IBM

Program Product

APL\360 Primer

Program Nos. 5734-XM1
            5736-XM1

This primer provides an introduction to the
APL/360 system and to the APL programming
language. It discusses the mechanics of using
the system, and shows how to write represent-
ative programs for a variety of applications.
The treatment is elementary. For a discussion
of advanced system features and available opera-
tions, refer to the APL/360 User's Manual
(GH20-0683).

ACKNOWLEDGEMENTS

This Primer was written by Paul Berry of the IBM Research Division, T. J. Watson Research Center, Yorktown Heights, New York. The text benefitted greatly from the suggestions, criticisms, and comments of the readers of the various early drafts, and especially those of A. D. Falkoff, K. E. Iverson, J. C. McPherson, L. M. Breed, and R. H. Lathwell. The final draft was completed with the technical assistance of Miss C. Conroy.

The text was entered at a terminal, and edited and composed by an APL\360 program. The master pages were then typed by the computer at the terminal, using an IBM courier typing element.

i

C O N T E N T S
= = = = = = = =
PART I.

PART I

1: INTRODUCTION

Why an APL Primer?

The APL\360 System puts an advanced computing system
within the reach of a wide range of users. APL\360 is
distinguished from earlier systems of this type by its speed
and power, and by the radical simplicity of the instructions
which control it. This combination makes APL well suited not
only to the advanced scientific or technical user, but also
to the occasional user and to the user with little or no
previous experience with computers.

This primer is intended to provide an introduction to
the APL\360 System and to the APL programming language. It
will show you the mechanics of using the system, and how to
write effective programs to cover a wide range of
applications. It explains in detail many points which the
experienced user will find obvious--and you may therefore
prefer to skip some portions.

Because this is a primer, little use will be made of a
number of the more advanced features of the system; the
primer doesn't describe all of the operations available, and
mentions only a few of the specialized applications that are
possible using APL. However, even at this rather elementary
level, you will already have at your command all you will
need for a wide range of uses--and frequently more than was
available even to the experienced users of earlier systems.
If you subsequently go on to more advanced material, you
will learn ways in which the programs included in this
primer could have been made neater or simpler or more
general. But that is beyond the scope of a primer. Complete
definitions of all of the operations in the APL language and
all of the features of the APL\360 System may be found in
the APL\360 User's Manual. Here we are concerned with provid-
ing you with a basic orientation to the way the system is
used, and arming you with the fundamental skills needed to
make APL\360 work effectively for you.

What is a Remote-Terminal Time-Sharing System?

APL\360 is a time-sharing system with remote terminals.
Let's consider each of those terms in turn.

"Remote terminal" means that you don't come to the
computer in order to use it. Instead, you use a special kind

of electric typewriter installed wherever it is convenient. This typewriter is equipped for tele-processing: in addition to serving as a regular office typewriter, it can be connected by telephone to a large computer located miles away. When you type something on the typewriter, what you type is transmitted to the computer. Within seconds you can receive the response transmitted over the telephone lines from a computer you have never seen. The computer is able to take control of the typewriter and cause it to type the computer's reply. A typewriter that is connected to a computer in this fashion is called a "terminal" of the computer. It is "remote" because the connection is by way of the public telephone lines, so that the typewriter can be located anywhere that a telephone line can reach.

"Time sharing" means that the central computer is capable of serving many customers at once. Actually it serves them in rotation; each one gets a tiny fraction of the computer's time, but the computer's operating speed is so high that often there is no appreciable delay between the time you type your request and the time the computer types its response. Even for problems of moderate size, a response may be received within a few seconds. Time sharing permits you to have apparently continuous use of a large computer, while paying a charge based on the far smaller amount of time that the computer is actually working on your problem. APL\360 has a conversational style, which permits you to make a request, get an immediate answer, then make another request, and so on. This would be prohibitively expensive if you had to tie up an entire computer to do it. Time sharing serves both to keep the computer in efficient use and to share its cost among many customers.

The letters APL designate the programming language that is the outgrowth of the work of K. E. Iverson, first at Harvard and then at IBM. The name comes from the initials of his book A Programming Language (New York: Wiley, 1962). APL\360 is the computing system which uses this language with IBM System/360 computers.

What Is a Computer Program?

A program is a set of instructions that tell a computer how to do something. A computer has to work from coded instructions which are usually stored inside it. When you want a job done, you must tell the computer precisely what you want it to do; no instructions, no work. The word

"program" has been used in this sense only since the advent of the computer. But the underlying idea of a set of precise instructions that are to be carried out literally and in sequence is older and more familiar. A cook book is an obvious example of an attempt to summarize, in order, those things that the cook must do in order to produce an unfamiliar dish. What is different about a computer program is the speed with which the computer can carry out its instructions, and the literal faithfulness with which the computer follows what it is told.

Sometimes the literalness of the computer requires you to be more precise than you would be if you were simply giving directions to a friend. If the instruction you give a computer can be carried out, the computer will carry it out, regardless of whether it is what you really had in mind. So you have to be careful to state your instructions in a way that correctly describes what you want. If the instruction is wrongly spelled, or otherwise impossible to accept as stated, the computer will stop and report what you instructed it and why it cannot proceed. Human beings might hazard a guess at what you meant by an incorrect instruction, but the computer doesn't.

The computer has to be able to understand the instructions you give it. Computers do not understand English; although they may be programmed to recognize a handful of English words, the natural language is too rich, too complex, and too ambiguous for them. Moreover, English is ill-suited to describe many of the things that you might want to ask a computer to do. Calculations can be described far more neatly, clearly, and briefly by the symbols of arithmetic. That is why we describe a calculation by a formula rather than in English words.

The designers of the traditional notations of arithmetic and algebra did not foresee all of the things you might want to ask a computer to do, and hence arithmetic and algebra do not contain all the symbols that are needed. This makes it necessary to have a special language for writing programs of computer instructions. That language is more extensive than conventional arithmetic, but much more restricted and precise than natural English. The language in which the computer is prepared to accept its instructions is its programming language.

What Is a Programming Language?

A programming language is the language in which you (the user) tell the computer what it must do. Most of this primer is concerned with APL, the programming language of the APL\360 System. A set of instructions written in APL can also be carried out by any person who knows the language: they don't have to be executed by a machine. A programming language is thus a way of stating a procedure, regardless of who or what actually executes the procedure.

Inside the hardware of the computer, all of its instructions and all of the data it works with are encoded as patterns of electronic pulses. This is the electronic language internal to the machine. You don't need to know anything about this language in order to use the computer. All of your communication with the computer will be in APL. The computer will then translate that into instructions in its own internal language, and then execute them. Internally, the machine works by carrying out only one very small and very simple step at a time. One APL instruction that you type may easily start a sequence of hundreds or even thousands of machine instructions before the work is completed. But these are executed so rapidly that the machine completes several thousand a second. The machine sets up its internal instructions in response to the brief instructions that you type in APL; you need never be concerned with the internal operation of the computer.

Power, Relevance, and Simplicity
In a Programming Language

A programming language should be relevant. That is, you should have to write only what is logically necessary to specify the job you want done. This may seem an obvious point, but many of the earlier programming languages forced the user to be concerned as much with the internal requirements of the machine as with his own statement of his problem. APL\360 takes care of those internal considerations automatically.

A programming language needs both power and simplicity. By power, we mean the ability to handle large or complicated tasks. By simplicity, we mean the ability to state what must be done briefly and neatly, in a way that is easy to read and easy to write. You might think that power and simplicity are competing requirements, so that if you have one you

can't have the other, but that is not so. Simplicity does
not mean that the computer is confined to doing simple
tasks, but that the user has a simple way to write his
instructions to the computer. The power of APL as a
programming language comes in part from its simplicity; it
is this simplicity that makes it simultaneously well suited
to the beginner and to the advanced user.

How to Read This Primer

If this is your first introduction to the use of
APL\360, after you've glanced through the primer to get a
general impression of its contents, it would probably be
wise to sit down at an APL\360 terminal with the book beside
you. Then you should try out the calculations and programs
in the text. Add variations or explorations of your own;
that's one of the advantages of a conversational system:
it's so easy to experiment. See for yourself how the system
responds to your instructions.

After this early stage, you will probably find it more
useful to come back to various passages as the need for them
arises; the table of contents and the index should help you
find what you need. You should also have a copy of the
APL\360 User's Manual, which gives complete but concise
descriptions of the APL operators and the features of the
APL\360 System. The manual includes coverage of a number of
advanced points which, for simplicity, are omitted from this
primer.

The two most distinctive and valuable characteristics
of the APL language are the way it treats arrays, and the
way it permits you to use a program as you would use a
mathematical function. Neither of these topics is mentioned
at all in Part I of the primer, since it seemed desirable to
lay a foundation of familiarity with other matters before
getting to them. But if you already feel familiar with these
topics and with their treatment in programming systems, you
may wish to look ahead to Chapter 16, where the treatment
of arrays is introduced, and to Chapter 25, where we take up
programs that can be used as functions. The examples in all
the earlier chapters may then be understood as applying also
to arrays of data, and could be written so that they behave
like functions.

APL\360 KEYBOARD

2: COMMUNICATING WITH THE COMPUTER

This chapter deals with some practical aspects of getting your terminal connected to the APL\360 System, communicating with the computer, and other such mundane but essential matters. If you are about to try a terminal for yourself, this chapter logically precedes the ones that describe the APL language and the way in which you carry out calculations in it. But if your interest is primarily in the APL language, you may wish to skip this chapter now, and return to it when you are ready to use an APL\360 terminal.

What Equipment Do You Need?

You need:

a communications terminal with an APL typeball

a data telephone or telephone coupler

an APL\360 account number.

The typewriter may be either an IBM 2741 or an IBM 1050. The 1050 permits the attachment of certain extra equipment, such as card readers and punches, but is slower in operation and more expensive. Thus unless you have special needs for the extra equipment, the 2741 is preferable, and that is the terminal to which this primer usually refers. A complete discussion of the alternative models and features usable with the APL\360 System may be found in Appendix E.

A data telephone may be rented from the telephone company. The system uses the Western Electric Dataset, Model 103-A2, or equivalent equipment.

The APL Typeface

The APL typing element provides both a full upper-case alphabet and the special symbols used in the APL programming language but not found on an ordinary office typewriter. The APL typeface was chosen so as to end confusion between the letter O and the number 0, or between the letter I and the number 1, or between the letter X and the sign that means multiplication. Three different styles of lettering

distinguish letters, numerals, and operation signs, as follows:

Alphabetics: * always capitalized
              * always italic
              * always condensed (higher than wide)
              * always with serifs

A B C D E F G H I J K L M N O P Q R S T U V W X Y Z

Numerals: * always upright
              * always condensed

1 2 3 4 5 6 7 8 9 0

Operators: * not condensed
              * upright (except for Greek letters)

$+ - \times \div * \lceil \lfloor \top \bot \wedge \vee = \neq < > \mid \rightarrow \leftarrow \iota \rho \epsilon \alpha \omega$

This typeface makes it quite clear whether any character is a letter, a numeral, or an operator sign. For instance, the phrase that indicates "the letter O times the letter X plus the letter I minus ten" can be typed

O×*X*+*I*-10

which leaves no doubt about which are the letters, which the numerals, and which the operator signs.

Establishing the Telephone Link

When you are ready to use APL\360 (assuming that you are using a 2741 terminal) you proceed as follows:

1. Turn on the typewriter power. Set the switch on the left side of the 2741 typewriter to COM (for "communicate") rather than to LCL (which stands for "local").

2. Depress the TALK button on the data telephone, and dial the number of the APL\360 computer.

3. When you hear a high-pitched tone, press the DATA button firmly and release it. Once DATA is pressed, you may cradle the telephone.

4. Type your sign-on.

When you first set the typewriter switch to COM, the keyboard is locked. When the connection to the computer is established, the keyboard will be unlocked, and you may type your sign-on. This is a right parenthesis followed by your user number and password (if any). When the computer receives your sign-on command, it acknowledges by typing these things: your <u>port</u> number (which of the computer's telephone extensions you reached), your name, the date and time, and the system identification. Sometimes you will also receive a broadcast message from the APL operator sent to all users who are signing on. Until your sign-on is accepted, you cannot do any other work. However, if you need to, you can type a message to the APL operator even before you're signed on (see p. 103). Here is a sample sign-on:

```
)4000000
OPR:  MERRY CHRISTMAS, HAPPY CHANUKKAH, OR SEASON'S GREETINGS.
009)        PCBERRY 12/23/67   39.12.37


    A P L \ 3 6 0
```

Once your sign-on has been acknowledged, you are ready to begin work. Most of the rest of the primer is devoted to explaining the kinds of calculations you can perform and programs you can write. The balance of this chapter deals with the mechanics of typing your instructions to the computer.

Whose Turn to Type?

You and the computer can't both type on the same typewriter at the same time. You have to take turns. <u>You</u> can type only when the keyboard is unlocked, whereas <u>the computer</u> can type only when the keyboard is locked. While it is the computer's turn to type, the keyboard remains locked, and you can't type anything.

When you complete the typing of an instruction, you have to let the computer know that you have finished. The carrier-return key serves to enter the instruction: that is, to signal the computer that you have finished typing, and that it should start interpreting and executing the instruction you have typed. When you hit the carrier-return key, three things happen:

1. The carrier returns to the left margin and the paper is moved up by one line.

2. The keyboard is locked.

3. The computer receives the signal that your message is complete.

Locking the keyboard serves two purposes: it keeps you from typing any more until the computer is ready for you, and it keeps the typewriter available for the computer's response to you. The computer never starts work on your instruction until it receives the signal that you have finished typing. Since that signal is the carrier return, each message you type must fit on a single line. But instructions in APL are so concise that you will rarely need as much as a whole line for an instruction.

As soon as the computer completes work on the instruction you typed, it does these three things:

1. Prints the result (if called for) and moves the paper up one line;

2. Indents by six spaces;

3. Unlocks the keyboard to await your next instruction.

Distinguishing Who Typed What

The paper in your typewriter will contain a complete record of your dialogue with the computer. When you read it, it is important to be able to tell who typed what. Because the computer makes the carrier space over by six spaces before the keyboard is unlocked, everything you type will ordinarily appear indented by six spaces, whereas what the computer types will ordinarily start at the left margin. (1050 terminals with two-color ribbon go a step further, by typing your part in red and the computer's part in black.)

Fixing Typing Errors as You Go

If _before you press the carrier return_ you notice a mistake in what you have typed, you have a chance to correct it before the computer starts to execute your instruction.

You can do that in the following way:

1. Backspace to the position of the leftmost character that is in error.

2. Press the ATTN key (this key is sometimes marked INDEX or LINEFEED).

When you do this, the computer types an inverted caret under the character in error, and spaces the paper up an additional line. Now that character and everything appearing to the right of it are considered "erased." You may resume typing the balance of your instruction. Suppose you type $A-B \times C$, and then you realize that the multiplication should have been a division. The "erasure" and correction would look like this:

```
A-B×C
   v
   ÷C
```

Overstrikes Not Allowed

Don't overstrike or X out any part of what you type. Except for certain APL characters which are always formed by overstriking, APL\360 cannot read overstruck characters. If you enter a statement which contains an illegal overstrike (i.e. if you type an illegal overstrike but don't erase it before you hit carrier return), the computer responds with an error message and a reproduction of your instruction up to the point that the illegal overstrike occurred. Like this:

```
    A-B÷⌿
CHARACTER ERROR
    A-B÷
       ∧
```

You will have to retype the line in which an overstrike occurs.

Visual Fidelity

While you are typing, you don't have to type each of the characters in order. For instance, you could leave extra spaces near the beginning of a line and then backspace over to that point and fill in the blanks. Your message is

interpreted by the computer the way it looks on the paper at the moment you press the carrier return. Within the line, the time sequence in which you hit the various keys doesn't matter. This principle can be summed up by the rule, "What you see goes in."

Interrupting the Computer

It may happen that you cause the computer to start typing a very long result, or working on a very lengthy (or even interminable) calculation. If you decide that you want to cut short what the computer is doing, there are two ways of stopping it. The easiest way is to use the "interrupt" feature supplied with some 2741 typewriters. If you have a 2741 thus equipped, pressing the ATTN key while your keyboard is locked will bring whatever the computer is doing for you to a halt.

If you have a 1050 terminal, or a 2741 that doesn't have the "interrupt" feature, you can produce the same effect as follows:

1. Uncradle the telephone.

2. Press the TALK button for a few seconds (you'll hear the high-pitched tone again).

3. Press the DATA button again.

If you have a 1050, that's all. If you have a 2741, the following additional step may be necessary:

4. If the carrier does not space over 6 spaces before unlocking, strike the carrier return repeatedly until it does.

3: BASIC OPERATIONS IN ARITHMETIC

The Workspace

As soon as your sign-on is completed, the computer puts at your disposal a block of its internal storage (or "memory"). This block of storage is called your workspace, since it is where all of your calculations take place. In it will be stored the definitions of programs that you enter, and the names and values of variables used in your calculations. Your workspace also includes locations used by the computer for the temporary storage of partial results while a calculation is in process, and specifications of a number of other items that affect the way your calculation is carried out or the way its result is printed. As you will see in Chapter 15, it is possible for a single user to have several different workspaces within the computer. However, only one of these is ever available for calculation at any one time. The one workspace which is currently available is called your active workspace.

The Two Modes of Operation: Execution vs. Definition

The computer has two modes of operation, called execution mode and definition mode. When the computer is in execution mode, it carries out any instruction immediately, as soon as you enter it. If you enter an arithmetic expression, the computer immediately responds with the result:

```
      12×13
156
```

Ordinarily, the computer is in execution mode; it is in execution mode when you first sign on, and it stays in execution mode unless you specifically direct it to switch to definition mode. When the computer is in definition mode, it does not execute the instruction that you enter, but stores it as part of the definition of a program. The instructions that make up the program are not executed until (at some later time, when you're back in execution mode) you call for execution of this program. How you enter the definition of a program is taken up in Chapter 6. The remainder of this chapter discusses the instructions you can use to get the computer to carry out some basic operations in arithmetic. These instructions could just as well be included as parts of a program, but the illustrations in this chapter show them being used in immediate execution.

Addition, Subtraction, Multiplication, and Division

These operations are familiar from everyday arithmetic. APL uses the familiar signs to indicate them:

+ - × ÷

The operation sign is typed between the numbers that are to be operated on, just as in arithmetic. For instance, if you want to multiply 1345.2 by 873.21, you simply type

      1345.2×873.21

The computer executes that instruction immediately, and replies with the answer:

1174642.092

Here are some more examples of simple instructions. Because the computer always indents by six spaces before unlocking the keyboard, the instructions you type always appear indented by six spaces, while the responses from the computer are typed starting at the left margin.

      176÷14.2
12.3943662

      17228.1-14.2
17213.9

      2+2
4

      5+0
5

      5
5

      4×1.25
5

      3÷32
0.09375

The Idea of a Variable:
Associating a Name with a Value, and Storing Them

You can store data, or the results of calculations, in your active workspace. A stored item of data is called a variable. Every variable has a name and a value; the computer associates the value with the name, and preserves that association in your active workspace. Whenever you refer to a variable by its name, the computer automatically supplies the value that has been associated with that name.

The symbol for assigning a value to a variable is the left-pointing arrow. If you enter the instruction

$SPEED \leftarrow 1088.5$

you cause the value 1088.5 to be associated with the name SPEED.

The left-pointing arrow causes the value of the expression to the right of the arrow to be stored under the name which appears immediately to the left of the arrow. This instruction may be read in several ways. You can read it as "SPEED is specified as 1088.5," or "SPEED is assigned the value 1088.5," or even "SPEED is 1088.5."

The variable SPEED is now stored in your active workspace. The computer doesn't type any specific acknowledgment that it has stored SPEED, but as soon as the variable's name and value have been stored in the workspace, the computer again indents and unlocks the keyboard.

A variable must always have both a name and a value; you can't create a variable which has a name but no value, and you can't store a value unless you assign it to a name.

Storing or Printing the Result of a Calculation

When you enter an instruction that calls for a calculation, as soon as the instruction is executed, the computer needs to know what to do with the result. There are three possibilities. All three are listed at the top of the next page, but the third one will not be discussed until the next chapter, where we take up compound expressions in a single instruction.

1. You can have the result printed. If you don't indicate that something else is to be done with the result of a calculation, the computer always assumes that you want to see it, and prints it.

2. You can have the result associated with a name, and stored in the workspace as a variable.

3. You can have the result of that operation used in another operation in the same instruction.

Using the Stored Value of a Variable

Once you have assigned a value to a variable, from then on whenever you refer to that variable's name, the computer supplies the associated value. If you simply type the name of a variable, the computer responds by printing its value:

    SPEED
1088.5

If you use the name of a variable in an instruction, the computer carries out the instruction, substituting the associated value wherever the name appears in the instruction. For instance, the value of SPEED is the speed of sound in air at 0 degrees Centigrade, expressed in feet per second. If you need to know how far sound travels in 15.5 seconds under those conditions, you can find out by the following instruction:

    15.5×*SPEED*
16871.75

Or, since multiplication is commutative (i.e. order doesn't matter), you could just as well enter:

    SPEED×15.5
16871.75

If you'd prefer to have that result stored, the following instruction assigns the result as the value of a variable called DISTANCE:

    DISTANCE←SPEED×15.5

And you could display the value of DISTANCE like this:

DISTANCE
16871.75

Possible Names for Variables

The name of a variable must begin with a letter of the
alphabet. After that it may have any combination of letters
or numerals, and any length. A name may not contain a space,
or any punctuation, or any of the symbols used for
operations. You may often find it helpful to select names
that have some mnemonic significance to you...but of course
the computer is unaffected by what names may or may not mean
in English. When naming a new variable, don't give it a name
that you want to keep in use for some other purpose in the
same workspace.

Storing a Result Under a Name
That Has Already Been Used

Suppose that at one point you type:

$X \leftarrow SPEED \times 8$

and then later on you type:

$X \leftarrow SPEED \div 8$

Each of these instructions calls for a, result to be stored
under the name X. What happens? The first time you use the
name X to the left of a specification arrow, a variable is
introduced, with the name X, and whatever value results when
the value of SPEED is multiplied by 8.

The next time you specify a value for X, that new value
replaces the former one. The value of SPEED is divided by 8,
and the result of that division becomes the value of X. The
old value is erased.

Clearly this would be the wrong way to write the
instructions if you really wanted to preserve both of those
results. To keep both, you must give them distinct names.
However, there are many situations in which it is convenient
to be able to replace one value of a variable by another
value stored under the same name. Suppose you want to count
how many times a task has been done. If, for example, you
have a variable called COUNT, you might have use for an

instruction which serves to update the counter, perhaps something like this:

COUNT←COUNT+1

Each time this instruction is executed, the computer adds 1 to whatever value it finds already associated with the name COUNT, and then stores the resulting value back under the name COUNT. (It should be noted that COUNT must have received its very first value in some other instruction: it can't <u>always</u> have been specified by referring to its own earlier value.)

Referring to a Variable Which Has No Value

You can assign a value to almost any name you like. But if you attempt to display or make use of the value of a variable before any value has been assigned to it, the computer is unable to supply an associated value, and can't proceed with the execution of your instruction. It reports the trouble by sending you an error message. For instance, suppose you have assigned a value to SPEED but not to INTERVAL, and then you enter an instruction which refers to INTERVAL. Your dialogue with the machine looks like this:

```
    INTERVAL×SPEED
VALUE ERROR
    INTERVAL×SPEED
    ∧
```

The first of those typed lines is your instruction. On the second line, the computer types its error message, indicating the kind of error it has found. On the third line the computer repeats the instruction as received. Then the computer types a caret under the point in the instruction at which it ran into trouble.

Examples of Arithmetic with Variables

The instruction

A×B

means that the operation of multiplication is to be performed on the value of A and the value of B. When the computer executes that instruction, it finds in the workspace the values of the variables A and B, and then

performs the operation, using those values. (The values associated with A and B in the workspace memory are not changed unless you specify that they should be.)

Suppose that A and B have been assigned the following values:

$A \leftarrow 6.25$
$B \leftarrow 144$

Then you can use those values in simple instructions, and the computer types results, like this:

$A + B$
150.25

$A + 1$
7.25

$B \div A$
23.04

$B - A$
137.75

$A \times B$
900

$900 \div B$
6.25

$Z \leftarrow 1 \div A$
$1 \div Z$
6.25

$A - A$
0

$B \div B$
1

)

This page intentionally left blank.

4: MORE ARITHMETIC OPERATIONS

Negation

If you place a minus sign in front of a number or variable, but nothing to the left of the minus sign, you get a result which has the same magnitude but opposite sign. For instance, if B has the value -17, then you get the negation of B like this:

```
        -B
17
```

Reciprocals

When you use a division sign in the same manner, it means that the reciprocal is to be found. If A still has the value 6.25, you can find the reciprocal of A like this:

```
        ÷A
0.16
```

Monadic and Dyadic Operators

Negation and the reciprocal are examples of __monadic__ operators. It is easy to distinguish them from __dyadic__ operators such as subtraction or division: the monadic operators have no value appearing to the left of them. That is, monadic operations such as negation and reciprocal each take only one argument, whereas subtraction and division take two arguments. The __arguments__ of an operator are the values it works on; an argument may be a variable, a number, or the value that results when an expression in parentheses is evaluated.

A dyadic operator is always written with the values on which it works (i.e. its arguments) on either side of it, as for instance in $A-B$. A monadic operator is always written with its argument to the right of the operator symbol, as in $-B$.

APL often uses the same symbol in two senses, one monadic and the other dyadic. You (and the computer) can always tell which sense is intended. If there is an argument immediately to the left of the operator sign, the operator is dyadic. Otherwise it is monadic.

Raising to a Power (Exponentiation)

In conventional arithmetic, exponentiation is indicated by writing the power to which a number is to be raised in a smaller typeface and placing it above the line. For instance, 2 raised to the 3rd power is written:

$$2^3$$

This is hard to type. Moreover, it seems odd that exponentiation has no symbol of its own, although addition, multiplication, division, etc., all have theirs. So APL uses a special symbol for exponentiation, placed between the number (or variable, expression, etc.) and the power to which it must be raised. The sign is $*$ and is located on the keyboard above the P (P for Power). For example:

```
      2*3
8
```

Here's an example of a calculation that uses exponentiation. It is based upon the familiar rules of compound interest. The names chosen for the variables should be self explanatory.

```
      PRINCIPAL←1045.28
      INTEREST←.03
      YEARS←17
      RATE←1+INTEREST
      MULTIPLIER←RATE*YEARS
      TOTAL←PRINCIPAL×MULTIPLIER
      TOTAL
1727.688573
```

This sequence of instructions estimates the total to which $1045.28 would grow if invested for 17 years at 3 per cent, compounded annually. The calculation could also be obtained in a single instruction, but that must wait until the next chapter.

Taking a Root

APL doesn't have any special sign for the extraction of a root. It doesn't need one. Taking the square root of a number is exactly the same thing as raising it to the one-half power. That's the way you write it in APL. If A has the value 144, you find the square root of A like this:

```
      A*0.5
12
```

Or you might get it this way:

        POWER←÷2
        A*POWER
12

    This procedure isn't confined to taking square roots. Any root can be extracted; for instance, you can find the fifth root of A by the following instruction:

        A*0.2
2.701920077

    The designers of musical instruments that are tuned to the "even tempered" scale (such as pianos) are faced with the problem of dividing an octave into 12 equal parts. The frequency of any note must be in a constant ratio to the note one semitone below it. Since it takes twelve semitones to make an octave, the ratio between one semitone and the next must be picked so that the product of all twelve of them will just make an octave. The semitone ratio is therefore the twelfth root of the octave ratio. Knowing that the octave ratio is exactly 2, you could find the size of an even-tempered semitone by the following two instructions. (Here again, this could also be done in a single instruction, as will be seen in the next chapter.)

        POWER←÷12
        2*POWER
1.059463094

Maximum: Taking the Larger of Two Numbers

    It is often convenient to be able to pick whichever is the larger of two numbers. APL includes an operation which does this. When the sign ⌈ is typed between two numbers (or variables that have numerical values) the computer selects whichever value is greater. If you type

        A⌈B

the computer examines what has been stored under those names. Then it takes whichever value is greater. (Recall that the values associated with A and B in the workspace remain unchanged.)

    Suppose that earlier calculations resulted in the following values for the variables ABC and XYZ:

ABC has the value 12345679, and
XYZ has the value 12345678

Then your dialogue with the computer might look like this:

ABC⌈XYZ
12345679

Consider an illustration in which this operation might be useful. Suppose you work for a department store. Each month, the store calculates for each of its customers how much he charged and how much he paid that month. You have a program which handles the billing. You calculate for each customer the value of a variable you call BALDUE, which is the difference between the total of the accumulated charges and the total of the accumulated payments for that customer.

The store charges each customer a service charge of 1.5% of the unpaid balance each month. You might find this charge by the following instruction:

CHARGE←BALDUE×.015

However, for one reason or another, some of the customers have overpaid their bills. For them, BALDUE is a negative number, and shows as a credit on their monthly statements. If you calculate the service charge by the instruction just shown, you'll be paying them interest at 1.5% per month whenever they overpay. Instead, the store prefers to calculate the service charge as 1.5% either of the balance due or of zero, whichever is greater. You can do this by using the following instructions:

CHARGEABLE←0⌈BALDUE
CHARGE←CHARGEABLE×.015

Minimum: Taking the Smaller of Two Numbers

In similar fashion, another primitive APL operator selects whichever is the smaller of the two values on either side of it. If ABC and XYZ have the same values as before, the lesser is selected by this instruction:

ABC⌊XYZ
12345678

The annual amount a wage-earner pays for FICA (social security) tax is based upon how much he earns. However, any income he has beyond $7800 a year doesn't count for social security purposes. The FICA tax rate is currently 4.4%. If a man's yearly gross income is called YEARGROSS, and has a value of $8320, then his annual FICA tax might be found this way:

```
    YEARGROSS←8320
    FICAINCOME←7800⌊YEARGROSS
    .044×FICAINCOME
343.2
```

The Floor and the Ceiling of a Number

You can disregard the fractional portion of a number and just consider the integer portion. You have a choice of two ways of doing this: by rounding down to the next smaller integer than the fraction, or by rounding up to the next larger integer. The operators which do this are called the floor and the ceiling. If A has the value 3.14159, then you get the floor of A as follows:

```
    ⌊A
3
```

and the ceiling of A like this:

```
    ⌈A
4
```

You will notice that <u>ceiling</u> is the meaning of the ⌈ symbol when it is used monadically; when it is used dyadically (i.e. with a value on either side of it) it means <u>maximum</u>. In the same way, ⌊ means <u>floor</u> when it is used monadically, but <u>minimum</u> when it is used dyadically.

```
    ⌈XYZ
```

means the ceiling of XYZ. If XYZ is already an integer, then the ceiling of XYZ has the same value as XYZ. But if XYZ has a fractional part, the ceiling is the next (algebraically) larger integer than XYZ.

```
    ⌊XYZ
```

means the floor of XYZ. If XYZ is already an integer, its floor has the same value. But if XYZ has a fractional part,

the floor of XYZ is the next (algebraically) smaller integer. In the case where XYZ has the value 3:

        ⌈XYZ
3
        ⌊XYZ
3

Rounding to the Nearest Integer

It is common practice to round numbers to the nearest integer. This means that when the fractional part is less than .5, the number is rounded down, but if the fraction is .5 or greater, the number is rounded up. This effect is produced if you first add .5 and then take the floor. Suppose A has the value 3.14159, and B has the value 3.5:

        X←.5+A
        ⌊X
3
        X←.5+B
        ⌊X
4

Summary of Arithmetic Operators Mentioned Thus Far

    $A+B$ means A plus B

    $A-B$ means A minus B
     $-B$ means the negation of B (i.e. 0 minus B)

    $A×B$ means A times B

    $A÷B$ means A divided by B
     $÷B$ means the reciprocal of B (i.e. 1 divided by B)

    $A⌈B$ means the maximum of A and B
     $⌈B$ means the ceiling of B

    $A⌊B$ means the minimum of A and B
     $⌊B$ means the floor of B

    $A*B$ means A raised to the Bth power.

Note: in conventional algebra, the expression ab means the product a×b. The multiplication sign is elided. But in APL, the multiplication sign must be explicitly entered wherever you want multiplication to occur.

5: SEVERAL OPERATIONS IN THE SAME INSTRUCTION

The preceding examples of APL instructions were all written so that only one arithmetic operator occurred on each line. APL\360 does not, of course, restrict you to writing only instructions with but a single operator. The examples were written that way so as to postpone for a moment the discussion of some issues that arise when there are several operations in the same instruction.

APL permits you to write any number of operators in the same instruction. But as soon as you write more than one operator, you have to be clear about the order in which they get executed. It makes a difference.

Conventional arithmetic has a number of rules for this. First of all, there is a hierarchy of operators. Some operators are always executed ahead of others, regardless of their position in the instruction. Exponentiation gets highest priority. Then multiplication and division are performed. Addition and subtraction are done last. There is also a rule to apply if the instruction contains several instances of the same operator. If those rules aren't appropriate to indicate the order in which you want to execute several operations, you can use parentheses to show that what falls within them gets priority.

APL employs not only the operators + - × ÷ but a great many others as well. You have made the acquaintance of * and ⌈ . There are several more. Moreover, in more advanced uses of APL, programs that you write yourself can be made to act like operators. You can see that attempting to have a hierarchical rule to determine which of all those operations should get done first could get very complicated.

Order in Which Operations Get Executed in APL

APL solves this problem by abolishing the hierarchy of operators altogether. Instead, the order of execution depends upon only two things:

1. Parentheses (in the usual way);

2. The order in which the operators appear in the instruction.

This is the rule: <u>An operator operates on everything to the right of it</u>. Stating it more formally, any operator takes as its right argument everything to the right of it.

This order is the same as the usual order of English speech. You may not have thought of English speech in that way, but it will be obvious in a brief example. Take the following English sentence:

"They oppose a rise in the price of products from farms."

To examine the meaning of this sentence, let's pose some questions about what applies to what.

Oppose what?   <u>A rise in the price of products from farms</u>.

What rise?   One in <u>the price of products from farms</u>.

Rise in what price?   The one <u>of products from farms.</u>

Price of what products?   The ones <u>from farms</u>.

Here each of the key elements in the sentence refers to all the rest of the sentence after it.

The same structure occurs in APL. Consider this example:

$$A+B\times C-D\div E$$

What is A added to? $B\times C-D\div E$

What is B multiplied by? $C-D\div E$

What is subtracted from C? $D\div E$

What is D divided by? E

Suppose you have to explain that English sentence to someone who knows about English grammar but doesn't know what the particular nouns in that sentence apply to. You'll find that your answer to each of the questions still involves everything that comes later in the sentence. If the questions are asked in the order we just used (starting at the left), the answer to the first one isn't immediately usable because the answer still refers to the rest of the sentence, and that hasn't been defined yet. To build up the

explanation in a logical way, you have to start with the last word, "farms." Then, using that, you can work back to "products," and that will permit you to get the meaning of "products from farms." And so on, working from right to left until the meaning of the whole sentence is established.

Similarly, when the computer executes an instruction such as

$$A + B \times C - D \div E$$

it starts by looking for the value associated with the name E. Then it looks up the value of D. That gives it enough information to evaluate D divided by E. Next it looks up the value of C, which gives it enough information to evaluate the difference between C and D-divided-by-E.

Thus although an APL instruction reads aloud easily from left to right, when the computer comes to execute the instruction, it executes the various operations in right-to-left order. This has become known as the "right to left rule." Notice that this doesn't mean that the computer reads the line backwards, only that when it executes the various operations within an instruction it does the rightmost one first, then the next rightmost one, and so on.

Use of Parentheses

In APL, you use parentheses in the usual way. That is, the operations inside the parenthesis are to be executed before operations outside. When the expression inside a parenthesis has been evaluated, the result must always be a value: you can't just put an operator symbol alone inside the parenthesis.

Consider the following expression in conventional arithmetic:

$$(a+b) \times (c+d)$$

The hierarchy of operators in conventional arithmetic would ordinarily cause the multiplication to be executed before the addition. But this order is overridden by the parentheses, which cause the two additions to be done first. In APL, you could write the instruction in exactly the same way:

$(A+B)\times(C+D)$

or you could also write it like this:

$(A+B)\times C+D$

Since in APL the rightmost operation is executed first anyway, you don't really need to use the right pair of parentheses--although it's all right to include them if you want to. If you enter the following instruction:

$(A+B)\times C+D$

here's how the computer will proceed. The rightmost operation is the addition of C and D, so it does that first. Moving leftwards, it finds that the next operation is the multiplication. But the left argument of the multiplication is a parenthesis. The computer suspends work on the multiplication until it has evaluated the expression in the parenthesis. When that is done, it comes back to the multiplication, and multiplies the sum-of-A-and-B by the sum-of-C-and-D.

Because of the order of execution of operations in APL, many instructions that would otherwise require parentheses no longer need them. When you enter an APL instruction, you can often arrange it so that the operators that you want executed first appear furthest to the right in your instruction. For instance, consider the instruction that in conventional arithmetic would have to be written as either

$(a+b)\times c \quad$ or $\quad c\times(a+b)$

In APL, that can be written without any parentheses:

$C\times A+B$

Where rearrangement doesn't eliminate the need for parentheses, you can still use them in the usual way. Parentheses within parentheses (sometimes called "nested" parentheses) are all right too. The computer starts first on the outermost parenthesis. If it finds another parenthesis inside the first one, it suspends work on the expression in the outer parenthesis until it has evaluated the inner one.

Rewriting the Earlier Examples
With Several Operations in the Same Instruction

The examples presented earlier in the text involved
only one arithmetic operation on each line. Now that we've
dealt with the question of order of execution when there are
several operations in the same instruction, we can rewrite
those earlier examples more neatly by using compound
expressions.

For instance, the total when an amount is invested at
compound interest can be stated as follows, (assuming that
the variables have the same values as before):

        $PRINCIPAL×(1+INTEREST)*YEARS$
1727.688573

The square root of XYZ can be written like this, using the
reciprocal sign, since the second root of XYZ is equivalent
to raising XYZ to the reciprocal-of-two power:

        $XYZ←12$
        $XYZ*÷2$
3.464101615

and in similar fashion the semitone ratio for an
even-tempered scale can be found by

        $2*÷12$
1.059463094

To prove that it really is the twelfth root, you could raise
the semitone ratio to the twelfth power. The result should
be the octave ratio, which is 2.

        $SEMITONE←2*÷12$
        $SEMITONE*12$
2

The charge on the balance due on a charge account becomes:

        $CHARGE←.015×0⌈BALDUE$

and the annual FICA tax becomes:

        $TAX←.044×7800⌊YEARGROSS$

Do You HAVE to Write Many Operations
In a Single APL Instruction?

As you gain experience in the use of APL, you will probably tend to use longer compound expressions in the instructions that you write. For one thing, it is often easier to understand a well-formed compound instruction than it is to trace through a sequence of simpler steps. Compare, for instance, the expression for the total resulting from compound interest used on the preceding page:

*TOTAL←PRINCIPAL×(1+INTEREST)*YEARS*

with the one-step-at-a-time sequence we used on page 22:

RATE←1+INTEREST

*MULTIPLIER←RATE*YEARS*

TOTAL←PRINCIPAL×MULTIPLIER

Nevertheless, it should be clear that whether you use a few long instructions or many short ones is up to you; you should write in the style that seems easiest to you.

6: ENTERING THE DEFINITION OF A PROGRAM
SO THAT IT CAN BE USED REPEATEDLY

If the work that you want done can be specified in an instruction that is brief and easy to type, you can get it done simply by entering that instruction. But if you want a more complex calculation, or one that you want to use repeatedly, you will certainly want to define a program to do the job. Then you can obtain execution of the program, regardless of the number of instructions it contains, simply by typing its name. This section tells you how to define a program.

Starting the Definition

You will recall that the computer has two modes, one for executing instructions immediately, and the other for storing the definition of a program. To start the definition, you must enter definition mode. The symbol ∇ (pronounced "del") takes you from one mode to the other. If you type a ∇ while you are in execution mode, it signals the computer that what follows is the definition of a program. If you type another ∇ while the computer is in definition mode, that signals the computer that you are finished with the definition of that program; the computer returns to execution mode.

To start the definition of a program, the first thing you do is type (on a single line) the symbol ∇ followed by the name of the program.*

When you press carrier return, the computer asks what you want as the first line of the definition. It does this by typing, in square brackets, the number 1. Then the

---

*At this point we are limiting the discussion to the simplest type of program, what you might call a "stand alone" program. A program of this type is executed simply by typing its name. That name, standing alone, is all that may appear in the instruction that causes the program to be executed. APL also permits you to define a program so that it may be used as part of a compound instruction, with other programs, operations, and variables all in the same instruction. Discussion of that type of program is deferred until Chapter 25.

computer spaces over until it has completed the usual
indentation of six spaces, and unlocks the keyboard to await
your definition of line 1. What you type at that point is
entered as the definition for line 1 of this program. Then
the computer types a 2 in square brackets, and awaits the
definition of line 2. It continues in this fashion until you
type another ∇ to indicate that you want to return to
execution mode.

This will be clear if we consider a simple program and
work through its definition step by step.

In order to. show where the typeball is when the
keyboard is unlocked, we use the following mark:

@

This mark does not, of course, appear on your paper; we
merely use it in this primer to show you where the typeball
is located at the moment when it becomes your turn to type.

Focal Length of a Lens: A Simple Calculation
To Illustrate Program Definition

Here is the formula for the focal length of a lens:

$$f = \frac{nr_1 r_2}{(n-1)[(n(r_1 + r_2) - t(n-1)]}$$

where   f   is the focal length
         n   is the index of refraction
         t   is the thickness of the lens
         r_1 and r_2 are the two radii of curvature.

Suppose that you want a program called FOCAL to compute
the focal length, F, from the stored values of variables
called N, T, R1, and R2. The program should both store F and
print F.

(Notice that it's all right for a name to include
numerals, as long as they aren't the first character in the
name, so that the names R1 and R2 are permissible. However,
they are individual names which do not mean that these are
the first and second members of a variable called R.
Indexing of a variable is introduced in Chapter 19.)

   Your first step in defining FOCAL is to type a del,
followed by the name of the program. When you do that, the
computer notes the name of the program, and asks what you
want as line 1 of the definition. Your paper looks like
this:

       ∇FOCAL
[1]   @

   You could start the calculation by finding the value of
the numerator of the fraction, and storing it. When you type
line 1, which might be as follows, the computer responds by
asking what you want on line 2:

       ∇FOCAL
[1]   NUM←N×R1×R2
[2]   @

   On line 2, you can calculate the denominator of the
fraction:

       ∇FOCAL
[1]   NUM←N×R1×R2
[2]   DENOM←(N-1)×(N×R1+R2)-T×N-1
[3]   @

   On line 3, you do the division and store the result
under the name F:

       ∇FOCAL
[1]   NUM←N×R1×R2
[2]   DENOM←(N-1)×(N×R1+R2)-T×N-1
[3]   F←NUM÷DENOM
[4]   @

   On line 4, you want to have F printed. You simply type
its name:

       ∇FOCAL
[1]   NUM←N×R1×R2
[2]   DENOM←(N-1)×(N×R1+R2)-T×N-1
[3]   F←NUM÷DENOM
[4]   F
[5]   @

   That's all the program needs.

When the computer asks for a definition of line 5, you type another del. The computer closes the definition of the program called FOCAL, stores it in the memory of the active workspace, and returns to execution mode.

```
        ∇FOCAL
[1]     NUM ← N×R1×R2
[2]     DENOM ← (N-1) × (N×R1+R2) - T×N-1
[3]     F ← NUM÷DENOM
[4]     F
[5]     ∇
        @
```

After the final del, the computer leaves definition mode. Therefore, as you can see in the example shown above, when it unlocks the keyboard for your next instruction, it again indents by six spaces, but this time without typing a line number (since now you're back in execution mode).

Sample Use of the Program Just Defined

The values of the variables N, T, R1 and R2 need not have been stored at the time you entered the definition of FOCAL, but they should be stored before you try to execute FOCAL. Once those values are in storage, you cause the computer to execute FOCAL simply by typing its name. The computer prints the value of F, as line 4 of the program directs.

```
        N← 1.3275
        T←  .375
        R1←8
        R2←7.85
        FOCAL
12.16918274
```

If you wish, you can set new values for the radii and then ask for a new execution of FOCAL. If you can still use the former values of N and T, you need not enter their values again, since they are retained in the workspace.

```
        R1 ← 8.1
        R2 ← 7.75
        FOCAL
12.16433832
```

Another Sample Program:
Efficiency of a Diesel Engine

One form of the equation for the theoretical efficiency of a Diesel engine is as follows:

$$EFF = 1 - \frac{1}{R^{\gamma-1}} \left[\frac{\left(\frac{V_3}{V_2}\right)^{\gamma} - 1}{\gamma \left(\frac{V_3}{V_2} - 1\right)} \right]$$

Using this formula, you would like to see how the theoretical efficiency varies over some range of the various parameters. You need a program that will compute EFF, the efficiency, from the stored values of those parameters. In your workspace, you can give them names based upon their representations in the formula; for instance, they might be R, GAMMA, V3, and V2.

There are various strategies for writing this program. To simplify your task, you might want to divide up the calculation into sections, and compute each section separately. Suppose you start by breaking the formula into sections A, B, and C, as follows:

$$EFF = 1 - \underbrace{\frac{1}{R^{\gamma-1}}}_{A} \left[\left.\frac{\left(\frac{V_3}{V_2}\right)^{\gamma} - 1}{\gamma \left(\frac{V_3}{V_2} - 1\right)}\right\}\!\!-\!\!\! B \atop \right\}\!\!-\!\!\! C \right]$$

Once you have calculated values for A, B, and C, you can get the efficiency by this instruction:

$EFF \leftarrow 1 - A \times B \div C$

This will probably be the last, or next-to-last, instruction of the program. Ahead of it you need instructions that will calculate A, B, and C. Part A is easily obtained with the following instruction:

$A \leftarrow \div R \star GAMMA - 1$

Notice that $V3 \div V2$ occurs twice in the formula, once in section B and once in section C. If you save the result the first time you do the division, you won't have to do the division twice. Hence before you evaluate B and C, you may

want to divide V3 by V2 and store the result; suppose you call it RATIO. Then B can be calculated by this instruction:

$B \leftarrow (RATIO*GAMMA)-1$

and C can be obtained by this instruction:

$C \leftarrow GAMMA \times RATIO-1$

Now the various steps can be put together into a program. Suppose the program has the name DIESEL. This version does not include a print instruction, although of course one could be added as line 6.

```
      ∇DIESEL
[1]   A←  ÷R*GAMMA-1
[2]   RATIO← V3÷V2
[3]   B← (RATIO*GAMMA) -1
[4]   C← GAMMA×RATIO -1
[5]   EFF← 1-A×B÷C
[6]   ∇
```

Writing the DIESEL Program in a Single Line

There are many ways this little program could be written. If you preferred to write an equivalent program which puts it all into one instruction, you could do it like this: Imagine the formula split into sections A, B, and C, as before, but this time instead of storing values for each of those variables, substitute the expression for A, B, and C directly in the first line. Let this short program be called S:

$$EFF \leftarrow 1 - \quad A \quad \times \quad B \quad \div \quad C$$

$\div R*GAMMA-1 \qquad ((V3÷V2)*GAMMA)-1 \qquad GAMMA×(V3÷V2)-1$

The one-line definition of S is therefore as follows:

```
      ∇ S
[1]   EFF←1-(÷R*GAMMA-1)×(((V3÷V2)*GAMMA)-1)÷GAMMA×(V3÷V2)-1
[2]   ∇
```

Sample Execution of DIESEL

```
      GAMMA←1.485
      V3←184
      V2←22
      R←15
      DIESEL
      EFF
0.4484434684
```

An Instruction in One Program
Can Call for the Execution of Another Program

The instructions in DIESEL might be rewritten so that each portion of the calculation is handled by another program. Since the diesel calculation was divided into parts A, B, and C, each of those might be calculated by a separate program; you might want to call them DOA, DOB, and DOC. Here's a program called DSL, written in that way:

```
      ∇ DSL
[1]     DOA
[2]     DOB
[3]     DOC
[4]     EFF←1-A×B÷C
      ∇
```

Of course, you can't tell what this definition means until definitions are supplied for the programs DOA, DOB, and DOC. Here they are:

```
      ∇ DOA
[1]     A←÷R*GAMMA-1
      ∇
```

```
      ∇ DOB
[1]     B←((V3÷V2)*GAMMA)-1
      ∇
```

```
      ∇ DOC
[1]     C←GAMMA×(V3÷V2)-1
      ∇
```

As far as you can tell when you use it, DSL works just like DIESEL. In a larger problem than this one, it is often convenient to be able to break the work up into modules

which are handled by separate sub-programs. As you will see
later, the ability of one program to call for the execution
of other programs becomes much more useful when those
programs can be written so that they have arguments in the
way that APL operators do. Then you can write compound
expressions involving calls to other programs. That topic
is discussed in Chapter 25.

It is sometimes convenient to define a program in which
the opening instructions set up the values that another
program is to use. A later instruction in the same program
may then call for the execution of the program that does the
actual calculation. Here is a definition for a program
called D, which sets new values for GAMMA, R, V3, and V2 by
modifying the earlier values as shown, and then calls for an
execution of DIESEL and for the printing of the final value
of EFF.

```
      ∇  D
[1]     GAMMA←1.01×GAMMA
[2]     V3←0.99×V3
[3]     V2←0.95×V2
[4]     R←R+0.04
[5]     DIESEL
[6]     EFF
      ∇

      D
0.4489094412

      D
0.4488170822
```

7: DISPLAYING OR CHANGING THE PROGRAM
AFTER YOU'VE DEFINED IT

Suppose you've defined a program DIESEL. You have typed all of your definition, and you've typed a final ∇ to indicate that the definition is ended. That has taken you back to execution mode. Perhaps you have even executed the program a few times. Now you decide that you want to change the definition. Perhaps you find a mistake in it, or some unnecessary lines; perhaps you wish to add some additional steps that you didn't think of before. How may you edit the stored definition?

Any time you edit the definition of a program (including just displaying it without changing it) you start out by typing a ∇ and the name of the program. For the "stand alone" type of program (the only kind introduced thus far) this is the same as the way you first started the definition of the program.

Adding Another Line

Whenever you enter definition mode and type the name of a program, the first thing the computer does is check the active workspace to see if there is already a definition for a program of that name. The first time you entered ∇DIESEL, the computer could find no prior definition for a program called DIESEL in the workspace. So it presumed you were starting a new definition, and asked what you wanted on line 1. When you first opened the definition of DIESEL, the opening dialogue went like this:

```
     ∇DIESEL
[1]  @
```

But when the computer finds that a definition of DIESEL has already been stored, it assumes that now you want to add to the stored definition. So it types the number of the line which comes next after the lines it already has, and awaits your definition for that new line.

```
     ∇DIESEL
[6]  @
```

After it gets the definition of that line, it asks you
for line 7, and so on until you once more enter a ▽ to
indicate that the definition is closed. You may recall that
the definition of DIESEL did not include an instruction to
print the value of EFF. Suppose you now add such an
instruction after the instructions that have already been
entered:

```
        ▽DIESEL
[6]     EFF
[7]     ▽
        @
```

Replacing a Line

Suppose that you don't like the definition that you
originally entered for line 3 of DIESEL; you want to replace
it with something else. If you once again enter definition
mode, since the computer now has six lines of definition for
DIESEL, it invites you to enter a definition for line 7. You
may override this suggested line number by typing a new line
number, in brackets as before. If you wish to redefine line
3, you now type [3] followed by whatever you would now like
to have on line 3. The new version of line 3 replaces the
old one.

```
        ▽DIESEL
[7]     [3] B←-1-V*GAMMA
[4]     ▽
```

Whenever you specify a new definition for a line which
already exists, your current definition replaces the earlier
one. After accepting your new definition for line 3, the
computer asks if you wish to revise line 4 also. If you
don't want to, you now type a ▽. Line 4, and all other lines
previously defined, remain unchanged.

Displaying What Is Already on a Line

Suppose you want to check up on what you wrote on a
line of your program. You want to see what was on line 3 of
DIESEL in order to decide whether to change it, or how. You
do this using the input-output symbol ▢, called "quad." Once
you are in definition mode, you type within brackets the
line number followed by a ▢. For example, to cause line 3 of
DIESEL to be displayed, at that point you enter [3▢]. This

is shown step by step in the following example. Notice that after it has shown you what is on line 3, the computer invites you to redefine line 3.

```
        ∇DIESEL
[7]     @
```

Step 1: Enter definition mode for the program called DIESEL. The computer already has six lines defined, so it asks what you want on line 7.

```
        ∇DIESEL
[7]     [3□]
[3]     B←(RATIO*GAMMA)-1
[3]     @
```

Step 2: Instead, you ask for a display of line 3. The computer types its stored definition of line 3, and then asks what you want as a new definition for line 3.

Step 3: Either--

```
        ∇DIESEL
[7]     [3□]
[3]     B←(RATIO*GAMMA)-1
[3]     B←-1-RATIO*GAMMA
[4]     @
```

(a) If you want to change line 3, type the new instruction for line 3. Then the computer asks what you want on line 4.

or--

```
        ∇DIESEL
[7]     [3□]
[3]     B←(RATIO*GAMMA)-1
[3]     [5□]
[5]     EFF←1-A×B÷C
[5]     @
```

(b) If you don't want to change line 3, but you now want to display some other line, type in brackets the number of the line you want to see next, followed by a □. The computer then shows you that line, and asks what you want as the new definition of that line.

or--

```
        ∇DIESEL
[7]     [3□]
[3]     B←(RATIO*GAMMA)-1
[3]     ∇
        @
```

(c) If you want to leave line 3 as it was, and leave definition mode, type a ∇. As always, all previously defined lines remain unchanged.

Displaying the Whole Stored Definition

Once you have the computer in definition mode, if you
use the ☐ symbol to get a line displayed but you don't say
which line you want, you get all of them. For instance,
entering

        ∇DIESEL
[7]   [☐]

causes the computer to print its entire stored definition of
the program DIESEL:

        ∇DIESEL
[7]   [☐]
    ∇ DIESEL
[1]   A←÷R*GAMMA-1
[2]   RATIO←V3÷V2
[3]   B←(RATIO*GAMMA)-1
[4]   C←GAMMA×RATIO-1
[5]   EFF←1-A×B÷C
[6]   EFF
    ∇
[7]   @

Notice that the computer even prints the initial ∇ with
which the definition starts, and another one to show where
the definition thus far stored comes to an end. These dels
that the computer types do not change the mode: only a del
that you type can do that. The first del you typed started
the definition mode; when you are ready, you will have to
type another del to get back to execution mode.

Notice too that after it has finished typing the entire
stored definition, the computer types a new line number,
inviting you to enter the definition of another line after
those already defined. As before, you don't have to enter
one if you don't want to.

Inserting a Line Between Lines that are Already Defined

Suppose that the line that you want to add doesn't come
at the end of the program. Perhaps you forgot to set up
something at the beginning of the program, or perhaps you

forgot an intermediate step somewhere in the middle. How can you insert a line between the existing lines of the program?

You interpolate a line by giving it an interpolated line number. Suppose you wish to insert a line so that it comes after line 1 but before line 2. You do that by assigning your line a decimal number between 1 and 2; 1.1 would do, or 1.5, or any other number with up to four decimal places and which is greater than 1 but less than 2. Negative line numbers aren't allowed, so if you want to insert a line ahead of the first line, assign it a line number between 0 and 1.

When you type ∇ followed by the name of the program, the computer, as before, asks what you want to add after the last line it now has in the definition. You decline this invitation; instead, you type a new line number, also in brackets. This new line number overrules the one typed by the computer. Suppose the program DSL now has 5 lines; you wish to insert a line saying

RATIO←V3÷V2

between lines 1 and 2. Here's what happens:

```
        ∇DSL
[6]     [1.5]RATIO←V3÷V2
[1.6]   @
```

As usual, after you enter your definition of that line, the computer responds by asking what you want as the definition of the next line. What is the "next" line in this situation? The computer determines the number of the "next" line by adding a 1 in the rightmost place of whatever number was typed. Since you typed [1.5], the machine asks next for line [1.6].

If you had given the line the number [2.0089], then the computer would have asked next for a definition of line [2.009]. Of course, you wouldn't have to give it one. You can always close the definition, or you can type any other line number you may want to insert next.

When you close the definition, the lines are all renumbered, and given line numbers that are consecutive integers (1, 2, 3, 4, ... etc). If you insert a single line between lines 1 and 2, that inserted line becomes line 2. The old line 2 becomes line 3, the old line 3 becomes line

4, and so on. If you now display the entire definition of DSL, you find that the inserted line has pushed down the lines that follow it:

```
       ∇DSL
[7]    [□]
    ∇ DSL
[1]    DOA
[2]    RATIO←V3÷V2
[3]    DOB
[4]    DOC
[5]    EFF←1-A×B÷C
[6]    EFF
    ∇
[7]    @
```

Deleting a Line of the Definition

Suppose that you decide that you don't want line 2 of program DSL in there after all. How can you remove it?

You can delete a line of the stored definition of a program by using the "erase" feature. You start out as if you were going to replace the definition of a line (see page 42). But when it comes time to type the new definition for that line, you simply press the ATTN key, followed immediately by carrier return. This combination erases that line from the stored definition. Then the computer asks what you want to do about the next line of the program. Erasure of line 2 of program DSL looks like this:

```
       ∇DSL
[7]    [2]
          ∨

[3]    @
```

When you type the final ∇ to leave definition mode, lines of the program which have nothing on them are dropped, and the other lines are moved up to fill the gaps. For instance, if you erase line 2, the old line 3 is moved up to become line 2, the old line 4 becomes lines 3, and so on.

Changing the Program's Header

The header of a program is the title line, containing the program's name. As you will see in Chapter 25, it may also contain several other items as well. It is quite possible to edit not only the instructions that make up the body of a program's definition, but also the header itself. In this way you can change the name of the program, if you wish. When you get to defined functions with results, arguments, or local variables (taken up in Chapter 25) you may also add, delete, or rename any of those.

The header of a program may be edited by asking to alter its line 0, and then proceding just as you would to change any other line of the program. Suppose you wish to rename the function DSL with the name DSL2:

```
        ∇DSL
[7]     [0□]
[0]     DSL
[0]     DSL2
[1]     ∇
```

Now you have created a program called DSL2, having the same definition that DSL used to have, and (indirectly) you've also erased DSL.

Erasing Programs or Variables Entirely from your Workspace

Suppose you are through with some programs or variables altogether. You may keep them in your workspace indefinitely if you wish. But if you no longer want them cluttering up your workspace, you may delete it using the system command)ERASE followed by the names of each of the programs (or variables) you want to erase. Suppose you want to erase the program called DIESEL and the variable GAMMA:

```
    )ERASE DIESEL GAMMA
```

Both the name and the entire definition of DIESEL are erased from the workspace. Similarly, both the name and the value of GAMMA are erased. If you try to use either of them, the computer no longer recognizes them:

```
    DIESEL
VALUE ERROR
    DIESEL
    ∧
```

You can't erase a program definition while it is pendent (i.e. while its execution is halted to await the result of another program). See page 85-86, where halted programs are discussed, and the discussion of definition errors in Appendix D.

8 : REPRESENTING NUMBERS

When you wish to enter a number into the computer, or when the computer prints the numerical value of a result, you have to have a system for representing numbers. You want the computer to understand what you type, and you need to understand what it reports. Internally, the computer represents numbers in the binary system, but with APL\360 you don't have to deal with the internal workings of the computer. Whenever you and the computer communicate, numbers are represented in the decimal system.

Within the decimal system, APL\360 uses either of two different forms. When you wish to enter a number, you may use whichever form is convenient for you. You may mix both forms in the same expression, if you like. The choice of one or the other form is purely a matter of convenience: it makes no difference to the computer's calculations. Similarly, each time the machine has to print a numerical value, it picks one or the other form in which to type it. In general, the computer picks whichever form yields the simplest representation. This choice of form is not made until the computer is ready to print, after its calculation has been completed.

Decimal Form

You can enter any number in the usual decimal form, using the period as a decimal point. If the number doesn't have any digits to the right of the decimal point, you don't need to type the decimal point either; to APL, it doesn't matter whether you enter 6, or 6.0, or even 06.00. Leading zeroes to the left of the decimal point, or trailing zeroes to the right of the decimal point, don't matter. However, the digits that represent a single number must not be separated by spaces or commas.

In the following examples, A, B, C, D, and E are given values that are entered in the standard decimal form:

$A \leftarrow$ 0

$B \leftarrow$ 1088.5

$C \leftarrow$ .00065

$D \leftarrow$ 186300000

$E \leftarrow$ 0.3

Exponential Form

When your work involves numbers that are very large or very small, it is often desirable to indicate these numbers by stating a value in some convenient range, and then multiplying it by the appropriate power of ten. For instance, Avogadro's number, which is the number of molecules in x grams of a substance whose molecular weight is x, is commonly written as 6.02×10^{23}.

A similar form exists in APL. It is called exponential form. In exponential form, Avogadro's number is written

     6.02E23

The E in the middle indicates that this is exponential form; the digits to the right of the E indicate the number of places that the decimal point must be shifted.

     6.02E23

means the same as

     602000000000000000000000.0 (point shifted 23 places).

That is, the digits to the right of the E indicate the power of 10 by which the number to the left of the E must be multiplied.

The estimated population of the world in 1964 could be written in any of the following ways; each results in the same value of the variable POP64:

     POP64←3.22E9

     POP64←3220E6

     POP64←3220000000

     POP64←3220000000.00

It is important to note that the letter E in a number such as 3.22E9 is a part of the name of that number, and not an operator. By contrast, when you enter 3.22×10*9 you are instructing the computer to perform a sequence of operations which, as it happens, will end up with the same value.

Negative Numbers

A negative number is indicated by the symbol that means "negative" placed in front of it. Negative two is written like this:

$^-2$

Note that the negative symbol is <u>not</u> the same as a minus sign. The minus sign denotes <u>the</u> operation of subtraction. The negative symbol is part of the name of all those numbers that lie below zero on the number line. Unfortunately, the distinction between the operation of subtraction and the names for numbers that are below zero has been muddled by the common practice of calling a negative number (for instance) "minus two," and using the minus sign for both purposes. APL avoids this confusion by using the minus sign only to mean the operations of either subtraction or negation, and the negative symbol only as part of the name of a negative number.

Notice that the operation - (subtraction), like all other APL operators, applies to everything to the right of it in an instruction. For instance, the instruction

7 - 2 + 3

means that the sum of 2 and 3 is to be subtracted from 7. By contrast, the negative symbol $^-$ is simply part of the representation of a single number. It doesn't apply to any other number but the one in which it occurs. Because it is not an operation at all, it can never be used alone, and it can never be used to operate on a variable. In this respect, the negative symbol is like the decimal point, or the exponential E: it has no meaning other than to help determine the value of the number represented by a particular cluster of digits. The decimal point, the negative sign, and the exponential E, may occur only as part of the representation of a number. You can't have any spaces separating these symbols from the other digits of the same number.

Negative Numbers in Exponential Form

The negative symbol can turn up in exponential form in just the same way as in other numbers. For instance, you can indicate the number negative two trillion by typing:

$^-2E12$

And you get negative 2.11684 trillion by entering:

    ‾2.11684E12

Very Small Numbers

In the exponential form, you can represent a very small number in the same fashion as a very large one. For large numbers, the decimal point is to be shifted to the right, so that 2E3 means 2000. For very small numbers the decimal point must be shifted not to the right but to the left. This is indicated by having a negative exponent. So you could write two trillionths like this:

    2E‾12

In the same fashion, you can write negative two trillionths like this:

    ‾2E‾12

Note that the two negative symbols that are in the representation of "negative two trillionths" occur independently. The first one means that the whole value of this number is negative. The second one means that this is a number with a very small magnitude.

Roughly speaking, APL\360 can work with numbers (positive or negative) whose magnitude ranges from a minimum of about 1E‾75 to a maximum of about 1E75.

Precision of Numbers

Internally, the computer represents numbers with a precision equivalent to about 16 decimal digits. Inevitably, any sequence of operations on values each of which requires the full precision will result in some cumulative error, so that the results (even though calculated to the equivalent of 16 decimal digits) are not necessarily that significant.

Number Display

Regardless of precision, it would be cumbersome to print all 16 digits every time you asked to see a number. So, when APL\360 prints a number, it prints only the 10 most significant digits, and suppresses trailing zeroes to the right of the decimal point. If you ask for the reciprocal of

3, the result that you see printed will show only ten places after the decimal point, even though the calculation was carried further than that:

```
      ÷3
0.3333333333
```

Similarly, 1 million divided by 3 is printed with ten digits:

```
      1E6÷3
333333.3333
```

And one million million million divided by 3 is also printed with ten digits:

```
      1E18÷3
3.333333333E17
```

But although only ten digits are printed, when a value is stored in the workspace, it is stored with APL\360's full precision, equivalent to about 16 decimal digits, and this is what is used in any calculation involving that stored value.

If for some reason you want your results printed with more or fewer digits, you can reset the number of digits printed in a particular workspace by using the command)DIGITS followed by any integer from 1 to 16.

Which Form Does the Computer Use?

When you are entering a numerical value, you may use whichever form you like (assuming that it is adequate to describe the number you want to enter). But when the computer types a number, it selects one of the forms according to its own preference. For instance, the computer always arranges numbers that it types in exponential form so that the left portion (the mantissa) is between 1 and 10, regardless of the way you entered the number:

```
      602E21
6.02E23
```

```
      1E6
1000000
```

```
      .0000000000000000000000000000000000000025
2.5E⁻37
```

```
      1.0
1
```

As you have just seen, it is quite possible that you
choose to represent a number in one form, while, when it
types, the computer represents the same value in the other
form. None of this makes any difference to the calculation,
since the way numbers are typed during input or output is
independent of the way they are represented inside the
computer during a calculation. For the computer, as for you,
the choice of one or the other form for writing numbers is
merely a matter of convenience in typing.

9: TESTING THE TRUTH OF A RELATIONSHIP

In the course of a calculation, there will be occasions when you want to know whether a particular relationship holds or not. You may want to test whether a counter has reached its maximum, or you may want to check whether a trial result is close enough to a desired standard of accuracy. Possibly you want to do something differently in your calculation, depending upon whether a particular condition is or is not met. APL includes operators which test whether two quantities are equal, as well as other relations.

The following APL operations test the truth of a relationship:

        <   less than

        ≤   less than or equal to

        =   equal to

        ≥   greater than or equal to

        >   greater than

        ≠   not equal to

Consider the following exchange:

```
A←12345678
B←12345679

A=B
```
0
```
A<B
```
1
```
A≥B
```
0

The computer always evaluates the truth of a relationship with 1 for true and 0 for false. Notice that because the result of testing one of these relationships is a number, it can be used in subsequent calculations.

$1+A \leq B$

2

Each time a relationship is tested, think of it this way: = means "Is it the case that A=B ?" (and similarly, "Is it the case that" A≤B, or A≠B, or whatever it is). If the answer is "Yes," the computer says 1; otherwise, 0.

Notice that these instructions do not tell the computer that A is less than B (or whatever the relation is). Nor do they instruct the computer to make A less than B. They test the truth of the relationship.

Example of Test for Equal

Suppose the correct answer to a problem has been stored as the value of a variable called RIGHT. Suppose that the answer supplied by a student has been stored under the name STUDENT. You need to keep track of the student's score. You want to add 1 to his score if his answer is the same as the right answer, and otherwise leave his score unchanged.

If the student got this problem right, then it is true that *STUDENT=RIGHT*. To add 1 to his score if and only if his answer is equal to the right answer, you could give this instruction:

SCORE ← SCORE + STUDENT=RIGHT

Then the amount added to SCORE will be 1 when the two values are equal, and 0 when they are different.

The example could be made slightly more complicated. Suppose that instead of adding 1 when the student is right, you wish to give some problems more weight than others. The weight for the current problem is stored under the name WEIGHT. If the student gets this problem right, you want to add WEIGHT to his score; otherwise, 0.

SCORE ← SCORE + WEIGHT × STUDENT=RIGHT

If the student's answer is equal to the right answer, then *STUDENT=RIGHT* has the value 1, so the amount that is added is *WEIGHT×1*. But if they are not equal, then the amount added is *WEIGHT×0*, which is 0.

How Close is Equal?

We have already mentioned that the computer stores the values of numbers out to about sixteen decimal digits. It is not programmed to handle greater precision than that. However, if you perform calculations on those stored numbers, there is almost certainly some loss of accuracy, so that although a result is carried to about sixteen digits, the final digits may become meaningless.

Whenever you ask a computer whether two quantities are equal, you have to qualify that question, and ask it (in effect), "Are these quantities equal <u>as nearly as it is reasonable to judge?</u>" APL\360 judges two quantities to be equal if the relative difference between them is less than 1 part in about ten million million (i.e. 1 part in $1E13$).

As we noted earlier, APL\360 types a maximum of ten significant digits. This means, in effect, that the typed answer is rounded to the nearest 1 part in $1E10$, and numbers which are not in fact equal may look alike when printed.

Occasionally this may cause some puzzlement. For instance, suppose you have two variables called A and B. If you ask for these two to be printed, exactly the same values are typed for both of them. But if you test to see if they are equal, you find that they are not!

```
A←÷3
B←.3333333333
A
0.3333333333
B
0.3333333333

A=B
0
```

This arises when the difference between A and B is large enough to be detected when the two stored values are compared, but not large enough to show up when they are typed. You can make the difference evident by finding the difference between A and B:

```
A-B
3.333332221E¯11
```

Notice that these are not mistakes, but the normal consequence of two facts:

1. No matter how you perform a computation or comparison, there must be some finite limit on the precision.

2. APL\360 does not normally display all the precision of which it is capable.

10: MORE OPERATIONS IN ARITHMETIC

So far we have considered the following arithmetic operations: addition, subtraction, multiplication, division, exponentiation, maximum, and minimum. . In this chapter we present capsule summaries of four other arithmetic operations, and four logical operations.

Absolute Value

Sometimes you want to consider the magnitude of a number without regard for whether it is positive or negative--that is, its absolute value. In conventional arithmetic, absolute value is often indicated by placing a vertical bar on either side of the name of a variable, thus:

$$|a|$$

In order to keep its syntax consistent, APL dispenses with the need to write the sign twice, and writes "the absolute value of A" like this:

$$|A$$

If A has a positive value, then $|A$ has the same value. But if A has a negative value, then $|A$ has the same magnitude but a positive sign.

Like every APL operator, $|$ operates on everything to the right of it, so that

$$|A+BZ \times Q$$

means "the absolute value of the sum of A and the product of BZ and Q"

Residue and Remainder

$$A|B$$

is read as "the A residue of B." The A residue of B is the smallest non-negative number that could be reached if you started out from the number B and added or subtracted the absolute value of A as·often as necessary. If A and B are both positive, this is the same as saying that the A residue of B is the remainder when B is divided by A.

If B is evenly divisible by A, then $A|B$ must be 0. By testing the truth of the relation $0=A|B$ you could decide whether B is divisible by A.

A program which prepares monthly statements includes a variable MO which contains the number of the current month. At the beginning of each new month, the program updates the stored values of MO. The months run from 1 to 12, so that the next month after month 12 is month 1. The following instruction would update the months correctly:

     $MO \leftarrow 1+12|MO$

For instance, at the end of March, MO has the initial value 3:

     $MO \leftarrow 3$
     $1+12|MO$
4

But at the end of December, when MO is 12, the same instruction has this result:

     $MO \leftarrow 12$
     $1+12|MO$
1

Powers of the Natural Constant e

If you type the symbol for exponentiation $\star$ with no left argument, APL presumes that the number which is to be raised to a power is the natural constant e. Thus e^a in APL is written $\star A$

The formula for the height of the Gaussian "normal curve of error" (when the total area under the curve is 1) provides that the height (i.e. frequency) Y of a deviation of T units from the mean may be found by the following formula:

$$Y = \frac{1}{\sqrt{2\pi}}\, e^{-T^2/2}$$

The reciprocal of the square root of two pi[*] is constant in this formula. Suppose we call that constant RTP; in APL, it may be found as $RTP \leftarrow \div(2 \times PI)\star 0.5$ Then the formula for Y becomes:

     $Y \leftarrow RTP \times \star -0.5 \times T \star 2$

---

[*]The value of pi is directly available from the monadic use of the circular function $\circ$ (see p. 66).

This might be embodied in a program called GAUSS:

```
      ∇ GAUSS
[1]   Y←RTP×*-0.5×T*2
      ∇
```

The height of the curve at its center, when the deviation T is zero, is found in the following execution:

```
      T←0
      GAUSS
      Y
0.3989422804
```

And the height when T is 2 units:

```
      T←2
      GAUSS
      Y
0.05399096651
```

Logarithms

The log of B to the base A is written:

```
      A⊛B
```

(The symbol for logarithm is formed by overstriking the circle ○ and the sign for exponentiation *.)

The common (i.e. base 10) logarithm of NUMBER can be found by the following instruction:

```
      NUMBER←20
      10⊛NUMBER
1.301029996
```

And the base 2 log of NUMBER is found this way:

```
      2⊛NUMBER
4.321928095
```

In order to approximate the responsiveness of human senses, radio engineers convert the power of an audible signal into units called decibels. The change of intensity, in decibels, measured with respect to an arbitrary reference power, is found from the formula

$$db = 20 \log_{10} \frac{power}{ref}$$

In APL, this becomes:

$$DB \leftarrow 20 \times 10 \circledast POWER \div REF$$

For a reference power of .002, an observed power of .08 is converted to decibels as follows:

```
POWER← .08
REF   ← .002
DB    ← 20×10⊛POWER÷REF
DB
32.04119983
```

Natural Logarithms

Just as the powers of e can be found by entering * with no left argument, so the log to the base e (the Napierian or natural logarithm) is found by entering the symbol for logarithm with no left argument. Hence the natural log of XYZ is found by the expression ⊛XYZ.

```
     ⊛2
0.6931471806
     XYZ←10
     ⊛XYZ
2.302585093
```

Antilogs

APL has no special symbol for the antilogarithm, since it can be found directly by exponentiation. The base 10 antilog of B is obtained by the instruction $10*B$, while the natural antilog of XYZ is found by $*XYZ$. For example:

```
     A←6
     B←7
     LOGA←⊛A
     LOGB←⊛B
     LOGPROD←LOGA+LOGB
     *LOGPROD
42
```

Logical Operations

The logical operations OR, AND and NOT operate only on zeroes or ones. Logical operations are most frequently used to form compound expressions about the truth of two or more

relationships. APL uses the number 1 to stand for "true" and the number 0 to stand for "false." Thus the logical operators can work on the result of any of the tests of relationship. But they aren't restricted to handling the results of relational tests; they can work on any values that contain only zeroes or ones, regardless of where those zeroes and ones came from.

Logical Or

Suppose A represents the truth of some relation, and B represents the truth of some other relation. Some condition you have in mind will be satisfied if either A or B is true. You can find the truth of "A or B" by the instruction

$A \vee B$

Suppose that in a particular program you are finding a solution by successive approximations. You will be satisfied if the result is correct within .0000001, but you will also be satisfied if the computer has already tried 100 approximations. You want to quit if either of those conditions is met. The first condition to test might be:

$1E^{-}7 \geq |LASTRESULT-NEWRESULT$

And the other one might be written this way:

$COUNT \geq 100$

An expression that yields a 1 if either of those conditions is true (i.e. has the value 1) is:

$(COUNT \geq 100) \vee 1E^{-}7 \geq |LASTRESULT-NEWRESULT$

In APL, as in logic, OR means the inclusive or: that is, you are satisfied if either one of the conditions is true, or if both of them are true.

Logical And

The instruction

$A \wedge B$

returns a 1 if and only if both A and B are 1. That is, $A \wedge B$ is true only when both A and B are true.

Let's return to the example in which we increase a student's score by 1 if his answer is equal to a right answer (page 56). Suppose now that this is a two-part question, and he has to have both parts right in order to get credit. If the student's two answers are called S1 and S2, and the correct answers are called R1 and R2, then you can keep track of his score by the following instruction:

$$SCORE \leftarrow SCORE + (S1=R1) \wedge S2=R2$$

In a certain jurisdiction, you can vote in school board elections if you are a citizen, and registered, and either a parent of a child in the local schools or a taxpayer to the school district. If CIT, REGD, PARENT, and TAXED are variables which indicate whether those conditions are met for an individual, you can combine them to test whether he is eligible to vote by the following expression:

$$ELIG \leftarrow CIT \wedge REGD \wedge PARENT \vee TAXED$$

Exclusive OR

In ordinary English speech, "or" often means "one or the other, but not both." Technically, this is the exclusive or. APL doesn't have a special symbol for exclusive or since the "unequal" operator provides this function. If A is a logical variable (i.e. is restricted to the values 0 or 1), and B is too, then

$$A \neq B$$

can have the value 1 if and only if one of those variables has the value 1 while the other has the value 0. The operation can be used to test whether any pair of values is unequal, including numerical values of any size, or even literal characters. But if the operation is applied to zeroes and ones, its effect is the same as an exclusive or.

In household electrical circuits, it is common practice to provide some lamps that may be turned on or off from either of two different switches--perhaps at the foot or the head of a staircase. The switches are arranged so that the current may flow when the two switches are in opposite positions. In that way, reversing the position of either switch always reverses the light. If the two switches are

called S and T, then the lamp (represented by the variable LAMP) is on (has a value 1) when:

$LAMP \leftarrow S \neq T$

NOT: Logical Negation

The operator ~ takes only one argument, which must be logical (i.e. must be composed exclusively of zeroes and ones), and produces a result of opposite truth. That is, the value of ~0 is 1, while the value of ~1 is 0.

Suppose a condition will be satisfied only if A is true and B is false. That can be tested by the result of this expression:

$A \wedge \sim B$

In constructing logical expressions involving the negation of some logical result, it may be handy to recall the equivalences given by De Morgan's rule:

Neither A nor B: $\sim A \vee B$ is equivalent to $(\sim A) \wedge \sim B$

Not both A and B: $\sim A \wedge B$ is equivalent to $(\sim A) \vee \sim B$

Getting a Variable's Sign with the Signum Function

The signum function is a handy way of obtaining the algebraic sign of a variable without regard to its magnitude. It returns a result of one when its argument is greater than zero, zero when it is zero, and negative one when the argument is negative. APL uses the monadic times sign to stand for the signum function:

```
      ×123.45
1
      ×¯667.89
¯1
```

The signum function may be useful in a variety of situations in which you have a three-condition test, such as "within range," "above range," or "below range." It is also useful in those situations in which it is more convenient to perform a calculation on the absolute value of a variable,

and then restore the appropriate sign at the end. Suppose you would like to convert a series of measurements to the log of their absolute value, but retain the original sign. You could do it like this:

$$LOGX \leftarrow (\times X) \times \circledast |X + X = 0$$

Sine, Cosine, and Tangent

In APL the trigonometric functions (which are also known as the <u>circular</u> functions) are provided by the circle symbol o. In order to get the necessary variety of functions using a single symbol, a left argument is used with the circle, as follows:

Sin A:	1oA	Arcsin A:	¯1oA
Cos A:	2oA	Arcos A:	¯2oA
Tan A:	3oA	Arctan A:	¯3oA

These are all defined in terms of radian measure, rather than degrees.

The circle with a left argument of zero is the function that gives the relation between the sine and cosine of an angle: 0oA is equivalent to (1-A*2)*0.5

The circle can also be used without an argument on the left, in which case it gives a multiple of the constant PI. For example, the number of radians in a circle is o2, and in a quarter circle (90 degrees), it is oo.5. If you are given an angle in degrees, and wish to find its cosine, you can write the single expression 2oo$A$÷180 since PI radians is equivalent to 180 degrees.

In Chapter 25 it is shown how a function can be defined to have an argument and a result. That capability can be used to obtain the circular functions in a convenient form, under their conventional names. For example, if a function SIN is defined as follows:

```
    ∇ Z←SIN A
[1]    Z←1oo A÷180
    ∇
```

it can be used like this:

```
    SIN 45
0.7071067812
```

11: CONTROLLING THE SEQUENCE
IN WHICH THE LINES OF A PROGRAM ARE EXECUTED

"Ordinary" Order of Execution

The ordinary order of execution of the lines of a program is to start at line 1, then do line 2, then line 3, and so on until the last line for which there is a definition. Inside each workspace, there is a line counter which tells the computer which line of the program it should execute next. When you call for a fresh execution of a program, it always starts out with line 1. In the usual course of events, in order to decide which line to do next, the computer simply adds 1 to the last value of the line counter.

In the programs which have been used as illustrations thus far, work always ended because the line counter moved up in the usual sequence until it came to a line that had not been defined. If a program has 4 lines, after the computer executes line 4, it sets its line counter to 5, and looks for line 5. When it finds that there isn't any line 5, it concludes that it has reached the end.

Branches

There are many situations in which you want to be able to tell the computer to go to some other line of the program, instead of the one that it would ordinarily do next. For instance, after a particular sequence of lines has been executed, you might want to have the computer go back and do them again with a different set of values. If the sequence that you want to have repeated starts at line 3, you might want to be able to tell the computer, "Go back to line 3." Or, if you want to repeat the sequence starting at line 3 only if a counter has not reached a particular value, you might want to say "Go back to line 3 if COUNTER is less than VALUE, otherwise stop."

An instruction which explicitly tells the computer which line to go to next is written with a right-pointing arrow, followed by an expression whose value is the number of the line that is next to be executed. Such an instruction is called a branch. The two examples mentioned in the last paragraph would be written like this:

$\rightarrow 3$
$\rightarrow 3 \times COUNTER < VALUE$

The second example, which depends for its effect on some condition that is tested, is often called a conditional branch. This and other forms of conditional branch will be discussed in a moment.

Branching Out of a Program

A branch to a line for which there is no definition always causes the computer to conclude that work on the program is finished, just as it does if the line counter is set to a line 1 greater than the last line of the program. If a program has 5 lines, the instruction

$\rightarrow 6$

will terminate work on it. So would $\rightarrow 99$, or $\rightarrow 678$. But the most obvious line number for which no instruction is ever defined is line 0. Hence, if for some reason a program needs an explicit instruction to end work, the instruction that's generally used is

$\rightarrow 0$

Naturally, you don't need to write $\rightarrow 0$ if the program comes to an end after the last line. (Although no line ever has a fractional number once function definition mode is ended, you can't use a branch to a fractional line number even to end execution of a program.)

Computed Branches

Instead of writing $\rightarrow 6$ you could just as well use this instruction:

$\rightarrow 2 \times 3$

The "go to" arrow means that the calculation on the right is to be performed, and the result of that calculation is to be used to reset the line counter for the current program.

Now suppose you give the instruction:

$\rightarrow 3 \times COUNTER < VALUE$

This calls for a test to see whether it is true that COUNTER is less than VALUE. If it is true, then the expression

COUNTER<VALUE has the value 1; otherwise, 0. Thus this instruction either means "Go to 3" or else it means "Go to 0 --i.e. exit from the program." Which of those meanings prevails depends in any instance upon whether it is true that COUNTER is less than VALUE.

The Factorial: An Example of a Program with a Branch

Suppose you want a program to compute factorials. The factorial of n is the product of the consecutive integers from 1 to n. You will need a counter; call it X. You will also need another variable F, to hold the result as it is developed. Start with X set equal to 1 and F also set equal to 1. (It's all right to write both of those in the same line.)

```
     ∇ FACTORIAL
[1]    F←X←1
```

Next increase X by 1. Then respecify F as the product of F and X.

```
     ∇ FACTORIAL
[1]    F←X←1
[2]    X←X+1
[3]    F←F×X
```

If N is the number whose factorial is to be computed, you now need an instruction that says "Go back to line 2 if it is true that X is less than N; otherwise go to 0."

```
     ∇ FACTORIAL
[1]    F←X←1
[2]    X←X+1
[3]    F←F×X
[4]    →2×X<N
[5]    ∇
```

Here is a sample execution of the program called FACTORIAL. First you set a value of N; then you call for execution of the program; finally you ask for display of the latest value of F.

```
     N←12
     FACTORIAL
     F
```

Program Loops

In the FACTORIAL program, the sequence of lines 2 to 4 is repeated as many times as required. A repeated segment of a program is called a loop. Whenever you write a program with a loop, there is some danger that a mistake in the program will cause the loop to be executed endlessly. For example, if the instruction on line 4 has requested a return to line 3 instead of to line 2, X would never be increased. The computer would return to line 3 indefinitely, because X would always be smaller than N. In this example, F would get larger and larger, being doubled at each repetition of line 3. Eventually the program would stop when the size of F exceeded the capacity of the computer.

Any time the computer seems to be taking longer to execute a program than you think it should, it is possible that it is in an endless loop. It is good practice to use the interrupt feature (see page 12) to stop it. If all is as it should be, you can tell the computer to resume where it left off by entering a branch instruction from the keyboard; this is discussed in more detail in Chapter 14, "What to Do When the Program Stops."

The Roots of a Quadratic: Another Example
Of a Program With a Conditional Branch Out

There are various ways of finding the roots of a quadratic equation. One of the best known goes as follows. Arrange the equation so that it is in the form

$$ax^2 + bx + c = 0$$

Then the roots are given by the formula:

$$\text{roots} = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

Suppose you want to write a program to calculate and print the values of the two roots. The problem that arises is this: the quantity $b^2 - 4ac$ (which is called the discriminant) may be negative. If the discriminant is negative, the roots are complex. If you woodenheadedly go ahead and try to calculate them anyway, you'll be in trouble when you try to take the square root of a negative discriminant. So you want to test whether the discriminant

is negative. For the moment, assume that when you find a negative discriminant, you want to terminate execution of the program, but if the discriminant is not negative, you'll go ahead with the calculation. (You're also in trouble if A is zero, since that would give you a 0 divisor, but let's not worry about that since if A is zero this isn't a quadratic equation.)

Here's an outline of the procedure you can use in a program to find the real roots of a quadratic. Suppose you call the program QUADROOTS.

    1. Calculate the discriminant and store it.

    2. If the discriminant is negative, go to 0 (i.e. quit).

    3. Otherwise, calculate and print the values of the two roots.

In writing this program, you can find the discriminant on line 1. Then line 2 is a branch: if the discriminant is negative, go to 0. Otherwise, go to 3. The program looks like this:

```
     ∇ QUADROOTS
[1]    DISC←(B*2)-4×A×C
[2]    →3×DISC≥0
[3]    (-B-DISC*0.5)÷2×A
[4]    (-B+DISC*0.5)÷2×A
     ∇
```

If a negative discriminant is encountered with this program, the computer will simply terminate execution without doing the calculation. Some procedures which are more general, and handle both the real and the complex roots, are, of course, possible, but aren't discussed here.

Branch or Continue

Line 2 of the QUADROOTS program says (in effect) "Go to 0 if the discriminant is negative, and otherwise go to line 3." It is more convenient to write an instruction which doesn't require you to know that the next instruction is on line 3. You would rather say, "If the branch is not taken, go to whatever line comes next." You can do that in the following way. The instruction

    →CONDITION/LINE

causes a branch to the line number specified by *LINE* if and only if *CONDITION* is true (i.e. has the value 1 rather than 0). When *CONDITION* is false (i.e. is 0), the program continues with the next line in the usual sequence.

This expression, involving the / sign, is actually an example of a much more general operation called <u>compression</u>, which is discussed in a later section. For the <u>moment it is</u> only necessary to note the way that it is used to provide an instruction which has the effect of "Branch if the tested condition is true, but otherwise continue in sequence."

Instructions which test whether the discriminant is negative, and go to 0 if it is but otherwise continue, could be written like this:

```
TEST←0>DISC
→TEST/0
```

Probably you don't want to create a stored variable called TEST on one line, and then branch on the next line. You can instead insert the formula for the condition being tested right into the branch instruction. However, now you must put parentheses around the expression for the test, so that the test is evaluated before deciding the branch:

```
→(0>DISC)/0
```

We can now go back to the QUADROOTS program and give it a different line 2, so that a display of the entire program now looks like this:

```
    ∇ QUADROOTS
[1]   DISC←(B*2)-4×A×C
[2]   →(0>DISC)/0
[3]   (-B-DISC*0.5)÷2×A
[4]   (-B+DISC*0.5)÷2×A
    ∇
```

Here is an example of the quadroots program in use. Suppose you need the roots of the following equation:

$$14x^2 - 2x = 18.6$$

Putting it into the form $ax^2+bx+c=0$, you find that

A is 14; B is -2; C is -18.6

Because the program presumes that values   of A, B, and C are
already   in   storage, you   must   enter   those values   before
calling for execution of the program:

     $A \leftarrow 14$
     $B \leftarrow ^-2$
     $C \leftarrow ^-18.6$

    Then you   call for   execution of the   QUADROOTS program
simply by typing its name.

     QUADROOTS
1.226276384
$^-$1.083419241

The Factorial Again: An Improved Version
Using Two Branch Instructions

    If   you try   executing the   factorial program   shown on
page 69   with small values of   N, you run into   some strange
results:

     $N \leftarrow 2$
     FACTORIAL
     F
2

     $N \leftarrow 1$
     FACTORIAL
     F
2

     $N \leftarrow 0$
     FACTORIAL
     F
2

    Something is wrong. The factorial of 1 should be 1. The
factorial of 0 is also defined to be 1. Where is the error?

    You will   recall that line   4 of the   FACTORIAL program
said, in effect, "Go back and multiply F by the next integer
if the counter X is less than the number N." But before the
computer   ever   gets   to   make that   test,   it   has   already
multiplied F by X+1, or 2,   regardless of the value of N. If
this program   is to work   properly for all   the non-negative
integers,   this   superfluous   multiplication   must   be
forestalled.

The answer lies in putting the test ahead of the loop.
That way, when appropriate, there may be zero repetitions of
the loop, since the test may cause the computer to skip
before it ever executes the instructions in the loop. The
branch instruction should therefore come right after line 1.
It should say, in effect, "Stop now if X is greater than or
equal to N. Otherwise continue with the instructions in the
loop." This may be written as:

$$\rightarrow(X \geq N)/0$$

As you will see by studying the program below, this test,
executed before the loop is entered, is the only test
necessary. The loop is closed by the instruction at line 5
to return to line 2, and test again. Here is the program as
revised:

```
     ∇ FACTORIAL
[1]    F←X←1
[2]    →(X≥N)/0
[3]    X←X+1
[4]    F←F×X
[5]    →2
     ∇
```

Sample executions of this program now give the correct
results:

```
     N←2
     FACTORIAL
     F
2
     N←1
     FACTORIAL
     F
1
     N←0
     FACTORIAL
     F
1
     N←12
     FACTORIAL
     F
479001600
```

This method of constructing a loop, with the test at
the beginning, is sometimes known as the "method of leading

decisions." While it requires two branch instructions (one at the beginning and one at the end of the loop), it will often keep you out of trouble and make for a neater program, as it does in the case just illustrated.

(It should also be noted that factorial is also available as a primitive operation in APL, so that, apart from this exercise, you wouldn't need to write a factorial program at all. See Appendix A.)

Techniques for programming with loops are discussed further in Chapter 22.

This page intentionally left blank.

12: ARRANGING THE WAY THE PROGRAM TYPES ITS OUTPUT

Frequently, you will want to write a program in such a way that the computer automatically types readable output, without your having to give special instructions each time. If you get much output printed by the computer, pretty soon you're going to want some headings to distinguish what is what.   You can instruct the computer to print alphabetic characters. Then you can arrange these as headings for the results of a program, or as any other message you may want to have typed. (The text of this primer was typed and arranged by a program running on APL\360.)

<u>Printing Text</u>

Literal text can be entered by using quote marks. If you type

'THIS IS A SAMPLE OF LITERAL TEXT'

you have entered a quotation. Since you haven't said what is to be done with the quotation, as usual the computer assumes that it should be printed. Your dialogue with the computer looks like this (first your instruction, then the computer's reply):

'THIS IS A SAMPLE OF LITERAL TEXT'
THIS IS A SAMPLE OF LITERAL TEXT

The quote marks mean that what you typed between them was a quotation. They aren't part of the quotation itself, so they do not appear when the computer types the quotation.

You can store a quotation in the same way that you store anything else. If you type

X←'IN 1492, COLUMBUS SAILED THE OCEAN BLUE.'

a variable named X is created in the workspace. Its value is that quotation. If you ask to have X typed, the dialogue will go like this:

X
IN 1492, COLUMBUS SAILED THE OCEAN BLUE.

Anything that you type between quotation marks is accepted as literal characters. Quoted text is not executed.

Operator signs, variable names, spaces, digits...if they are in quotes, they are just so many literal characters, with no meaning to the computer as names, operators, or numbers. Any character you can print from the keyboard can be included inside the quote. The computer will either store this string of characters, or print it, as you direct. In this way you can put together captions and headings that will make your output easier to understand.

A quotation must have a quote mark at the beginning and one at the end. Once you use one quote mark, everything that you type after that is a part of the quotation until you reach another quote mark. This fact occasionally trips an inexperienced user. He types one quote mark, and then changes his mind and decides to do something else. He types what he thinks is an instruction to the computer, and meanwhile the computer is still compiling the quotation he started but never finished.

Lines of a program which call for the printing of quoted text can be used to get a program to print headings. For instance, in the QUADROOTS program, ahead of the lines that calculate and print the two roots, you could insert lines which call for the printing of appropriate text. Here is a revised version of that program. Lines 3 and 5 now call for the printing of headings.

```
      ∇ QUADROOTS
[1]    DISC←(B*2)-4×A×C
[2]    →(0>DISC)/0
[3]    'THE VALUE OF THE FIRST ROOT IS'
[4]    (-B-DISC*0.5)÷2×A
[5]    'THE VALUE OF THE SECOND ROOT IS'
[6]    (-B+DISC*0.5)÷2×A
      ∇
```

Here is a sample execution of QUADROOTS, as revised:

```
      A←14
      B←¯2
      C←¯18.6
      QUADROOTS
THE VALUE OF THE FIRST ROOT IS
1.226276384
THE VALUE OF THE SECOND ROOT IS
¯1.083419241
```

As a further variation, you can have the program type another quotation to indicate what has happened when it finds that the discriminant is negative. To do this, you have to make the following changes in the QUADROOTS program:

1. On line 2, if it is true that DISC is negative, instead of branching to 0, branch to a line which contains some suitable quotation.

2. At the end of the program, add that quotation. It is to be typed only when DISC is negative.

3. When there are real roots, line 6 is still the last executable line of the program. After the computer executes line 6, you want it to finish work without running into the quotation about complex roots. So you should insert a branch to 0 after line 6.

Here's the revised program, followed by two sample executions to illustrate the alternative headings:

```
     ∇ QUADROOTS
[1]    DISC←(B*2)-4×A×C
[2]    →(0>DISC)/8
[3]    'THE VALUE OF THE FIRST ROOT IS'
[4]    (-B-DISC*0.5)÷2×A
[5]    'THE VALUE OF THE SECOND ROOT IS'
[6]    (-B+DISC*0.5)÷2×A
[7]    →0
[8]    'ROOTS COMPLEX; CALCULATION TERMINATED.'
     ∇

       A←10
       B←12
       C←22
       QUADROOTS
ROOTS COMPLEX; CALCULATION TERMINATED.

       A←10
       B←¯22
       C←4
       QUADROOTS
THE VALUE OF THE FIRST ROOT IS
2
THE VALUE OF THE SECOND ROOT IS
0.2
```

Results and Heading Appearing on the Same Line

A neater output is sometimes obtained when the heading and the result are typed so that they appear on the same line. This is called "mixed output." A line of a program which calls for mixed output has the following characteristics:

1. Different items to appear on the same line are separated by semicolons.

2. An item within a line of mixed output may be a variable, literal characters, or the result of an expression.

3. If blank spaces are to appear between the items, the blanks must be specifically included as parts of the quotations. Mixed output printing does not automatically supply spaces between the items.

4. A line of mixed output may not be stored as a single variable.

Here is yet another version of the QUADROOTS program, this time written to use mixed output, followed by sample executions that show the same two problems used on page 78.

```
      ∇ QUADROOTS
[1]    DISC←(B*2)-4×A×C
[2]    →(0>DISC)/6
[3]    ' FIRST ROOT: ';(-B-DISC*0.5)÷2×A
[4]    'SECOND ROOT: ';(-B+DISC*0.5)÷2×A
[5]    →0
[6]    'ROOTS COMPLEX;  CALCULATION TERMINATED.'
      ∇
      A←10
      B←12
      C←22
      QUADROOTS
ROOTS COMPLEX; CALCULATION TERMINATED.

      B←¯22
      C←4
      QUADROOTS
 FIRST ROOT: 2
SECOND ROOT: 0.2
```

13: LINE LABELS FOR EASIER BRANCHING

In each of the examples of a branch instruction introduced thus far, you had to know the number of the line you were branching to. For instance, in writing the instruction

$\rightarrow(0>DISC)/6$

you had to know that the instruction you wanted next was on line 6. But as you saw in the discussion of inserting a line in a program, or deleting a line of a program, it is possible that the instruction which used to be the sixth one in the program will be moved up or down as lines are inserted or deleted ahead of it. In that case, you'd have to rewrite the branch instruction each time so that it always showed the correct number of the line you want to branch to.

There is an easier way to handle this problem. You can create a variable which is automatically assigned a value that is the number of the line at which a particular instruction is located. When you write a branch instruction, you write it in terms of that name. If the discriminant is negative in the QUADROOTS program, you want the computer to go to the line that deals with complex roots, wherever that line may be. Suppose you give that line the name COMP. Then you write the branch instruction like this:

$\rightarrow(0>DISC)/COMP$

A variable like COMP, whose value is the line number for a particular line of a program, is called a label. You show the computer what line the label goes with by typing the label and a colon in front of that instruction.

If the instruction at COMP asks for the printing of a message saying that the roots are complex, where formerly you had

[6]   'ROOTS COMPLEX; CALCULATION TERMINATED.'

now, with a label on this instruction, it looks like this:

[6]   COMP: 'ROOTS COMPLEX; CALCULATION TERMINATED.'

and COMP becomes a variable whose value is 6.

The computer automatically sets the values of labels each time you leave definition mode for that program, so that after each revision of a program each label again shows the correct position of the line to which it is attached. Because a label is a variable, it is necessary that a label have a name distinct from the name of any program, or any other variable in the same workspace.

14: WHAT TO DO WHEN THE PROGRAM STOPS

While you enter the definition of a program, the computer stores the definition, line by line, in the active workspace. It doesn't make any check to see whether your definition makes sense. You won't discover whether the definition is satisfactory until you try executing it on a few examples. It's a good idea to start by running a problem for which you already know the right answer. If the definition is correct, the computer will run through your program without mishap, and you will get the appropriate results. But if some of your definition is in error, your mistake will come to light in any of the following three ways:

1. The computer stops without finishing work on your program because it has come across an instruction that cannot be executed.

2. The computer doesn't stop work on your program in a reasonable time, probably because you've mistakenly given it an endless task. If a simple program doesn't produce results in a second or two, you'd better press the ATTN key to interrupt the computer.

3. The program runs, but the result it produces isn't what it should be. Your definition is acceptable to the computer, but it isn't what you really wanted.

The first of these three is probably the most common. Mistakes of this kind also come to light first, since if the computer can't execute the instruction at all, it doesn't get a chance to reveal any of the other kinds of error.

Halt When an Instruction in Your Program Can't Be Executed

If the computer finds that it cannot execute an instruction in your program, here's what it does:

1. It types an error message. This identifies the type of trouble the computer ran into as it tried to execute the instruction.

2. It types the name of the program and the number of the line on which it was working when the trouble

was encountered, together with the complete
instruction on that line.

3. It types a caret to show you how far along in the
instruction it had gone (working through the
operations from right to left) when the trouble
was encountered.

The error message is the computer's report telling you
what type of trouble it has run into. There are ten
categories that you might possibly encounter during the
execution of a program. Here are three of the more common
errors:

Value error means that your instruction refers to a
variable which has not been assigned a value in this
workspace.

Syntax error means that your instruction violates the
rules of APL syntax, by such things as mis-matching
parentheses, or failing to show what operation is to be
performed on a pair of variables, or failing to provide an
argument for an operator.

Domain error means that you have given an APL operator
an argument that is outside the domain of values that that
operator can handle. You would have such an error if you
were inadvertently dividing by zero, or trying to do
arithmetic on a literal character.

There is an extensive summary of error messages in
Appendix D. You may want to look through that appendix
briefly, and then refer to it again as the need arises.

A Program Error Doesn't Mean
That Execution Is All Over

The cure for a great many program errors is to rewrite
the defective instruction. You can do this without having to
abandon execution of the program, and without having to
start over from the beginning.

Whenever the computer encounters an unexecutable line
in a program, it halts the work and prints an error message.
But that doesn't mean that execution is all over. The
execution is suspended for whatever corrections you wish to
make. The computer awaits a branch instruction from you to

tell it where to resume work on the suspended program. This fact has two important consequences.

First, while execution is suspended, you may perform any calculation. You can display the values of variables used in your program, or almost any others.

You can enter the definition of a new program, or edit the definition of almost any program. In particular, you can usually edit the definition of the suspended program, and thus correct the mistake that produced the error. (There is one restriction. You can't edit the definition of a program whose execution is pendent-- i.e. whose execution has been started but has not been either terminated or suspended. A program can be pendent only if an instruction within it calls for the execution of another program, and that other program has been halted. (See the discussion of definition errors, in Appendix D.)

Second, sooner or later you should tell the computer where to resume work on the suspended program, or else terminate work by the instruction →0. The computer will wait indefinitely for your instruction telling it where to resume. If you decide to save this workspace and resume work on it another day, the computer will save along with the workspace the list of programs whose execution is still pending. You aren't required to dispose of these halted executions...but it's a good idea, since they take up some space in the workspace and if you don't dispose of them you may gradually accumulate a large number of them (see the discussion of depth errors, Appendix D).

Resuming Execution

If you wish to resume execution of a suspended program at the place where work was halted, you enter a right arrow and the number of the line shown in the error message. If work was halted because of an error on line 3, the instruction

→3

causes the computer to resume work where it left off. Alternatively, you can resume execution at any other line of the program, by entering a right arrow followed by the number (or the label) for the line at which you want work to be resumed.

As usual, if you enter the instruction to branch to zero, or to any line for which there is no definition, execution is terminated.

Where Was Work Suspended?

It is important to keep track of which programs are halted, in what sequence, and at what point. You can find this out by displaying the state indicator. This gives you a list of the names of the halted programs in order, starting with the most recently halted one. For each program, the state indicator shows the line on which work was halted (and on which work should be resumed if you wish to continue in sequence). You cause the state indicator to be displayed by entering the command)SI.

The system distinguishes two different types of program halt. When a program halts because it has run into an unexecutable instruction, or because you've interrupted it by pushing the ATTN key, it is suspended. Work on it can't resume until you enter a branch instruction from the keyboard.

By contrast, a program may also be halted because one of its instructions calls for execution of another program, and that other program has been halted. The earlier program is waiting for the program it called to finish; if and when that called program is terminated, the earlier one will automatically resume execution. A program that is halted while waiting for another program to terminate is pendent. A pendent program may not be edited or deleted.

The state indicator shows you which programs are suspended by typing an asterisk after their names. The ones without asterisks are pendent.

```
      )SI
AREA[1]  *
WORK[2]  *
REPEAT[7]
```

In this example, three programs are halted. The most recently started program appears first on the list; when you type an instruction to resume execution, it always refers to the program at the top of the state indicator list.

The programs called WORK and AREA are suspended, but the program called REPEAT is pendent. This indicates that

REPEAT has not itself been suspended, but is held up because execution of the program called WORK has not been terminated. If and when the execution of WORK is completed, execution of REPEAT will resume automatically. Evidently the execution of WORK was initiated not by an instruction entered directly from the keyboard, but by the instruction located at line 7 of REPEAT.

Terminating Execution of All Halted Functions

Anytime you enter from the keyboard a right-pointing arrow followed by a line number (or label), it is taken as an instruction to resume work on the most recently halted program on the line you indicated. Sometimes you want to abandon all executions that are halted, and start over again. This is especially true if you have inadvertently accumulated a large number of halted executions, or if you have a pendent function that resumes automatically at an inconvenient spot whenever you terminate execution of the program ahead of it.

The state indicator may be cleared back to the last previous suspended function by entering a single right-pointing arrow with no value to the right of it, like this:

→

After you've entered a single right arrow, if you now ask once more to see the state indicator and it is not entirely blank, you can continue to clear it by entering another right-pointing arrow. When it is entirely cleared, any program can then be displayed or edited, and any program can then be erased.

Changing Individual Characters Within a Line

Chapter 7 explained how to edit a program by adding new lines, deleting lines, or substituting new instructions for individual lines of a program. In addition, APL\360 gives you a way to change, delete, or insert particular characters within a line, without having to retype the entire line. This is particularly handy for long or complicated lines, because it means that you don't run the risk of introducing new mistypings while correcting others. The procedure has three phases:

1. The computer types the line the way the definition is now.

2. Under this display, you enter marks indicating which characters are to be omitted, and where new characters are to be inserted.

3. The computer retypes the line, leaving out the characters you indicated, and inserting spaces where you requested. Then the computer positions the typeball at the leftmost inserted space, and unlocks the keyboard. It does this without advancing the paper, so that your insertions can be typed right into the blank spaces left in the line. Now when you press the carrier return key, the line is entered the way it appears on the paper (even though the computer typed some of those characters and you typed the others).

Illustration: Using Character Editing
To Correct a Mistake in a Program

Suppose you have a program called AREA, which calculates the area of a segment of a circle in terms of a constant called PI, and the variables ANGLE and RADIUS. Here is the definition:

```
     ∇ AREA
[1]    A←(PI×RADIUS)*2×ANGEL÷360
     ∇
```

Notice that there are two mistakes: first, ANGLE has been mistakenly spelled ANGEL; second, the parentheses are in the wrong place. They should surround the expression RADIUS*2. The first mistake makes the line impossible to execute (unless there happens to be a value for ANGEL in the workspace). The second mistake won't prevent execution, but it will make the result unreasonable.

Suppose you call for execution of this program called AREA. The computer encounters the mistaken reference to ANGEL, which (let's suppose) has never been defined. It types the following error message:

```
     AREA
VALUE ERROR
AREA[1] A←(PI×RADIUS)*2×ANGEL÷360
              ∧
```

As you look at the display of the offending line, you recognize the trouble, and you also notice that the parentheses are wrong. Here's how you could make all of those corrections.

You re-enter definition mode by typing a ∇ and the name of the program. Then, instead of defining an additional line 2, you ask for character editing of line 1. To get character editing, you type, in brackets, these three things:

1. The number of the line that you want to edit.

2. A quad symbol.

3. Some integer indicating your guess about the position in the line at which you will start making changes. Your guess needn't be very accurate; the fact that some number appears after the quad is what tells the computer that you want to start character editing.

Here's how your paper appears after you ask for character editing of line 1. (As before, the symbol @ indicates where the typeball is located at the moment that the keyboard unlocks.)

```
        ∇AREA
[2]     [1□12]
[1]     A←(PI×RADIUS)*2×ANGEL÷360
        @
```

Now you are in a special editing mode. Anything you type now will be considered as your instructions about the way the displayed line is to be edited. Wherever you type a /, the character immediately above it will be deleted. Wherever you type a numeral 1 to 9, that many spaces will be inserted in front of the character immediately above the numeral. In addition, a letter A means 5 spaces, B 10 spaces, C 15 spaces, and so on.

In the current example, you want to remove both the parentheses from their present positions, and you also want to remove the E in ANGEL. You want to insert one space for a left parenthesis in front of RADIUS, and another space in front of the multiplication sign that follows 2. You also want to insert one space in front of the division sign, for the final E in ANGLE.

After you've indicated which characters are to be deleted, and where you want space for insertions, the computer responds by retyping the line with the changes you've asked for. Then the computer backs up the typeball until it is opposite the first of the inserted spaces (without moving the paper up to a new line) and awaits your insertions. You may type anywhere on the line as long as you don't illegally overstrike the characters that are already there.

Step 1: Enter definition mode and ask for character editing of line 1 (shown on opposite page).

Step 2: Indicate which characters are to be deleted and where space is to be left for insertions.

```
        ∇ AREA
[2]     [1□12]
[1]     A←(PI×RADIUS)*2×ANGEL÷360
            /   1       /  1    //2
```

Step 3: The computer retypes the line, deleting the characters that you marked with /, and inserting spaces as you indicated in step 2.

```
        ∇ AREA
[2]     [1□12]
[1]     A←(PI×RADIUS)*2×ANGEL÷360
            /   1       /  1    //2
[1]     A←PI× RADIUS*2 ×ANG  ÷360
```

Step 4: You insert the missing characters in the spaces provided. (Unless you have a 1050 terminal with 2-color ribbon, it is now impossible to distinguish which characters you typed and which characters the computer typed. They all become part of the definition of that line.)

```
        ∇ AREA
[2]     [1□12]
[1]     A←(PI×RADIUS)*2×ANGEL÷360
            /   1       /  1    //2
[1]     A←PI×(RADIUS*2)×ANGLE÷360
```

Step 5: The computer invites you to define another line. Since there is nothing more to do, you type a ∇ and leave definition mode. This is illustrated on the next page.

```
        ∇ AREA
[2]     [1□12]
[1]     A←(PI×RADIUS)*2×ANGEL÷360
         /   1     /  1    //2
[1]     A←PI×(RADIUS*2)×ANGLE÷360
[2]     ∇
        @
```

<u>Step 6</u>: You resume execution with a branch instruction which tells the computer to resume work at the line on which it was suspended (i.e. in this case, at line 1).

```
        ∇ AREA
[2]     [1□12]
[1]     A←(PI×RADIUS)*2×ANGEL÷360
         /   1     /  1    //2
[1]     A←PI×(RADIUS*2)×ANGLE÷360
[2]     ∇
        →1
        A
2245.867725
```

After successful resumption, the result is found.

<u>Editing Characters in the Program Header</u>

Character editing can be applied to the program's header just as it is to any line in the body of the definition by referring to it as line 0 of the definition. This permits you to alter the spelling of the name of a program, and (more important) to alter the syntax governing a defined function, or the names of local variables within it--a topic that isn't discussed until Chapter 25.

<u>Tracing the Execution of a Program</u>

If a program comes out with a result that isn't what you expected, it is useful to check up on what was done on certain lines of the program. This is called <u>tracing</u> the execution of that program. The instruction

```
    TΔWORK ← 3 4 7
```

means that, until you instruct otherwise, the computer

should trace the execution of the program called WORK on lines 3, 4, and 7. When the computer traces, each time it executes a traced line, it prints the name of the program, the number of the line, and the result of that instruction.

To discontinue tracing, you type

$T\Delta WORK \leftarrow 0$

Suspending Execution Part Way Through

Sometimes it is useful to check up on what a program has been doing by suspending its execution at some intermediate point. You can cause the program WORK to suspend immediately before the execution of line 2 or line 8 by typing the following instruction:

$S\Delta WORK \leftarrow 2\ 8$

When it suspends execution, the computer types the name of the program and the number of the line which it is about to execute. After you've done whatever you wanted to do while execution was suspended there, you can resume execution in the usual way. That is, by typing a right-pointing arrow and the number of the line which is to be executed next.

To discontinue stopping, you type

$S\Delta WORK \leftarrow 0$

Stop or Trace Can Be Controlled by the Program Itself

A trace or stop instruction can be made part of a program. For instance, you might want to trace the execution of line 5 of the program WORK if and only if some variable B has a value greater than 1.5. That could be done by the following instruction within the program:

[4] $T\Delta WORK \leftarrow 5 \times B > 1.5$

If you change the definition of a line within a program, the revision will also discontinue tracing or stopping on that line. So after revising some lines within a program definition, you should restate which lines you want to have traced or stopped.

15: SYSTEM COMMANDS

APL is a language for describing mathematical procedures. APL\360 is a system for executing procedures written in the APL language. Most of what we have discussed so far has dealt with the operators of the APL language, and how you may define and execute programs using those operators. In addition to using the APL language itself, you also need to be able to give instructions directly to the computer. These concern such practical matters as signing on and off, saving your workspace for future use, borrowing variables or programs from other libraries, or establishing passwords that lock your account or your workspaces against unwarranted use by others. None of these matters is part of a mathematical procedure, and so none of them is dealt with in the APL language. However, the APL\360 System has a family of instructions called system commands, by which these and similar instructions to the computer are given. A few of them have already been introduced. This chapter pulls together some of the other system commands you are likely to need. You won't want them all at once, of course, but you should read through this chapter and come back to it as specific needs arise later.

Distinguishing System Commands from Other Instructions

A system command always starts with a right parenthesis. The right parenthesis was selected because no conceivable expression in arithmetic starts with a right parenthesis, and thus system commands can be readily distinguished from other instructions. Anything that you type which starts with a right parenthesis is treated as a system command.

A system command can never occur as part of a program. Whenever you enter a system command, it is executed at once, if possible, and in any case is never accepted as part of the definition of a program.

Signing On

Signing on has already been described in Chapter 2. In a way, it is the simplest of the system commands, since it consists of nothing but the right parenthesis and your user number. A sign-on will only be accepted if you aren't signed on already. If you mistakenly type another sign-on after you're already signed on, the computer rejects it, and reminds you that you're "already signed on."

If you have established a sign-on password, after your user number you must type a colon and then the password.

Before your sign-on is accepted by the computer, you cannot do any work. However, you can send a message to the operator (see page 103).

Signing Off

When you have finished working, you should sign off. There are several ways to do this. If next time you sign on you'll want to resume work right where you left off, you should sign off with the command)*CONTINUE* (see page 99). Otherwise, you sign off simply by entering the command:

)*OFF*

The system responds by typing a statement showing the amount of time you worked. Then it signs you off, and breaks the telephone connection.

Holding the Telephone Line for the Next User

If another user would like to sign on from the same terminal right after you, you can save him the trouble of dialling the computer by instructing the computer to hold the telephone connection after you have signed off. You can do this by typing the command:

)*OFF HOLD*

This causes the computer to wait 60 seconds before it disconnects the telephone line. During that time, another user may sign on without having to redial the telephone.

Establishing a Sign-On Password

When you sign off, you may also, if you wish, establish a password which will thenceforth be required whenever you sign on. You do this in the following way. At the end of the sign-off instruction (which may be)*OFF* or)*OFF HOLD* or)*CONTINUE* , mentioned below, see page 99) you type a colon followed by any single word. For instance, the password SHAZAM would be established by signing off like this:

)*OFF:SHAZAM*

From now on, whenever you sign on, you will have to type not

only a right parenthesis and your user number, but also a
colon and the password SHAZAM. This lock remains in effect
until, at some subsequent sign-off, you specify some other
lock. If you sign off with a colon but you don't indicate a
key, that will mean that from now on no password is needed
(and the colon isn't needed either).

Saving a Workspace

After you have done some calculations or defined some
programs, you may want to stop and resume work at some later
time or some other day. There are two ways to do this: by
using the command to save your workspace, or by signing off
with the command continue.

If you give the system command to save your currently
active workspace, the computer makes a complete copy (on
magnetic disc) of everything that's in your workspace at the
moment you give the save command. The entire contents of the
workspace is saved: programs, stored data, the list of
programs awaiting execution--all of it.

The save command does not alter what is in your active
workspace, but causes the computer to save an exact
duplicate of it.

When a workspace is saved, it must have a name. This
name will be used to locate it when you subsequently ask to
retrieve a copy of the saved workspace. Once a workspace has
been named, you have only to enter the command *)SAVE* and the
computer saves a duplicate of your active workspace under
that name. The computer reports to you the date and time at
which the save took place, and the workspace's name. For
example, if your workspace has been named ACCOUNT, then the
dialogue when you save it will look like this:

)SAVE
 12.45.26 07/28/68 *ACCOUNT*

The most recently saved workspace named ACCOUNT replaces any
earlier workspace of that name in storage. In effect, you
have revised the earlier saved workspace, replacing it with
a later version, whose content may be the same or totally
different, but which has the same name.

Naming a Workspace

You may give your active workspace a new name by using the command)*WSID* (for "workspace identification") followed by the name you've selected. To give your active workspace the name TABLE, you enter:

)*WSID TABLE*

If the workspace already had a name, the computer reports to you what the former name was:

)*WSID TABLE*
WAS ACCOUNT

The name of a workspace can be any single word which starts with a letter of the alphabet and has any letters or numerals in the remainder. The computer only reads the first 11 characters in a workspace name.

If you don't recall the name of your currently active workspace, you may inquire by entering)*WSID* with no name after it, and the computer types the workspace name.

The commands to save a workspace and to name it can be combined by stating a name with the)*SAVE* command. However, the computer won't accept this command if you attempt to save a workspace which now has one name under a name which is already in use for another saved workspace. This prevents you from accidentally replacing one workspace with a quite different one.

Your Library of Saved Workspaces

The collection of all of the workspaces that you have saved is referred to as your library of saved workspaces. You may execute a save only into your own private library. There is no way for you to save a workspace so that it becomes part of some one else's private library. (However, it is possible under some circumstances to save your active workspace into a public library, that does not belong to any individual user.)

The system command)*LIB* causes the computer to type the names of the workspaces currently stored in your private library.

Occasional users will probably not require a library of more than one or two workspaces. Heavy users may well need more. Each user is assigned a quota which is the maximum number of workspaces he may save. The system won't let you save another workspace if your quota is used up. If there's no room in your library for something you want to save, you should drop a saved workspace (see below, p. 99), or ask the operator how to arrange for a larger library.

Bringing a Saved Workspace into the Active Area

In order to use a previously saved workspace, you have to give the system command to <u>load</u> that workspace. This causes the computer to load into your active area a complete copy of the entire saved workspace. Your active workspace is now restored so that it is exactly the way it was at the moment the workspace was saved. Anything in the active area before you gave the load command is replaced by the material from the saved workspace.

After a load command, the computer confirms that a copy of the saved workspace has indeed been loaded by typing the date and time at which it was saved. Like this:

```
    )LOAD ACCOUNT
SAVED   12.45.26 07/28/68
```

You may load a saved workspace into your active area as often as you wish. Each time, the active area will be restored so that it is again exactly the way it was at the moment when that workspace was saved.

Loading a Workspace from a Public Library

An APL\360 System may have several <u>public libraries</u> in which have been stored workspaces that <u>contain programs</u> or data that may be generally useful to many users. Public libraries are assigned numbers less than 1000, while the private libraries of individuals have numbers greater than 1000. Anyone may load one of these workspaces and thus acquire a wide variety of special programs that are already written, tested, and ready for use. Or a group of users whose work is related may use this means for having some workspaces that can easily be loaded by any of their members. A listing of public library workspaces is available, together with directions for the use of prepared programs that have been stored in them.

To load a workspace from one of the public libraries, you type the system command *)LOAD* followed by the number of the public library and the name of the workspace. For instance, if Public Library 1 contains a workspace called NEWS, you can get a copy of it by the following command:

>   *)LOAD 1 NEWS*
> *SAVED*   23.13.46 07/29/68

Now if you ask for the name of the active workspace, the computer supplies its library number as well:

>   *)WSID*
> 1 *NEWS*

Saving a Workspace Into a Public Library

If you save a workspace into a public library, that still counts as part of your ration of saved workspaces. However, it does not show on your list of saved workspaces when you use the *)LIB* command. The public libraries permit you to save a workspace so that other people can load it without knowing your private user number. You should consult the operator if you believe you need to save into a public library.

A workspace can be saved into a public library only if the following conditions are met:

1. The number of workspaces now saved on your behalf (either in your private library or in a public library) is less than your quota of saved workspaces.

2. The name proposed for this saved workspace is not already in use in this public library.

3. The operator has established a Common Library number for the use of your group.

The only person who can drop or revise a workspace in a public library is the person who originally stored it there.

Loading a Workspace
From the Private Library of Another User

You can load a workspace from the private library of another user only if he has told you his private library

number, and the name of the workspace. Private library numbers are regarded as confidential, and can not be obtained from the system or from the operator. Notice that even if he permits you to load a workspace from his library, this merely gives you a copy of his workspace, while leaving his saved version inviolate. If you wish, you may lock a saved workspace against unauthorized use (see Appendix A, p. 207).

Dropping a Workspace From Your Library

You can drop a saved workspace from your library by the system command)*DROP* followed by the workspace name. The system confirms that the workspace has been dropped:

```
     )DROP ACCOUNT
11.01.13 07/30/68
```

When a workspace is dropped, it is removed from the magnetic disc file of saved workspaces. This has no effect at all on whatever is in your active area. Note also that you do not need to have the dropped workspace in your active area at the time it is dropped.

Automatic Saving:
Signing Off with "Continue"

The system command)*CONTINUE* causes your currently active workspace to be saved under the name CONTINUE. Any earlier version of CONTINUE is replaced by the the newly saved workspace, even though the active workspace had a different name. Then you are signed off.

The workspace named CONTINUE is intended primarily as temporary storage until the next time you sign on. It doesn't count as part of your quota of saved workspaces. It is the only workspace in your library which is not protected against replacement by a workspace that formerly had some other name. Any time you give the command)*CONTINUE* or)*SAVE CONTINUE*, the earlier version of CONTINUE is replaced by whatever is in your currently active workspace.

If the last time you signed off you signed off with)*CONTINUE*, or if your work was terminated by an automatic continue (see next paragraph), the next time you sign on, the workspace named CONTINUE will be loaded into your active area automatically, without your having to make a specific

request for it. However, there is an exception. It is
possible to lock a workspace so that it can be loaded only
if an appropriate key accompanies the load command. If you
saved CONTINUE with a lock, or if the workspace that was
active when an automatic save was executed on your behalf
was locked, then CONTINUE will have the same lock, and won t
be loaded until you make the appropriate request.

Automatic Continue When a Connection is Broken

If the computer notices that your telephone connection
has been broken but you haven't signed off, it automatically
executes a)CONTINUE on your behalf. This happens if trouble
with the telephone lines accidentally disconnects you. It
also happens if you simply turn off the power or hang up
your telephone without signing off, or if for some reason
the computer operator has to disconnect your line. However,
if you haven't yet loaded or saved a workspace, and haven't
yet stored anything in the clear workspace you got when you
signed on, the computer does not save CONTINUE.

A new version of CONTINUE is saved any time the
computer detects that your line has been disconnected after
you have done some savable work. This protects you against a
loss of whatever is in your active workspace when you are
disconnected.

If you were working in a locked workspace, its
lock now applies to the saved workspace named CONTINUE as
well.

It is not usually possible for the computer to execute
a)CONTINUE if the System is halted by a malfunction of the
computer itself. In that case, the contents of your active
workspace is lost.

Diagram Summarizing Information Flow
Between You, Your Active Workspace, and Saved Workspaces

     The diagram on the next page represents the flow of information between you and the computer. It summarizes the following points:

1. You can see or use only programs or data that are in your currently active workspace.

2. You can save only what is in your currently active workspace.

3. You can save only into your own library (except for the special case in which you save into a common library).

4. You can load into your active area from your own library, from a common library, or (provided you have the necessary information) from the library of another user.

APL\360 DATA FLOW AND MEMORY STRUCTURE
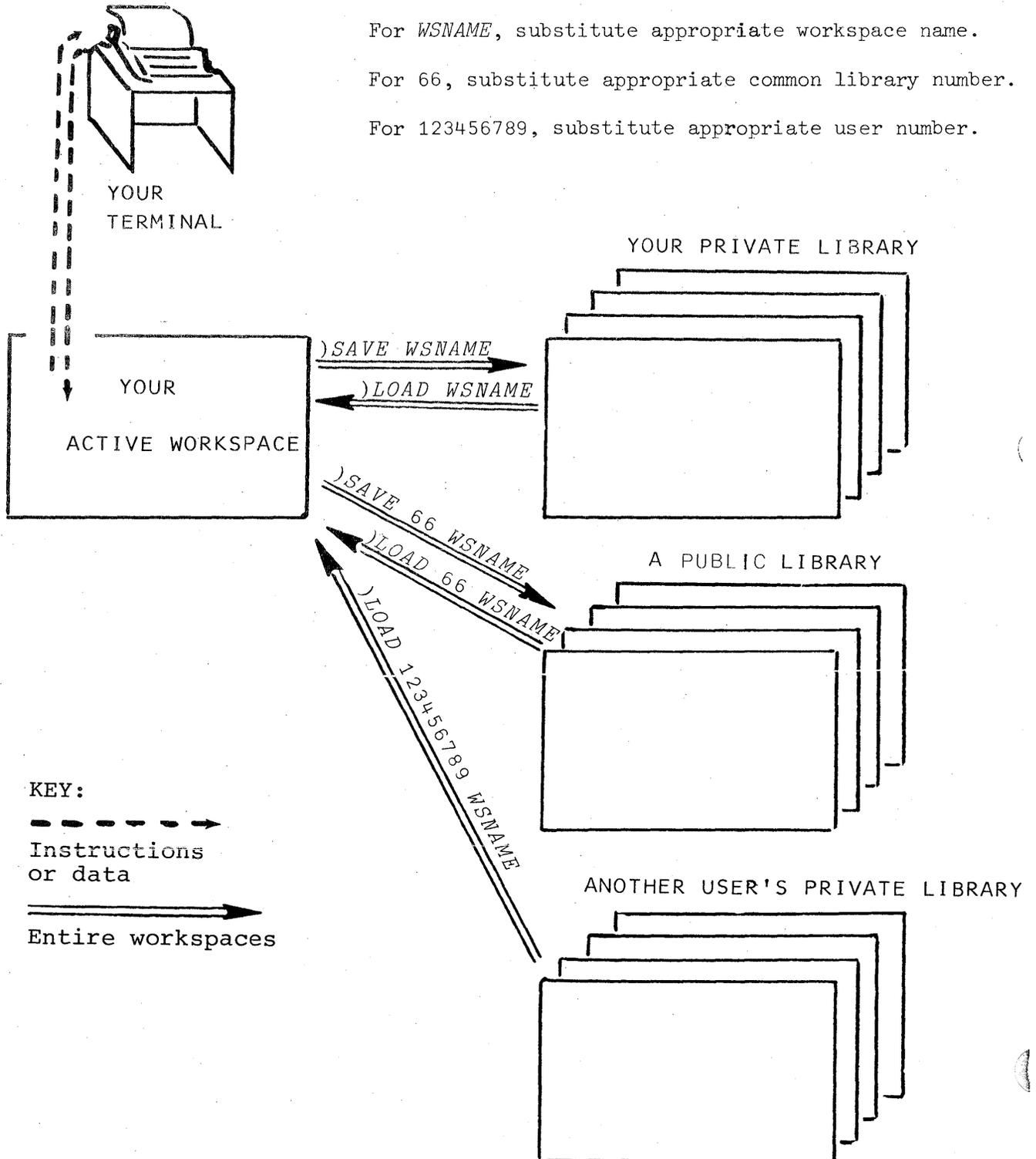
For *WSNAME*, substitute appropriate workspace name.

For 66, substitute appropriate common library number.

For 123456789, substitute appropriate user number.

YOUR
TERMINAL

YOUR

ACTIVE WORKSPACE

)SAVE WSNAME

)LOAD WSNAME

)SAVE 66 WSNAME

)LOAD 66 WSNAME

)LOAD 123456789 WSNAME

YOUR PRIVATE LIBRARY

A PUBLIC LIBRARY

ANOTHER USER'S PRIVATE LIBRARY

KEY:

Instructions
or data

Entire workspaces

Clearing the Active Workspace

If, after doing some work, you wish to start over with a new, blank workspace (such as you get each time you sign on), you enter the command

)CLEAR

The entire contents of the active area is replaced by a clear workspace.

Sending a Message to the Operator

There are two commands which permit you to send a brief message to the APL\360 System Operator. One of these sends a message and leaves your keyboard locked so that your terminal is ready to accept the reply. The other presumes that there isn't going to be a reply, and unlocks your keyboard as soon as the message is delivered. These two commands are)OPR and)OPRN. The letter N stands for "no reply."

Sending a message to the operator and signing on are the only things you can do before the system has accepted your sign-on.

Suppose you want to send the operator a message which says, "Why does computer say 'Not saved' when I try to save my workspaced ACCT?" That message calls for a reply, so you enter:

)OPR WHY DOES IT SAY 'NOT SAVED' WHEN I TYPE)SAVE ACCT ?

The operator receives your message, preceded by a number showing the line you are using, and also the symbol R, which indicates that you are awaiting a reply. When your message is delivered, the computer types "OK" at your terminal, but doesn't unlock the keyboard. When the operator sends you a reply, it is typed with the letters "OPR:" in front of it:

OPR: ACTIVE WS HAS DIFFERENT NAME AND NAME 'ACCT' ALREADY IN USE

After a reply has been received, your terminal again indents and unlocks, ready for whatever you want to enter next.

Suppose now you wish to send a message to the operator which requires no reply. This one you enter with the)OPRN command:

)OPRN THAT WAS IT! THANKS

When it is delivered, the operator will receive the message without the symbol *R*, and will know that you aren't awaiting a reply. At the same time, your terminal will again indent and unlock.

If your terminal remains locked because a message hasn't been delivered, or because you haven't received a reply, you can unlock it by pressing ATTN. That will cancel a message if it hasn't yet been delivered.

Broadcast Messages

The system operator has the power to send out broadcast messages to all users when some special situation warrants. Such a "public address" message appears at the terminal preceded by the letters *PA!*.

16: VECTORS:
PARALLEL PROCESSING OF THE ELEMENTS OF ARRAYS

In science or business, calculations frequently involve not just one number but a whole array of them. APL gets much of its power and simplicity from its approach to the processing of arrays.

1. A single name can stand for an entire array of values.

2. The basic operations which apply to single values can be applied with equal ease to the processing of entire arrays.

A one-dimensional array is called a vector, and a two-dimensional array is called a matrix, while arrays of higher dimensions do not have common names. APL\360 can handle arrays of numbers (or of literal characters) having almost any number of dimensions, but in this primer the discussion is generally limited to vectors.

A vector is one-dimensional in the following sense: the various numbers or characters that make up its elements are arranged in a single chain. Any element can be identified by its position in the chain. Since a vector has only one dimension (its length) a single index-number suffices to identify any element within it, by specifying how far along from the beginning that element is located.

Entering a Vector of Numbers

If you enter

$A \leftarrow 1$ 2.5 7 11

A is specified to be a vector of four numbers. Each of those numbers is an element of the vector A. As you enter the four numbers, you have to type them with at least one space between them. Whenever you enter numbers separated solely by spaces (that is, with no operator sign between them) they are assumed to be consecutive elements of a vector. This applies only to numbers; you can't do it with variables.

Notice that you don't have to say in advance that there are going to be four elements in the vector called A, or even that A is going to be a vector. The computer notices

that you have entered four values for A, and automatically makes A a four-element vector.

If you ask to see what has been stored under the name A, the computer responds by typing all of the elements, like this:

```
      A
1   2.5   7   11
```

Parallel Processing of Vectors

If A is a vector of four numbers, and B is another vector which also consists of four numbers, then the instruction $A+B$ causes the computer to add the first number in A to the first number in B, and the second number in A to the second number in B, and so on. Four separate additions are performed, and so the result is also a vector of four numbers. The four additions are done in parallel fashion; as far as you can see, the answers to the four separate problems are obtained simultaneously.

```
      B←10   20   30   40
      A+B
11   22.5   37   51
```

The same sort of element-by-element parallel processing can be obtained with any of the other arithmetic operators. For instance :

```
      3   5   9 * 4   3   2
81   125   81
```

```
      B⌈19   20   21   22
19   20   30   40
```

```
      A×B
10   50   210   440
```

```
      ÷A
1   0.4   0.1428571429   0.09090909091
```

```
      ⌊A
1   2   7   11
```

```
      ⊛B
2.302585093   2.995732274   3.401197382   3.688879454
```

Using Parallel Processing
In Some of the Problems Introduced Earlier

On page 34, we presented a short program to calculate and print F, the focal length of a lens, as a function of the following variables:

N, the refractive index of the glass
T, the thickness of the lens
R1 and R2, the radii of curvature

The example on page 36 shows how this program calculates a value for F, provided that the values of the variables N, T, R1 and R2 are already specified in the workspace. That very same program, without any change, can just as well calculate any number of F's in parallel, provided now that N, T, R1, and R2 are arrays of the same size. For instance, here is an example in which N, T, R1 and R2 are five-element vectors. Because those variables are five-element vectors, the five focal lengths are calculated simultaneously, as another five-element vector. (This same program could just as well handle vectors of any length, or matrices, or even multi-dimensional arrays, if that's what you should need.)

```
N  ←   1.32   1.32   1.32   1.32
T  ←    .65    .65    .65    .65
R1←    8.1    8.2    8.3    8.4
R2←    7.29   7.38   7.47   7.56
FOCAL
12.1141666   12.26217727   12.41018832   12.55819974
```

In similar fashion, the DIESEL program can process any number of efficiency problems at once, provided the necessary input variables are vectors of compatible length. Here's a sample showing three done at once:

```
R  ←     8.5      9.7     10.9
V3←     22.8      25       32
V2←    140       143      145
GAMMA←    1.35     1.38     1.42
DIESEL
0.61772102   0.6630320875   0.7074198645
```

Vectors Must Have Matching Lengths

In the last paragraph, we remarked that the various vectors must be of compatible length. If you enter an

instruction such as the following:

```
1   2   3 + 17   18   19   20
```

the computer finds 3 elements in the first vector, and 4 elements in the second one. Which element is supposed to be matched with which? The problem is ambiguous. The computer cannot proceed, so it types the following error message:

```
      1   2   3 + 17   18   19   20
LENGTH ERROR
      1 2 3 + 17 18 19 20
            ^
```

Generally speaking, whenever an operation is to be performed on two vectors, the vectors must have the same length (i.e. the same number of elements).

Extending a Single Number
To Match the Length of a Vector

Ordinarily, when an operation is performed on two vectors, they have to be of the same length. But there is one important exception to this rule. The exception is when one of the operands is a vector but the other one is a single number. Whenever a single number enters into an arithmetic operation with a vector, the single number is extended to match the length of the vector. For instance, if you enter

```
1   3   5   7   9 + 2
```

the computer finds that one argument of the addition is a vector of five elements, while the other argument is the single number 2. It treats the instruction as if it were

```
1   3   5   7   9 + 2   2   2   2   2
```

In effect, the single number 2 is replicated until it is a vector of the same length as the other vector. Here are some examples of operations involving a vector and a single number.

Take the square roots of nine numbers simultaneously:

```
      2   4   9   16   25   36   49   64   83 * 0.5
1.414213562   2   3   4   5   6   7   8   9.110433579
```

Convert four angles in degrees to radians:

```
      1   15   22.5   45 × 2×PI÷360
0.01745329252   0.2617993878   0.3926990817   0.7853981634
```

Is it true that some single number stored under the name C
is divisible by each of four prime integers?

```
      C←20937
      0=3 5 7 11|C
1  0  1  0
```

The single number B divided by each of the four elements of
the vector H:

```
      B←28
      H←0.014  9E⁻11  3.5  0.0357142857142857
      B÷H
2000  3.111111111E11  8  784
```

Two raised to each of the powers 0 through 12:

```
      2*0 1 2 3 4 5 6 7 8 9 10 11 12
1  2  4  8  16  32  64  128  256  512  1024  2048  4096
```

The largest integer whose square is less than or equal to
10, 20, 30, 40, 50, 60, 70, 80, 90, and 100:

```
      ⌊10 20 30 40 50 60 70 80 90 100*.5
3  4  5  6  7  7  8  8  9  10
```

The frequencies of all the semitones in the octave starting
with middle C (whose frequency is 262), rounded to the
nearest integer:

```
      ⌊0.5+ 262 × 2 * 0 1 2 3 4 5 6 7 8 9 10 11 12 ÷12
262  278  294  312  330  350  371  393  416  441  467  495  524
```

To make quite clear the way the computer extends a
single number to match the dimensions of an array, it may be
useful to examine that last example in detail. The first
operation to be executed is, as usual, the rightmost one. It
is a division. The dividend is a vector of thirteen
elements, the numbers 0 through 12. The divisor is the
single number 12. So the computer replicates the number 12
until there are thirteen of them. Then it executes the
thirteen divisions, producing thirteen quotients.

The next operation is exponentiation. The left argument is the single number 2, but the right argument is the 13-element vector of quotients arising from the division. So the 2 is replicated until it is also a 13-element vector of 2s, and then the thirteen exponentiations are executed, producing thirteen results.

The next operation is multiplication. The left argument is the single number 262, while the right argument is the vector of thirteen results from exponentiation. Once again the single number is replicated to match the length of the vector, the thirteen multiplications are performed, and thirteen products found.

The next operation is addition. Its left argument is the single number .5, and its right argument is the vector of products resulting from the last operation. As before, this produces a vector of thirteen results.

The computer reaches the last operation: it must take the floor of the thirteen results coming from the addition. Once these thirteen integers are found, there is no further instruction telling what to do with them, so the computer prints them.

Parallel Processing Requires All the Elements To Be Treated in the Same Way

We've mentioned two programs that were originally written to work on single numbers, but which turn out to work just as well on vectors of numbers. This depends on the fact that each of the elements in those vectors was treated in the same way. It isn't always obvious how this can be done.

In the program to find quadratic roots, the first step was to find a value for the discriminant (b^2-4ac) and store it under the name DISC. But after that the program did either of two different things, depending upon whether DISC was found to be positive or negative. If you were to give this program a whole vector to work on at once, DISC would be a vector. Some of its elements might be positive and others negative. They would generate a whole vector of line numbers to which the program should branch. However, it isn't possible to branch to several different places at once, and therefore the program would not produce the results you want.

To use parallel processing of vectors, you have to have a procedure that can be applied uniformly to all of the elements in a vector. Even if some elements of the vector must be treated in one way and others in another, it is often possible to devise a single procedure which has that effect. In the next paragraph, the same problem is handled first by branching, and then by a formula that applies a single procedure to all the elements of a vector.

Adjusting a Formula To Facilitate Work with Vectors

During 1967, New York State gave the following formula for income tax on a weekly paycheck. P is the pay; E is $13 for each exemption. (Although the State didn't say so, presumably the tax is rounded to the nearest penny, and is never negative, even if you have little pay and many exemptions.)

Net Income $350 or less: Tax=(.018+.000105(P-E))(P-E)-.48

Net Income Exceeds $350: Tax=.09(P-E)-12.80

If you need a program to handle only one person's tax, you could write it with a branch. (To simplify rounding, this program treats income in pennies rather than in dollars.) In this program, EX is the number of exemptions, PAY is the amount paid, TI is the taxable income, and T is the amount of tax owed.

```
      ∇ TAX
[1]     TI←PAY-1300×EX
[2]     →(TI>35000)/OVER
[3]     T←0⌈⌊0.5+ ̄48+TI×0.018+TI×1.05E ̄6
[4]     →0
[5]     OVER: T←0⌈⌊0.5+ ̄1280+TI×0.09
      ∇
```

This program has two separate instructions (line 3 and line 5), only one of which is executed in any use of the program. The branch instruction at line 2 decides for any single instance which of them will be executed, 3 or 5.

To take advantage of vector operations, you need a single formula which works for any execution, so that no branch is necessary. Suppose you calculate the tax rate by multiplying the alternative rates by 0 or 1, depending on whether the taxable income is or is not over $350. This is

done in the program called TAXES. The variables EX, TI, and
PAY have the same meaning as before. HI has the value 1 for
a person who is in the high income bracket (taxable income
over $350.00), and 0 otherwise. LO is the negation of HI.

```
      ∇ TAXES
[1]    LO←~HI←35000<TI←PAY-1300×EX
[2]    T←0⌈⌊0.5+(HI×¯1280+TI×0.09)+LO×¯48+TI×0.018×1.05E¯6
      ∇
```

Line 1 of TAXES may be read this way: LO is the
negation of HI, which is the truth of 35000 is less than TI,
which is PAY minus 1300 times EX.

Line 2 computes the tax as the sum of two quantities.
The one within parentheses will always have the value 0
whenever income is low, since the values are all multiplied
by HI, and that will be 0 for all persons who are not in the
high income bracket. The other quantity comes from the
expression to the right of the parentheses. Here the values
will always be 0 for anyone who is not in the low bracket.
When the values are added together, for any individual, the
component that isn't multiplied by 0 should be his correct
tax.

Whenever they are asked to calculate the tax for only
one person, these two programs give the same answer. But
when they are asked to calculate a whole vector of taxes,
the first program, called TAX, will have to decide its
branch solely on the basis of the first element of those
vectors. This may be inappropriate for the other elements,
and so answers other than the first may be wrong. The second
program, called TAXES, does not involve a branch, and can be
applied correctly to arrays of any size.

A Vector in a Branch Instruction

Whenever the value to the right of a right-pointing
arrow is a vector, the computer branches to the value of the
first element of the vector, and ignores the rest.

17: "REDUCING" A VECTOR:
APPLYING THE SAME OPERATION TO ALL THE ELEMENTS

It is often useful to have the sum of all the elements in a vector, or the product of all of them, or the maximum of all of them, and so on. APL has a simple procedure for applying the same operation cumulatively to all the elements of a vector. This operation is called "reduction." It reduces a vector of numbers down to a single number that represents their sum, their product, their maximum, and so on, as the case may be.

Summation

In conventional notation, the capital sigma (Greek for S) means that you are to take the sum of the specified members of an array. To sum them all, you have to specify that the summing starts with the first element and then goes on summing the consecutive elements until it gets to the last one. You write it like this:

$$\sum_{i=1}^{n} A_i$$

In APL, you can sum all the elements of a vector called A (regardless of how few or how many elements A has) by typing:

    +/A

The / sign means that the operation on the left of it is to be applied to all the elements along the last dimension of the array on the right. Since vectors have only one dimension anyway, this means summing all the elements. Thus, if A is a vector, like this:

    A←1 2 3 4 5 6 7 8 9 10 11 12 13 14 15

then +/A is executed by the computer as if it were

    1+2+3+4+5+6+7+8+9+10+11+12+13+14+15

When you find the sum of the elements of A, your dialogue with the computer goes like this:

    +/A

120

Notice that you don't have to tell the computer the dimensions of A. It reduces the last dimension by applying the operation all the way along that dimension; when the elements of A are arranged in one dimension, in a vector, the computer finds the sum of all of them, for however many there are.

In speaking, $+/A$ is read as "plus reducing A," or "plus over A," or simply "the sum of A."

Product

In conventional notation, the product of all of the elements in a vector is written with the Greek letter pi (P for product):

$$\prod_{i=1}^{n} A_i$$

In APL, you get the product of all of the elements of a vector called A by entering:

$\times/A$

That is read as "times over A," or "the product of A." If A has values as follows:

$A\leftarrow1$ 2 3 4 5 6 7 8 9 10 11 12 13 14 15

then the computer treats $\times/A$ as if it were:

$1\times2\times3\times4\times5\times6\times7\times8\times9\times10\times11\times12\times13\times14\times15$

Here's how you get the product of all the elements of A:

$\times/A$
$1.307674368E12$

Maximum Reduction: Looking for the Largest

To select the single largest element in a vector, you reduce the vector by the maximum operator, like this

$\lceil/A$

If BALDUE is the vector of the balances due for all of the customers of a store,

BALDUE←62.15   127   4.42   18.65 814.10 76.42 118.50 6.01

⌈/BALDUE gives the amount owed by the customer who has the biggest bill.

      ⌈/BALDUE
814.1

Minimum Reduction: Looking for the Smallest

In similar fashion, ⌊/VECTOR selects the (algebraically) smallest element from a vector. For instance, if ROOT1 contains the vector of all of the first roots of a set of equations,

ROOT1← 0.4815 ¯.085236 16.442 0.000625 ¯4 3.17215

⌊/ROOT1 selects whichever value is the smallest.

      ⌊/ROOT1
¯4

OR Reduction: Looking for "Any"

Suppose you need to know whether a particular value exists anywhere in a long vector. Suppose, for instance, you want to know if any element of the vector V is equal to the single number Q. If you type

      V=Q

you will have a vector of zeroes and ones indicating for every element of V whether or not it is equal to Q. You don't want to examine all those zeroes and ones: you want to reduce them to a single result, either 1 or 0, by applying the logical OR operation so that it puts an OR between each of the elements:

      0∨0∨0∨0∨0∨0∨1∨0∨0∨0∨0∨0∨0∨1∨0∨ ...∨0∨0∨0∨1∨0∨0∨0

Thus the instruction that you need is typed like this:

      ∨/V=Q

The result is 1 if there is a 1 anywhere in that vector, and 0 if and only if every element is 0.

Suppose N is a vector of integers. You want to know if any of them is a perfect square. If an element of N is a perfect square, then its square root is an integer. In that case, rounding the root off to the nearest integer won't make any difference. Then, if you square the rounded-off roots of N, wherever an element of N was a perfect square, you should be able to get back to the original value of N. The following expression tests to see if that condition is met for any elements of N:

    N←103 117 142 121 135 176 149 169 128 156 118 124 133
    ∨/N=($\lfloor N$*.5)*2
1

And if you need to know not just whether any of them are perfect squares, but how many, you can find that by reducing the expression N=($\lfloor N$*.5)*2 by plus instead of OR:

    +/N=($\lfloor N$*.5)*2
2

AND Reduction: Looking for "All"

Suppose you want to know if every one of a set of equations has real roots. The vector of discriminants for these equations has been stored as the variable DISC. Then

    $DISC$>0

is a vector of zeroes and ones, indicating for each element of DISC whether it is true that DISC is equal to or greater than 0. The operation AND placed between every element will return the result 1 if and only if every element is 1, and otherwise 0. Thus to find out if the test is true for every element of DISC, you enter:

    ∧/$DISC$≥0

Suppose you have a vector called KEY, and another vector called LOCK. Both vectors have the same length. You need to know whether every element of KEY is equal to the corresponding element of LOCK:

    ∧/KEY=$LOCK$

```
     KEY←1.01 1.763 1.808 1.2346 1.2272 1.8095 1.1
     LOCK←1.01 1.763 1.898 1.2346 1.2272 1.8095 1.1
     ∧/KEY=LOCK
0
```

Evidently at least one of the elements of KEY does not match an element of LOCK.

<u>Example Using the Sum of Products:</u>
<u>Price Times Quantity Ordered</u>

Suppose that PRICE is a variable which contains the price list for the various items sold by a store, and Q1 and Q2 are vectors indicating the quantities of the various items ordered by Customer 1 and Customer 2. Then the total bill for Customer 1 is the sum of the product of PRICE and Q1, while the total bill for Customer 2 is the sum of the product of PRICE and Q2.

```
     PRICE ←  .66   1.40   27.10   2.39   14.00  7.60   8.45  2.80
     Q1    ←  0     0      2       1      0      0      0     0
     Q2    ←  12    7      0       5      0      0      0     10
     +/Q1×PRICE
56.59
     +/Q2×PRICE
57.67
```

<u>The Area Under a Curve</u>

One simple approach to finding the area under a curve is to divide it into a great many small trapezoids and then find the sum of the areas of all of them. Suppose you want to find the area under the curve produced by some function F of X for all the values of X between 0 and 1. You might get a suitably fine division by splitting that interval into 100 parts. Counting both end points, that makes 101 values. Suppose now that you have stored under the name FX the vector of the 101 values of F of X as X varies from 0.00 up to 1.00 in steps of .01. The area of any one of the trapezoids is the average of the two values of FX that bound it, times the width of the interval, which is .01. You don't actually have to average all those adjacent pairs; you can get the same effect by simply using FX times the width, provided that you first divide the first and last elements of FX by 2. Suppose that D is a vector whose first and last elements are 2, with 99 ones in between. Then you get the area under the curve by the instruction:

```
     AREA ← +/FX×WIDTH÷D
```

This page intentionally left blank.

18: GENERATING ARRAYS
AND FINDING THEIR DIMENSIONS

As you have seen, in APL, arithmetic operations apply not only to single numbers, but also to entire arrays. Array-handling requires a variety of manipulations for which conventional arithmetic makes no provision. Therefore, in addition to the arithmetic operators (most but not all of which have been introduced) APL includes several other operators specifically designed for manipulating arrays. They can generate an array of a given size and structure, tell you the size of an array, pick out certain elements from an array, find where particular elements are located within an array, selectively throw away some elements and keep others, and so on.

Generating an Array by Reshaping

In order to build an array, you have to specify two things:

1. The shape that the array is to have: the number of dimensions, and the length of each.

2. The values that are to be assigned to each of the elements of the new array.

The APL operator which reshapes an array (or a single element) to form a new array with a new shape is rho, the Greek form of the letter R, which looks like this: ρ. The reshaping operator ρ is dyadic. The left argument determines the shape of the resulting array, and the right argument provides the values for the various elements. As the left argument of ρ you enter one number for each of the dimensions to be generated, indicating the length that that dimension is to have. Since for the time being we are limiting the discussion to vectors, which are one-dimensional arrays, in the examples ρ will have only one number as its left argument.

The values of the elements of the new array are taken from whatever values appear as the right argument of ρ. The instruction

$7 \rho A$

means that a seven-element vector is to be generated. Its seven values are to be supplied from whatever values are

found stored under the name A. It doesn't matter whether A
is an array, or what structure A has--just so long as it has
at least one value that can be used in the new array. If A
has more than seven elements, just the first seven are
taken. If A has fewer than seven elements, its elements are
repeated as often as needed to provide seven entries in the
new vector. The following examples may make this clear:

```
      7ρ1  2  3
1  2  3  1  2  3  1

      2ρ1  2  3
1  2

      10ρ1.3
1.3   1.3   1.3   1.3   1.3   1.3   1.3   1.3   1.3   1.3
```

Vectors of Literal Characters

On page 77, we mentioned that the value of a variable
can be quoted alphabetic letters (or numerals, or any sign
from the keyboard). Although no mention of it was made at
the time, a quotation with several letters in it is in fact
an array. Just as a one-dimensional array of numbers is a
numerical vector, so a one-dimensional array of literal
characters is a literal vector. Each element of a literal
vector is a single literal character. When the computer
prints this sort of a vector, the elements are typed without
any extra space between them. They are typed without
additional spaces for two reasons:

1. Since an element of an array of literals can only
   contain one literal character, there isn't any
   need to insert spaces to distinguish where one
   ends and the next begins.

2. Spaces which occur as part of a quotation are
   characters, just like any other character that can
   be entered from the keyboard.

The reshaping operator, ρ, can also be used to generate
vectors of literal characters. For instance:

```
      6ρ'A'
AAAAAA

      15ρ'THE QUICK BROWN FOX JUMPS OVER THE LAZY DOG'
THE QUICK BROWN
```

A vector of literal characters can include numerals or operator signs as well as alphabetic letters.

```
      60ρ'→← * →← ⎕ '
→← * →← ⎕ →← * →← ⎕ →← * →← ⎕ →← * →← ⎕ →← * →← ⎕ →← * →← ⎕
```

An Array Can Have Zero Length

When you use ρ to generate a vector, the number of elements you ask for (i.e. the length of the vector) can be anything you want, provided that it isn't negative, isn't fractional, and doesn't generate a vector so large that there's no room for it in the workspace. In particular, it's all right to have a vector of length zero. This vector has no elements at all. If you ask to see such a vector printed, there is nothing to see. The computer prints a line on which nothing is written: a blank line.

A quotation which has no letters in it is sometimes useful when you want a program to insert a blank line between some portions of the typed output. You can get a blank line typed if one instruction in your program simply calls for the printing of a vector of zero letters (or zero numbers, for that matter). Typing two quote marks side by side, with nothing between them, will also generate a vector of zero length. The following expressions are equivalent:

```
   0ρ' '      0ρ'A'      0ρ0      0ρA      0ρ1E60
```

Generating Consecutive Integers

The operator ι is called iota, which is the Greek form of the letter I. Like most other APL operators, ι has both a dyadic and a monadic use. The dyadic (i.e. two-argument) use is explained in Chapter 20. The monadic use of iota (i.e. with a right argument but no left argument) generates consecutive integers. The right argument must be a single positive integer.

```
      ι12
1  2  3  4  5  6  7  8  9  10  11  12
```

One way to think of this use of iota is to say that it generates all the index numbers for a vector of a given length. Index numbers are always consecutive integers. The first element of a vector is element number 1, and similarly the consecutive integers generated by iota also start with

1. (The point is worth noting because, as is mentioned in the Appendix, p. 207, it is also possible to have index numbers which start at 0, and in that case iota generates integers which start at 0 also.)

```
      2.5×ι12
2.5   5   7.5   10   12.5   15   17.5   20   22.5   25   27.5   30

      5-ι10
4   3   2   1   0   ‾1   ‾2   ‾3   ‾4   ‾5

      ÷ι5
1   0.5   0.3333333333   0.25   0.2
```

The index-generating iota is very handy when you want to refer to a consecutive block of numbers. You could get the first 35 powers of 2 simply by typing this instruction:

```
      2*ι35
2   4   8   16   32   64   128   256   512   1024   2048   4096
      8192   16384   32768   65536   131072   262144   524288
      1048576   2097152   4194304   8388608   16777216
      33554432   67108864   134217728   268435456   536870912
      1073741824   2147483648   4294967296   8589934592
      1.717986918E10   3.435973837E10
```

(Note that when the computer finds that a vector is too long to fit on a line, it continues it on the next line for as many lines as it needs, but it indents the continuation lines to show that they're still part of the same vector. The width used for typing results may be adjusted; see Appendix, p. 208.)

Finding Out the Shape of an Array

If you use ρ with no left argument, it no longer means that an array should be generated. Now it asks the computer to report on the shape of the array that is the right argument of ρ. If A is a vector with eight elements, and you ask for ρA, the computer responds by typing one number (because A is a one-dimensional array). The one number it types is 8, which is the length of A's one dimension.

```
      A←186   17   .00165   3.14159   1.26E45   3   2E‾9   .00001
      ρA
8
      B←0ρA
      ρB
0
```

```
      ρ 1 3 5 7 9 11
6
      ρ'1 3 5 7 9 11'
12
```

There are many useful expressions which use ι and ρ together. Suppose you'd like to have a vector of consecutive integers which matches the length of another vector called A. A is a vector with 13 elements. You can get the correct number of consecutive integers by entering this instruction:

```
      ιρA
1  2  3  4  5  6  7  8  9  10  11  12  13
```

Perhaps you'd rather have the integers run backwards to zero. The place-values for the successive columns in the representation of a number are found by raising the base of the number system to the 0th power for the last column, the first power for the next column, and so on. For base 10, the values of the first 6 columns would be found like this:

```
      10*6-ι6
100000  10000  1000  100  10  1
```

And in base 8 they'd be:

```
      8*6-ι6
32768  4096  512  64  8  1
```

Suppose you want integers that depart from 500 in steps of 8. You enter:

```
      500+8×ι6
508  516  524  532  540  548
```

The expression ιN always results in a vector of length N.

```
      ρι1000
1000
      ρι5
5
      ρι0
0
```

Thus, still another way to get an empty vector is to enter the instruction:

        ι 0

                    ◄(Here the computer prints a blank line)

What Is the Shape of a Single Number?

The answer to this question depends upon whether the single number is an array or not. Suppose you generate an array which has one dimension, and the length of that dimension is 1. When you ask for the length of that array, the answer will be 1:

        A←1ρ5
        ρA
1

By contrast, if you simply store a single number under the name A, without involving any of the operations that generate arrays, then A is not an array. Like a point in geometry, which is presumed to have no length, breadth or height, a single number or literal character, unless produced by some array-generating operation, has no dimensions, and is called a scalar. If you ask for its length, the length is neither 1 nor 0: length just isn't an attribute of a scalar.

When you ask for the shape of a scalar, the result is itself an empty vector (a vector with no elements).

        A←5
        ρA
                    ◄(Here the computer prints a blank line)

When you use the shape operator rho to find the dimensions of a variable, the result is always a vector. This vector has one element for each dimension of the variable you asked about. If you ask for the shape of a three dimensional array--a topic we're not otherwise mentioning in this primer--you get a vector of three elements, one element for each dimension of the array. If you ask for the shape of a vector, you get back a vector of one element, for the one dimension (length) of the vector. If you ask for the shape of a scalar, which has no dimensions, the result is a vector of no elements: an empty vector.

Another Example Using Parallel Processing of Vectors: The Correlation Coefficient

The correlation coefficient is the average product of two vectors of scores. The average of the elements of a vector V is readily found by the expression:

$$(+/V) \div \rho V$$

And the average of the product of the vectors X and Y is:

$$(+/X \times Y) \div \rho X$$

However, this simple definition requires that the vectors X and Y be in "standard" form. Scores are standardized if they are arranged so that their average is zero and their standard deviation is 1. Since scores are seldom found already standardized, the first step is to standardize them, by reducing each score by the mean of its group, and then dividing each score by the standard deviation of the group. The steps needed to calculate the correlation between two vectors of scores called X and Y are therefore as follows:

1. From X and Y, subtract their respective means.

2. Divide X and divide Y by their respective standard deviations. Once the means have been subtracted, the standard deviation is the square root of the average of the squares.

3. Find the correlation coefficient as the average product of the standardized scores.

The program called CORR (next page) presumes that the scores are already stored in X and Y, and that X and Y are vectors of the same length. The standard deviations are stored under the names SDX and SDY, and the standard scores are stored under the names XX and YY. The correlation coefficient is given the name R. Once XX and YY have been set up, the key formula appears on line 10 of the definition.

```
      ∇ CORR
[1]   (ρX);' OBSERVATIONS'
[2]   XX←X-MEANX←(+/X)÷ρX
[3]   YY←Y-MEANY←(+/Y)÷ρY
[4]   SDX←((+/XX*2)÷ρX)*0.5
[5]   SDY←((+/YY*2)÷ρY)*0.5
[6]   XX←XX÷SDX
[7]   YY←YY÷SDY
[8]   'X: MEAN ';MEANX;'  STANDARD DEVIATION ';SDX
[9]   'Y: MEAN ';MEANY;'  STANDARD DEVIATION ';SDY
[10]  'CORRELATION ';R←(+/XX×YY)÷ρX
      ∇
```

Here is a sample execution of CORR. The values of X and Y are taken from an illustration involving the reciprocity of affection among "steady" couples (S. Diamond, Information and Error, Basic Books, 1959, p. 167).

```
      X←2 8 7 5 4 4 3 2 5 6 7 3
      Y←5 6 5 5 6 3 4 3 3 6 7 2


      CORR
12 OBSERVATIONS
X: MEAN 4.666666667  STANDARD DEVIATION 1.92930615
Y: MEAN 4.583333333  STANDARD DEVIATION 1.497683396
CORRELATION: 0.6152566398
```

19: SELECTING PARTICULAR ELEMENTS FROM AN ARRAY
BY USING INDEX NUMBERS

Once an array exists, you may want to refer not to the whole thing but just to the elements in certain positions within it. This procedure is called indexing. (Because historically the index values were written in a smaller type face and set below the line, index numbers are often loosely called "subscripts.") In APL, index numbers must be integers; they are enclosed in brackets and written after the array to which they apply.

```
      A←1.11 1.22 1.33 1.44 1.55 1.66 1.77
      A[2]
1.22
      A[3 3 1 5]
1.33  1.33  1.11  1.55

      B←2 4 2 6 1
      A[B]
1.22  1.44  1.22  1.66  1.11

      QQ←'ABCDEFGHIJKLMNOPQRSTUVWXYZ '
      QQ[20 8 5 27 3 1 20 27 19 1 20 27 15 14 27 20 8 5 27 13 1 20]
THE CAT SAT ON THE MAT

      6.11 6.22 6.33 6.44 6.55 6.66 6.77 6.88[6 6 1 3 2]
6.66  6.66  6.11  6.33  6.22
```

If you use an index which refers to an element which doesn't exist in the array, the computer is unable to execute the instruction, and reports an "index error."

```
      A[8]
INDEX ERROR
      A[8]
      ∧
```

Respecifying Certain Elements Within an Array

An indexed variable may also appear on the left of a specification arrow. Then the result on the right is stored in the indicated positions within the array on the left, while the rest of the array on the left remains unchanged.

```
      A[3 1]←7E30 7E10
      A
7E10  1.22  7E30  1.44  1.55  1.66  1.77
```

You can't index an array until after the entire array has been specified. Suppose that no value has been assigned to the name Z. Then an attempt to store some values as particular elements within Z would be an error:

```
      Z[3 4] ← 18 46
VALUE ERROR
      Z[3 4]← 18 46
      ∧
```

The Index Numbers May Result from an Expression

Indices (i.e. whatever is inside the brackets) may include expressions, provided that when those expressions are finally evaluated, they turn out to have values that are valid indices for the vector.

```
      QQ←'ABCDEFGHIJKLMNOPQRSTUVWXYZ '
      X←4 9 16 25 36

      QQ[⌊.5+X÷2]
BEHMR

      QQ[X*0.5]
BCDEF

      QQ[10+X[3 1 2]]
ZNS

      ZED←18 ¯2 31.1 ¯118.4 ¯4.96E27 0.2
      SIGN←'+-'
      SIGN[1+0>ZED]
+-+--+
```

Indexing an Expression

The thing from which elements are selected does not have to be a variable. A constant vector may be indexed:

```
      2 3 5 7 11 13 17 19 [7 2 4 2]
17 3 7 3

      'ABCDEFGHIJKLMNOPQRSTUVWXYZ '[12 15 15 11 27 16 1]
LOOK PA
```

Similarly, an expression may be indexed, provided you enclose it in parentheses:

```
      (- 1 2 3 4 5 *0.5)[2 1 3]
¯1.414213562   ¯1   ¯1.732050808
```

Indexing by An Empty Vector of Indices

A vector of 0 index numbers (i.e. an empty vector inside the brackets) refers to none of the elements of an array, and therefore it produces an empty vector of results. But that is not an illegal operation.

    $A[0\rho 1\ 2\ 3]$
                    ◄(Here the computer prints a blank line)

If selection by indexing is summarized as $R \leftarrow A[X]$, in which A is an array, R is the result, and X represents whatever index numbers are used for the selection, then it is always true that

    $(\rho R) = \rho X$

This means that it is possible to index a vector by a matrix, or indeed by any array all of whose elements are valid indices for the vector. But that goes beyond the scope of this primer.

Indexing a Matrix

Matrices don't get much attention in this primer. Nevertheless, it may be useful to describe how you select particular elements from within a multidimensional array.

Within the brackets, the index numbers for the various dimensions are separated by semicolons. Suppose M is a 3 by 4 matrix of consecutive integers, generated like this:

    $M \leftarrow 3\ 4\rho\iota 12$

If you ask to see the values of M, they are printed in the usual matrix form. Note that the computer prints one blank line before printing a matrix.

    M

```
 1    2    3    4
 5    6    7    8
 9   10   11   12
```

If you want to refer to the element in row two, column three, you enter:

    M[2;3]
7

If you would like the third and fourth elements in that row, you enter:

    M[2;3 4]
7   8

Similarly, if you would like the elements in column four, rows one and two and one, you enter:

    M[1 2 1;4]
4   8   4

You use the same procedure to select a sub-matrix from within M. If you want the matrix of those elements which are on rows two and three and columns one, two, and one of M, you enter:

    M[2 3;1 2 1]

    5   6   5
    9 10   9

Now the result is a two by three matrix.

If you fail to state any value for one or more of the dimensions of the array that is being indexed, the computer assumes that you want all of that dimension. For instance, to get all of row two of M, you enter:

    M[2;]
5   6   7   8

Or to get all of columns four and one, you enter:

    M[;4 1]

    4   1
    8   5
   12   9

Note that you still have to type the semicolon: it's needed to make clear which dimension is which.

20: FINDING THE INDEX NUMBERS THAT LOCATE
PARTICULAR ELEMENTS WITHIN A VECTOR

Suppose that A is a vector which has the following values:

$A \leftarrow 1.2916\ 1.3184\ 1.2196\ 1.1629\ 1.2619\ 1.2961\ 1.1326$

and B has a single value:

$B \leftarrow 1.2619$

Then the instruction

$A \iota B$

means "Where in A can you find the value of B?" This is the dyadic (two-argument) use of iota. The instruction is read as "A iota B" or "the A-index of B."

The computer responds with the index number that shows which element of A has the same value as B:

$A \iota B$
5

If you would like to know where in A its largest value is located, that can be found too:

$A \iota \lceil /A$
2

And the smallest value likewise:

$A \iota \lfloor /A$
7

Finding Several Indices at Once

Suppose instead of being a single number, B is itself an array. In that case the computer will look up the A-index of each of the elements of B in turn. Like this:

$B \leftarrow 1.2619\ 1.2916\ 1.2961$
$A \iota B$
5   1   6

Notice that the result always has in it one element for each
element in the right argument of iota. If the instruction is
in the form $X \leftarrow A \iota B$, then it is always true that

$$(\rho X) = \rho B$$

Indeed, the right argument of iota may be a matrix or a
multi-dimensional array: the result of $A \iota B$ always has the
same shape as B. But this point is not pursued in this
primer.

Indexing Works Just as Well
For Arrays of Literal Characters

Iota can also be used to look up the position in which
a literal character is located. For instance, suppose A and
B are vectors of literal characters as follows:

```
A←'ABCDEFGHIJKLMNOPQRSTUVWXYZ'
B←'CAT'
```

Then the locations in A at which the values of B can be
found can be obtained by the following instruction:

```
      AιB
3  1  20
```

And similarly the index numbers for various other literal
characters can be found:

```
      A←'THE QUICK BROWN FOX JUMPS OVER THE LAZY DOG'
      Q←'CAT SAT ON MAT'
      X←AιQ
      X
8  37  1  4  25  37  1  4  13  15  4  23  37  1
      A[X]
CAT SAT ON MAT

      A[Aι'HELLO IS WHAT A MIRROR SAYS']
HELLO IS WHAT A MIRROR SAYS

      NUM←'1234567890'
      NUMι'1776'
1  7  7  6

      NUMι'1890'
1  8  9  10
      10|NUMι'1890'
1  8  9  0
```

```
      10*3 2 1 0
1000 100 10 1

      +/(10|NUMι'1890') × 10*3 2 1 0
1890
```

To prove that that last result is a number, whereas '1890' is literal, you can try adding 1 to each of them:

```
      1 + '1890'
DOMAIN ERROR
      1+'1890'
       ∧

      1 + +/(10|NUMι'1890') × 10*3 2 1 0
1891
```

Looking for the Index Number
Of a Value that Isn't There

Suppose that one of the values in the right argument of an iota simply isn't represented anywhere in the left argument. What number does the computer return as the index of this nonexistent element?

For a value that isn't represented anywhere in the vector to the left of an iota, the computer responds with the first illegal index for that vector. For instance, suppose that A is a vector of seven elements with the following values:

```
      A←11  12  13  14  22  77  18
```

Then the possible index numbers for this vector are the integers 1, 2, 3, 4, 5, 6, 7. The first "illegal" index for this array is 8. If you ask for the index of a value that isn't anywhere in the vector A, the computer responds by saying that it is at location 8. For example:

```
      Aι77 15
6  8

      'ABCDEFGHIJKLMNOPQRSTUVWXYZ' ι '?'
27

      1 2 3 4 5 ι '5'
6
```

The Index for a Value That Occurs
At Several Locations in the Vector

Suppose you ask where in the vector HIK there is an element with the value 666. And suppose that HIK in fact has three elements with that value. The computer responds by giving you the location of the first occurrence of 666 in HIK. Like this:

$HIK \leftarrow 18$ 66 618 666 627 216 616 666 624 466 424 666
$HIK \iota 666$
4

(You have already seen an example of this, since the vector 'THE QUICK BROWN FOX JUMPS OVER THE LAZY DOG', used on page 132, contains several duplications.)

The computer looks up an index number for every element of the right argument of iota. It does so even if the right argument contains duplicates. For instance, suppose the left argument of iota contains two occurrences of 'X', while the right argument is 'XXX'. What are the index numbers of 'XXX'? The computer will respond with three index numbers, one for each of the elements in 'XXX'. Those three numbers will all be the same: they will all be the index of the first 'X' in the left argument. Thus:

$'SEX EX MACHINA' \iota 'XXX'$
3   3   3

An Example Using Iota to Find Index Numbers:
Evaluating Hexadecimal Representations

The internal work of some computers is performed in base-16 arithmetic, sometimes called hexadecimal arithmetic. The System/360 computer operates this way internally, although of course you don't see that when you're using APL\360. But the people who work closely with such machines have to have some familiarity with the way numbers act and look when they are represented in base 16. When numbers are represented to that base, as usual, the rightmost column is the ones column. But the next column is the 16s column, and the one to the left of that is the 256s, and so on. A problem arises because any one column can contain any of the numbers 0-15: that is, a single column may contain a number which ordinarily would require two places to write. To solve this problem, the digits 10-15 are represented by the letters A through F.

If you're working with hexadecimal arithmetic, you may often find that you'd like to be able to translate a hexadecimal representation back into the more familiar decimal. Suppose, for example, you are faced with the number whose hexadecimal representation is 8A2F. How much is that in decimal?

The index-finding operator gives you an easy way to work this out. You ask the computer where in a standard set of digits the digits 8A2F are to be found. Suppose the standard digits are set up as follows:

$DIGITS\leftarrow$'0123456789ABCDEF'

Then you find the positions of the literals 8A2F this way:

$DIGITS$ ι'8A2F'
9 11 3 16

Evidently '8' is in the 9th position, '2' is in the 3rd position, and so on. They're all off by 1 because 0 is in the first position. That's easily remedied by subtracting 1 from the result. Now you have only to multiply each of those values by the appropriate powers of 16, and sum. The program DH (for "decimal from hex") stores under the name VALUE the value of the hexadecimal number represented by the literal vector HEX. The vector CV on line 1 contains the place value for each column; it doesn't matter how long HEX is, since CV is generated so that it has the same length.

```
    ∇ DH
[1]    CV←16*(ρHEX)-ιρHEX
[2]    VALUE←+/CV×¯1+'0123456789ABCDEF'ιHEX
    ∇

    HEX←'8A2F'
    DH
    VALUE
35375
    HEX←'1001'
    DH
    VALUE
4097
```

APL includes two operators, ι and τ, which convert numbers to their representation in any base, or vice versa. These operators would further simplify the DH program, but they are discussed only in the Appendix, page 195.

This page intentionally left blank.

21: CATENATION:
BUILDING A VECTOR BY CHAINING ITEMS TOGETHER

You can chain together two vectors to make a single vector by using the catenation operator. The symbol for this operation is the comma, placed between the vectors which are to be catenated. The number of elements in the resulting vector is the sum of the number of elements in the two items that are catenated.

```
      18  2.5  3,3  14  1E7
18  2.5  3  3  14  1E7
```

```
      12  13, 13
12  13  13
```

Here's what happens when you catenate two vectors called QS and HT:

```
      QS←1  2  3
      HT←102 105
```

```
      XX←QS,HT
      XX
1  2  3  102  105
```

The things that are to be catenated may be either vectors or scalars (dimensionless single values). For instance:

```
      QS,6.02E23
1  2  3  6.02E23
```

```
      ρQS,6.02E23
4
```

```
      2,HT
2  102  105
```

```
      A←2*0.5
      B←2*÷3
```

```
      A,B
1.414213562  1.25992105
```

```
      A,A
1.414213562  1.414213562
```

```
H←'NOW IS THE TIME '
H,'FOR ALL GOOD MEN'
NOW IS THE TIME FOR ALL GOOD MEN
```

Note that when you want to form a vector from numbers (or literal characters) that you already know (rather than from stored variables) you don't need to use the catenation operator. You can form those vectors simply by typing their values with no operation sign between them. Thus

```
18,2,40
```

has the same effect as

```
18   2   40
```

and similarly

```
'A', 'P', 'P', 'L', 'E'
```

has the same effect as

```
'APPLE'
```

When you enter a numerical vector simply by typing spaces between the successive elements, the machine at once treats those numbers as a single vector. By contrast, if you type commas between the elements, then the commas indicate the operation of catenation, and they are executed according to the usual rules governing order of execution. For instance

```
18 2 10×4
```

means that the vector 18 2 10 is to be multiplied by four, whereas

```
18, 2, 10×4
```

means that the product of 10 and 4 is to be catenated to 2 and then to 18.

A vector must always be either entirely made up of numbers or entirely made up of literal characters. Therefore you can't catenate a number to a literal character. Literals are not in the domain of things that can be catenated to

numbers, and vice versa. If you try it, the computer will respond with an error message as follows:

```
    NUMB← 1 5 9
    LIT← '1 FIVE 9'
    NUMB,LIT
DOMAIN ERROR
    NUMB,LIT
    ∧
```

Building a Vector of Results
By Catenating the Latest Result to the Earlier Ones

Suppose you have a program that works through a series of problems by doing them one at a time. One way of accumulating the answers is to catenate each new result onto the vector of the results previously obtained. If the most recent result is in a variable called LATEST, and all the former ones are in a vector called RESULT, somewhere in the program you want an instruction like this:

```
    RESULT←RESULT,LATEST
```

The very first time that this instruction is executed, there won't be any old result. Therefore, before you get to the point at which you instruct the computer to catenate the latest result onto the vector of earlier results, the program should have a separate instruction which gives RESULT an initial value. Since before you start there aren't any results, the appropriate way to initialize this variable is to make it an empty vector, by an instruction such as this:

```
    RESULT←ι0
```

Example Using Catenation: Accumulating Primes

Here is a simple program which finds prime numbers by considering the odd integers in turn. The number being considered at any moment is called T, for Trial. The primes that have been found are in P. Whenever another T is found to be prime, it is catenated to P. The core of this program is the proposition that a number is prime if it cannot be divided evenly by any prime number smaller than itself.

```
     ∇ PR
[1]     P←1+T←1
[2]     TEST:→(END≤ρP)/PRINT
[3]     ADD:→(∨/0=P|T←T+2)/ADD
[4]     P←P,T
[5]     →TEST
[6]     PRINT:P
     ∇
```

On line 1, initial values are set for P and T. This program starts by assuming that it is already known that 2 is prime, so line 1 sets P to 2. T is initially set to 1 because the successive values of T are going to be increased by 2, and each increase is made before the test to see whether T is prime. The first T that will actually be tested is 3.

Line 2 is labelled, because it is the beginning of a loop. The loop starts with a test, to see whether a variable called END is less than or equal to the number of primes already found. If not, the work of testing another T continues. But if P has grown so that its length is equal to END, the program branches to an instruction called PRINT, which calls for printing of the accumulated primes.

Line 3 is a one-line loop. The line will be repeated indefinitely, each time with the value of T raised by 2, until a value of T is found which is not divisible by any of the primes already found. The instruction says, in effect, "Branch to the line labelled ADD (i.e. repeat this line) if it is true that any of the P residues of T is zero."

The program gets to line 5 only after it has found that none of the P residues of T is zero--that is, when T has been found to be a new prime number. T is catenated to the primes already found in the vector P.

After that, the program returns to the line labelled TEST, to see whether it has yet found enough primes.

Here is a sample execution, finding the first sixteen prime numbers:

```
     END←16
     PR
2   3   5   7   11   13   17   19   23   29   31   37   41   43   47   53
```

Making Any Variable Into a Vector

Occasionally it is useful to be able to turn a scalar into a one-element vector. For instance, suppose you have a program that works on whatever values have been stored under the name INPUT. In order to find out how many elements there are in INPUT, you might use $\rho INPUT$. But if INPUT was specified as a single dimensionless value (i.e. a scalar) rho of INPUT will be an empty vector. You won't be able to use its numerical value, since it has none. The remedy is first to convert INPUT so that it is always a vector, by using the ravel operator.

When the comma is used monadically (i.e. with no left argument) it ravels whatever is to the right of it. That is, it converts its argument to a vector, no matter what structure the argument previously had. If applied to a scalar, the ravel operator produces a one-element vector.

When a program asks for the shape of a variable that may be a vector or may be a scalar, it is prudent first of all to make sure that the variable is a vector by an instruction something like this:

$INPUT \leftarrow ,INPUT$

When the ravel operator is applied to multidimensional arrays, it produces vectors made by taking all of the elements in the multidimensional array and arranging them into a single vector. If the variable you ravel is a vector already, the result is the same vector, without change.

Inserting New Elements
Between Existing Elements of a Vector

Suppose V is a long vector. It contains, perhaps, several hundred elements. Now you find that you would like to insert several new elements between what are now the 135th and 136th elements. The new version of V can be assembled if you can catenate together these three vectors:

1. The vector containing V's elements 1 through 135.

2. The vector that is to be inserted; call it INSERT.

3. The vector containing V's elements from 136 to the end.

Getting the first 135 elements of V is easy: you just ask for $V[\iota 135]$. The INSERT (we'll assume) you have already.

There are several ways of getting the elements from 136 to the end; here's one. You index V by the consecutive integers starting after 135, and going until they reach the last element of V. Like this:

$V[135+\iota(\rho V)-135]$

The complete expression to reassemble a new V, longer than the old by the number of inserted elements, becomes:

$V[\iota 135]$, *INSERT*, $V[135+\iota(\rho V)-135]$

The same technique can be used to delete elements from within a vector. Suppose that you wish to keep elements 1 through D, but delete the N elements that follow element D. Then you want to keep all that remains after element D+N.

The formula for the first part of the new vector is:

$V[\iota D]$

and the formula for the remaining part is:

$V[D+N +\iota(\rho V)-D+N]$

so the formula for the whole new vector becomes:

$V[(\iota D), D+N+\iota(\rho V)-D+N]$

Here's an example. The vector G is a string of literals, like this:

$G\leftarrow$'*NOW IS THE TIME FOR ALL GOOD MEN TO COME TO OUR AID*'

You decide to keep elements 1 through 23, omit the next five, and then retain the rest. Like this:

$G[(\iota 23), 28+\iota(\rho G)-28]$
NOW IS THE TIME FOR ALL MEN TO COME TO OUR AID

There are several other techniques for inserting elements within a vector, or removing some. Some of them involve operations that have not yet been introduced, or which are mentioned only in the Appendix.

Building Pascal's Triangle:
An Example Using Catenation

The famous triangle that bears Pascal's name starts out like this:

```
              1
           1     1
        1     2     1
     1     3     3     1
  1     4     6     4     1
```

Each row has one more element than the row above it. The value of each element is the sum of the two elements nearest it in the row above. The triangle has many interesting properties; among the best known is the fact that each row provides a set of binomial coefficients. That is, the values of the nth row are the coefficients for the expansion of $(a+b)^{n-1}$.

You can construct Pascal's triangle in the following way. Notice that to get any row, you leave the elements on the end unchanged (they are always 1), and add all the pairs of adjacent elements. The fourth row is 1 3 3 1. You can get the fifth row by the following addition:

```
  1     3     3     1
        1     3     3     1

  1     4     6     4     1
```

In APL terms, this can be written as follows. First catenate a zero at one end of the row. Then add to that the same row but with a zero catenated at the other end. Like this:

$P \leftarrow (0,P) + P,0$

Here is a program that prints the first N rows of Pascal's triangle.

```
      ∇ PASCAL
[1]    P←1
[2]    P
[3]    →2×N≥ρ P←(0,P)+P,0
      ∇
```

```
        N←12
        PASCAL
1
1    1
1    2    1
1    3    3    1
1    4    6    4    1
1    5    10   10   5    1
1    6    15   20   15   6    1
1    7    21   35   35   21   7    1
1    8    28   56   70   56   28   8    1
1    9    36   84   126  126  84   36   9    1
1    10   45   120  210  252  210  120  45   10   1
1    11   55   165  330  462  462  330  165  55   11   1
```

The triangle, as printed by this program, turns out to be an erect right triangle rather than the more convention‑al form, but a triangle nonetheless. If you'd really like it Christmas-tree shaped, here's an alternate version which inserts a calculated number of spaces at the left--just enough to make the triangle symmetrical about its vertical axis. A sample execution of this program appears below it.

```
     ∇  TRI
[1]     P←,1
[2]     SPACES←(0⌈MIDPAGE-⌊0.5×+/3+⌊10⊛P)ρ' '
[3]     SPACES;P
[4]     →2×N≥ρ P←(0,P)+P,0
     ∇
```

```
        MIDPAGE←30
        TRI
                                1
                             1    1
                          1    2    1
                       1    3    3    1
                    1    4    6    4    1
                 1    5    10   10   5    1
              1    6    15   20   15   6    1
           1    7    21   35   35   21   7    1
        1    8    28   56   70   56   28   8    1
     1    9    36   84   126  126  84   36   9    1
  1    10   45   120  210  252  210  120  45   10   1
1    11   55   165  330  462  462  330  165  55   11   1
```

22: LOOPS

A repeated sequence of instructions is called a loop. Loops have already been discussed briefly on p. 70, and there's a loop in the program which accumulates prime numbers (pp. 139-40). This chapter brings together some points to be observed in writing programs with loops.

Exit from a Loop

Whenever a program contains a loop, you must provide the computer some way of knowing when to stop. It needs to have a test which tells when it has executed the instructions in the loop a sufficient number of times. The exit test is a branch instruction, written so that the line to which the computer branches depends upon whether the loop has been sufficiently repeated.

Here is a program called GCD. It finds G, the greatest common divisor of two numbers N1 and N2, by the Euclidean algorithm. The method depends upon the fact that a number which is an even divisor of both N1 and N2 must also be an even divisor of the remainder when N2 is divided by N1.

If there is no remainder when N2 is divided by N1, then we can immediately conclude that N1 is itself the greatest common divisor. But if N1 doesn't go into N2 evenly, then the g.c.d. must be some smaller number; in particular, it must be some number that is a factor of the remainder. So we look next for the g.c.d. of N1 and the remainder when N2 is divided by N1.

The program starts off by arbitrarily supposing that one member of the pair is the g.c.d., and assigns G the value of N1.Then on line 2 we have a test to see whether the G-residue of N2 is zero. This is the exit test: if the remainder is zero, then we may let G stand as the value of the g.c.d.

But if the remainder is not zero, then we respecify N1 as the remainder, and N2 as G (i.e. what N1 used to be), and try again.

```
      ∇  GCD
[1]     G←N1
[2]     →(0=N1←G|N2)/0
[3]     N2←G
[4]     →1
      ∇
```

In this example, the entire program is a single loop. The conditional branch in line 2 either permits another iteration to proceed or terminates the work. The loop is "closed" by an unconditional branch back to the first line of the program.

Is the program sure to reach an exit eventually, no matter what the values of N1 and N2? As long as N1 and N2 are integers, at each iteration G will be smaller. If the test at line 2 is not satisfied earlier, eventually G will be 1, and the 1-residue of any integer is zero. At that point the instruction on line 2 will result in a branch to zero, and the work will be complete.

Here is a sample execution when N1 is 1155 and N2 is 12298.

$N1 \leftarrow 1155$
$N2 \leftarrow 12298$
GCD
G
11

You might want to trace the execution of GCD. For instance, what happens if you transpose the two values so that N1 is initially the larger of the two?

Leading Decisions

As we remarked on page 74, sometimes the right number of times to repeat the execution of a loop is zero times--i.e. the work in the loop should not be executed at all. For that reason, it is better practice (whenever possible) to put the instruction which decides whether the loop should be entered at the beginning, rather than at the end.

Loops which involve repeating a sequence a specified number of times require a way of counting how many times the loop has been repeated. This is commonly done by using a variable whose sole function is to count which iteration of the loop is now in progress. Counters are not always needed; there are some calculations in which you could deduce how many times the work had been repeated without having an explicit counter, or others in which you want to have the loop repeated indefinitely until some other condition is satisfied, as in the preceding example.

Before the computer gets to the instructions that will be repeated in the loop, you will need to specify the initial values of the counter (if one is used), and perhaps the initial value of a result. Setting these initial values is often called "initializing" a loop.

Standard Procedure for Writing a Loop With a Counter

There are many ways of writing loops. The outline that follows isn't the only way things can be done. But it is quite general, and is recommended for many situations.

1. Pick some convenient name for the variable which is to be used as a counter. Any name not already in use for something else will do. I, J and K are conventional favorites for counters.

2. Give the counter its initial value. This should be one less than the first value that will be used inside the loop. The reason for this will be apparent when we get to step 4.

3. Give an initial value to the variable which contains the result of work on this loop, if that is appropriate. This should be the value that you want left as the result if it turns out that the loop is not executed at all.

4. Label the first line of the loop, so that you can come back to it easily. The labelled instruction is the test that decides whether the computer will continue on through the loop or will skip on to the next part of the program. This means that the form of the test is this:

   "If it is true that work on the loop is finished, branch to another part of the program. Otherwise, continue into the loop."

   Thus you want the computer to go ahead with executing the loop when the tested condition is false.

   The test can be combined with the instruction that increases the counter, so that that needn't take a separate instruction:

LABEL: →(*TOTAL*<*COUNTER*←*COUNTER*+1)/*NEXTPART*

Now the test instruction says in effect: "If it is true that the desired number of iterations of this loop is less than the new (augmented) value of the counter, you are about to overshoot, so branch. But if not, continue through the loop."

5. Write the working instructions for the loop. If you need to pick out individual elements of a vector, the counter may be used to index them.

6. After the last repeated instruction in the loop, branch (unconditionally) back to the labelled line that contains the exit test, at the beginning of the loop.

A summary of procedures for writing loops should also be accompanied by a caution: there are a great many situations in which the array-processing capabilities of APL make it unnecessary to use a loop. Earlier programming languages, which lacked provision for the parallel processing of the elements of an array, could express procedures on arrays only by the writing of loops. If you've had prior experience with one of these languages, or if you are writing an APL program by transcribing the procedure from another language, you may find that you've written a program with loops that aren't necessary. Such a program will still work in APL; it just won't be as concise or elegant as it might have been, nor as efficient in its use of the computer's time.

An Iterative Program to Print an Interest Table

An interest table shows the amount to which an initial sum will grow at various rates after each of the intervals at which interest is compounded. Suppose that the various columns of the table are the various rates of interest, while the rows are the successive compoundings. If PRINC is a scalar, containing the principal sum, and RATES is a vector of interest rates, while YEARS is the number of years for which interest is annually compounded, a simple program to generate the table might be as follows:

```
      ∇ INT1
[1]     I←0
[2]     →(YEARS<I←I+1)/0
[3]     PRINC×(1+RATES)*I
[4]     →2
      ∇
```

Here's an execution of INT1, for five years and three different rates of interest:

```
PRINC←100
YEARS←5
RATES←.05 .06 .07
INT1
105   106   107
110.25   112.36   114.49
115.7625   119.1016   122.5043
121.550625   126.247696   131.079601
127.6281562   133.8225578   140.2551731
```

The output reveals a problem: because the various lines of output were printed independently, each line is spaced for a convenient display of the numbers appearing on that line, but without regard to alignment with the other lines. So now let's modify the program to take care of that difficulty.

Alignment of Output in Columns

If you want the successive lines of output from a program to be vertically aligned, you have a choice of two procedures:

1. Instead of printing each line separately, one at each iteration of the loop, accumulate them until they can all be printed as a matrix. The computer automatically aligns the decimal points in the columns of a matrix.

2. Print each line separately, but instead of having the computer print the values directly, convert the numerical values to literal characters in a fixed format. There are many ways this can be done; one possibility is illustrated on page 151.

Interest Table with Output as a Matrix

The program INT2, shown overleaf, accumulates OUTPUT as a long vector, until the very last instruction, which restructures that vector as a matrix. The matrix has one more row than there are years, and one more column than there are rates. That permits the top row to show what the rates are, and the leftmost column to number the years. The zero in the top left corner doesn't do anything, but a

matrix must always have some value for every one of its elements.

```
    ∇ INT2
[1]   OUTPUT←0,RATES
[2]   I←0
[3]   I2LOOP:→(YEARS<I←I+1)/I2PRINT
[4]   OUTPUT←OUTPUT,I,PRINC×(1+RATES)*I
[5]   →I2LOOP
[6]   I2PRINT:(1+YEARS,ρRATES)ρOUTPUT
    ∇
```

INT2 instructs that OUTPUT is to be printed as a matrix. The width of the columns is therefore sufficient to accommodate any value that might appear, given the usual rules for the representation of numbers. Since APL\360 usually prints up to ten significant digits, the columns are spaced widely enough to accommodate numbers that long. It is possible to alter the maximum number of digits printed by the system with the command)DIGITS (see page 53), and that will make a corresponding adjustment in the column width of matrices.

```
    PRINC←100
    YEARS←5
    RATES←.05 .06 .07
    INT2
```

0	0.05	0.06	0.07
1	105	106	107
2	110.25	112.36	114.49
3	115.7625	119.1016	122.5043
4	121.550625	126.247696	131.079601
5	127.6281562	133.8225578	140.2551731

Interest Table with Fixed Format on Each Line

The program INT3 generates each row of the interest table independently. Then INT3 calls on another program called PRINT to do the actual typing of the result.

The definition of PRINT does not really concern us at this point, although it is shown as a footnote to page 152. (PRINT is a program which takes two arguments, like an APL dyadic operator. Functions with arguments are discussed in Chapter 25. This one prints the value of whatever expression appears to the right of it. The left argument indicates the maximum number of digits to be printed. All numbers are

printed with a decimal point and two places after the point,
a format that is appropriate for typing sums of money. The
definition of PRINT makes use of the representation encoding
operator, which is otherwise discussed only in the Appendix,
p. 195.)

Suffice it to say that PRINT sets up a vector of
literal characters to represent the various values within
the right argument of PRINT, always assigning the same field
width for each element, and always putting the decimal
points in the same position. Because the format is always
the same, regardless of the values that are printed, the
successive printings of the various rows of the table always
have the same horizontal spacing, and so the columns are
aligned even though printed independently.

Apart from its use of PRINT, INT3 is identical to INT1.
Here is the definition, followed by a sample execution of
the same problem used in the two preceding examples.

```
      ∇ INT3
[1]    5 PRINT RATES
[2]    J←0
[3]    ' '
[4]    I3LOOP:→(YEARS<J←J+1)/0
[5]    5 PRINT PRINC×(1+RATES)*J
[6]    →I3LOOP
      ∇

      PRINC←100
      YEARS←5
      RATES←.05 .06 .07

      INT3
  0.05      0.06      0.07

 105.00   106.00   107.00
 110.25   112.36   114.49
 115.76   119.10   122.50
 121.55   126.25   131.08
 127.63   133.82   140.26
```

A Footnote: the PRINT Program

The program which does the printing for INT3 is listed below. Unless you have some immediate need to use PRINT, or are especially interested in its definition, you should skip this note and go on to the next page. Similar programs may well be available through the system library. You may often find yourself making use of a program whose inner workings are quite unknown to you--so it isn't essential at this point to trace through what happens in PRINT. But if you need it, here it is.

All of the names appearing in the program are local variables (see Chapter 25). DGTS is the maximum number of digits to appear in a printed number. FLD is the total field for representing one number. SHP is the shape (i.e. rho) of the right argument, X; it will be used at the end to reshape the literals back into a shape that matches the original shape of X.

PL is a vector showing the number of digits to be used for each of the elements of X, but always at least 3 and never more than DGTS. RSLT, starts out as a vector of blanks and decimal points, ready to receive the encoded representations as they are calculated. REP is the representation of a single element, including its sign, and I is the counter, that controls the iterations of the loop that generates the result for each number.

In the last line, the literal vector RSLT is restructured. The last dimension of the literal array is FIELD times longer than the last dimension of X.

```
      ∇ RSLT←DGTS PRINT X;POS;FLD;SHP;PL;REP;I
[1]    POS←(ιFLD-3), ‾1 0 +FLD←3+DGTS
[2]    SHP←ρX
[3]    RSLT←(FLD×ρX←,X)ρ(DGTSρ' '),'.    '
[4]    PL←DGTS⌊3⌈3+⌊10⊛(X=I←0)+|X
[5]    PRT1:→((ρX)<I←I+1)/PRT2
[6]    REP←(1+0>X[I]),3+(PL[I]ρ10)⊤⌊0.5+100×|X[I]
[7]    RSLT[(FLD×I-1)+(-ρREP)↑POS]←' ‾0123456789'[REP]
[8]    →PRT1
[9]    PRT2:RSLT←(((‾1+0=ρSHP)↓SHP),FLD×‾1↑1,SHP)ρRSLT
      ∇
```

Paying the Mortgage

Suppose that a loan is to be repayed so that the payments are always of the same size, and at regular intervals. Suppose that the principal sum and the interest rate are fixed. For a given number of payments, you can solve for the size that each must be. Conversely, given the amount paid each time, you can solve for the number of payments to pay off the entire debt.

It turns out that the size of each flat-rate payment can be found, at least approximately, without using a loop. The following program does that. PRINC is the principal sum borrowed, T is the number of times a payment will be made, and RATE is the interest in one time interval. The program is approximate since it does not include the effects of rounding to the nearest penny at each iteration.

```
    ∇ SIZE
[ 1]    PAYMENT←(PRINC×(1+RATE)*TIMES)÷+/(1+RATE)*TIMES-ιTIMES
    ∇
```

Suppose that the principal to be borrowed is $17,300, the interest rate is .055 per year, and it is to be paid in 240 monthly installments (i.e. 20 years). Then RATE should be one twelfth of .005.

```
    PRINC←17300
    RATE←.055÷12
    TIMES←240
    SIZE
    PAYMENT
119.0045043
```

Next we consider a program which counts the number of payments needed to pay off the mortgage. This is an iterative program, and so it can include at each iteration the correction for rounding to the nearest cent. As usual, the program that follows contains a loop and a counter. But the exit-test is whether the balance due has been reduced to zero, while the counter keeps track of the number of iterations needed. At the same time, the program notes the amount of the last payment, since that may be for the odd amount due at the end.

For the first execution, let's see if the approximation obtained as the result of the program called SIZE does indeed pay off the mortgage in exactly 20 years.

```
      ∇ REPAY
[1]     BAL←PRINC×100
[2]     PAY←PAYMENT×100
[3]     COUNT←0
[4]     RPLOOP:→(0≥BAL←⌊0.5+BAL×1+RATE)/RPEND
[5]     BAL←BAL-LASTPAY←PAY⌊BAL
[6]     COUNT←COUNT+1
[7]     →RPLOOP
[8]     RPEND:'TOTAL OF ';COUNT;' PAYMENTS'
[9]     'OF WHICH ';COUNT-1;' ARE ';PAYMENT
[10]    'AND THE LAST IS ';LASTPAY÷100
      ∇
      PRINC←17300
      RATE←.055÷12
      PAYMENT←119
      REPAY
TOTAL OF 241 PAYMENTS
OF WHICH 240 ARE 119
AND THE LAST IS 1.99
```

As it turns out, the payment of $119.00 is not quite sufficient to pay off everything in 240 months. Let's try again with a slightly larger payment.

```
      PAYMENT←119.01
      REPAY
TOTAL OF 240 PAYMENTS
OF WHICH 239 ARE 119.01
AND THE LAST IS 116.67
```

An Iterative Program for Finding Prime Factors

Suppose NUMBER is a scalar integer. You need to find all of the prime factors of NUMBER. This isn't just a matter of finding which primes are factors of NUMBER, since you also want to know how many times any particular prime is a factor.

The program called PF finds prime factors. It presumes that you already have in the workspace a vector called PRIME, which contains all the prime numbers you are likely to need, in ascending order.

On line 1, X is given the same value as NUMBER. As factors are extracted, X will be reduced by dividing it by each new factor as it is found, but NUMBER will be left unchanged.

```
      ∇ PF
[1]    X←|NUMBER
[2]    FACTORS←ιI←0
[3]    NEWTF:→((X*0.5)<TF←PRIME[I←I+1])/OVER
[4]    TRYTF:→(0≠TF|X)/NEWTF
[5]    FACTORS←FACTORS,TF
[6]    X←X÷TF
[7]    →TRYTF
[8]    OVER:FACTORS←FACTORS,(X≠1)ρX
[9]    →(1=ρFACTORS)/PR
[10]   'PRIME FACTORS OF ';NUMBER;':   ';FACTORS
[11]   →0
[12]   PR:NUMBER;' IS PRIME'
      ∇
```

On line 2, I is given an initial value of 0, and FACTORS (which will contain the result) is made an empty vector.

The line labelled NEWTF tests to see whether work has been completed. TF (for "trial factor") is selected as the next prime number from the vector PRIME. Then TF is compared with the square root of X. At the first iteration, X has the same value as NUMBER, but in subsequent iterations X is the quotient after NUMBER has been divided by each of the factors already found. If TF is larger than the square root of X, TF can't be a factor of X, and nor can any other number larger than TF, so it is safe to conclude that no new value of TF, other than X itself, is going to be a factor of X. If there are no new factors to be found, the program branches to OVER.

But if there may be factors yet unfound, the program proceeds to the next line in sequence, which is labelled TRYTF. Here the current value of TF is tested to see if it is a factor of X. If it is, the program catenates TF to the FACTORS already found, divides X by TF, and branches back to TRYTF. Note that it does not go back to the line labelled NEWTF, since the old value of TF may still be a factor of the newly-divided X.

Only when it is established that TF is not a factor of X does the program return to NEWTF, and select the next prime number as a value for TF.

When the computer reaches the line labelled OVER, there are two possible situations. If the successive values of TF which were catenated to form the vector called FACTORS

indeed account for all the prime factors of NUMBER, then the value of X must be 1. That is, X now has the value that you get when NUMBER is successively divided by all of its factors.

But it is also possible for the program to reach the line labelled OVER if the last value of X is a prime different from any of those so far used in TF. In that case, the test on line 3 will correctly reveal that there is no other factor of X smaller than the square root of X. In this case, however, X itself should be counted among the prime factors of NUMBER.

So on line 8, to the end of the vector FACTORS is catenated either one or zero extra elements, having the value X.

Notice that line 8, although not formally a branch instruction, has the effect of either catenating the value of X or not catenating it, depending upon whether or not the value of X is 1.

Line 9 represents a small refinement in the output. If only one prime factor has been found, then the NUMBER was a prime, and the result may be printed in a different format.

Here are some examples of the PF program at work:

```
      NUMBER←1505
      PF
PRIME FACTORS OF 1505:   5   7   43

      NUMBER←1728
      PF
PRIME FACTORS OF 1728:   2   2   2   2   2   2   3   3   3

      NUMBER←12345679
      PF
PRIME FACTORS OF 12345679:   37   333667

      NUMBER←333667
      PF
333667 IS PRIME
```

23: COMPRESSION:
SELECTING SOME ELEMENTS FROM A VECTOR
AND OMITTING OTHERS

Suppose you have a vector named V. You would like to generate a new vector that contains some of the elements from V, but omits others. For instance, you want to keep all those that are greater than zero, while omitting those that aren't. The APL operator that does this is called compression. The sign for compression is a slash--the same sign that is used for reduction. The two operations, compression and reduction, are easily distinguished by the fact that in reduction the slash has an operator sign immediately to the left of it, whereas in compression there must be an expression resulting in a vector of zeroes and ones in that position.

The way compression works is this: wherever there is a 1 in the vector on the left, the corresponding element of the array on the right is retained. But where there's a 0 on the left, the corresponding element on the right is omitted. The left argument must contain a 0 or a 1 for each element on the right; that is, the two vectors must have the same length.

Suppose that as the right argument of a compression you have a vector called V, composed like this:

$$V \leftarrow 1.1 \quad \bar{}2.2 \quad 3.3 \quad 4.4 \quad \bar{}5.5 \quad 6.6 \quad 7.7 \quad 8.8$$

You want to keep all the elements from V except the second and fifth. So as a left argument for compression you need a vector that has the same length as V, and all of whose elements have the value 1 except the second and the fifth, which must be zero.

```
      1 0 1 1 0 1 1 1/V
1.1   3.3   4.4   6.6   7.7   8.8
```

If the selection vector (i.e. the left argument of compression) is made up entirely of ones, then all the elements from the array on the right are preserved:

```
      1 1 1 1 1 1 1 1/V
1.1   ¯2.2   3.3   4.4   ¯5.5   6.6   7.7   8.8
```

Conversely, if the selection vector consists of nothing but zeroes, then none of the elements on the right is selected, and so the result is an empty vector:

$$0\ 0\ 0\ 0\ 0\ 0\ 0\ 0/V$$
◀(Here the computer prints a blank line)

In general, the vectors on either side of the compression sign must be of the same length, so that the ones and zeroes on the left can be matched one-to-one with the elements on the right. However, if either argument is a single element, as usual the computer first replicates it until it matches the length of the vector on the other side. Thus a single 1 on the left of a compression keeps everything from the vector on the right:

$$1/V$$
1.1   ¯2.2   3.3   4.4   ¯5.5   6.6   7.7   8.8

and a single 0 on the left of a compression selects none of the elements on the right.

$$0/V$$
◀(Here the computer prints a blank line)

Similarly, if a selection vector having several ones and zeroes is used to compress a single number, you get that number selected once for each of the ones on the left:

$$1\ 0\ 1\ 1\ 0/6.02$$
6.02   6.02   6.02

In fact, whenever the left argument of a compression contains more than one element, the length of the result is the same as the number of ones in the selection vector.

Tests of the Truth of a Relationship Provide the Zeroes and Ones Needed to Control Compression

You will recall that when the computer tests whether a relationship is true, it responds with 1 for true, and 0 for false. These ones and zeroes are just what is needed for the selection vector during compression. For instance, suppose you would like to keep from V only those elements that are greater than some constant X. The expression

$$V > X$$
0   0   0   1   1   1   1   1

generates a response for each element in V. That response is
1 for each element of V that is greater than X, and 0 for
each that is not. This expression can be used directly in
the compression, like this:

$$(V>X)/V$$
4.4   6.6   7.7   8.8

(Evidently X was something smaller than 4.4, but greater
than 3.3)

$$(V\leq 0)/V$$
$^-$2.2 $^-$5.5

Example: Compression and the Sieve of Eratosthenes

Our earlier program for finding prime numbers
considered at each iteration whether a single number N was
or was not a prime. If it was, it was catenated to the list
of primes found already. Then N was increased by 2, and
checked again. A different procedure was proposed by
Eratosthenes around 200 BC. He suggested that you start with
all the integers (or as many as you have patience for) and
successively cross out all those divisible by various
divisors. The numbers that remain when all possible divisors
have been tried are the primes.

You don't have to try all possible divisors; once a
number has dropped through the sieve, it doesn't need to be
considered as a divisor either. After you finish with one
trial divisor, the next trial divisor is the next higher
number from among the potential primes still remaining.

Here is a program to find primes by the sieve method.
In the earlier program, the test for finishing work was
whether sufficient primes had been found. But with the sieve
method it is easier to count how many numbers are in the
sieve at first; you can't say in advance exactly how many of
them will turn out to be prime. So the test for stopping is
whether you've reached a divisor so high that it couldn't
possibly divide any of the remaining numbers in the initial
set. The square root of the largest number in the sieve is
such a number. The initial divisor is 2, and
values for the potential primes are the integers from 2 to
N.

```
      ∇  ERATOS
[1]     PP←1+ιN-1
[2]     LAST←N*÷D←2
[3]     →(LAST<D)/ERPRINT
[4]     PP←((D=PP)∨0≠D|PP)/PP
[5]     D←PP[1+PPιD]
[6]     →3
[7]     ERPRINT:PP
      ∇
```

The compression on line 4 is the sieve. An element of PP is retained if it meets either of two conditions: if it is equal to D, the divisor, or if it is not exactly divisible by D.

After each use of the sieve, D is respecified as the next of the potential primes now remaining (line 5).

Here is a sample execution showing selection of the numbers that are prime up to 20.

```
      N←20
      ERATOS
2   3   5   7   11   13   17   19
```

It might be interesting to trace the execution of ERATOS to see how many different trial divisors are used before all the prime numbers up to 20 can be found. If you trace the execution of lines 4 and 5, you will see on line 4 the potential primes as they are sifted until only genuine primes remain, and on line 5 the successive values of D following the initial value of 2. (Tracing was discussed on page 91.) To start tracing, you enter:

```
      TΔERATOS←4 5
```

Now when you execute ERATOS, you see the values after each execution of lines four and five, before the final printing of the result.

```
      ERATOS
ERATOS[4] 2   3   5   7   9   11   13   15   17   19
ERATOS[5] 3
ERATOS[4] 2   3   5   7   11   13   17   19
ERATOS[5] 5
2   3   5   7   11   13   17   19
```

Two iterations were needed. After the initial value of 2, D took on the value 3 and then 5. No compression was done for D=5, since that value already exceeds the square root of N.

How many iterations would it take to select all the primes up to 1000? This time, if we just trace line 5, we shall see each of the successive values of D, and then the answer:

```
      TΔERATOS←5
      N←1000
      ERATOS
ERATOS[5]  3
ERATOS[5]  5
ERATOS[5]  7
ERATOS[5]  11
ERATOS[5]  13
ERATOS[5]  17
ERATOS[5]  19
ERATOS[5]  23
ERATOS[5]  29
ERATOS[5]  31
ERATOS[5]  37
2   3   5   7   11   13   17   19   23   29   31   37   41   43   47   53
        59   61   67   71   73   79   83   89   97   101   103   107   109
        113   127   131   137   139   149   151   157   163   167   173
        179   181   191   193   197   199   211   223   227   229   233
        239   241   251   257   263   269   271   277   281   283   293
        307   311   313   317   331   337   347   349   353   359   367
        373   379   383   389   397   401   409   419   421   431   433
        439   443   449   457   461   463   467   479   487   491   499
        503   509   521   523   541   547   557   563   569   571   577
        587   593   599   601   607   613   617   619   631   641   643
        647   653   659   661   673   677   683   691   701   709   719
        727   733   739   743   751   757   761   769   773   787   797
        809   811   821   823   827   829   839   853   857   859   863
        877   881   883   887   907   911   919   929   937   941   947
        953   967   971   977   983   991   997
```

Evidently eleven iterations sufficed. By contrast, the earlier program PR would have taken 499 iterations to find that many primes.

Another Program Using Catenation and Compression:
Sorting the Elements of a Vector

Sorting the elements of a vector so that they are arranged in ascending order is a classical problem to which there are a great many solutions. Indeed, a sorting function is directly available in APL as a primitive function (see the next section). But suppose it were not: how might a sorting program be written? Here is one which uses compression to find which elements should go first, and catenation to reassemble them into a new, ordered vector. The steps in the procedure are as follows:

1. Call the vector that is to be sorted UNS (for unsorted). Call the sorted vector that results ORD (for ordered). Start with ORD being an empty vector.

2. Test to see whether any elements remain in UNS. If there are none, exit.

3. Set up the logical vector WHICH, with a 1 corresponding to each element of UNS that is equal to the minimum of UNS.

4. Compress UNS by WHICH. That is, pick out from UNS those elements that are equal to its minimum. Catenate them to those already found in ORD.

5. Compress UNS by the negation of WHICH. That is, respecify UNS to be all those elements that were not selected.

6. Return to line 2.

```
      ∇ SORT
[1]    ORD←0ρUNS←,UNSORTED
[2]    →(0=ρUNS)/0
[3]    WHICH←UNS=⌊/UNS
[4]    ORD←ORD,WHICH/UNS
[5]    UNS←(~WHICH)/UNS
[6]    →2
      ∇
```

Here is a sample execution of SORT:

```
      UNSORTED←18  43  6  22  17  6  44  29  8  19  24  17  32
      SORT
      ORD
6  6  8  17  17  18  19  22  24  29  32  43  44
```

A More General Form of the Sorting Program

  The program called SORT starts with a vector that may
be in scrambled order and produces a vector with the same
values arranged in ascending order. Sometimes it is more
useful to produce as your result not the values themselves
arranged in order, but the index numbers which, if used to
index the scrambled vector, would order it. The advantage of
doing it that way is that, once you have the ordered index
numbers, you can then apply them not only to the original
scrambled vector, but to any other vector of the same
length. For instance, suppose FINAL is a vector of the
grades obtained by a class of students, and ID is a vector
of their identification numbers. Then you could arrange ID
in an order based upon the order of their grades. (Or, when
you get into multidimensional arrays, you could have their
names arranged as the rows of a matrix, and print their
names in an order determined by their grades.)

  To do that, you again find a logical vector WHICH. But
now instead of using it to select values from the scrambled
vector, you use it to select index numbers. Now you remove
elements from the vector of index numbers as well as from
UNS. But you still iterate until all the elements of UNS are
used up. Here is such a definition:

```
      ∇ SORTX
[1]     ORDX←0ρINDEX←ιρUNS←,UNSORTED
[2]     →(0≥ρUNS)/0
[3]     WHICH←UNS=⌊/UNS
[4]     ORDX←ORDX,WHICH/INDEX
[5]     UNS←(~WHICH)/UNS
[6]     INDEX←(~WHICH)/INDEX
[7]     →2
      ∇
```

  Using SORTX, in order to put the elements of a variable
called Y into order, you have to index Y by ORDX, the vector
of ordered indices that the program produces:

```
      UNSORTED←Y←18  6  24  72  14  27  6  31  17  14  20
      SORTX
      Y[ORDX]
6   6   14   14   17   18   20   24   27   31   72
```

In the next example, a vector called FINAL contains the grades for a class of students. Their names are stored as the rows of the matrix NAMES. The program called REPORT prints both the names and grades in rank order by grade.

```
      NAMES

BRENNER,  WILLIAM
DRISCOLL,  KEITH
GALTO,  JULIE
KURTZBERG,  BURTON
ROTHWELL,  DAVIS
STRONG,  VERA
SUGARMAN,  DAVID
THOMPSON,  EDWARD
WATSON,  EDWIN
YANG,  TSIAO

       FINAL
73   80   79   84   90   85   76   94   62   80

      ∇ REPORT
[1]    I←0
[2]    UNSORTED←FINAL
[3]    SORTX
[4]    →( (ρFINAL)<I←I+1)/0
[5]    NAMES[ORDX[I];];'   '; FINAL[ORDX[I]]
[6]    →4
      ∇


      REPORT
WATSON,  EDWIN          62
BRENNER,  WILLIAM       73
SUGARMAN,  DAVID        76
GALTO,  JULIE           79
DRISCOLL,  KEITH        80
YANG,  TSIAO            80
KURTZBERG,  BURTON      84
STRONG,  VERA           85
ROTHWELL,  DAVIS        90
THOMPSON,  EDWARD       94
```

The Primitive Functions for Sorting

The result produced by the program SORTX can be obtained directly with the APL operator called grade. If the argument of grade is a numerical vector, then the result is the indices that will arrange that vector in either ascending or descending order. The operator which arranges in ascending order is called grade up, and is formed by overstriking the delta Δ and the vertical line | otherwise used for absolute value. The operator which arranges the values in descending order is called grade down, and is formed by overstriking with the del instead of the delta: ⍒.

Using this primitive operator, an expression which would arrange the scores on a final exam called FINAL would be:

   FINAL[⍋FINAL]

In order to print the matrix NAMES in ascending order of the vector FINAL (assuming there is a row of NAMES for each element of FINAL) the instruction would be:

   NAMES[⍋FINAL;]

Why the Branch-or-Continue Instruction Includes a Compression

On page 71, we remarked that a conditional branch instruction may be written

   →CONDITION/LINE

The value of CONDITION is logical (i.e. either 1 or 0). The result of the compression is therefore either the value of the variable called LINE (when CONDITION is 1), or else an empty vector (when CONDITION is 0).

A branch to an empty vector is no branch at all: it is taken to mean "Continue with the next instruction in sequence."

Thus a branch-or-continue instruction is any instruction in which a right-pointing arrow is followed by an expression which, when evaluated, yields either the number of a line to which the program is to branch, or else an empty vector if no branch is to be taken.

Compression is not the only operator which would give that effect. Recall that ⍳0 also produces an empty vector. So another form of the branch instruction can be written as follows. Suppose LINE is a label for the line to which the program is to branch if it is true that X is smaller than the square root of Y. Otherwise the program should continue in sequence. You could get that by using compression (as is done in almost all the illustrative programs in this primer) like this:

$\rightarrow(X<Y*0.5)/LINE$

Or you could get it by this instruction, which has the advantage of putting the label at the beginning rather than the end:

$\rightarrow LINE\times\iota X<Y*0.5$

You can read that instruction as "Branch to LINE if X is less than the square root of Y."

24: THE PROGRAM ASKS FOR INPUT, GETS IT, AND THEN PROCEEDS

The quad symbol □ stands for input and output. If a quad appears immediately to the left of a specification arrow, it means that the value to the right of the arrow is to be printed. You don't often need this sort of explicit instruction to print something, since the computer prints a value automatically anytime you fail to specify what else is to be done with it.

If a quad symbol appears anywhere else in an instruction (that is, anywhere but immediately to the left of a specification arrow), it means that the computer should at that point ask for input from the terminal. Suppose you enter an instruction like this:

    Z←□

The value of Z is to be specified as whatever value is entered from the terminal in response to the quad. For this reason, input in response to a quad is called evaluated input. To show that it is requesting input, the computer types a quad and colon at the left margin, and then indents and unlocks. The value of the expression that you type now is taken as the value of □; in this case, that value is now assigned to the variable Z.

Here's how it looks: first the instruction containing a quad. Then the quad typed by the computer, to show that it is requesting input. Then your response to that request. Finally, if you ask to see the value of Z, you find that the value of the expression you entered at the quad has indeed been assigned to Z.

        Z←□
    □:
        2×3
        Z
    6

Anytime a quad occurs in an instruction, when the computer reaches that point in its evaluation of the instruction, it goes to the terminal for input, evaluates what you enter then, and then returns to the original instruction. Suppose you enter this instruction:

    X×□+A×B

Recall that the computer performs the rightmost operation
first, so first it finds the product of A and B. Then it
encounters the ⬚ symbol; the product of A and B is to be
added to the value of ⬚. Whatever you enter now becomes the
value used in the instruction. If you enter 6, that value is
added to the product of A and B. But if you enter an
expression, that entire expression is evaluated at once, and
its result becomes the value used.

In the illustrations that follow, A has the value 5, B
has the value 2, and X has the value 1.

$$X \times \square + A \times B$$
⬚:
      6
16

$$X + \square * .5$$
⬚:
      2
2.414213562

$$X + \square$$
⬚:
      $A \times B$
11

$$\sqcup \div X$$
⬚:
      $A * \iota 5$
5   25   125   625   3125

Example of Input to a Program:
Crystal Lattice Problem

In the examples used in earlier chapters, the data
needed for a particular program had to be assigned to
variables before execution of the program. It may be more
convenient to have the program ask for the data it needs as
it goes along. You can do that by using the ⬚ in the
program. For instance, here is a program intended for work
with some problems in the geometry of crystal lattices. The
program finds D, the distance between adjacent planes of a

hexagonal crystal, as a function of 5 parameters. The first two, A and C, are constant for a given compound. The other three, called H, K, and L, are integers which identify the set of planes under consideration. In conventional notation, D can be found from the following formula:

$$\frac{1}{D^2} = \frac{4}{3}\left(\frac{H^2 + HK + K^2}{A^2}\right) + \frac{L^2}{C^2}$$

At the bottom of the page you will find an APL program which first asks for the values of A and C (as a single 2-element vector) and then asks for an HKL combination. After printing the value of D, the program returns and asks for a new HKL combination. It will keep repeating until you enter a scalar instead of a vector for HKL.

Notice that when the □ asks you for input, you're free to enter numerical values, or an expression, or the name of a variable. For instance, suppose you may want to work repeatedly with germanium oxide. You could store the values of A and C for germanium oxide under the name GEO2. Similarly, since the program ends by testing to see if HKL is a scalar, you could store a scalar under the name END, and henceforth END will suffice to indicate the end of your execution of the program. Both of these points are illustrated on the next page.

```
    ∇ HEXAGONAL
[1]    'SPECIFY A AND C (IN ANGSTROMS)'
[2]    AC←□
[3]    'SPECIFY  H K L'
[4]    HKL←□
[5]    →(0=ρρHKL)/0
[6]    D←÷(+/ 4 4 4 3 ×HKL[1 2 2 3]×HKL[1 1 2 3]÷3×AC[1 1 1 2]*2)*0.5
[7]    'D IS ';D;' ANGSTROMS'
[8]    →4
    ∇

    GEO2←4.987 5.652

    END←0
```

```
        HEXAGONAL
SPECIFY A AND C (IN ANGSTROMS)
⎕:
        GEO2
SPECIFY  H K L
⎕:
        1 0 0
D IS 4.318868689 ANGSTROMS
⎕:
        1 0 1
D IS 3.431678812 ANGSTROMS
⎕:
        1 1 0
D IS 2.4935 ANGSTROMS
⎕:
        1 0 2
D IS 2.364743232 ANGSTROMS
⎕:
        1 1 1
D IS 2.281351922 ANGSTROMS
⎕:
        2 0 0
D IS 2.159434344 ANGSTROMS
⎕:
        2 1 1
D IS 1.568280737 ANGSTROMS
⎕:
        END
```

Ordinarily, the system types ⎕: (followed by a line-feed) each time it requests evaluated input. It is possible, if you prefer, to substitute any other characters that you may prefer for the ⎕:. This can be done by using the program SFEI (for "symbol for evaluated input"). That program is to be found in one of the public workspaces supplied with the APL\360 System. See the APL\360 Manual, discussion of Evaluated Input in part three, and discussion of Library Functions in part four.

Input as Literal Characters

The symbol ⍞ is called quote-quad; it is formed by overstriking the quad symbol with a quote mark. Quote-quad asks for input in the same way that quad does, but with two important differences:

1. When the computer requests input from a ⍞, it simply unlocks the keyboard with the typeball at the left margin. It doesn't print a quad symbol, and it doesn't indent.

2. Whatever you enter in response to a ⍞ is accepted as literal characters. If you enter just one character, it goes in as a literal scalar. If you enter any other number of characters, they go in as a literal vector. In particular, if you don't type anything but a carrier return, a vector of length 0 is entered.

Suppose you would like to build up a list of names. The list might be a long literal vector containing all the various names. At the same time that you enter new items to the list, you want to keep a record of where each entry starts and how long it is. Then later on you can recover a name from the list by indexing the vector called LIST. The following pair of programs illustrate how this might be done.

The first one compiles a vector of literal entries. It continues to accept new entries until you enter an empty vector. Line 6 tests the length of the entry, and terminates execution when the length is 0.

The second program prints entries from that vector --those that are indicated by the values of NO. Actually, it prints only one entry at a time. A counter called J steps through the various elements of NO. Since NO is indexed by J, NO has to be a vector. So the first line of PRINT ravels NO.

```
      ∇ ENTER
[1]   LENGTH←LIST←⍳START←,0
[2]   ρSTART
[3]   →(0=ρENTRY←⍞)/0
[4]   LENGTH←LENGTH,ρENTRY
[5]   START←START,ρLIST←LIST,ENTRY
[6]   →2
      ∇


      ∇ PRINT
[1]   NO←,NO
[2]   J←0
[3]   →((ρNO)<J←J+1)/0
[4]   LIST[START[NO[J]]+⍳LENGTH[NO[J]]]
[5]   ''
[6]   →3
      ∇
```

Below, there are samples of the execution of these two programs. Notice that, since the input is literal characters, any character on the typeball can be included in the input. Indeed, you don't have to use an APL ball at all, but you can type input with any other typeball that fits your terminal. In the following example, a script typeball was fitted while the names were being entered and then again when printing of the names numbered 4, 2, and 5 was asked for.

```
        ENTER
1
Mr. and Mrs. John H. Hoe, 245 Center Street, Plainesville, Michigan
2
Miss Barbara Halverson, 12245 South Broadway, Alameda, Olkahoma
3
Dr. Harold Jacobs, RFD 4, Bartontown, New Jersey
4
Mr. Jonathan Lester, 614 24th Avenue NW, Cedar Falls, Iowa
5
Mr. and Mrs. J. Q. Walden, Trade Center, Pt. Barrow, Alaska
6
IBM Research Center, Yorktown Heights, New York 10598
7
```

```
        NO←4 2 5
        PRINT
```

Mr. Jonathan Lester, 614 24th Avenue NW, Cedar Falls, Iowa

Miss Barbara Halverson, 12245 South Broadway, Alameda, Olkahoma

Mr. and Mrs. J. Q. Walden, Trade Center, Pt. Barrow, Alaska

25: DEFINED FUNCTIONS
THAT HAVE ARGUMENTS AND RESULTS

Up to now, the discussion of how to write a program has dealt only with what on page 33 we called "stand alone" programs. The instruction that calls for the execution of such a program always consists of just one word: the name of the program. With that sort of program, the data the program works on must either be stored in the workspace before you execute it, or else entered from the typewriter when the program calls for input. However, APL provides for some other forms of definition which are more powerful and often far more useful than the simple type to which discussion has been confined. This chapter is devoted to introducing these more general forms of program definition.

The Idea of a Function

To a mathematician, a function is a correspondence between one set of values (the domain) and another (the range). This correspondence can be represented in various ways. One way would be to have a table in which each value of the domain appears beside the corresponding value of the range.

Another way to represent a function is to state an algorithm (or procedure) by which, given any particular value within the domain as input, you (or a computer) could determine the corresponding value as an output, or result. In APL, a program is considered to be the algorithmic definition of a function, and a program may be used like a function, provided it is properly defined.

The Arguments and the Result of a Function

The operations of arithmetic are functions; if you perform an addition, you start with the addends (the input) and you follow a procedure which gives you the sum (the output, or the result). The input values to a function are called its arguments. In the instruction $3+4$, the function is addition, and the arguments are 3 and 4. You have already seen that the primitive functions of APL (each of which has its own symbol) are always written in one of two forms: for a function of two arguments, the function symbol always appears between the two arguments (like $A+B$, or $A*B$, and so on). For a function of only one argument, the argument appears to the right of the function symbol.

Suppose A, B, and C are variables. Consider the instruction

$$A + B \div C$$

It contains two primitive functions, addition and division. The division function has two arguments: B and C; B is the dividend, since it's on the left of the $\div$ sign, and C is the divisor, since it's on the right.

What are the arguments of the addition function? The left argument of + is A. The right argument of + is whatever result you get when you finish executing the division of B by C. The point is important : an instruction which calls for the execution of two functions depends upon the fact that the first returns a result which then becomes the argument of the second.

Programs as the Definitions of Functions

A program is a statement of a procedure. It generally works on some input data, and processes the input until it produces a result; the value of the result depends on what the input values are: i.e. the result is a function of the input. So it is perfectly reasonable and consistent to think of a program as a function.

If the system in which you're working has a primitive operator for everything you ever want to do, you never need to write programs. A program is a way of telling the computer the procedure it must follow in order to evaluate a function that it doesn't otherwise have.

APL uses the general word "function" to refer both to the operators that are primitive to the language, and to the programs that APL users write. A program is simply a user-defined function.

When you use a defined function, it would be very handy to be able to use it in the same way that you use primitive functions. For instance, you'd like to be able to say what function is to be used, what values it is to work on, and what is to be done with the result, all in the same instruction.

Suppose that you sometimes need to calculate the resistance RR of several resistors in parallel. Their

resistances, considered separately, are stored as the elements of a vector called R. In conventional notation, the formula for RR is:

$$\frac{1}{RR} = \frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3} + \frac{1}{R_4} + \cdots + \frac{1}{R_n}$$

There is no APL primitive which, when applied to R, gives you RR. So you would like to define a function which does that. Suppose that function is called PR (for "parallel resistance"). Before we discuss how to write a definition for this function PR, consider how you would like to be able to use it.

To find the parallel resistance for the vector R, you'd like to be able to enter simply:

   PR R

Or to find the parallel resistance of a resistor of 800 ohms and another of 1200 ohms, you'd like to be able to enter:

   PR 1200 800

You'd like to get back the answer simply by entering the instruction:

   PR 1200 800
480

Or conversely you'd like to be able to assign the result of *PR R* to a variable, like this:

   RESIST←PR R

just as you would if PR were an APL primitive.

This description implies that PR, just like a primitive operator, takes as its argument whatever comes to the right of it in the instruction. Like a primitive operator, it returns a result that may be stored, or passed on to the next operator to the left, or printed if neither of the other two is indicated.

It is a simple matter to write the definition of PR so that it behaves in this way. Indeed, every one of the program definitions used in the various examples in the

early chapters of this primer could be written in that way, and would thereby become a great deal more convenient to use.

The Definition of a Function
That Takes an Argument and Returns a Result

The joint resistance of several resistors in parallel may be found as the reciprocal of the sum of the reciprocals of the separate resistances. In APL, that is:

$RR \leftarrow \div + / \div R$

Here is the definition for the function PR:

```
     ∇ RR←PR R
[1]    RR←÷+/÷R
     ∇
```

It differs from the definitions that appeared in the earlier examples in two ways:

1. Its header (that is, the top line which contains the ∇ symbol and the name of the function) now includes some other items which serve to indicate that this function takes one argument and returns a result.

2. The definition does not contain any statement calling for the printing of the result. Now that the function has a formal result, the result will be printed automatically whenever the instruction calling for execution of this function doesn't indicate some other use for the result.

The header of a function definition always stipulates the name of the function. At the same time, the header serves as a paradigm, illustrating the syntax that is to govern the way this function will be used.

If the header includes a specification arrow (with some name to the left of it) it means that the function returns a result. That result may be stored (as illustrated in the header), or passed on to some other function appearing further to the left in the same instruction, or printed, just like the result of a primitive function.

If in the header the name of the function appears with one or two other names next to it, those other names indicate the arguments of the function. When you use the function, you must provide a variable or expression next to the name of the function, in the positions illustrated in the paradigm. As with the primitive functions, if there's one argument, it comes after the function, and if there are two they go on either side of the name of the function.

GCD: A Simple Function of Two Arguments

On page 145 we gave a definition for a program to find the greatest common divisor of two numbers N1 and N2. Leaving the body of the definition exactly as it was, we can write a second version with a different header, making N1 and N2 the arguments of GCD, and G the result.

```
      ∇  G←N1  GCD  N2
[1]      G←N1
[2]      →(0=N1←G|N2)/0
[3]      N2←G
[4]      →1
      ∇
```

Now to find the greatest common divisor of 1155 and 12298, you enter those values with the name GCD between them:

```
      1155 GCD 12298
11
```

Six Possible Forms for a Function Header

A function may or may not return a result, and it may have one argument, two arguments, or no arguments. That makes six possibilities. If the header contains a specification arrow, then the function returns a result, and the name to the left of the arrow is the name used within the function to identify the result.

To the right of the arrow (if any) there may be one, two or three names. If there's only one, it is the name of the function. If there are two, the one on the right is the name of the argument, and the one on the left is the function name. If there are three, the one in the middle is the name of the function, and those around it are the names of the two arguments.

What Happens When the Computer Executes
A Function with Arguments or a Result

Consider what the machine does when you ask for an execution of the function PR. Here is the definition of PR:

```
    ∇  RR←PR R
[1]    RR←÷+/÷R
    ∇
```

Here is an instruction that calls for its execution:

```
    RESIST←PR 800 200
```

When the computer encounters the name PR, it finds that in this workspace PR is a function. Checking the header of function PR, it finds that PR has one argument, named R. So the computer creates a new local variable called R, whose value is the vector 800 1200. Then it carries out the work specified in the body of the function definition, using the new local meaning of R wherever that name may occur.

When the computer finds that it has no further work to do in the execution of PR, it again consults the header: does this function require a formal result? In our case, the answer is yes; there is a result, called RR. The computer takes the latest value of RR, and reports that as the result. What must be done with the result? The computer returns to the instruction which called for this execution of PR, and finds that the result is to be assigned as the value of a variable called RESIST, and does that.

As soon as the execution of PR is complete and its result has been reported, the variables R and RR which were created during this execution of the function have no further use. They cease to exist; they are removed from the workspace.

Suppose you had entered the instruction RR←PR R. In that case, the argument of PR happens to have the name R and the result happens to be assigned to a variable called RR. As far as the computer is concerned, it is merely a coincidence that the names are the same as those occurring in the header of PR. Your instruction refers to the meanings of R and RR outside the function. During execution, the computer goes ahead and as usual creates new local variables with the names R and RR, keeping those distinct from the meanings of R and RR outside this definition.

A Simple Function of Two Arguments:
Area of a Segment of a Circle

Suppose that you need to calculate the areas of sector of circles. For each sector, you know its radius and the angle it subtends. You would like to have a function called CA (for "circular area") so that when you need the area of a sector whose radius is 415 feet and whose angle is 42 degrees, you have only to enter the instruction:

        415 *CA* 42

The function needs two arguments and should return a result. You might as well give them names which will be easy to interpret if you subsequently check back to see what is in this definition. Let's assume that the angle is given in degrees, rather than in radians, and that PI has been assigned the value 3.14159.

        ∇ *AREA←RADIUS CA DEGREE*
    [1]   *AREA←(PI×RADIUS*2)×DEGREE÷360*
        ∇

Here is the area (in square feet) for the problem we just mentioned (415 feet, 42 degrees):

        415 *CA* 42
    63123.70607

This defined function works just as well if the arguments are arrays. However, the arguments must either have the same dimensions, or at least one of them must have only one element:

        112 240 88 *CA* 45 110 70
    4926.017281   55291.0307   4730.540405

        100 *CA* 45 55 60 90
    3926.990817   4799.655443   5235.987756   7853.981634

        10 20 30 40 *CA* 90
    78.53981634   314.1592654   706.8583471   1256.637061

        144 200 *CA* 30 45 60
    LENGTH ERROR
    CA[1] *AREA←(PI×RADIUS*2)×DEGREE÷360*
            ∧

The last example on the preceding page illustrates
several things. To begin with, you can't specify two radii
and three angles, at least not with this definition of CA.
But notice some additional points:

1. Execution of CA has not been abandoned, but
   suspended. You can take some corrective action and
   resume work.

2. Since an execution of CA has been started but not
   finished, and no more recent function is in
   execution, you can display the variables RADIUS
   and DEGREE, containing the value of the arguments
   for this execution of CA.

3. While execution is suspended, you can alter the
   definition itself, or the values of the arguments.
   In this case, it would be useful to respecify one
   or the other of the arguments so they're the same
   length, and then resume execution.

In the following example, the instruction and the
computer's response are repeated from the bottom of the
preceding page, so that the entire exchange is visible in
one place. That doesn't mean that the same problem was
started over again.

144 200 *CA* 30 45 60	Your instruction
LENGTH ERROR *CA[1] AREA←(PI×RADIUS*2)×DEGREE÷360* ∧	Error Message
RADIUS 144 200 *DEGREE* 30 45 60	You ask for display of each of the arguments of CA
DEGREE←2ρDEGREE →1	You respecify one of the arguments and resume execution
5428.672105 15707.96327	Result is printed

DEGREE
VALUE ERROR
    DEGREE
    ∧
	Now that execution is
	complete, the variable
	DEGREE no longer
	exists.

Another Example with Two Arguments:
Converting Pounds to Dollars

The British use a currency with three units: pounds, shillings, and pence. There are 12 pence in a shilling, and 20 shillings in a pound. The dollar value of a pound varies; during 1967, it went from $2.80 per pound to $2.40 per pound. Here is a function which calculates the dollar value of an expression in pounds, shillings, and pence. It is called SL, for "dollars from pounds." (S stands for dollars; the British use L for pounds.)

```
     ∇ S←RATE SL BRIT
[1]    BRIT←3ρ((0⌈3-ρBRIT)ρ0),BRIT
[2]    S←RATE×+/BRIT÷ 1 20 240
     ∇
```

The first line of the program respecifies its own argument. First it inserts up to three zeroes ahead of BRIT, so as to fill the high-order positions with zeroes if an amount is stated solely in pence, or in pence and shillings with no pounds. Then it takes the first three elements of the resulting vector. If the argument contained three elements to begin with, this won't produce any change. But if the argument stated only the pence, the argument will be respecified as a vector whose first two elements are zero.

On line two, the argument (now assured of having three elements) is divided by 1 20 240, converting all three columns to pounds. Then those are summed, and the sum is multiplied by the other argument (the exchange rate.)

```
      2.80 SL 14 7 6
40.25
```

```
      2.42 SL 10 6
1.2705
```

As written, this function won't process several different British amounts at once, since no matter how long the BRIT vector starts off, the program always converts it to three elements which it presumes to represent a single sum of money. But the function will accept any number of exchange rates:

```
      4.20 2.80 2.40 2.10 SL 0 13 8
2.87   1.913333333   1.64   1.435
```

Compound Expressions Using Defined Functions:
Another Approach to the Correlation Coefficient

The great advantage of permitting defined functions to have arguments and results is that you can use them in compound expressions, just as you can write compound expressions involving the primitive operators. As a simple example, let's return to the correlation coefficient, which we discussed earlier on pages 125-126.

The correlation coefficient is defined as the average product of two vectors of scores, provided that the scores are in standard form. You could therefore write a simple one-line program for the correlation coefficient like this:

```
     ∇ R←X CORR Y
[1]    R←AVG(STD X)×STD Y
     ∇
```

Clearly, this definition depends upon having definitions for AVG and STD; it also depends upon the fact that each of them can take a right argument and can return a result.

To define AVG, you could treat its argument as a vector, and divide the sum of the elements by the number of elements:

```
     ∇ R←AVG X
[1]    R←(+/X)÷ρX
     ∇
```

To standardize a vector of scores, first you center them (that is, reduce each of them by their average) and then you divide them by their standard deviation:

```
     ∇ R←STD X
[1]    R←(CTR X)÷SD X
     ∇
```

Centering the elements of a vector means this:

```
     ∇ R←CTR X
[1]    R←X-AVG X
     ∇
```

Finally, you need a definition for standard deviation. It is the square root of the average of the squares of the centered scores:

```
    ∇ R←SD X
[1]    R←AVG(CTR X)*2
[2]    R←R*0.5
    ∇
```

In this definition of SD, the header declares that inside this function, its result will be called R. Line 1 specifies a value for R, and then line 2 respecifies R with a new value. When this function is executed, it will return as its result the last value of R arising from this particular execution of SD (i.e. the one stored in response to line 2 of the definition).

You should keep in mind that the set of definitions just presented is devised to illustrate one approach to programming, making maximum use of sub-programs and compound expressions. For a problem of this scale, perhaps you wouldn't really want to break the main program into quite so many parts. Moreover, this particular illustration doesn't give you the most economical way of doing the work in terms of the computer's internal operations; the average of the scores is computed more than once, and there are other such minor extravagances. But these definitions do illustrate a style of programming which starts from the most general description of the procedure, and then fills in the other definitions as they are required. This makes for a highly readable program, and one which corresponds closely to the original English description of the procedure.

Notice that every one of the definitions on the last two pages uses the name X for one of its arguments. Each of these X's refers only to the argument of a single execution of that particular function. There is no problem of overlap, even though the same name occurs as an argument in each of the functions. When these functions are used to process variables stored in your workspace, there is no need for those variables to be called X--nor is there any reason why they should not be called X.

Suppose your workspace contains two vectors of scores, called H and W. You can examine their averages, their standard deviations, and the correlation between them with instructions such as those on the next page:

$(\rho H),\rho W$ 100 100	Lengths of H and W
$(AVG\ H),AVG\ W$ 801.0938 545.57689	Averages of H and W
$(SD\ H),SD\ W$ 77.41809831 49.21222692	Standard deviations of H and W
$AVG\ STD\ H$ $2.183586645E^-14$	Average of standardized H (Close enough to theoretical 0)
$SD\ STD\ H$ 1	Standard deviation of standardized H (As it should be)
$H\ CORR\ W$ 0.5312751892	Correlation of H with W
$H\ CORR\ -W$ $^-0.5312751892$	Correlation of H with -W (Same, but opposite sign)
$H\ CORR\ H+W$ 0.9276489928	Correlation of H with sum of H and W
$H\ CORR\ H-W$ 0.7758668562	Correlation of H with difference between H and W
$H\ CORR\ H$ 1	Correlation of H with itself

Changing a Function's Syntax
After You've Entered its Definition

Suppose you have entered the definition of a function without arguments or result, and you decide you would prefer to have it take an argument and return a result. Or you decide that you'd prefer to convert a function of two arguments so that it becomes a monadic function. How can you do that?

You do it by editing the header, in the same way as you edit any other line of a program (see pages 47 and 91).

Variables that are Local
To the Execution of a Function

When you write a definition so that the function has arguments or a result, you cause the creation of some variable names which are not permanently stored in the workspace, but which exist only during execution of that function. They are called local variables. The arguments and the result of a function are automatically local variables. If you wish, you can also make other variables local to a function (see the next page).

The variables named as the arguments of a function get their values as soon as the computer starts an execution of that function, even before it starts to execute line 1. The result, and any other variables local to the function, get their values only by being specified by some instruction within the function. Thus the result gets its value only if and when the instructions within the program assign it one--possibly never, if you don't include the appropriate instructions in the definition.

Global vs Local Variables

Unless the header of a function specifically indicates otherwise, APL\360 assumes that all variables are global variables. A global variable is one that is generally available to any calculation or any function in the workspace. (The only exception is this: access to the global meaning of a name is blocked while the computer is executing a function to which that name is local.) All the variables mentioned in all the chapters before this one were global variables.

A name becomes local if it is mentioned in the header of a function. Then it exists only while that function's execution remains incomplete.

Whenever a name occurs within the body of a function definition, it falls into one of these two categories:

1. If the name is local to that function (i.e. it appears in the function header), then the local meaning is understood.

2. Otherwise it refers to the next higher local meaning, if any. That is, it refers to the local

meaning in the most recently called function whose
execution was started earlier than this one, but
hasn't yet been completed. If there aren't any
such higher local meanings of a name, then the
global meaning is understood.

Displaying the Value of a Local Variable

Since the value of a local variable disappears as soon
as the computer finishes executing that function, the only
time you can ever display the value of a local variable is
while execution of the function to which it belongs is still
incomplete. That is precisely the point: the local variable
is available if you need to check up on it while debugging a
program, but doesn't clutter up the workspace when normal
execution of the function is completed.

A variable local to a function that is suspended may
still be inaccessible if its name is also local to any more
recently called function. Putting it the other way, the
value that you can use or display is always the most recent
local meaning of the name. That's also what you get if you
copy from that workspace (see the next chapter).

A local variable is not merely local to the function in
which it occurs, but local to each specific execution of
that function. If you start executing a function and it is
suspended, and then you start a new execution of the same
function and that too is suspended, you can see and use only
the most recent meanings of the local variables. Of course,
as execution of the more recently called functions is
completed, the next earlier meaning of each will again be
accessible.

Additional Local Variables
Other than the Arguments or Result

A program may involve temporary variables that are of
no further interest once execution is complete. If you
prefer, you can make them local to the function in which
they're used. As many extra names as you like can be made
local by listing them in the header, to the right of the
name of the function and the right argument (if any). They
are set off from the rest of the header, and from each
other, by semicolons (see, for example, the definition of
PRINT, p. 152).

A Mystification to Avoid

Every now and again an APL user forgets to tell the computer what should be done with a function whose execution has been suspended. Ordinarily this may not matter much, but if the suspended function uses a local variable whose name is also used for a function or for a global variable, you may think you're referring to the function or to the global variable, and instead you're getting the value of the local variable from within that suspended function. But the problem is easily avoided: don't leave suspended executions hanging around unresolved any longer than necessary. You can always check to see which functions remain suspended by displaying the state indicator (using the command)*SI*). If any of those functions contain local variables, it may be important to you to know which variables are local to which functions. For this purpose you can use the command)*SIV*. This command causes the state indicator to be displayed in the usual way, but to the right of each halted program, the computer types a list of the variables that are local to that program.

Editing the Definition of a Function
That Has Arguments, a Result, or Local Variables

When you reopen the definition of a function, whether to change it or just to display it, enter ∇ followed solely by the name of the function. You should not re-enter the entire function header. An attempt to do so will be rejected as a "definition error."

If you wish to make local variables global, or global variables local, you can do so by editing the header of the function to which they belong in the usual way (see pages 47 and 91).

Spaces Separate a Function from its Arguments

When you use functions which take arguments or return results, it is possible to construct an expression in which several names occur next to each other, or the name of a function occurs next to a number which is its argument. So you have to make clear to the computer where each name begins and ends.

A numerical digit may be part of the name of a variable or function, provided that the first character of the name is a letter of the alphabet. That means that FN6, for instance, is an allowable name, so the computer must be able to distinguish between *FN* 6 10 (meaning the function FN with an argument of 6 10) and *FN*6 10 (meaning the function FN6 with an argument of 10).

APL uses spaces as delimiters, to mark where the name of a function or variable begins and ends. When a name is used in an expression, it must be separated from another name or a number by one or more spaces. Since the symbols used for the primitive operators can never occur in names, it isn't necessary to enter spaces next to them, but you may if you wish.

26: TRANSFERRING PROGRAMS OR DATA
FROM SAVED WORKSPACES TO THE ACTIVE WORKSPACE

There are several situations in which it is handy to be able to transfer programs or data between workspaces. You may wish to load a package of programs from a public library into your own workspace, for use with the data you have already stored there. Or you may wish to set aside data you have produced, leaving it in a different workspace until at some future time you need it and recall it to the active area. APL\360 provides a family of related commands which let you copy items from saved workspaces into your active workspace, while leaving unchanged most of what was previously in the active workspace.

If you give the command to copy an entire saved workspace, the computer reads into your active workspace all of the programs and all of the global variables from whatever saved workspace you name. But, unlike the load command, copying (as far as possible) leaves unchanged the programs or variables already in your active area. For instance, the command

)COPY ACCOUNT

copies into your active workspace the definition of every program now in your saved workspace called ACCOUNT, and every global variable in ACCOUNT. (There are some things that are not copied; we'll get to that in a moment.)

Alternatively, instead of copying all the global variables and programs from a workspace, you may copy a single program or a single variable. The command

)COPY ACCOUNT OBJECT

copies into your active workspace the program (or global variable) called OBJECT from your workspace named ACCOUNT.

When a copy is completed, the computer acknowledges by typing the time and date at which the workspace from which you're copying was saved.

Grouping of Functions and Variables

It is often the case that what you want to copy is a package of related functions and data items, which are more

than a single item but less than the entire contents of a
workspace. To facilitate copying groups of items, APL\360
permits you to collect together whatever programs or
variables you wish and treat them (for some purposes) as a
single object. A group can be made up of the names of
functions, or of global variables, or of some of each. A
group can also include another group.

Suppose in your active area you have the functions
AREA, TAX, and TRAJECTORY, and the variables PAYROLL,
WITHHOLDING, and ELEVATION. You believe it would be handy
(once this workspace has been saved) to be able to treat as
a group the functions called AREA and TRAJECTORY, and the
variable called ELEVATION. You anticipate that you'll want
this group to include the variable RADII for which you
haven't supplied any value yet. So you need to create a
group comprising those members, and give it a name. Suppose
you want the group to be called TRIG. You do that by using
the command)*GROUP*, followed first by the name the group is
to have and then the names of its members. Like this:

)*GROUP TRIG AREA TRAJECTORY ELEVATION RADII*

When a group is formed, it is simply a group of names.
It doesn't matter if nothing of that name exists in the
workspace at the time the group is formed. You can define
the group in advance; then, at any subsequent time that you
enter a program or variable of that name, it will
automatically be included in any reference you make to the
group.

The useful thing about the group, of course, is that
once the workspace containing it has been saved, you can
copy the entire group simply by referring to the group name.
If the group TRIG exists in a saved workspace, you can copy
it by the command

)*COPY TRIG*

Any time you ask to copy a group, you get the following
things:

    1. The group name and the list of the names of its
       members.

    2. The definitions or values of as many of the group
       members as have them in the source workspace.

Listing the Names of Groups and their Members

The system command)*GRPS* causes the computer to type a list of the groups for which there are definitions in your active workspace.

The system command)*GRP* followed by the name of a group causes the computer to type the list of names that are members of the group (whether or not there are values or definitions for those names).

Dispersing a Group

If you use the command to form a group, name the group, but then don't list any members for it, you have dispersed the group. The various variables or functions whose names were members of the group are unchanged.

Erasing the Members of a Group

The)*ERASE* command, if applied to the name of a group, erases not only the group-name, but also the values or definitions of all the objects that were members of the group. This is a handy way to erase an entire package that is no longer wanted. But watch out: erasing a group isn't the same thing as dispersing it!

Protected and Unprotected Copying

What happens if you execute a copy, and try to copy into the workspace a program or variable whose name was already in use for some other program or variable in the active workspace?

There are two different forms of the copy command. They differ only with respect to what they do about this problem.

The command)*COPY* (the only one we've mentioned so far) causes the copied value or definition to replace whatever value or definition that object formerly had in the active workspace.

The other form of the copy command is the protected copy. The command)*PCOPY* causes objects to be copied only if their names are not in use as the names of groups, functions, or global variables in the active workspace. If an object that you asked to have copied is not copied because of this protection, the computer prints a list of the objects that weren't copied. (Even with an unprotected

copy, you can't copy an object that has the same name as a pendent function, since that would cause the definition of the pendent function to be erased, and that isn't ever permitted.)

Copying an Entire Workspace into a Cleared Workspace

You might suppose that clearing the active area (using the)CLEAR command), copying an entire workspace into it, and then naming it with the same name as the workspace from which you copied... you might suppose that that sequence would give you the same end result that you'd get if you simply loaded that workspace. But there are some differences which on occasion may be useful. They depend upon what is not copied when you copy an entire workspace.

What Is Copied

1. All global variables, all functions, and all group definitions in the source workspace.

What Isn't Copied

1. The list of functions awaiting execution in the source workspace (the state indicator).

2. Any local variable.

3. The index origin, maximum digits printed, and width of the source workspace.

4. The internal symbol table of the source workspace.

You can't see the symbol table; it is a dictionary by which the computer identifies names used in a workspace. It includes all the function or variable names ever used in that workspace, even names that have since been deleted. Even though you no longer have any use for those entries in the table, they still take up space. The symbol table has a total capacity of 256 names; if it should get full, you will encounter the message SYMBOL TABLE FULL. Hence if you have a workspace that has had many names and much use, it may be advisable to save it, enter the system command)CLEAR, and then copy it into the cleared workspace. Then you will have to drop the saved version of that workspace, and then save the active workspace to replace it.

Copying an entire workspace leaves behind the list of programs whose execution is halted. This would serve to abandon all halted program executions... but that can also be done by simply entering right-pointing arrows with nothing to the right of them as explained on p. 87.

Commands that Summarize What's In your Workspace

The following commands are useful in summarizing the contents of the active workspace:

)*VARS*   List of global variables.

)*FNS*   List of functions.

)*GRPS*   List of groups.

)*SI*   State indicator (list of halted programs).

)*SIV*   State indicator with list of local variables in each halted program.

This page intentionally left blank.

APPENDIX A: NOTES ABOUT
WHAT HASN'T BEEN MENTIONED

For the purposes of this primer, we have deliberately
refrained from mentioning some APL operators and certain
features of the APL\360 System.   To keep things   in
perspective, we present here a list of topics which have
received little if any attention in this primer.

Encoding and Decoding the Representations of Numbers

The encoding operator ⊤ converts the value of a
number into its representation in any number system.   The
left argument is a vector which specifies the base, one
element for each column of the representation. For instance,
1277 expressed in 7 columns of base 3 could be found by:

```
      (7ρ3) ⊤ 1277
1  2  0  2  0  2  2
```

Mixed bases are allowed: 105246 inches   expressed in miles,
yards, feet, and inches is found by:

```
   0 1760  3 12 ⊤ 105246
1  1163  1  6
```

The decoding operator ⊥ does the converse: it reduces a
representation vector in any number system to a value.   The
left argument specifies the base.   It may be either a vector
of the same length as the right argument, or a single number
which is then extended to match the length of the right
argument. The base-8 value of 1 7 7 6 is found by:

```
      8 ⊥ 1 7 7 6
1022
```

The number of seconds in 14 days, 12 hours, 20 minutes, and
57 seconds is found by:

```
      0 24 60 60 ⊥ 14 12 20 57
1254057
```

Encoding and decoding have a variety of uses apart from
explicit shifts of number system. For instance, the indices
of an array A might be considered as the base ρA encoding of
$\iota \times / \rho A$. The integer and fractional portions of a number N may
be separated by taking the 0 1 ⊤ N.

Factorial

In conventional notation, factorial A is written: A! Following the uniform syntax rules, APL places the operator first and its argument to the right:

$!A$

The $!$ symbol is formed by overstriking the quote and the period. When A is a positive integer, $!A$ is equivalent to $×/\iota A$.

$!A$ is also defined when A is not an integer. It is then equivalent to the gamma function:

$(!A) \equiv \Gamma A+1$

Combinations Operator

When the $!$ symbol is used dyadically, it indicates the combination operator. $A!B$ means the number of possible combinations of B things taken A at a time. Where A and B are positive integers and B is not less than A, the value of $A!B$ is $(!B)÷(!A)×!B-A$

Residue Function With Non-Integral Left Argument

The definition for residue given on page 59 does not exclude having a fractional left argument:

$1.5|4.2$
1.2

Nor and Nand

The symbols for AND $\wedge$, and OR $\vee$, may be overstruck with the NOT symbol to form NOR and NAND.

$A\barwedge B$ is equivalent to $\sim A\wedge B$.

$A\barvee B$ is equivalent to $\sim A\vee B$.

Unlike $\wedge$ and $\vee$, the operators $\barwedge$ and $\barvee$ are not associative, and in general

$\barwedge/X$ is not equivalent to $\sim\wedge/X$.

Multidimensional Arrays

Rectangular arrays can have any number of dimensions. All arithmetic operators extend automatically on an element-by-element basis to arrays of any rank (i.e. any number of dimensions).

Indexing of Multidimensional Arrays

The values of the indices in each dimension are separated by semicolons. The elements selected are those at the intersection of the specified positions in each dimension. Thus

$A[2\ 4;2\ 3\ 8]$

means those elements of A located at the intersection of positions 2 and 4 of the first dimension with positions 2 3 8 of the second dimension. If A is a 4 by 9 matrix of literals, like this:

ABCDEFGHI
JKLMNOPQR
STUVWXYZ1
234567890

the expression $A[\ 2\ 4;\ 2\ 3\ 8]$ selects from A the following matrix:

KLQ
349

The dimensions of an array produced by indexing are given by catenating the rank-vector (i.e. rho) for the indices of each dimension considered separately. Thus, if an array X is indexed by an expression in which A represents the indices for the first dimension, B represents the indices for the second dimension, and so on, like this: $X[A;B;C]$, then the dimensions of the resulting array are $(\rho A),(\rho B),\rho C$. Notice that if any of the terms A, B, C, etc. is a scalar, then ρA or ρB or ρC will be an empty vector, and hence the result will not have any extent in that dimension. For this reason, $X[2;3;5]$ is a scalar, while $X[,2;,3;,5]$ is a 1-by-1-by-1 three-dimensional array.

Matrix Products

APL provides for three general forms of matrix multiplication. The simple element-by-element product of two matrices A and B is obtained directly from the instruction

$A \times B$

It is necessary that A and B have the same rank, and the same length in each dimension, except (as usual) for the case in which either A or B has only one element.

Generalized Matrix Product

In matrix algebra, the "matrix product" (or "inner product") of two matrices A and B is found by a procedure in which the element i;j of the result is the sum of the products of the elements in the ith row of A with those in the jth column of B. APL indicates these two component operations explicitly, and writes the conventional matrix product of A and B like this:

$A + . \times B$

The advantage of this notation is that it permits the user to substitute any other dyadic arithmetic operator for + or ×, thus generalizing matrix product to permit such forms as

$A \times . \div B$ $A\lfloor .\lceil B$ $A + . \star B$ $A \vee . \neq B$

For instance, $A\lceil .-B$ returns for the i;j element of the result the maximum difference between the pairs of elements in the ith row of A and the jth column of B. $A\wedge .=B$ returns a 1 where each row of A is equal in all its elements to a column of B. Many other matrix products are possible and useful.

APL permits matrix products on arrays of any rank, subject only to the restriction that the last dimension of the left argument must match the length of the first dimension of the right argument. In the result, these matched dimensions disappear, and the dimensions of the result become all-but-the-last-dimension-of-A catenated to all-but-the-first-dimension-of-B.

Outer Product

An outer product requires that each element of A operate on every element of B. The result is a higher order array. Its dimensions are the dimensions of A catenated to the dimensions of B. Outer product is written in a form that resembles the standard matrix product, but with a null symbol ∘ replacing the first operator:

$A \circ . + B$

The outer product of the vector 2 3 4 and the vector 4 5 6 7 is the following 3-by-4 matrix:

        2 3 4 ∘.× 4 5 6 7

     8 10 12 14
    12 15 18 21
    16 20 24 28

Transposition of an Array

An array can be restructured so that its coordinates appear in a permuted order. If M is a matrix, then $\lozenge M$ transposes the rows and columns of M. If A is an array having more than two dimensions, $\lozenge A$ simply transposes the last two, leaving the others unchanged.

More elaborate transpositions may be obtained by using an explicit left argument for the transposition operator. Suppose that a result R is to be obtained from the Q transposition of an array A, by this expression:

$R \leftarrow Q \lozenge A$

Q, the left argument of the transposition operator, is a vector containing one element for each of the dimensions of A.

The values appearing in Q indicate the dimensions of the resulting array. Let $\rho\rho R$ indicate the rank of the desired result. Then the vector Q must contain values which are limited to values from the sequence $\iota\rho\rho R$, and which contain each value in $\iota\rho\rho R$ at least once.

If the same value occurs more than once in Q, it means that the indicated dimension of R is to be formed from more

than one dimension of A. Suppose, for instance, that A is a
four dimensional array, and you ask for 2 2 1 2 ⍉A.   The
result will be a two dimensional array in which the second
dimension is formed from the first, second, and fourth
dimensions of A. The elements selected are those for which
the first, second, and fourth coordinates of A have the same
value: that is, from A's diagonal through the first, second,
and fourth dimensions. A diagonal has as many elements as
the shortest of the dimensions from which it is taken.

    If A is a matrix, the main diagonal is obtained by the
instruction:

    1 1 ⍉A

Reversal of an Array
====================

    The elements lying in one of the dimensions of an array
can be restated so that they are in reverse order. If the
dimension to be reversed is not stated, it is presumed to be
the last dimension. A different dimension may be indicated
by indexing the reversal

    Suppose A is the same 4 by 9 literal matrix introduced
on page 197:

        A

ABCDEFGHI
JKLMNOPQR
*STUVWXYZ*1
234567890

        ⍉*A*

IHGFEDCBA
RQPONMLKJ
1*ZXYWVUTS*
098765432

        ⍉[1]*A*

234567890
*STUVWXYZ*1
JKLMNOPQR
ABCDEFGHI

Notice that reversing a matrix in both of its dimensions is not the same as transposing it:

$\phi\phi[1]A$

```
098765432
1ZYXWVUTS
RQPONMLKJ
IHGFEDCBA
```

ϕA

```
AJS2
BKT3
CLU4
DMV5
ENW6
FOX7
GPY8
HQZO
IR10
```

Rotation of an Array

An array may be restructured so that the elements are rotated by a specified amount in one of the dimensions of the array. For a vector, a single integer specifies the amount of rotation; a positive rotation is a left shift, while a negative rotation is a right shift:

```
3φ'ABCDEFGHIJKLMNOPQRSTUVWXYZ'
DEFGHIJKLMNOPQRSTUVWXYZABC
```

```
¯7φ'ABCDEFGHIJKLMNOPQRSTUVWXYZ'
TUVWXYZABCDEFGHIJKLMNOPQRS
```

When a multidimensional array is rotated, the computer presumes that rotation is to take place in the last dimension, unless the operator is indexed to indicate some other dimension. The coefficient of rotation may be a single number or an array whose dimensions are the dimensions of A with the dimension of rotation omitted. If a 4 by 9 matrix is rotated in its last dimension, the rotation coefficients may therefore be a single number or a four-element vector. If it is rotated in the first dimension, the coefficients may be a single number or a 9-element vector.

        0 1 2 3ϕ*A*

ABCDEFGHI
KLMNOPQRJ
UVWXYZ1ST
567890234

        1 1 2 2 3 3 4 4 5ϕ[1]*A*

JKUV67GHR
ST45EFPQ1
23CDNOYZ0
ABLMWX89I

     Rotation   provides   an   alternative   procedure   for
selecting a consecutive sequence of elements from the middle
or the end of a long vector (see pp. 141-2). For instance,
the 116 elements of V which follow element 135 can be
obtained by the expression:

     116ρ135ϕ*V*

Compressing a Multi-Dimensional Array

     Compression may be applied to an array of any number of
dimensions. If no dimension is specified, it is presumed to
be the last. The logical selection vector must have the same
length as the array in the dimension being compressed.
Consider the literal array A, which is the same 4 by 9
matrix as before:

     A

ABCDEFGHI
JKLMNOPQR
STUVWXYZ1
234567890

If the last dimension (columns) is compressed, the selection
vector must have a length of nine. Columns 3 and 6 are
omitted by the following compression:

     1 1 0 1 1 0 1 1 1/*A*

ABDEGHI
JKMNPQR
STVWYZ1
2356890

Compression along any dimension other than the last is indicated by indexing the compression symbol. In the case of a matrix, the only other dimension is the first. Compression to remove the second of the four rows of A is achieved this way:

          1 0 1 1/[1]A

ABCDEFGHI
STUVWXYZ1
234567890

Expansion of an Array

An array A may be expanded in one of its dimensions by the insertion of zeroes (or blanks, as appropriate) in designated positions between the elements. The expansion operator is the backslash \, with a left argument that is a logical vector. The number of ones in the left argument must be the same as the length of the dimension of A that is being expanded. The zeroes in the left argument indicate where the extra zeroes or spaces must be inserted.

Expanding a numerical vector:

          1 1 0 1 0 1\1.1 1.2 1.3 1.4
1.1 1.2 0 1.3 0 1.4

Expanding a literal vector:

          1 1 0 1 0 1\'*ABCD*'
AB C D

Expanding the literal matrix A in its last dimension (columns):

          1 1 0 1 0 1 1 1 1 0 0 1 1\A

AB C DEFG   HI
JK L MNOP   QR
ST U VWXY   Z1
23 4 5678   90

Random Numbers: Roll and Deal

You can generate random numbers by using the operator whose symbol is *?*. This operator has two forms, one monadic and the other dyadic. The monadic operator *?* generates an array of independent random integers—the sort of thing you might get from rolling a handful of dice. The dyadic operator *?* produces a number of elements randomly selected from a single population of consecutive integers—the sort of thing you might get by dealing from a deck of cards.

First consider <u>roll</u>. In the expression

?B

each element of the array B must be a positive integer. The result is an array of the same dimensions as B, with each of its elements a random integer in the range between 1 and the value of the corresponding element of B. For instance, the instruction *?6 10* produces a vector of two elements, the first between 1 and 6, and the second between 1 and 10.

```
      ?6    10
1  8
```

```
      ?4  7ρ100

   46   54   22    5   69   68   94
   39   52   83    4    6   53   68
    1   39    7   42   69   59   94
   85   53   10   66   42   71   92
```

Now consider <u>deal</u>. Both the left and the right argument of *?* must be a single positive integer. The result of *A?B* is a random deal of *A* elements from the population ι*B*. That means that the elements are selected from ι*B* without replacement, so that no two elements of the result are ever the same. Thus in the expression *A?B*, *A* may not be greater than *B*. And if *A=B*, you are getting a random permutation of ι*B*.

```
      2?5
4  1
```

```
      10?10
7  3  2  6  1  9  10  5  4  8
```

System and Program Information

APL\360 permits you access to several items of information concerning the status of programs in your workspace and the status of the APL System. These are called "I-beam" functions; the I-beam symbol ι is formed by overstriking $\top$ and $\bot$. The following I-beam functions are currently available to users:

ι19 Cumulative keyboard-unlocked time since sign-on in 60ths of a second

ι20 Time of day in 60ths of a second.

ι21 Your compute time since sign-on in 60ths of a second.

ι22 The amount of unused space remaining in your workspace, in bytes.

ι23 The number of users currently signed on.

ι24 Your connect time since sign-on in 60ths of a second.

ι25 Today's date.

ι26 The number of the line of the function currently being executed.

ι27 A vector containing the line numbers of all functions whose execution has been started but not yet completed, with the most recent function first.

Library Functions

Each APL\360 System includes a common library of workspaces containing defined functions for a variety of special purposes. These workspaces may be loaded into your active area, or individual functions from within them may be copied and incorporated into your own library. The use of these pre-written library functions in effect provides extra operations, in addition to those available as primitive operators. While the contents of individual libraries may

vary, they may likely include provision for such things as trigonometric functions, complex arithmetic, graph plotting, format control for output, text editing, matrix inversion, and so on. Some systems may use these libraries for notices of system modification or the posting of schedules of operating hours. For details of the library functions available, you should consult the operator of the system you are using.

As distributed, APL\360 provides a public library space 1 *WSFNS* which contains the locked definitions of a number of special functions. These functions have the same effect as several of the system commands for workspace control, such as control of the maximum number of digits printed in displays of numbers, the index origin, width of printed output, and so on. The fact that these are defined functions means that you can copy them into your own workspace, and then call for their execution in programs that you write yourself, thereby giving a program control of these matters. By contrast, the usual system commands can never be included as part of a program's definition.

Additional System Commands

Change of Origin

Ordinarily APL operates in 1-origin indexing. However, the system command *)ORIGIN* 0 sets the index origin to 0. The first element of a vector is now *A[*0*]*, and the index-finding and index generating operators both start counting at 0 rather than at 1. The following operators follow the index origin: random; indexing; index-finding and generating; transpose.

The index origin is not changed by saving a workspace. If desired, the index origin may be reset to 1 by the command *)ORIGIN* 1.

Locking a Function

The definition of a function may be locked. Once it is locked, it can not be displayed or edited; when an error is encountered within it, the computer reports the line on which it was working when it found it was unable to continue, but does not print the offending line.

A function is locked by closing its definition with ⍫, formed by a del overstruck with the tilde (for "not"), at the time of its initial definition or at any subsequent time when the definition has been reopened. Locking a function is not reversible. A locked function may be copied, and may be deleted, but it may never again be displayed or edited.

Locking a Workspace

When the command to save a workspace is given, you may add a password. The workspace can not be loaded, nor can anything in the workspace be copied, unless the subsequent *)LOAD* or *)COPY* commands are accompanied by the password. The command

)SAVE WORK:SESAME

saves the workspace under the name WORK and establishes the password SESAME. Subsequent *)LOAD* or *)COPY* commands must include the colon and the password. The password remains in effect until the next time that the workspace is saved.

The system operator has no way of finding out what password was used, and no means to override a workspace password. There is no remedy for a lost password. However, a locked workspace may be dropped even if you don't know its password.

Width Control

The width of the output field to the terminal may be set by the command)WIDTH followed by an integer in the range 30 to 130. This controls the maximum length of an output line in that workspace. When you sign on, the initial empty workspace has a width of 120 characters.

Setting the width has no effect on the line-length you may type in, and no effect upon the length of messages sent to you from the operator or from other users, and no effect upon register dumps caused by system malfunction. But all other output is broken into lines that fit within the limits set.

Messages Between Terminals

You can exchange brief messages with other users of the system while they are signed on. Messages are directed by the number of the line to which the users are attached (generally identified by the numbers 1 up to the maximum capacity of the system). The command

)MSG 68 HAPPY BIRTHDAY!

sends the message to the terminal currently signed on to port 68.

Messages may or may not wait for a reply; the distinction is the same as for messages directed to the computer operator (see p. 103).

When you receive a message, the source terminal's line number is typed first, followed by a colon and then the message. If the terminal sending the message expects a reply, the message when delivered is preceded by an underlined R.

Identifying Who Else is Signed On

The system command)PORTS gives you a list of the ports currently in use, plus the first three characters in the names of their current users.

APPENDIX B
SUMMARY OF SYSTEM COMMANDS

To illustrate the way the various system commands work, the summary that follows shows them arranged into groups according to the type of effect they produce. To illustrate the format in which they are used, samples are shown. These samples refer to various fictitious names. Naturally, you should substitute for these names the ones that make sense in your context. The names used are as follows:

123456789	Sign-on and library number of a fictitious user.
601	Number of a fictitious public library.
WSNAME	Name of a workspace.
FUNC	Name of a function (program).
VARB	Name of a global variable.
*GP*1	Name of a group.
*OBJ*1	Name of a global object (could be a function, a group, or a global variable).
*OBJ*2	Name of another global object.
PSST	Password used as key to a locked workspace.
SESAME	Password used as key to a locked sign-on number.

Terminal Control Commands

1. Sign On. Right parenthesis followed by user number.

)123456789

2. Sign On with Key. If the sign-on number is locked, the sign-on number must be followed with a colon and then the key you specified earlier.

)123456789:*SESAME*

3. Sign Off. At the time you sign off, you may or may not exercise any or all of three options: (a) to have your active workspace saved under the name CONTINUE; (b) to set a new lock for subsequent sign-ons; (c) to have the computer hold the telephone connection (if you're using a dial-up system) for 30 seconds in order to receive a new sign-on. That means there are 2×2×2 ways to sign off. The three options are listed separately, and then their eight permutations are shown.

3a. Saving the Active Workspace for Continued Work. If you want to go right on working where you left off at the start of the next work session, you can sign off with the command)CONTINUE. This causes the active workspace to be saved under the name CONTINUE. If the active workspace was loaded from a locked workspace, the same lock now applies to the workspace named CONTINUE. Next time you sign on, if CONTINUE was not locked, it will be loaded automatically as soon as you sign on for the next work session. But if you don't want to have the active workspace saved, you just sign off with the command)OFF. Then the active workspace is lost.

3b. Holding the Telephone Connection. If the word HOLD appears following either)OFF or)CONTINUE, the telephone connection (if there is one) is held for 30 seconds. Otherwise the telephone connection is broken following sign-off.

3c. Establishing a New Sign-On Lock. If the sign-off command (whatever it is) is followed by a colon, the one word after the colon becomes the new lock thenceforth in effect at sign-on. If there's nothing after the colon, then no password will be required.

The eight possible sign-off commands are therefore:

)OFF

)OFF:SESAME

)OFF HOLD

)OFF HOLD:SESAME

)CONTINUE HOLD

)CONTINUE HOLD:SESAME

4. Statement of Time and Work. Any of the eight forms of sign-off causes the computer to type a statement of the time at which you were signed off, and the time you were connected, and the time used by the central processing unit (CPU) during actual calculations. These are also shown cumulatively since the date of the last system accounting.

5. Forced Sign-Off. If there is a break of more than a few seconds in your line connection to the computer, or if the System Operator in some emergency situation so directs, the computer executes an automatic *)CONTINUE* on your behalf. However, you should <u>not</u> sign off by simply hanging up the telephone or turning <u>off</u> the terminal, since this leaves the line you were using open to calls dialled in during the few seconds before the computer decides you must have been disconnected.

Workspace Control Commands

The following commands affect only the active workspace.

1. Clearing the Active Workspace. The command

)CLEAR

clears the active workspace. The resulting active workspace contains no definitions, and has origin 1, width 120, and prints up to 10 digits in numerical output.

2. Loading a Workspace. The load command has the following format: First, the word LOAD; second, the library number from which a workspace is to be taken (if that is not your own library); third, the name of the workspace; fourth, (if the workspace is locked) the key, consisting of a colon and the password. For instance:

)LOAD WSNAME	Load the workspace WSNAME from your own private library.
)LOAD WSNAME:PSST	Load the locked workspace WSNAME from your own private library.
)LOAD 601 WSNAME	Load the workspace WSNAME from public library 601.

3. Copy. The copy commands cause some but not all of the material in a saved workspace to be transferred to the active workspace. The source workspace (i.e. the one from which the copied material is taken) is identified in the same way as for the load command.

If another name appears after the workspace identification, the named object (global variable, function, or group) is copied. But if no name appears after the workspace identification, all global variables, functions, and groups in the source workspace are copied.

During a protected copy, objects are copied only if their names are not already in use as the names of global variables, functions, or groups in the active workspace. But during an unprotected copy, a copied object replaces any object having the same name in the active workspace.

Some possible copy commands are as follows:

)COPY WSNAME   Copy all global objects from your workspace named WSNAME.

)PCOPY WSNAME   Copy from WSNAME all global objects whose names don't overlap with the names of global objects in the active workspace.

)PCOPY 601 WSNAME:PSST GP1   Copy from locked workspace WSNAME of public library 601 the membership list of the group GP1, plus the definitions of all of the functions and global variables named within GP1, or named in any of the groups named within GP1, provided that their names don't overlap the names of global objects in the active workspace.

4. Forming and Dispersing Groups. This command causes a group to be formed from the functions, global variables or groups named.

)GROUP GP1 FUNC VARB   Form a group with the name GP1, composed of the names FUNC and VARB, and whatever values or definitions those may have in the active workspace now or later.

```
)GROUP GP1 GP1 OBJ1        Redefine   the   group   GP1   so
                          that   it includes   all   the former
                          members of GP1 and also the object
                          OBJ1.

)GROUP GP1                 Redefine   the   group   GP1   so
                          that it has no members  (i.e. dis-
                          perse GP1).
```

6. Erase. Global variables, functions or groups may be erased by the command)*ERASE* followed by all the objects to be erased. The command

>)*ERASE FUNC VARB GP1*

causes the erasure of the function FUNC, the variable VARB, the group definition GP1, and the values or definitions of all the members of GP1 which are now in existence. You are not permitted to erase a pendent function.

7. Set Index Origin. The index origin for array operations may be set to 0 or 1 by the command

>)*ORIGIN* 0   or   )*ORIGIN* 1

8. Set Digits. The maximum number of digits printed for a numerical value is normally 10. It may be reset to any value between 1 and 16 by a command such as

>)*DIGITS* 12

9. Width Control. The maximum number of characters on a line of output may be set to any integer between 30 and 130. The clear workspace uses a width of 120. To set width to 80:

>)*WIDTH* 80

10. Renaming the Active Workspace. If the command)*WSID* is used with a name after it, it serves to rename the active workspace with the indicated name. This does **not** alter the lock for this workspace, if any.

Following a change of origin, width, digits, or workspace identification, the system reports what these were before the change.

Library Control Commands

Workspaces may be placed in a library by the save command, and removed from a library by the drop command. Loading or copying material from a library workspace has no effect upon the library, but serves to duplicate in the active workspace material from the library workspace.

1. Saving a Workspace. The command

)SAVE

causes your active workspace to be saved with whatever name and whatever lock it now has. Until renamed, a workspace bears the name of its source. The command save cannot be used for a workspace with no name, or a name from a library other than your own.

The command

)SAVE WSNAME

assigns the name WSNAME to the active workspace, and saves the workspace under that name in your private library. But the command can't be accepted if it would result in changing the name of the active workspace to a name already in use as the name of a workspace in your library.

The command

)SAVE WSNAME:PSST

assigns the indicated lock at the same time that the workspace is named and saved.

2. Dropping a workspace. The command

)DROP WSNAME

removes that workspace from your library. Locks and keys are not used with the drop command. Dropping a workspace from your library has no effect upon the active workspace.

Inquiry Commands

1. Library List. You get a list of the names of work-spaces in your own library by the command

)*LIB*

2. Library List for a Public Library. You get a list of the workspaces saved in a public library by inserting its library number after the command)*LIB*

)*LIB* 601

3. State Indicator. You get a list of all halted functions in your active workspace by the command

)*SI*

The list is typed with the most recently halted function at the top. The number in brackets is the line about to be exe-cuted. Functions marked with an asterisk are suspended.

4. State Indicator with Local Variables. The command

)*SIV*

gives you the state indicator as before, but with a list of the variables local to each of the halted functions.

5. Name of the Active Workspace. The command

)*WSID*

gives you the name of the active workspace.

6. Names of Objects in the Active Workspace. There are three commands which cause the computer to type a list of objects defined in the active workspace. The list is printed in alphabetical order. If you name any single letter of the alphabet after one of these commands, the list is typed in order starting with the letter you named.

)*VARS*	List of global variables.
)*FNS*	List of functions.
)*GRPS*	List of groups.

)*VARS R*   List of global variables,
            in alphabetic order starting with
            the letter R.

7. List the Members of a Designated Group. The command

)*GRP GP*1

gives you the list of the members of the group called GP1.
Note that the fact that a name appears on this list is not
proof that a function, group, or variable of that name has
been defined.

8. Ports in Use and their Users. The system command

)*PORTS*

gives you a list (in numerical order) of the computer's
ports currently in use, and the first three letters of the
identification of their current users. For the sake of pri-
vacy, the system refrains from giving the full name.

9. Port of a Specific User. If you'd like to know at
what port (if any) a particular user is signed on, you can
inquire by using the command)*PORT* followed by the first
three letters of his name. For instance:

)*PORT PCB*

Communication Commands

)*OPR ANY TEXT*   Send message to Operator; keep
            keyboard locked to receive reply.

)*OPRN ANY TEXT*   Send message to Operator; no reply.

)*MSG* 123 *ANY TEXT* Send message to indicated port;
            lock keyboard to await reply.

)*MSGN* 123 *ANY TEXT* Send message to indicated port; no
            reply.

Incoming messages awaiting reply are prefixed by R̲:

The system types *SENT* when an outgoing message has been sent.

APPENDIX C: TABLE OF APL OPERATORS

Standard Scalar Operators

The following operators return a scalar result when their arguments are scalars. They may also be applied on an element-by-element basis to arrays of any rank, provided that where used dyadically, either both arguments have the same rank and the same length in every dimension, or that at least one argument has only one element. Any of the scalar dyadic operators may be used in reduction, or in the generalized inner and outer products.

$X+Y$	X plus Y
$+Y$	Y (no change)
$X-Y$	X minus Y
$-Y$	Minus Y
$X \times Y$	X times Y
$\times Y$	Signum of Y
$X \div Y$	X divided by Y
$\div Y$	Reciprocal of Y
$X * Y$	X to the Yth power
$* Y$	e to the Yth power
$X \lceil Y$	Maximum of X and Y
$\lceil Y$	Ceiling of Y
$X \lfloor Y$	Minimum of X and Y
$\lfloor Y$	Floor of Y
$X \mid Y$	X residue of Y
$\mid Y$	Absolute value of Y
$X \circledast Y$	Log of Y to the base X
$\circledast Y$	Natural log of Y

$\circ Y$	PI times Y
$X \circ Y$	Trigonometric functions and inverses
$X ! Y$	Number of combinations of Y things taken X at a time
$! Y$	Y factorial; Gamma of Y-1
$? Y$	Random equi-probable selection of an integer from ιY
$X < Y$	X less than Y
$X \leq Y$	X less than or equal to Y
$X = Y$	X equals Y
$X \geq Y$	X greater than or equal to Y
$X > Y$	X greater than Y
$X \neq Y$	X not equal to Y
$X \wedge Y$	X and Y
$X \vee Y$	X or Y
$X \vee Y$	Neither X nor Y
$X \wedge Y$	Not both X and Y (X nand Y)
$\sim Y$	Not Y

Generalized Matrix Operations

In the entries below, the symbol $\circledcirc$ stands for "any standard scalar dyadic operator."

$X + . \times Y$	Ordinary matrix product of X and Y
$X \circledcirc . \circledcirc Y$	Generalized inner product of X and Y
$X \circ . \circledcirc Y$	Generalized outer product of X and Y

Generalized Reduction

⊙/Y	The ⊙ reduction along the last dimension of Y
⊙/[Z]Y	The ⊙ reduction along the Zth dimension of Y

Compression and Expansion

X/Y	X (logical) compressing along the last dimension of Y
X/[Z]Y	X (logical) compressing along the Zth dimension of Y
X\Y	X (logical) expanding along the last dimension of Y
X\[Z]Y	X (logical) expanding along the Zth dimension of Y

Other Operators

$X\rho Y$	Reshape Y to have dimensions X
ρY	Dimension of Y
$X[Y]$	The elements of X at locations Y
$X\iota Y$	Locations of Y within vector X
ιY	The first Y consecutive integers (follows index origin)
$X\epsilon Y$	Each element of X is a member of Y
$X\top Y$	Representation of Y in number system X
$X\bot Y$	Value of the representation Y in number system X
$X?Y$	X integers selected without replacement from ιY
$X\phi Y$	Rotation by X along the last dimension of Y
$X\phi[Z]Y$	Rotation by X along the Zth dimension of Y

ϕY Reversal along the last dimension of Y

$\phi[Z]Y$ Reversal along the Zth dimension of Y

$X\phi Y$ Transpose by X of the coordinates of Y

ϕY Ordinary transpose of Y (tranposing last two coordinates only)

X,Y Y catenated to X

$,Y$ Ravel of Y (make Y a vector)

$X\uparrow Y$ Take the first (or last) X elements of Y

$X\downarrow Y$ Leave the first (or last) X elements of Y

$X\leftarrow Y$ X specified by Y: the name X receives the value of Y. Value and dimensions of Y are passed on unchanged to to the next operator to the left of X, if any.

ΔX Grade up of X

∇X Grade down of X

Symbols Having Special Functions

The following symbols are not operators, but may appear in APL expressions with the sense indicated below:

( )   Parentheses. Expression within them is to be evaluated before being used as the argument of an operator or defined function

$\rightarrow X$ Branch to X. Where X is a scalar or a vector, branch to $1\rho X$; where X is an empty vector, go to the next line in sequence.

$\square \leftarrow X$ Print the value of X. The value of X is also passed on to the next operator further to the left.

$X\leftarrow \square$ Request input. Value of $\square$ is the resulting value after expression entered is evaluated.

$X\leftarrow \square$ Request input. Value of $\square$ is entire input text as literal characters, up to but not including carrier return.

$'XYZ'$ The literal characters XYZ.

Comment. Precedes an unexecuted line of comments
in a function definition.

Illegal character having the special property of
halting a request for literal input. Formed by
overstriking the characters *O U T*.

APPENDIX D
TRIALS AND ERRORS

One of the advantages of a conversational computing system is that it becomes very easy to experiment. If you don't know what the result of an instruction will be, you can try it and see. Indeed, you could discover for yourself the effect of all the various APL operators just by trial and error--plus a certain amount of patience and ingenuity. "Try it and see" is a practical strategy with a conversational computing system because the result comes back so rapidly. If you're in doubt about what the computer does in some particular case, or what an unfamiliar operator does, you are encouraged to experiment.

Of course, trial and error does entail some risks. One risk is that you will incorrectly generalize the results of your experimentation; that's why primers and manuals exist. Another risk is that you run into some errors that you don't understand, because they go beyond the topics to which you've been introduced. This appendix lists a great many of the possible error messages, even including some whose significance may not at first be clear, and some which are not otherwise discussed at all in this primer.

Form of Error Messages

When you enter an instruction, first the computer has to read it, then it has to execute it. Two types of errors arise because the computer is unable to read your instruction. One is a transmission error, which may be due to electrical faults or noisy transmission lines. The other is a character error, which arises when the transmission is technically adequate, but still doesn't refer to an allowable APL character.

Once the computer has received your instruction, it starts work on executing it. If you have entered an instruction which the computer cannot execute, it stops work on that instruction and sends you an error message. Each error message consists of three lines. The first identifies the type of error that the computer has encountered. The second restates the instruction as the computer understands it. The third shows a caret mark indicating where the computer ran into trouble. If the trouble was an instruction that could not be executed, the caret shows how far the computer had worked (proceeding from right to left) when it found it could go no further.

Errors Are Described from the Computer's Point of View

Errors arise in various ways. You may have misunderstood the proper use of an operation. You may have tried to carry out a sequence of instructions in the wrong order. You may have forgotten what value is associated with a variable. A great many errors are simply mistypings. The computer, of course, has no way of knowing what you intend. It executes each of the instructions you enter as best it can, until it encounters something that it cannot execute. Then it reports the trouble that it has encountered. The computer's classification of the error is, of necessity, written from its own point of view, since it can't very well guess how the error departs from what you privately had in mind.

For example, if you misspell the name of a variable, the computer may read this as a reference to some other variable, and it will report an error only if the value of that other variable makes the instruction impossible to execute. It can't stop and tell you "spelling error," even if that is how the error really arose.

Similarly, if you put a parenthesis in the wrong place, or leave one out by mistake, the computer can only tell you what problem it encountered as it tried to execute the instruction that you did enter. Thus, while the computer reports the type of error it has found, it can't tell you what you should have typed; you have to figure that out for yourself.

Generally speaking, when the computer finds an error in an instruction, you have to reenter the entire instruction. The value of an intermediate expression within the instruction is not saved, unless of course your instruction specifically directs that it should be stored as the value of a named variable. This arises only when there is a specification arrow further to the right (and hence executed earlier) than the caret that indicates where the trouble is. If the result of an intermediate step has been stored, you need only reenter (correctly!) the part of the instruction that appears to its left.

If the instruction that's causing the trouble is a line within a program, you may ask to have the line retried-- presumably after you've taken some steps to correct whatever was wrong. Correcting and restarting a program are discussed in Chapter 14.

Resend (Transmission Error)

Because of some malfunction of equipment between you and the computer, the computer may receive a garbled or unreadable transmission. When this happens, the computer immediately requests that your last transmission be repeated.

If you are using a 1050 terminal, you will notice that a little lamp marked RESEND lights up and stays lit. (It remains lit until you manually extinguish it, by pressing the button beside it.) The carrier does not indent, but this time unlocks at the left margin, waiting for you to retype the last transmission.

The 2741 terminal does not have a resend lamp, so the computer types the word RESEND. Here again the keyboard unlocks without the usual indentation of six spaces, and waits for you to reenter your last instruction--that is, to be more precise, you must reenter everything since the last time you pressed the carrier return key.

Character Error

Your message contains an illegal overstrike. The computer types back as much as it can of the instruction as received, up to the first unacceptable character. The computer makes no start on executing any part of an instruction containing a character error. The caret mark indicates where the instruction is unreadable rather than where it is unexecutable. You have to reenter the entire instruction.

Value Error

Your instruction refers to a variable for which no value can be found in this workspace. This may arise because you have failed to assign a value to that variable, or because you have misspelled the variable name so that the computer does not recognize it. In that case, you may correct the situation by entering a value for the missing variable, or correcting the misspelled name.

You may also encounter a value error if you have confused the local and global meanings of a name, and are

getting one when the other was intended. See the discussion of local variables, and an avoidable mystification, page 186. Displaying the state indicator)*SI* and branching to 0 may resolve the difficulty by taking you out of the function to which the name is local.

Value errors may also arise if you attempt to make use of the result of a defined function, but the function definition fails to provide one. You can remedy this by rewriting the function definition so as to provide an explicit result, or (if it already has one) by making sure that the body of the definition in fact specifies a value for the result before execution of the function is complete.

Domain Error

You have entered an instruction which asks an APL operator to operate on a value outside the domain that that operator can handle. You may get a domain error if you try to divide by zero, or to do arithmetic on a value which is not a number, or to perform an operation whose execution would develop a result too large to be handled. You will also get a domain error if you attempt to catenate a literal vector with a numeric vector, or to insert literal elements into a numeric array, or numeric elements into a literal array.

Syntax Error

You have entered an expression whose syntax is impossible. Some examples of impossible syntax are:

1. Two variable names are juxtaposed with no indication of the operation that is to be performed on them.

2. An operator symbol is used with no indication of a value on which it is to operate.

3. A parenthesis or bracket is opened but not closed, or closed but not opened.

4. A defined function is used in a way that is inconsistent with the syntax specified in its header.

You will have to correct the instruction and reenter it.

Rank Error

The rank of a variable is the number of dimensions it has. You have entered an instruction which uses variables of different rank for an operation which requires that the ranks be matched, or you have used a variable whose rank is too large for the particular operation. While the scalar operators extend to arrays of any rank, a number of the other operators, such as ⊥, ⊤, or ι, can take arguments only of rank 1 or rank 0.

Length Error

You have entered an instruction involving two arrays whose lengths do not match properly.

Definition Error

You have entered an instruction which employs the symbol ∇ improperly. This symbol is used to begin the definition of a function, or to revise (edit) an existing function. A definition error is reported if you use a ∇ in one of the following improper ways:

1. The ∇ is not the first character in the instruction.

2. You attempt to reopen the definition of a function whose execution has been started but not finished, and which is not suspended (i.e. it is waiting for the result of some other function which it calls and which has been suspended). Check the state indicator by entering)SI.

3. You attempt to start a new definition for a function whose header contains a result, an argument, or a local variable when a definition for a function of that name already exists in the workspace.

4. While in definition mode, you enter a defective request to edit a line of the function.

Depth Error

It is possible to start the execution of one program before the execution of earlier programs is complete. This may happen if you suspend execution of a program and enter a new instruction from the keyboard, or if one program itself contains instructions calling for the execution of other programs, or of itself. A depth error occurs when too many calls to execute functions are stacked up at one time. The two common causes of depth errors are these:

1. You have a program which inadvertently calls for its own execution and produces an infinte recursion. This occasionally arises because you enter definition mode to display the definition, find the definition satisfactory, and then enter the name of the function in order to execute it. If you did this without leaving definition mode, you have appended to the definition a new line calling for its own execution.

2. After a program is suspended in mid-execution, you keep starting new executions without disposing of those already started but not yet completed.

After a depth error, the easiest way to get rid of all those pending executions is to enter a single right pointing arrow, which serves to clear the state indicator. The same effect is also achieved by saving the workspace, clearing, copying the saved version, and then renaming the active area appropriately.

Label Error

You have used a colon improperly. Within the definition of a function, the colon separates a label from the instruction. Only one colon may appear on any line, and it must have one and only one label to the left of it. The name may not be a name that is already in use as the name of a function; it is unwise to use as a label a name that has any other use in the same workspace. Any use of a colon outside definition mode gives rise to a label error.

WS Full

You have entered an instruction which requires more storage in your workspace than is now available. This may arise because you have assigned to a variable a value that involves a large array, or because some of the intermediate steps in your calculation (even though not assigned to a variable) require too much space even for temporary storage during the calculation. You should check over the list of variables in the workspace to see if some may be removed when no longer needed. You should also check the state indicator for functions whose execution is suspended, since space is required for the values of their arguments or of other local variables occurring within them, and these are stored separately for each execution of a function.

System Error

The computer has detected some fault in its own internal operations, either because of mechanical difficulties or because of errors in its own system program. System errors should be brought to the attention of the system operator. As an aid to diagnosis, in several instances the computer will cause the typing of hexadecimal numbers representing the state of the computer when the trouble was encountered. This block of eighteen 8-character words consisting of the numeral 0-9 or the letters A-F is called a "register dump." The register dump, together with the printout showing what you were doing before the trouble occurred, should be sent to the system operator.

APPENDIX E
EQUIPMENT YOU NEED TO USE APL\360

The physical equipment needed to use APL\360 is a terminal equipped with an APL typing element, and a means of connecting the terminal to the APL computer. APL\360 can be used with either of two kinds of terminal: the IBM 2741 or the IBM 1050. Each of these includes a tele-processing typewriter, equipped to transmit what you type to a remote computer, or to receive and type messages transmitted from the computer.

The 1050 System

In addition to the basic keyboard, typewriter, and the associated electronic controls, the 1050 System may be obtained with a card reader, a punch which may be operated either from the 1050 keyboard or on-line from the computer, or with units for punching or reading paper tape. Whether or not these will be useful will depend upon the type of use you have for them; if you want to process data which are readily available in punched form, the reader may be useful. If your hours of access to the APL computer are limited, it may be useful to punch cards off-line and then read them. But you should note that the APL\360 System is primarily a conversational system, neither requiring nor adapted to batch-processing from a terminal.

The APL\360 system does not explicitly select one or another form of input or output, but accepts input from, or transmits output to, whichever units of the 1050 system are connected at a particular time.

Note that cards or tape punched elsewhere must use an encoding compatible with the reader at which they are read. Because APL\360 is a conversational system, each input, even those from card or tape readers, must include the symbols for "carrier return" and "end of block" (in that order). If these codes are not generated automatically by the terminal equipment (as they may be in some installations), they must be typed or punched explicitly.

It is preferable to use a 1050 terminal equipped with Time-out Suppress Feature, but not equipped with the Request Feature. An automatic end-of-block signal is useful only if the EOB signal is emitted after the carrier return.

The APL typing element for use with the 1050 terminal is part number 1167988.

The 2741 Terminal

The 2741 terminal is basically a typewriter, and includes no provision for card or paper — tape attachments. Although the typing speeds of the two types of terminal are identical, the 2741 eliminates delay between the typing of successive lines, and therefore has a markedly faster response for conversational use, or for the typing of output involving many lines.

It is recommended that a 2741 terminal intended for use with APL 360 be ordered with the Interrupt Feature.

The 2741 terminal is manufactured with either of two systems for encoding the typing element. For terminals built with the standard SELECTRIC®keyboard correspondence, the APL typing element is number 1167987. Such a terminal may also use any of the typing elements intended for SELECTRIC typewriters.

Some 2741 terminals are built with the PTTC/EBCD correspondence, which is also employed in the 1050. These terminals require typing element number 1167988, and are compatible with elements used in some other computer systems.

When a new terminal is delivered, it is provided with keytops to match the typing element with which it was ordered. Terminals ordered with an APL typeball will show APL characters on the keytops. Terminals which were ordered with some other typing element, and which therefore have keytops showing a different character set, may be converted with a stick-on conversion kit, or by use of an overlay or a keyboard map.

Coupling to the Transmission Line

A device is required to couple the terminal to the line running to the central computer. Where this is done by dial-up over telephone circuits, a Western Electric 103-A2 Dataset may be used. Other devices providing the same encoding are possible, including acoustic couplers which may be used with any voice telephone circuit, and which are not electrically connected to the telephone equipment.

The equipment needed to couple the terminal to the line may depend upon the equipment used by the central computer, so you should check this out with the installation to which you expect to be connected. In some cases, direct wiring rather than a telephone circuit may be possible, and then you'd need the appropriate modulator-demodulator instead of a Dataset or acoustic coupler.
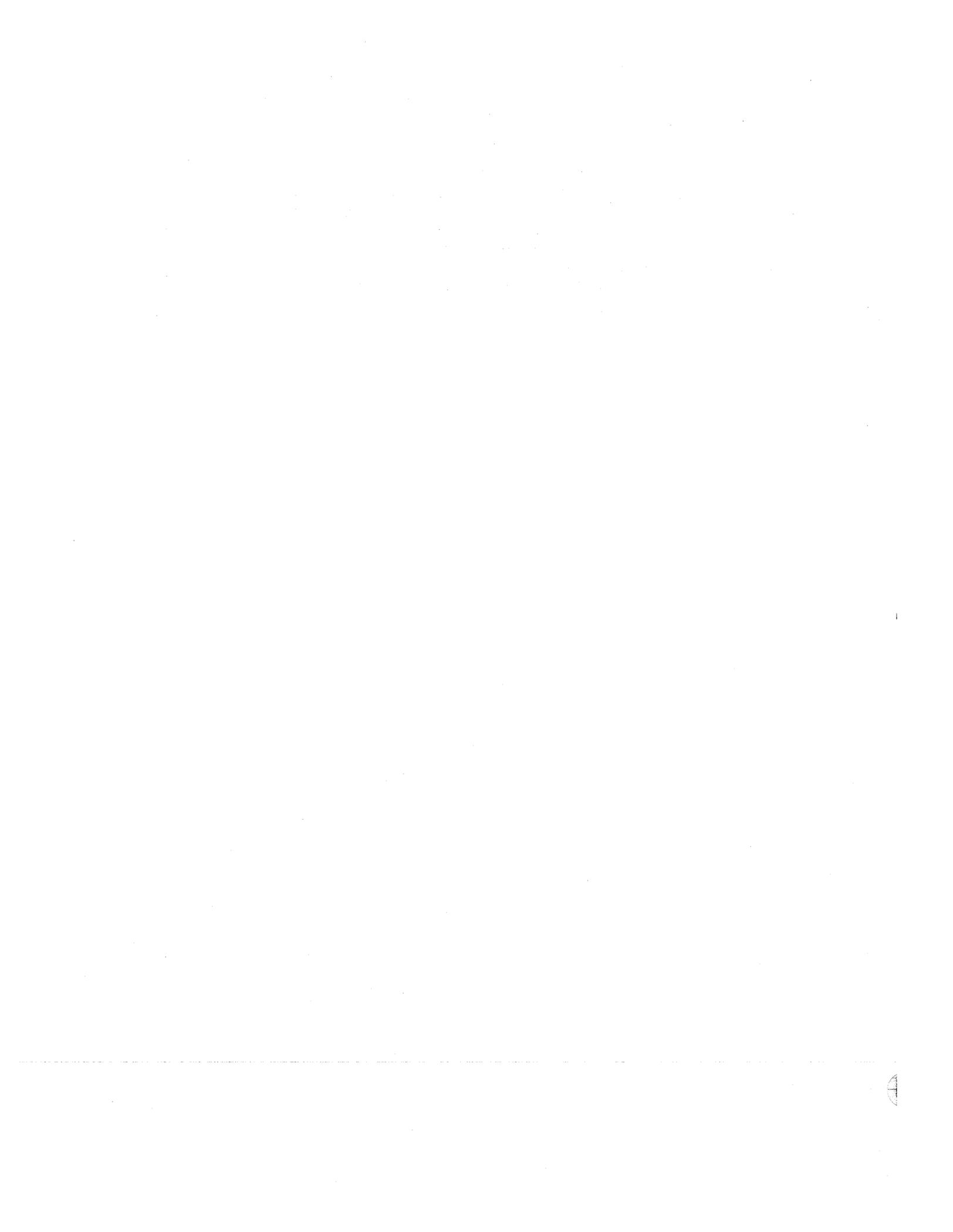
INDEX

A table of all the system commands appears on pages 209-216 and a table of all the APL operators on pages 217-221. The index does not contain any references to the entries in those tables.

GH20-0689-1

APL）SCU Primer Printed in U.S.A. GH20-0689-1

APL\360 Primer   GH20-0689-1

Program Numbers 5734-XM1 and 5736-XM1

Please comment on the usefulness and readability of this publication, suggest additions and deletions, and list specific errors and omissions (give page numbers). All comments and suggestions become the property of IBM. If you wish a reply, be sure to include your name and address.

COMMENTS

fold   fold

fold   fold

● Thank you for your cooperation. No postage necessary if mailed in the U.S.A.
  FOLD ON TWO LINES, STAPLE AND MAIL.

GH20-0689-1

YOUR COMMENTS PLEASE...

Your comments on the other side of this form will help us improve future editions of this publication. Each reply will be carefully reviewed by the persons responsible for writing and publishing this material.

Please note that requests for copies of publications and for assistance in utilizing your IBM system should be directed to your IBM representative or the IBM branch office serving your locality.

fold   fold

fold   fold

APL\360 Primer   Printed in U.S.A.   GH20-0689-1

IBM ®

International Business Machines Corporation
Data Processing Division
112 East Post Road, White Plains, N.Y. 10601
[USA Only]

IBM World Trade Corporation
821 United Nations Plaza, New York, New York 10017
[International]

APL\360 Primer

GH20-0689-1

Program Numbers 5734-XM1 and 5736-XM1

Please comment on the usefulness and readability of this publication, suggest additions and deletions, and list specific errors and omissions (give page numbers). All comments and suggestions become the property of IBM. If you wish a reply, be sure to include your name and address.

COMMENTS

fold

fold

fold

fold

● Thank you for your cooperation. No postage necessary if mailed in the U.S.A.
FOLD ON TWO LINES, STAPLE AND MAIL.

YOUR COMMENTS PLEASE...

Your comments on the other side of this form will help us improve future editions of this publication. Each reply will be carefully reviewed by the persons responsible for writing and publishing this material.

Please note that requests for copies of publications and for assistance in utilizing your IBM system should be directed to your IBM representative or the IBM branch office serving your locality.

fold   fold

fold   fold

IBM
®

International Business Machines Corporation
Data Processing Division
112 East Post Road, White Plains, N.Y. 10601
[USA Only]

IBM World Trade Corporation
821 United Nations Plaza, New York, New York 10017
[International]