$, called the *base*, and $n \geq 1$, called the *significance*, let the significance space, S_β^n, be the following set of real numbers:

(1)
$$S_\beta^n = \left\{ x \mid |x| = \sum_{i=1}^{n} \alpha_i \beta^{j-i}, j, \alpha_i \text{ integers}, \right.$$
$$\left. 0 \leq \alpha_i \leq \beta - 1 \quad \text{for } 1 \leq i \leq n \right\}.$$

S_β^n may be interpreted as the *space of normalized β-ary numbers of significance* n (i.e. containing n β-ary digits). Note that our definition includes the number zero in S_β^n.

Using the natural ordering of the real numbers, every number in S_β^n, other than zero, will have both a next smallest and next largest neighbor in S_β^n. These numbers play an important part in the discussion of conversion, so that we introduce the following notation.

DEFINITION. For $x \in S_\beta^n$, $x \neq 0$, x' is the *successor* of x in S_β^n iff

$$(2) \qquad x' = \min\{y \mid y > x, \, y \in S_\beta^n\}.$$

To avoid ambiguity we occasionally use $[x]'$ for x'. Thus in S_{10}^4 we have $5123' = 5124$, $[-.18 \times 10^{15}]' = -.1799 \times 10^{15}$, $9.999' = 10$, and $70700' = 70710$.

COROLLARY 1. *In S_β^n for any integer i,*

$$(3) \qquad [\beta^i(1 - \beta^{-n})]' = \beta^i, \qquad [\beta^i]' = \beta^i(1 + \beta^{1-n}).$$

COROLLARY 2. *For $x \in S_\beta^n$ and any integer i,*

$$(4) \qquad \beta^i[x]' = [\beta^i x]'.$$

The relative difference between a number in S_β^n and its successor is a useful measure of the spacing between numbers in S_β^n. Note that the uniform significance in S_β^n yields a variable relative difference over the elements of S_β^n. The relative difference is given by the gap function on S_β^n, which is defined as follows to depend only on the magnitude of the numbers involved and not on their sign.

DEFINITION. The *gap* in S_β^n at $y \in S_\beta^n$ for $y \neq 0$ is denoted by $\gamma_\beta^n(y)$ where

$$(5) \qquad \begin{aligned} \gamma_\beta^n(y) &= (y' - y)/y & \text{for } y > 0, \\ \gamma_\beta^n(y) &= \gamma_\beta^n(-y) & \text{for } y < 0. \end{aligned}$$

When there is no ambiguity, $\gamma(y)$ may be used for $\gamma_\beta^n(y)$. Although the difference, $y' - y$, grows monotonically with $y \in S_\beta^n$, the gap function, $\gamma_\beta^n(y)$, varies between fixed bounds for all $y \neq 0$. We now find the smallest, $g(S_\beta^n)$, and the largest, $G(S_\beta^n)$, values that the gap function attains for $y \in S_\beta^n$, $y \neq 0$.

LEMMA.

$$(6) \qquad \begin{aligned} g(S_\beta^n) &= \min \{\gamma_\beta^n(y) \mid y \in S_\beta^n, \, y \neq 0\} = 1/(\beta^n - 1), \\ G(S_\beta^n) &= \max \{\gamma_\beta^n(y) \mid y \in S_\beta^n, \, y \neq 0\} = 1/\beta^{n-1}. \end{aligned}$$

PROOF. For any $y = \alpha_1\beta^{j-1} + \alpha_2\beta^{j-2} + \cdots + \alpha_n\beta^{j-n} \in S_\beta^n$ with $\alpha_1 \neq 0$,

$$
\gamma_\beta^n(y) = (y' - y)/y = \beta^{j-n}/(\alpha_1\beta^{j-1} + \alpha_2\beta^{j-2} + \cdots + \alpha_n\beta^{j-n})
$$

(7)

$$
= 1/(\alpha_1\beta^{n-1} + \alpha_2\beta^{n-2} + \cdots + \alpha_n).
$$

The minimum on the right occurs when $\alpha_i = \beta - 1$ for all i, and the maximum occurs when $\alpha_1 = 1$, $\alpha_2 = \alpha_3 = \cdots = \alpha_n = 0$.

Thus $1/\beta^{n-1} \geq \gamma_\beta^n(y) \geq 1/(\beta^n - 1)$ for all $y \in S_\beta^n$, $y > 0$, and since $\gamma_\beta^n(-y) = \gamma_\beta^n(y)$, the bounds hold for all $y \in S_\beta^n$, $y \neq 0$. Since $\gamma_\beta^n(\beta^n - 1) = 1/(\beta^n - 1)$ and $\gamma_\beta^n(\beta^{n-1}) = 1/\beta^{n-1}$, both bounds are achieved and the lemma is proved.

III. **A variation of Kronecker's theorem.** We shall need the following number theoretic result.

THEOREM 1. *If β, $\nu > 1$ are integers such that $\beta^i \neq \nu^j$ for any positive integers i, j, then the space of rational numbers of the form β^i/ν^j, where i, j are positive integers, is dense in the positive real line.*

PROOF. Note that $\log_\nu \beta$ must be irrational since $\beta^i \neq \nu^j$ for any positive integers i, j. From Kronecker's theorem [2, Theorem 438, p. 373] we have that for any real α and any $\epsilon > 0$, there exist positive integers i, j such that

(8)
$$
\alpha - \epsilon < i \log_\nu \beta - j < \alpha + \epsilon
$$

so that

(9)
$$
\nu^{\alpha-\epsilon} < \beta^i/\nu^j < \nu^{\alpha+\epsilon}.
$$

Since α and $\epsilon > 0$ are arbitrary, this proves the theorem.

IV. **Conversion mappings.** Conversion of the real numbers into the significance space S_β^n constitutes a mapping, and the following definition characterizes the two most prevalent means of effecting conversions.

DEFINITION. The *truncation conversion mapping, T_β^n,* and the *rounding conversion mapping, R_β^n,* of the real numbers into S_β^n are defined for all integers $n \geq 1$ and $\beta \geq 2$ as follows:

TRUNCATION CONVERSION.

$$
T_\beta^n(x) = \max\{y \mid y \leq x, y \in S_\beta^n\} \text{ for } x > 0,
$$

(10)
$$
= \min\{y \mid y \geq x, y \in S_\beta^n\} \text{ for } x < 0,
$$

$$
= 0 \quad \text{ for } x = 0.
$$

ROUNDING CONVERSION.

$$R_\beta^n(x) = \min\{y \mid x < (y + y')/2, \, y \in S_\beta^n\} \qquad \text{for } x > 0,$$

$$(11) \qquad\qquad = \min\{y \mid x \leqq (y + y')/2, \, y \in S_\beta^n\} \qquad \text{for } x < 0,$$

$$= 0 \qquad \text{for } x = 0.$$

NOTE. The above definitions are for the usual truncation and rounding by magnitude with sign appended, so that $T_\beta^n(-x) = -T_\beta^n(x)$ and $R_\beta^n(-x) = -R_\beta^n(x)$ for all x.

It follows that the conversion of one significance space into another may be specified by a restricted mapping, for example, $R_\nu^m \colon S_\beta^n \to S_\nu^m$. A considerable simplification in the theory of base conversion occurs in this example when both β and ν are variants of a common base, e.g., $\beta = \gamma^j$, $\nu = \gamma^i$, since then many numbers are common to both S_β^n and S_ν^m, and in addition a practical conversion procedure may be described which involves simply loooking at blocks of digits of numbers represented in the base γ. A deeper theory is needed to treat conversion between bases which are not variants of a common base, a condition which is specified for bases β, ν, by requiring that $\beta^i \neq \nu^j$ for any positive integers i, j. With these preliminaries we now state and prove our major result.

BASE CONVERSION THEOREM. *If $\beta^i \neq \nu^j$ for any integers $i, j > 0$, then the truncation (rounding) conversion mapping of S_β^n into S_ν^m, i.e., $T_\nu^m \colon S_\beta^n \to S_\nu^m (R_\nu^m \colon S_\beta^n \to S_\nu^m)$*

   (1) *is one-to-one iff $\nu^{m-1} \geqq \beta^n - 1$,*

   (2) *is onto iff $\beta^{n-1} \geqq \nu^m - 1$.*

PROOF. The following eight statements will be proved separately:

   (i) $\nu^{m-1} \geqq \beta^n - 1 \Rightarrow T_\nu^m \colon S_\beta^n \to S_\nu^m$ is one-to-one.

   (ii) $\nu^{m-1} < \beta^n - 1 \Rightarrow T_\nu^m \colon S_\beta^n \to S_\nu^m$ is not one-to-one.

   (iii) $\nu^{m-1} \geqq \beta^n - 1 \Rightarrow R_\nu^m \colon S_\beta^n \to S_\nu^m$ is one-to-one.

   (iv) $\nu^{m-1} < \beta^n - 1 \Rightarrow R_\nu^m \colon S_\beta^n \to S_\nu^m$ is not one-to-one.

   (v) $\beta^{n-1} \geqq \nu^m - 1 \Rightarrow T_\nu^m \colon S_\beta^n \to S_\nu^m$ is onto.

   (vi) $\beta^{n-1} < \nu^m - 1 \Rightarrow T_\nu^m \colon S_\beta^n \to S_\nu^m$ is not onto.

   (vii) $\beta^{n-1} \geqq \nu^m - 1 \Rightarrow R_\nu^m \colon S_\beta^n \to S_\nu^m$ is onto.

   (viii) $\beta^{n-1} < \nu^m - 1 \Rightarrow R_\nu^m \colon S_\beta^n \to S_\nu^m$ is not onto.

   (i) $\nu^{m-1} \geqq \beta^n - 1 \Rightarrow T_\nu^m \colon S_\beta^n \to S_\nu^m$ is one-to-one:

Assume that $\nu^{m-1} \geqq \beta^n - 1$. From the lemma:

$$(12) \qquad\qquad\qquad g(S_\beta^n) \geqq G(S_\nu^m).$$

Given $x, y \in S_\beta^n$, $x \neq y$, we must show $T_\nu^m(x) \neq T_\nu^m(y)$. If either x or y is zero or if x and y are of opposite sign the result is immediate. Fur-

thermore if both x and y are negative and $T_\nu^m(x) = T_\nu^m(y)$, then also $T_\nu^m(-x) = T_\nu^m(-y)$. Hence it is sufficient to consider the case $x > y > 0$.

$$x = y + \{(x - y)/y\}y$$

(13)
$$\geqq T_\nu^m(y) + g(S_\beta^n)T_\nu^m(y) \geqq T_\nu^m(y) + G(S_\nu^m)T_\nu^m(y)$$

$$\geqq T_\nu^m(y) + \{\{[T_\nu^m(y)]' - T_\nu^m(y)\}/T_\nu^m(y)\}T_\nu^m(y) = [T_\nu^m(y)]'$$

Thus $T_\nu^m(x) \geqq [T_\nu^m(y)]'$ and $T_\nu^m(x) \neq T_\nu^m(y)$.

(ii) $\nu^{m-1} < \beta^n - 1 \Rightarrow T_\nu^m: S_\beta^n \to S_\nu^m$ is not one-to-one.

Let $T_\nu^m: S_\beta^n \to S_\nu^m$ and assume $\nu^{m-1} < \beta^n - 1$. Thus

(14)
$$0 < \nu^{1-m} - \beta^{-n} - \beta^{-n}\nu^{1-m}$$

so that

(15)
$$1/(1 - \beta^{-n}) < (1 + \nu^{1-m}).$$

By Theorem 1 there exist integers $i, j > 0$, such that

(16)
$$1/(1 - \beta^{-n}) < \beta^i/\nu^j < 1 + \nu^{1-m}.$$

Thus

(17)
$$\nu^j < \beta^i(1 - \beta^{-n}) < \nu^j(1 + \nu^{1-m})(1 - \beta^{-n}) < \nu^j(1 + \nu^{1-m})$$

and using Corollary 1 of §II

(18)
$$\nu^j < \beta^i(1 - \beta^{-n}) < [\beta^i(1 - \beta^{-n})]' = \beta^i < \nu^j(1 + \nu^{1-m}) = [\nu^j]'.$$

The inequalities above imply $T_\nu^m(\beta^i(1-\beta^{-n})) = T_m^\nu(\beta^i)$, so that $T_\nu^m: S_\beta^n \to S_\nu^m$ is not one-to-one.

(iii) $\nu^{m-1} \geqq \beta^n - 1 \Rightarrow R_\nu^m: S_\beta^n \to S_\nu^m$ is one-to-one.

Assume $\nu^{m-1} \geqq \beta^n - 1$, therefore $g(S_\beta^n) \geqq G(S_\nu^m)$. Now if we assume $R_\nu^m: S_\beta^n \to S_\nu^m$ is *not* one-to-one then there exists a $y \in S_\beta^n$, which may be taken positive, such that $R_\nu^m(y') = R_\nu^m(y)$. Choose $p \in S_\nu^m$ such that $p' = R_\nu^m(y)$. Thus

(19)
$$(p' + p)/2 \leqq y < y' < (p'' + p')/2$$

and

(20)
$$\gamma(y) = (y' - y)/y < \frac{p'' - p}{p' + p}$$

$$= \left(\frac{p'}{p' + p}\right)\left(\frac{p'' - p'}{p'}\right) + \left(\frac{p}{p' + p}\right)\left(\frac{p' - p}{p}\right)$$

$$\leqq \left(\frac{p'}{p' + p}\right)G(S_\nu^m) + \left(\frac{p}{p' + p}\right)G(S_\nu^m)$$

$$= G(S_\nu^m) \leqq g(S_\beta^n).$$

But $\gamma(y) < g(S_\beta^n)$ is a contradiction, so that R_ν^m is one-to-one.

   (iv) $\nu^{m-1} < \beta^n - 1 \Rightarrow R_\nu^m : S_\beta^n \to S_\nu^m$ is not one-to-one.

Assume $\nu^{m-1} < \beta^n - 1$, and since ν, β, m, n are integers

$$(21) \qquad \nu^{m-1} < \beta^n - 3/2.$$

Thus

$$(22) \qquad 0 < \nu^{1-m} - \beta^{-n} - (3/2)\nu^{1-m}\beta^{-n},$$

$$(23) \qquad \frac{1 + (1/2)\nu^{1-m}}{1 - \beta^{-n}} < 1 + (3/2)\nu^{1-m}.$$

By Theorem 1 there exist integers $i, j > 0$ such that

$$(24) \qquad (1 + (1/2)\nu^{1-m})/(1 - \beta^{-n}) < \beta^i/\nu^j < (1 + (3/2)\nu^{1-m})$$

and

$$(25) \qquad \nu^j(1 + (1/2)\nu^{1-m}) < \beta^i(1 - \beta^{-n}) < \beta^i < \nu^j(1 + (3/2)\nu^{1-m})$$

so that $R_\nu^m(\beta^i(1 - \beta^{-n})) = R_\nu^m(\beta^i)$.

This proves (iv) and completes the proof of the conditions for a one-to-one conversion mapping.

   (v) $\beta^{n-1} \geqq \nu^m - 1 \Rightarrow T_\nu^m : S_\beta^n \to S_\nu^m$ is onto.

Given $\beta^{n-1} \geqq \nu^m - 1$ or equivalently $G(S_\beta^n) \leqq g(S_\nu^m)$, assume that $T_\nu^m : S_\beta^n \to S_\nu^m$ is *not* onto. Then there must exist $x \in S_\beta^n$, $y \in S_\nu^m$, where $x, y > 0$ may be assumed, such that

$$(26) \qquad T_\nu^m(x) \leqq x < y < y' \leqq T_\nu^m(x') \leqq x'$$

(where the successors are taken in the appropriate spaces). Thus $\gamma_\beta^n(x) > \gamma_\nu^m(y)$, which contradicts $G(S_\beta^n) \leqq g(S_\nu^m)$. Hence T_ν^m is onto.

   (vi) $\beta^{n-1} < \nu^m - 1 \Rightarrow T_\nu^m : S_\beta^n \to S_\nu^m$ is not onto.

Given $g(S_\nu^m) < G(S_\beta^n)$, then by (ii) the truncation map of S_ν^m into S_β^n is not one-to-one, so that for some $y \in S_\nu^m$, $x \in S_\beta^n$ (again $x, y > 0$ may be assumed)

$$(27) \qquad x \leqq y < y' < x'.$$

Now if $y > x$, then $T_\nu^m : S_\beta^n \to S_\nu^m$ is not onto. If $y = x$ by Theorem 1 choose $i, j > 0$ such that

$$(28) \qquad x'/y' > \nu^j/\beta^i > 1$$

and using Corollary 2 of §II

$$(29) \qquad \beta^i x < \nu^j y < [\nu^j y]' < [\beta^i x]'.$$

Hence $\nu^j y \in S_\nu^m$ is not covered and T_ν^m is not onto.

(vii) $\beta^{n-1} \geqq \nu^m - 1 \Rightarrow R_\nu^m \colon S_\beta^n \to S_\nu^m$ is onto.

Assume $R_\nu^m \colon S_\beta^n \to S_\nu^m$ is *not* onto. Then for some $x \in S_\beta^n$, $y \in S_\nu^m$, where $x, y > 0$ may again be assumed,

$$(30) \qquad x < (y' + y)/2 < (y'' + y')/2 \leqq x'$$

so that

$$(31) \qquad \frac{x' - x}{x} > \frac{y'' - y}{y' + y} = \frac{y'}{y' + y} \gamma_\nu^m(y') + \frac{y}{y' + y} \gamma_\nu^m(y).$$

Thus $G(S_\beta^n) > g(S_\nu^m)$ and $\beta^{n-1} < \nu^m - 1$, which proves (vii).

(viii) $\beta^{n-1} < \nu^m - 1 \Rightarrow R_\nu^m \colon S_\beta^n \to S_\nu^m$ is not onto.

Given $\beta^{n-1} < \nu^m - 1$, the integrity of β, ν, n, m implies $\beta^{n-1} < \nu^m - 3/2$, so that

$$(32) \qquad \frac{\nu^m - 1/2}{\beta^{n-1} + 1} < \frac{\nu^m - 3/2}{\beta^{n-1}}.$$

By Theorem 1 we may choose $i, j > 0$ such that

$$(33) \qquad \frac{\nu^m - 1/2}{\beta^{n-1} + 1} < \frac{\beta^i}{\nu^j} < \frac{\nu^m - 3/2}{\beta^{n-1}}.$$

Hence

$$(34) \qquad \begin{aligned} \beta^{i+n-1} &< \nu^j(\nu^m - 3/2), \\ [\beta^{i+n-1}]' &= \beta^i(\beta^{n-1} + 1) > \nu^j(\nu^m - 1/2). \end{aligned}$$

Thus

$$R_\nu^m(\beta^{i+n-1}) < \nu^j(\nu^m - 1) \quad \text{and} \quad R_\nu^m([\beta^{i+n-1}]') \geqq \nu^{j+m} = [\nu^j(\nu^m - 1)]'$$

so that $\nu^j(\nu^m - 1) \in S_\nu^m$ is not covered and $R_\nu^m \colon S_\beta^n \to S_\nu^m$ is not onto.

This completes the proof of the Base Conversion Theorem.

We close with the following two corollaries.

COROLLARY 1. *When $\beta^i \neq \nu^j$ for any positive integers i, j, then neither the truncation conversion mapping nor the rounding conversion mapping of S_β^n into S_ν^m is both one-to-one and onto.*

Indeed, a stronger result than Corollary 1 can be shown for arbitrary bases β, ν. This result called the *Floating Point Incompatibility Principle*, states that the truncation or rounding conversion mapping of S_β^n into S_ν^m is both one-to-one and onto only when the mapping is the identity, i.e., $\beta = \nu$ and $n = m$.

COROLLARY 2. *The truncation (rounding) conversion mapping of S_β^n into S_γ^m is onto if and only if the truncation (rounding) conversion mapping of S_γ^m into S_β^n is one-to-one.*

REFERENCES

1. D. W. Matula, *Base conversion mappings*, Proc. Amer. Fed. Information Processing Soc. **30** (1967), 311–318.

2. G. H. Hardy and E. M. Wright, *An introduction to the theory of numbers*, 3rd ed., Clarendon Press, Oxford, 1954.

WASHINGTON UNIVERSITY