

Chapter 1

Introduction

The purposes of this initial chapter are to present some basic definitions about infinite electrical networks, to show by examples that their behaviors can be quite different from that of finite networks, and to indicate how they approximately represent various partial differential equations in infinite domains. Finally, we explain how the transient responses of linear RLC networks can be derived from the theory of purely resistive networks; this is of interest because most of the results of this book are established in the context of resistive networks.

1.1 Notations and Terminology

Let us start by reviewing some symbols and phraseology so as to dispel possible ambiguities in our subsequent discussions. We follow customary usage; hence, this section may be skipped and referred to only if the need arises. Also, an Index of Symbols is appended for the more commonly occurring notations in this book; it cites the pages on which they are defined.

Let X be a set. X is called denumerably infinite or just denumerable if its members can be placed in a one-to-one correspondence with all the natural numbers: $0, 1, 2, \dots$. X is called countable if it is either finite or denumerable. In this book the set of branches of any network will always be countable.

The notation $\{x \in X : P(x)\}$, or simply $\{x : P(x)\}$ if X is understood, denotes the set of all $x \in X$ for which the proposition $P(x)$ concerning x is true. We will at times be constructing infinite sets, such as infinite

paths through graphs, by selecting elements from collections of sets; when doing so, we will be tacitly invoking the axiom of choice [1]. $|X|$ denotes the cardinality of X . \emptyset denotes the void set. A singleton is a set having exactly one member x and may be denoted by $\{x\}$. If Y is also a set, then $X \cup Y$ and $X \cap Y$ are the usual union and intersection sets. However, $X \subset Y$ (instead of $X \subseteq Y$) will indicate that all elements of X are members of Y , and X may equal Y ; if X is smaller than Y , X is said to be a proper subset of Y . $X \setminus Y$ is the set of all members of X that are not in Y . The notation $x, y, z, \dots \in X$ means that all the elements x, y, z, \dots are members of X ; $x, y, z, \dots \notin X$ means that none are in X . If X is a set in a metric space, X° is the interior of X , and \bar{X} is the closure of X . However, if α is a complex number, $\bar{\alpha}$ is its complex conjugate. If X is a linear space, $-X = \{y: y = -x, x \in X\}$.

$\{x_i\}_{i \in I}$ denotes a collection of indexed elements, where the index i traverses the set I . If I is a set of integers, $\{x_i\}_{i \in I}$ is a sequence with the order induced by those integers. A sequence may be represented in several ways: If, for example, I is the set of natural numbers, we may write $\{x_i\}_{i=0}^\infty$ or $\{x_0, x_1, x_2, \dots\}$ or (x_0, x_1, x_2, \dots) or just $[x_i]$. More generally, if a and b are two integers with $a < b$ and $b - a = n - 1$, $\{x_i\}_{i=a}^b$ is a finite sequence or synonymously an n -tuple. If $a = -\infty$ and/or $b = \infty$, then $\{x_i\}_{i=a}^b$ denotes an infinite sequence with I being the set of all integers from a to b ; it is understood that there are no entries $x_{-\infty}$ or x_∞ . A sequence is often called a vector. The support of a vector whose components are members of a linear space is the set of indices for its nonzero components; it may happen that those indices correspond to branches of a network, in which case we think of the support as a set of branches. Furthermore, $\cup_{i=a}^b X_i$ and $\cap_{i=a}^b X_i$ denote respectively the union and intersection of all members X_i of a sequence $\{X_i\}_{i=a}^b$ of sets. The Cartesian product $X_1 \times \dots \times X_n$ of a finite number of sets is the set of all n -tuples (x_1, \dots, x_n) , where $x_i \in X_i$ and $i = 1, \dots, n$. A partition of a set X is a collection of subsets X_i of X such that every $x \in X$ is a member of one and only one X_i .

N denotes the set $\{0, 1, 2, \dots\}$ of all natural numbers, and Z the set $\{\dots, -1, 0, 1, \dots\}$ of all integers. We use Greek iota for the imaginary unit: $\iota = \sqrt{-1}$. R^n and C^n denote respectively the real and complex, n -dimensional, Euclidean spaces. Thus, an arbitrary point $x \in R^n$ (or $x \in C^n$) is an n -tuple $x = (x_1, \dots, x_n)$ of real (respectively, complex) numbers x_i . The Euclidean norm of x is $\|x\| = [\sum |x_i|^2]^{1/2}$. At times we use a different norm for x : $\|x\|_1 = |x_1| + \dots + |x_n|$. A compact subset of R^n or C^n is a closed bounded set. A point $x \in R^n$ is called a

1.1 Notations and Terminology

3

lattice point or an n -dimensional integer if every component x_i of x is an integer. Z^n denotes the set of lattice points in R^n . Furthermore, l_{2r} and l_2 denote respectively the real and complex, infinite-dimensional, Hilbert coordinate spaces. Any member x of these spaces is an infinite sequence of real or respectively complex numbers x_i such that $\|x\| = [\sum |x_i|^2]^{1/2} < \infty$. In this case we say that the infinite sequence x is quadratically summable (instead of square summable). The inner product of $x, y \in X$ is $(x, y) = \sum x_i \bar{y}_i$. Actually, the symbol (x, y) has several meanings: an open interval in R^1 , or a point in R^2 , or a two-element sequence, or an inner product; we will be explicit as to which meaning we happen to be using. Also, $\text{sgn}(x)$ denotes the sign of $x \in R^1$.

A rule f that assigns one or more elements in a set Y to each element in another set X is called a relation or correspondence. Thus, f determines a subset of $X \times Y$ called the graph of f and consisting of all 2-tuples (x, y) , where $y = fx$ is an element assigned to x by f . We also use the notation $f: X \rightsquigarrow Y$ as well as $f: x \mapsto y$ to denote the rule, the sets for which the rule is defined, and the typical elements related by the rule. The domain of f is the set of all elements x for which f is defined, in this case X ; x is called the independent variable. The range of f is the subset of Y consisting of all elements y that are assigned by f to members of X ; y is called the dependent variable.

A function f is a relation $f: X \rightsquigarrow Y$ that assigns precisely one element in Y to each member of X ; in other words, f is a function if and only if the equations $y = fx$ and $w = fx$ imply that $y = w$. We may emphasize this property by referring to f as single-valued (in contradistinction to a multivalued function, which perforce is a relation). Synonymous with function are operator, mapping, and transformation. We say that $f: x \mapsto y$ maps, carries, transforms, or converts x into y and that f is a function on or from X into Y or a mapping of X into Y . We also say that f is Y -valued. When Y is R^1 or C^1 , the phrase Y -valued is replaced by real-valued or complex-valued respectively, and f is sometimes called a *functional*. If W is a subset of X , the symbol $f(W)$ denotes the set $\{y \in Y: y = fx, x \in W\}$. The function g that is defined only on W and coincides with f on W is called the restriction of f to W .

A function f is said to be one-to-one or injective and is also called an injection if the equations $fx = y$ and $fu = y$ imply that $x = u$. In this case we have the inverse function $f^{-1}: y \mapsto x$, which maps the range of f into X . A function $f: X \rightsquigarrow Y$ is said to be onto or surjective and is also called a surjection if the range of f coincides with Y . If $f: X \rightsquigarrow Y$

is both injective and surjective, it is said to be bijective or a bijection.

We denote the elements in the range of f by the alternative notations $y = fx = f(x) = \langle f, x \rangle$. On occasion, it will be convenient to violate this symbolism by using $f(x)$ to denote the function f rather than its range value, as is commonly done in classical mathematics. Whenever we do so, it will be clear from the context what is meant. On still other occasions, we use the dot notation $f(\cdot) = \langle f, \cdot \rangle$ in order to indicate where the independent variable should appear.

The symbols sup and inf indicate respectively the supremum and infimum of a subset of R^1 , whereas max and min indicate respectively the maximum and minimum of a finite subset of R^1 . On the other hand, ess sup and ess inf indicate the essential supremum and essential infimum of the range of a measurable function. Also, lim sup and lim inf symbolize the limit superior and limit inferior.

Our notation for asymptotic behavior is standard. Let $X \subset R^n$. Let f and g be two mappings of X into R^1 . Assume that $x \in X$ passes to a limit $a \in X$, and consider the ratio $f(x)/g(x)$. The notation $f(x) = o(g(x))$ means that the ratio tends to 0, $f(x) = O(g(x))$ means that the ratio remains bounded, and $f(x) \sim g(x)$ means that the ratio tends to 1.

The span of a subset X of a linear space is the set of all (finite) linear combinations of elements in X . When X and Y are Hilbert spaces, the symbol $[X; Y]$ denotes the linear space of all continuous linear mappings of X into Y ; $[X; Y]$ is supplied with the uniform – synonymously, norm – topology [94, page 410].

When X and Y are Euclidean spaces, any linear mapping $f : X \rightsquigarrow Y$ has a unique representation as a matrix with respect to the natural bases in X and Y . We shall often use the same symbol to represent f and its matrix when those bases are understood. Similarly, when X and Y are the Hilbert coordinate space, a member of $[X; Y]$ and its natural matrix representation may be denoted by the same symbol.

Electrical units and their symbols are as follows: volt V, ampere A, ohm Ω , mho \mathcal{U} , farad F, and henry H. We use some of these symbols (but in italic font) for other purposes as well. The context in which they are used should dispel any possible confusion.

♣ denotes the end of a proof or example.

1.2 Countable Graphs

In the mathematical sense an infinite electrical network is obtained by placing a certain analytical structure upon an infinite graph. With regard

1.2 Countable Graphs

5

to graphs, we gather in this section most of the definitions that we shall employ.

Let K and J be finite or denumerably infinite index sets. Let $\mathcal{N} = \{n_k\}_{k \in K}$ be an indexed set, where n_k and n_m are distinct if $k \neq m$. The members of \mathcal{N} are called *nodes*. Also, let $\mathcal{B} = \{b_j\}_{j \in J}$ be an indexed set of two-entry families $b_j = \{n_k, n_m\}$ of nodes. The members of \mathcal{B} are called *branches*. That each b_j is a “family” instead of a set means that $k = m$ is allowed in $b_j = \{n_k, n_m\}$; such a branch $\{n_k, n_k\}$ will be called a *self-loop*. Also allowed is $b_i = b_j$ with $i \neq j$; in this case b_i and b_j are said to be in *parallel*. We will at times suppress the indexings of \mathcal{N} and \mathcal{B} , but it will be understood that such indexings exist.

We say that $b_j = \{n_k, n_m\}$ is *incident to* n_k and to n_m and similarly n_k and n_m are *incident to* b_j . A node is *isolated* if it is not incident to any branch. A node is called *infinite* if there is an infinity of branches incident to the node; otherwise, it is called *finite*. The cardinality of the set of branches incident to a node is called the *degree* of that node.

A *countable graph* \mathcal{G} is defined as the pair $\mathcal{G} = (\mathcal{B}, \mathcal{N})$. \mathcal{G} is called a *finite graph* if both \mathcal{B} and \mathcal{N} are finite sets; otherwise, it is called *countably infinite*. \mathcal{G} is also called *locally finite* if all its nodes are finite. If $n \in \mathcal{N}$ or $b \in \mathcal{B}$, we shall say that n is in \mathcal{G} or b is in \mathcal{G} and that \mathcal{G} *has* or *contains* n or b .

This definition of \mathcal{G} is quite standard. However, it is not amenable to a generalization we shall make in Chapter 5, where the idea of a transfinite graph is introduced. For that purpose we present in the next section a different definition, which will be equivalent to the present one if isolated nodes are disallowed.

Given the graph $\mathcal{G} = (\mathcal{B}, \mathcal{N})$, let \mathcal{N}_* be a subset of \mathcal{N} and let \mathcal{B}_* be a subset of \mathcal{B} such that all the nodes of all the branches of \mathcal{B}_* appear in \mathcal{N}_* . Then, the graph $(\mathcal{N}_*, \mathcal{B}_*)$ is called a *subgraph* of \mathcal{G} . If \mathcal{B}_* is given and \mathcal{N}_* is the set of all nodes to which the branches in \mathcal{B}_* are incident, then $\mathcal{G}_* = (\mathcal{N}_*, \mathcal{B}_*)$ is called the *subgraph induced by* \mathcal{B}_* or a *branch-induced subgraph*. Similarly, if \mathcal{N}_* is given and \mathcal{B}_* is the set of all branches all of whose nodes belong to \mathcal{N}_* , then \mathcal{G}_* is called the *subgraph induced by* \mathcal{N}_* or a *node-induced subgraph*.

Let \mathcal{H} and \mathcal{K} be two subgraphs of a given graph \mathcal{G} . The *union* $\mathcal{H} \cup \mathcal{K}$ is the subgraph whose node set (branch set) is the union of the node sets (respectively, branch sets) of \mathcal{H} and \mathcal{K} . The *intersection* $\mathcal{H} \cap \mathcal{K}$ is defined the same way except that “union” is replaced by “intersection.” When $\mathcal{H} \cap \mathcal{K}$ has at least one node, we say that \mathcal{H} and \mathcal{K} *intersect* or *meet*. $\mathcal{H} - \mathcal{K}$ denotes the subgraph of \mathcal{H} induced by all the nodes in \mathcal{H} that

are not in \mathcal{K} . At times, we replace \mathcal{K} by a node set \mathcal{S} , in which case it is understood that $\mathcal{H} - \mathcal{S} = \mathcal{H} - \mathcal{K}$, where now \mathcal{K} is the subgraph $\{\emptyset, \mathcal{S}\}$. On the other hand, if b is a branch of \mathcal{H} , $\mathcal{H} - b$ denotes the subgraph obtained by removing b from the branch set of \mathcal{H} but leaving the node set unaltered. Finally, $\mathcal{H} \subset \mathcal{K}$ denotes that \mathcal{H} is a subgraph of \mathcal{K} .

Let \mathcal{M} be subgraph of \mathcal{G} . A *partition* $\{\mathcal{M}_p\}$ of \mathcal{M} is a collection of subgraphs \mathcal{M}_p of \mathcal{M} such that every branch of \mathcal{M} appears in one and only one \mathcal{M}_p and $\mathcal{M} = \bigcup \mathcal{M}_p$. A branch b is said to be *adjacent* to a subgraph \mathcal{K} if b is not in \mathcal{K} and at least one node of b is in \mathcal{K} . On the other hand, a node n_0 and a subgraph \mathcal{M} are called *adjacent* if n_0 is not in \mathcal{M} and there exists a branch joining n_0 to a node of \mathcal{M} . A branch is said to *join* two subgraphs of \mathcal{G} if those subgraphs do not meet and b has a node in each of the subgraphs.

A *path* is an alternating sequence of nodes n_{k_m} and branches b_{j_m}

$$\{\dots, n_{k_m}, b_{j_m}, n_{k_{m+1}}, b_{j_{m+1}}, \dots\} \quad (1.1)$$

wherein the indices m are restricted to the integers (as distinct from the transfinite ordinals) and number the elements of (1.1) sequentially as indicated, no node appears more than once, and each branch is incident to the two nodes immediately preceding and succeeding it in the sequence. Consequently, no branch appears more than once in the sequence. The sequence may be either finite, or one-way infinite, or two-way infinite, in which case it is called a *finite path*, or a *one-ended path*, or an *endless path* respectively. Moreover, if the sequence terminates in either direction, it is required that it terminate at a node; that node is called an *end node* of the path. A finite path with the terminal nodes n_a and n_b is called an $n_a n_b$ -*path*. If the path has at least one branch, it is called *non-trivial*; a *trivial path* is a singleton $\{n_{j_m}\}$. Two paths are called *totally disjoint* if no node is contained in both paths.

A *loop* is defined exactly as is a finite nontrivial path except that its two end nodes are required to be the same node. Thus, a self-loop $\{n, n\}$ can be viewed as a special case of a loop in the following way: $\{n, \{n, n\}, n\}$.

Two nodes n_a and n_b are said to be *connected* if there is a path that contains both nodes; this is the same as saying that there is a finite path with n_a and n_b as its two end nodes. If this is true for every two nodes in \mathcal{G} , then \mathcal{G} itself is called *connected*. Two branches are said to be *connected* if their nodes are connected. A *component* of a graph is a subgraph that is maximal with respect to connectedness. A *forest* \mathcal{F} is a graph containing no loops; \mathcal{F} is called a *tree* if it is also connected.

1.3 0-Graphs

7

When \mathcal{F} is a subgraph of \mathcal{G} , it is said to be *spanning in \mathcal{G}* if it contains every node in \mathcal{G} . An *end node* of \mathcal{F} is a node whose degree with respect to \mathcal{F} is 1, and its incident branch is called an *end branch*.

1.3 0-Graphs

In order to facilitate our discussion of transfinite graphs in Chapter 5, we now define a special kind of (finite or infinite) graph in an unusual way. The reasons why we resort to an unconventional approach to graphs are the following. The conventional definition of a graph starts with a set of nodes and then defines the branches as unordered pairs of nodes. This is the reverse of how a physical electrical network is usually constructed. Indeed, beginning with a variety of electrical elements, each having two or more terminals, one constructs the nodes of a physical network by shorting the terminals together; in particular, starting from a set of branches one defines the nodes as shorts between various branch ends. Moreover, for our purposes there is a more forceful misgiving about the conventional definition; it arises when one tries to extend it to transfinite graphs, as was done in [177]. To accommodate connections between conventional nodes and the extremities of a transfinite graph, one is led to defining a whole hierarchy of “extended branches” involving pairs of nodes and extremities of various ranks. This results in a ponderous set of definitions.

Some simplification can be achieved if we formalize how a physical network is usually put together. We can view each terminal of an electrical element as being an “elementary tip” and define a branch as being a pair of such tips. Nodes are then sets of such tips shorted together. Next, the extremities of an ordinary infinite graph are considered to be “tips of rank zero,” and a “transfinite graph of rank one” is constructed by shorting together elementary tips and tips of rank zero. This process can be continued to obtain transfinite graphs of higher ranks. In the end, branches remain as they are in an ordinary graph and nodes are merely shorts between various kinds of “tips.”

Adopting this alternative approach, we now present our specialized definition of an ordinary (finite or infinite) graph. Let \mathcal{T} be a finite set with an even number of elements or a denumerably infinite set. Call each element of \mathcal{T} an *elementary tip* or just a *tip* for short. Partition \mathcal{T} into subsets of two elementary tips each and call each subset a *branch*. Thus, a branch has the form $\{t_1, t_2\}$, where the t_1 and t_2 are its two tips. \mathcal{B} denotes the set of all branches; it is countable. As before, we usually index the branches with the set J ; thus, $\mathcal{B} = \{b_j : j \in J\}$.

Also, partition \mathcal{T} in an arbitrary way: $\mathcal{T} = \cup n_k$. Thus, $n_k \cap n_m$ is void if $k \neq m$. The subsets n_k will be called *0-nodes* or simply *nodes*. (Later on, we will view an n_k having two or more members as being an electrical connection that “shorts together” the tips in n_k .) As a terminology we will need in Chapters 3 and 5, we say that a 0-node *embraces* itself as well as its tips. We will often index the 0-nodes with the set K . \mathcal{N}^0 will denote the set of 0-nodes: $\mathcal{N}^0 = \{n_k : k \in K\}$.

A *0-graph* is defined as the pair $\mathcal{G}^0 = (\mathcal{B}, \mathcal{N}^0)$. Note that we do not display \mathcal{G}^0 as the triplet $(\mathcal{T}, \mathcal{B}, \mathcal{N}^0)$ simply because \mathcal{T} can be recovered as the union of all the branches. More important, observe that the present definition implies a well-defined mapping from every branch to either a single 0-node or to a pair of 0-nodes, namely, the 0-node or 0-nodes in which the two tips of the branch are found. This means that a 0-graph becomes a graph when we use that mapping to define the *incident* node or nodes for each branch. It follows that all the definitions pertaining to a graph can be transferred to 0-graphs. This we do. Later on, we shall refer to a path (1.1) in a 0-graph as a *0-path* and will say that it *embraces* itself, all its elements, and also all the elementary tips in its branches. Also, note that there are no isolated nodes according to this unconventional definition of a graph.

However, a subgraph of a 0-graph may not be a 0-graph because, for instance, a subgraph may have isolated nodes. Nonetheless, we can modify the definition of a branch-induced subgraph to generate a 0-graph from a specified subset \mathcal{B}_* of \mathcal{B} . Let \mathcal{T}_* be the union of the branches in \mathcal{B}_* (i.e., the set of all tips of the branches of \mathcal{B}_*). Delete from each node in \mathcal{N}^0 all tips that are not in \mathcal{T}_* . The set of all such reduced but nonvoid nodes is denoted by \mathcal{N}_*^0 . Finally, $\mathcal{G}_*^0 = (\mathcal{B}_*, \mathcal{N}_*^0)$ is a 0-graph; we call it the *reduction of \mathcal{G}^0 induced by \mathcal{B}_** or a *\mathcal{B}_* -reduced 0-graph* or simply a *reduced 0-graph*.

For 0-graphs *0-connectedness* means the same thing as connectedness. A *0-section* of a 0-graph \mathcal{G}^0 is a reduction of \mathcal{G}^0 induced by a maximal set of branches that are pairwise 0-connected. At this stage of our definitions, a 0-section is equivalent to a component of \mathcal{G}^0 , but in Section 3.2 we will generalize the idea of connectedness and thereby introduce a new meaning for a 0-section.

Before leaving this section, it is worth emphasizing that the introduction of 0-graphs anticipates our construction of transfinite graphs but is not essential. Another exposition of transfinite graphs is given in [182] wherein the role of the 0-graphs is borne by ordinary graphs.

1.4 Electrical Networks

An *electrical network* is defined by assigning to each branch of a given graph or 0-graph several electrical parameters connected together in a specified way. Moreover, an orientation is assigned to each branch with respect to which the direction of the branch current and branch voltage is measured. That current or voltage is positive when the orientation and direction agree and is negative when they disagree.

A *resistive branch with independent voltage and current sources* is indicated in Figure 1.1 in two entirely equivalent ways. The symbols in that diagram denote real numbers, but we will also use r , e , and h to denote the kind of parameter at hand – thereby avoiding a more cumbersome notation. The branch's orientation is the same as that indicated by the arrow for the branch current i or the plus-to-minus direction for the branch voltage v . r is the *branch resistance*, which is always a positive number (nonzero and finite); its reciprocal $g = 1/r$ is the *branch conductance*. e is the *branch voltage source* measured positively as a voltage rise (from $-$ to $+$) with respect to the branch's orientation. h is the *branch current source* measured positively in the direction opposite to the branch's orientation. These quantities are related by *Ohm's law*:

$$v + e = r(i + h) \quad (1.2)$$

or equivalently

$$i + h = g(v + e). \quad (1.3)$$

The two connections of Figure 1.1 are equivalent because e and h are both independent of any other voltage or current. The customary units of volts (V), amperes (A), ohms (Ω), and mhos (\mathcal{U}) for voltages, currents, resistances, and conductances are adopted throughout. Actually, since r and g are real numbers which enter as multiplying factors in (1.2) and (1.3), we have in fact defined a linear element. More generally, however, r and g may be mappings of the real line R^1 into R^1 ; when these functions are different from multiplications by constants, we have instead a nonlinear resistance $r(\cdot)$ and a nonlinear conductance $g(\cdot)$. By rearranging (1.2) and (1.3), we can represent any branch in two more equivalent forms. *Thevenin's form* is shown in Figure 1.1(b), and *Norton's form* is shown in Figure 1.1(c). All these forms impose exactly the same relationship between v , i , e , and h .

There are other kinds of branches. A *pure voltage source* is shown in Figure 1.2(a). For this kind of branch, it is required that n_1 and n_2 be distinct nodes. Also, $v = -e$, and e is required to be nonzero; on

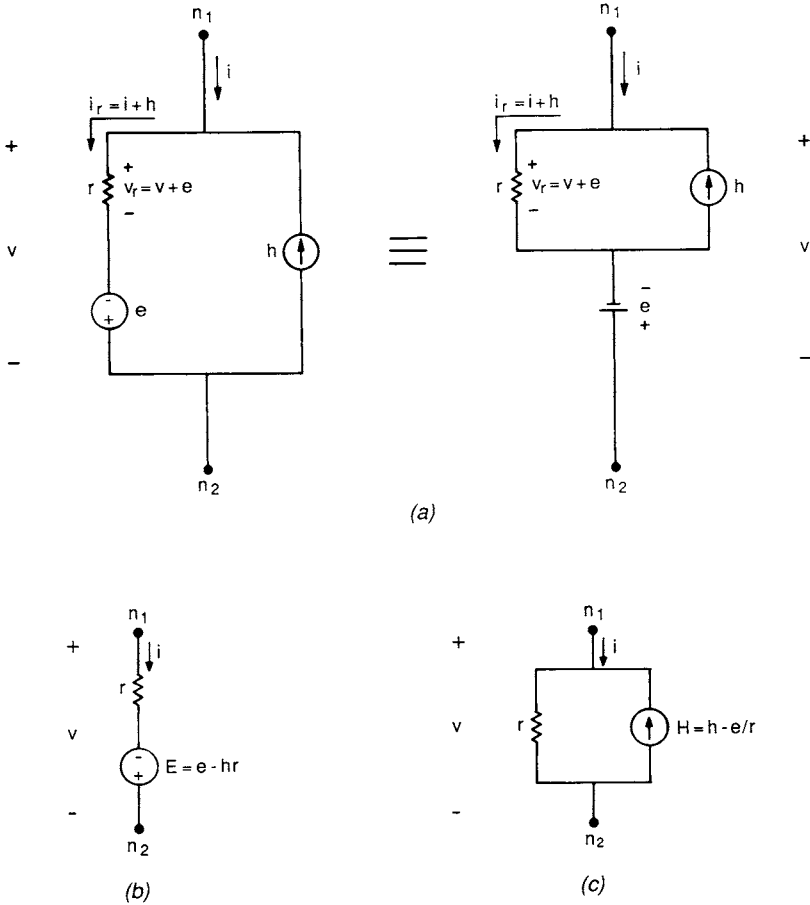


Figure 1.1. (a) Two equivalent forms of a resistive branch. The indicated quantities, other than the nodes n_1 and n_2 , are real numbers and r is positive. (Also shown are two alternative symbols for the voltage source e . Both are used in this book.) (b) The Thevenin equivalent form. (c) The Norton equivalent form.

the other hand, i is determined by the network in which the branch appears. When e is set equal to zero, this branch becomes a *short circuit* or synonymously a *short*; we will not allow a short circuit to be a branch and will coalesce n_1 and n_2 into a single node by replacing n_1 and n_2 if need be.

A *pure current source* is indicated in Figure 1.2(b). Now, $i = -h$, where h is required to be nonzero; in this case, it is v that is determined by the rest of the network. When h is set equal to zero, we get an