Contents

Fitting Curves and Surfaces with Monotonic and Non-Monotonic Four Parameter Equations

John Mandel*

National Bureau of Standards, Washington, DC 20234

This is a series of three papers in which methods are presented, with instructions on computational details, on the empirical fitting of tabulated data. Part I deals with fitting functions of a single argument; Part II with functions of two arguments; and Part III with functions of three or more arguments.

Key words: Curve fitting; empirical fitting; surface fitting.

Part I. Fitting Functions of a Single Argument

1. Introduction

The ready availability of calculators and computers has had a profound effect on the use of tables of complicated functions. For example, in statistical work one may be required in a specific computer program, to call on critical values of the F distribution for specified degrees of freedom in the numerator and the denominator, at specified levels of significance. It is totally impractical to store the entire F table in the memory of the computer, but it is entirely feasible to let the computer calculate the required value by a suitable approximation formula.

Similar situations occur for physical or chemical properties that are tabulated as functions of temperature, pressure, wave-length, etc.

The object of this paper is to present a widely applicable procedure for finding empirical representations of tabulated values. The tabulated values are of course assumed to be derived from reasonably smooth functions of the arguments. The approximation formulas are expected to generate values that are practically interchangeable with the corresponding tabulated values.

We will present the procedure in three parts. Part I is concerned with the empirical fitting of curves, i.e., functions of a single argument. Part II deals with functions of two arguments. The case of functions of more than two arguments is discussed in Part III.

Part I consists of two sections. In section 1, we deal with monotonic functions, and in section 2, with functions that have a single maximum or a single minimum.

2. Monotonic functions

2.1 The general formula

Polynomials, which are widely used for empirical fitting, have well-known shortcomings for the fitting of monotonic functions: they often have undesirable maxima, minima, and inflection points. The formula we propose in this section applies to monotonic functions, with or without a single inflection point in the range over which the curve is fitted. The formula is

$$y_i = y_o + A(x_i - x_o) |x_i - x_o|^{B-1}; \quad -\infty < x < +\infty \tag{1}$$

*National Measurement Laboratory.

where (x_i, y_i) are the coordinates of the points to be fitted by a monotone curve; and x_o, y_o, A, and B are four parameters the values of which have to be estimated. Note that B need not be an integer.

A simpler formula would be:

$$y = y_o + A(x - x_o)^B \tag{2}$$

but this formula presents difficulties for $x < x_o$, because of the ambiguity of defining $(x - x_o)^B$ for negative values of $x - x_o$. Equation (1) is totally free of this shortcoming. We will refer to eq (1) as the "four parameter equation" for monotonic functions and denote it by the symbol MFP.

2.2 Nature of the MFP function

Table 1 presents the properties of the *MFP* function in diagrammatic form. The function is defined for the entire range $x = -\infty$ to $x = +\infty$. Note that for $B < 0$, the curve is discontinuous at x_o and consists essentially of two branches, each of which is free of points of inflection. Note also that in this case ($B < 0$), the curve has finite asymptotes, equal to y_o, both at $+\infty$ and $-\infty$.

TABLE 1. *MFP: Nature of Function*

$$y = y_0 + A(x - x_0) |x - x_0|^{B-1}$$

		Value of y	
	x	$A < 0$	$A > 0$
	$-\infty$	y_0	y_0
$B < 0$	x_0	$\{{+\infty \atop -\infty}$	$\{{-\infty \atop +\infty}$
	$+\infty$	y_0	y_0

		Value of y	
	x	$A < 0$	$A > 0$
	$-\infty$	$+\infty$	$-\infty$
$B > 0$	x_0	y_0	y_0
	$+\infty$	$-\infty$	$+\infty$

$B = 0$	y is a constant
$B = 1$	y is a linear function of x.

For $B > 0$, y becomes infinite both at $x = -\infty$ and $x = +\infty$ and has a point of inflection at x_o.

For $B = 1$, the curve becomes a straight line and for $B = 0$, it becomes a constant. It is worth noting that the curve is *increasing* when $AB > 0$ and *decreasing* when $AB < 0$.

It is apparent that by choosing appropriate portions of the curve, with the proper parameter values, great flexibility is available, and it may therefore be expected that the curve will provide good fits for many sets of empirical data representing monotone functions with no more than one point of inflection. This does not mean, of course, that it will provide satisfactory fits for all sets of monotone data.

Figures (1a) through (1e) show some examples of curves that were generated by eq (1). The four parameters are given for each case. The figures demonstrate the flexibility that can be achieved through the use of this general formula.

2.3 Method of fitting

The procedure we use for finding the four parameters consists of two steps: (a) finding initial values for x_o and B; (b) iterating, using the Gauss-Newton procedure, to improve these estimates.

2

FIGURE 1-a.

$x_o = -.39$

$y_o = 230$

$A = -195$

$B = -.14$

FIGURE 1-b.

$x_o = 1$

$y_o = 25$

$A = -15$

$B = .2$

FIGURE 1-c.

$x_o = 1$

$y_o = 20$

$A = 15$

$B = .2$

FIGURE 1-d.

$x_o = 3.5$

$y_o = 15$

$A = 1.5$

$B = 1.8$

FIGURE 1-e.

$x_o = 5$

$y_o = 2600$

$A = -264$

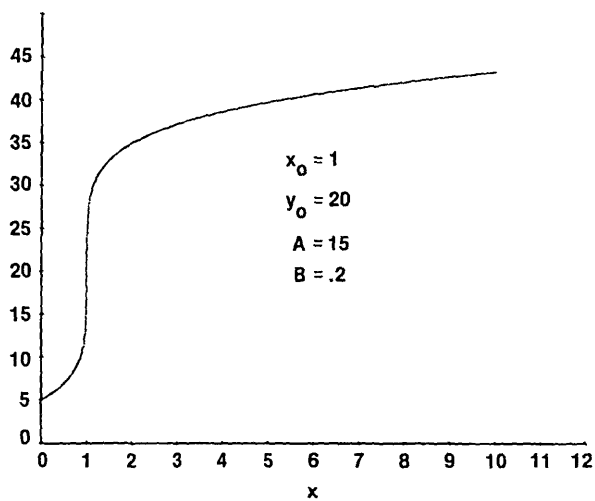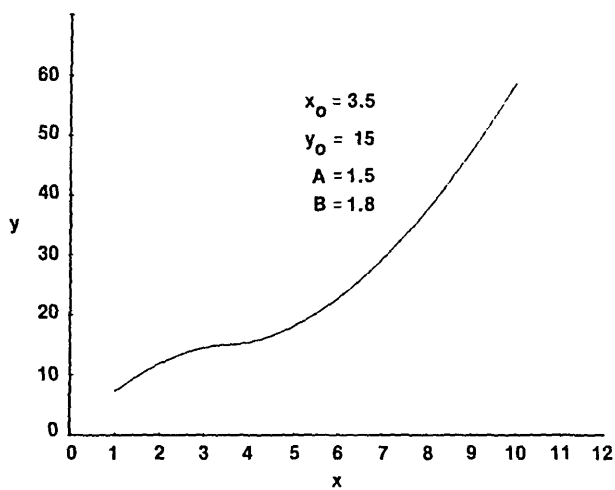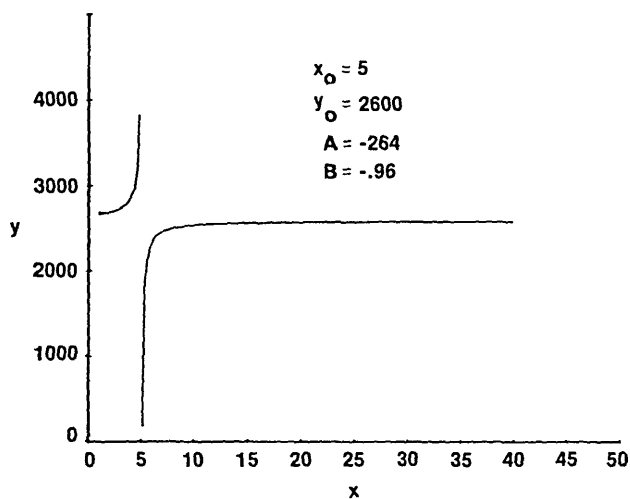$B = -.96$

3
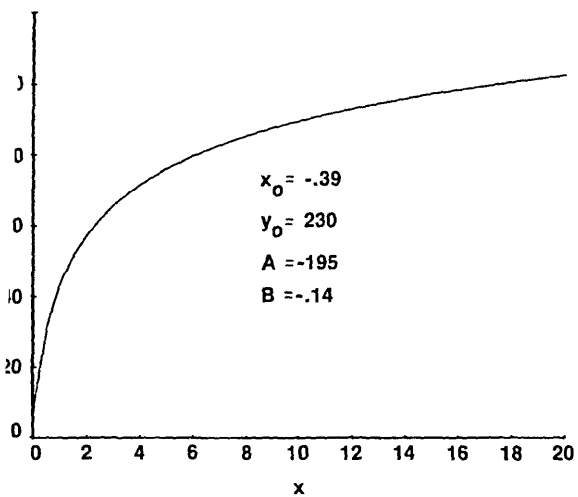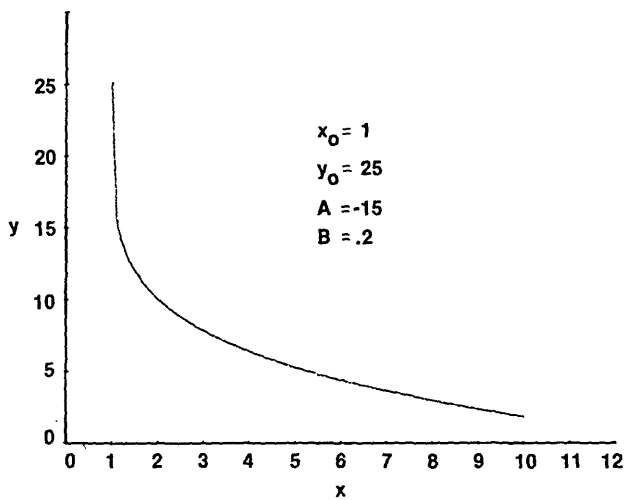
Both steps are presented with a great deal of detail, in spite of the fact that standard procedures can be found in the literature for non-linear fitting. The reason for this is the intent to make this paper essentially self-contained.

Once x_o and B are known, define

$$z_i \equiv (x_i - x_o) |x_i - x_o|^{B-1} \tag{3}$$

Then, eq (1) becomes:

$$y_i = y_o + A z_i \tag{4}$$

which is the equation of a straight line in y_i versus z_i, for which the intercept y_o and the slope A are readily estimated by linear regression.

The variable z_i can be regarded as a reexpression of x_i in a transformed scale. The new scale must be such that z_i is linearly related to y_i. For the purpose of comparing different pairs of (x_o, B) in achieving a good fit, a convenient measure is therefore the correlation coefficient between z_i and y_i. We will use this measure throughout the paper with the understanding that it is merely a comparative measure for the adequacy of a (x_o, B) pair of values, and that we are not concerned here with the statistical properties of this measure.

2.4 Initial values for x_o and B

Differentiating y with respect to x, in Eq (1), gives the relation:

$$\frac{dy}{dx} \equiv y'_x = AB |x - x_o|^{B-1} \tag{5}$$

Dividing eq (1) by eq (5) yields:

$$\frac{y - y_o}{y'_x} = \frac{x - x_o}{B}$$

or:

$$y = y_o + \frac{1}{B} (xy'_x) - \frac{x_o}{B} (y'_x) \tag{6}$$

Now x and y are given for N points of the curve, and y'_x can be approximately calculated for the midpoints of the intervals between successive x-values. The value of y can also be estimated approximately at these midpoints. This yields $N-1$ sets of values x, y, and y'_x from which xy'_x can be calculated. A multiple linear regression of y on xy'_x and on y'_x, allowing for the constant term y_o, then gives estimates of the coefficients y_o, $\frac{1}{B}$ and $\frac{x_o}{B}$. We ignore the first and use the two others to estimate x_o and B.

2.5 Illustrative example for finding initial values of x_o and B

An important statistical application of empirical curve and surface fitting is to represent standard statistical tables by formulas that can be used for ready interpolation or for incorporation into computer programs. Our first example is the two-tail 5 percent critical value of Student's t, for values of v, the degrees of freedom, ranging from 2 to ∞. The data used for the fit consist of 20 selected pairs of (v, t_c), where t_c is the critical value in question [2]. The data are given in table 2. We substituted 10,000 for $v = \infty$.

The calculations for the initial values estimation are shown in table 3. The midpoints of x and y are denoted x_m and y_m. The derivative y'_x is approximated by $\Delta y / \Delta x$ (column 5). For example:

4

TABLE 2. *Two-Tail 5 percent Critical Values of Student's t Statistic*

Degrees of Freedom (ν)	Student's t
2.0	4.3027
3	3.1825
4	2.7764
5	2.5706
6	2.4469
7	2.3646
8	2.3060
9	2.2622
10	2.2281
12	2.1788
14	2.1448
16	2.1199
18	2.1009
20	2.0860
25	2.0595
30	2.0423
40	2.0211
60	2.0003
120	1.9799
10000	1.9600

TABLE 3. *Calculations for Initial Values of x_o and B.*
(Data in Table 2)

(1) x_i	(2) y_i	(3) x_m	(4) y_m	(5) $\dfrac{\Delta y}{\Delta x}$	(6) $x_m \dfrac{\Delta y}{\Delta x}$
2.0	4.3027	2.5	3.74260	− 1.120200	− 2.800500
3.	3.1825	3.5	2.97945	− .406100	− 1.421350
4.	2.7764	4.5	2.67350	− .205800	− .926100
5.	2.5706	5.5	2.50875	− .123700	− .680350
6.	2.4469	6.5	2.40575	− .082300	− .534950
7.	2.3646	7.5	2.33530	− .058600	− .439500
8.	2.3060	8.5	2.28410	− .043800	− .372300
9.	2,2622	9.5	2.24515	− .034100	− .323950
10.	2.2281	11.0	2.20345	− .024650	− .271150
12.	2.1788	13.0	2.16180	− .017000	− .221000
14.	2.1448	15.0	2.13235	− .012450	− .186750
16.	2.1199	17.0	2.11040	− .009500	− .161500
18.	2.1009	19.0	2.09345	− .007450	− .141550
20.	2.0860	22.5	2.07275	− .005300	− .119250
25.	2.0595	27.5	2.05090	− .003440	− .094600
30.	2.0423	35.0	2.03170	− .002120	− .074200
40.	2.0211	50.0	2.01070	− .001040	− .052000
60.	2.0003	90.0	1.99010	− .000340	− .030600
120.	1.9799	5060.0	1.96995	− .000002	− .010191
10000.	1.9600				

$$\frac{3.1825 - 4.3027}{3 - 2} = -1.1202$$

Finally, the value xy'_x is approximated by $x_m \dfrac{\Delta y}{\Delta x}$ (column 6).

The multiple linear regression is carried out by regressing column (4) on a column consisting of unity for all rows (for the coefficient y_o) and on columns (6) and (5). Thus, the first observational equation is:

$$3.7426 = y_o (1) + \frac{1}{B} (-2.8005) - \frac{x_o}{B} (-1.1202)$$

5

The normal equations are (omitting the symmetrical elements below the diagonal in the $x'x$ matrix):

	y_o	$\frac{1}{B}$	$-\frac{x_o}{B}$
44.00215000	19.0	−8.86179170	−2.15789201
−25.67782645		12.14223318	4.10274947
−6.95277684			1.49195207

Solving these 3 equations in 3 unknowns, we obtain for the second and third coefficients:

$$\frac{1}{B} = -.93471932$$

$$-\frac{x_0}{B} = .75296356$$

which give the estimates:

$$\hat{x}_0 = .80555$$

$$\hat{B} = -1.069839$$

Calculating the corresponding z_i for all x_i by eq (3) and regressing y_i on z_i, we obtain the estimates

$$\hat{y}_o = 1.96424556$$

$$\hat{A} = 2.82544745$$

The correlation coefficient between z_i and y_i, for the pair ($x_o = .80555$, $B = 1.069839$), is 0.999993, indicating a quite successful fit. We now develop a Gauss-Newton iteration process to achieve, if possible, an even better fit.

2.6. Gauss-Newton iteration to improve the fit

The iteration is carried out on the standardized vectors of y and z, defined as follows:[1]

$$v_i \equiv \frac{y_i - \bar{y}}{S_y}, \text{ where } S_y = \sqrt{\sum_i (y_i - \bar{y})^2} \tag{7}$$

$$t_i \equiv \frac{z_i - \bar{z}}{S_z}, \text{ where } S_z = \sqrt{\sum_i (z_i - \bar{z})^2} \tag{8}$$

Define:

$$D_i \equiv v_i - t_i \text{ if } \sum_i v_i t_i > 0 \tag{9}$$

$$D_i \equiv -v_i - t_i \text{ if } \sum_i v_i t_i < 0 \tag{10}$$

Then the equations yielding the corrections for x_o and B, denoted by Δx_o and ΔB, are:

$$D_i = \frac{\partial t_i}{\partial x_o} \Delta x_o + \frac{\partial t_i}{\partial B} \Delta B \tag{11}$$

---

[1] t_i, as defined by eq (8) should not be confused with Students's t used for the illustrative example.

6

The partial derivatives $\frac{\partial t_i}{\partial x_o}$ and $\frac{\partial t_i}{\partial B}$ are readily evaluated using eqs (3), (7), and (8). The results are:

$$\frac{\partial t_i}{\partial x_o} = \frac{1}{S_z} (P_i - \bar{P} - E \cdot t_i) \tag{12}$$

$$\frac{\partial t_i}{\partial B} = \frac{1}{S_z} (Q_i - \bar{Q} - F \cdot t_i) \tag{13}$$

where

$$P_i \equiv -B|x_i - x_o|^{B-1} \tag{14}$$

$$Q_i \equiv z_i \ln|x_i - x_o| \tag{15}$$

$\bar{P}$ and $\bar{Q}$ are the averages of the P_i and the Q_i, and

$$E \equiv \sum_i P_i t_i \tag{16}$$

$$F \equiv \sum_i Q_i t_i \tag{17}$$

Using these equations, $\frac{\partial t_i}{\partial x_o}$ and $\frac{\partial t_i}{\partial B}$ are computed for all i and a multiple linear regression is carried out, in accordance with eq (11), of D_i on $\frac{\partial t_i}{\partial x_o}$ and $\frac{\partial t_i}{\partial B}$. The coefficients are Δx_o and ΔB.

To avoid "overshooting," it is advisable to correct x_o and B by only a fraction of Δx_o and ΔB, say $(\Delta x_o)/4$ and $(\Delta B)/4$. Thus, the new values for x_o and B are

$$\text{new } x_o = x_o + \frac{\Delta x_o}{4} \tag{18}$$

$$\text{new } B = B + \frac{\Delta B}{4} \tag{19}$$

Using these new values, z_i is recalculated for all i and the entire process is repeated. Iteration continues until practical convergence is reached.

Using the above equations, the entire procedure is readily programmed on the computer. The calculations are simple and rapid.

Referring to our illustrative example (table 2), and starting with the initial values $x_o = .80555$ and $B = -1.069839$, we obtain after 30 iterations: $x_o = .836464$ and $B = -1.055240$. The correlation between z and y is now 0.9999964. The fitted values are correct to within 2 or 3 units in the third place. Further iterations do not improve the fit and result in only minute changes in x_o and B. Table 4 lists the four parameters of the final fit and the fitted values. Figure 2 is a graph of the experimental points and of the fitted curve.

2.7. Additional remarks

Occasionally, the initial values for x_o and B, obtained by the method described in section 4 above, are unsatisfactory and the iteration process may fail to converge. One possible remedy is to interchange x and y in the formula. Indeed, the basic formula $y = y_o + A(x - x_o)^B$ can be written:

$$x = x_o + \frac{1}{A^{1/B}} (y - y_o)^{1/B} \tag{20}$$

indicating the same form for expressing x in terms of y as vice-versa.

If this advice leads to a satisfactory fit, one can do further iterations in the original form (y as a function of x), using the values of x_o and B obtained by use of the inverted formula after a few iterations.

7

TABLE 4. *Fit of Data of Table 2*

Equation: $y = 1.963000 + 2.745822 (v - .836464)^{-1.055240}$

Degrees of Freedom (v)	Tabular Value	Fitted Value
2	4.3027	4.3032
3	3,1825	3.1792
4	2.7764	2.7775
5	2.5706	2.5725
6	2.4469	2.4487
7	2.3646	2.3659
8	2.3060	2,3068
9	2.2622	2.2625
10	2.2281	2.2281
12	2.1788	2.1783
14	2.1448	2.1439
16	2.1199	2.1188
18	2.1009	2.0997
20	2.0860	2.0847
25	2.0595	2.0583
30	2.0423	2.0411
40	2.0211	2.0203
60	2.0003	2.0004
120	1.9799	1.9807
10,000 (∞)	1.9600	1.9632
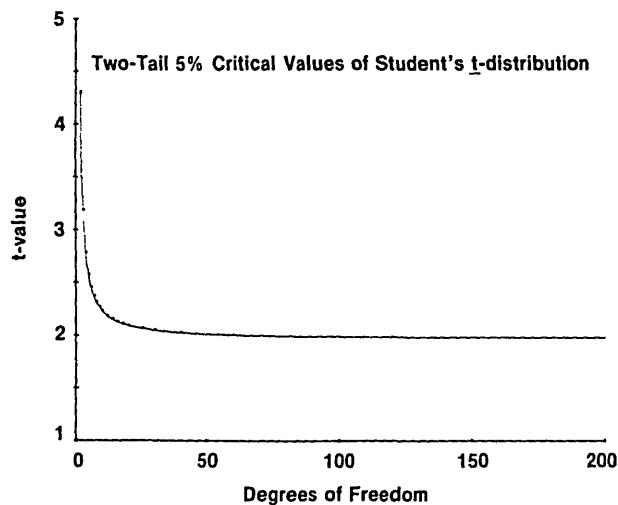
FIGURE 2. Two-tail 5 percent critical values of Student's *t*-distribution.

It is also possible to simply "guess" initial values. It happens occasionally that after an initial tendency to diverge (decreasing correlation coefficient), the process suddenly reverses itself and leads to a good fit.

As noted earlier, the general fitting equation given by (1), yields good results for many sets of monotonic data, but cannot be guaranteed to be satisfactory in all cases.

8

3. Non-Monotonic functions

3.1. The general formula

We will deal with functions that present a single maximum or a single minimum in the range in which they are to be fitted. Polynomials of order 2 (quadratics) appear to be the most plausible candidates for such functions. The general equation for the quadratic is:

$$y = a + bx + cx^2 \tag{21}$$

In practice, however, it is often found that even for functions presenting only a single maximum or a single minimum, eq (21) provides a very poor fit indeed. For example, the five points represented by the first two columns of table 5, when fitted by a quadratic, give the least squares fit of column 3 of that table. A fairly obvious device for improving the fit is to express the x variable in a transformed scale. To be effective, the scale transformation must be non-linear. A simple example is given by the relation:

$$y = a + bx^\alpha + c(x^\alpha)^2 \tag{22}$$

We will use a slight modification of eq (22), to allow for a frequently occurring case, namely the case in which a logarithmic transformation of x is indicated.

TABLE 5. *A Non-Monotonic Function*

x	y	Quadratic Fit of y[a]
25	.513089	.503630
10	.551180	.684080
5	.516202	.329230
1	.063119	−.104050
0.5	−.403100	−.167547

[a] Value $\hat{y} = a + bx + cx^2$, resulting from unweighted least squares fit.

It can readily be shown that

$$ln\ x = \lim_{\alpha \to 0} \frac{x^\alpha - 1}{\alpha} \tag{23}$$

Thus, by writing:

$$y = a + b\left(\frac{x^\alpha - 1}{\alpha}\right) + c\left(\frac{x^\alpha - 1}{\alpha}\right)^2;\ 0 < x < \infty \tag{24}$$

we obtain a general formula that includes the transformation of x to its logarithm (for very small α) as well as all cases covered by eq (22) (for any other α).

Equation (24) is the formula we propose for fitting non-monotonic functions of a positive valued argument, with a single maximum or minimum. We refer to this equation as the Quadratic Four Parameter (QFP) formula. The unsolved problem is to find the value of α that provides the best fit by eq (24). But first we study the properties of the QFP function.

3.2. Nature of the QFP function

A diagrammatic presentation of the properties of the QFP function is given in Table 6.

Note that the curve can be monotonic under certain conditions, namely for $b\alpha/2c \geq 1$. If not monotonic, the curve is finite at one end of the range of x and ∞ at the other end. Within this range, it passes through a local minimum or a maximum, which always occurs at $x = (1 - \frac{b\alpha}{2c})^{1/\alpha}$. Here again, as for the *MFP*, a great deal of flexibility is available.

9

TABLE 6. *QFP: Nature of Function.*

$$y = a + b\left(\frac{x^{\alpha}-1}{\alpha}\right) + c\left(\frac{x^{\alpha}-1}{\alpha}\right)^2$$

(If $\frac{b\alpha}{2c} \geq 1$, curve is monotonic)

Case of $\frac{b\alpha}{2c} < 1$

Value of x	Value of y		
	$\alpha > 0$	$\alpha < 0$	α very close to zero[b]
$x = 0$	$a - \dfrac{b}{\alpha} + \dfrac{c}{\alpha^2}$	$+\infty (c>0)$ $-\infty (c<0)$	$+\infty (c>0)$ $-\infty (c<0)$
$x = x_m$[a]	$a - \dfrac{b^2}{4c}$	$a - \dfrac{b^2}{4c}$	$a - \dfrac{b^2}{4c}$
$x = +\infty$	$+\infty (c>0)$ $-\infty (c<0)$	$a - \dfrac{b}{\alpha} + \dfrac{c}{\alpha^2}$	$+\infty (c>0)$ $-\infty (c<0)$

[a] $x_m = (1 - \frac{b\alpha}{2c})^{1/\alpha}$; for $\alpha \approx 0$, $x_m = e^{-b/2c}$

[b] In that case: $y = a + b \ln x + c(\ln x)^2$

3.3. The fitting procedure

Appropriate α-values are often found in the range -2 to $+2$. Since a computer program can readily be written that fits eq (24) for any given value of α, and since the calculations are quite rapid on any programmable desk calculator, or minicomputer, a trial and error search procedure is a reasonable way to obtain an initial value for α. It is then easy to apply a Gauss-Newton iteration process to zero in on the best value for α. We propose the following procedures:

1. Let (25)

$$u \equiv \frac{x^{\alpha}-1}{\alpha}$$

Then:

$$y = a + bu + cu^2 \qquad (26)$$

2. Choose a small set of values of α between -2 and $+2$. For each value of α, calculate u for each x, fit eq (26), and calculate the correlation coefficient ϱ between y_i and the corresponding fitted value $\hat{y}_i$. Also calculate:

$$k = \frac{c}{b} \qquad (27)$$

With a properly-written program, step 2 should take very little time on a programmable desk calculator and much less on a minicomputer.

3. Having found a value of α that gives a ϱ-value of 0.99 or better, in absolute value,[2] this α and the corresponding k may be taken as the starting values for the iteration process, which is presently described:

4. Equation (26) is written in the form:

$$y = a + b(u + ku^2) \qquad (28)$$

[2] For data of relatively low precision, it may not be possible to achieve $\varrho = .99$. We are particularly concerned, in this paper, with data of high precision, such as tabulated mathematical functions, or high precision physical or chemical data.

5. Let

$$z_i \equiv u_i + ku_i^2 \quad (i = 1 \text{ to N}) \tag{29}$$

Then

$$y_i = a + bz_i \tag{30}$$

6. Using the same approach as for the monotonic four-parameter (MFP) curve fit, and defining v_i, t_i, and D_i in the same way (eqs (7) to (10)), we obtain, as before:

$$\frac{\partial t_i}{\partial \alpha} = \frac{1}{S_z} [P_i - \bar{P} - E \cdot t_i] \tag{31}$$

$$\frac{\partial t_i}{\partial k} = \frac{1}{S_z} [Q_i - \bar{Q} - F \cdot t_i] \tag{32}$$

However, the quantities P and Q are now defined by the following relations:

$$P_i \equiv \frac{1 + 2ku_i}{\alpha} [(1 + \alpha u_i)\ln x_i - u_i] \tag{33}$$

$$Q_i \equiv u_i^2 \tag{34}$$

The quantities E and F are defined, as before, by eqs (16) and (17).
In the place of eq (11), we now have:

$$D_i = \frac{\partial t_i}{\partial \alpha} \Delta\alpha + \frac{\partial t_i}{\partial k} \Delta k \tag{35}$$

The regression of D_i on $\frac{\partial t_i}{\partial \alpha}$ and $\frac{\partial t_i}{\partial k}$ yields the "corrections" $\Delta\alpha$ and Δk. Again, it is advisable to use only a fraction, say one-fourth, of these quantities at each iteration.

The iteration process is continued until further impovement of the fit becomes negligible.

3.4. An illustration

For illustration of the QFP process, let us return to the example of table 5. Trying first a few values between -1 and 1, one finds readily that a high correlation ($\varrho = .999263$) is obtained for $\alpha = -0.1$, with a corresponding $k = .243550$. Using these approximations as initial estimates for the iteration process, the latter rapidly converges to a ϱ-value of 0.9999988, giving the final estimates (for eq (24)):

$$\alpha = -.198499$$
$$a = .062839$$
$$b = .522744$$
$$c = -.140142$$

The effectiveness of this fitting procedure can be judged by examining table 7, in which a number of simple transformations are compared with the one resulting from our fitting procedure. Note that the logarithmic transformation is represented by $\alpha = 0$; what is actually meant is that instead of the transformation to the logarithm, eq (23) could have been applied, using a very small value of α, such as 0.001 or 0.0001. It is apparent that a dramatic improvement in the fit results from using the value of α that is given by the iteration process.

Contrary to what may be believed, the (x,y) data in table 5 were not "made up." They represent one of the eigenvectors obtained in the process of fitting a table of a statistical function of three arguments by an empirical formula (see part III). The x-values are the values of one of the three arguments: level of significance.

TABLE 7. *Comparison of Quadratic Fits*

		y fitted, using quadratic function of			
x	y observed	x	$\ln x$[a]	x^{-1}	$x^{-0.198499}$
25	0.513089	0.503630	0.495354	0.560887	0.513351
10	.551180	.684080	.578775	.530048	.550271
5	.516202	.329230	.525325	.478823	.517023
1	.063119	−.104050	.014030	.076830	.062839
0.5	−.403100	−.167547	−.372926	−.406150	−.402995
	α	1	0	−1	−.198499
	ϱ	.89630	.99675	.99687	.999999

[a] $\ln x$ corresponds to $\alpha = o$. See text.

3.4. An illustration from the physical sciences and a simple stratagem

Table 8 is taken from a tabulation of the density d,[3] in grams per milliliter, of ordinary water, for values of temperature ranging from −5 to 30 °C [2].[4] For ease of calculations, we use the coded values $y = 10^{-6} (d - .998)$. It is well known that this property has a maximum in the vicinity of 4 °C. We consequently try to fit the data by the Quadratic Four Parameter (QFP) formula.

TABLE 8. *Density of Ordinary Water as a Function of Temperature*

Temperature (°C)	Density (in y-units[a])	Temperature (°C)	Density (in y-units[a])
−5	1283	5	1992
−4	1441	6	1968
−3	1578	7	1930
−2	1694	8	1877
−1	1790	9	1809
0	1868	10	1728
1	1927	15	1129
2	1968	20	234
3	1992	25	−925
4	2000	30	−2322

[a] $y = 10^6 (d - 0.998)$

We run into a minor difficulty in that our range of x-values includes negative values. This can be remedied in this case by taking as the independent variable $x = t + \beta$, where β is chosen so as to make all $x−$ values positive. We use $\beta = 10$ for our example. Thus:

$$x = t + 10$$
$$y = 10^6 (d - 0.998)$$

Trial of a few values for α shows that $\alpha = 0.8$ gives a fit with a correlation coefficient of 0.999967 (whereas for $\alpha = 1$, which corresponds to a "no transformation" quadratic fit, the correlation is 0.99894). The corresponding k value is $k = -.055752$. Application of the iteration process leads rapidly to the best parameter values:

$$\alpha = .824103$$
$$a = 242.267451$$
$$b = 373.749212$$
$$c = 19.881303$$

with a correlation $\varrho = .999990$.

---

[3] To avoid confusion with the correlation coefficient, we use the symbol d, rather than the conventional ϱ for density.

[4] Figures in brackets refer to literature references located at the end of each of the three parts of this paper.

Table 9 presents a comparison of the simple quadratic fit (no transformation) with the QFP. For ease of evaluation, the tabulated value is the residual $y - \hat{y}$ (fitted minus observed).

The QFP can reliably be used for interpolation. Use of an empirical formula for *extrapolation* is of course always risky and should be done if at all, with great caution. The usefulness of the formula for interpolation is demonstrated in table 10, which shows both fits for values of x *not* included in the fitting process (but *within* the range of the x values used for the fit).

TABLE 9. *Comparison of QFP and QLS Fits for Data of Table 6*[a]

Temperature (°C)	Residual		Temperature (°C)	Residual	
	QFP	QLS		QFP	QLS
−5	−10	93	5	−3	−35
−4	−1	53	6	−3	−26
−3	4	20	7	−4	−17
−2	6	−4	8	−5	−5
−1	6	−22	9	−4	7
0	5	−35	10	−4	20
1	3	−43	15	2	77
2	2	−45	20	8	97
3	0	−45	25	8	47
4	−1	−41	30	−7	−97

[a]QLS = quadratic least squares fit (unweighted)

TABLE 10. *Comparison of QFP and QLS Fits for Interpolation*

Temperature (°C)	Residual		Temperature (°C)	Residual	
	QFP	QLS		QFP	QLS
11	−3	32	21	9	94
12	−2	45	22	9	86
13	−1	57	23	9	77
14	1	68	24	8	64
16	4	86	26	5	26
17	5	92	27	3	2
18	6	96	28	1	−26
19	7	97	29	−4	−60

4. Conclusion for Part I

We have presented two formulas for the empirical fitting of functions of a single argument. The first applies to monotonic functions; the second can be used for monotonic functions under certain conditions, but its main use is for functions that have a single minimum or a single maximum. These formulas turn out to be useful also in the fitting process of functions of two or more arguments, as will be shown in Parts II and III of this paper. The fitting of the two functions is straightforward and can be readily programmed on computers and even on programmable calculators.

5. References

[1] *Symbols, Definitions and Tables for Industrial Statistics and Quality Control,* Rochester, Institute of Technology, Rochester, N.Y. (1958).
[2] *Handbook of Chemistry and Physics,* 56th Edition, CRC Press, Cleveland, Ohio (1975-1976).

Part II: Fitting Functions of Two Arguments

1. Introduction

Let z_{ij} be a function of two arguments x_i and y_j. The form of the function is unknown, but a set of data is available in the form of a rectangular array, in which the row "labels" are x_i and the column "labels" y_j. The tabulated value corresponding to x_i and y_j is z_{ij}. We assume that each cell of this two-way table is "filled"; i.e.: there are no cells for which the value of z_{ij} is missing.

An example will clarify matters. Table 1 is a portion of the table of 5 percent critical values of the "Studentized Range" [1]. This is a relatively short portion of the complete table but will suffice for illustration of the method, and will help to show its power.

The principle of the procedure is to first find the "Singular Value Decomposition" (SVD) [2] [1] of the matrix representing the two-way table, using only the z_{ij} values (but not the x_i and the y_j), and then to relate the parameters of the SVD to the x_i and the y_j, using the methods of part I of this paper.

The SVD [2] is a technique for developing the following relation:

$$z_{ij} = \theta_1 u_{1i} v_{1j} + \theta_2 u_{2i} v_{2j} + \ldots + \theta_p u_{pi} v_{pj} \tag{1}$$

or, more compactly:

$$z_{ij} = \sum_{k=1}^{p} \theta_k u_{ki} v_{kj} \tag{2}$$

where the θ_k are positive constants and the u_{ki} and v_{kj} are vectors such that

$$\sum_i u_{ki}^2 = \sum_j v_{kj}^2 = 1, \text{ for all } k \tag{3a}$$

and

$$\sum_i u_{ki} u_{k'i} = \sum_j v_{kj} v_{k'j} = 0, \text{ for } k \neq k' \tag{3b}$$

It can be shown that p is equal to the rank of the z-matrix, and this rank is, in turn, equal at most to the number of rows or the number of columns of the table, whichever is smaller. The θ_k are the square roots of the "eigenvalues" of the $z^T z$ matrix, where z^T denotes the transpose of the matrix z, and the v_{kj} are the corresponding "eigenvectors." The u_{ki} are the eigenvectors of the zz^T matrix. The θ's are called the *singular values* of the matrix z.

2. The SVD technique

Algorithms for finding the SVD are readily available and a number of computer programs have been written for this purpose [14,15,16,17].

In most cases, it is not necessary to consider all p terms on the right hand side of eq (1). In fact, the first 2 or 3 terms are often sufficient to give an excellent approximation for the z_{ij}. From a practical viewpoint, it is easy to judge at what point the SVD can be terminated.

We illustrate the procedure with the data of table 1. Denote the residuals, after fitting the first q terms on the right-hand side of eq (2), by $(d_{ij})_q$. Thus:

$$(d_{ij})_q = z_{ij} - \sum_{k=1}^{q} \theta_k u_{ki} v_{kj} \tag{4}$$

It can be shown that:

---

[1] The SVD procedure is intimately related to the Method of Principal Components [3]; its use, in either of these two forms for the analysis of two-way tables of data has been discussed in a number of places [4,5,6,7,8,9,10,11,12,13]. Figures in brackets refer to literature references, listed on page 19.

14

$$\sum_i \sum_j (d_{ij})^2_q = \sum_i \sum_j z^2_{ij} - \sum_{k=1}^{q} \theta^2_k \qquad (5)$$

For the data of table 1, we find

$$\theta_1 = 38.034666; \theta_2 = 1.392726; \theta_3 = .064557$$

and

$$\sum_i \sum_j z^2_{ij} = 1448.579715.$$

Applying eq (5), for $q = 3$, we find:

$$\sum_i \sum_j (d_{ij})^2_{q=3} = 0.000044$$

This sum of squares of residuals applies to 42 observations; the residuals are not independent. Nevertheless, we can calculate an "average square residual per observation,"[2] which will be equal to

$$\frac{0.000044}{42} = 1.05 \times 10^{-6} = (0.0010)^2$$

Thus, after fitting the first three terms of the SVD, the residuals will be of the order of 0.001. Since the data are given with 3 decimals, such a fit is, for most practical purposes, quite satisfactory.

Table 2 lists the θ values and the eigenvectors u_i and v_j for the first three terms.

TABLE 1. *Five Percent Critical Values of the Studentized Range*

ν \\ n	2	3	5	10	20	60	100
4	3.927	5.040	6.287	7.826	9.233	11.240	12.090
8	3.261	4.041	4.886	5.918	6.870	8.248	8.843
20	2.950	3.578	4.232	5.008	5.714	6.740	7.187
40	2.858	3.442	4.039	4.735	5.358	6.255	6.645
120	2.800	3.356	3.917	4.560	5.126	5.929	6.275
∞	2.772	3.314	3.858	4.474	5.012	5.764	6.085

TABLE 2. *θ-Values and Eigenvectors for Studentized Range (table 1)*

ν	u_1	u_2	u_3	n	v_1	v_2	v_3
4	.586984	−.684908	−.393566	2	.199994	.550013	.514706
8	.439355	−.132755	.588796	3	.246064	.489632	.178017
20	.366344	.194757	.446619	5	.296980	.365874	−.212600
40	.343350	.317316	.117282	10	.357146	.167917	−.490644
120	.328090	.407791	−.239725	20	.411696	−.032030	−.447995
∞	.320429	.456467	−.477211	60	.489830	−.320941	.071843
				100	.523336	−.437783	.459993
$\theta_1 = 38.034666,$		$\theta_2 = 1.392726,$		$\theta_3 = .064557$			

3. Fitting the structural parameters

Our next task is to express the u_i and v_j (which we can call the *structural parameters*) as functions of ν and n respectively. This is a curve fitting problem, and can be attacked by the methods developed in part I of this paper.

---

[2] No attempt is made to compute a "standard deviation of fit" with an appropriate number of degrees of freedom. The fit is purely empirical and is not based on a mathematical model; the proposed "average square residual per observation" is to be understood in that spirit.

The method of section 2 of part I is used for eigenvectors that are monotonic functions of their arguments, while section 3 is applied to eigenvectors that present a single maximum or minimum. In general, the fit for the u_{3i} and v_{3j} vectors need not be as good as those for the first two sets of vectors, since the third term, having a small multiplier θ_3, contributes only a small part to the SVD.

The entire computational procedure is summarized in Table 3 which lists the fitted vectors u_{ki} and v_{kj} obtained, for $k = 1$ and 2, by the MFP procedure and for $k = 3$, by the QFP fitting procedure. Table 4 lists the fitted values, which may be compared to the values in table 1.

TABLE 3. *Parameters of Fits*

	MFP Fits			
Vector	x_0	y_0	A	B
u_1	.981602	.320232	.766066	−.954958
v_1	.569248	1.335141.	−1.167675	−.079057
u_2	.523691	.459599	−3.326912	−.856578
v_2	−5.477500	−.964295	3.392462	−.398982

	QFP Fits			
Vector	α	a	b	c
u_3	−.776878	−24.645495	47.554916	−22.368740
v_3	−.032413	1.449792	−1.543742	.307466

TABLE 4. *Fitted Values for Data of Table 1*

v \ n	2	3	5	10	20	60	100
4	3.923	5.044	6.291	7.823	9.231	11.242	12.089
8	3.261	4.038	4.889	5.921	6.867	8.247	8.848
20	2.954	3.572	4.233	5.012	5.712	6.733	7.183
40	2.862	3.435	4.040	4.739	5.358	6.253	6.648
120	2.804	3.348	3.917	4.564	5.126	5.930	6.283
∞	2.776	3.308	3.859	4.477	5.007	5.755	6.081

4. Interpolation

The total number of parameters used for the fit is 27: four for each of the two vectors in each of the three terms of the SVD, in addition to the three θ-values. The actual number is less, since each θ can be incorporated into the coefficients of one of the two corresponding vectors, and further simplification is possible when the entire equation is algebraically reduced. However, this is unimportant for two reasons.

In the first place, for purposes of programming the calculations, any additional manipulation to reduce the number of parameters is unnecessary.

Secondly, and this is an important point—the often-made assertion that "it is absurd to fit a set of data with as may parameters, or almost as many parameters as there are data" can not be justified. The fact is that by fitting the 42 values of table 1, we have obtained a formula that fits the 5 percent critical value table of the Studentized Range for *all* values of v *from 4 to infinity* and *all* values of n *from 2 to 100*, using only 27 parameters. That this indeed so, can be verified by applying our fitting algorithm to *any* pair of v and n values (v = 4 to ∞, n = 2 to 100).

The values of Table 5 illustrate this point for some selected pairs of values.

The procedure we propose is in many cases, a powerful and reliable interpolation algorithm.

16

TABLE 5. *Examples of Fitted Values for (ν,n) Not Included in Fit.*[a]

		Studentized Range	
ν	n	Tabular	Fitted
3	3	5.910	5.945
3	8	8.853	8.863
3	50	13.360	13.352
6	4	4.896	4.897
6	12	6.789	6.789
6	70	9.370	9.375
13	2	3.055	3.059
13	17	5.931	5.929
13	90	7.667	7.663
60	4	3.737	3.734
60	50	5.958	5.956
60	80	6.303	6.305
500	3	(3.314)	3.317
700	40	(5.498)	5.517
1000	90	(6.020)	6.042

[a] Values in parentheses correspond to ν values not found in tables. The values in parentheses are for $ν = \infty$.

AN EXAMPLE FROM THE PHYSICAL SCIENCES

Table 6 lists values of the quantity $(n^2-1)/(n^2+2)$ for benzene at various values of pressure and wavelength. The values were derived from table 1 of ref. [18]. The symbol n represents refractive index. The function $(n^2-1)/(n^2+2)$ is chosen, in prefernce to n, because of the Lorentz-Lorenz equation:

$$\frac{n^2-1}{n^2+2} \frac{1}{D} = f(\lambda) \qquad (6)$$

where D represents density and λ wavelength. Since for a fixed mass, the density is a function of pressure, we can replace D by $\phi(P)$, where P denotes pressure. Equation (6) can then be written as:

$$\frac{n^2-1}{n^2+2} = \phi(P) \cdot f(\lambda) \qquad (7)$$

According to this equation, the quantity $(n^2-1)/(n^2+2)$ is a multiplicative function of two factors, one depending on pressure only, and the other on wavelength only. It can be shown that if this is true, all but the first term of SVD of table 6 represent merely experimental error. The first term of the SVD, on the other hand represents the quantity $\phi(P) \cdot f(\lambda)$.

TABLE 6. *Refractive Index of Benzene at 34.5 °C as a Function of Pressure and Wavelength*

	Tabulated Value $= \dfrac{n^2-1}{n^2+2}$							
	Wavelength							
Pressure	*6678*	*6438*	*5876*	*5086*	*5016*	*4922*	*4800*	*4678*
1.	0.287528	0.288222	0.290224	0.294366	0.294869	0.295531	0.296490	0.297546
246.2	.293514	.294242	.296296	.300494	.301012	.301707	.302652	.303743
484.8	.298497	.299225	.301278	.305558	.306072	.306737	.307738	.308830
757.2	.303242	.303964	.306057	.310383	.310917	.311597	.312595	.313713
1107.7	.308426	.309152	.311276	.315688	.316232	.316921	.317930	.319058

17

Eventhough our procedure makes no pretenses to anything but empirical fitting, this set of data provides us with an opportunity to examine the agreement between a physical theory (the Lorentz-Lorenz relation) and a set of experimental data.

The first three θ values of the SVD of table 6 are:

$$\theta_1 = 1.92330513; \theta_2 = 0.00009801; \theta_3 = 0.00002689$$

Note the very large drop from θ_1 to θ_2, indicating that one multiplative term in the SVD should represent the data quite well. More exactly, we find:

$$\sum_i \sum_j (d_{ij})_1^2 = \sum_i \sum_j z_{ij}^2 - \theta_1^2$$

$$= 3.699102657 - (1.92330513)^2 = 3.4 \times 10^{-8}$$

The average square residual per observation is:

$$\frac{3.4 \times 10^{-8}}{40} = 8.5 \times 10^{-10} = (2.9 \times 10^{-5})^2$$

Thus, one single multiplicative term reproduces the data of table 6 to about 3 units in the 5th place. It is easily verified that addition of a second multiplicative term fails to significantly improve this fit. The precision of a measurement of n in this study is no better than 1 to 3 units in the fifth place [18]. Applying the law of propagation of errors, it is easily seen that the same statement holds for the quantity $(n^2-1)/(n^2+2)$.

We now have the model:

$$\frac{n^2-1}{n^2+2} = \theta u_i v_j + \text{error} \tag{8}$$

where u_i is a function of pressure only, and v_j a function of wavelength only. Thus, eq (8) is equivalent to eq (7), as required by the Lorentz-Lorenz theory.

The fit of u_i as a function of pressure and v_j as a function of wavelength can be accomplished by the four-parameter curve. Table 7 lists the parameters of the two curves as well as the fitted values, using eq (8). A comparison of tables 6 and 7 confirms the satisfactory quality of the fit.

TABLE 7. *Fitted Values for Data of Table*

Equation: $\dfrac{n^2-1}{n^2+2} = 1.923305 \, [.671861 - 1.483692(P+1480.809979)^{-.249049}] \cdot [.333889 + .091736 \, (\dfrac{\lambda}{1000} - 1.817095)^{-1.237982}]$

				Wavelength				
Pressure	6678	6438	5876	5086	5016	4922	4800	4678
1	0.287552	0.288247	0.290238	0.294364	0.294840	0.295519	0.296470	0.297513
246.2	.293563	.294272	.296305	.300517	.301003	.301696	.302667	.303732
484.8	.298467	.299188	.301255	.305537	.306032	.306736	.307723	.308806
757.2	.303229	.303961	.306061	.310412	.310915	.311630	.312633	.313733
1107.7	.308387	.309133	.311268	.315693	.316204	.316931	.317952	.319070

5. Conclusion for Part II

By combining the Singular Value Decomposition technique with the curve fitting procedures developed in part I, it is possible to obtain excellent empirical fits for many sets of data in which the dependent

18

(response) variable is displayed as a two-way table and the rows and columns represent levels of the two independent(regressor) variables, respectively.

The procedure consists in performing an SVD on the matrix of values of the response variable and then fitting the vectors of parameters, which are functions of the rows or of the columns, but not of both, to the corresponding regressor variables.

6. References

[1] Harter, H. L., Tables of Range and Studentized Range, Annals of Mathematical Statistics, **31**, No. 4, 1122-1147, 1960.

[2] Rao, C. R., *Linear Statistical Inference and Its Applications,* (John Wiley & Sons, N.Y., 1973).

[3] Hotelling, H., Analysis of a Complex of Statistical Variables into Principal Components, J. of Educational Psychology **24**, 417-441, 498-520 (1933).

[4] Church, A., Analysis of Data when the Response is a Curve, Technometrics **8**, 229-246 (1966).

[5] Gollob, H. F., A Statistical Model which combines Features of Factor Analytic and Analysis of Variance Techniques, Psychometrika **33**, 73-116 (1968).

[6] Jolicoeur, Pierre and J. E. Mosimann, Size and Shape Variation in the Painted Turtle. A Principal Component Analysis, Growth **24**, 339-354 (1960).

[7] Mandel, John, The Partitioning of Interaction in Analysis of Variance, J. of Research of the National Bureau of Standards—B. Mathematical Sciences 73B, 309-328 (1969).

[8] Mandel, John, Distribution of Eigenvalues of Covariance Matrices of Residuals in Analysis of Variance, J. of Research of the National Bureau of Standards—B. Mathematical Sciences 73B, 149-154 (1970).

[9] Mandel, John, A New Analysis of Variance Model for Non-Additive Data, Technometrics **13**, 1-18 (1971).

[10] Mandel, John, Principal Components, Analysis of Variance and Data Structure, Statistica Neerlandica **26**, 119-129 (1972).

[11] Simonds, J. L., Application of Characteristic Vector Analysis to Photographic and Optical Response Data, J. Opt. Soc. Am. **53**, 968-974 (1963).

[12] Snee, R. D., On the Analysis of Response Curve Data, Technometrics **14**, 47-62 (1972).

[13] Wernimont, Grant, Evaluating Laboratory Performance of Spectrophotometers, Anal. Chem. **39**, 554-562 (1967).

[14] Garbow, B. J., et al., Matrix Eigensystem Routines—EISPACK Guide Extension. Lecture Notes in Computer Science, Vol. **51**. Springer—Verlag, New York/Heidelberg/Berlin, 1977.

[15] Smith, B. T., et al., Matrix Eigensystem Routines—EISPACK Guide, Second Edition. Lecture Notes in Computer Science, Vol. **6**. Springer—Verlag, New York/Heidelberg/Berlin, 1976.

[16] Sparks, D. N. and Todd, A. D., Algorithm AS 60: Latent Roots and Vectors of a Symmetric Matrix. Applied Statistics, Vol. **22** (1973), pp. 260-265. Corrigendum: Applied Statistics, Vol. **23** (1974), pp. 101-102.

[17] Wilkinson, J. H. and Reinsch, C., editors, *Handbook for Automatic Computation, Volume II, Linear Algebra.* (Springer-Verlag, New York/Heidelberg/Berlin, 1971).

[18] Waxler, R. M., Weir, C. E., and H. W. Schamp, Jr., Effect of Pressure and Temperature Upon the Optical Dispersion of Benzene, Carbon Tetrachloride and Water, J. Res. Nat'l. Bureau of Standards—A. Physics and Chemistry, *68A*, No. 5, 489-498, 1964.

Part III: Fitting Functions of Three or More Arguments

1. Introduction

The first two papers in this series (Parts I and II) dealt with ordinary curve and surface fitting, i.e., with the fitting of functions of one or two arguments. In the latter case, it was assumed that the data were in the form of a two-way table with no cells missing. Similarly, we will assume in this paper, that each value of the function to be fitted is associated with a combination of the levels of three or more arguments, all combinations being present, and each one being associated with a single value of the function. In other words, we assume a "complete factorial" with no replications per cell. Of course, if one or more cells contain more than a single observation, one can substitute the average for these replicates. For purposes of empirical fitting, this should be quite acceptable, provided the precision of the single observations is satisfactory.

We present the method in terms of a single example, a function of three arguments. Generalization to functions of more than three arguments should be self-evident. However, the method may become cumbersome, and is not recommended as a first choice in these cases.

19

2. Illustration: Fitting the F table

Table 1 is a portion of the table of critical values of the F distribution for the levels of significance P, of 25, 10, 5, and 1 percent, and for degrees of freedom, both in the numerator and in the denominator, of 4, 6, 60, 120, and ∞. The table, taken from ref. [1], has 100 "observations," but covers an infinite range of both sets of degrees of freedom, ν_1 and ν_2. We fully intend the empirical fit to be acceptable over this doubly-infinite range, and for all values of P between 1 and 25 percent.

TABLE 1. *F-Table. Data for Fit*

ν_1	ν_2	P—Level of Significance, in percent P			
		25	10	5	1
4	4	2.06	4.11	6.39	15.98
	6	1.79	3.18	4.53	9.15
	60	1.38	2.04	2.53	3.65
	120	1.37	1.99	2.45	3.48
	∞	1.35	1.94	2.37	3.32
6	4	2.08	4.01	6.16	15.21
	6	1.78	3.05	4.28	8.47
	60	1.35	1.87	2.25	3.12
	120	1.33	1.82	2.17	2.96
	∞	1.31	1.77	2.10	2.80
60	4	2.08	3.79	5.69	13.65
	6	1.74	2.76	3.74	7.06
	60	1.19	1.40	1.53	1.84
	120	1.16	1.32	1.43	1.66
	∞	1.12	1.24	1.32	1.47
120	4	2.08	3.78	5.66	13.56
	6	1.74	2.74	3.70	6.97
	60	1.17	1.35	1.47	1.73
	120	1.13	1.26	1.35	1.53
	∞	1.08	1.17	1.22	1.32
∞	4	2.08	3.76	5.63	13.46
	6	1.74	2.72	3.67	6.88
	60	1.15	1.29	1.39	1.60
	120	1.10	1.19	1.25	1.38
	∞	1.00	1.00	1.00	1.00

3. Approach

The method is simply stated. First, we combine two of the three factors, in this case ν_1 and ν_2, pretending it to be, for the time being, a single factor. We then apply the SVD (singular value decomposition) to the two-way table thus obtained. Each eigenvector will be a function of either P or of the combination of a particular ν_1 and a particular ν_2. This latter type of eigenvector is then entered into a two-way table, as a function of ν_1 and ν_2. This two-way table is, itself, subjected to a SVD, with resulting eigenvectors that are functions of ν_1 or ν_2, taken singly. The problem is thus reduced to the fitting of a number of curves (functions of a single argument).

4. Details of the fitting process

1. First step: SVD of 25 × 4 table.

The 25 rows are the combinations of the five levels of ν_1 and the five levels of ν_2; the four columns represent the four levels of the factor P (see table 1).

The SVD of this table, carried out to three terms, is represented by[1]

$$y_{it} = \theta_1 u_{1i} v_{1t} + \theta_2 u_{2i} v_{2t} + \theta_3 u_{3i} v_{3t} \qquad (1)$$

with the values (see Parts I and II for notation and terminology):

$$\sum_i \sum_t y_{it}^2 = 1939.001 \qquad (2)$$

$$\theta_1 = 43.70505824 \qquad (3)$$

$$\theta_2 = 5.34641595 \qquad (4)$$

$$\theta_3 = .53288922 \qquad (5)$$

From these values we derive:

$$\sum_i \sum_t (d_{it})_3^2 = 7.495692 \times 10^{-4}$$

which gives an average square residual per observation of:

$$\frac{7.495692 \times 10^{-4}}{5 \times 5 \times 4} = 7.495692 \times 10^{-6} = (.00274)^2$$

Thus, the fit will be good to approximately 3 units in the third place, provided that all the eigenvectors are fitted to an equivalent degree of approximation.

2. *Second step: Fitting the v vectors*

All v vectors are functions of a single variable, P, as shown in table 2. They are readily fitted by the methods of Part 1, with the results shown in Table 3.

TABLE 2. *v-Vectors as Functions of P*

P	v_1	v_2	v_3
25	0.161663	0.567164	0.750537
10	.275861	.537558	−.155183
5	.395108	.412703	−.613514
1	.861193	−.468007	.190293

TABLE 3. *Fit of v-Vectors*

a) MFP Fits
$$y = y_0 + A(x - x_0) \,|x - x_0|^{B-1}$$
where $x = P; y = v$

Vector	x_0	y_0	A	B
v_1	0.104080	−0.102867	0.923714	−0.388978
v_2	−6.628102	.568912	−9496.645	−4.489745

b) QFP Fit
$$y = a + b\left(\frac{x^\alpha - 1}{\alpha}\right) + c\left(\frac{x^\alpha - 1}{\alpha}\right)^2$$
where $x = P; y = v$

Vector	α	a	b	c
v_3	−0.237709	0.190293	−1.847334	0.931911

---

[1] The critical F values are temporarily represented by the symbol y_{it}.

3. Third Step: SVD of the u vectors

Each u vector is a function of v_1 and v_2 as shown by tables 4, 5, and 6.

To avoid confusion, we will denote the eigenvectors resulting from the SVD of each u vector by the symbols A_j and B_k for u_1, C_j and D_k for u_2, and E_j and G_k for u_3. We find that, to obtain sufficient precision, the SVD for u_1, requires three terms, while for u_2 and u_3 two terms suffice; thus:[2]

$$u_1 = T_{11}A_{1j}B_{1k} + T_{12}A_{2j}B_{2k} + T_{13}A_{3j}B_{3k} \tag{6}$$

$$u_2 = T_{21}C_{1j}D_{1k} + T_{22}C_{2j}D_{2k} \tag{7}$$

$$u_3 = T_{31}E_{1j}G_{1k} + T_{32}E_{2j}G_{2k} \tag{8}$$

TABLE 4. u_1-Vector as a Function of v_1 and v_2

$v_2 \backslash v_1$	4	6	60	120	∞
4	0.406210	0.388401	0.352024	0.349916	0.347548
6	.247943	.231426	.196782	.194521	.192350
60	.112774	.098616	.063326	.060227	.056489
120	.108349	.094350	.058259	.054485	.050072
∞	.104083	.090175	.052868	.048419	.038755

TABLE 5. u_2-Vector as a Function of v_1 and v_2

$v_2 \backslash v_1$	4	6	60	120	∞
4	−.273800	−.232084	−.153928	−.149371	−.144944
6	.058346	.084441	.132782	.135561	.139113
60	.227296	.231801	.224040	.221888	.218938
120	.229914	.232483	.220851	.216840	.212030
∞	.230595	.233936	.216705	.210834	.196284

TABLE 6. u_3-Vector as a Function of v_1 and v_2

$v_2 \backslash v_1$	4	6	60	120	∞
4	0.054105	0.101229	0.149334	0.154646	0.159300
6	−.352902	−.284130	−.137823	−.118086	−.109860
60	−.259816	−.119461	.163908	.180098	.215083
120	−.227942	−.098102	.195804	.216706	256403
∞	−.206583	−.088254	.221559	.247171	.323018

4. Fourth Step: Fitting the vectors A, B, C, D, E, and G.

The vectors A, C, and E are functions of v_1 only, while B, D, and G are functions of v_2 only, as shown in table 7. Again, we use the methods of Part I to fit these vectors to their corresponding arguments, with the results shown in table 8.

5. Fifth step: Fit of F as a function of P, v_1, and v_2.

By substituting for u_1, u_2, and u_3 in eq. (1), their expressions as given by eqs (6), (7), and (8), one readily obtains an expression for y_{it} as a function of quantities that are either constants (the θ and the T), or functions of a single argument (P, v_1, and v_2). Since the latter have all already been fitted in terms of their respective arguments, the problem is solved, except for the routine multiplications and additions involved in eqs (1), (6), (7), and (8). A program can readily be written to obtain the value of y_{it}, that is, of F, for any v_1, v_2, and P, using eqs (1), (6), (7), (8) and the MFP or QFP fits shown in Tables 3 and 8.

---

[2] The square roots of the eigenvalues are represented by the letter T, to avoid confusion with the θ of eq. (1).

TABLE 7. *SVD Vectors as Functions of ν_1 or ν_2*
(Eqs. (6), (7), (8))

ν_1	A_1	A_2	A_3	C_1	C_2	E_1	E_2
4	.828136	−.362993	−.216246	−.436484	−.712491	.220226	.407258
6	.479297	.070208	.361271	.243432	−.673184	.053854	−.898992
60	.179612	.482627	.431289	.506592	−.166174	.511488	−.141436
120	.168194	.520863	.266325	.501735	−.105627	.554120	−.025071
∞	.154640	.599231	−.752183	.491779	−.020205	.616385	.072944

ν_2	B_1	B_2	B_3	D_1	D_2	G_1	G_2
4	.511101	.609203	.238929	.481716	.687371	−.473303	.661606
6	.481313	.367238	−.143676	.475017	.314111	−.195838	.558564
60	.416240	−.301559	−.508432	.434081	−.344513	.426098	.311762
120	.411520	−.379435	−.342844	.426716	−.377055	.471649	.281945
∞	.405595	−.509016	.739069	.414481	−.409871	.577595	.271280

TABLE 8. *Fit of Vectors of Table 7*

a) *MFP Fits*

$$y = y_0 + A(x-x_0)\,|x-x_0|^{B-1}$$
where $x = \nu_1$ or ν_2; y = vector fitted

Vector	x_0	y_0	A	B
A_1	2.333242	.154595	1.080698	−0.925484
A_2	3.160494	.609721	−.893793	−.483655
B_1	−.251639	.405481	.366518	−.859647
B_2	−.228726	−.518182	2.767959	−.622996
D_1	−9.389754	.414477	.470463	−.749928
D_2	−.114913	−.409594	4.845992	−1.050171
G_1	.678625	.581829	−2.303516	−.650175
G_2	−1.463892	.269022	2.026138	−.966873

b) *QFP Fits*

$$y = a + b\left(\frac{x^\alpha-1}{\alpha}\right) + c\left(\frac{x^\alpha-1}{\alpha}\right)^2$$
where $x = \nu_1$; y = vector fitted

Vector	α	a	b	c
A_3	−.428299	−6.868927	9.408257	−2.907025
B_3	−.325922	3.754828	−4.383540	1.108433
C_1	−1.196614	−34.856565	87.329516	−53.883595
C_2	−.830262	3.094441	−9.025929	5.346705
E_1	−1.010583	15.626475	−37.446675	22.511319
E_2	−1.126615	93.240278	−235.519021	147.086636

5. Results

Table 9 shows the fitted value for each entry of table 1. As expected, the fit is very good. The sum of squares of the residuals for 100 values is 0.0109. This gives a root mean square deviation per value of 0.0104.

6. Interpolation

We mentioned earlier that the final fit should be adequate not only in reproducing the values of the original table, but also as an interpolation formula. In table 10, a comparison is made between values of F as given by the Biometrika Tables, and those given by our empirical fit, *for combination of P, ν_1, and ν_2 not*

included in table 1 (the basis for our formula). Note, in particular, the values for $P = 2.5$ percent, a level that was totally absent from table 1.

The sum of squares of residuals for these 80 values is 0.01812. Thus, the root mean square deviation per value is 0.048. Interpolation would of course be better if a larger table of F values had been used for the fitting process.

TABLE 9. *F Table—Fitted Values*

ν_1	ν_2	P			
		25	10	5	1
4	4	2.07	4.11	6.40	15.98
	6	1.78	3.18	4.52	9.15
	60	1.39	2.04	2.54	3.66
	120	1.37	1.98	2.45	3.47
	∞	1.36	1.95	2.39	3.32
6	4	2.07	4.00	6.16	15.21
	6	1.79	3.07	4.29	8.47
	60	1.35	1.87	2.26	3.12
	120	1.32	1.80	2.16	2.95
	∞	1.31	1.76	2.10	2.82
60	4	2.07	3.78	5.69	13.66
	6	1.75	2.77	3.74	7.06
	60	1.19	1.40	1.54	1.84
	120	1.15	1.32	1.42	1.65
	∞	1.11	1.23	1.30	1.47
120	4	2.08	3.78	5.66	13.56
	6	1.74	2.74	3.70	6.96
	60	1.17	1.35	1.47	1.73
	120	1.13	1.26	1.35	1.53
	∞	1.08	1.16	1.21	1.31
∞	4	2.09	3.77	5.64	13.46
	6	1.73	2.72	3.66	6.89
	60	1.15	1.29	1.38	1.60
	120	1.10	1.18	1.23	1.36
	∞	1.01	1.02	1.00	.95

TABLE 10. *F Table—Interpolated Values*[a]

P(%)		P									
		25		10		5		2.5		1	
ν_1	ν_2	Tab.	Fit	Tab.	Fit	Tab.	Fit	Tab.	Fit	Tab.	Fit
5	5	1.89	1.87	3.45	3.41	5.05	4.99	7.15	7.06	10.97	10.93
	15	1.49	1.53	2.27	2.33	2.90	2.98	3.58	3.59	4.56	4.61
	30	1.41	1.42	2.05	2.07	2.53	2.57	3.03	3.00	3.70	3.73
	40	1.39	1.39	2.00	2.00	2.45	2.47	2.90	2.87	3.51	3.53
15	5	1.89	1.86	3.24	3.20	4.62	4.57	6.43	6.34	9.72	9.69
	15	1.43	1.48	1.97	2.06	2.40	2.51	2.86	2.89	3.52	3.57
	30	1.32	1.34	1.72	1.76	2.01	2.06	2.31	2.28	2.70	2.72
	40	1.30	1.31	1.66	1.68	1.92	1.95	2.18	2.14	2.52	2.53
30	5	1.88	1.85	3.17	3.14	4.50	4.44	6.23	6.13	9.38	9.34
	15	1.40	1.45	1.87	1.96	2.25	2.35	2.64	2.67	3.21	3.26
	30	1.28	1.30	1.61	1.64	1.84	1.89	2.07	2.05	2.39	2.40
	40	1.25	1.27	1.54	1.56	1.74	1.77	1.94	1.90	2.20	2.21
40	5	1.88	1.85	3.16	3.12	4.46	4.40	6.18	6.07	9.29	9.25
	15	1.39	1.43	1.85	1.93	2.20	2.30	2.59	2.60	3.13	3.17
	30	1.27	1.29	1.57	1.61	1.79	1.84	2.01	1.98	2.30	2.32
	40	1.24	1.25	1.51	1.53	1.69	1.72	1.88	1.83	2.11	2.12

[a] Tab. = tabulated, from [1]

Fit = fitted by the procedure of this paper.

The total number of parameters is the sum of 10 (one for each θ or T), and 4×17 (four for each of the three eigenvectors v, and four for each of the 14 vectors occuring in eqs (5), (6), and (7)); i.e., 78. As mentioned in Part II, this number can be somewhat reduced through algebraic manipulation, but this is unnecessary for a fitting process carried out on a programmable calculator or on a computer.

We finally repeat our previous assertion (see also Part II) that these 78 parameters fit not only a table of 100 observations (Table 1), (which would be a waste of time) but actually *any F* value, for *P* between 1 and 25 percent, and for ν_1 and ν_2 between 4 and ∞.

7. Conclusion for Part III

Through repeated application of the procedure given in Part II, it is possible to fit functions of more than two arguments, provided the data appear as a complete factorial. This is accomplished by first combining all combinations of two or more factors into one factor until a two-way table is obtained. The parameter vectors of the SVD of this table are then expressed as two-way tables themselves and further SVD's are carried out. The procedure is simple in principle but can become quite cumbersome in practice. It is not recommended for functions of more than three arguments, unless no other appropriate fitting procedure is available.

8. References

[1] Pearson, E. S., and H. O. Hartley, Editors, *Biometrika Tables for Statisticians*, (Cambridge University Press, London, 1970).

The Refractivity of Air

Frank E. Jones*

National Bureau of Standards, Washington, DC 20234

July 23, 1980

The air density equation of Jones, Edlén's dispersion formula for standard air, and Edlén's empirically-derived expressions for the effects of CO_2 abundance and water vapor partial pressure on refractivity have been combined into a simplified equation for the refractivity of air, and estimates have been made of uncertainties in calculated refractivity. Under ambient conditions typical of metrology laboratories, the agreement between the simplified equation and Edlén's formulation is well within the uncertainty in each. The simplified equation is valid in the visible region.

Key words: Air density; index of refraction of air; refractivity of air; wavelength of light in air.

1. Introduction

In metrological applications of wavelengths of light in air, it is necessary to calculate the wavelength at ambient conditions of temperature (T), pressure (P), effective water vapor partial pressure (e'), and CO_2 abundance (x_{CO_2}), using the refractive index of air under these conditions. The relation between λ_{vac}, the vacuum wavelength, λ_{air}, the wavelength in air, and n, the refractive index of air, is $\lambda_{vac} = n \lambda_{air}$. Edlén [1] [1] has derived a dispersion formula for standard air ($T = 288.15K$, $P = 101325$ Pa, $e' = 0$, $x_{CO_2} = 0.0003$ by volume) and a formulation for the refractivity of ambient air, $(n - 1)_{tpf}$. Edlén's formulation is in general use in metrology. Jones [2] has recently published a reformulation of the equation for the density of air and applied it to the transfer of the mass unit. It is the purpose of the present paper to combine the air density equation, Edlén's dispersion formula for standard air, and Edlén's empirically-derived expressions for the effects of CO_2 abundance and water vapor partial pressure on refractivity, and in so doing to develop a simpler formulation and to estimate uncertainties in the calculated refractivity.

The Edlén 1966 [1] dispersion formula for standard air is

$$(n-1)_s \times 10^8 = 8342.13 + 2406030 \, (130-\sigma^2)^{-1} + 15997 \, (38.9 - \sigma^2)^{-1}, \quad (1)$$

where n is the refractive index, σ is the vacuum wave number, $(1/\lambda_{vac})$, in μm^{-1} and standard air is dry air at 288.15K,

101325 Pa and a CO_2 abundance of 0.0003 by volume. Edlén [1] expressed the refractivity, $(n-1)_{tp}$ of dry air at temperature t (in °C) and pressure p (in torr) as

$$(n-1)_{tp} = K_\lambda \, D_{tp}, \quad (2)$$

where K_λ [3] is a dispersion factor which is independent of t and p, and the density factor, D_{tp}, is

$$D_{tp} = p \, (1 + \epsilon_t \, p)/\left\{(1 + \alpha t)\left[1 - \frac{(n-1)_{tp}}{6}\right]\right\}, \quad (3)$$

where $\alpha = 1/273.15$ and ϵ_t is a factor which multiplies p in an expression for the nonideality of the gas. By substituting suitable values, (3) becomes

$$D_{tp} = p \, [1 + p \, (0.817 - 0.0133 \, t) \times 10^{-6}]/(1 + 0.0036610 \, t). \quad (4)$$

For air with a CO_2 abundance of x by volume, Edlén derived

$$(n - 1)_x = [1 + 0.540 \, (x - 0.0003)] \, (n - 1)_s, \quad (5)$$

and,

$$n_{tph} - n_{tp} = -h \, (5.722_4 - 0.0457 \, \sigma^2) \times 10^{-8} \quad (6)$$

for the difference in refractive index of moist air holding h torr of water vapor at a total pressure p. (To avoid using the same symbol for two different quantities, in the present work h has been substituted for Edlén's f).

From (4) and the relation

$$(n - 1)_{tp} = (n - 1)_s \, D_{tp}/D_s, \quad (7)$$

27

Edlén's general formula is

$$(n - 1)_{tp} = \frac{p\,(n - 1)_s}{720.775} \cdot$$

$$\frac{[1 + p\,(0.817 - 0.0133\,t) \times 10^{-6}]}{[1 + 0.0036610\,t]}, \tag{8}$$

where D_s ($= 720.775$) is the density factor for standard air.

Equations (5), (6), and (8) are generally combined in the calculation of the refractivity of moist air, in the visible region.

2. Present Formulation

In the following, the air density equation derived by Jones [2] will be incorporated into a refractivity equation. The density of moist air, ϱ, is given by [2]

$$\varrho = \frac{P\,M_a}{RTZ}\left[1 - \left(1 - \frac{18.0152}{M_a}\right)\frac{U}{100}\frac{f\,e_s}{P}\right], \tag{9}$$

where P is the pressure in Pa, M_a is the apparent molecular weight of dry air, R is the universal gas constant, T is the temperature in kelvins, Z is the compressibility factor (the non-ideality of the air-water vapor mixture is reflected in the departure of Z from 1), U is the relative humidity in percent, and f is the enhancement factor (a factor which expresses the fact that the effective saturation vapor pressure of water in air is greater than the saturation vapor pressure, e_s, of pure phase over a plane surface of pure ordinary liquid water). Tables of Z, e_s and f are provided in the appendix of the present paper.

The Lorentz-Lorenz [4,5] formulation of the Clausius-Mossotti [6,7] equation can be expressed as

$$\frac{n^2 - 1}{n^2 + 2} = C\,\frac{\varrho_a}{M_a}, \tag{10}$$

the left side of which can be approximated [1] by $\frac{2}{3}(n-1) \cdot [1 - (n-1)/6]$. Therefore,

$$(n-1) = C'\,\frac{\varrho_a}{M_a}\left[1 - \frac{(n-1)}{6}\right]^{-1}, \tag{11}$$

where ϱ_a and M_a are the density and apparent molecular weight, respectively, of dry air and C and C' are constants. Since $\varrho_a = PM_a/RTZ$ [2], (11) becomes

$$(n-1) = \frac{C'P}{RTZ\left[1 - \frac{(n-1)}{6}\right]}, \tag{12}$$

and for standard air,

$$(n-1)_s = \frac{C'P_s}{RT_sZ_s\left[1 - \frac{(n-1)_s}{6}\right]}. \tag{13}$$

By dividing (12) by (13),

$$(n-1) = \frac{\left[1 - \frac{(n-1)_s}{6}\right]}{\left[1 - \frac{(n-1)}{6}\right]}\frac{P/TZ}{P_s/T_sZ_s}(n-1)_s. \tag{14}$$

By substituting the appropriate values of P_s (101325 P_a), T_s (288.15K) and Z_s (0.99958 from table 1 in the appendix), (14) becomes

$$(n-1) = 0.0028426\,\frac{P}{TZ}\frac{\left[1 - \frac{(n-1)_s}{6}\right]}{\left[1 - \frac{(n-1)}{6}\right]}(n-1)_s, \tag{15}$$

which, when rearranged, becomes

$$(n-1)^2 - 6(n-1) + 0.0170556\,(n-1)_s \cdot$$
$$\left[1 - \frac{(n-1)_s}{6}\right]\frac{P}{TZ} = 0. \tag{16}$$

The appropriate square root of (16) is

$$(n-1) = 3 - \left\{9 - 0.0028426\,(n-1)_s \cdot\right.$$
$$\left.[6 - (n-1)_s]\frac{P}{TZ}\right\}^{1/2}. \tag{17}$$

We shall return now to Edlén's development and combine (2) with (3):

$$(n-1)_{tp} = K_\lambda D_{tp} = \frac{K_\lambda p\,(1 + \epsilon_t p)}{(1 + \alpha t)\left[1 - \frac{(n-1)_{tp}}{6}\right]}. \tag{18}$$

$(1 + \epsilon_t p)$ is recognized to be $1/Z$, $(1 + \alpha t) = T/273.15$, and $p = 760\,P/101325$; therefore,

$$(n-1)_{tp} = \frac{760 \times 273.15}{101325}\frac{K_\lambda P}{TZ}\frac{1}{\left[1 - \frac{(n-1)_{tp}}{6}\right]}. \tag{19}$$

By comparing (19) with (12), $K_\lambda R(760 \times 273.15)/101325$ is seen to correspond to C'.

It remains now to combine (17) with Edlén's empirically-derived expressions for the effects of CO_2 abundance, (5), and water vapor partial pressure, (6), to arrive at the general expression:

$$(n-1) = 3 - \left\{9 - (n-1)_s\,[6 - (n-1)_s] \cdot\right.$$
$$\left.0.0028426\,\frac{P}{TZ}\right\}^{1/2}$$
$$- f e_s\,\frac{U}{100}\,(0.042922 - 0.000343\,\sigma^2) \times 10^{-8}, \tag{20}$$

where e_s is in Pa. Equation (20) corresponds to (8) combined with (5) and (6), i.e. Edlén's formulation [1]. The agreement between the refractivity of moist air calculated using (20)

28

and Edlén's formulation is illustrated for $T = 293.15K$, $P = 101325$ Pa, $U = 50$, $x_{CO_2} = 0.00043$, $Z = 0.99963$ (from table 1), $f = 1.0041$ (from table 2), $e_s = 2338$ Pa (from table 3) and $\lambda_k = \sigma^{-1} = 0.6329912714\ \mu m$ for an iodine stabilized helium-neon laser [8]. Using (20), $(n - 1) = 27131._0 \times 10^{-8}$; using Edlén's formulation $(n - 1)_{tph} = 27131._3 \times 10^{-8}$. For a more extreme case ($T = 288.15K$, $P = 70000$ Pa, $U = 50$, $x_{CO_2} = 0.00080$, $Z = 0.99971$, $f = 1.0030$, $e_s = 1705$ Pa, (for the same wavelength), (20) gives $(n - 1) = 19069._6 \times 10^{-8}$, and the Edlen formulation gives $(n - 1)_{tph} = 19068._1 \times 10^{-8}$. As will be demonstrated in the next section, the difference between the results for the two formulations is well within the uncertainty of each.

Equation (15) can be approximated by

$$(n-1) = 0.0028426\ \frac{P}{TZ}\ (n-1)_s; \qquad (21)$$

in the first of the above examples, the resulting change is 0.02×10^{-8} which is negligible. Equation (20) then becomes

$$(n-1)_{TPe}' = 0.0028426\ \frac{P}{TZ}\ (n-1)_x$$

$$-\ f\, e_s\ \frac{U}{100}(0.042922 - 0.000343\ \sigma^2) \times 10^{-8}, \quad (22)$$

where the subscript TPe' follows Edlén's convention, $e' = fe_sU/100$. For a CO_2 abundance of 0.0003 by volume and a vacuum wavelength of $0.6329912714\ \mu m$ (22) becomes

$$(n-1)_{TPe}' = (78.603\ \frac{P}{TZ} - 0.042066\, fe_s\ \frac{U}{100}) \times 10^{-8}. \quad (23)$$

The variation of CO_2 abundance, x, can be incorporated in (23) by multiplying 78.603 by $[1 + 0.540\ (x - 0.0003)]$. At NBS, a constant value of 1.0042 can be used for f [2] with negligible effect on calculated $(n-1)_{TPe}'$. Equation (23) then becomes

$$(n-1)_{TPe}' = (78.603\ \frac{P}{TZ} - 0.042243\ e_s\ \frac{U}{100}) \times 10^{-8}. \quad (24)$$

3. Estimation of Uncertainties

We follow the suggested practice of Eisenhart [9, 10] in stating separately the random and systematic components of the estimated uncertainties. The stated random component is one standard deviation; the stated systematic component is one-third of the half-width of the interval between the bounds on the systematic error.

The uncertainties in calculated $(n-1)_{TPe}'$ due to estimated uncertainties [2] in P, T, Z, U, f, e_s, and x can be estimated from equation (22). We shall not attempt to estimate the uncertainties in Edlén's [1] dispersion formula

for standard air and his expressions for the effects of CO_2 abundance and water vapor partial pressure. The state-of-the-art in pressure measurement [11] permits the measurement of pressure in a laboratory with a random relative uncertainty of less than ± 0.02 percent, calibration of pressure measuring instruments against a primary standard of pressure contributes a systematic relative uncertainty of about ± 0.003 percent. The corresponding uncertainties in $(n-1)_{TPe}'$, in the first example above are ± 5.4×10^{-8} and ± 0.8×10^{-8}.

The measurement of temperature in the air path is potentially as critical as the pressure measurement, in terms of its effect on the uncertainty in the calculated $(n-1)_{TPe}'$; it is possible to make only a rough estimate of the uncertainty in the temperature measurement. If the vicinity of the path were instrumented with a network of thermopile junctions, the measurements would be expected to have a standard deviation of about ± 0.05K [12] and a systematic uncertainty of the order the ± 0.01K. The corresponding uncertainties in $(n-1)_{TPe}'$ in the first example are ± 4.6×10^{-8} and ± 0.9×10^{-8}.

The estimated systematic relative uncertainty in the compressibility factor, Z, for the first example is ± 0.0017 percent. The corresponding uncertainty in $(n-1)_{TPe}'$ is ± 0.5×10^{-8}.

The uncertainty in calculated $(n-1)_{TPe}'$ due to humidity measurement can be estimated from the second term in (22). The state-of-the-art in humidity measurement [13] permits the measurement of relative humidity, U, with a random uncertainty of ± 0.5 percent relative humidity and a systematic uncertainty of ± 0.3 percent relative humidity. The corresponding uncertainties in $(n-1)_{TPe}'$ in the first example are ± 0.5×10^{-8} and ± 0.3×10^{-8}. The uncertainties contributed by uncertainties in f and e_s are negligible [2].

The uncertainty in calculated $(n-1)_{TPe}'$ due to a variation in CO_2 abundance, x, can be estimated from (5). In the first example, a variation in x of ± 0.0001 corresponds to a systematic uncertainty in $(n-1)_{TPe}'$ of ± 1.5×10^{-8}.

The overall random uncertainty in $(n-1)_{TPe}'$, estimated by combining the random uncertainties by quadrature, is ± 7.1×10^{-8}. The overall systematic uncertainty, estimated by combining the addition, is ± 2.5×10^{-8}. The systematic uncertainty due to variation in CO_2 abundance is necessarily not included. It should be emphasized that these uncertainties are based on the *best possible measurements* of P, T and U.

4. Direct Determination of Air Density

In 1967, Bowman and Schoonover [14] used a pair of stainless steel weights (one of which was hollow) of nearly equal mass but of grossly different volume to make direct

determination of the air density in a balance case, thus avoiding the uncertainties in the parameters and environmental variables in air density calculations. A similar scheme will be used in the transfer of the mass unit [15].

Having estimated the uncertainty in calculated $(n-1)_{TP e'}$ due to the uncertainties in the various variables to be about $\pm 1 \times 10^{-7}$ at the level of the equivalent of 1 standard deviation, it is of interest to estimate how much improvement would result from the *direct* determination of air density, ϱ, if practicable. From (9),

$$\frac{P}{TZ} = \frac{\varrho R}{M_a} \frac{1}{\left[1 - (1 - \frac{18.0152}{M_a}) \frac{U}{100} \frac{f e_s}{P} \right]}, \quad (25)$$

where $M_a = 28.963 + 12.011 (x_{CO_2} - 0.00033)$; recalling that ϱ is the density of moist air. By substituting (25) in (22),

$$(n-1)_{TP e'} = 0.0028426 \frac{\varrho R}{M_a} \cdot$$

$$\frac{(n-1)_x}{\left[1 - (1 - \frac{18.0152}{M_a}) \frac{U}{100} \frac{f e_s}{P} \right]}$$

$$- f e_s \frac{U}{100} (0.042922 - 0.000343 \, \sigma^2) \times 10^{-8}. \quad (26)$$

The uncertainties in the various parameters in (25), other than ϱ and $(n-1)_x$, are taken from [2]. The resulting overall uncertainty in the calculated $(n-1)_{TP e'}$ are $\pm 1.9 \times 10^{-8}$ random and $\pm 1.8 \times 10^{-8}$ systematic. The uncertainty due to the effect on M_a of a variation of x_{CO_2}, 1.1×10^{-8} per 0.0001, has necessarily not been included. It can be concluded that even it the uncertainty in a direct determination of ϱ were negligible, the uncertainty in $(n-1)_{TP e'}$ due to the uncertainties in the various variables and parameters would be reduced by a factor of about 2.5. The major contributors to the uncertainty in $(n - 1)_{TP e'}$ are the uncertainties in R, M_a and U.

5. Conclusions

Jones's air density equation [2], Edlén's [1] dispersion formula for standard air, and Edlén's empirically-derived expressions for the effects of CO_2 abundance and water vapor partial pressure on refractivity have been combined into a simple refractivity of air equation, and estimates have been made of uncertainties in calculated refractivity.

The general equation is (22), which is valid in the visible region; tables of Z, f and e_s have been included in the appendix of this paper. The overall estimated uncertainty is about $\pm 1 \times 10^{-7}$ at the level of the equivalent of 1 standard deviation. The major contributors to the uncertainty in refractivity are the uncertainties in the measurements of pressure and temperature. The magnitude of the uncertainty due to variation in CO_2 concentration can approach that of the uncertainties due to the pressure and temperature measurements. Therefore, the CO_2 concentration should be treated as a variable and should be observed.

If it were practicable to make a direct measurement of air density representative of the air path, the uncertainty in calculated refractivity due to the uncertainties in the various variables and parameters would be reduced by a factor of about 2.5.

---

6. References

[1] Edlén, B. The refractive index of air. Metrologia. 2: 71-80; 1966.

[2] Jones, F. E. The air density equation and the transfer of the mass unit. J. Res. Nat. Bur. Stand. (U.S.). 83(5): 419-428; 1978 September- October.

[3] Barrell, H.; Sears, J. E., Jr. The refraction and dispersion of air for the visible spectrum. Phil. Trans. Roy. Soc. London. A238: 1-64; 1939.

[4] Lorentz, H. A. Ueber die Beziehung zwischen der Fortpflanzungsgeschwindigkeit des Lichtes und der Korperdichte. Ann. Physik u. Chemie. N. F. 9: 641-665; 1880.

[5] Lorenz, L. Ueber die Refractionsconstante. Ann. Physik u. Chemie. N. F. 11: 70-103; 1880.

[6] Clausius, R. Die Mechanische Warmetheorie, Vol. 2. Braunschweig: Friedrich Vieweg und Sohn; 1879. 62-97.

[7] Mossotti, O. F. Memorie di matematica e di fisica della Societa Italiana delle Scienze Residente in Modena. (26)(2): 49-74; 1850.

[8] Schweitzer, W. G.; Kessler, E. G., Jr.; Deslattes, R. D.; Layer, H. P.; Whetstone, J. R. Description, performance, and wavelengths of iodine stabilized lasers. Appl. Opt. 12(12): 2927-2938; 1973.

[9] Eisenhart, C. Expression of the uncertainties of final results. Science. 160: 1201-1204; 1968.

[10] Eisenhart, C. Realistic evaluation of the precision and auccuracy of instrument calibration systems. J. Res. Nat. Bur. Stand. (U.S.). 67C: 161-187; 1963 April-June.

[11] Heydemann, P. L. M. Private communication.

[12] Roeser, W. F.; Wensel, H. T. *Temperature, its measurement and control in science and industry.* New York: Reinhold Publ. Corp.; 1941. p. 311.

[13] Wexler, A. Measurement of humidity in the free atmosphere near the surface of the earth. Meteorological Monographs. 11(33): 262-282; 1970.

[14] Bowman, H. A.; Schoonover, R. M.; Jones, M. W. Procedure for high precision density determinations by hydrostatic weighing. J. Res. Nat. Bur. Stand. (U.S.). 71C(3): 179-198; 1967 July-September.

[15] Schoonover, R. M. Private communication.

[16] Hyland, R. W. A correlation of the second virial coefficients and enhancement factors for moist air. J. Res. Nat. Bur. Stand. (U.S.). 79A(4): 551-560; 1975 July-August.

[17] Wexler, A.; Greenspan, L. Vapor pressure equation for water in the range 0 to 100 °C. J. Res. Nat. Bur. Stand. (U.S.). 75A(3): 213-230; 1971 May-June.

7. Appendix

TABLE 1. Compressibility factor, Z, for air containing reasonable amounts of CO_2 [2]

Temperature (Celsius)	Pressure (pascals)	(mm Hg)	Relative Humidity in Percent				
			0	25	50	75	100
15.0	70000	525.0	.99971	.99970	.99968	.99967	.99965
	75000	562.5	.99969	.99968	.99966	.99965	.99963
	80000	600.0	.99966	.99966	.99964	.99963	.99961
	85000	637.6	.99964	.99963	.99962	.99961	.99959
	90000	675.1	.99962	.99961	.99960	.99959	.99957
	95000	712.6	.99960	.99959	.99958	.99957	.99955
	100000	750.1	.99958	.99957	.99956	.99955	.99953
	101325	760.0	.99958	.99957	.99956	.99955	.99953
	105000	787.6	.99956	.99955	.99954	.99953	.99951
	110000	825.1	.99954	.99953	.99952	.99951	.99949
16.0	70000	525.0	.99971	.99970	.99969	.99967	.99965
	75000	562.5	.99969	.99968	.99967	.99965	.99963
	80000	600.0	.99967	.99966	.99965	.99963	.99962
	85000	637.6	.99965	.99964	.99963	.99962	.99960
	90000	675.1	.99963	.99962	.99961	.99959	.99958
	95000	712.6	.99961	.99960	.99959	.99957	.99956
	100000	750.1	.99959	.99958	.99957	.99956	.99954
	101325	760.0	.99959	.99958	.99956	.99955	.99953
	105000	787.6	.99957	.99956	.99955	.99954	.99952
	110000	825.1	.99955	.99954	.99953	.99952	.99950
17.0	70000	525.0	.99972	.99971	.99970	.99968	.99966
	75000	562.5	.99970	.99969	.99968	.99966	.99964
	80000	600.0	.99968	.99967	.99966	.99964	.99962
	85000	637.6	.99966	.99965	.99964	.99962	.99960
	90000	675.1	.99964	.99963	.99962	.99960	.99958
	95000	712.6	.99962	.99961	.99960	.99958	.99956
	100000	750.1	.99960	.99959	.99958	.99956	.99954
	101325	760.0	.99960	.99959	.99957	.99956	.99954
	105000	787.6	.99958	.99957	.99956	.99954	.99953
	110000	825.1	.99956	.99955	.99954	.99952	.99951
18.0	70000	525.0	.99973	.99972	.99970	.99968	.99966
	75000	562.5	.99971	.99970	.99968	.99966	.99964
	80000	600.0	.99969	.99968	.99966	.99964	.99962
	85000	637.6	.99967	.99966	.99964	.99963	.99960
	90000	675.1	.99965	.99964	.99962	.99961	.99959
	95000	712.6	.99963	.99962	.99961	.99959	.99957
	100000	750.1	.99961	.99960	.99959	.99957	.99955
	101325	760.0	.99961	.99960	.99958	.99957	.99955
	105000	787.6	.99959	.99958	.99957	.99955	.99953
	110000	825.1	.99957	.99956	.99955	.99953	.99951
19.0	70000	525.0	.99973	.99972	.99971	.99968	.99966
	75000	562.5	.99972	.99970	.99969	.99966	.99964
	80000	600.0	.99970	.99968	.99967	.99965	.99963
	85000	637.6	.99968	.99967	.99965	.99963	.99961
	90000	675.1	.99966	.99965	.99963	.99961	.99959
	95000	712.6	.99964	.99963	.99961	.99960	.99957
	100000	750.1	.99962	.99961	.99959	.99958	.99956
	101325	760.0	.99962	.99960	.99959	.99957	.99955
	105000	787.6	.99960	.99959	.99958	.99956	.99954
	110000	825.1	.99958	.99957	.99956	.99954	.99952
20.0	70000	525.0	.99974	.99973	.99971	.99969	.99966
	75000	562.5	.99972	.99971	.99969	.99967	.99964
	80000	600.0	.99970	.99969	.99967	.99965	.99963
	85000	637.6	.99969	.99967	.99966	.99964	.99961
	90000	675.1	.99967	.99966	.99964	.99962	.99959
	95000	712.6	.99965	.99964	.99962	.99960	.99958
	100000	750.1	.99963	.99962	.99960	.99958	.99956
	101325	760.0	.99963	.99961	.99960	.99958	.99956
	105000	787.6	.99961	.99960	.99958	.99956	.99954
	110000	825.1	.99959	.99958	.99957	.99955	.99953
21.0	70000	525.0	.99975	.99973	.99971	.99969	.99966
	75000	562.5	.99973	.99972	.99970	.99967	.99964
	80000	600.0	.99971	.99970	.99968	.99966	.99963
	85000	637.6	.99969	.99968	.99966	.99964	.99961
	90000	675.1	.99968	.99966	.99965	.99962	.99960
	95000	712.6	.99966	.99965	.99963	.99961	.99958
	100000	750.1	.99964	.99963	.99961	.99959	.99956
	101325	760.0	.99964	.99962	.99961	.99959	.99956
	105000	787.6	.99962	.99961	.99959	.99957	.99955
	110000	825.1	.99960	.99959	.99958	.99956	.99953

Temperature (Celsius)	Pressure (pascals)	(mm Hg)	Relative Humidity in Percent				
			0	25	50	75	100
22.0	70000	525.0	.99975	.99974	.99972	.99969	.99966
	75000	562.5	.99974	.99972	.99970	.99968	.99964
	80000	600.0	.99972	.99971	.99969	.99966	.99963
	85000	637.6	.99970	.99969	.99967	.99964	.99961
	90000	675.1	.99968	.99967	.99965	.99963	.99960
	95000	712.6	.99967	.99965	.99963	.99961	.99958
	100000	750.1	.99965	.99964	.99962	.99960	.99957
	101325	760.0	.99965	.99963	.99961	.99959	.99956
	105000	787.6	.99963	.99962	.99960	.99958	.99955
	110000	825.1	.99962	.99960	.99958	.99956	.99954
23.0	70000	525.0	.99976	.99975	.99972	.99969	.99966
	75000	562.5	.99974	.99973	.99971	.99968	.99964
	80000	600.0	.99973	.99971	.99969	.99966	.99963
	85000	637.6	.99971	.99969	.99967	.99965	.99962
	90000	675.1	.99969	.99968	.99966	.99963	.99960
	95000	712.6	.99968	.99966	.99964	.99962	.99959
	100000	750.1	.99966	.99964	.99962	.99960	.99957
	101325	760.0	.99965	.99964	.99962	.99960	.99957
	105000	787.6	.99964	.99961	.99961	.99959	.99956
	110000	825.1	.99963	.99961	.99959	.99957	.99954
24.0	70000	525.0	.99977	.99975	.99973	.99969	.99965
	75000	562.5	.99975	.99973	.99971	.99968	.99964
	80000	600.0	.99973	.99972	.99970	.99967	.99963
	85000	637.6	.99972	.99970	.99968	.99965	.99962
	90000	675.1	.99970	.99969	.99966	.99964	.99960
	95000	712.6	.99968	.99967	.99965	.99962	.99959
	100000	750.1	.99967	.99965	.99963	.99961	.99957
	101325	760.0	.99966	.99965	.99963	.99960	.99957
	105000	787.6	.99965	.99962	.99962	.99959	.99956
	110000	825.1	.99964	.99962	.99960	.99957	.99954
25.0	70000	525.0	.99977	.99976	.99973	.99970	.99965
	75000	562.5	.99976	.99974	.99971	.99968	.99964
	80000	600.0	.99974	.99972	.99970	.99967	.99963
	85000	637.6	.99973	.99971	.99968	.99965	.99962
	90000	675.1	.99971	.99969	.99967	.99964	.99960
	95000	712.6	.99969	.99968	.99965	.99962	.99959
	100000	750.1	.99968	.99966	.99964	.99961	.99958
	101325	760.0	.99967	.99966	.99963	.99961	.99957
	105000	787.6	.99966	.99964	.99962	.99960	.99956
	110000	825.1	.99965	.99963	.99961	.99958	.99955
26.0	70000	525.5	.99978	.99976	.99973	.99970	.99965
	75000	562.5	.99976	.99975	.99972	.99968	.99964
	80000	600.0	.99975	.99973	.99970	.99967	.99963
	85000	637.6	.99973	.99971	.99969	.99966	.99961
	90000	675.1	.99972	.99970	.99967	.99964	.99960
	95000	712.6	.99970	.99968	.99966	.99963	.99959
	100000	750.1	.99969	.99967	.99964	.99961	.99958
	101325	760.0	.99968	.99966	.99964	.99961	.99957
	105000	787.6	.99967	.99965	.99963	.99960	.99956
	110000	825.1	.99966	.99964	.99961	.99959	.99955
27.0	70000	525.0	.99979	.99977	.99974	.99969	.99964
	75000	562.5	.99977	.99975	.99972	.99968	.99963
	80000	600.0	.99976	.99974	.99971	.99967	.99962
	85000	637.6	.99974	.99972	.99969	.99966	.99961
	90000	675.1	.99973	.99971	.99968	.99964	.99960
	95000	712.6	.99971	.99969	.99966	.99963	.99959
	100000	750.1	.99970	.99968	.99965	.99962	.99958
	101325	760.0	.99969	.99967	.99965	.99961	.99957
	105000	787.6	.99968	.99966	.99964	.99960	.99956
	110000	825.1	.99966	.99965	.99962	.99959	.99955
28.0	70000	525.0	.99979	.99977	.99974	.99969	.99964
	75000	652.5	.99978	.99976	.99972	.99968	.99963
	80000	600.0	.99976	.99974	.99971	.99967	.99962
	85000	637.6	.99975	.99973	.99970	.99966	.99961
	90000	675.1	.99973	.99971	.99968	.99965	.99960
	95000	712.6	.99972	.99970	.99967	.99963	.99959
	100000	750.1	.99970	.99968	.99966	.99962	.99958
	101325	760.0	.99970	.99968	.99965	.99962	.99957
	105000	787.6	.99969	.99967	.99964	.99961	.99956
	110000	825.1	.99967	.99965	.99963	.99959	.99955

Pressure, pascals	t, C			
	15	20	25	30
70 000	1.0030	1.0031	1.0032	1.0034
75 000	1.0032	1.0033	1.0034	1.0035
80 000	1.0033	1.0034	1.0035	1.0037
85 000	1.0035	1.0036	1.0037	1.0038
90 000	1.0036	1.0037	1.0038	1.0040
95 000	1.0038	1.0039	1.0040	1.0041
100 000	1.0039	1.0040	1.0042	1.0043
101 325	1.0040	1.0041	1.0042	1.0043
105 000	1.0041	1.0042	1.0043	1.0045
110 000	1.0043	1.0043	1.0045	1.0046

TABLE 3. Values of saturation water vapor pressure, e_s, calculated
using formulation of Wexler and Greenspan [17]

e_s, pascals

Temperature, C

	15	16	17	18	19	20	21	22	23	24	25	26	27
.00	1705	1818	1938	2064	2197	2338	2487	2644	2810	2985	3169	3363	3567
.05	1711	1824	1944	2070	2204	2346	2495	2652	2818	2994	3178	3372	3577
.10	1716	1830	1950	2077	2211	2353	2503	2660	2827	3003	3188	3382	3588
.15	1722	1836	1956	2083	2218	2360	2510	2669	2836	3012	3197	3392	3598
.20	1727	1841	1962	2090	2225	2367	2518	2677	2844	3021	3207	3402	3609
.25	1733	1847	1968	2097	2232	2375	2526	2685	2853	3030	3216	3413	3619
.30	1738	1853	1975	2103	2239	2382	2533	2693	2861	3039	3226	3423	3630
.35	1744	1859	1981	2110	2246	2390	2541	2701	2870	3048	3235	3433	3641
.40	1749	1865	1987	2116	2253	2397	2549	2709	2879	3057	3245	3443	3651
.45	1755	1871	1994	2123	2260	2404	2557	2718	2887	3066	3255	3453	3662
.50	1761	1877	2000	2130	2267	2412	2565	2726	2896	3075	3264	3463	3673
.55	1766	1883	2006	2136	2274	2419	2573	2734	2905	3085	3274	3473	3683
.60	1772	1889	2012	2143	2281	2427	2580	2743	2914	3094	3284	3484	3694
.65	1778	1895	2019	2150	2288	2434	2588	2751	2922	3103	3294	3494	3705
.70	1783	1901	2025	2157	2295	2442	2596	2759	2931	3112	3303	3504	3716
.75	1789	1907	2032	2163	2302	2449	2604	2768	2940	3122	3313	3515	3727
.80	1795	1913	2038	2170	2310	2457	2612	2776	2949	3131	3323	3525	3738
.85	1801	1919	2044	2177	2317	2464	2620	2785	2958	3140	3333	3535	3749
.90	1806	1925	2051	2184	2324	2472	2628	2793	2967	3150	3343	3546	3759
.95	1812	1931	2057	2190	2331	2480	2636	2801	2976	3159	3353	3556	3770

The Configurations $3d^n4p + 3d^{n-1}4s4p + 3d^{n-2}4s^24p$ in the First Spectra of the Iron Group

Charles Roth

Department of Mathematics, McGill University, Montreal, Quebec

September 1, 1980

Energy levels and Lande g-factors for the configurations $3d^n4p + 3d^{n-1}4s4p + 3d^{n-2}4s^24p$ in the first spectra of the iron group were calculated and compared with experimental values, in both general and individual treatments. The calculations were done in intermediate coupling taking into account explicity the interactions between configurations, as well as complete effective interactions of the core, and effective interactions $d - p$. Due to a successful starting point based on Hartree-Fock calculations for the Slater parameters, as well as the insertion of the effective interactions, considerable improvement was obtained compared to previous results.

On fitting 1537 levels using 67 free interaction parameters a mean error of 182 cm^{-1} was obtained. Altogether 3652 energy levels were calculated including all the levels for the configurations $3d^{n-2}4s^24p$ across the sequence. It was shown that all interaction parameters could be expressed either as linear functions, or linear functions with small quadratic corrections, of the atomic number. There was general qualitative agreement between the values of the parameters calculated using the semi-empirical method and those calculated using Hartree-Fock methods. There remained some isolated terms with large deviations. These are attributed to be due to the interactions with the configurations $(3d + 4s)^n5p$, that were not considered explicitly in this analysis.

Tables comparing the experimental and calculated energy levels and Lande g-factors, as well as detailed analyses for each spectrum are given in another paper.

Key words: Iron group elements; least squares optimization; theoretical spectroscopy.

1. Introduction.

Traditionally, theoretical spectroscopists consider the radial Slater integrals as unknown parameters, obtaining their values empirically by fitting the experimental data to the calculated energy levels, and then performing least-squares optimization calculations.

For the even configurations $3d^n + 3d^{n-1}4s$, and the odd configurations $3d^n4p$ in the second and third spectra of the iron group, the results were excellent, [1-4].[1] Furthermore, it was shown that the radial parameters are either linear functions, or linear functions with small quadratic corrections, of the atomic number. For the odd configurations in neutral atoms, the interactions between configurations are very strong. Thus the algebraic matrices of the configurations $(d+s)^n p$ were calculated and checked by the author, [5-9]. Theoretical investigations were then performed for the configurations $(3d + 4s)^n4p$ in neutral atoms of calcium, scandium, titanium, vanadium, chromium, manganese, iron, cobalt and nickel, [10-16].

Although the results were good (average r.m.s. error of 210cm^{-1}); a very disturbing feature of the results was the fact that the behavior of the final values of the radial parameters was generally far from linear.

It would be highly anomalous to have the radial parameters behave so irregularly in the first spectra. Thus, in order to overcome this discrepancy it was essential to have improved *initial* values of the radial parameters. Hence the radial parameters were first calculated using the Hartree-Fock method. Least-squares optimization calculations were then performed on these parameters *forcing* them to behave linearly, with possibly at most small quadratic corrections, as functions of the atomic number. The values thus computed were then compared with those obtained previously in individual treatments by the author, [10-16]. Whenever the Hartree-Fock values were *uniformly* higher or lower than those of the previous results, [10-16], appropriate scaling factors were utilized on the linearized Hartree-Fock values, and the latter were than used as initial parameters for this investigation.

---

[1] Figures in brackets indicate literature references at the end of this paper.

In this project were included the electrostatic and spin-orbit interactions of the individual configurations $3d^n4p$, $3d^{n-1}4s4p$ and $3d^{n-2}4s^24p$; the explicit electrostatic interactions between configurations $3d^n4p-3d^{n-1}4s\,4p$, $3d^{n-1}4s4p-3d^{n-2}4s^24p$ and $3d^n4p-3d^{n-2}4s^24p$; and the complete two-and three-body effective interactions of the core d electrons, as well as two-body mixed effective interactions between the $3d$ and $4p$ electrons.

The initial values of the radial parameters were then used to multiply the algebraic matrices on tape and the resulting matrices were diagonalized. Besides the eigenvalues, the diagonalization routine also yields the derivatives of the eigenvalues with respect to the parameters, the squares of the eigenvectors (percentage compositions) and the calculated Lande g values. The appropriate experimental levels were then fitted to the eigenvalues, and using the derivatives obtained in the diagonalization, least squares optimization calculations were performed. In these calculations, the improved values of the theoretical energy levels, the corrected values of the parameters including their statistical deviations and the sum of the squares of the differences between the observed and the calculated levels, were obtained. The rms error is then defined as

$$\Delta = \sqrt{\frac{\sum\limits_{i=1}^{n} \Delta_i^2}{n-m}}$$

where the Δ_i are the differences between the observed and calculated levels, n is the number of known levels and m is the number of free parameters. The mean error is quite different from the mean deviation

$$\delta = \sqrt{\frac{\sum\limits_{i=1}^{n} \Delta_i^2}{n}}$$

as the former takes into account the statistical effect of the number of free parameters. Hence in order for a new parameter to have physical significance, it should cause an essential decrease in the rms error, and not simply a decrease in the mean deviation.

The value of Δ is also given by the least-squares routine. The same derivatives can be used for several variations in the least squares, either imposing different conditions on the parameters, inserting the experimental levels with different assignments, or even rejecting some levels from consideration. The parameters of that variation which yielded the best results were used to perform new diagonalizations. This iterative process was continued until mathematical convergence was attained. In the present project four complete iterations were required.

The use of the same assumptions and the same approximations in all the spectra made it possible to obtain a consistent set of interaction parameters and compare the results obtained from the spectra of different elements. Due to a successful choice of the initial values of the radial parameters, it was shown that the final values can *indeed* be expressed as simple functions of the atomic number. A consistent use of such interpolation formulas for all parameters, combined eleven problems, formerly independent, into one problem. This result, which is significant by itself, very much improved the reliability of the results for those spectra where the experimental data is still scarce, and which are thus most in need of reliable predictions of the unknown levels.

This is particularly true for the configurations $3d^{n-2}4s^24p$. For each individual element there in not a sufficient number of experimental levels in order to predict even approximately the remaining levels. However, by considering a general treatment ALL the levels of the configurations $3d^{n-2}4s^24p$ for the entire sequence were calculated.

For completeness and comparison, individual least squares (ILS) were also performed for each element.

The procedures followed, a description of the various interactions considered, and an analysis of the results and significance of the different parameters are contained in this work. The tables comparing the experimental and calculated energy levels, values for all the theoretical levels specifying their percentage compositions, as well as detailed analyses for each spectrum are given in another paper, [17].

34

2. Effective electrostatic interactions

For the odd configurations in the first spectra of the iron group, both strong and weak configuration interactions are significant. The former arise when the perturbing and perturbed configurations are energetically close to each other and there is strong coupling of the configurations by the Coulomb field. These were taken into account by *explicitly* considering the configurations $3d^n4p$, $3d^{n-1}4s4p$, $3d^{n-2}4s^24p$ and the electrostatic interactions between them. Weak interactions occur when the perturbing configurations are well separated from the perturbed configuration, and the coupling of the Coulomb field is weak.

The individual weak interactions may not be significant, but their cumulative influence may be quite large, due to the increasing density of states as the continuum is approached. As it would be completely futile to consider each of these effects individually, the aim should be to modify the energy matrices of the principal configurations so that the major part of *all* the weakly perturbing configurations be included.

By first order perturbation theory, different configurations do not interact. In second order only those configurations interact that differ in the quantum numbers of at most two electrons. Bacher and Goudsmit, [18], have shown that the terms of the configuration l^n may be expressed as linear combinations of the terms of l^2, so that the perturbation of l^n by *all* the configurations differing from it by the state of two electrons, and being distant from it, can be accounted for by suitably modifying the terms of l^2. Hence these perturbations can be described by two-body effective interactions.

The first correction of this kind for the configurations d^n was the $\alpha L(L+1)$ correction introduced by Trees, [19-20], in the configurations $3d^54s$ of MnII and FeIII. Trees introduced his correction empirically, but Racah, [21], showed that the above effects can be described by a model or effective interaction of the form

$$2\alpha(l_1 \cdot l_2) + \beta q_{12}$$

where q_{12} is the seniority operator, [22]. For the configuration d^n this becomes

$$\alpha[L(L+1)-6n] + \beta Q$$

where

$$\begin{aligned}Q(n,v) &= \tfrac{1}{4}(n-v)(4l+4-n-v)\\ &= \tfrac{1}{4}(n-v)(12-n-v)\end{aligned} \tag{1}$$

is the total seniority operator. Here n is the number of d electrons in the configurations d^np, v is the seniority of the d^n core term, and l is 2 as we are dealing with d electrons. The constant $-6n\alpha$ is usually incorporated into the height of the configuration. Racah, [21], showed that the $\alpha L(L+1)$ and the βQ corrections form a complete set of two-body effective interactions for the d^n configurations. This is due to the fact that together with the Slater integrals $F_0(d^2)$, $F_2(d^2)$ and $F_4(d^2)$, they form a set of five independent parameters that can represent the five terms of d^2.

Bacher and Goudsmit, [18], also showed that if the far-lying perturbing configuration differs from l^n by the state of only one electron, its effect can be described by expressing the terms of l^n as linear combinations of l^2, and modifying the values of these terms. Hence in the linear theory, the Hamiltonian in this case must be augmented by additional three-body interactions.

Rajnak and Wybourne, [23], obtained explicit formulas for the effective interactions representing the perturbation of an l^n configuration by far-lying configurations differing from it by one or two electrons or holes. Racah and Stein, [24], subsequently, developed an elegant method that considerably simplified the calculations of Rajnak and Wybourne.

If A and B represent the perturbed and perturbing configurations, respectively, and if G is the operator representing the Coulomb energy of repulsion between the electrons, $\sum_{i<j} \frac{e^2}{r_{ij}}$, then the matrix elements of the second-order perturbation produced by B on A are given approximately by

$$(A\psi \mid W_2 \mid A\psi') = -\frac{1}{\Delta E} \sum_{\psi''} (A\psi \mid G \mid B\psi'')(B\psi'' \mid G \mid A\psi'), \tag{2}$$

where ΔE is the distance between the centers of gravity of the two configurations, which are assumed to be well separated. According to Racah and Stein, [24], the operator G in the first factor is replaced by a "curtailed" operator g, whose matrix elements $(A'\psi \mid g \mid B'\psi'')$ are equal to those of G if $A' = A$ and $B' = B$, and vanish otherwise. Similarly, the operator G in the second factor is replaced by $\tilde{g}$, defined analogously to g. Then

$$(A\psi \mid W_2 \mid A\psi') = -\frac{1}{\Delta E} \sum_{B'\psi''} (A\psi \mid g \mid B'\psi'')(B'\psi'' \mid \tilde{g} \mid A\psi'), \tag{3}$$

where the summation is over the complete set $B'\psi''$. Thus,

$$(A\psi \mid W_2 \mid A\psi') = -\frac{1}{\Delta E}(A\psi \mid \tilde{g}g \mid A\psi'), \tag{4}$$

and hence the electrostatic interaction between the configuration A and all the other distant configurations may be simply expressed as an effective interaction within the configuration A given by

$$W_2 = -\frac{\tilde{g}g}{\Delta E} \tag{5}$$

Then using either the above method or that of Rajnak and Wybourne, [23], we obtain that the correction term W_2, that must be added to the Hamiltonian of ℓ^n caused by the perturbation of ℓ^n by $\ell^{n-1}\ell'$configurations is given by

$$W_2 = -\sum_{kk'} P(kk'; \ell\ell, \ell\ell') \, \phi(kk'; \ell\ell, \ell\ell') \tag{6}$$

where

$$P(kk'; \ell_a\ell_b, \ell_c\ell_d) = X(k; \ell_a\ell_b, \ell_c\ell_d) \cdot X(k'; \ell_a\ell_b, \ell_c\ell_d)/\Delta E, \tag{7}$$

and

$$X(k; \ell_a\ell_b, \ell_c\ell_d) = (\ell_a \| C^{(k)} \| \ell_c)(\ell_b \| C^{(k)} \| \ell_d) R^k(\ell_a\ell_b, \ell_c\ell_d), \tag{8}$$

ΔE is the distance between the perturbing configuration and ℓ^n,

$$\phi(kk', \ell\ell, \ell\ell') = -\sum_{k''} (2k'' + 1) \left\{ {kk'k'' \atop \ell\ell\ell'} \right\} [U^{(k)} \times U^{(k'')} \times U^{(k')}]^{(0)} \ldots\ldots$$
$$+ [\delta_{\ell\ell'}/(2\ell+1)] [(U^{(k)} \cdot U^{(k)}) + (U^{(k')} \cdot U^{(k')}) - n/(2\ell+1)]. \tag{9}$$

For the perturbation W_2' of ℓ^n by $\ell'^{(4\ell'+1)}\ell^{n+1}$ we similarly obtain

$$W_2' = -\sum_{kk'} P(kk'; \ell\ell, \ell\ell') \, \phi'(kk'; \ell\ell, \ell\ell'), \tag{10}$$

where

$$\phi'(kk'; \ell\ell, \ell\ell') = -\sum_{k''} (-1)^{(k''+1)} (2k''+1) \left\{ {kk'k'' \atop \ell\ell\ell'} \right\} \cdot$$
$$[U^{(k)} \times U^{(k'')} \times U^{(k')}]^{(0)} - [2\delta_{kk'}/(2k+1)] (U^{(k)} \cdot U^{(k)}). \tag{11}$$

The P are radial parameters; the ϕ and ϕ' are their coefficient operators.

The matrix elements of ℓ_n can be calculated by Racah algebra. The variables k and k' are even and nonzero integers that must satisfy the usual triangular conditions of the 6-j symbols. The variables k'' can assume all integral values consistent with the triangular conditions for the 6-j symbols.

The parameter T represents the perturbation of the configuration $3d^n$ by the configuration $3s3d^{n+1}$. It was first considered by Trees, [25], when he investigated the configuration $3s^2 3p^6 3d^6$, and took into account its

36

interaction with the configuration $3s3p^63d^7$. Shadmi, [26] extended the work of Trees to all configurations $3d^n + 3d^{n-1}4s + 3d^{n-2}4s^2$ in the sequence of the second spectra of the iron group by introducing a three-body effective interaction between $3d$ electrons which represented the perturbation of a configuration of the type $3s^23d^n$ by the configuration $3s3d^{n+1}$. Roth [3–4] included the parameters α, β and T in the configurations $3d^n4p$ in the second and third spectra of the iron group, as well as for the configurations $3d^34p + 3d^24s4p$ in V II, [27].

Further important investigations were carried out by Shadmi, Stein, Oreg, Caspi, Goldschmidt and Starkand [28–30].

Now from (10) with $k = k' = 2$ here, we have $W_2' = -P(22;3d3d,3d3s)\phi'(22;dd,ds)$.

$$P(22 ; 3d3d,3d3s) = \frac{X(2 ; 22,20)^2 \, [R^2(dd,ds)]^2}{\Delta E}$$

$$= \frac{(2\,\|\,C^{(2)}\,\|\,2)^2(2\,\|\,C^{(2)}\,\|\,0)^2[R^2(dd,ds)]^2}{\Delta E}$$

$$= \frac{10}{7} \frac{[R^2\,(dd,ds)]^2}{\Delta E} \text{ . Thus,}$$

$$W_2' = -\frac{10}{7} \frac{[R^2(dd,ds)]^2}{\Delta E} \, \phi'(22 ; dd,ds)$$

$$= -\frac{10}{7} \frac{(35)^2H^2}{\Delta E} \, \phi'(22 ; dd,ds)$$

$$= -1750 \, \frac{H^2}{\Delta E} \, \phi'(22; dd,ds),$$

where we used $H = \frac{R^2(dd,ds)}{35}$ according to Racah, [22]. However, in order to be consistant with the definition of

$$T = \frac{H^2}{\Delta E}$$

used in a previous work by the author, [7], we get for the coefficient operator of T the expression

$$t = -1750 \, \phi'(22; dd,ds). \tag{12}$$

The parameters T_x, T_y, T_z represent the perturbation of the configuration $3d^n$ by a configuration of the type $3d^{n-1}n'd$, where $n' \geqslant 4$. Similarly, to the above result for t their coefficient operators t_x, t_y, t_z are

$$t_x = -1750 \, \phi(22; dd,dd'), \tag{13a}$$
$$t_y = -1750 \,[\phi(24; dd,dd') + \phi(42; dd,dd')], \tag{13b}$$
$$t_z = -1750 \, \phi(44; dd,dd'). \tag{13c}$$

In terms of the P, the parameters are given by

$$T = (^1/1750)P(22; 3d3d, 3d3s), \tag{14a}$$
$$T_x = (^1/1750)P(22; 3d3d, 3dn'd), \tag{14b}$$
$$T_y = (^1/1750)P(24; 3d3d, 3dn'd), \tag{14c}$$
$$T_z = (^1/1750)P(44; 3d3d, 3dn'd). \tag{14d}$$

Only the parameters T and T_x were used as T_y and T_z depend upon the other parameters.

2.1. Effective interactions in the configuration $l^n l'$

First we define the tensor operators $u^{(k)}$, $v^{(k)}$, $w^{(k)}$, $\widetilde{w}^{(k)}$, $y^{(k)}$, $\widetilde{y}^{(k)}$, $z^{(k)}$ and $\widetilde{z}^{(k)}$ as the tensor operators of order k whose only non-vanishing reduced matrix elements are

$$(nl\|u^{(k)}\|nl) = 1, \tag{15a}$$

$$(n'l'\|v^{(k)}\|n'l') = 1, \tag{15b}$$

$$(n'l'\|w^{(k)}\|n''l'') = (n''l''\|\widetilde{w}^{(k)}\|n'l') = 1, \tag{15c}$$

$$(nl\|y^{(k)}\|n''l'') = (n''l''\|\widetilde{y}^{(k)}\|nl) = 1, \tag{15d}$$

$$(nl\|z^{(k)}\|n'l') = (n'l'\|\widetilde{z}^{(k)}\|nl) = 1. \tag{15e}$$

Then in terms of the quantities P and X defined in (7) and (8), the curtailed operators g and $\widetilde{g}$ representing the effective interactions of the configuration $l^n l'$, may be written as

$$g = \sum_k [X(k; ll', ll'') \sum_{i \neq j} (u_j^{(k)} \cdot w_i^{(k)})$$

$$+ X(k; ll', l''l) \sum_{i \neq j} (y_i^{(k)} \cdot \widetilde{z}_j^{(k)})], \tag{16}$$

and

$$\widetilde{g} = \sum_{k'} \left[X(k'; ll', ll'') \sum_{i \neq j} (u_j^{(k')} \cdot \widetilde{w}_i^{(k')}) + X(k'; ll', l''l) \sum_{i \neq j} (\widetilde{y}_i^{(k')} \cdot z_j^{(k')}) \right]. \tag{17}$$

Upon substituting (16) and (17) into (5) we obtain after considerable manipulation

$$W_2 = W_A + W_B + W_C, \tag{18}$$

where

$$W_A = -\sum_{kk't} P(kk'; ll', ll'') (2t+1) \left\{ \begin{matrix} k & k' & t \\ l' & l' & l'' \end{matrix} \right\} \cdot \sum_{ijs} [u_i^{(k')} \times (u_j^{(k)} \times v_s^{(t)})]^{(0)}, \tag{19}$$

$$W_B = -\sum_{kk't} P(kk'; ll', l''l) (2t+1) \left\{ \begin{matrix} k & k' & t \\ l' & l' & l \end{matrix} \right\} \left\{ \begin{matrix} k & k' & t \\ l & l & l'' \end{matrix} \right\} \cdot \sum_{i \neq j} (u_i^{(t)} \cdot v_j^{(t)}), \tag{20}$$

$$W_C = -\sum_{kk't} Q(kk'; ll', ll'') (2t+1) \left\{ \begin{matrix} k & k' & t \\ l' & l' & l \end{matrix} \right\} \cdot \left\{ \sum_{i \neq j \neq s} (-1)^{k+k'+t} [u_i^{(k)} \times \right.$$

$$\left([z_j^{(k')} \times \widetilde{z}_s^{(t)}]^{(k)} + \right.$$

$$\left. + [z_j^{(t)} \times \widetilde{z}_s^{(k')}]^{(k)}]^{(0)} + 2 \sum_{i \neq j} \left\{ \begin{matrix} k & k' & t \\ l' & l & l \end{matrix} \right\} (z_i^{(t)} \cdot \widetilde{z}_j^{(t)}) \right\}, \tag{21}$$

$$Q(kk'; ll', ll'') = X(k; ll', ll'') X(k'; ll', l''l)/\Delta E. \tag{22}$$

W_B represents a two-body interaction and for even values of t is a linear combination of the coefficients of the Slater parameters $F_t(ll')$. However, new parameters are obtained for odd values of t. The second term of W_c, also representing a two-body interaction, yields new parameters with t equal to $l + l' + 1$. Hence a Hamiltonian $H_{eff.}^2$, containing the two-body contribution to W_2 may be written

$$H_{eff.}^2 = -\sum_t F_t \sum_{i \neq j} (u_i^{(t)} \cdot v_j^{(t)}) - \sum_t G_t \sum_{i \neq j} (z_i^{(t)} \cdot \widetilde{z}_j^{(t)}). \tag{23}$$

The two terms are referred to as the direct and exchange effective Slater parameters. The matrix elements of $H_{eff.}^2$ are diagonal in S and L, and are independent of J and M.

W_A represents a three-body interaction if $i \neq j \neq s$. Otherwise, it contains mixtures of one and two-body terms which either vanish or have been previously considered. Hence taking this term and the first term of

W_c, the Hamiltonian representing the three-body effective electrostatic direct and exchange interactions, may be written

$$H^3_{eff} = - \sum_{k'kt} \cup (k'kt) \sum_{i \neq j \neq s} [u_i^{(k')} \times u_j^{(k)} \times v_s^{(t)}]^{(0)} - \sum_{k'kt} V(k'kt) \sum_{i \neq j \neq s} (-1)^{k'+k+t} \cdot$$

$$[u_i^{(k)} \times \{(z_j^{(k')} \times \widetilde{z}_s^{(t)})^{(k)} + (z_j^{(t)} \times \widetilde{z}_s^{(k')})^{(k)}\}]. \tag{24}$$

The coefficients of F_t in the configuration d_p^n are found by calculating the matrix elements

$$(d^n(\alpha_1 S_1 L_1) p\, SLJM | \sum_{i \neq j} (u_i^{(t)} \cdot v_j^{(t)}) | d^n (\alpha_1' S_1' L_1') p\, S'L'J'M').$$

Since this involves the reduced matrix element $(p \| v^{(t)} \| p)$, t may take only the values 0, 1 and 2 because of the triangular relationship $\delta(1,t,1)$ between the two p angular momenta and t. For t equal to 0 and 2 the Slater parameters F_0 and F_2 are included elsewhere (see sec. 3). Thus, the only new direct effective parameter for $d^n p$ as well as for $d^{n-1}sp$ and $d^{n-2}s^2 p$ is F_1.

Similarly for G_t, the calculation of the matrix elements

$$(d^n(\alpha_1 S_1 L_1) p\, SLJM | \sum_{i \neq j} (z_i^{(t)} \cdot \widetilde{z}_j^{(t)}) | d^n (\alpha_1' S_1' L_1') p\, S'L'J'M')$$

is required. Due to the reduced matrix element $(d \| z^{(t)} \| p)$ with the triangular relation $\delta(2,t,1)$, t may take on the values 1, 2 or 3. Since G_1 and G_3 are already included elsewhere, the only new exchange effective parameter is G_2.

The three-body effective interactions involve 9 independent parameters. They were not included in this investigation as their contributions are not expected to be significant (see sec. 4).

3. Parameters

The algebraic matrices of $(d+s)^n p$ comprise the electrostatic and spin orbit interaction matrices of the configurations $d^n p$, $d^{n-1}sp$ and $d^{n-2}s^2 p$; the electrostatic interactions between configurations $d^n p - d^{n-1}sp$, $d^{n-1}sp - d^{n-2}s^2 p$, $d^n p - d^{n-2}s^2 p$; the complete two and three-body effective interactions of the core d electrons, as well as two-body mixed effective interactions between the d and p electrons. The energy matrix (for a particular n) is then a linear combination of these matrices, the coefficients of which are parameters to be discussed below. Unprimed quantities denote the configuration $d^n p$, primes denote $d^{n-1}sp$ and double primes denote $d^{n-2}s^2 p$.

A, A', A''—the heights of the configurations,

$$S' = A' - A, \quad S'' = A'' - A.$$

B, B', B''—linear combinations of the Slater parameters $F^2(d,d)$ and $F^4(d,d)$:

$$B = \frac{1}{441} [9F^2(dd) - 5F^4(dd)] = F_2(dd) - 5F_4(dd) \,, \text{[31]}.$$

C, C', C''—multiples of the Slater parameter $F^4(d,d)$:

$$C = \frac{5}{63} F^4(dd) = 35F_4(dd) \,, \text{[31]}.$$

G_{ds}—the parameter of the $d - s$ interaction in the configuration $d^{n-1}sp$:

$$G_{ds} = \frac{1}{5} G^2(ds) \,, \text{[22]}.$$

39

F_2, F_2', F_2''—parameters of the direct part of the d-p interaction:

$$F_2 = \frac{1}{35} F^2(dp) , [31].$$

G_1, G_1', G_1''—parameters of the exchange part of the d-p interaction:

$$G_1 = \frac{1}{15} G^1(dp) , [31].$$

G_3, G_3', G_3''—parameters of the exchange part of the d-p interaction:

$$G_3 = \frac{3}{245} G^3(dp) , [31].$$

G_{ps}—the parameter of the $p-s$ interaction in the configuration $d^{n-1}sp$:

$$G_{ps} = \frac{1}{3} G^1(ps) , [32].$$

H—the parameter of the $d^n - d^{n-1}s$ interaction:

$$H = \frac{R^2(dd,ds)}{35} , [22].$$

H'—the parameter of the $d^{n-1}s - d^{n-2}s^2$ interaction, defined the same as H.

J—the parameter of the direct part of the $d^n p - d^{n-1}sp$ interaction:

$$J = \frac{R^2(dp,sp)}{5} , [33].$$

J'—the parameter of the direct part of the $d^{n-1}sp - d^{n-2}s^2p$ interaction, defined the same as J.

K—the parameter of the exchange part of the $d^n p - d^{n-1}sp$ interaction:

$$K = \frac{R^1(dp,ps)}{3} , [33].$$

K'—the parameter of the exchange part of the $d^{n-1}sp - d^{n-2}s^2p$ interaction, defined the same as K.

G^*—the parameter of the $d^n p - d^{n-2}s^2p$ interaction:

$$G^* = \frac{R^2(dd,ss)}{5} , [22].$$

The spin-orbit contribution to the Hamiltonian has the form (p. 120, TAS, [32],)

$$\sum_i \xi(r_i) (\ell_i \cdot s_i)$$

where r_i is the distance of the i th electron from the nucleus,

$$\xi(r_i) = \frac{1}{2\mu^2 c^2} \frac{1}{r_1} \frac{\partial U(r_i)}{\partial r_i} , \text{ where}$$

$U(r_i)$ is the potential in which the i th electron moves and μ is the reduced mass of the electron. Then,

$$\zeta_{n\ell} = \hbar^2 \int_0^\infty R^2(n\ell)\xi(r)\,dr,$$

where $\dfrac{R(n\ell)}{r}$ is the radial part of the wave function. Then

ζ_d, ζ_d', ζ_d''—parameters of the spin-orbit interaction of the d-electrons.

ζ_p, ζ_p', ζ_p''—parameters of the spin-orbit interaction of the p-electron.

The following parameters were discussed in detail in the previous section:

   α, α', α''—parameters of the $L(L+1)$ correction.
   β, β', β''—parameters of the Q correction.
   T, T', T'' parameters of the three-body effective
   T_x, T_x', T_x'' interactions of the core d-electrons.
   F_1, F_1', F_1''—parameters of the direct part of the effective interactions.
   G_2, G_2', G_2''—parameters of the exchange part of the effective interactions.

The calculated Lande g-factors given by

$$g = \frac{3}{2} + \frac{S(S+1) - L(L+1)}{2J(J+1)}, \quad [31]$$

were also computed. Whenever observed g-values exist, they provide very useful information to help fit the experimental levels.

Racah and Shadmi [1,2,9] have shown that if $M(d^n)$, $M'(d^{n-1}s)$ and $M''(d^{n-2}s^2)$ are the centers of gravity of d^n, $d^{n-1}s$ and $d^{n-2}s^2$, i.e., the weighted averages of the terms of these configurations, then the parameters

$$D' = M'(d^{n-1}s) - M(d^n)$$

and

$$D'' = M''(d^{n-2}s^2) - M'(d^{n-1}s)$$

are linear functions with small quadratic corrections of n for the configurations $(3d+4s)^n$ in the second spectra of the iron group.

Now in $d^n p$ we must consider n interactions $d-p$, $n(n-1)/2$ interactions $d-d$. From page 200, TAS, [32], the center of gravity of dp is

$$M(dp) = F_0(dp) - \left(G_1 + \frac{7}{2}\,G_3\right). \tag{25}$$

From equation (78) of Racah [31], the center of gravity of d^2 is

$$M(d^2) = A + \frac{7}{9}\,(C-2B). \tag{26}$$

Also, (p.197, TAS, [32],)

$$M(ds) = F_0(ds) - \frac{G_{ds}}{2}, \tag{27}$$

$$M(ps) = F_0(ps) - \frac{F_{ps}}{2}. \tag{28}$$

Since H, H', J, J', K, K' and G^* don't have diagonal elements, they need not be considered for the centers of gravity.

Now, by taking into account α, we have

$$M(d^2) = A + \frac{7}{9}(C-2B) + \frac{32}{3}\alpha. \tag{29}$$

As explained in the previous section, the $L(L+1)$ correction has to be considered in the form $[L(L+1) - 6n]\alpha$. Thus, write

$$M(d^2) = A + \frac{7}{9}(C-2B) + 12\alpha - \frac{4}{3}\alpha. \tag{30}$$

For the βQ correction we note that since Q is a two-body operator, and for d^2 we have a contribution of $1/9$, the net contribution to d^n will be $\frac{n(n-1)}{2}\left(\frac{1}{9}\right)$. Since T represents a three-body interaction, its contribution will involve a cubic in n. Since there is no interaction of s^2d^n with sd^{n+1} for $n = 0,9,10$, the polynomial must vanish for those n. For $n = 1$ the contribution is easily seen to be 70, and thus the contribution to d^n is $(35/36)\,n\,(9-n)\,(10-n)$. Directly from the algebraic matrices of the parameters T_x, T'_x and T''_x it was seen that these parameters do not contribute to the center of gravity.

Finally, the direct part of the effective interaction for $d^n p$ has no contribution to the center of gravity, whereas the exchange part gives nG_2.

Thus, we obtain

$$M(d^n p) = A + \frac{7n(n-1)}{18}(C-2B) - n\left(G_1 + \frac{7}{2}G_3\right) + \frac{2n}{3}(10-n)\alpha +$$
$$\frac{n(n-1)}{18}\beta + \frac{35}{36}n(9-n)(10-n)T + nG_2. \tag{31}$$

$$M'(d^{n-1}sp) = A' + \frac{7}{18}(n-1)(n-2)(C'-2B') - (n-1)\left(G'_1 + \frac{7}{2}G'_3\right) - \frac{(n-1)}{2}G_{ds} - \frac{1}{2}G_{ps} +$$
$$+ \frac{2}{3}(n-1)(11-n)\alpha' + \frac{(n-1)(n-2)}{18}\beta' + \frac{35}{36}(n-1)(10-n)(11-n)T' + (n-1)G'_2. \tag{32}$$

Since for $d^{n-2}s^2p$ the interactions $d-s$ and $s-p$ are constant and thus can be incorporated into the height of the configuration, we have

$$M''(d^{n-2}s^2p) = A'' + \frac{7}{18}(n-2)(n-3)(C''-2B'') - (n-2)\left(G''_1 + \frac{7}{2}G''_3\right) - \frac{2}{3}(n-2)(12-n)\alpha'' +$$
$$+ \frac{(n-2)(n-3)}{18}\beta'' + \frac{35}{36}(n-2)(11-n)(12-n)T'' + (n-2)G''_2. \tag{33}$$

The above expressions must be modified for complementary configurations ($n > 5$). The matrices of B, C, α, and β for the configurations $d^n p$, $d^{n-1}sp$, and $d^{n-2}s^2p$ are equal to the corresponding matrices of the complementary configurations $d^{10-n}s^2p$, $d^{11-n}sp$, and $d^{12-n}p$, respectively, [7]. Also, the matrices of G_{ds} and G_{ps} for $d^{n-1}sp$ are equal to the corresponding matrices of $d^{11-n}sp$. The matrices of T and G_2 must be replaced for complementary configurations by those of T_c and G_{2c}, respectively, [7]. Since the weighted average of the terms of d^9p is given by

$$M(d^9p) = M(d^9s^2p) = F_0 + \left(G_1 + \frac{7}{2}G_3\right), \quad [32, \text{p. } 200]$$

we obtain

$$M(d^{12-n}p) = A + \frac{7}{18}(n-2)(n-3)(C-2B) + (n-2)\left(G_1 + \frac{7}{2}G_3\right) + \frac{2}{3}(n-2)(12-n)\,\alpha +$$

$$+ \frac{(n-2)(n-3)}{18}\beta + \frac{35}{36}(12-n)(n-2)(n-3)\,T_c + (12-n)\,G_{2c}. \tag{34}$$

$$M'(d^{11-n}sp) = A' + \frac{7}{18}(n-1)(n-2)(C'-2B') + (n-1)\left(G_1' + \frac{7}{2}G_3'\right) - \frac{(n-1)}{2}G_{ds} -$$

$$- \frac{1}{2}G_{ps} + \frac{2}{3}(n-1)(11-n)\,\alpha' + \frac{(n-1)(n-2)}{18}\beta' + \frac{35}{36}(11-n)(n-1)(n-2)\,T_c' + (11-n)\,G_{2c}'. \tag{35}$$

$$M''(d^{10-2}s^2p) = A'' + \frac{7}{18}n(n-1)(C''-2B'') + n\left(G_1'' + \frac{7}{2}G_3''\right) + \frac{2}{3}n(10-n)\,\alpha'' +$$

$$+ \frac{n(n-1)}{18}\beta'' + \frac{35}{36}(10-n)\,n\,(n-1)\,T_c'' + (10-n)\,G_{2c}''. \tag{36}$$

The checks of the algebraic matrices were described previously by the author, [8,9]. In section 5, [9], we had

$$t + t_c + 160Q + 15L(L+1) - 80c = 240 - 40(n-2)(n-3). \tag{37}$$

In analogy to the above result we obtain

$$14t_x + 14t_{xc} - 1280Q - 15L(L+1) - 60b + 640c = 320n^2 - 1180n + k, \tag{38}$$

where

t, t_c, t_x, t_{xc}, b, and c are the algebraic matrices of the parameters T, T_c, T_x, T_{xc}, B, and C, respectively, and k is a numerical constant. The checks are possible because the sum of either t and t_c, or t_x and t_{xc}, gives rise only to two-body and one-body terms, that may be expressed as linear combinations of the algebraic matrices associated with the two and one-body parameters β, α, B, and C of the configuration d^n. The algebraic matrices of the parameters T, T', T'', T_c, T_c', T_c'', T_x, T_x', T_x'', T_{xc}, T_{xc}', and T_{xc}'' were added to the previously calculated matrices and the above checks utilized.

In the general treatment, with the exception of the centers of gravity $M(d^np)$, the parameters are *required* to satisfy the interpolation formula

$$P = P_0 + P_1(n-6) + P_2[(n-6)^2 - 10], \tag{39}$$

where P_0, P_1, and P_2 are the general parameters and replace the individual parameter P for all the spectra of the sequence; n is the number of $3d$ and $4s$ electrons for each spectrum. As an example, in table 42, B_0, B_0', B_0'' refer to the constant terms; B_1, B_1', B_1'' to the linear terms; B_2, B_2', B_2'' to effectively the quadratic terms, for the configurations d^np, $d^{n-1}sp$ and $d^{n-2}s^2p$, respectively, in the general treatment.

The above form of accounting for linearity with a quadratic correction, was chosen in order to make the parameters as nearly orthogonal as possible.

Since the parameter $M(d^np)$ incorporates setting the value of the lowest energy level to zero, it is not expected to behave regularly as a function of n. Consequently, when M, M', and M'' were replaced with

$$D' = M'(d^{n-1}sp) - M(d^np)$$

$$D'' = M''(d^{n-2}s^2p) - M'(d^{n-1}sp),$$

the latter were compelled to satisfy the relation (39).

4. Results

The source of the experimental data for all the elements with the exception of manganese is "Atomic Energy Levels", Vols. I and II by C.E. Moore, referred to as AEL, [35]. For manganese, the experimental results of Catalan, Meggers and Garcia-Riquelme, [36] were used. Details indicating the levels included, as well as those that were rejected, are provided in another paper, [17]. On fitting 1537 levels using 67 free interaction parameters a mean error of 182cm^{-1} was obtained. Altogether 3652 energy levels were calculated including all the levels for the configurations $3d^{n-2}4s^24p$ across the sequence.

Tables 1-41^2 yield the results for each of the parameters individually, whereas tables 42 and 43 give the results of the general interpolation formulas, together with the centers of gravity $M(d^np)$, $(0 \leqslant n \leqslant 10)$, $M_{11}(d^9s^2p)$ and $M_{12}(d^{10}s^2p)$, for the various significant stages of this project.

The error attributed to a parameter defines the range within which it may vary without the mean error increasing. In particular, if zero lies within the range of the value of a parameter and $\pm$ its error, then that particular parameter may be fixed at zero, which is equivalent to excluding it in the first place.

In this project, Hartree-Fock calculations were performed for all the configurations $3d^n4p$, $3d^{n-1}4s4p$ and $3d^{n-2}4s^24p$. Details of these calculations and results are given in another paper, [38].

With the exception of the results obtained from the Hartree-Fock calculations, the values of particular parameters pertaining to the three configurations $3d^n4p$, $3d^{n-1}4s4p$ and $3d^{n-2}4s^24p$ are in arithmetic progression, since there is insufficient experimental data for $3d^{n-2}4s^24p$ to determine parameters pertaining only to those configurations. Hence even in results arising from the least-squares computations these parameters do not have attached a $\pm$ error value. Furthermore, also in those cases where a parameter has a value in the least-squares either equal to the one directly preceding it (denoted, specifically, for the sake of clarity by $EQ.$), or if it is maintained at a fixed difference (unless specified otherwise, the differences that appears in the diagonalization, and denoted by F.D.), there is no $\pm$ error attached to the value of the parameter. In those cases where a parameter either changes sign or is ill-defined in the least-squares, it is fixed at either its value in the diagonalization, or at a value of zero (denoted by FIX). This procedure may also be followed if one is interested to study the effect of a particular parameter either on other parameters or in reducing the mean error, by considering two variations, one with the parameter left intact and in the other maintained at a value of zero.

The column PREVIOUS RESULTS refers to individual treatments without two- and three-body effective interactions, performed by the author for calcium-nickel, [10-16]. For the case of copper, the excellent results of Martin and Sugar, [38], for the configurations $3d^94s4p + 3d^84s^24p$ of CuI, were considered.

Results for the following parameters were obtained by the Hartree-Fock calculations: B,B',B'',C,C',C'', $F_2,F_2',F_2'',G_1,G_1',G_1'',G_3,G_3',G_3'',G_{ps},G_{ds},G^*,H,H',J,J',K,K',\zeta_d,\zeta_d',\zeta_d''$, ζ_p,ζ_p', and ζ_p'' as indicated in Tables 3-32. Very interesting and important conclusions from these results, not known or anticipated previously, were that G^* is significantly larger than G_{ds}, whereas H' is appreciably smaller than H. The Hartree-Fock results for J' and K' in ZnI $3d^{10}4s4p + 3d^94s^24p$ were anomalous to the other values, and thus excluded from consideration.

All entries of the parameters are in cm^{-1}.

4.1 Initial values

These entries appear in DIAGON 1 in Tables 1-40, and in the column INITIAL in table 42 for the general interpolation parameters.

For the parameters D' and D'', least-squares optimization calculations were performed on those values obtained in the columns PREVIOUS RESULTS. Even when allowing for quadratic corrections, the deviations were quite large, especially near the center of the period, where the elements have the largest number of levels. Consequently, least-squares were performed, *weighing* the entry of each element proportionally to the number of its levels. When $(D')_2$ and $(D'')_2$, the quadratic corrections were allowed to be free, their values were very different with both assuming large errors. When they were restricted to be equal the common value was -87 ± 95. Consequently, initially D' and D'' were taken to be linear functions of the atomic

---

^{2} All tables and figures are placed at the end of this paper, beginning on page 54.

number, as indicated in tables 1, 2, 42, and 43. Entries are given to two decimal places to ascertain that the interpolation relation, (39), for the parameters be satisfied.

From previous results obtained for the even configurations in the iron group by Racah, Shadmi, Oreg, Stein and Caspi, [1,2,28,29], and for the odd configurations by the author, [3,4], the parameters B and C behaved as linear functions of the atomic number.

For the parameters B, B', and B'' using the values obtained from the Hartree-Fock calculations, and demanding that they be in arithmetic progression, the best straight lines yielded

$$B = 976 + 84(n-6) \tag{40a}$$

$$B' = 1083 + 77(n-6) \tag{40b}$$

$$B'' = 1190 + 70(n-6). \tag{40c}$$

The Hartree-Fock results were on the average greater by a factor of 1.27 from the spectroscopic results. Hence the initial values given in tables 3-5, and 42 are the above values divided by 1.27.

Exactly as for the B,s, we obtain, using the Hartree-Fock results

$$C = 3558 + 310(n-6) \tag{41a}$$

$$C' = 3980 + 280(n-6) \tag{41b}$$

$$C'' = 4402 + 250(n-6). \tag{41c}$$

Again, the initial values for these parameters given in Tables 6-8, and 42 are the above values divided by the same factor of 1.27.

For F_2, F_2', F_2'', least-squares were performed on the Hartree-Fock results demanding that they be in arithmetic progression. When quadratic corrections were allowed, the mean error was 5.0cm^{-1}, with values of 1.2 $\pm$ 0.3cm^{-1} and 0.7$\pm$0.2cm^{-1} for $(F_2)_2$ and $(F_2')_2$, respectively. When $(F_2)_2$ and $(F_2')_2$ were set equal, they had a common value of 1.0 $\pm$ 0.3cm^{-1}, with the mean error rising to 7.9cm^{-1}. When both were zero, the mean error only rose to 9.0cm^{-1}. Hence this variation was adopted, the parameters were restricted to be linear with initial values given in tables 9-11 and 42. However, the possibility of allowing for quadratic corrections in subsequent iterations and variations was permitted.

For the parameters G_1, G_1', and G_1'', when least-squares were performed on the Hartree-Fock results, the quadratic corrections were significant. When the parameters were forced to be linear, the mean error was 68 cm^{-1}, whereas with $(G_1)_2$ assuming a value of 7.8$\pm$1.0cm^{-1} and $(G_1')_2$ a value of 3.5$\pm$0.6cm^{-1} the mean error was reduced to 21 cm^{-1}. The initial values are given in tables 12-14 and 42. As for the F_2 parameters no scaling factor was used between the spectroscopic and the Hartree-Fock results.

From the definition of the G_3 parameters, it is clear that they should be allowed the same degree of freedom as the G_1 parameters. Hence the initial values given in tables 15-17 and 42, were obtained by allowing quadratic corrections in the least squares of the Hartree-Fock results, and then using a scale factor of 2 to divide the latter. This common factor was obtained by comparing the Hartree-Fock and spectroscopic results.

For the parameter G_p, the quadratic correction was very significant, reducing the mean error in the least-squares of the Hartree-Fock results from 79cm^{-1} to 27cm^{-1}. The values given in tables 18 and 42 are those obtained after dividing the least squares results by a common factor of 1.58.

For G_{ds} and G^*, least-squares calculations on the Hartree-Fock results decreased the value of the mean error from 91cm^{-1} to 38cm^{-1}, after allowing a quadratic correction. However, the values of $(G_{ds})_2$ and $(G^*)_2$ were sufficiently close to each other, 7.3 $\pm$ 1.5cm^{-1} and 8.9 $\pm$ 1.6cm^{-1}, respectively, that these were compelled to be equal in the initial computations. The latter were obtained by dividing the results from the least-squares by a common factor of 1.1.

For H and H', in the least-squares of the Hartee-Fock results, the values of H_1 and H_1' were $-11.5 \pm$ 1.6cm^{-1} and $-13.5 \pm$ 1.2cm^{-1}, respectively, whereas H_2 and H_2' were 2.9 $\pm$ 0.2cm^{-1} and 2.5 $\pm$ 0.2cm^{-1}, respectively. Consequently, both pairs of parameters were set equal. Then the mean error was only 5.8cm^{-1} and the results, without a scaling factor, are given in tables 21, 22, and 42.

45

Unlike the case of the H parameters, in the least-squares calculations of the Hartree-Fock results, J_1 and J_1' had to be different, as their values were $-28 \pm 8\text{cm}^{-1}$ and $14 \pm 7\text{cm}^{-1}$, respectively. However, J_2 and J_2' were sufficiently close to have a common value of $11 \pm 2\text{cm}^{-1}$. This quadratic correction was important in reducing the mean error from 98cm^{-1} to $41\ \text{cm}^{-1}$. The initial values in tables 23,24, and 42 are the least-square values divided by a common factor of 1.67 for both J and J'.

For K and K' the situation was very similar as that for J and J'—linearity with a common quadratic correction—here $24.5 \pm \text{cm}^{-1}$, yielded a mean error of 62cm^{-1} compared with 142cm^{-1}, when the parameters were forced to be linear. The initial values for K were obtained by using a scale factor of 1.28, whereas for K' the scale factor was 1.10. Hence, initally K_2 had a value of 19.15cm^{-1} and K_2', a value of 22.29cm^{-1} in the tables 25, 26 and 42.

In previous general treatments dealing with the spectra of the iron group [1-4,27,28], the parameters of the spin-orbit interaction generally required a quadratic correction. After performing several variations in the least-squares of the Hartree-Fock results, it was seen that the parameters $\zeta_d, \zeta_d', \zeta_d''$ were required to have a quadratic correction, whereas the parameters $\zeta_p, \zeta_p',$ and ζ_p'' could, initially, be linear.

The values of $(\zeta_d)_2$ and $(\zeta_d')_2$ in the least-squares of the Hartree-Fock results were $6.9 \pm 0.2\ \text{cm}^{-1}$ and $8.1 \pm 0.2\text{cm}^{-1}$, respectively. Hence they were set equal.

The mean error was then only 8.7cm^{-1}, reduced from 32cm^{-1} when the ζ_d parameters were linear. The values obtained from those least-squares calculations were used directly without a scaling factor and are given in Tables 60-62, and 75.

When least squares were performed on the Hartree-Fock results of $\zeta_p, \zeta_p',$ and ζ_p'', demanding that they be in arithmetic progression, the value of $(\zeta_p)_2$ was $-0.2 \pm 0.3\text{cm}^{-1}$ and that of (ζ_p') was $0.3 \pm 0.2\text{cm}^{-1}$. Thus, these parameters were fixed at zero, and initially $\zeta_p, \zeta_p',$ and ζ_p'' were linear. Then, the mean error only increased from 7.9cm^{-1} to 8.2cm^{-1}. By comparing the Hartree-Fock and spectroscopic results, a common scale factor of 0.8 was used to divide the values of the least-squares, in order to obtain the initial values given in tables 30-32 and 42.

From the parameters of the effective interactions, previous values were available only for the α's. As this parameter behaved irregularly, a common average value of 60 was taken as the initial value of α, α', and α'' for all elements in the period. The initial values of β, β', β'', T, T', T'', T_x, T_x', T_x'', F_1, F_1', F_1'', G_2, G_2', and G_2'' were zero for all elements.

4.2 Preliminary iterations

For the general least-squares routine a maximum capacity for 101 parameters could be achieved.

In the first two iterations the parameters D', D'', G_1, G_1', G_3, G_3', G_{ps}, K, K', ζ_d, ζ_d', ζ_p, and ζ_p' were allowed independent quadratic variations as functions of the atomic number.

The parameters F_2, F_2', G_{ds}, G^*, H, H', J, and J' were permitted a quadratic variation with the restrictions

$$(F_2)_2 = (F_2')_2 \,,$$
$$(G_{ds})_2 = G_2^* \,,$$
$$H_2 = H_2' \,,$$
$$J_2 = J_2' \,.$$

The parameters $B, B', C, C', \alpha, \alpha', F_1, F_1', G_2$, and G_2' were allowed to vary linearly.

The parameters $\beta, \beta', T, T', T_x$, and T_x' varied linearly with the restrictions

$$\beta_1 = \beta_1'.$$
$$T_1 = T_1',$$
$$(T_x)_1 = (T_x')_1.$$

The remaining 13 places were reserved for the centers of gravity $M_0 - M_{12}$.

In the third iteration the restrictions

$$(G_{ds})_2 = G_2^*,$$
$$H_2 = H_2',$$
$$J_2 = J_2',$$

were not included. Thus, *all* the parameters for the interactions between configurations were permitted independent quadratic variations. The required places were obtained by realizing from the first two iterations that F_2 need vary only linearly and it was also possible to impose the restrictions

$$(G_1)_2 = (G_1')_2 \text{ and}$$
$$(G_3)_2 = (G_3')_2.$$

The initial parameters of the first variation are given in the column DIAGON 1, (tables 1-40), whereas the final values are given in the column GLS1, (table 42). These parameters, with some modifications, were used for the second iteration and given in the column DIAGON 2. The results of the most successful of many variations in the general least-squares are given in the column GLS2, which are essentially the ones used in the next iteration and given in the column DIAGON 3. Again the results of the best variation are given in the columns GLS3, and then used with some modifications for the next iteration.

The values of the least-squares where all 101 parameters are allowed to be free are given in the columns GLS 1a, GLS 2a, and GLS 3a.

From the results of GLS1a it is clear that we should have

$$D_2'' = D_2'$$
$$(F_2)_2 = 0 \,(\text{FIX})$$
$$(G_1')_2 = (G_1)_2 = 0 \,(\text{FIX})$$
$$(G_3')_2 = (G_3)_2$$
$$G_0^* - (G_{ds})_0 = 708 \,(\text{F.D.})$$
$$G_1^* = (G_{ds})_1$$
$$H_0 - H_0'' = 34 \,(\text{F.D.})$$
$$K_2' = K_2$$
$$(\zeta_d')_2 = (\zeta_d)_2$$
$$(\zeta_p')_1 = (\zeta_p)_1$$
$$(\zeta_p')_2 = (\zeta_p)_2 = 0 \,(\text{FIX})$$
$$\alpha_0' = \alpha_0$$
$$\alpha_1' = \alpha_1$$
$$\beta_0' = \beta_0$$
$$T_0' = T_0$$
$$(T_x')_0 = (T_x)_0$$
$$(F_1')_0 = (F_1)_0$$
$$(F_1')_1 = (F_1)_1 = 0 \,(\text{FIX})$$
$$(G_2')_0 = (G_2)_0$$
$$(G_2')_1 = (G_2)_1,$$

which can be seen either directly from the results or by noting that otherwise the parameters are ill-defined. With the above restrictions there then remained 78 free parameters in that variation which was used for the next stage of the computations, GLS 1. In GLS 1, the sum of the squares of the deviations decreased from 2.83879×10^9 to 5.97895×10^7. Besides the changes in the M's, the parameter that had the greatest effect was C_1; while changing from $244\,\text{cm}^{-1}$ to $408 \pm 8\,\text{cm}^{-1}$ it reduced the sum of the squares of the deviations from 3.29325×10^8 to 2.57289×10^8.

In a variation where G_0^*, J_0', K_0' were free, but in addition to the restriction of GLS 1, there were the conditions

$$G_1^* = (G_{ds})_1$$
$$H_1' = H_1$$
$$J_1' = J_1$$
$$K_1' = K_1$$

the parameters β_0, T_0, $(G_1')_2$, $(G_{ps})_2$, and H_2 were badly defined. In each case the $\pm$ error of the parameter exceeded its actual value.

In another variation, with the same restrictions as in GLS 1, but leaving β_1, $(F_1)_1$ and $(G_2')_0$ free, the values of these parameters were

$$\beta_1 = -29 \pm 37$$
$$(F_1)_1 = 0.8 \pm 2.2$$
$$(G_2')_0 = 36 \pm 7.$$

As $(G_2)_0$ had the value of 24 ± 8 it is clear that β_1 and $(F_1)_1$ should be zero whereas

$$(G_2)_0 = (G_2')_0.$$

In the second iteration when all 101 parameters are left free, GLS 2a, the G_3, F_1, and G_2 parameters behave erratically, whereas H_0', J_1', K_2', and $(\zeta_d'')_2$ change sign. From these results and many other variations it became apparent that for the final variation in this iteration we should have

$$D_2'' = D_2'$$
$$(F_2)_2 = 0\,(\text{FIX})$$
$$(G_1')_2 = (G_1)_2 = 0\,(\text{FIX})$$
$$(G_3')_0 = (G_3)_0$$
$$(G_3')_1 = (G_3)_1$$
$$(G_3')_2 = (G_3)_2 = 0\,(\text{FIX})$$
$$G_1^* = (G_{ds})_1$$
$$H_0 - H_0' = 40\,(\text{F.D.})$$
$$H_1' = H_1$$
$$J_1' = J_1$$
$$K_1' = K_1$$
$$K_2' = K_2$$
$$(\zeta_d')_2 = (\zeta_d)_2$$
$$(\zeta_p')_2 = (\zeta_p)_2$$
$$\alpha_0' = \alpha_0$$
$$\alpha_1' = \alpha_1$$
$$\beta_0' = \beta_0$$
$$T_0' = T_0$$
$$(T_x')_0 = (T_x)_0$$
$$(F_1')_0 = (F_1)_0$$
$$(F_1')_1 = (F_1)_1$$
$$(G_2')_1 = (G_2)_1$$

with 77 free parameters remaining.

The sum of the squares of the deviations is decreased from 2.06466×10^8 to 5.71710×10^7. The parameters having the greatest effect were $(G_3)_0$, in changing from 27 cm^{-1} to 10 ± 1 cm^{-1} it reduced the sum of the deviations from 1.47723×10^8 to 1.14285×10^8; $(G_3)_1$ from -7 cm^{-1} to 2.2 ± 0.7 cm^{-1} and reduced $\Sigma(\Delta_i)^2$ from 1.14285×10^8 to 8.71367×10^7, and $(G_{ps})_2$, from 0 to -33 ± 5 cm^{-1}, and reduced $\sum_i (\Delta_i)^2$ from 5.97160×10^7 to 5.75448×10^7.

48

In the variation with parameters given in the column GLS 2b, the values of G_1^*, H_1', J_1', K_1', T_0', $(G_1)_2$, and $(G_3)_2$ are allowed to be free, whereas

$$G_0^* - (G_{ds})_0 = 708\,(\text{F.D.})$$
$$J_0' - J_0 = 280\,(\text{F.D.})$$
$$K_0' - K_0 = 1020\,(\text{F.D.}).$$

The results clearly show that G_1^* can't vary independently. H_1' and H_1 are too far apart with H_1' not well-defined. J_1 and J_1', as well as K_1 and K_1', have opposite signs, which is unacceptable. T_0' is well-defined but need not be different from T_0. Allowing $(G_1)_2$ and $(G_3)_2$ to vary has no significant effect. The reduction of the mean error from 203.1 cm^{-1} to 200.3 cm^{-1}, certainly does not justify allowing the increase in the number of parameters as then they do not behave reasonably.

In a variation identical to GLS 2, except that $D_2' \neq D_2''$, the values of these two parameters were -120 ± 5 cm^{-1} and -127 ± 11 cm^{-1}, respectively, showing clearly that they should be equal.

A variation in which G_1^*, H_1', J_1', K_1' are free, and all other conditions the same as in GLS 2, indicated that none of these parameters should be free, as they were either of different sign than the corresponding values G_1, H_1, J_1, and K_1, or were ill-defined.

In GLS 3a, with all 101 parameters free, the sum of the squares of the deviations was reduced from 9.62040×10^7 to 5.79187×10^7. However, several parameters such as G_0^*, G_1^*, G_2^*, H_0', J_0', J_1' and J_2' were badly defined.

From the results of GLS 3a, as well as many other variations wherein only a smal number of restrictions are imposed in each case, the following restrictions were imposed in the final variation of the third iteration, GLS 3, yielding 68 free parameters:

$$D_2'' = D_2'$$
$$(F_2)_1 = (F_2)_1$$
$$(G_1')_1 = (G_1)_1$$
$$(G_1)_2 = 0\,(\text{FIX})$$
$$(G_3')_1 = (G_3)_1 = 0\,(\text{FIX})$$
$$(G_3)_2 = 0\,(\text{FIX})$$
$$G_1^* = (G_{ds})_1 = -10\,(\text{FIX})$$
$$G_2^* = (G_{ds})_2$$
$$H_0' = 84\,(\text{FIX})$$
$$H_1' = H_1$$
$$H_2' = H_2 = 3\,(\text{FIX})$$
$$J_1' = J_1$$
$$J_2' = J_2$$
$$K_1' = K_1$$
$$K_2' = K_2$$
$$(\zeta_p')_1 = (\zeta_p)_1$$
$$(\zeta_p')_2 = (\zeta_p)_2$$
$$(\zeta_d')_2 = (\zeta_d)_2$$
$$\alpha_0' = \alpha_0$$
$$\alpha_1' = \alpha_1$$
$$\beta_0' = \beta_0$$
$$T_0' = T_0$$
$$(T_x')_0 = (T_x)_0$$
$$(T_x)_1 = -0.3\,(\text{FIX})$$
$$(F_1')_0 = (F_1)_0$$
$$(F_1')_1 = (F_1)_1 = 2\,(\text{FIX})$$
$$(G_2')_0 = (G_2)_0$$
$$(G_2')_1 = (G_2)_1 = 0\,(\text{FIX})$$

49

The sum of the squares of the deviations was reduced from 8.28089×10^7 to 5.12809×10^7. The only parameter that had a very significant effect was J_0, which changed from 1600 cm^{-1} to 1125 ± 39 cm^{-1}, thereby reducing the sum of the squares of the deviations from 7.63469×10^7 to 6.39959×10^7. Although all the parameters had reasonable well-defined values, the change in the sum of the squares of the deviations was such that a further iteration was required.

In GLS 3b, there are 76 free parameters. The same conditions apply as for GLS 3a, with the exception of $(F_2')_1$, $(\zeta_p')_1$, $(T_z)_1$, $(F_1)_1$, $(G_2)_1$ being free; $G_1^* = (G_{ds})_1$, but not fixed at -10 cm^{-1}; $(G_3')_1 = (G_3)_1$, but not fixed at 0; $H_2' = H_2$, but not fixed at 3 cm^{-1}.

Although the parameters are very reasonable, the mean error actually increased from 196.7 cm^{-1} to 198.1 cm^{-1}, indicating that the improvement caused by these parameters was more than counterbalanced by the fact that their number increased.

In GLS 3c there are 72 free parameters. The same conditions apply as for GLS 3a, with the exception of H_0', K_1', and $(F_1)_1$ being free; $G_1^* = (G_{ds})$, but not forced at -10 cm^{-1}. Although the mean error decreases to 194.5 cm^{-1}, the parameters K_1' and $(F_1)_1$ change sign, while the value of 51 ± 12 cm^{-1} is definitely too small for H_0'.

In a variation where J_1' was allowed to vary freely, although its value changed from 50 cm^{-1} to the unreasonable value of -304 ± 44 cm^{-1}, the sum of the squares of the deviations only decreased as a consequence from 6.74632×10^7 to 6.67194×10^7. Furthermore, such a value for J_1' would cause J' to become negative for Cu I. Clearly, then J_1' should equal J_1 as in GLS 3.

In a variation where D_2' was not equal D_2'' and $(F_1)_0$ not equal to $(F_1')_0$, the resulting values were

$$D_2' = -117 \pm 5 \text{ cm}^{-1}$$
$$D_2'' = -109 \pm 11 \text{ cm}^{-1}$$
$$(F_1)_0 = -13 \pm 6 \text{ cm}^{-1}$$
$$(F_1')_0 = -9 \pm 6 \text{ cm}^{-1},$$

and thus as in GLS 3 these parameters should be respectively equal to each other.

The values of the parameters of the individual least squares of the third iteration are given in tables 3–40.

4.3. Final iteration.

For this iteration the 101 places for the parameters were the same as for the previous iteration.

From the final results of that iteration, GLS 3, as well as several different variations in this iteration, it became evident that 67 parameters should be free in the final results, GLS 4. Thus, the following 34 restrictions were imposed:

$$D_2'' = D_2'$$
$$(F_2')_1 = (F_2)_1$$
$$(G_1')_1 = (G_1)_1$$
$$(G_1)_2 = 0 \text{ (FIX)}$$
$$(G_3')_1 = (G_3)_1 = 0 \text{ (FIX)}$$
$$(G_3)_2 = 0 \text{ (FIX)}$$
$$G_0^* - (G_{ds})_0 = 231 \text{ (F.D.)}$$
$$G_1^* = (G_{ds})_1 = 0 \text{ (FIX)}$$
$$G_2^* = (G_{ds})_2$$
$$H_0' = 84 \text{ (FIX)}$$
$$H_1' = H_1$$
$$H_2' = H_2 = 0 \text{ (FIX)}$$
$$J_1' = J_1$$
$$J_2' = J_2$$
$$K_1' = K_1$$
$$K_2' = K_2$$
$$(\zeta_p')_1 = (\zeta_p)_1$$

50

$$(\xi_p')_2 = (\xi_p)_2$$
$$(\zeta_d')_2 = (\zeta_d)_2$$
$$\alpha_0' = \alpha_0$$
$$\alpha_1' = \alpha_1$$
$$\beta_0' = \beta_0$$
$$T_0' = T_0$$
$$(T_x')_0 = (T_x)_0$$
$$(T_x)_1 = 0\,(\text{FIX})$$
$$(F_1')_0 = (F_1)_0$$
$$(F_1')_1 = (F_1)_1 = 0\,(\text{FIX})$$
$$(G_2')_0 = (G_2)_0$$
$$(G_2')_1 = (G_2)_1 = 0\,(\text{FIX}).$$

Hence in the final results the parameters D', D'', G_{ps}, G_{ds}, G^*, J, J', K, K', ζ_d, ζ_d', ζ_d'', ζ_p, ζ_p' and ζ_p'' required quadratic corrections, with the restrictions noted above. In particular, G_{ds} and G^* did not have linear terms. All the other parameters were linear functions of the atomic number. Furthermore, T_x, T_x', T_x'', F_1, F_1', F_1'', G_2, G_2', and G_2'' were constant for all spectra across the entire period.

The sum of the squares of the deviations only decreased from 5.83692×10^7 to 4.87994×10^7 for all the parameters. In *each* case the change in a parameter from the value in the diagonalization to its value in the least squares was less than the statistical error associated with the parameter. All the parameters have very reasonable values and are well-defined. The comparisons between these final results and those obtained from the Hartree-Fock calculations as well as those of previous results, are indicated graphically at the end of this section.

In GLS4a, where all the 101 parameters are free, although the mean error is only $171.3\,\mathrm{cm}^{-1}$, the parameters are not nearly as well defined as those of GLS4. Furthermore, the values of J_0' at $598 \pm 271\,\mathrm{cm}^{-1}$, J_1' at $-376 \pm 127\,\mathrm{cm}^{-1}$, $(F_1')_1$ at $-11 \pm 4\,\mathrm{cm}^{-1}$, $(F_1'')_1$ at $-20\,\mathrm{cm}^{-1}$ (causing both F' and F'' to assume positive and negative values for different elements of the period), are not reasonable.

In GLS4b, the parameter G_0^* is allowed to be free and changes from $1820\,\mathrm{cm}^{-1}$ to a value of $1649 \pm 83\,\mathrm{cm}^{-1}$, thereby decreasing the sum of the squares of the deviations only from 5.40826×10^7 to 5.38969×10^7. Hence G^* is then uniformly higher than G_{ds} by only $56\,\mathrm{cm}^{-1}$. Thus, the significant difference between G^* and G_{ds} obtained in the Hartree-Fock calculations and the initial iterations becomes considerably less pronounced here. Nevertheless, in view of the other considerations above, G_0^* was maintained in the final variation at $1826\,\mathrm{cm}^{-1}$, a fixed difference of $231\,\mathrm{cm}^{-1}$ above the value of $(G_{ds})_0$.

In GLS4c, there are 71 free parameters, the same restrictions are imposed as in GLS4, with the exception that G_0^*, H_0', J_1', and $(F_1)_1$ are left free. Although the mean error is reduced from $182.2\,\mathrm{cm}^{-1}$ to $176.1\,\mathrm{cm}^{-1}$ and the values for G_0^* of $1674 \pm 81\,\mathrm{cm}^{-1}$ and $(F_1)_1$ of $-4 \pm 2\,\mathrm{cm}^{-1}$ are reasonable, the value for H_0' of 42 ± 15 cm^{-1} is much too low, whereas the value for J_1' of $-231 \pm 37\,\mathrm{cm}^{-1}$ is definitely unacceptable.

In another variation, with the same restrictions as in GLS4, but $(F_1)_1$ free, the value of this parameter was $-5 \pm 2\,\mathrm{cm}^{-1}$. Coupled with a value of $(F_1)_0$ of $-11 \pm 4\,\mathrm{cm}^{-1}$, the parameter F_1 assumes negative and positive values for different spectra of the period, which is unacceptable.

In GLS4d, the following additional 26 levels, which although having deviations exceeding the maximum tolerated for the other iterations ($\sim 600\,\mathrm{cm}^{-1}$), had deviations below $1000\,\mathrm{cm}^{-1}$ when inserted into the least squares.

Ti

1. The level $3d^2 4s(b^2P)4py^3S$ at $35439.43\,\mathrm{cm}^{-1}$.
2. The three levels $3d^3(a^2G)4pt^3F_{2,3,4}$ at $38451.29\,\mathrm{cm}^{-1}$, $38544.38\,\mathrm{cm}^{-1}$ and $38670.73\,\mathrm{cm}^{-1}$.
3. The level $3d^3(a^2G)4px^1G$ at $38959.53\,\mathrm{cm}^{-1}$.
4. The level $3d^3(a^2P)4pw^1D$ at $39265.80\,\mathrm{cm}^{-1}$.
5. The three levels $3d^3(a^2P)4pr^3D$ at $40556.70\,\mathrm{cm}^{-1}$, $40670.60\,\mathrm{cm}^{-1}$, and $40844.19\,\mathrm{cm}^{-1}$.

V

1. The level 1° at $34019.12\,\mathrm{cm}^{-1}$.
2. The four levels of $3d^4(a^3G)4pt^4F$ at $41389.49\,\mathrm{cm}^{-1}$, $41428.93\,\mathrm{cm}^{-1}$, $41492.29\,\mathrm{cm}^{-1}$, and $41599.36\,\mathrm{cm}^{-1}$.

Cr

1. The level $3d^44s(a^2P)4px^3S$ at 60084.09cm^{-1}.

Mn

1. The four levels $3d^6(^3G)4pw^4H$ at 63395.45cm^{-1}, 63444.61cm^{-1}, 63457.85cm^{-1}, and 63363.54cm^{-1}.

Fe

1. The five levels $3d^44s(a^4D)4px^5D$ at 39625.829cm^{-1}, 39969.880cm^{-1}, 40231.365cm^{-1}, 40404.544cm^{-1}, and 40491.32cm^{-1}.

Cu

1. The two levels $3d^94s(^1D)4p''\ ^2P$ at 56343.74cm^{-1} and 58364.73cm^{-1}.

The restrictions are the same as in the case of GLS4, and as expected, the parameters are well-defined and reasonable. However, the mean error increases substantially from 182.2cm^{-1} to 199.7cm^{-1}.

In GLS4e there are 65 parameters. In addition to the restrictions of GLS4, the conditions

$$(F_1)_0 = (F_1')_0 = (F_1'')_0 = 0(\text{FIX})$$

$$(G_2)_0 = (G_2')_0 = (G_2'')_0 = 0(\text{FIX})$$

are imposed. As a consequence, the parameters of the two-body effective interactions $d-p$: F_1, F_1', F_1'', G_2, G_2', and G_2'' are eliminated. Surprisingly, the mean error only increases from 182.2cm^{-1} to 183.5cm^{-1}. The elimination of these parameters also affects very slightly the values of the other parameters, and thus their overall effects are minimal indeed. Consequently, the higher order three-body effective interactions $d-p$ were not included (see the end of sec. 2).

In the variation GLS4f, there are 64 free parameters. In addition to the elimination of the F_1 and G_2 parameters, as in GLS4e, the condition

$$(T_x)_0 = (T_x')_0 = (T_x'')_0 = 0(\text{FIX})$$

was imposed. Hence the three-body effective interaction parameters of the d-electrons, T_x, T_x', and T_x'' were eliminated. The mean error rose to 189.1cm^{-1}, indicating that the T_x parameters have a considerably greater effect than the F_2 and G_1 parameters combined. The significant increase in the value of

$$\beta_0 = \beta_0' = \beta_0''$$

from -527 ± 28cm^{-1} in GLS4 to the present value of -673 ± 31cm^{-1}, should be noted.

In GLS4g, there are 62 parameters. In addition to the elimination of the F_1, G_2, and T_x parameters, as in GLSf, the conditions

$$T_0 = T_0' = T_0'' = 0(\text{FIX})$$

$$T_1 = T_1' = T_1'' = 0(\text{FIX})$$

are imposed. Hence both parameters of the effective interactions of the d-electrons are eliminated. As expected, the effects here are very pronounced, the mean error rising from 189.1cm^{-1} to 203.7cm^{-1}. Thus the effect of T is twice as great as T_x, and 11 times as large as F_2 and G_1 combined! As expected, the elimination of T has profound effects on the other remaining effective intersection parameters:

$$\alpha_0 = \alpha_1' = \alpha_0''$$

increases from 30 ± 3cm^{-1} to 61 ± 2cm^{-1}, whereas

$$\beta_0 = \beta_0' = \beta_0''$$

changes from $-673 \pm 31 \mathrm{cm}^{-1}$ to $-336 \pm 32 \mathrm{cm}^{-1}$.

In GLS 4h there are also 62 free parameters. The F_2 and G_1 parameters are included, but instead the restrictions

$$\beta_0 = \beta_0' = \beta_0'' = 0(\mathrm{FIX})$$

$$\beta_1 = \beta_1' \quad \beta_1'' = 0(\mathrm{FIX})$$

are imposed.

Hence, β, β', and β'' are eliminated. The results again deteriorate, the mean error rising from $203.7 \mathrm{cm}^{-1}$ to $210.8 \mathrm{cm}^{-1}$. As expected, the F_1 and G_2 parameters have similar values as in GLS4 and GLS4d.

$$\alpha_0 = \alpha_0' = \alpha_0''$$

has the higher value of $71 \pm 1 \mathrm{cm}^{-1}$, incorporating partially the other effective parameters.

In GLS4i, there are only 60 free parameters. Here, in addition to the restrictions of GLS4, the parameters β, β', β'', T, T', T'', T_x, T_x', T_x'', F_1, F_1', F_1'', G_2, G_2' and G_2'' are eliminated. The mean error rises to $211.4 \mathrm{cm}^{-1}$. Comparing this value with that of GLS4h (Δ of $210.8 \mathrm{cm}^{-1}$) emphasizes again the minimal effects of the parameters F_2 and G_1. α_0 again has the high value of $72 \pm 1 \mathrm{cm}^{-1}$.

In conclusion, starting from the 67 parameters in GLS4, the elimination of F_2 and G_1 increases the mean error by only $1.3 \mathrm{cm}^{-1}$, T_x by a further $5.6 \mathrm{cm}^{-1}$, T by an additional $14.6 \mathrm{cm}^{-1}$, and β by an additional $7.7 \mathrm{cm}^{-1}$.

The values of the parameters of the individual least squares of the final iteration are given in tables 3–40. More detailed discussions of the individual least squares are given in the paper dealing with the correspondence of the energy levels of the various configurations $(3d+4s)^n 4p$, [17].

4.4 Plots of parameters

Figures 1–17 describe graphically the variation of the parameters as functions of the atomic number for the cases of the previous results, Hartree-Fock computations and the final general least-squares (GLS4). In the case of the previous results, PR, the entries are from individual least squares and are indicated by a dot, •; whenever a particular parameter varied freely the $\pm$ value of the error is also given. For the case of Hartree-Fock computations, HF, the entries are given by a solid triangle, ▲; whereas a solid square, ■, denotes the results of the final GLS.

Since T_x, T_x', T_x'', F_1, F_1', F_1'', G_2, G_2', G_2'' are constant in the final results, their graphs clearly need not be drawn. Graphs of all the other parameters are given.

The sets of parameters B, B', B''; C, C', C''; F_2, F_2', F_2''; G_1, G_1', G_1''; G_3, G_3', G_3''; ζ_d, ζ_d', ζ_d''; ζ_p, ζ_p', ζ_p''; are in arithmetic progression in the least squares. Consequently, in order to indicate explicitly this interdependence, separate graphs were drawn where each set of three parameters appears together.

The graphs vividly illustrate and contrast the often irregular behavior of the parameters in the previous results, with the smooth, regular and very reasonable variations obtained in the present project for both the Hartree-Fock computations and, of course, for the general least squares. The contrasts are particularly prominent for the parameters of the interactions between configurations. There is general qualitative agreement between the Hartree-Fock results and those of the GLS. The only notable exceptions are the parameters F_2, F_2', and F_2'', as they increase slowly in the GLS as functions of the atomic number, whereas in the Hartree-Fock computations they decrease.

53

5. References

[1] Racah, G., and Shadmi, Y., Bull. Res. Council Israel, **8F**, No. 1, 15(1959).

[2] Shadmi, Y., Bull. Res. Council Israel, **10F**, No. 3, 109(1962).

[3] Roth, C., J. Res. Nat. Bur. Std., **72A** (Phys. and Chem.) No. 5, 505(1968).

[4] Roth, C., J. Res. Nat. Bur. Std., **73A** (Phys. and Chem.) No. 2, 125(1968).

[5] Roth, C., J. Math. Phys. **9**, 686(1968).

[6] Roth, C., J. Math. Phys. **9**, 1832(1968).

[7] Roth, C., J. Math. Phys. **10**, 1038(1969).

[8] Roth, C., J. Res. Nat. Bur. Stand., **75B**(Math. Sci.) Nos. 1 & 2, 31 (Jan.-June 1971).

[9] Roth, C., J. Res. Nat. Bur. Stand., **76B**(Math. Sci.) Nos. 1 & 2, 61 (Jan.-June 1972).

[10] Roth, C., J. Res. Nat. Bur. Stand., **73A**(Phys. and Chem.) No. 5, 497 (Sept.-Oct. 1969).

[11]. Roth, C., J. Res. Nat. Bur. Stand., **74A**(Phys. and Chem.) No. 2, 141(Mar.-Apr. 1970).

[12] Roth, C., J. Res. Nat. Bur. Stand., **74A**(Phys. and Chem.) No. 2, 157(Mar.-Apr. 1970).

[13] Roth, C., J. Res. Nat. Bur. Stand., **74A**(Phys. and Chem.) No. 4, 507(July-Aug. 1970).

[14] Roth, C., J. Res. Nat. Bur. Stand., **74A**(Phys. and Chem.) No. 2, 181 (Mar.-Apr. 1970).

[15] Roth, C., J. Res. Nat. Bur. Stand., **74A**(Phys. and Chem.) No. 5, 703 (Sept.-Oct. 1970).

[16] Roth, C., J. Res. Nat. Bur. Stand., **74A**(Phys. and Chem.) No. 5, 715 (Sept.-Oct. 1970).

[17] Roth, C., Atomic Data and Nuclear Data Tables. (In press).

[18] Bacher, R.F., and Goudsmit, S., Phys. Rev. **46**, 948(1934).

[19] Trees, R.E., Phys. Rev. **83**, 756(1951).

[20] Trees, R.E., Phys. Rev. **84**, 1089(1951).

[21] Racah, G., Phys. Rev. **85**, 381(1952).

[22] Racah, G., Phys. Rev. **63**, 367(1943).

[23] Rajnak, K., and Wybourne, B.G., Phys. Rev. **132**, 280(1963).

[24] Racah, G., and Stein, J., Phys. Rev. **156**, 58(1967).

[25] Trees, R.E., Phys. Rev. **129**, 1220(1963).

[26] Shadmi, Y., Phys. Rev. **139**, A43(1965).

[27] Roth, C., J. Res. Nat. Bur. Stand., **73A**(Phys. and Chem.) No. 2, 159(1969).

[28] Shadmi, Y., Oreg. J., and Stein, J., J. Opt. Soc. Am **58**, 909(1968).

[29] Shadmi, Y., Caspi, E., and Oreg, J., J. Res. Nat. Bur. Stand., **73A**(Phys. and Chem.) No. 2, 173(1969).

[30] Goldschmidt, Z.B., and Starkand, J., J. Phys. B, L141(1970).

[31] Racah, G., Phys. Rev. **62**, 438(1942).

[32] Condon, E.U., and Shortley, G.H., *The Theory of Atomic Spectra* (Cambridge University Press, 1935), referred to as TAS.

[33] Rosenzweig, N., Phys. Rev. **88**, 580(1952).

[34] Racah, G., Bull. Res. Council Israel 3, 290(1954).

[35] Moore, C.E., Atomic Energy Levels, NBS Circ. 467, I(1949) ibid., II(1952).

[36] Catalan, M.A., Meggers, W.F., and Garcia-Riquelme, J. Res. Nat. Bur. Stands. (U.S.) **68A**(Phys. and Chem.) No. 1, 9(1964).

[37] Roth, C. (To be published).

[38] Martin, W.C., and Sugar, J., J. Opt. Soc. Am. 59, 1266(1969).

TABLE 1. *Results for the parameter D'*

El.	Previous results	DIAGON 1	DIAGON 2	DIAGON 3	DIAGON 4	Final GLS
Ca	−18902	−17851.26	−19140	−19497	−19247	−19252
Sc	−14492	−13958.87	−14420	−14681	−14506	−14525
Ti	−10511	−10066.48	−9940	−10119	−9999	−10026
V	−5504	−6174.09	−5700	−5811	−5726	−5755
Cr	−2275	−2281.70	−1700	−1757	−1687	−1712
Mn	1554	1610.69	2060	2043	2118	2103
Fe	6185	5503.08	5580	5589	5689	5690
Co	8939	9395.47	8860	8881	9026	9049
Ni	12367	13287.85	11900	11919	12129	12180
Cu		17180.24	14700	14703	14998	15083

TABLE 2. *Results for the parameter D″*

El.	Previous results	DIAGON 1	DIAGON 2	DIAGON 3	DIAGON 4	Final GLS
Sc	4833	4817.60	4880	5140	5028	4950
Ti	12010	10768.97	11440	11740	11585	11532
V	17228	16720.34	17760	18086	17908	17886
Cr	20565	22671.71	23840	24178	23997	24012
Mn	27138	28623.08	29680	30016	29852	29910
Fe	36419	34574.45	35280	35600	35473	35580
Co	41960	40525.82	40640	40930	40860	41022
Ni	47690	46477.18	45760	46006	46013	46236
Cu	51395	52428.55	50640	50828	50932	51222
Zn		58379.92	55280	55396	55617	55980

TABLE 3. *Results for the parameter B*

El.	Previous results	Hartree-Fock	DIAGON 1	DIAGON 2	DIAGON 3	DIAGON 4	ILS 3	Final ILS	Final GLS
Sc	529 ± 6	615.95	501	526	515	509	536 ± 7	518 ± 4	510
Ti	554 ± 7	724.20	568	578	568	561	555 ± 6	549 ± 6	561
V	579 ± 12	816.61	635	630	621	613	547 ± 12	570 ± 10	612
Cr	677 ± 7	900.84	702	682	674	665	652 ± 8	666 ± 7	663
Mn	800 ± 14	980.13	769	734	727	717	722 ± 15	739 ± 14	714
Fe	738 ± 9	1056.03	836	786	780	769	768 ± 11	759 ± 4	765
Co	833 ± 8	1129.51	903	838	833	821	841 ± 16	842 ± 10	816

TABLE 4. *Results for the parameter B′*

El.	Previous results	Hartree-Fock	DIAGON 1	DIAGON 2	DIAGON 3	DIAGON 4	ILS 3	Final ILS	Final GLS
Ti	651 ± 7	850.52	670	677	669	665	655 ± 7	655 ± 6	664
V	730 ± 5	934.80	731	739	729	726	723 ± 6	715 ± 5	724
Cr	755 ± 5	1013.87	792	801	789	787	769 ± 8	776 ± 7	784
Mn	861 ± 4	1089.50	853	863	849	848	843 ± 6	843 ± 7	844
Fe	943 ± 7	1162.72	914	925	909	909	901 ± 5	904 ± 3	904
Co	956 ± 6	1234.15	975	987	969	970	983 ± 10	1001 ± 12	964
Ni	1024 ± 7	1304.17	1036	1049	1029	1031	1021 ± 7	1017 ± 7	1024

TABLE 5. *Results for the parameter B″*

El.	Previous results	Hartree-Fock	DIAGON 1	DIAGON 2	DIAGON 3	DIAGON 4	ILS 3	Final ILS	Final GLS
V	881	1041.24	827	848	837	839	899	860	833
Cr	833	1116.47	882	920	904	909	886	886	903
Mn	922	1189.31	937	992	971	979	964	948	973
Fe	1148	1260.59	992	1064	1038	1049	1034	1049	1043
Co	1079	1330.49	1047	1136	1105	1119	1135	1160	1113
Ni	1109 (FIX)	1399.33	1102	1208	1172	1189	1164	1175	1183
Cu	1000 (FIX)	1467.33	1157	1280	1239	1259	1239 (FIX)	1259 (FIX)	1253

TABLE 6. *Results for the parameter C*

El.	Previous results	Hartree-Fock	DIAGON 1	DIAGON 2	DIAGON 3	DIAGON 4	ILS 3	Final ILS	Final GLS
Sc	714 ± 69	2221.17	1826	1266	1222	1240	762 ± 69	866 ± 47	1281
Ti	1661 ± 33	2623.15	2070	1674	1655	1677	1741 ± 31	1747 ± 30	1714
V	2084 ± 23	2965.90	2314	2082	2088	2114	2204 ± 42	2228 ± 38	2147
Cr	2445 ± 15	3277.50	2558	2490	2521	2551	2596 ± 51	2535 ± 48	2580
Mn	2772 ± 17	3569.86	2802	2898	2954	2988	3121 ± 71	3108 ± 63	3013
Fe	3310 ± 29	3849.30	3046	3306	3387	3425	3404 ± 31	3435 ± 25	3446
Co	3744 ± 63	4119.39	3290	3714	3820	3862	3852 ± 84	3797 ± 70	3879

TABLE 7. *Results for the parameter C'.*

El.	Previous results	Hartree-Fock	DIAGON 1	DIAGON 2	DIAGON 3	DIAGON 4	ILS 3	Final ILS	Final GLS
Ti	2319 ± 57	3131.03	2473	2343	2257	2285	2380 ± 38	2354 ± 38	2318
V	2587 ± 19	3443.16	2693	2647	2611	2636	2658 ± 29	2722 ± 30	2666
Cr	2918 ± 12	3735.06	2913	2951	2965	2987	3029 ± 39	2999 ± 37	3014
Mn	3140 ± 9	4013.59	3133	3255	3319	3338	3410 ± 57	3413 ± 50	3362
Fe	3509 ± 14	4282.74	3353	3559	3673	3689	3687 ± 18	3688 ± 16	3710
Co	3875 ± 18	4545.00	3573	3863	4027	4040	3994 ± 35	3954 ± 34	4058
Ni	4187 ± 54	4801.79	3793	4167	4381	4391	4394 ± 45	4439 ± 43	4406

TABLE 8. *Results for the parameter C"*

El.	Previous results	Hartree-Fock	DIAGON 1	DIAGON 2	DIAGON 3	DIAGON 4	ILS 3	Final ILS	Final GLS
V	3090	3878.88	3072	3212	3134	3158	3112	3216	3185
Cr	3391	4155.83	3268	3412	3409	3423	3462	3463	3448
Mn	3508	4423.76	3464	3612	3684	3688	3699	3718	3711
Fe	3708	4685.03	3660	3812	3959	3953	3970	3940	3974
Co	4006	4941.20	3856	4012	4234	4218	4136	4111	4237
Ni	4287	5193.31	4052	4212	4509	4483	4522	4531	4500
Cu	4500 (FIX)	5442.15	4248	4412	4784	4748	4784 (FIX)	4748 (FIX)	4763

TABLE 9. *Results for the parameter F_2*

El.	Previous results	Hartree-Fock	DIAGON 1	DIAGON 2	DIAGON 3	DIAGON 4	ILS 3	Final ILS	Final GLS
Ca	128 ± 2	306.13	273.34	200	151	178	171 ± 24	190 ± 27	170.5
Sc	201 ± 8	258.97	259.70	200	156	180	212 ± 9	200 ± 11	173
Ti	153 ± 9	235.96	246.05	200	161	182	163 ± 6	167 ± 5	175.5
V	160 ± 8	220.50	232.40	200	166	184	182 ± 7	168 ± 6	178
Cr	187 ± 10	208.61	218.75	200	171	186	196 ± 8	189 ± 5	180.5
Mn	193 ± 7	198.79	205.10	200	176	188	206 ± 9	178 ± 8	183
Fe	173 ± 10	190.32	191.45	200	181	190	205 ± 7	202 ± 7	185.5
Co	218 ± 10	182.79	177.80	200	186	192	221 ± 10	226 ± 9	188
Ni	166 ± 13	175.92	164.15	200	191	194	167 ± 21	170 ± 15	190.5

TABLE 10. *Results for the parameter F_2'*

El.	Previous results	Hartree-Fock	DIAGON 1	DIAGON 2	DIAGON 3	DIAGON 4	ILS 3	Final ILS	Final GLS
Sc	284 ± 8	300.16	305.77	300	279	291	297 ± 11	285 ± 12	289
Ti	286 ± 8	290.88	298.09	300	281.5	293	318 ± 5	317 ± 5	291.5
V	282 ± 6	283.55	290.41	300	284	295	286 ± 6	274 ± 6	294
Cr	275 ± 5	277.22	282.73	300	286.5	297	287 ± 5	297 ± 5	296.5
Mn	314 ± 4	271.46	275.06	300	289	299	308 ± 10	290 ± 10	299
Fe	305 ± 5	266.08	267.38	300	291.5	301	315 ± 4	322 ± 5	301.5
Co	303 ± 7	260.91	259.70	300	294	303	305 ± 8	311 ± 7	304
Ni	311 ± 7	255.92	252.02	300	269.5	305	300 ± 9	304 ± 8	306.5
Cu	329 ± 12	251.02	244.35	300	299	307	319 ± 7	319 ± 9	309

TABLE 11. *Results for the parameter F_2''*

El.	Previous results	Hartree-Fock	DIAGON 1	DIAGON 2	DIAGON 3	DIAGON 4	ILS 3	Final ILS	Final GLS
Ti	419	354.97	350.13	400	402	404	473	468	407.5
V	404	352.41	348.42	400	402	406	390	380	410
Cr	363	349.86	346.71	400	402	408	378	405	412.5
Mn	435	347.26	345.01	400	402	410	410	402	415
Fe	437	344.57	343.31	400	402	412	425	442	417.5
Co	388	341.77	341.60	400	402	414	389	396	420
Ni	456	338.86	339.89	400	402	416	433	434	422.5
Cu	480 (FIX)	335.82	338.18	400	402	418	422	432	425
Zn		332.70		400	402	420			427.5

TABLE 12. *Results for the parameter G_1*

El.	Previous results	Hartree-Fock	DIAGON 1	DIAGON 2	DIAGON 3	DIAGON 4	ILS 3	Final ILS	Final GLS
Ca	394 ± 2	559.20	501.60	231	305	328	419 ± 29	423 ± 32	327.5
Sc	335 ± 9	367.63	405.13	226	289	309	344 ± 10	347 ± 11	309
Ti	283 ± 10	290.99	324.32	221	273	290	287 ± 5	288 ± 5	290.5
V	229 ± 8	247.17	259.17	216	257	271	251 ± 7	251 ± 7	272
Cr	224 ± 7	217.86	209.69	211	241	252	243 ± 8	241 ± 7	253.5
Mn	227 ± 7	196.41	175.88	206	225	233	220 ± 8	230 ± 9	235
Fe	202 ± 8	179.74	157.73	201	209	214	237 ± 7	238 ± 5	216.5
Co	196 ± 9	166.21	155.25	196	193	195	199 ± 9	199 ± 10	198
Ni	165 ± 9	154.84	168.43	191	177	176	172 ± 11	171 ± 10	179.5

TABLE 13. *Results for the parameter G_1'*

El.	Previous results	Hartree-Fock	DIAGON 1	DIAGON 2	DIAGON 3	DIAGON 4	ILS 3	Final ILS	Final GLS
Sc	327 ± 12	328.91	349.14	330	323	336	354 ± 13	344 ± 16	335
Ti	288 ± 10	293.92	309.30	311	307	317	328 ± 6	325 ± 6	316.5
V	281 ± 8	271.04	276.55	292	291	298	301 ± 6	294 ± 7	298
Cr	236 ± 5	254.27	250.87	273	275	279	239 ± 7	262 ± 5	279.5
Mn	245 ± 6	241.08	232.26	254	259	260	232 ± 9	247 ± 8	261
Fe	245 ± 7	230.18	220.73	235	243	241	248 ± 6	255 ± 5	242.5
Co	199 ± 7	220.78	216.28	216	227	222	210 ± 8	213 ± 7	224
Ni	223 ± 9	212.51	218.90	197	211	203	227 ± 12	226 ± 12	205.5
Cu	284 ± 2	205.00	228.59	178	195	184	206 ± 13	208 ± 16	187

TABLE 14. *Results for the parameter G_1''*

El.	Previous results	Hartree-Fock	DIAGON 1	DIAGON 2	DIAGON 3	DIAGON 4	ILS 3	Final ILS	Final GLS
Ti	293	311.36	294.28	401	341	344	369	362	342.5
V	333	300.52	293.93	368	325	325	351	337	324
Cr	248	292.02	292.05	335	309	306	235	283	305.5
Mn	263	284.95	288.64	302	293	287	244	264	287
Fe	288	278.81	283.73	269	277	268	259	272	268.5
Co	202	273.28	277.31	236	261	249	221	227	250
Ni	281	268.19	269.37	203	245	230	282	281	231.5
Cu	284 (EQ.)	263.39	259.91	170	229	211	240	235	213
Zn		258.81	248.94	137	213	192			194.5

TABLE 15. *Results for the parameter G_3*

El.	Previous results	Hartree-Fock	DIAGON 1	DIAGON 2	DIAGON 3	DIAGON 4	ILS 3	Final ILS	Final GLS
Ca	0 (FIX)	66.67	30.42	62	5	15	5 (FIX)	15 (FIX)	13
Sc	5 ± 3	45.84	24.96	55	5	15	8 ± 6	4 ± 5	13
Ti	10 ± 3	37.21	20.38	48	5	15	8 ± 3	7 ± 3	13
V	13 ± 2	32.11	16.67	41	5	15	18 ± 3	14 ± 3	13
Cr	16 ± 1	28.61	13.82	34	5	15	10 ± 3	7 ± 2	13
Mn	18 ± 1	25.98	11.85	27	5	15	24 ± 3	14 ± 3	13
Fe	20 ± 2	23.90	10.75	20	5	15	17 ± 3	18 ± 2	13
Co	14 ± 2	22.18	10.52	13	5	15	13 ± 4	13 ± 3	13
Ni	14 ± 4	20.70	11.16	6	5	15	7 ± 5	6 ± 5	13

TABLE 16. *Results for the parameter G_3'*

El.	Previous results	Hartree-Fock	DIAGON 1	DIAGON 2	DIAGON 3	DIAGON 4	ILS 3	Final ILS	Final GLS
Sc	5 (EQ.)	45.63	24.04	60	15	23	18 ± 6	4 (EQ.)	21
Ti	10 (EQ.)	41.50	21.68	53	15	23	21 ± 3	22 ± 3	21
V	13 (EQ.)	38.66	19.72	46	15	23	24 ± 3	17 ± 3	21
Cr	16 (EQ.)	36.50	18.16	39	15	23	29 ± 3	30 ± 2	21
Mn	18 (EQ.)	34.73	16.99	32	15	23	30 ± 4	22 ± 4	21
Fe	20 (EQ.)	33.23	16.21	25	15	23	20 ± 2	21 ± 4	21
Co	14 (EQ.)	31.90	15.82	18	15	23	14 ± 3	15 ± 2	21
Ni	14 (EQ.)	30.71	15.84	11	15	23	11 ± 6	12 ± 6	21
Cu	51 ± 1	29.62	16.24	4	15	23	31 ± 6	35 ± 7	21

TABLE 17. *Results for the parameter G_3''*

El.	Previous results	Hartree-Fock	DIAGON 1	DIAGON 2	DIAGON 3	DIAGON 4	ILS 3	Final ILS	Final GLS
Ti	10	47.75	22.98	58	25	31	34	37	29
V	13	46.23	22.77	51	25	31	30	20	29
Cr	16	44.98	22.49	44	25	31	48	53	29
Mn	18	43.89	22.12	37	25	31	36	30	29
Fe	20	42.90	21.67	30	25	31	23	24	29
Co	14	42.00	21.13	23	25	31	15	17	29
Ni	14	41.14	20.52	16	25	31	15	18	29
Cu	51 (EQ.)	40.32	19.82	9	25	31	41	43	29
Zn		39.53	19.03	2	25	31			29

TABLE 18. *Results for the parameter G_p.*

El.	Previous results	Hartree-Fock	DIAGON 1	DIAGON 2	DIAGON 3	DIAGON 4	ILS 3	Final ILS	Final GLS
Ca	4977 ± 19	8451.47	5384.41	5326	4736	5059	5135 ± 89	4917 ± 86	5129
Sc	5970 ± 82	8924.04	5631.23	5564	5250	5449	5906 ± 76	5850 ± 71	5484
Ti	5395 ± 97	9294.41	5856.49	5802	5704	5803	5873 ± 63	5900 ± 66	5809
V	6022 ± 80	9605.22	6060.20	6040	6098	6121	6000 ± 66	5936 ± 67	6104
Cr	6155 ± 45	9873.36	6242.34	6278	6432	6403	6479 ± 55	6493 ± 52	6369
Mn	6631 ± 60	10107.40	6402.92	6516	6706	6649	6641 ± 61	6498 ± 54	6604
Fe	7116 ± 58	10313.14	6541.94	6754	6920	6859	6726 ± 37	6746 ± 36	6809
Co	7038 ± 56	10493.43	6659.39	6992	7074	7033	6991 ± 51	7005 ± 50	6984
Ni	7027 ± 60	10651.78	6755.28	7230	7168	7171	7011 ± 76	7011 ± 72	7129
Cu	8425 ± 17	10789.03	6829.62	7468	7202	7273	7994 ± 96	8003 ± 92	7244
Zn		10908.11	6882.39	7706	7176	7339			7329

TABLE 19. *Results for the parameter G_d.*

El.	Previous results	Hartree-Fock	DIAGON 1	DIAGON 2	DIAGON 3	DIAGON 4	ILS 3	Final ILS	Final GLS
Sc	1943 ± 68	2007.15	1840.72	1574	1670	1637	1787 ± 62	1930 ± 65	1649
Ti	1719 ± 56	1876.97	1728.73	1506	1625	1581	1538 ± 37	1560 ± 66	1590
V	1584 ± 33	1795.01	1637.57	1470	1590	1541	1531 ± 29	1563 ± 30	1541
Cr	1590 ± 19	1739.54	1567.26	1466	1565	1517	1513 ± 24	1474 ± 24	1514
Mn	1532 ± 22	1700.66	1517.80	1494	1550	1509	1515 ± 34	1501 ± 37	1505
Fe	1536 ± 24	1672.86	1489.17	1554	1545	1517	1561 ± 19	1556 ± 18	1514
Co	1607 ± 33	1652.90	1481.39	1646	1550	1541	1554 ± 46	1557 ± 41	1541
Ni	1626 ± 53	1638.76	1494.45	1770	1565	1581	1590 ± 68	1589 ± 65	1590
Cu	1550 (FIX)	1629.02	1528.35	1926	1590	1637	1274 ± 86	1250 ± 83	1649

TABLE 20. *Results for the parameter G^**

El.	Previous results	Hartree-Fock	DIAGON 1	DIAGON 2	DIAGON 3	DIAGON 4	ILS 3	Final ILS	Final GLS
Sc	1943 (EQ.)	3223.33	2855.34	2282	2070	1868	2187 ± 62	2161 (F.D.)	1880
Ti	1719 (EQ.)	2936.23	2689.36	2214	2025	1812	1938 ± 37	1789 (F.D.)	1821
V	1584 (EQ.)	2751.21	2544.23	2178	1990	1772	2233 ± 91	1794 (F.D.)	1772
Cr	1590 (EQ.)	2622.68	2419.94	2174	1965	1748	1784 ± 87	1882 ± 82	1745
Mn	1532 (EQ.)	2529.16	2316.48	2202	1950	1740	1915 ± 34	1732 (F.D.)	1736
Fe	1536 (EQ.)	2458.99	2233.87	2262	1945	1748	1961 ± 19	1787 (F.D.)	1745
Co	1607 (EQ.)	2405.28	2172.11	2354	1950	1772	1954 ± 46	1788 (F.D.)	1772
Ni	1626 EQ.)	2363.57	2131.18	2478	1965	1812	1990 ± 88	1820 (F.D.)	1821
Cu		2330.91	2111.11	2634	1990	1868	1674 ± 101	1481 (F.D.)	1880

TABLE 21. *Results for the parameter H*

El.	Previous results	Hartree-Fock	DIAGON 1	DIAGON 2	DIAGON 3	DIAGON 4	ILS 3	Final ILS	Final GLS
Sc	275 ± 18	323.61	313.20	158	238	224	331 ± 27	310 ± 31	215
Ti	175 ± 7	267.45	267.36	144	197	204	184 ± 5	184 ± 4	192
V	150 ± 6	224.55	226.76	130	162	184	165 ± 6	166 ± 5	169
Cr	157 ± 4	190.00	191.38	116	133	164	166 ± 5	161 ± 5	146
Mn	70 ± 6	161.19	161.23	102	110	144	100 ± 8	121 ± 9	123
Fe	85 ± 6	136.54	136.31	88	93	124	92 ± 5	90 ± 5	100
Co	72 ± 6	115.04	116.61	74	82	104	82 ± 6	81 ± 6	77
Ni	154 ± 33	95.99	102.15	60	77	84	118 ± 22	123 ± 18	54

TABLE 22. *Results for the parameter H'*

El.	Previous results	Hartree-Fock	DIAGON 1	DIAGON 2	DIAGON 3	DIAGON 4	ILS 3	Final ILS	Final GLS
Ti	175 (EQ.)	189.71	199.65	104	141	144	141 (FIX)	124 (F.D.)	153
V	150 (EQ.)	152.87	159.04	90	106	124	35 ± 18	106 (F.D.)	130
Cr	157 (EQ.)	122.62	123.66	76	77	104	84 ± 15	61 ± 13	107
Mn	70 (EQ.)	97.00	93.51	62	54	84	54 (FIX)	61 (F.D.)	84
Fe	85 (EQ.)	74.76	68.59	48	37	64	37 (FIX)	30 (F.D.)	61
Co	72 (EQ.)	55.13	48.90	34	26	44	26 (FIX)	21 (F.D.)	38
Ni	154 (EQ.)	37.58	34.44	20	21	24	21 (FIX)	63 (F.D.)	15
Cu	0 (FIX)	21.65	25.20	6	22	4	22 (FIX)	4 (FIX)	0 (FIX)

TABLE 23. *Results for the parameter J*

El.	Previous results	Hartree-Fock	DIAGON 1	DIAGON 2	DIAGON 3	DIAGON 4	ILS 3	Final ILS	Final GLS
Ca	575 ± 20	2607.28	1485.69	1480	1260	1237	1260 (FIX)	1237 (FIX)	1112
Sc	1877 ± 96	2239.09	1367.98	1330	1040	1088	1668 ± 91	1579 ± 85	1066
Ti	1251 ± 53	2041.11	1263.71	1220	880	979	1326 ± 44	1313 ± 44	1036
V	972 ± 48	1907.24	1172.89	1150	780	910	992 ± 46	946 ± 40	1022
Cr	954 ± 32	1806.73	1095.51	1120	740	881	1061 ± 35	990 ± 31	1024
Mn	1294 ± 34	1726.42	1031.57	1130	760	892	1292 ± 56	1016 ± 55	1042
Fe	1183 ± 41	1659.47	981.08	1180	840	943	1004 ± 39	1025 ± 37	1076
Co	1245 ± 50	1601.79	944.02	1270	980	1034	1224 ± 55	1210 ± 55	1126
Ni	1144 ± 134	1550.86	920.42	1400	1180	1165	906 ± 89	913 ± 81	1192
Cu	(2446 ± 365)	1505.08	910.25	1570	1440	1336	1440 (FIX)	1336 (FIX)	1274

TABLE 24. *Results for the parameter J'*

El.	Previous results	Hartree-Fock	DIAGON 1	DIAGON 2	DIAGON 3	DIAGON 4	ILS 3	Final ILS	Final GLS
Sc	1877 (EQ.)	2250.32	1371.30	1610	1380	1540	2008 (F.D.)	2031 (F.D.)	1596
Ti	1251 (EQ.)	2134.09	1292.59	1500	1220	1431	1666 (F.D.)	1765 (F.D.)	1566
V	972 (EQ.)	2052.61	1227.31	1430	1120	1362	1675 (F.D.)	1337 ± 89	1552
Cr	954 (EQ.)	1990.63	1175.48	1400	1080	1333	1401 (F.D.)	1442 (F.D.)	1554
Mn	1294 (EQ.)	1940.85	1137.10	1410	1100	1344	1632 (F.D.)	1468 (F.D.)	1572
Fe	1183 (EQ.)	1899.24	1112.15	1460	1180	1395	1344 (F.D.)	1477 (F.D.)	1606
Co	1245 (EQ.)	1863.20	1100.65	1550	1320	1486	1564 (F.D.)	1662 (F.D.)	1656
Ni	1144 (EQ.)	1831.30	1102.59	1680	1520	1617	1246 (F.D.)	1365 (F.D.)	1722
Cu	(2446 (EQ.))	1802.11	1117.97	1850	1780	1788	1780 (FIX)	1788 (FIX)	1804
Zn		(1393.74)	1146.79	2060	2100	1999			1902

TABLE 25. *Results for the parameter K.*

El.	Previous results	Hartree-Fock	DIAGON 1	DIAGON 2	DIAGON 3	DIAGON 4	ILS 3	Final ILS	Final GLS
Ca	3795 ± 32	4824.07	3535.35	2710	2875	2930	2922 ± 98	2571 ± 83	3009
Sc	2551 ± 95	4001.34	3215.28	2515	2680	2741	2743 ± 85	2638 ± 82	2812
Ti	2415 ± 48	3590.97	2933.52	2370	2525	2586	2613 ± 30	2611 ± 29	2645
V	2468 ± 48	3324.70	2690.06	2275	2410	2465	2555 ± 42	2520 ± 39	2508
Cr	2311 ± 30	3130.26	2484.91	2230	2335	2378	2423 ± 33	2350 ± 29	2401
Mn	2599 ± 36	2977.82	2318.05	2235	2300	2325	2597 ± 46	2358 ± 52	2324
Fe	2459 ± 45	2852.49	2189.51	2290	2305	2306	2211 ± 51	2236 ± 36	2277
Co	2331 ± 55	2745.36	2099.27	2395	2350	2321	2328 ± 63	2323 ± 53	2260
Ni	2072 ± 135	2651.31	2047.34	2550	2435	2370	1980 ± 96	1982 ± 89	2273
Cu	(5090 ± 125)	2566.78	2033.70	2755	2560	2453	4122 ± 97	3864 ± 90	2316

TABLE 26. *Results for the parameter K′*

El.	Previous results	Hartree-Fock	DIAGON 1	DIAGON 2	DIAGON 3	DIAGON 4	ILS 3	Final ILS	Final GLS
Sc	3059 (F.D.)	3944.13	3670.74	3535	3480	3639	3543 (F.D.)	3536 (F.D.)	3759
Ti	2987 (F.D.)	3715.02	3427.99	3390	3325	3484	3413 (F.D.)	3509 (F.D.)	3592
V	3104 (F.D.)	3561.34	3229.81	3295	3210	3363	3342±73	2828±89	3455
Cr	3011 (F.D.)	3448.62	3076.21	3250	3135	3276	3353±88	3458±87	3348
Mn	3363 (F.D.)	3360.71	2967.18	3255	3100	3223	3397±74	3256 (F.D.)	3271
Fe	3287 (F.D.)	3288.84	2902.72	3310	3105	3204	3011±79	3134 (F.D.)	3224
Co	3223 (F.D.)	3227.54	2882.84	3415	3150	3219	3128±87	3221 (F.D.)	3207
Ni	3028 (F.D.)	3173.85	2907.54	3570	3235	3268	2780 (F.D.)	2880 (F.D.)	3220
Cu	(5090 (EQ.))	3124.84	2976.79	3775	3360	3351	4922 (F.D.)	4762 (F.D.)	3263
Zn		(2275.24)	3090.63	4030	3525	3468			3336

TABLE 27. *Results for the parameter ζ_d*

El.	Previous results	Hartree-Fock	DIAGON 1	DIAGON 2	DIAGON 3	DIAGON 4	ILS 3	Final ILS	Final GLS
Ca	18±9	27.94	47.87	10	1	0	1 (FIX)	0 (FIX)	0 (FIX)
Sc	58±21	59.25	60.81	80	49	41	53±26	34±28	13
Ti	114±29	97.59	90.04	150	103	90	99±25	97±24	73.5
V	141±22	145.93	135.56	220	163	147	148±34	172±33	139
Cr	247±24	206.17	197.37	290	229	212	235±26	238±25	209.5
Mn	321±28	280.14	275.46	360	301	285	215±41	228±41	285
Fe	410±23	369.69	369.85	430	379	366	367±38	341±23	365.5
Co	517±17	476.79	480.52	500	463	455	447±48	461±45	451
Ni	603±19	603.50	607.48	570	553	552	582±52	580±50	541.5

TABLE 28. *Results for the parameter ζ_d'*

El.	Previous results	Hartree-Fock	DIAGON 1	DIAGON 2	DIAGON 3	DIAGON 4	ILS 3	Final ILS	Final GLS
Sc	58 (EQ.)	79.12	80.69	100	35	42	78±42	79±35	15
Ti	114 (EQ.)	121.83	116.21	180	102	99	108±23	99±22	88.5
V	141 (EQ.)	174.62	168.01	260	175	164	149±19	160±20	167
Cr	247 (EQ.)	239.97	236.11	340	254	237	296±75	223±52	250.5
Mn	328 (EQ.)	319.65	320.49	420	339	318	368±51	343±55	339
Fe	410 (EQ.)	415.55	421.16	500	430	407	446±24	439±30	432.5
Co	517 (EQ.)	529.64	538.12	580	527	504	535±21	541±20	531
Ni	603 (EQ.)	663.96	671.37	660	630	609	639±22	640±22	634.5
Cu	836±3	820.74	820.91	740	739	722	860±48	847±42	743

TABLE 29. *Results for the parameter ζ_d''*

El.	Previous results	Hartree-Fock	DIAGON 1	DIAGON 2	DIAGON 3	DIAGON 4	ILS 3	Final ILS	Final GLS
Ti	114	146.57	142.38	210	101	108	117	101	103.5
V	141	205.25	200.46	300	187	181	150	148	195
Cr	247	275.91	274.85	390	279	262	357	208	291.5
Mn	328	361.53	365.52	480	377	351	521	458	393
Fe	410	464.00	472.48	570	481	448	525	537	499.5
Co	517	585.31	595.72	660	591	553	623	621	611
Ni	603	727.55	735.26	750	707	666	696	700	727.5
Cu	836 (EQ.)	892.91	891.09	840	829	787	950	912	849
Zn		1083.68	1063.20	930	957	916			975.5

TABLE 30. *Results for the parameter ζ_p*

El.	Previous results	Hartree-Fock	DIAGON 1	DIAGON 2	DIAGON 3	DIAGON 4	ILS 3	Final ILS	Final GLS
K		37.03	10	145	232				197
Ca	87±16	36.66	45.69	40	109	177	141±67	139±65	155
Sc	105±56	44.52	54.35	70	83	134	147±55	125±53	123
Ti	114±94	52.28	63.02	100	67	103	93±51	82±48	101
V	140±78	60.06	71.70	130	61	84	96±70	95±60	89
Cr	184±51	67.91	80.36	160	65	77	118±56	118±51	87
Mn	212±60	75.88	89.04	190	79	82	0 (FIX)	82 (FIX)	95
Fe	200±68	83.96	97.70	220	103	99	157±57	106±52	113
Co	236±53	92.14	106.38	250	137	128	111±58	187±54	141
Ni	255±51	100.38	115.05	280	181	169	263±71	270±71	179
Cu	375±8	108.71	123.71	310	235	222	208±94	201±82	227

TABLE 31. *Results for the parameter ζ_p'*

El.	Previous results	Hartree-Fock	DIAGON 1	DIAGON 2	DIAGON 3	DIAGON 4	ILS 3	Final ILS	Final GLS
Ca	89 (EQ.)	63.40	77.59	100	244	333	277 (F.D.)	295 (F.D.)	294
Sc	105 (EQ.)	79.93	101.90	130	222	290	190±55	185±52	262
Ti	114 (EQ.)	96.20	126.22	160	210	259	176±59	192±57	240
V	140 (EQ.)	112.88	150.55	190	208	240	156±57	137±51	228
Cr	184 (EQ.)	130.15	174.86	220	216	223	230±49	268±40	226
Mn	212 (EQ.)	148.07	199.19	250	234	238	370±52	377±46	234
Fe	200 (EQ.)	166.68	223.50	280	262	255	339±56	278±52	252
Co	236 (EQ.)	185.90	247.82	310	300	284	270±62	246±61	280
Ni	255 (EQ.)	205.76	272.15	340	348	325	324±63	323±62	318
Cu	375 (EQ.)	226.17	296.46	370	406	378	408±74	388±68	366
Zn		247.11	320.79	400	474	443			424

TABLE 32. *Results for the parameter ζ_p''*

El.	Previous results	Hartree-Fock	DIAGON 1	DIAGON 2	DIAGON 3	DIAGON 4	ILS 3	Final ILS	Final GLS
Sc	105	131.98	149.95	190	361	446	333	245	401
Ti	114	157.90	189.42	220	353	415	259	302	379
V	140	185.31	229.40	250	355	396	216	179	367
Cr	184	214.35	269.36	280	367	389	342	418	365
Mn	212	245.08	309.34	310	389	394	740	772	373
Fe	200	277.51	349.30	340	421	411	521	450	391
Co	236	311.61	389.28	370	463	440	429	305	419
Ni	255	347.35	429.25	400	515	481	385	376	457
Cu	375	384.69	469.22	430	577	534	608	575	505
Zn		423.56	509.19	460	649	599			563
Ga		463.94	549.16	490	731	676			631

TABLE 33. *Results for the parameter $\alpha = \alpha' = \alpha''$*

El.	Previous results	DIAGON 1	DIAGON 2	DIAGON 3	DIAGON 4	ILS 3	Final ILS	Final GLS
Sc	50 (FIX)	60	23	31	28	54±9	28 (FIX)	19
Ti	43±4	60	29	33	30	19±6	18±5	22.5
V	54±2	60	35	35	32	37±6	24±5	26
Cr	63±1	60	41	37	34	37±6	32±4	29.5
Mn	73±2	60	47	39	36	16±6	15±5	33
Fe	74±2	60	53	41	38	40±3	39±3	36.5
Co	71±3	60	59	43	40	39±5	37±5	40
Ni	83±7	60	65	45	42	55±7	52±7	43.5
Cu		60	71	47	44		44 (FIX)	47

TABLE 34. *Results for the parameter $\beta = \beta' = \beta''$*

EL.	DIAGON 1	DIAGON 2	DIAGON 3	DIAGON 4	ILS 3	FINAL ILS	FINAL GLS
Sc	0	−250	−20	−308	−20 (FIX)	−308 (FIX)	−435
Ti	0	−250	−110	−350	−173 ± 76	−243 ± 69	−458
V	0	−250	−200	−392	−507 ± 67	−740 ± 68	−481
Cr	0	−250	−290	−434	−506 ± 68	−426 ± 62	−504
Mn	0	−250	−380	−476	−634 ± 72	−637 ± 68	−527
Fe	0	−250	−470	−518	−470 (FIX)	−518 (FIX)	−550
Co	0	−250	−560	−560	−560 (FIX)	−560 (FIX)	−573
Ni	0	−250	−650	−602	−650 (FIX)	−602 (FIX)	−596
Cu	0	−250	−740	−644	−740 (FIX)	−644 (FIX)	−619

TABLE 35. *Results for the parameter $T = T' = T''$*

EL.	DIAGON 1	DIAGON 2	DIAGON 3	DIAGON 4	ILS 3	FINAL ILS	FINAL GLS
Sc	0	−4.5	−1.9	−1.7	−1.9 (FIX)	−1.7 (FIX)	−2.7
Ti	0	−4.1	−2.2	−2.1	−2.7 ± 0.8	−2.6 ± 0.7	−2.9
V	0	−3.7	−2.5	−2.5	−1.9 ± 0.6	−2.7 ± 0.6	−3.1
Cr	0	−3.3	−2.8	−2.9	−2.6 ± 0.6	−3.2 ± 0.5	−3.3
Mn	0	−2.9	−3.1	−3.3	−5.6 ± 0.9	−5.6 ± 0.7	−3.5
Fe	0	−2.5	−3.4	−3.7	−3.3 ± 0.4	−3.1 ± 0.3	−3.7
Co	0	−2.1	−3.7	−4.1	−4.0 ± 0.8	−4.3 ± 0.6	−3.9
Ni	0	−1.7	−4.0	−4.5	−4 (FIX)	−4.5 (FIX)	−4.1
Cu	0	−1.3	−4.3	−4.9	−4.3 (FIX)	−4.9 (FIX)	−4.3

TABLE 36. *Results for the parameter $T_x = T'_x = T''_x$*

EL.	DIAGON 1	DIAGON 2	DIAGON 3	DIAGON 4	ILS 3	FINAL ILS	FINAL GLS
Sc	0	−1.8	−1.4	−2.8	−1.4 (FIX)	−1.4 (FIX)	−2.6
Ti	0	−2.1	−1.7	−2.8	−1.7 (FIX)	−2.8 (FIX)	−2.6
V	0	−2.4	−2.0	−2.8	−2.0 (FIX)	−2.8 (FIX)	−2.6
Cr	0	−2.7	−2.3	−2.8	−2.3 (FIX)	−4.0 ± 0.8	−2.6
Mn	0	−3.0	−2.6	−2.8	−2.6 (FIX)	−2.8 (FIX)	−2.6
Fe	0	−3.3	−2.9	−2.8	−2.9 (FIX)	−2.8 (FIX)	−2.6
Co	0	−3.6	−3.2	−2.8	−3.2 (FIX)	−6.0 ± 0.8	−2.6
Ni	0	−3.9	−3.5	−2.8	−3.5 (FIX)	−2.8 (FIX)	−2.6
Cu	0	−4.2	−3.8	−2.8	−3.8 (FIX)	−2.8 (FIX)	−2.6

TABLE 37. *Results for the parameter $F_1 = F'_1 = F''_1$*

EL.	DIAGON 1	DIAGON 2	DIAGON 3	DIAGON 4	ILS 3	FINAL ILS	FINAL GLS
Ca	0	−10	−45	−6	−45 (FIX)	−6 (FIX)	−13
Sc	0	−10	−43	−6	−43 (FIX)	−46 ± 3	−13
Ti	0	−10	−41	−6	−12 ± 4	−11 ± 4	−13
V	0	−10	−39	−6	−10 ± 3	−13 ± 4	−13
Cr	0	−10	−37	−6	−19 ± 4	−6 (FIX)	−13
Mn	0	−10	−35	−6	−30 ± 5	−14 ± 4	−13
Fe	0	−10	−33	−6	−30 ± 4	−30 ± 3	−13
Co	0	−10	−31	−6	−19 ± 6	−14 ± 6	−13
Ni	0	−10	−29	−6	−42 ± 6	−38 ± 6	−13
Cu	0	−10	−27	−6	−27 (FIX)	−6 (FIX)	−13
Zn	0	−10	−25	−6			−13

TABLE 38. *Results for the parameter G_2*

EL.	DIAGON 1	DIAGON 2	DIAGON 3	DIAGON 4	ILS 3	FINAL ILS	FINAL GLS
Ca	0	1	0	13	0 (FIX)		9
Sc	0	5	0	13	27±9	14±4	9
Ti	0	9	0	13	0 (FIX)	0 (FIX)	9
V	0	13	0	13	21±6	11±3	9
Cr	0	17	0	13	0 (FIX)	4±3	9
Mn	0	21	0	13	24±9	3±3	9
Fe	0	25	0	13	30±6	28±4	9
Co	0	29	0	13	0 (FIX)	5±3	9
Ni	0	33	0	13	0 (FIX)	9±4	9

TABLE 39. *Results for the parameter G_2'*

EL.	DIAGON 1	DIAGON 2	DIAGON 3	DIAGON 4	ILS 3	FINAL ILS	FINAL GLS
Sc	0	5	15	13	27 (EQ.)	14 (EQ.)	9
Ti	0	9	15	13	15 (FIX)	0 (FIX)	9
V	0	13	15	13	21 (EQ.)	11 (EQ.)	9
Cr	0	17	15	13	15 (FIX)	4 (EQ.)	9
Mn	0	21	15	13	24 (EQ.)	3 (EQ.)	9
Fe	0	25	15	13	30 (EQ.)	28 (EQ.)	9
Co	0	29	15	13	15 (FIX)	5 (EQ.)	9
Ni	0	33	15	13	15 (FIX)	9 (EQ.)	9
Cu	0	37	15	13	15 (FIX)	13 (FIX)	9

TABLE 40. *Results for the parameter G_2''*

EL.	DIAGON 1	DIAGON 2	DIAGON 3	DIAGON 4	ILS 3	FINAL ILS	FINAL GLS
Ti	0	9	30	13	30 (FIX)	0 (FIX)	9
V	0	13	30	13	21	11	9
Cr	0	17	30	13	30 (FIX)	4	9
Mn	0	21	30	13	24	3	9
Fe	0	25	30	13	30	28	9
Co	0	29	30	13	30 (FIX)	5	9
Ni	0	33	30	13	30 (FIX)	9	9
Cu	0	37	30	13	30 (FIX)	13 (FIX)	9
Zn	0	41	30	13			9

TABLE 41. *Results for the mean error Δ in the ILS.*

EL.	Previous results	ILS 3	FINAL ILS
Ca	22.8	52.7	45.6
Sc	126.4	112.3	108.7
Ti	261.4	129.8	124.0
V	215.8	164.2	161.1
Cr	183.1	137.6	132.4
Mn	169.9	161.2	157.6
Fe	213.4	159.2	152.8
Co	164.2	146.3	140.2
Ni	131.1	112.5	106.3
Cu	12[a]	89.2[b]	82.8[b]

[a] Based on 18 levels.
[b] Based on 23 levels.

TABLE 42. *General interpolation parameters preliminary iterations*

PAR.	INITIAL	GLS 1	GLS 1a	GLS 2	GLS 2a	GLS 2b	GLS 3	GLS 3a	GLS 3b	GLS 3c
D'_0	1610.69	860±31	543±108	773±32	713±67	795±40	948±37	585±79	906±38	974±39
D''_0	28623.08	28482±62	26086±236	28746±81	28674±107	28350±92	28682±81	28723±101	28665±87	28527±*110*
D'_1	3892.39	3641±16	3550±42	3673±16	3614±22	3661±16	3688±16	3487±29	3703±18	3709±16
D''_1	5951.37	5719±21	5278±164	5711±20	5898±44	5738±27	5738±20	5832±80	5731±29	5740±30
D'_2	0	−123±4	−133±5	−128±5	−130±5	−123±5	−117±5	−135±9	−120±5	−115±5
D''_2	0	−123	−238±67	−128	−127±13	−123	−117	−94±12	−120	−115
B_0	769	734±8	745±13	727±6	734±6	731±5	717±5	735±8	712±6	719±5
B'_0	853	863±6	854±7	849±4	849±4	852±3	848±3	840±5	845±4	848±3
B''_0	937	992	963	971	964	973	979	945	978	977
B_1	67	52±2	57±3	53±2	57±2	55±2	52±2	60±3	52±2	52±2
B'_1	61	62±2	59±2	60±2	57±2	59±2	61±2	57±2	61±2	60±2
B''_1	55	72	61	67	57	63	70	54	70	68
C_0	2802	2898±31	2889±50	2954±30	2896±42	2935±25	2988±21	2919±42	2998±25	2988±26
C'_0	3133	3255±28	3258±49	3319±26	3325±22	3306±20	3338±19	3361±37	3340±19	3332±20
C''_0	3464	3612	3627	3684	3754	3677	3688	3803	3682	3676
C_1	244	408±8	402±14	433±7	419±13	429±8	437±8	413±11	429±9	438±9
C'_1	220	304±7	317±12	354±7	353±11	347±9	351±8	357±11	351±9	352±9
C''_1	196	200	232	275	287	265	265	301	273	266
$(G_3)_1$	−1.54	−7.0±2.0	0.1±2.2	2.2±0.7	1.8±1.1	0.3±0.5	0 (FIX)	−0.8±1.1	0.8±0.6	0 (FIX)
$(G'_3)_1$	−0.97	−7.0±1.8	−2.2±2.0	2.2	−2.8±0.8	0.3	0	−3.5±1.2	0.8	0
$(G''_3)_1$	−0.40	−7.0	−4.3	2.2	−7.4	0.3	0	−6.2	0.8	0
$(G_3)_2$	0.44	1.4±2.0	−0.9±0.3	0 (FIX)	−1.2±0.3	1.4±0.2	0 (FIX)	−1.1±0.3	0 (FIX)	0 (FIX)
$(G'_3)_2$	0.20	−1.4	−1.1±0.3	0	−1.0±0.3	−1.4	0	−1.1	0	0
$(G''_3)_2$	−0.04	−1.4	−1.3	0	−0.8	−1.4	0	−1.1	0	0
$(G_{pa})_0$	6295.11	6516±37	6451±76	6406±36	6517±42	6494±33	6469±33	6597±46	6443±34	6515±33
$(G_{pa})_1$	149.80	238±13	244±38	244±12	258±16	292±13	228±9	287±20	230±10	259±11
$(G_{pa})_2$	−10.78	−2±3	−24±5	−33±5	−24±5	−11±4	−18±4	−16±6	−19±4	−15±4
$(G_{da})_0$	1622.01	1654±28	1626±45	1599±28	1647±30	1724±25	1589±22	1641±34	1580±23	1589±24
$(G^*)_0$	2420.66	2362 (F.D.)	3030±425	1985±76	2350±183	2432 (F.D.)	1664±68	4913±365	1707±72	1767±71
$(G_{da})_1$	−39.05	44±8	24±27	−11±7	16±8	6.5±4.2	−10 (FIX)	13±8	−10±5	−5.5±3.3
$(G^*)_1$	−93.02	44	430±86	−11	148±83	179±45	−10	373±94	−10	−5.5
$(G_{da})_2$	10.42	16±3	9.1±4.0	5.0±3.2	12±3	17±3	8.7±2.1	13.5±3.8	6.6±2.5	8.2±2.7
$(G^*)_2$	10.42	16	9	5.0	12	17	8.7	216±70	6.6	8.2
H_0	187.37	102±9	120±9	140±9	136±7	139±7	144±3	122±10	119±7	144±3
H'_0	119.65	68 (F.D.)	−117±47	106 (F.D.)	−54±21	105 (F.D.)	88 (F.D.)	43±54	84 (FIX)	51±12
H_1	−27.53	−14±5	−23±5	−20±2	−21±2	−22±2	−20±2	−21±2	−22±2	−20±2
H'_1	−27.53	−14	−78±11	−20	−48±9	−5.6±4.0	−20	−30±16	−22	−20
H_2	2.61	1.0±0.8	−0.7±0.8	3.2±1.1	1.8±0.9	3.0±0.9	3 (FIX)	−0.1±1.2	−0.2±1.0	3 (FIX)
H'_2	2.61	1.0	−0.7	3.2	1.8	3.0	3	10±6	−0.2	3
J_0	1098.80	1328±48	981±80	1058±47	1047±50	1076±47	1125±39	1021±68	1118±41	1063±42

Table 42. *General Interpolation Parameters Preliminary Iterations   continued.*

PAR.	INITIAL	GLS 1	GLS 1a	GLS 2	GLS 2a	GLS 2b	GLS 3	GLS 3a	GLS 3b	GLS 3c
J_0'	1204.32	1612±67	832±307	1362±62	583±172	1356 (F.D.)	1702±60	784±229	1676±70	1687±72
J_1	−57.21	30±9	29±41	55±9	37±13	45±10	9.3±5.8	−11±18	6.4±5.1	−10±6
J_1'	−31.67	38±12	−322±70	55	−319±54	−46±21	9.3	−308±78	6.4	−10
J_2	6.72	20±5	4±5	59±9	7.6±5.6	6.7±5.1	5.5±3.2	5.2±7.0	5.1±4.7	0.6±4.9
J_2'	6.72	20	4	59	7.6	6.7	5.5	−53±29	5.1	0.6
K_0	2509.57	2485±53	2563±76	2497±50	2451±46	2412±35	2491±39	2336±73	2449±41	2437±42
K_0'	3190.04	3506±88	2788±213	3288±73	2839±110	3432 (F.D.)	3489±58	2696±190	3426±62	3525±61
K_1	−147.70	30±7	−46±37	−18±7	−44±13	−59±10	−57±10	−96±22	−53±11	−85±11
K_1'	−86.74	34±11	−185±98	−18	−138±36	39±14	−57	−193±75	−53	17±7
K_2	19.15	26±4	34±5	18±5	17±5	11±4	13±4	7.6±6.9	9.0±4.2	9.4±4.3
K_2'	22.29	26	−54±11	18	−31±9	11	13	−43±19	9.0	9.4
$(F_1)_0$	205.11	200±7	175±10	176±5	173±5	181±4	188±4	181±5	189±5	187±4
$(F_1')_0$	275.06	299±4	292.5±6	289±4	288±4	299±3	299±3	289±4	301±3	301±3
$(F_1'')_0$	345.01	398	410	402	403	417	410	397	413	415
$(F_1)_1$	−13.65	5.9±2.1	5.2±4.0	5.7±1.9	7.3±1.4	7.6±1.6	1.9±1.0	6.4±2.0	7.5±1.7	3.7±1.1
$(F_1')_1$	−7.68	4.1±1.5	2.9±3.1	2.3±1.4	−2.9±1.3	1.7±1.2	1.9	−3.4±2.1	−0.4±1.1	3.7
$(F_1'')_1$	−1.71	2.3	0.6	−1.1	−13.1	−4.2	1.9	−13.2	−8.3	3.7
$(F_1)_2 = (F_1')_2 =$										
$(F_1'')_2$	0	0 (FIX)	0±0.4	0 (FIX)	−2.1±1.6	0 (FIX)	0 (FIX)	0 (FIX)	0 (FIX)	0 (FIX)
$(G_1)_0$	254.21	205±7	241±9	225±4	237±5	229±4	233±4	233±6	229±4	230±4
$(G_1')_0$	267.65	254±5	265±7	259±3	271±4	261±4	260±3	259±4	258±3	262±3
$(G_1'')_0$	281.09	303	289	293	305	293	287	285	287	294
$(G_1)_1$	−25.98	−5±2	−17±5	−19±2	−15±2	−17±2	−19±1	−17±2	−16±1	−18±1
$(G_1')_1$	−15.07	−19±2	−16±5	−14±2	−16±2	−19±1	−19	−17±2	−16	−18
$(G_1'')_1$	−4.16	−33	−15	−9	−17	−21	−19	−17	−16	−18
$(G_1)_2$	7.83	0 (FIX)	2.5±0.6	0 (FIX)	1.8±0.6	1.1±0.4	0 (FIX)	0.9±0.3	0 (FIX)	0 (FIX)
$(G_1')_2$	3.54	0	0.6±0.7	0	1.2±0.5	1.1	0	0 (FIX)	0	0
$(G_1'')_2$	−0.75	0	−1.3	0	0.6	1.1	0	0	0	0
$(G_3)_0$	16.22	26±3	4.1±5.9	10±1	−2.0±3.2	2.5±1.7	15±2	1.9±3.1	13±2	14±2
$(G_3')_0$	18.96	32±3	11±3	10	11±3	2.5	23±1	11±3	24±2	24±1
$(G_3'')_0$	21.70	38±2	18	10	24	2.5	31	20	35	34
$(\zeta_4)_0$	356.97	360±39	337±45	331±26	352±24	339±21	322±24	354±31	326±26	332±25
$(\zeta_4')_0$	401.94	421±20	367±19	369±18	368±15	370±14	358±16	363±21	372±17	368±17
$(\zeta_4'')_0$	446.92	482	397	407	384	401	394	372	418	404
$(\zeta_4)_1$	86.24	70±12	73±26	78±10	81±8	76±8	84±8	87±11	82±9	85±9
$(\zeta_4')_1$	92.53	79±11	96±15	89±6	90±4	89±5	87±5	90±6	91±6	90±6
$(\zeta_4'')_1$	98.81	88	119	100	99	102	90	93	100	95
$(\zeta_4)_2$	8.14	2.3±1.8	2.2±3.1	2.7±1.6	6.9±2.9	2.8±1.9	3.7±1.4	8.5±4.0	4.8±2.4	4.9±2.4
$(\zeta_4')_2$	8.14	2.3	3.3±2.3	2.7	2.3±2.3	2.8	3.7	2.5±3.0	4.8	4.9

66

PAR.	INITIAL	GLS 1	GLS 1a	GLS 2	GLS 2a	GLS 2b	GLS 3	GLS 3a	GLS 3b	GLS 3c
$(\zeta'')_2$	8.14	2.3	4.4	2.7	-2.3	2.8	3.7	-3.5	4.8	4.9
$(\zeta_\rho)_0$	89.03	190±61	127±68	129±49	123±39	103±38	142±38	123±51	137±40	125±41
$(\zeta)_0$	199.19	252±39	281±39	284±38	273±29	279±30	298±30	280±38	298±32	305±32
$(\zeta'')_0$	309.34	314	435	439	423	455	454	437	459	485
$(\zeta_\rho)_1$	8.67	31±11	37±35	19±10	20±14	16±9	11±8	18±16	9.5±7.1	10±7
$(\zeta)_1$	24.32	31	21±20	23±9	24±11	23±10	11	22±14	19±8	10
$(\zeta'')_1$	39.97	31	5	27	28	30	11	26	29	10
$(\zeta_\rho)_2$	0	0 (FIX)	8.3±4.3	5.0±4.0	5.0±4.3	3.2±2.8	6.2±2.7	7.2±5.6	4.0±2.6	4.6±2.5
$(\zeta)_2$	0	0	3.9±3.2	5.0	2.4±3.2	3.2	6.2	4.0±4.2	4.0	4.6
$(\zeta'')_2$	0	0	0	5.0	-0.2	3.2	6.2	0.8	4.0	4.6
α_0	60	47±5	42±10	39±3	45±6	42±3	36±3	40±7	37±3	37±3
α_0'	60	47	48±6	39	39±3	42	36	35±4	37	37
α_0''	60	47	54	39	33	42	36	30	37	37
α_1	0	7.6±1.2	8.5±2.0	1.3±1.1	2.0±1.9	1.2±1.1	1.7±1.1	4.4±2.5	2.8±1.0	1.1±1.2
α_1'	0	7.6	5.8±1.7	1.3	1.1±1.6	1.2	1.7	1.4±2.2	2.8	1.1
α_1''	0	7.6	3.1	1.3	0.2	1.2	1.7	-1.6	2.8	1.1
β_0	0	-253±41	-247±196	-381±66	-299±100	-321±52	-476±48	-280±115	-476±51	-460±50
β_0'	0	-253	-301±138	-381	-365±57	-321	-476	-507±73	-476	-460
β_0''	0	-253	-355	-381	-431	-321	-476	-734	-476	-460
$\beta_1=\beta_1'=\beta_1''$	0	0 (FIX)	-28±39	-122±43	-121±38	-115±34	-42±22	-24±39	-21±16	-53±19
T_0	0	-2.9±0.6	-2.2±0.9	-3.1±0.3	-2.3±0.5	-2.7±0.4	~3.3±0.3	-3.2±0.7	-3.2±0.3	-3.3±0.3
T_0'	0	-2.9	-1.7±0.6	-3.1	-3.4±0.3	-3.2±0.3	~3.3	-3.5±0.4	-3.2	-3.3
T_0''	0	-2.9	-1.2	-3.1	-4.5	-3.7	-3.3	-3.8	-3.2	-3.3
$T_1=T_1'=T_1''$	0	0.4±0.2	0.3±0.2	-0.4±0.2	-0.3±0.2	-0.3±0.2	~0.4±0.2	-0.2±0.2	-0.3±0.2	-0.4±0.2
$(T)_0$	0	-3.1±0.9	-4.7±1.2	-2.6±0.6	-3.1±0.8	-2.4±0.5	~2.8±0.6	-3.5±1.0	-2.6±0.6	-2.7±0.6
$(T')_0$	0	-3.1	-3.2±0.9	-2.6	-2.1±0.6	-2.4	-2.8	-1.9±0.7	-2.6	-2.7
$(T'')_0$	0	-3.1	-1.7	-2.6	-1.1	-2.4	-2.8	-0.3	-2.6	-2.7
$(T)_1=(T')_1=(T'')_1$	0	-0.3±0.2	-0.3±0.2	-0.3±0.2	-0.3±0.2	-0.3±0.2	0 (FIX)	0.1±0.3	0.1±0.3	0 (FIX)
$(F)_0$	0	-12±4	-40±11	-35±5	-38±9	-18±5	-6.2±4.1	-28±8	-8.1±3.7	-7.0±3.8
$(F')_0$	0	-12	-18±3	-35	-33±8	-18	-6.2	-27±9	-8.1	-7.0
$(F'')_0$	0	-12	4	-35	-28	-18	-6.2	-26	-8.1	-7.0
$(F_1)_1$	0	0 (FIX)	2.8±3.8	2.6±1.9	7.8±4.2	3.1±2.3	2 (FIX)	-1.6±2.8	-5.0±1.6	-3.9±2.0
$(F')_1$	0	0	-7.6±2.6	2.6	-9.1±2.8	3.1	2	-12±4	-5.0	-3.9
$(F'')_1$	0	0	-18	2.6	-26	3.1	2	-22	-5.0	-3.9
$(G)_0$	0	21±5	-4.2±10	-6.6±5.1	-7.4±5.3	2.8±4.0	13±3	-5.4±6.9	3.2±3.0	5.9±3.1
$(G')_0$	0	21	29±9	-6.6	25.2±4.8	23±4	13	19±6	3.2	5.9
$(G'')_0$	0	21	62	-6.6	57.8	43	13	43	2.3	5.9
$(G)_1$	0	4.7±1.2	-0.2±2.1	0.8±1.1	4.3±2.2	2.0±1.6	0 (FIX)	-0.4±1.7	1.3±1.1	0 (FIX)
$(G')_1$	0	4.7	6.0±2.0	0.8	-0.1±2.2	2.0	0	-1.9±1.8	1.3	0
$(G'')_1$	0	4.7	12	0.8	-4.5	2.0	0	-3.4	1.3	0

TABLE 42. *General Interpolation Parameters Preliminary Iterations   continued.*

PAR.	INITIAL	GLS 1	GLS 1a	GLS 2	GLS 2a	GLS 2b	GLS 3	GLS 3a	GLS 3b	GLS 3c
M_0	13020	13030 ± 159	13049 ± 170	13051 ± 180	13048 ± 179	13037 ± 161	13073 ± 126	13065 ± 172	13060 ± 131	13061 ± 131
M_1	35800	37641 ± 96	37643 ± 119	37840 ± 98	37580 ± 127	37600 ± 89	37651 ± 88	37700 ± 174	37682 ± 94	37655 ± 95
M_2	35070	35281 ± 86	35278 ± 109	35332 ± 86	35252 ± 120	35298 ± 80	35333 ± 83	35304 ± 96	35339 ± 87	35344 ± 86
M_3	38690	38515 ± 84	38486 ± 108	38552 ± 84	38504 ± 116	38571 ± 80	38521 ± 82	38467 ± 96	38514 ± 84	38545 ± 85
M_4	43540	43757 ± 88	43807 ± 102	43798 ± 82	43796 ± 110	43863 ± 80	43724 ± 83	43721 ± 98	43722 ± 87	43748 ± 86
M_5	59160	58916 ± 83	59013 ± 90	58888 ± 79	58911 ± 100	58978 ± 77	58794 ± 78	58878 ± 96	58798 ± 86	58820 ± 84
M_6	70910	67536 ± 78	67669 ± 77	67465 ± 72	67504 ± 87	67551 ± 71	67382 ± 70	67607 ± 92	67390 ± 75	67400 ± 68
M_7	51750	51974 ± 67	52177 ± 61	51869 ± 74	51934 ± 69	51930 ± 60	51779 ± 58	51936 ± 79	51792 ± 67	51786 ± 65
M_8	42130	42285 ± 51	42548 ± 42	42158 ± 54	42230 ± 47	42183 ± 43	42071 ± 47	42481 ± 69	42085 ± 48	42065 ± 49
M_9	31370	31914 ± 73	32269 ± 103	31784 ± 88	31865 ± 113	31766 ± 67	31639 ± 71	31970 ± 102	31642 ± 74	31618 ± 75
M_{10}	30700	32238 ± 75	32681 ± 65	32118 ± 84	32189 ± 71	32039 ± 68	31857 ± 70	32227 ± 91	31846 ± 78	31812 ± 77
M_{11}''	94502	92090 ± 183	91674 ± 310	92033 ± 191	92430 ± 239	91861 ± 164	92261 ± 120	92580 ± 118	92142 ± 128	92166 ± 127
M_{12}''	551	570 ± 160	527 ± 173	597 ± 180	561 ± 181	594 ± 162	583 ± 108	566 ± 171	598 ± 116	591 ± 119
Δ		205.0	193.9	200.3	188.2	203.1	196.7	187.8	198.1	194.5

TABLE 43. *General Interpolation Parameters Final Iteration*

PAR.	GLS 4	GLS 4a	GLS 4b	GLS 4c	GLS 4d	GLS 4e	GLS 4f	GLS 4g	GLS 4h	GLS 4i
D_0'	963 ± 31	612 ± 105	920 ± 36	895 ± 38	929 ± 36	965 ± 32	968 ± 33	1007 ± 34	1039 ± 35	$1041\pm;37$
D_0''	28770 ± 72	28599 ± 98	28689 ± 82	28745 ± 81	28683 ± 74	2873 ± 77	28739 ± 78	28784 ± 77	28704 ± 79	28703 ± 83
D_1'	3701 ± 15	3592 ± 46	3691 ± 17	3690 ± 17	3699 ± 17	3706 ± 17	3706 ± 16	3723 ± 17	3720 ± 19	3724 ± 19
D_1''	5784 ± 32	5818 ± 58	5736 ± 35	5754 ± 32	5748 ± 31	5791 ± 33	5794 ± 32	5814 ± 31	5813 ± 33	5822 ± 35
D_2'	-114 ± 4	-126 ± 10	-116 ± 5	-114 ± 5	-116 ± 5	-115 ± 5	-117 ± 5	$\pm120\pm5$	-120 ± 6	-121 ± 6
D_2''	-114	-100 ± 15	-116	-114	-116	-115	-117	-120	-120	-121
B_0	714 ± 4	727 ± 7	714 ± 5	716 ± 5	714 ± 5	712 ± 5	700 ± 4	705 ± 4	714 ± 4	712 ± 4
B_0'	844 ± 3	839 ± 4	842 ± 3	841 ± 3	842 ± 3	844 ± 3	839 ± 3	843 ± 3	855 ± 3	855 ± 3
B_0''	974	951	970	966	970	976	978	981	996	998
B_1	51 ± 2	57 ± 3	51 ± 2	51 ± 2	51 ± 2	51 ± 2	45 ± 2	51 ± 2	52 ± 2	52 ± 2
B_1'	60 ± 1	57 ± 2	60 ± 2	58 ± 2	59 ± 2	61 ± 2	55 ± 2	63 ± 1	63 ± 1	63 ± 1
B_1''	69	57	69	65	67	71	65	75	74	74
C_0	3013 ± 19	2938 ± 46	3005 ± 27	3010 ± 26	3005 ± 22	3016 ± 22	3099 ± 20	2916 ± 15	2839 ± 15	2843 ± 13
C_0'	3362 ± 16	3376 ± 26	3359 ± 23	3364 ± 23	3359 ± 21	3361 ± 18	3433 ± 17	3267 ± 12	3190 ± 10	3191 ± 9
C_0''	3711	3814	3713	3718	3713	3706	3767	3618	3541	3539
C_1	433 ± 7	426 ± 15	441 ± 11	448 ± 11	440 ± 8	437 ± 8	450 ± 8	413 ± 7	412 ± 7	414 ± 8
C_1'	348 ± 7	362 ± 13	358 ± 10	362 ± 10	358 ± 8	350 ± 8	362 ± 8	322 ± 7	319 ± 6	320 ± 6
C_1''	263	298	275	276	276	263	274	231	226	226
$(F_2)_0$	183 ± 3	190 ± 5	187 ± 4	186 ± 4	186 ± 4	180 ± 4	179 ± 4	180 ± 4	181 ± 5	178 ± 4
$(F_2')_0$	299 ± 3	293 ± 4	298 ± 3	290 ± 3	298 ± 3	299 ± 3	298 ± 3	298 ± 3	297 ± 4	297 ± 3
$(F_2'')_0$	415	396	410	394	410	418	417	416	413	416
$(F_2)_1$	2.5 ± 1.1	7.7 ± 1.8	2.5 ± 1	2.0 ± 1.0	2.1 ± 1.1	2.3 ± 1.1	2.2 ± 1.1	2.3 ± 1.1	2.5 ± 1.2	2.4 ± 1.2
$(F_2')_1$	2.5	-2.2 ± 2.0	2.5	2.0	2.1	2.3	2.2	2.3	2.5	2.4
$(F''_2)_1$	2.5	-12.1	2.5	2.0	2.1	2.3	2.2	2.3	2.5	2.4
$(F_2)_2 = (F_2')_2 =$										
$(F_2'')_2$	0 (FIX)	0 (FIX)	0 (FIX)	0 (FIX)	0 (FIX)	0 (FIX)	0 (FIX)	0 (FIX)	0 (FIX)	0 (FIX)
$(G_1)_0$	235 ± 3	236 ± 5	233 ± 4	237 ± 4	234 ± 4	228 ± 4	229 ± 4	230 ± 4	237 ± 4	231 ± 4
$(G_1')_0$	261 ± 3	260 ± 4	251 ± 3	259 ± 3	261 ± 3	256 ± 3	256 ± 3	256 ± 3	260 ± 3	256 ± 3
$(G_1'')_0$	287	284	269	281	288	284	283	282	283	281
$(G_1)_1$	-19 ± 1	-17 ± 2	-19 ± 1	-18 ± 1	-17 ± 1	-19 ± 1	-18 ± 1	-18 ± 1	-19 ± 1	-19 ± 1
$(G_1')_1$	-19	-18 ± 2	-19	-18	-17	-19	-18	-18	-19	-19
$(G_1'')_1$	-19	-19	-19	-18	-17	-19	-18	-18	-19	-19
$(G_1)_2$	0 (FIX)	0.7 ± 0.3	0 (FIX)	0 (FIX)	0 (FIX)	0 (FIX)	0 (FIX)	0 (FIX)	0 (FIX)	0 (FIX)
$(G_1')_2$	0	0.7	0	0	0	0	0	0	0	0
$(G_1'')_2$	0	0.7	0	0	0	0	0	0	0	0
$(G_3)_0$	13 ± 1	4.6 ± 2.8	14 ± 2	14 ± 1	14 ± 1	13 ± 1	13 ± 1	13 ± 1	14 ± 2	14 ± 1
$(G_3')_0$	21 ± 1	11.1 ± 2.6	22 ± 1	19 ± 2	22 ± 1	22 ± 1	22 ± 1	22 ± 1	20 ± 2	21 ± 1
$(G_3'')_0$	29	17.6	30	24	30	31	31	31	26	28
$(G_3)_1$	0 (FIX)	-0.5 ± 1.0	0 (FIX)	0 (FIX)	0 (FIX)	0 (FIX)	0 (FIX)	0 (FIX)	0 (FIX)	0 (FIX)
$(G_3')_1$	0	-3.5 ± 1.1	0	0	0	0	0	0	0	0

TABLE 43. *General Interpolation Parameters Final Iteration—continued*

PAR.	GLS 4	GLS 4a	GLS 4b	GLS 4c	GLS 4d	GLS 4e	GLS 4f	GLS 4g	GLS 4h	GLS 4i
$(G_s^*)_1$	0	−6.5	0	0	0	0	0	0	0	0
$(G_s)_2$	0 (FIX)	−1.1±0.2	0 (FIX)	0 (FIX)	0 (FIX)	0 (FIX)	0 (FIX)	0 (FIX)	0 (FIX)	0 (FIX)
$(G_s)_2$	0	−1.1	0	0	0	0	0	0	0	0
$(G_s^*)_2$	0	−1.1	0	0	0	0	0	0	0	0
$(G_\rho)_0$	6454±20	6601±49	6447±34	6555±37	6445±23	6446±23	6463±22	6462±23	6458±24	6469±26
$(G_\rho)_1$	220±10	257±18	222±10	280±13	221±10	222±10	221±10	219±11	217±11	219±11
$(G_\rho)_2$	−15±3	−4.6±5.6	−14±4	−10±4	−15±4	−17±4	−17±4	−19±4	−14±4	−16±4
$(G_\mu)_0$	1595±14	1660±30	1593±26	1587±25	1592±15	1579±17	1583±15	1584±15	1616±16	1601±18
$(G^*)_0$	1826 (F.D.)	3107±801	1649±83	1674±81	1829 (F.D.)	1810 (F.D.)	1814 (F.D.)	1815 (F.D.)	1847 (F.D.)	1832 (F.D.)
$(G_\mu)_1$	0 (FIX)	14±9	0 (FIX)	0 (FIX)	0 (FIX)	0 (FIX)	0 (FIX)	0 (FIX)	0 (FIX)	0 (FIX)
$(G^*)_1$	0	544±241	0	0	0	0	0	0	0	0
$(G_\mu)_2$	8.7±1.2	15±3	8.1±2.9	6.8±2.6	8.1±1.9	9.0±1.7	9.0±1.6	8.7±1.8	9.0±1.7	9.2±1.8
$(G^*)_2$	8.7	82±62	8.1	6.8	8.1	9.0	9.0	8.7	9.0	9.2
H_0	123±2	114±8	123±3	124±3	123±2	124±3	124±3	122±3	123±3	124±3
H_0'	84 (FIX)	88±47	84 (FIX)	42±15	84 (FIX)	84 (FIX)	84 (FIX)	84 (FIX)	84 (FIX)	84 (FIX)
H_1	−23±1	−24±2	−22±1	−23±1	−22±1	−23±1	−24±1	−24±1	−24±1	−24±2
H_1'	−23	−11±14	−22	−23	−22	−23	−24	−24	−24	−24
H_2	0 (FIX)	−1.3±1.1	0 (FIX)	0 (FIX)	0 (FIX)	0 (FIX)	0 (FIX)	0 (FIX)	0 (FIX)	0 (FIX)
H_2'	0	9.0±5.0	0	0	0	0	0	0	0	0
J_0	1122±21	1090±61	1134±47	1039±41	1123±25	1126±25	1127±24	1118±23	1126±24	1129±26
J_0'	1652±38	598±271	1693±70	975±74	1666±39	1695±36	1708±37	1685±40	1645±42	1689±43
J_1	26±3	−15±17	14±9	−2.8±1.9	19±6	11±6	15±5	−1.8±5.2	7.0±5.3	6.7±5.2
J_1'	26	−376±127	14	−231±37	19	11	15	−1.8	7.0	6.7
J_2	8.4±2.1	8.2±6.2	7.2±5	−1.4±3.1	6.1±3.0	6.9±3.2	7.5±3.1	5.0±3.2	8.1±3.1	6.6±4.2
J_2'	8.4	−32±27	7.2	−1.4	6.1	6.9	7.5	5.0	8.1	6.6
K_0	2474±29	2404±66	2508±36	2468±40	2485±28	2468±29	2467±28	2445±29	2476±30	2470±31
K_0'	3421±38	2984±209	3446±46	3182±51	3426±34	3463±38	3467±39	3435±37	3426±39	3467±41
K_1	−62±6	−93±20	−54±13	−76±13	−60±7	−61±8	−61±7	−67±7	−63±8	−62±9
K_1'	−62	−167±70	−54	−76	−60	−61	−61	−67	−63	−62
K_2	15±3	12±6	17±5	14±5	14±3	12±4	13±4	7.9±4.1	15±5	12±5
K_2'	15	−5.6±19	17	14	14	12	13	7.9	15	12
$(\zeta_d)_0$	310±19	343±28	319±25	327±24	323±18	311±17	308±18	315±19	314±20	315±21
$(\zeta_d)_0$	364±16	360±19	355±17	367±17	359±14	365±15	367±16	365±17	366±16	366±19
$(\zeta_d)_0$	418	377	391	407	395	419	426	415	418	417
$(\zeta_d)_1$	78±5	82±10	82±9	84±9	82±6	77±7	77±8	78±8	77±9	76±10
$(\zeta_d)_1$	91±4	90±5	87±6	89±5	87±5	90±6	91±6	92±6	92±6	91±6
$(\zeta_d)_1$	104	98	92	94	92	103	105	106	107	106
$(\zeta_d)_2$	2.6±1.2	5.5±3.5	1.9±1.6	3.2±2.1	2.4±1.5	2.5±1.8	2.3±1.7	3.0±1.5	2.8±1.6	2.7±1.4
$(\zeta_d)_2$	2.6	1.8±2.8	1.9	3.2	2.4	2.5	2.3	3.0	2.8	2.7
$(\zeta_d)_2$	2.6	−1.9	1.9	3.2	2.4	2.5	2.3	3.0	2.8	2.7

TABLE 43. *General Interpolation Parameters Final Iteration—continued*

PAR.	GLS 4	GLS 4a	GLS 4b	GLS 4c	GLS 4d	GLS 4e	GLS 4f	GLS 4g	GLS 4h	GLS 4i
$(\zeta_p)_0$	143±26	139±46	149±43	135±42	150±26	137±28	140±25	135±27	130±29	123±30
$(\zeta'_p)_0$	284±19	278±34	290±34	279±33	290±19	269±24	269±21	269±22	279±25	264±27
$(\zeta''_p)_0$	425	417	431	423	430	401	398	403	428	405
$(\zeta_p)_1$	13±5	17±17	15±8	13±8	18±6	14±6	15±5	16±6	15±7	16±7
$(\zeta'_p)_1$	13	19±13	15	13	18	14	15	16	15	16
$(\zeta''_p)_1$	13	21	15	13	18	14	15	16	15	16
$(\zeta_p)_2$	5.2±1.7	5.4±5.0	5.6±3.2	5.3±3.0	5.6±1.9	5.3±2.1	5.5±2.0	5.4±1.9	5.4±2.0	5.5±2.1
$(\zeta'_p)_2$	5.2	3.6±3.8	5.6	5.3	5.6	5.3	5.5	5.4	5.4	5.5
$(\zeta''_p)_2$	5.2	1.8	5.6	5.3	5.6	5.3	5.5	5.4	5.4	5.5
α_0	33±2	42±6	34±3	33±3	34±3	34±3	30±3	61±2	71±1	72±1
α'_0	33	32±3	34	33	34	34	30	61	71	72
α''_0	33	22	34	33	34	34	30	61	71	72
α_1	3.7±0.9	3.0±2.3	1.6±1.2	0.7±1.1	1.5±1.2	3.1±1.1	2.9±1.2	6.8±1.1	7.1±0.8	6.9±0.7
α'_1	3.7	0.8±2.0	1.6	0.7	1.5	3.1	2.9	6.8	7.1	6.9
α''_1	3.7	−1.4	1.6	0.7	1.5	3.1	2.9	6.8	7.1	6.9
β_0	−527±28	−380±102	−523±37	−534±38	−526±31	−521±30	−673±31	−336±32	0 (FIX)	0 (FIX)
β'_0	−527	−549±65	−523	−534	−526	−521	−673	−336	0	0
β''_0	−527	−718	−523	−534	−526	−521	−673	−336	0	0
$\beta_1=\beta'_1=\beta''_1$	−23±11	−59±44	−65±29	−89±31	−40±13	−22±12	−51±18	−7.2±8.1	0 (FIX)	0 (FIX)
T_0	−3.5±0.3	−2.6±0.7	−3.4±0.3	−3.5±0.3	−3.4±0.3	−3.5±0.3	−3.8±0.3	0 (FIX)	0 (FIX)	0 (FIX)
T'_0	−3.5	−3.7±0.4	−3.4	−3.5	−3.4	−3.5	−3.8	0	0	0
T''_0	−3.5	−4.8	−3.4	−3.5	−3.4	−3.5	−3.8	0	0	0
$T_1=T'_1=T''_1$	−0.2±0.1	−0.3±0.2	−0.4±0.2	−0.4±0.2	−0.4±0.2	−0.3±0.2	−0.3±0.2	0 (FIX)	0 (FIX)	0 (FIX)
$(T_x)_0$	−2.6±0.6	−3.1±0.9	−2.6±0.6	−2.5±0.5	−2.6±0.5	−2.7±0.6	0 (FIX)	0 (FIX)	0 (FIX)	0 (FIX)
$(T'_x)_0$	−2.6	−2.0±0.6	−2.6	−2.5	−2.6	−2.7	0	0	0	0
$(T''_x)_0$	−2.6	−0.9	−2.6	−2.5	−2.6	−2.7	0	0	0	0
$(T_x)_1 =$										
$(T'_x)_1=(T''_x)_1$	0 (FIX)	0.1±0.3	0 (FIX)	0 (FIX)	0 (FIX)	0 (FIX)	0 (FIX)	0 (FIX)	0 (FIX)	0 (FIX)
$(F_x)_0$	−13±3	−14±9	−8.6±4.1	−21±5	−11±4	0 (FIX)	0 (FIX)	0 (FIX)	−13±5	0 (FIX)
$(F'_x)_0$	−13	−26±9	−8.6	−21	−11	0	0	0	−13	0
$(F''_x)_0$	−13	−38	−8.6	−21	−11	0	0	0	−13	0
$(F_x)_1$	0 (FIX)	−1.9±4.3	0 (FIX)	−4.0±2.1	0 (FIX)	0 (FIX)	0 (FIX)	0 (FIX)	0 (FIX)	0 (FIX)
$(F'_x)_1$	0	−11±4	0	−4.0	0	0	0	0	0	0
$(F''_x)_1$	0	−20	0	−4.0	0	0	0	0	0	0
$(G_x)_0$	9±3	4.2±4.0	12±3	10±3	12±3	0 (FIX)	0 (FIX)	0 (FIX)	9±4	0 (FIX)
$(G'_x)_0$	9	19±4	12	10	12	0	0	0	9	0
$(G''_x)_0$	9	34	12	10	12	0	0	0	9	0
$(G_x)_1$	0 (FIX)	0.6±2.4	0 (FIX)	0 (FIX)	0 (FIX)	0 (FIX)	0 (FIX)	0 (FIX)	0 (FIX)	0 (FIX)
$(G'_x)_1$	0	−2.0±2.9	0	0	0	0	0	0	0	0
$(G''_x)_1$	0	−4.6	0	0	0	0	0	0	0	0

TABLE 43. *General Interpolation Parameters Final Iteration—continued*

PAR.	GLS 4	GLS 4a	GLS 4b	GLS 4c	GLS 4d	GLS 4e	GLS 4f	GLS 4g	GLS 4h	GLS 4i
M_0	13065 ± 99	13057 ± 156	13066 ± 113	13062 ± 109	13060 ± 101	13062 ± 103	13061 ± 101	13060 ± 102	13059 ± 107	13057 ± 106
M_1	37650 ± 69	37629 ± 161	37646 ± 79	37644 ± 80	31647 ± 66	37652 ± 68	37658 ± 66	37681 ± 68	37678 ± 71	37681 ± 72
M_2	35331 ± 72	35300 ± 151	35337 ± 81	35332 ± 76	35339 ± 71	35332 ± 72	35343 ± 71	35356 ± 73	35365 ± 76	39230 ± 75
M_3	38522 ± 70	38524 ± 136	38533 ± 80	38539 ± 79	38533 ± 69	38524 ± 68	38544 ± 69	38538 ± 70	38566 ± 73	38569 ± 74
M_4	43725 ± 71	43773 ± 105	43738 ± 80	43750 ± 76	43732 ± 70	43722 ± 72	43742 ± 71	43719 ± 73	43771 ± 75	43770 ± 76
M_5	58792 ± 59	58887 ± 126	58808 ± 75	58819 ± 72	58798 ± 58	58789 ± 64	58786 ± 63	58764 ± 62	58837 ± 63	58836 ± 63
M_6	67352 ± 54	67506 ± 98	67367 ± 69	67364 ± 68	67355 ± 56	67347 ± 57	67330 ± 56	67302 ± 57	67379 ± 56	67374 ± 55
M_7	51757 ± 49	52013 ± 91	51772 ± 67	51761 ± 61	51756 ± 51	51754 ± 52	51720 ± 51	51701 ± 52	51789 ± 52	51783 ± 53
M_8	42032 ± 47	42398 ± 88	42063 ± 50	42042 ± 48	42042 ± 42	42027 ± 48	41994 ± 49	41961 ± 47	42016 ± 48	42008 ± 48
M_9	31585 ± 55	32095 ± 102	31644 ± 78	31619 ± 74	31614 ± 56	31580 ± 59	31558 ± 60	31493 ± 59	31511 ± 61	31504 ± 61
M_{10}	31759 ± 59	32429 ± 91	31868 ± 79	31829 ± 76	31823 ± 60	31752 ± 62	31758 ± 63	31652 ± 61	31653 ± 63	31646 ± 64
M''_{11}	92641 ± 94	92492 ± 146	92289 ± 103	92306 ± 102	93322 ± 96	92668 ± 102	92631 ± 101	92697 ± 103	92640 ± 101	92664 ± 99
M''_{12}	572 ± 102	559 ± 153	579 ± 105	573 ± 103	583 ± 101	569 ± 104	571 ± 103	573 ± 102	577 ± 104	574 ± 102
Δ	182.2	171.3	180.2	176.1	199.7	183.5	189.1	203.7	210.8	211.4
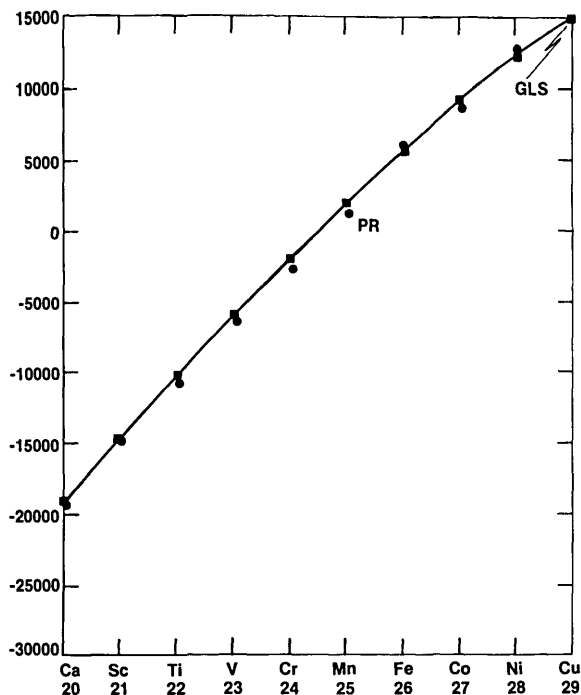
FIGURE 1. Plots of D' versus atomic number.
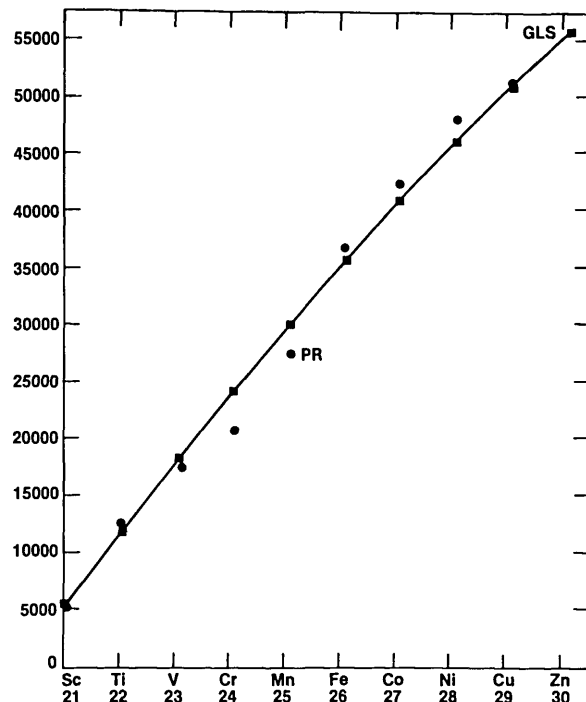
FIGURE 2. Plots of D'' versus atomic number.
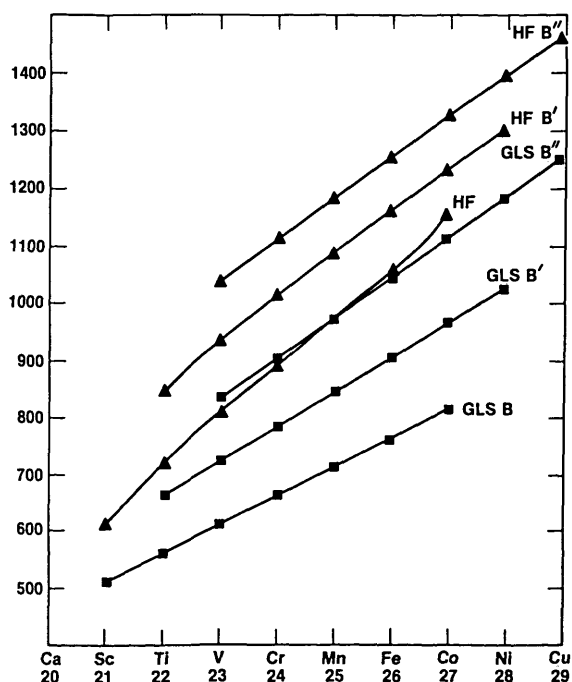
FIGURE 3. Plots of B, B', and B'' versus atomic number.
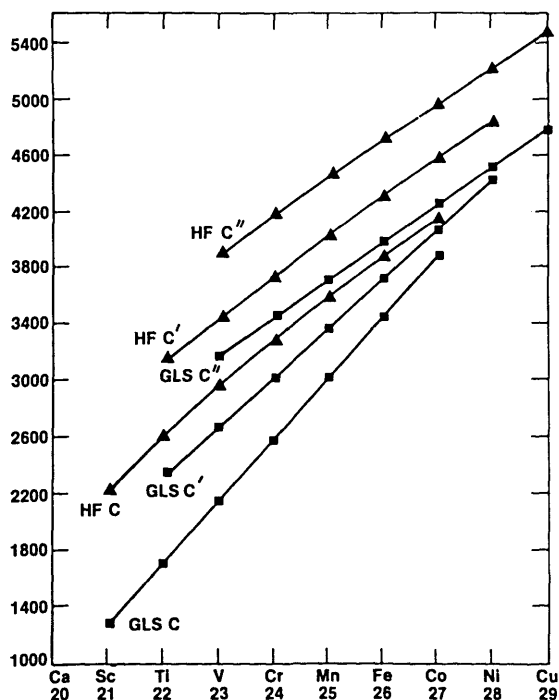
FIGURE 4. Plots of C, C', and C'' versus atomic number.

73
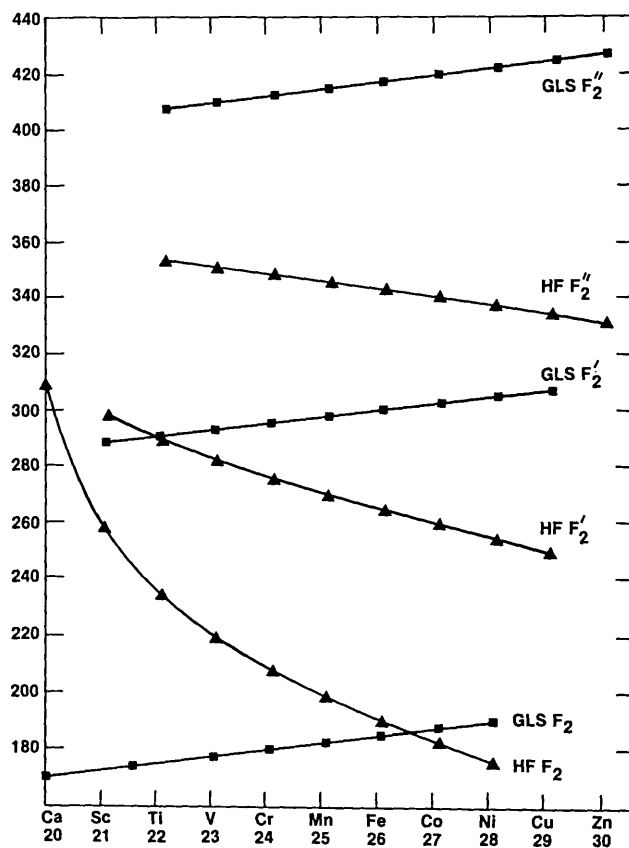
FIGURE 5. Plots of F_2, F_2', and F_2'' versus atomic number.
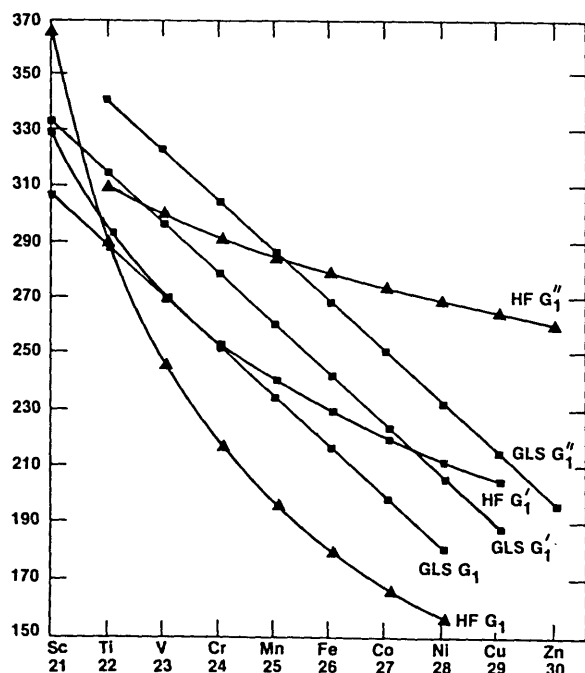
FIGURE 6. Plots of G_1, G_1', and G_1'' versus atomic number.
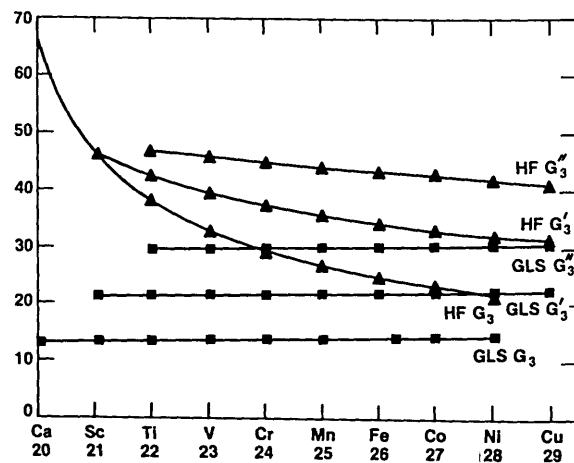
FIGURE 7. Plots of G_3, G_3', and G_3'' versus atomic number.

74
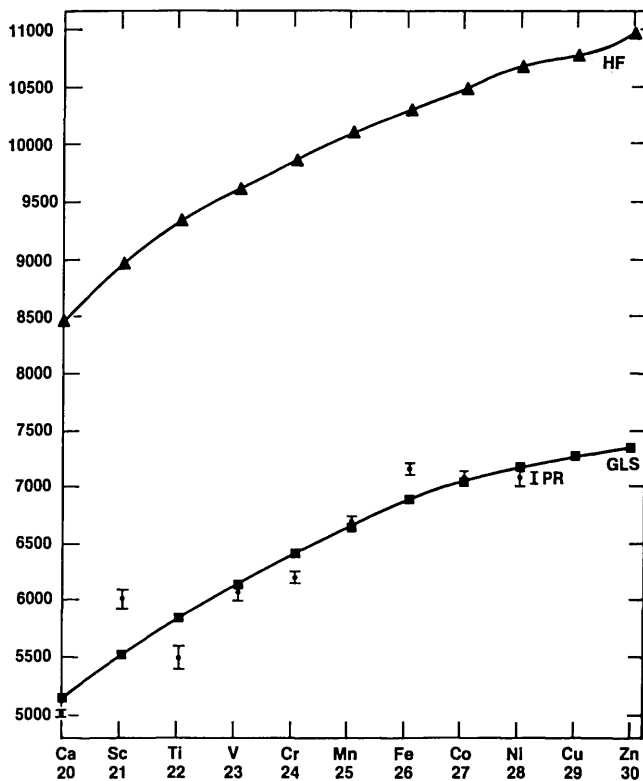
FIGURE 8. Plots of G_{ps} versus atomic number.
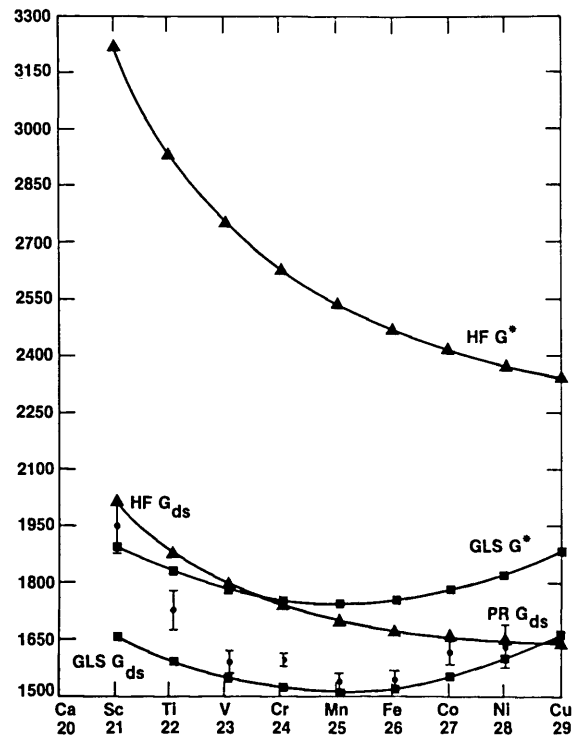
FIGURE 9. Plots of G_{ds} and G^* versus atomic number.
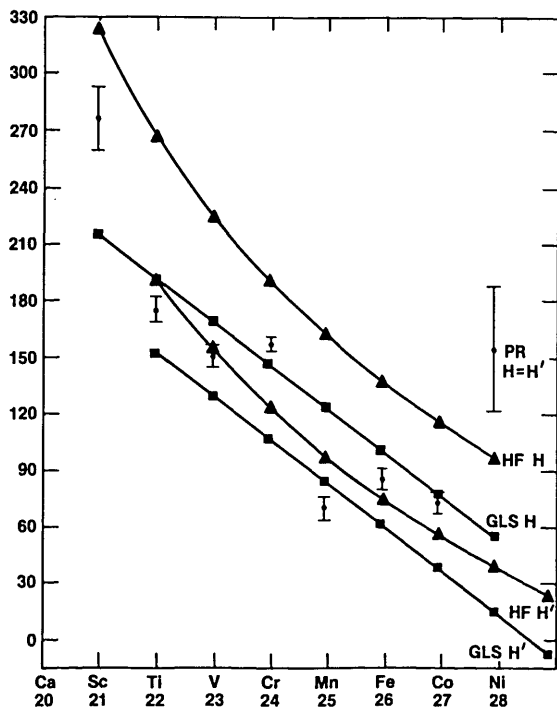
FIGURE 10. Plots of H and H' versus atomic number.
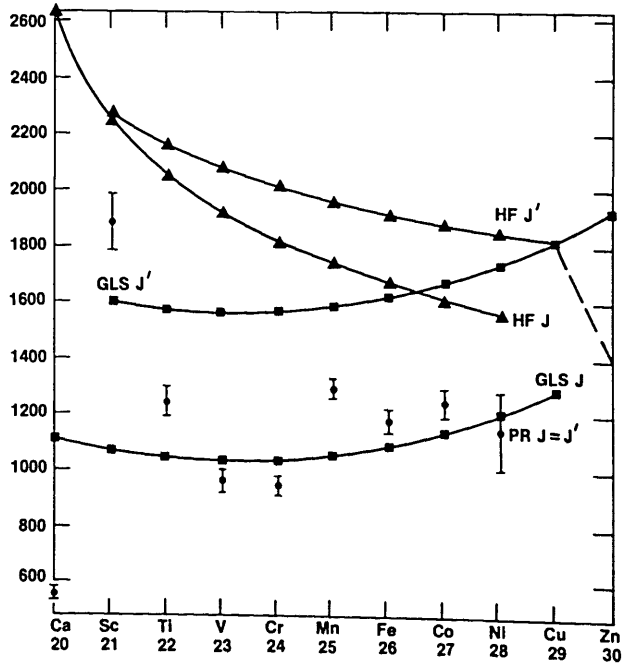
FIGURE 11. Plots of J and J' versus atomic number.

75
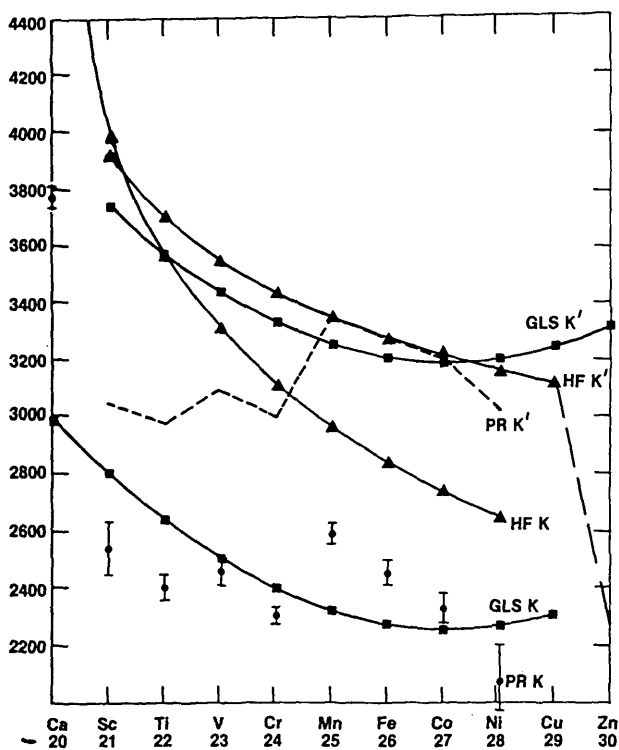
FIGURE 12. Plots of K and K' versus atomic number.
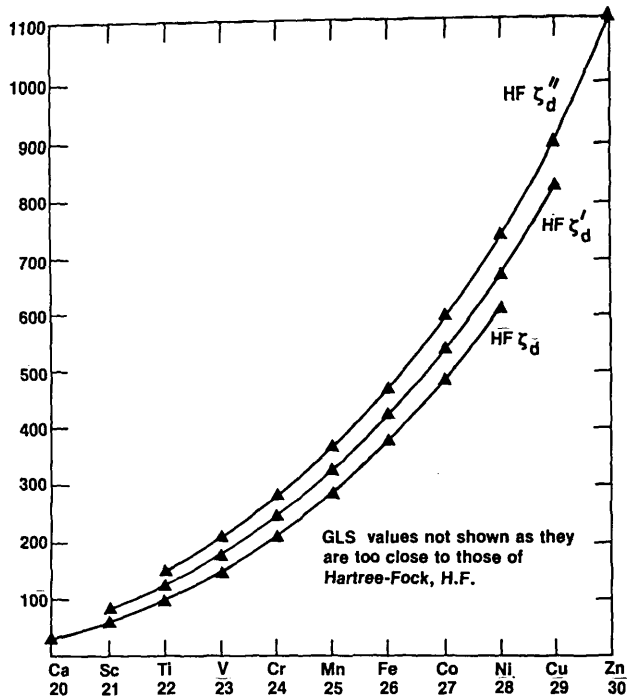
FIGURE 13. Plots of ζ_d, ζ_d', and ζ_d'' versus atomic number.
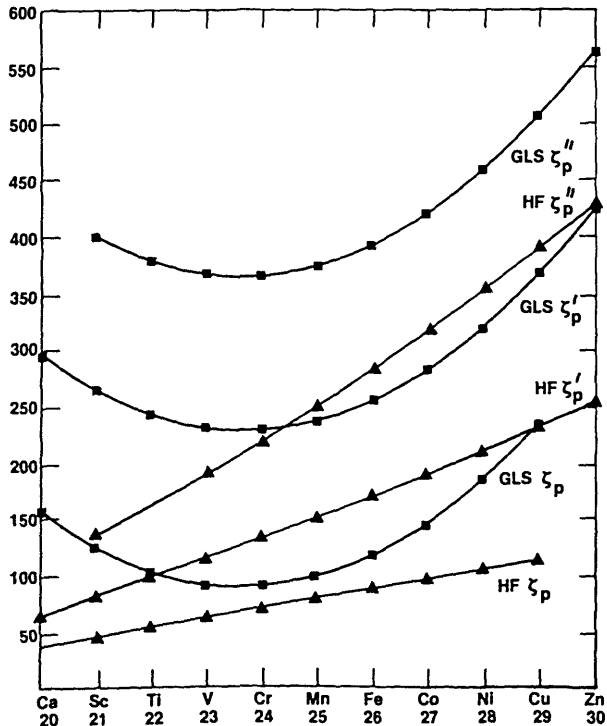
FIGURE 14. Plots of ζ_p, ζ_p', and ζ_p'' versus atomic number.
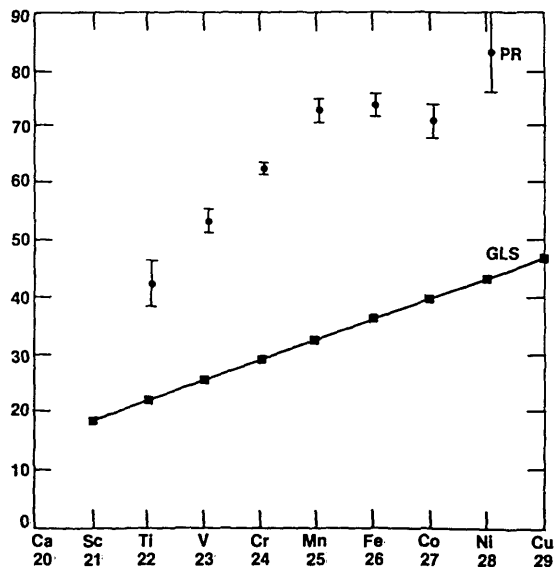
FIGURE 15. Plots of $\alpha = \alpha' = \alpha''$ versus atomic number.

76
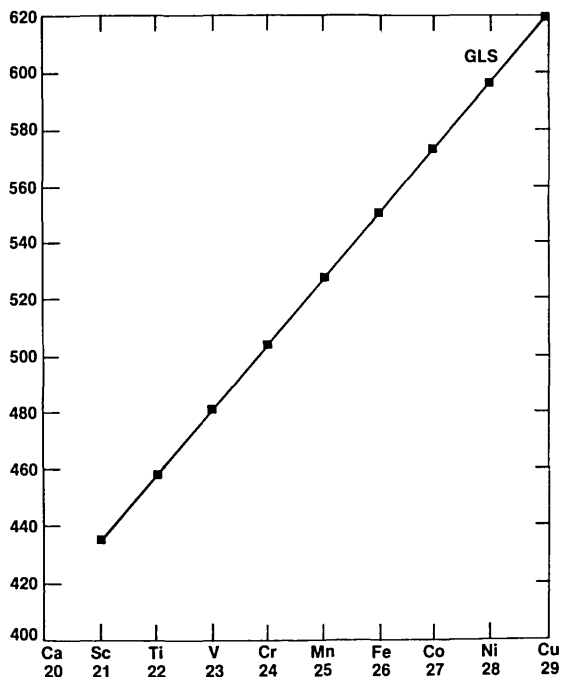
FIGURE 16. Plots of $\beta = \beta' = \beta''$ versus atomic number.
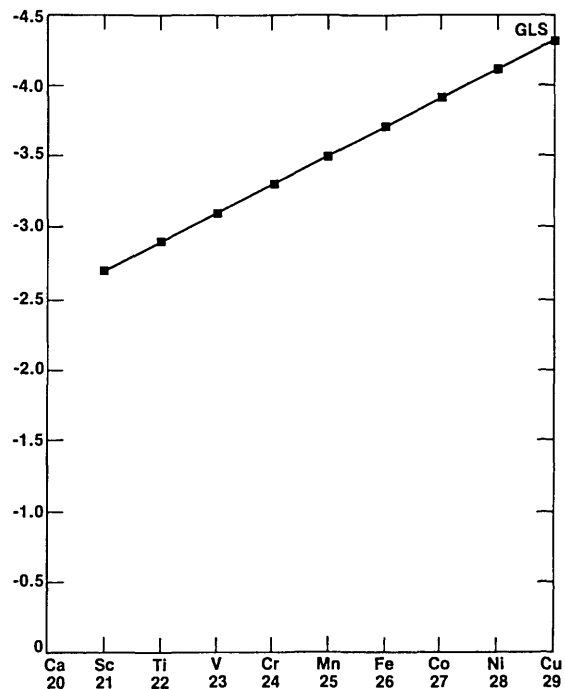
FIGURE 17. Plots of $T = T' = T''$ versus atomic number.

77

An Isoparametric Finite Element Model for Large-Strain Elastostatics

David S. Malkus*

Department of Mathematics, Illinois Institute of Technology, Chicago, IL 60616

and

E. R. Fuller, Jr.†

National Bureau of Standards, Washington, DC 20234

This paper describes a simple finite element model for large-strain elastostatics. The realization of the model in a small-scale computer-code is described. The purpose of the model is to produce test problems for research on the application of penalty techniques in nonlinear elasticity. For this reason the code must balance the requirements of reasonable flexibility with those of computational economy. The current code employs multilinear isoparametric elements. The model is capable of generalization to a variety of element types. The solution method employed is that of incremental loading combined with the Newton-Raphson method. Symmetric, banded systems of equations are produced which are solved in-core. Two- and three-dimensional symmetric bodies which are isoparametric images of a reference "brick" may be modeled. An example comparing two- and three-dimensional models of a "dogbone"—shaped A.S.T.M. rubber tensile-test specimen is presented. The results shed some light on the nature of stress-concentrations which occur in specimens of this geometry.

Key words: Checkerboard pressure; elastostatics; finite elements; in-core solver; isoparametrics; Mooney-Rivlin material; penalty method; plane-stress; nonlinear equation-solvers; strain invariants; tensile-test specimen.

1. Introduction

The finite element solution of the equations of elastostatic equilibrium for incompressible bodies poses special problems not encountered in the compressible case. These problems devolve from the fact that the constraint of material incompressibility turns the problem into a constrained minimization problem. There is a long history of attempts to find computationally efficient methods to impose the constraint within the context of the finite element method [1–5].[1] One of the most promising approaches has proved to be the "penalty method" [4–14]. This technique is one in which the discrete equations of the finite element method are modified to be able to incorporate a very large bulk modulus. Methods were devised [3] by which this could be done without incurring the disastrous loss of accuracy which accompanied early attempts to allow large bulk moduli in the finite element equations [1]. The state of the practice of penalty methods is now refined enough to allow the bulk modulus to be so large that practically any level of compressiblity can be achieved in this fashion. Many problems in incompressible elasticity and fluid dynamics are being successfully attacked using penalties [4–14].

The theory of penalty methods has not, however, caught up with practice. The stability and accuracy of finite element penalty formulations depend crucially on the choice of element types and numerical integra-

[1] Figures in brackets indicate literature references at the end of this paper.

tion formulas [3–8]. Mathematical theories which can give predicted convergence rates and stability criteria exist for problems which are linear or those for which at least the constraint equations are linear [13,14]. Most of the successful applications of penalty methods are in such problems [6,8,11,12,14]. When the constraint is nonlinear—such as the constraint of material incompressibility under finite strain—the situation is quite different. Therefore problems of large-strain, incompressible elasticity are currently under intense scrutiny both in theory and in numerical experiments [9,10,15–19].

The purpose of this paper is to describe a relatively simple finite element model designed for the investigation of penalty methods in finite elasticity. The idea in development of the model is to balance the need for sufficient generality with the need for computational economy. Problems with moderately distorted geometry may be solved so that the interaction between isoparametric transformations and various numerical integration schemes may be observed. On the other hand, since it is desired to run numerous test cases over and over again varying key parameters, the model incorporates only the simplest of isoparametric mesh descriptions. The model is realized in a small-scale code in which all equations can be solved in-core.

The most important feature of the current model is that it has parallel two- and three-dimensional versions. This is designed to confront one of the major obstacles to numerical experimentation in finite elasticity: the unavailability of exact solutions for all but the most simplified problems. The current model allows computation of solutions for thin bodies using a fully three-dimensional formulation. This can be compared with plane-stress computations for the same body. In the plane-stress case, the incompressibility is imposed exactly in the model. In addition, the plane-stress equations assume a linear displacement variation through the thickness of the body. This can be matched exactly by the 3-D model using but one element through the thickness. Therefore the degree of success of the penalty method in imposing incompressibility can be evaluated by comparsion of the 2- and 3-D results. In this paper, we will deal with a simple, but interesting numerical test problem which illustrates such a comparison.

2. Finite element formulation

The kind of material properties we shall incorporate in our model are those of rubber-like, hyperelastic, isotropic materials. We adopt a total Lagrangian description similar to [19]. The static equilibria of such idealized materials are solutions to variational boundary-value problems similar to those described in [20]. We have found that we can derive the finite element equations of equilibrium very compactly if we base our derivations on differentiation with respect to displacement-gradients, as opposed to Green's strains [19,21]. For this reason we are naturally led to the use of the Lagrange stress tensor in our constitutive equation.

2.1. The variational formulation

Let u_i be the displacement field of a body with reference configuration Ω_o in a Cartesian system with coordinates, x^i. If $u_{i,j} = (\partial u_i / \partial x_j)$ is the displacement-gradient tensor, we shall make use of the deformation-gradient tensor

$$J_{ij} = \delta_{ij} + u_{i,j} \tag{1}$$

and the right Cauchy-Green tensor

$$G_{ij} = J_{ki} J_{kj} \tag{2}$$

The constitutive equation for the materials under consideration will be of the form

$$\tau_{ij} = \frac{\partial U}{\partial J_{ji}} = \frac{\partial U}{\partial u_{j,i}} \tag{3}$$

where τ_{ij} is the Lagrange stress tensor and U an energy per unit volume in the undeformed state.

The energy density, U, is assumed to be expressed as a function of the three principal invariants of G_{ij}. If we denote by ψ_{ij} the matrix of cofactors,

$$\psi_{ij} \equiv Co(J_{ij}) \tag{4}$$

the three principal invariants of G_{ij} are given by:
$$
\begin{aligned}
\text{I} &= \text{Trace }[G_{ij}] = J_{ki}J_{ki} \\
\text{II} &= \text{Trace }[Co(G_{ij})] = \psi_{ij}\psi_{ij} \\
\text{III} &= |G_{ij}| = |J_{ij}|^2
\end{aligned}
\tag{5}
$$

where $|\cdot|$ denotes the determinant.

We consider problems whose solutions are characterized by the minima of the variational principle

$$\pi(u_i) = \int_{\Omega_0} [U(\text{I},\text{II},\text{III}) - g_i u_i]dx^1 dx^2 dx^3 - \int_{\Gamma_\sigma^o} f_i u_i ds \tag{6}$$

where g_i is the body-force and f_i the surface tractions, both referred to the undeformed state. The surface of Ω_o, denoted Γ_o, is divided into a load-surface Γ_σ^o and a surface with prescribed displacements, Γ_u^o. All loads are assumed to be dead loads.

2.2. Isoparametric finite elements

We shall consider a particular kind of finite element mesh. It will be composed of only one type of element, which is defined on a reference domain Ω_R (the unit cube, for example), and is mapped into its location in the global mesh by a transformation based on the element shape-functions defined on Ω_R. A variety of moderately complex bodies may be modeled in this fashion, and with the development of some simple notation, the model can be described without much difficulty.

The process of constructing a finite element assemblage involves piecing together small domains to form a large domain. Similarly, the global trial functions are the unions of element trial functions which are supported (i.e., are nonzero) only on individual elements. The global trial functions we describe are compatible for elasticity problems [22,23]. This means that they are continuous across interelement boundaries, which is assured by requiring that the values of the trial function agree on shared boundaries.

a. The element transformations

Let the ith component of the local coordinates of Ω_R be ξ^i, and select ξ_n^i as coordinates of the nth of n_R nodal points of the prototypical element. Construct n_R shape functions, ϕ_r, defined for all values of ξ^i. These shape functions are polynomials on Ω_R, such that

$$
\begin{cases}
\phi_r(\xi^i) = 0 & \xi^i \notin \Omega_R \\[2mm]
\phi_r(\xi_s^i) = \delta_{rs} & r,s = 1, \ldots, n_R
\end{cases}
\tag{7}
$$

according to the procedure outlined in [22,23] and illustrated by an example in figure 1.

For each of the $1 \leqslant e \leqslant N_e$ elements in the global mesh, define a vector-valued map

$$T_i^e \equiv t_{ir}^e \phi_r \tag{8}$$

The summation convention is used on the non-tensorial index $r = 1, \ldots, n_R$, but is never used on the index e. Equation (8) defines the isoparametric transformation of each element from Ω_R to its location in the global mesh. Note that the global nodal locations are then

$$T_i^e(\xi_s^j) = t_{ir}^e \tag{9}$$

Examples of typical transformations are illustrated in figure 2.

We assume that the t_{ir}^e are chosen to satisfy certain basic properties: (1) that compatibility for elasticity problems is satisfied by the global shape functions, (2) that the totality of the transformations T_i^e defines an
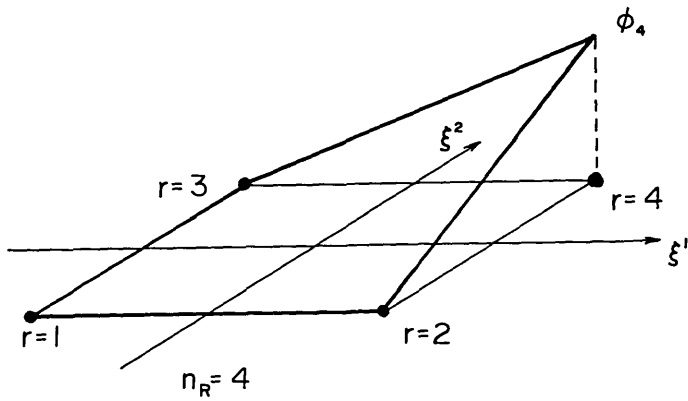
FIGURE 1. A finite element shape function on a reference four-node element.
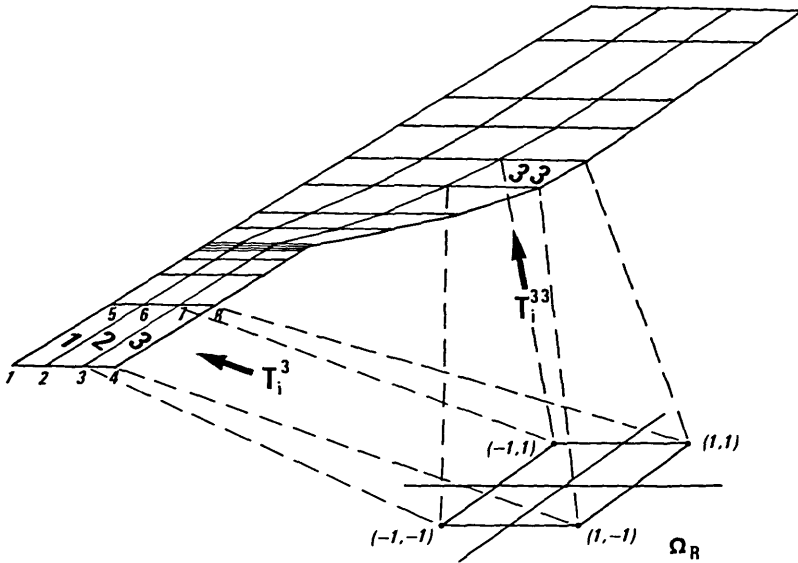
FIGURE 2. The transformation by T_i^r from the reference element on Ω_R to the global mesh on Ω_o.

assemblage of elements such that their domain $\overline{\Omega}_o$ with boundary $\overline{\Gamma}{}^o$ approximates Ω_o well enough so that Ω_o − $\overline{\Omega}_o$ is negligible for the purposes of the model, (3) that T_i^r is an invertible transformation at every point of element e, with inverse S_i^e. The inverse S_i^e is extended to $\overline{\Omega}_o$ by

$$ S_i^e(x^j) = \begin{cases} \xi^i, \ni T_j^e(\xi^i) = x^j, \ x^j \in \text{element } e \\[2mm] \text{any point not in } \Omega_R, \text{ otherwise} \end{cases} \tag{10} $$

The meshes which we will employ in our model are simple enough that the connectivity of the global mesh can be described by a closed form, integer-valued function, $A(e,r)$, called the "nodal alignment function". To each element e, $1 \leqslant e \leqslant N$, and r, $1 \leqslant r \leqslant n_R$, A assigns the global ordering number of the node with coordinates t_{ir}: For more complicated meshes, A is well-defined, but may be determined from a table of data, rather than an arithmetic expression. In terms of $A(e,r)$, the global shape-functions have the form

$$ \Phi_I = \bigcup_{\substack{e,r \\ \ni A(e,r)=I}} \{\phi_r[S_i^e(\bullet)]\} \tag{11} $$

82

In terms of (11), it can be seen that the purpose of the extension of S_i^e defined in (10) is merely to assure via (7) that whenever x^j is not in any of the elements with $A(e,r) = I$, then $\Phi_I(x^j)$ is well-defined and equal to zero.

A typical Φ_I and its construction is illustrated in figure 3. The finite element trial space, S^h, is then the space of all trial functions which satisfy the displacement boundary conditions on $\tilde{\Gamma}_u^*$ and have the form

$$u_i(x^j) = u_i^I \, \Phi_I \, (x^j) \tag{12}$$

The sum on I is taken over all $1 \leq I \leq N$ global ordering numbers, and u_i^I is the ith component of the nodal-value vector at node I. Equations (7) and (11) imply that if x_I^j are the coordinates of the Ith node in the global ordering,

$$u_i(x_I^j) = u_i^I \tag{13}$$
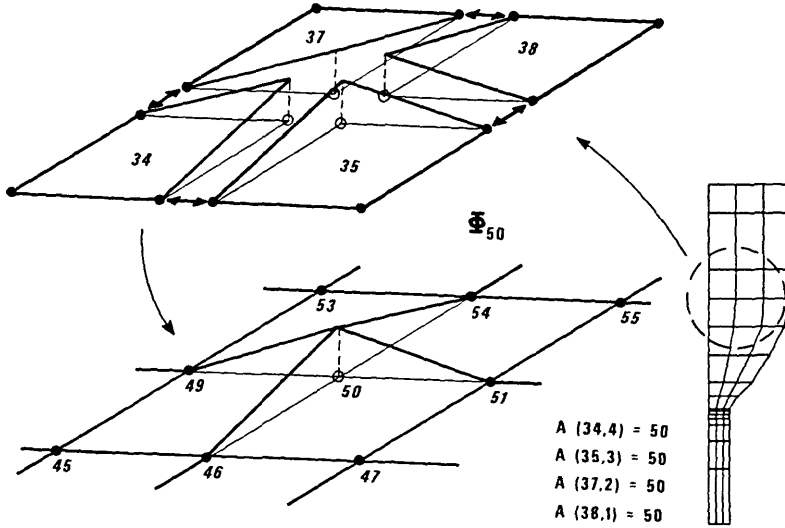
FIGURE 3. Alignment of local and global numberings, showing the role of $A(e,r)$ in piecing together the global shape function Φ_{50} from the ϕ_r on Ω_R.

The restriction of u_i to element e can be determined from the vectors of local nodal values in element e, u_i^{er}, $r = 1, \ldots, n_R$:

$$u_i^{er} = u_i^I, \qquad A(e,r) = I \tag{14}$$

Then by (11) through (14), the restrictions of u_i and $u_{i,j}$ to element e, denoted by, u_i^e and $u_{i,j}^e$, as functions of the reference coordinates, are

$$\begin{cases} u_i^e(x^j) = u_i^e \, [T_j^e(\xi^k)] = u_i^{er} \, \phi_r(\xi^k) \\[2mm] u_{i,j}^e(x^k) = [(\partial T^e)^{-1} u_i^{er} \, \dfrac{\partial \phi_r}{\partial \xi^l} \, C_{ij}^e] \, (\xi^k) \end{cases} \tag{15}$$

where $\xi^k = S_k^e \, (x^j)$ and ∂T^e and C_{ij}^e are defined by

83

$$\begin{cases} \partial T^e_{ij} = \partial T^e_i / \partial \xi^j \\ \\ \partial T^e = |\partial T^e_{ij}| \\ \\ C^e_{ij} = Co\,(\partial T^e_{ji}) \end{cases} \tag{16}$$

To avoid confusion, differentiation with respect to the x^i will use the comma notation, and differentiation with respect to the ξ^i will use $\partial/\partial \xi^i$.

b. The element energy expressions

For U evaluated at any $u_i \in S^h$

$$\int_{\bar{\Omega}_o} U\, dx^1 dx^2 dx^3 = \Sigma_e \int_{\Omega_R} U_e\, \partial T^e d\xi^1 d\xi^2 d\xi^3 \tag{17}$$

where

$$\begin{cases} U_e = U(\mathrm{I}_e, \mathrm{II}_e, \mathrm{III}_e) \\ \\ \mathrm{I}_e = \mathrm{I}\,[u^e_{i,j}\,(\xi^k)],\ \text{etc.} \end{cases} \tag{18}$$

The integral is evaluated by a quadrature rule with weights w_k and evaluation points η^i_k on Ω_R. We define elemental weights on e by

$$w^e_k = w_k\, \partial T^e\,(\eta^i_k) \tag{19}$$

The final energy expression is then

$$\int_{\Omega_o} U\, dx^1 dx^2 dx^3 \cong \Sigma_e\, \Sigma_k\, w^e_k\, U_e\,(\eta^i_k) \tag{20}$$

c. Modifications for penalty methods

The numerical integration scheme of (20) is adequate for a compressible energy density. For incompressible or nearly incompressible energy densities, following [8–10,15,19], we have

$$U(\mathrm{I},\mathrm{II},\mathrm{III}) = W(\mathrm{I},\mathrm{II}) + \frac{z}{2}\, G^2\,(\mathrm{III}-1) \tag{21}$$

where z is a bulk modulus scale factor; z may also be interpreted as the penalty enforcing III–1 to be arbitrarily close to zero for material incompressibility [8–10,15,19]. In the plane-stress case J_{ij} is assumed to have a form which enforces III $= 1$ so that $U(\mathrm{I},\mathrm{II},\mathrm{III}) = W(\mathrm{I},\mathrm{II})$. This will be fully described in section 3.2. It allows the numerical integration scheme of (20) to be applied in the plane-stress case.

For the 3-D model, two integration formulas are required, the one in (20) applied to the W terms in (21), and a "reduced" formula, with weights r_l and evaluation points ζ^i_l applied to the G^2 terms [8–10,22]. Equation 19 has analogue:

$$r^e_l = r_l\, \partial T^e(\zeta^i_l) \tag{22}$$

and (20) is replaced by:

$$\int_{\Omega_o} U\, dx^1 dx^2 dx^3 \cong \Sigma_e\, [\Sigma_k w^e_k W_e(\eta^i_k) + \frac{z}{2} \Sigma_l\, r^e_l G^2_e\,(\zeta^i_l)] \tag{23}$$

where W_e and G_e are defined as in (18).

84

2.3. Derivatives of the energy-density

The load terms in (6) from body and surface forces may be treated exactly as in linear finite element analysis [22,23]. Combining this with (20) or (23) gives a numerically integrated variational principle to be minimized over S^h. We will denote this principle by the same $\pi(u_i)$ and understand that throughout the following discussion the minimum is to be taken over S^h. The equations of equilibrium of the finite element model are then

$$F_i^I \equiv \frac{\partial \pi}{\partial u_i^I} = \sum_{\substack{e,r \\ \ni A(e,r) \,=\, I}} \Sigma_k w_k^e \frac{\partial W_e}{\partial u_i^{er}}(\eta_k^e) + f_i^I = 0$$

$$i = 1, 2 \text{ and } I = 1, 2, \ldots, N \tag{24}$$

in the plane stress case, and

$$F_i^I \equiv \frac{\partial \pi}{\partial u_i^I} = \sum_{\substack{e,r \\ \ni A(e,r) \,=\, I}} \left[\Sigma_k w_k^e \frac{\partial W_e}{\partial u_i^{er}}(\eta_k^e) \right.$$

$$\left. + \frac{z}{2} \Sigma_t \frac{\partial}{\partial u_i^{er}} [G_e^2](\zeta_t^e) \right] + f_i^I = 0 \tag{25}$$

$$i = 1, 2, 3 \text{ and } I = 1, 2, \ldots, N$$

For the 3-D case.

$$f_i^I = \sum_{\substack{e,r \\ \ni A(e,r) \,=\, I}} f_i^{er} \tag{26}$$

where f_i^{er} is the ith component of the distributed nodal force from body and surface forces acting on node r of element e [22,23]. Thus f_i^I is just the usual finite element load vector, and (24–26) express the fact that the assembly of the equilibrium equations of the nonlinear model obeys the same rule as assembly of load vectors in linear analysis [22,23].

The equations of equilibrium, $F_i^I = 0$, will be solved by a combination of incremental loading with the Newton-Raphson method [22]. More details of these procedures will be described in section 4. Both incremental loading and the Newton-Raphson method require calculation of the "tangent stiffness matrix" [22],

$$\mathbf{K}_T = \left[\frac{\partial F_i^I}{\partial u_j^J} \right] \tag{27}$$

Since $\mathbf{K}_T$ is the matrix of second derivatives with respect to u_i^I of π, which is a continuously differentiable function of the nodal-values, $\mathbf{K}_T$ is a symmetric "hessian" matrix.

$$\frac{\partial F_i^I}{\partial u_j^J} = \frac{\partial F_j^J}{\partial u_i^I} = \sum_{\substack{e,r,s \\ \ni A(e,r)=I \\ \ni A(e,s)=J}} \Sigma_k w_k^e \frac{\partial^2 W_e}{\partial u_i^{er} \partial u_j^{es}}(\eta_k^e) \tag{28}$$

in the plane-stress case and

$$\frac{\partial F_i^I}{\partial u_j^J} = \frac{\partial F_j^J}{\partial u_i^I} = \sum_{\substack{e,r,s \\ \ni A(e,r)=I \\ \ni A(e,s)=J}} \left[\Sigma_k w_k^e \frac{\partial^2 W_e}{\partial u_i^{er} \partial u_j^{es}}(\eta_k^e) \right. \tag{29}$$

$$\left. + \frac{z}{2} \Sigma_t r_t \frac{\partial^2}{\partial u_i^{er} \partial u_j^{es}} [G_e^2](\zeta_t^e) \right]$$

in the 3-D case.

In other words, (27–29) imply that $\mathbf{K}_T$ assembles by the same rule by which stiffness matrices assemble in linear analysis [22,23]. Beyond that, $\mathbf{K}_T$ will have the same banded structure which a stiffness matrix in

linear analysis would have, using the same mesh [22,23]. This follows because the band-structure is determined by

$$\omega \equiv \max_{\substack{e,r,s \\ A(e,r)=I \\ \ni A(e,s)=J}} |I-J| \tag{30}$$

$\mathbf{K}_T$ will have $(2\omega+1)d$ bands, where $d = 3$ for the 3-D model and $d=2$ for plane-stress. Clearly ω is determined by the form of A and not the form of the energy expression. Boundary conditions are also imposed on $\tilde{\Gamma}_u^*$ in the same manner as in linear analysis [22,23].

a. Separation of tensorial factors

In what follows, let U_e represent generically either a compressible element energy density, $U(I_e, II_e, III_e)$, or one of the terms of an incompressible element energy density, W_e or G_e. From the previous section we may conclude that

$$\begin{cases} \dfrac{\partial U_e}{\partial u_i^{er}} = \dfrac{\partial U_e}{\partial u_{k,\ell}^e} \dfrac{\partial u_{k,\ell}^e}{\partial u_i^{er}} = \dfrac{\partial U_e}{\partial u_{i,j}^e} (\partial T^e)^{-1} \dfrac{\partial \phi_r}{\partial \xi^m} C_{mj}^e \\[4mm] \dfrac{\partial^2 U_e}{\partial u_i^{er} \partial u_k^{es}} = \dfrac{\partial^2 U_e}{\partial u_{i,j}^e \partial u_{k,\ell}^e} (\partial T^e)^{-2} \dfrac{\partial \phi_r}{\partial \xi^m} \dfrac{\partial \phi_s}{\partial \xi^n} C_{mj}^e C_{n\ell}^e \end{cases} \tag{31}$$

We point out again that there is no summation on e; in fact, since in what follows we will always be concerned with elemental expressions, we will drop the index e. At this point the strain-energy derivatives are seen to be the sum of terms which are separated into factors, the second of which,

$$(\partial T)^{-1} \frac{\partial \phi_r}{\partial \xi^m} C_{mj} \quad \text{or} \quad (\partial T)^{-2} \frac{\partial \phi_r}{\partial \xi^m} \frac{\partial \phi_s}{\partial \xi^n} C_{mj} C_{n\ell} \tag{32}$$

is determined from the elemental isoparametrics and local shape functions, and is no different in character from the corresponding factors in linear isoparametric problems [22,23]. What distinguishes the finite elasticity case is the complexity of the first factors. It is these derivatives of the energy density with respect to displacement-gradients to which we now turn our attention. Actually, we may further restrict our attention to the derivatives of strain-invariants, since

$$\begin{cases} U_{ij} = \dfrac{\partial U}{\partial u_{i,j}} = \dfrac{\partial U}{\partial I} \dfrac{\partial I}{\partial u_{i,j}} + \dfrac{\partial U}{\partial II} \dfrac{\partial II}{\partial u_{i,j}} + \dfrac{\partial U}{\partial III} \dfrac{\partial III}{\partial u_{i,j}} \\[4mm] U_{ijk\ell} = \dfrac{\partial^2 U}{\partial u_{i,j} \partial u_{k,\ell}} = \dfrac{\partial^2 U}{\partial I^2} \dfrac{\partial I}{\partial u_{i,j}} \dfrac{\partial I}{\partial u_{k,\ell}} + \dfrac{\partial U}{\partial I} \dfrac{\partial^2 I}{\partial u_{i,j} \partial u_{k,\ell}} \end{cases} \tag{33}$$

$$+ \text{ similar terms for II and III}$$

$\dfrac{\partial U}{\partial I}, \dfrac{\partial^2 U}{\partial I^2}, \dfrac{\partial U}{\partial II} \ldots$, etc., depend on the form of the energy density for a specific material and are not usually difficult to evaluate. So the problem is to compute the derivatives of the invariants.

b. Derivatives of the strain-invariants

Derivatives of the strain invariants, I_{ij}, $I_{ijk\ell}$, II_{ij}, ..., etc., are defined in a fashion similar to U_{ij} and $U_{ijk\ell}$. The expression for I_{ij} follows immediately from equation (5), as does III_{ij}, by remembering that the derivative of a determinant with respect to the matrix itself is the matrix of cofactors. Next we observe that the derivative of the matrix of cofactors with respect to the matrix itself can be written

$$\frac{\partial \psi_{ij}}{\partial J_{k\ell}} = \epsilon_{ikm} \epsilon_{j\ell n} J_{mn} \tag{34}$$

86

where ϵ_{ijk} is the Cartesian permutation symbol. Using the chain-rule and collecting terms by application of symmetries gives an expression for II_{ij}. So we have

$$
\left\{
\begin{array}{l}
I_{ij} = 2J_{ij} \\[4pt]
II_{ij} = 2\epsilon_{ikm}\epsilon_{j\ell n}J_{mn}\psi_{k\ell} \\[4pt]
III_{ij} = 2|J_{ij}|\psi_{ij}
\end{array}
\right.
\tag{35}
$$

Calculating $I_{ijk\ell}$ is trivial. Using the chain-rule and the identity $\epsilon_{ijp}\epsilon_{k\ell p} = \delta_{ik}\delta_{j\ell} - \delta_{i\ell}\delta_{kj}$ leads readily to an expression for $II_{ijk\ell}$. To find $III_{ijk\ell}$, use can be made of the inverse of J_{ij}, $(J_{ij})^{-1} = \psi_{ji}/|J_{k\ell}|$. Then III_{ij} becomes

$$
III_{ij} = 2|J_{ij}|^2(J_{ji})^{-1}
\tag{36}
$$

(note the transposition). Making use of the identity

$$
\frac{\partial(J_{ji})^{-1}}{\partial J_{k\ell}} = -(J_{\ell i})^{-1}(J_{jk})^{-1} = \psi_{i\ell}\psi_{kj}/|J_{mn}|^2
\tag{37}
$$

leads to a simple expression for $III_{ijk\ell}$. Accordingly we have

$$
\left\{
\begin{array}{l}
I_{ijk\ell} = 2\delta_{ik}\delta_{j\ell} \\[8pt]
II_{ijk\ell} = 2\epsilon_{ikm}\epsilon_{j\ell n}\psi_{mn} + 2\delta_{ik}\delta_{j\ell}J_{mn}J_{mn} \\[4pt]
\qquad\quad - 2\delta_{ik}J_{m\ell}J_{mj} - 2\delta_{j\ell}J_{im}J_{km} + 2J_{ij}J_{k\ell} \\[8pt]
III_{ijk\ell} = 4\psi_{ij}\psi_{k\ell} - 2\psi_{i\ell}\psi_{kj}
\end{array}
\right.
\tag{38}
$$

Although the terms with the ϵ_{ijk} may look forbidding, they may be computed using no floating point operations other than "change sign," and the same calculations can be used for II_{ij} and $II_{ijk\ell}$. Alternative expressions (of comparable complexity) for II_{ij} and $II_{ijk\ell}$ can be derived without ϵ_{ijk} using the identity: $II = (1/2)(I^2 - G_{ij}G_{ij})$.

3. Two- and three-dimensional problems

In this section we describe the assumptions which are made in reducing a three-dimensional problem in finite elasticity to a two-dimensional problem. We describe how this reduction is incorporatd in the framework of the results of section 2. We also will describe the similarity and differences between the element shape-functions and the isoparametric transformations in the two- and three-dimensional cases. It will be easier to describe the implementation of the isoparametric transformations in two-dimensions. The three-dimensional implementation can be viewed as a straight-forward generalization of the 2-D case. Finally, as will be explained in more detail in section 5, linear elastic problems serve an important purpose in computer code verification. They are also interesting in their own right. We therefore conclude this section with a discussion of two- and three-dimensional linear elasticity problems. The linear two-dimensional model incorporates compressible, nearly incompressible and incompressible materials in states of plane stress or plane strain. Penalties are used to enforce plane-strain incompressibility.

3.1. Three-dimensional problems

The results of section 2 were derived in full three-dimensional form. They may be applied directly to the coding of the element matrix, element load-vector and assembly subroutines [22,23]. Details of how the mesh is described and how the isoparametric transformations are generated will be described later in this section.

But the two-dimensional, plane-stress model requires additional development. It may be thought of as a three-dimensional model in which specific simplifying assumptions are imposed on the stress and strain tensors. To see how this is accomplished, it is useful to look first at those tensors in the fully three-dimensional case and observe an example of a particular material model.

a. The general incompressible formulation

In the 3-D case J_{ij}, G_{ij}, τ_{ij} and all the tensors of the previous section are of dimension three in each rank. The Lagrange stress tensor can be written in terms of (35)

$$\tau_{ij} = \frac{\partial U}{\partial u_{j,i}} = \frac{\partial U}{\partial \mathrm{I}}\, \mathrm{I}_{ji} + \frac{\partial U}{\partial \mathrm{II}}\, \mathrm{II}_{ji} + \frac{\partial U}{\partial \mathrm{III}}\, \mathrm{III}_{ji} \tag{39}$$

For incompressible, 3-D problems, U is assumed to have the form (21), so (39) becomes

$$\tau_{ij} = \frac{\partial W}{\partial \mathrm{I}}\, \mathrm{I}_{ji} + \frac{\partial W}{\partial \mathrm{II}}\, \mathrm{II}_{ji} + zG(\mathrm{III}{-}1)\frac{\partial G}{\partial \mathrm{III}}\, \mathrm{III}_{ji} \tag{40}$$

Recall that z is a large penalty parameter forcing $G(\mathrm{III}{-}1)$ to be small in the finite element solution. Therefore the pressure is given by

$$h = zG(\mathrm{III}{-}1)\frac{\partial G}{\partial \mathrm{III}} \tag{41}$$

so that (40) gives an incompressible constitutive equation in the $z \to \infty$ limit.

For stress calculations from the finite element model, it is more physically meaningful to compute the Euler stress tensor, referred to the strained body:

$$\sigma'_{ij} = J_{ki}\tau_{kj}/\,|J_{mn}| \tag{42}$$

Equation (35) and the fact that $(J_{ji})^{-1} = \psi_{ij}/|J_{ij}|$ with (41–42) imply

$$\sigma'_{ij} = \Big[\frac{\partial W}{\partial \mathrm{I}} J_{kl}\mathrm{I}_{jk} + \frac{\partial W}{\partial \mathrm{II}} J_{kl}\mathrm{II}_{jk} + 2h\mathrm{III}\delta_{ij}\Big]/\,|J_{mn}| \tag{43}$$

Then denote by B_{ij} the left Cauchy-Green tensor and by C_{ij} its inverse

$$B_{ij} = J_{ik}J_{jk}$$

$$C_{ij} = \psi_{kj}\psi_{ki}/\,|J_{mn}|^2 \tag{44}$$

We can rewrite II as

$$\mathrm{II} = C_{ii}|G_{mn}| \tag{45}$$

from which it easily follows using (37) that

$$J_{ki}\mathrm{II}_{jk} = 2\mathrm{III}[C_{kk}\delta_{ij} - C_{ij}] \tag{46}$$

which leads to

$$\sigma'_{ij} = 2\sqrt{\mathrm{III}} \Big[\frac{\partial W}{\partial \mathrm{I}} B_{ij} - \frac{\partial W}{\partial \mathrm{II}} C_{ij} + \bar{h}\,\delta_{ij}\Big] \tag{47}$$

in which

$$\bar{h} = \text{III}h + \frac{\partial W}{\partial \text{II}} C_{ii} \tag{48}$$

Finally it is desired to have a stress-free undeformed state, but for $J_{ij} = \delta_{ij}$ (47) gives

$$\sigma'_{ij} = h_o \delta_{ij} = \left(\frac{\partial W}{\partial \text{I}} + 2\frac{\partial W}{\partial \text{II}}\right) \delta_{ij} \tag{49}$$

so letting $H = \bar{h} - h_o$

$$\sigma_{ij} = 2\sqrt{\text{III}} \left[\frac{\partial W}{\partial \text{I}} B_{ij} - \frac{\partial W}{\partial \text{II}} C_{ij} + H\delta_{ij}\right] \tag{50}$$

All constituents of (50) are easy to compute given J_{ij}. Symmetry implies only six of the components of σ_{ij} need be calculated.

b. Application to Mooney-Rivlin materials

A simple constitutive equation of the form (40) can be obtained by taking

$$\begin{cases} \dfrac{\partial W}{\partial \text{I}} = c_1, \qquad \dfrac{\partial W}{\partial \text{II}} = c_2 \\[2mm] G(x) = x \end{cases} \tag{51}$$

This is a penalty function version of the Mooney-Rivlin constitutive equation [9,10,19, and 21]; c_1 and c_2 are material constants. This equation is intended to model exactly incompressible materials, and in fact, it has some continuum mechanical inadequacies if significant compressive deformation is allowed [10]. Therefore it should only be employed with large values of z. The exactly incompressible, plane-stress version of the constitutive model has been found to give reasonable agreement with experimental results in some problems [21]. How well the 3-D penalty version succeeds will be discussed below.

The symmetric stress tensor for Mooney-Rivlin material is given by

$$\begin{cases} \sigma_{ij} = 2\sqrt{\text{III}} \, [c_1 B_{ij} - c_2 C_{ij} + H\delta_{ij}] \\[2mm] H = z(\text{III}-1)\text{III} + c_2 C_{ii} - c_1 - 2c_2 \end{cases} \tag{52}$$

It should be noted that this differs slightly from the H used elsewhere [9,19,21], because $c_2 C_{ii} \delta_{ij}$ is considered a pressure term. This means that H defined in (52) has the value $c_2 - c_1$ in the undeformed state. [In [19], since the $c_2 C_{ii} \delta_{ij}$ term is included with the deviatoric terms, the pressure in the undeformed state is $-c_1 - 2c_2$.] A quantity of more interest than H, then, is the pressure above the undeformed state:

$$\bar{H} = z(\text{III}-1)\text{III} + c_2(C_{ii}-3) \tag{53}$$

3.2. Finite plane stress

a. The general incompressible formulation

The assumptions in the plane-stress model are that
1. The body is very thin, with uniform thickness, $2t$, in the x^3-coordinate. How thin the body must be will be dealt with further below.
2. The applied forces are in the x^1-x^2 plane; the upper and lower surfaces in the x^1-x^2 plane are free surfaces.
3. $\sigma_{31} = \sigma_{32} = \sigma_{33} = 0$ throughout the thickness, i.e. for $-t \le x^3 \le t$.

4. The deformation is essentially biaxial, i.e.

$$
\begin{aligned}
u_1(x^1,x^2,x^3) &= u_1(x^1,x^2) \\
u_2(x^1,x^2,x^3) &= u_2(x^1,x^2) \\
u_3(x^1,x^2,x^3) &= [\Theta(x^1,x^2)-1]x^3
\end{aligned}
\tag{54}
$$

for an unknown function $\Theta(x^1,x^2)$.

5. The body is exactly incompressible.

This model is discussed in more detail in [21]. An example of such a deformation is shown schematically in figure 4.

$$
u_3(x^1,x^2,x^3) = [\Delta^1(x^1,x^2)-1]\,x^3
$$
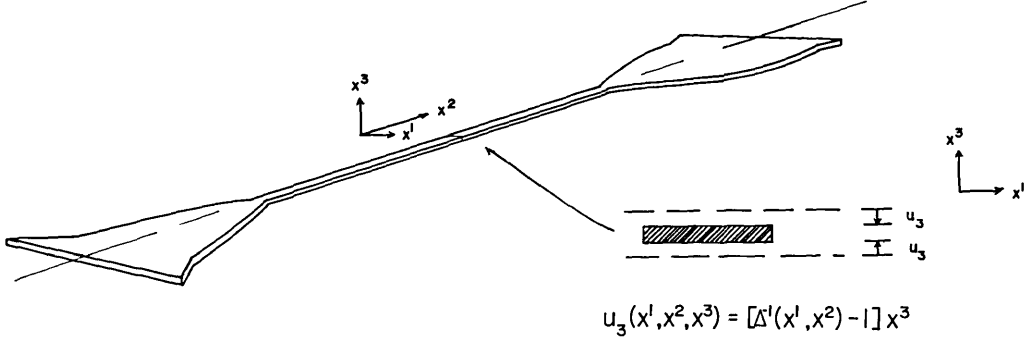
FIGURE 4. Plane-stress assumptions for the deformation through the thickness of a thin body.

These assumptions imply that the deformation-gradients have the form

$$
J_{ij} =
\begin{bmatrix}
1+u_{1,1} & u_{1,2} & 0 \\
u_{2,1} & 1+u_{2,2} & 0 \\
x^3\Theta_{,1} & x^3\Theta_{,2} & \Theta
\end{bmatrix}
\tag{55}
$$

Assumption 1 is further specified by the requirement that the body be thin enough for $t\Theta_{,1}$ and $t\Theta_{,2}$ to be neglected. Note that this is not an absolute requirement on the thickness of the body, since if $\Theta_{,1}$ and $\Theta_{,2}$ are small enough, it can compensate for a larger t. On the other hand, very rapid spatial variations of $\Theta(x^1,x^2)$ could invalidate the assumption, even for small t.

Let $K_{\alpha\beta}$ be defined for $\alpha,\beta = 1,2$

$$
K_{\alpha\beta} =
\begin{bmatrix}
1+u_{1,1} & u_{1,2} \\
u_{2,1} & 1+u_{2,2}
\end{bmatrix}
\tag{56}
$$

We will use Greek subscripts to indicate tensors which are of dimension 2 in each rank. Imposing the thinness assumption leads to

$$
J_{ij} = [K_{\alpha\beta} \oplus \Theta]
\tag{57}
$$

where " $\oplus$ " means "matrix direct sum" of the 2x2 matrix, $K_{\alpha\beta}$, with the 1x1 matrix, Θ. That is

$$
J_{ij} =
\begin{bmatrix}
K_{\alpha\beta} & \vdots & 0 \\
& \vdots & 0 \\
\cdots & \vdots & \cdots \\
0 \quad 0 & \vdots & \Theta
\end{bmatrix}
\tag{58}
$$

90

Assumption 5 is enforced by taking $\Theta = \Delta^{-1}(x^1,x^2)$ where

$$\Delta(x^1,x^2) = |K_{\alpha\beta}| \tag{59}$$

Now we can rewrite the invariants of G_{ij} and their derivatives, using (35–38) and the special form (58). In (58) we have mixed tensor and matrix notation in an obvious way. In what follows, we wish to construct tensors of rank 4 from tensors of rank 2 given in matrix notation. We adopt the convention that a matrix with subscripts appended on the lower right defines a tensor of rank 2 with components given by the matrix entries. The first subscript is the row index and the second the column index, as usual. For example, in terms of (58), ψ_{ij} of cofactors of J_{ij} is

$$\psi_{ij} = \{[\Delta^{-1}CoK_{\alpha\beta}] \oplus \Delta\}_{ij}. \tag{60}$$

Similarly

$$
\left\{
\begin{aligned}
& I = K_{\alpha\beta}K_{\alpha\beta} + \Delta^{-2} \\[4pt]
& II = \Delta^{-2}CoK_{\alpha\beta}CoK_{\alpha\beta} + \Delta^2 \\[4pt]
& I_{ij} = 2[K_{\alpha\beta} \oplus \Delta^{-1}]_{ij} \\[4pt]
& II_{ij} = 2\{[\Delta^{-2}K_{\alpha\beta} + \Delta CoK_{\alpha\beta}] \oplus [\Delta^{-1}CoK_{\alpha\beta}CoK_{\alpha\beta}]\}_{ij} \\[4pt]
& III_{ij} = 2\{[\Delta^{-1}CoK_{\alpha\beta}] \oplus \Delta\}_{ij} \\[4pt]
& I_{ijkl} = 2\delta_{ik}\delta_{jl} \\[4pt]
& II_{ijkl} = 2\{\epsilon_{ikm}\epsilon_{jln}\psi_{mn} + \delta_{ik}\delta_{jl}I \\[4pt]
& \qquad\qquad - \delta_{ik}[K_{\gamma\alpha}K_{\gamma\beta} \oplus \Delta^{-2}]_{jl} - \delta_{jl}[K_{\alpha\gamma}K_{\beta\gamma} \oplus \Delta^{-2}]_{ik} \\[4pt]
& \qquad\qquad + J_{ij}J_{kl}\}\}
\end{aligned}
\right. \tag{61}
$$

Equation (58) has made $III = 1$ and the energy density is not a function of III, so III_{ijkl} is not needed. III_{ij} would only be needed in the event one wished to calculate τ_{ij}. By enforcing plane-stress assumption 3, the pressure can be determined directly. It should be observed that (58) and (59) already imply that $\sigma_{31} = \sigma_{32} = 0$, since (58) implies

$$
\left\{
\begin{aligned}
& B_{ij} = [K_{\alpha\gamma}K_{\beta\gamma} \oplus \Delta^{-2}]_{ij} \\[4pt]
& C_{ij} = [\Delta^{-2}CoK_{\gamma\alpha}CoK_{\gamma\beta} \oplus \Delta^2]_{ij}
\end{aligned}
\right. \tag{62}
$$

Substituting (62) in (50), setting $\sigma_{33} = 0$ and solving for H implies

$$H = -\frac{\partial W}{\partial I}\Delta^{-2} + \frac{\partial W}{\partial II}\Delta^2 \tag{63}$$

The reduced constitutive law may be now written

$$
\left\{
\begin{aligned}
& \sigma_{ij} = 2\left[\frac{\partial W}{\partial I}B_{ij} - \frac{\partial W}{\partial II}C_{ij} + H\delta_{ij}\right] \\[4pt]
& H = -\frac{\partial W}{\partial I}\Delta^{-2} + \frac{\partial W}{\partial II}\Delta^2
\end{aligned}
\right. \tag{64}
$$

Assumption 3 is enforced by choosing plane finite elements of the type illustrated in figures 1–3 and by replacing every occurrence of $u_{\alpha,3}$ or $u_{3,\alpha}$ by zero ($\alpha = 1,2$) and $u_{3,3}$ by Θ in the energy expressions of section 2.2. Any explicit occurrence of x^3 is replaced by $x^3 = 0$, since we assume negligible variation of all quantities through the thickness. The x^3-integral in the multiple integrals of the energy expressions integrates to a factor of $2t$ in the energy expressions. This may be ignored since this will not change the minimizing displacement field. The result of the reduction is that there is no dependence on x^3 or t in the plane-stress model.

A reduced displacement field is used which has only two components, u_α; u_3 is a function of u_1 and u_2. Therefore in section 2.3, all derivatives with respect to nodal values of u_3 vanish. But it is important to note that in (31), terms of the form $\dfrac{\partial u_{3,3}}{\partial u_\alpha^{er}}$ do not vanish. They have a new form, introduced by the plane-stress assumption of the form of $u_3(x^1,x^2)$. For notational convenience, define

$$\phi_{rm} \equiv (\partial T)^{-1}\frac{\partial \phi_r}{\partial \xi^k}\, C_{km} \tag{65}$$

This is the rth element shape function's derivative with respect to Cartesian spatial variable x^m, and for plane stress $m = 1$ or 2. Let $L = $ I, II or III and u_α^r $\alpha = 1$ or 2 be the αth component of the rth local nodal value in an element

$$
\left\{
\begin{aligned}
\frac{\partial L}{\partial u_\alpha^r} &= \frac{\partial L}{\partial u_{i,j}}\frac{\partial u_{i,j}}{\partial u_\alpha^r} = L_{ij}\frac{\partial u_{i,j}}{\partial u_\alpha^r} \\[2mm]
\frac{\partial^2 L}{\partial u_\alpha^r \partial u_\beta^s} &= \frac{\partial^2 L}{\partial u_{i,j}\partial u_{k,\ell}}\frac{\partial u_{i,j}}{\partial u_\alpha^r}\frac{\partial u_{k,\ell}}{\partial u_\beta^s} + L_{ij}\frac{\partial^2 u_{i,j}}{\partial u_\alpha^r \partial u_\beta^s} \\[2mm]
&= L_{ijk\ell}\frac{\partial u_{i,j}}{\partial u_\alpha^r}\frac{\partial u_{k,\ell}}{\partial u_\beta^s} + L_{ij}\frac{\partial^2 u_{i,j}}{\partial u_\alpha^r \partial u_\beta^s}
\end{aligned}
\right. \tag{66}
$$

Since trial functions are linear in the nodal-value coefficients u_α^r, if i and $j \neq 3$ we have

$$\frac{\partial u_{i,j}}{\partial u_\alpha^r} = \delta_{i\alpha}\phi_{rj}, \qquad \frac{\partial^2 u_{i,j}}{\partial u_\alpha^r \partial u_\beta^s} = 0 \tag{67}$$

$u_{3,3}$ is not linear in u_α^r and is a function of the in-plane displacement gradients. Therefore contributions to the chain-rule sums in (66) from $i,j,k,$ and $\ell = 1$ or 2 can be computed directly from (61), and (67) which hold for these values of the subscripts $i,j,k,$ and ℓ. The contributions to (66) from $i = j = 3$ and/or $k = \ell = 3$ may be found by using L_{ij} and $L_{ijk\ell}$ determined by (61) and the following facts:

$$
\left\{
\begin{aligned}
\frac{\partial u_{3,3}}{\partial u_\alpha^r} &= \frac{\partial \Delta^{-1}}{\partial u_\alpha^r} = -\Delta^{-2}\frac{\partial \Delta}{\partial u_\alpha^r} \\[2mm]
\frac{\partial \Delta}{\partial u_\alpha^r} &= (-1)^{\alpha+1}(\phi_{r\alpha}J_{3-\alpha\ 3-\alpha} - \phi_{r\ 3-\alpha}J_{3-\alpha\ \alpha}) \\
&\qquad\qquad\text{(no sum on }\alpha\text{)} \\[2mm]
\frac{\partial^2 u_{3,3}}{\partial u_\alpha^r \partial u_\beta^s} &= 2\Delta^{-3}\frac{\partial \Delta}{\partial u_\alpha^r}\frac{\partial \Delta}{\partial u_\beta^s} - \Delta^{-2}\frac{\partial^2 \Delta}{\partial u_\alpha^r \partial u_\beta^s} \\[2mm]
\frac{\partial^2 \Delta}{\partial u_\alpha^r \partial u_\beta^s} &= (1-\delta_{\alpha\beta})(-1)^{\alpha+1}(\phi_{r\alpha}\phi_{s\beta} - \phi_{r\beta}\phi_{s\alpha}) \\
&\qquad\qquad\text{(no sum on }\alpha,\beta\text{)}
\end{aligned}
\right. \tag{68}
$$

b. Application to Mooney-Rivlin materials

When the constitutive equation with $\dfrac{\partial W}{\partial I}$ and $\dfrac{\partial W}{\partial II}$ given as in (51) is employed in the plane-stress case (64) becomes

$$
\left\{
\begin{aligned}
\sigma_{ij} &= 2[c_1 B_{ij} - c_2 C_{ij} + H\delta_{ij}] \\[2mm]
H &= -c_1\Delta^{-2} + c_2\Delta^2
\end{aligned}
\right. \tag{69}
$$

92

$\bar{H}$, the pressure above the undeformed state, given by

$$\bar{H} = c_1(1-\Delta^{-2}) + c_2(\Delta^2-1) \tag{70}$$

3.3. Shape functions and isoparametrics

Here we discuss the specific choice of shape functions, ϕ_r, and the isoparametric transformations, T_i^e, which are described generally in section 2.2. It will be easier to describe the two-dimensional case first and think of the 3-D case as the appropriate generalization.

a. Two dimensions

The reference element, Ω_R, is the square in the $\xi^1-\xi^2$ plane with $-1 \leq \xi^i \leq 1$. On this domain,

$$\phi_r(\xi^1,\xi^2) = \frac{(\xi^1+\xi_r^1)(\xi^2+\xi_r^2)}{4\,\xi_r^1\,\xi_r^2} \tag{71}$$

where $(\xi_1^1, \xi_1^2) = (-1,-1)$, $(\xi_2^1, \xi_2^2) = (1,-1)$, $(\xi_3^1, \xi_3^2) = (-1,1)$ and $(\xi_4^1, \xi_4^2) = (1,1)$. These element shape functions look exactly like the one illustrated in figure 1.

To construct the isoparametric transformation, T_i^e, all that is needed is to determine the nodal locations, t_{ir}^e, of (8). Then the isoparametric transformation of Ω_R is determined throughout element e via (8). The nodal locations t_{ir}^e give the $i = 1,2$ coordinates in the global mesh of the nodes $r = 1,2,3,4$ of element e. From this information, the rest of the computations implementing the ideas of section 2 can be carried out once the nodal alignment function, $A(e,i)$, is defined. To do this, we chose to restrict our attention to curved bodies which are described as a class of differentiable, invertible tensor transformations of the reference "brick" of elements illustrated in figure 5. It is clear that the nodal alignment of such a body is the same as the nodal alignment of the reference brick. Therefore the global element and nodal numbering schemes and $A(e,i)$ may be determined from the reference brick. We have chosen the numbering scheme illustrated in figure 5. In terms of these numberings, $A(e,i)$ can be expressed as a simple, closed form, integer expression.
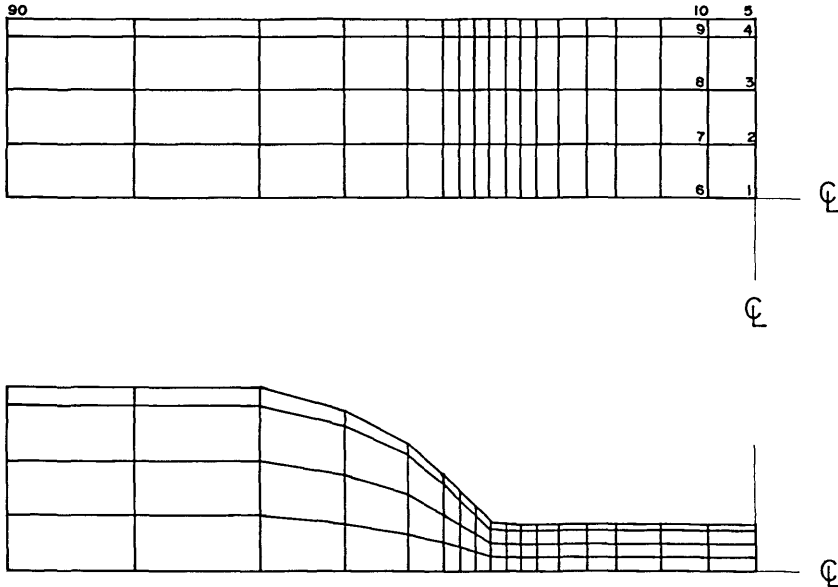
FIGURE 5. A two-dimensional reference brick before and after curving. Numbers indicate beginning and ending of global numbering scheme.

93

As a further simplification, we consider only bodies which have symmetry lines on the lower boundary and rightmost boundary as illustrated in figure 5. The transformation of the reference brick is generated by defining a cubic spline [24] which interpolates to given data defining the location of the upper boundary of the curved body (the leftmost boundary is not curved). This spline is used to define a transformation which varies linearly in the interior of the reference brick in such a way that the reference brick and curved body have the same axes of symmetry. This means that the coordinates, t_{ir}, of the nodal locations in the curved body can be obtained by multiplying the appropriate coordinate of the corresponding nodes in the reference brick by a factor which depends on the given cubic spline and the distance of that node from the symmetry line in the reference brick.

b. Three dimensions

The ϕ_r are trilinear polynomials on Ω_R, which is the cube such that $-1 \leq \xi^i \leq 1$.

$$\phi_r(\xi^1, \xi^2, \xi^3) = \frac{(\xi^1 + \xi_r^1)(\xi^2 + \xi_r^2)(\xi^3 + \xi_r^3)}{8\,\xi_r^1\,\xi_r^2\xi_r^3} \tag{72}$$

where ξ_r^i are the coordinates of the corners of Ω_R.

The t_{ir} are determined analogously to the way they are determined in 2-D. In 3-D, the back, rightmost and bottom surfaces of the reference brick are symmetry surfaces, as illustrated in figure 6. The leftmost surface is not curved. The remaining two surfaces are defined by interpolating splines and define transformations which vary linearly in the interior in such a way that the symmetry planes of reference and curved domain coincide. In the current code a further simplification is obtained by requiring that the surface splines be tensor products of one-dimensional splines. This restricts the type of surface which can be described but allows the use of a simple, one-dimensional spline code.
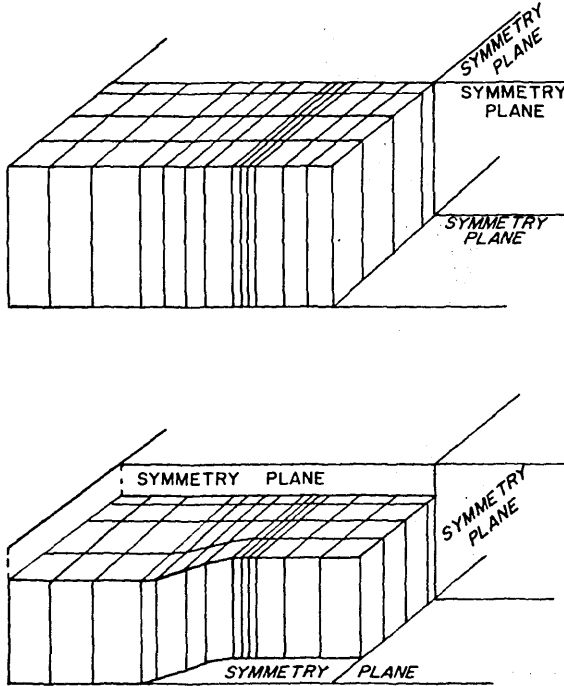
FIGURE 6. The three dimensional generalization of the curving procedure illustrated in figure 5.

3.4. Linear elasticity

The current code can compute finite element solutions to standard 2- and 3-D linear isotropic elasticity problems [22]. The same shape-functions and mesh generation are used as for the nonlinear problems. The linear solutions can be compared to small-strain nonlinear solutions. This allows determination of the load and/or strain levels at which the linear and nonlinear elastic models depart significantly in their predictions. It is also interesting to observe the nature of the differences in the linear and nonlinear solutions. A simple analysis shows that the Mooney constitutive laws (52) and (69) have shear modulus

$$\mu = 2(c_1 + c_2) \tag{73}$$

in the zero-strain limit. The incompressibility condition determines the other constitutive parameter of linear, isotropic elasticity which give a correspondence between the material properties of the linear and nonlinear models in the small-strain limit.

For comparison with nonlinear plane-stress solutions, it is natural to look at linear plane-stress solutions. As with the nonlinear constitutive laws, the linear plane-stress model enforces incompressibility exactly by assumption, without penalties. However, by a well-known re-interpretation of the material properties, every plane-stress solution is also a plane-strain solution [25]. Incompressible plane-strain requires a penalty to enforce zero volume change. Therefore the current code can be used to evaluate the behavior of two-dimensional elements with penalties, at least in linear elasticity.

a. Three-dimensional linear problems

Three-dimensional linear elasticity with a penalty to enforce incompressibility is treated exactly as described in [5]. In terms of the infinitesimal strain tensor $e_{ij} = 1/2(u_{i,j} + u_{j,i})$, the constitutive law is

$$\sigma_{ij} = 2\mu\{e_{ij} + (ze_{ii})\delta_{ij}\} \tag{74}$$

μ is the shear modulus and z is the penalty parameter

$$z = \frac{\nu}{1-2\nu} \tag{75}$$

ν is Poisson's ratio, which has the value 1/2 for exact incompressibility. By taking

$$h = \frac{z}{\nu} e_{ii} \tag{76}$$

equation (74) becomes

$$\sigma_{ij} = 2\mu \{e_{ij} + \nu h\delta_{ij}\} \tag{77}$$

which is the usual constitutive law for linear, isotropic elasticity, which is valid for incompressible material if $\nu = 1/2$ [1].

b. Linear plane-stress and plane-strain

The linear plane-strain constitutive law is used to model the cross-section of a body which is very long in the x^3-direction [25]. The problem is two-dimensional since it is assumed that $u_i(x^1,x^2,x^3) = u_i(x^1,x^2)$ and $e_{33} = 0$. Substituting this into (74–77) gives a 2-D constitutive law for which

$$h = \frac{z}{\nu} (e_{11} + e_{22}) \tag{78}$$

Then as $\nu \to 1/2$, z is a penalty parameter enforcing $e_{11} + e_{22} = 0$ in the incompressible limit. The energy expressions and the finite element model are very similar to 3-D linear elasticity, but $e_{33} = 0$, and there is no x^3-dependence. If we take $\alpha, \beta = 1,2$ the plane-strain constitutive law is

$$\sigma_{\alpha\beta} = 2\mu\{e_{\alpha\beta} + \frac{\nu}{1-2\nu} \delta_{\alpha\beta}e_{\gamma\gamma}\} \tag{79}$$

For any value of $\nu \neq \frac{1}{2}$, if we re-interpret a solution obtained using (79) by defining

$$\nu = \frac{\eta}{1 + \eta}$$

$$e_{33} = -\frac{\eta}{1 - \eta}(e_{11} + e_{22}) \tag{80}$$

it is easy to see that we have a linear plane-stress solution satisfying $\sigma_{33} = 0$, with apparent Poisson's ratio η, and shear modulus μ [25]. In particular, if we want the plane-stress solution to be incompressible, we choose $\eta = \frac{1}{2}$. This corresponds to $\nu = \frac{1}{3}$ in (79), or $z = 1$. The current code then can be used to obtain an incompressible plane-stress solution by setting $\nu = \frac{1}{3}$, and an incompressible plane-strain solution by the penalty approach, taking ν close to $\frac{1}{2}$, but not exactly equal to $\frac{1}{2}$. Since they involve quite different geometric assumptions, the two kinds of 2-D incompressible solutions usually are quite different. The same kind of comparisons between the plane-stress solutions and 3-D solutions are possible as were for nonlinear elasticity.

4. Nonlinear equation-solving procedures

One of the principles which has guided us in the construction of this finite element model is that it should retain as much of the structure of a linear elasticity model as possible. Other than in the assumption of somewhat simplified loading, this has been done without making any restriction on the degree of nonlinearity in the energy expressions. The point is that this can be done because the energy principle (6) which characterizes solutions in our chosen continuum model admits to the generalized Rayleigh-Ritz procedures described above. Because trial function are linear combinations of the basis functions, and because of the linearity of the integral in the energy expressions—features which are common to linear and nonlinear elastic models—the assembly of elemental expressions and much of the generation of these expressions are essentially linear procedures. Much of the current computer code is the same code which would be generated to solve linear elasticity problems alone.

To solve the nonlinear equations (24) or (25), many of the common procedures involve generation of a sequence of linearizations using the tangent stiffness matrix (27). A sequence of linear systems is solved with $\mathbf{K}_T$ on the left-hand side. We have seen that $\mathbf{K}_T$ has the same banded structure in the linear and nonlinear case. Furthermore, it is a consequence of variational calculus [26] that at the solution to (24) or (25), $\mathbf{K}_T$ is positive-definite (we assume our problem is well-posed). It is a guiding principle of the nonlinear iteration schemes described here to retain positive definiteness of $\mathbf{K}_T$ throughout the whole sequence of iterates. First of all, this will allow even more common code between the linear and nonlinear procedures: the same stable positive-definite, linear-equation solvers can be used. Second—and in our view more important—is that departure of the sequence of iterates from the region of positive-definiteness of $\mathbf{K}_T$ has serious implications for the stability of the whole solution process—both in a numerical and analytical sense. Loss of positive definiteness of $\mathbf{K}_T$ means that the numerical inversion procedure is less stable in the sense that small rounding errors are more likely to be magnified [23]. But potentially more serious is the fact that loss of positive-definiteness of $\mathbf{K}_T$ is an indication that the current iterate in the nonlinear scheme is outside the region of local convexity surrounding the desired solution [21]. The iteration scheme runs a danger of becoming "lost" and failing to converge.

The numerical results described in the next section show that the possibility of the iteration scheme becoming lost in a region of nonconvexity of the energy is sometimes a very real one—in spite of the precau-

tions we take. One of the most important conclusions which can be drawn from the current state of research into penalty methods is finite elasticity is a negative one: penalties greatly increase the likelihood of departure from the convex region. Before we discuss this further we describe the iteration schemes we employ.

4.1. Incremental loading

Iteration schemes require an "initial guess" solution at which to begin. The degree of nonlinearity in equation (24) or (25) is, to a large extent, dependent on the intensity of the loads as reflected in the magnitude of the load terms. A useful strategy in producing an initial guess at the desired load-level is to begin from the unloaded state with $u_i \equiv 0$, and increment the load gradually to the desired level. At each step of this "incremental loading" process [21,22], an approximate solution is generated which is an initial guess for the next load-level. When the final load level is achieved, the approximate solution at that load can be used as an initial guess for another iterative method which produces the final, refined solution. The strength of this strategy lies in the fact that for small enough load increments, the approximate solution at each load stays in a region of convexity of the energy. Furthermore, if the errors which accumulate during the loading process tend to cause departure from the region of convexity, the user has three options: (1) refine the load increment; (2) use the other iterative method to eliminate the accumulating errors at intermediate load-levels; or (3) combine (1) and (2).

Let u represent the nodal-value vector of displacements in (24) or (25). Let the strain energy terms in these equations be denoted by the vector $P(u)$ and the load terms by F. If $0 \leq \lambda \leq 1$ is a scalar, at any stage of the incremental loading process we solve the nonlinear equation

$$P(u + \Delta u) + (\lambda + \Delta\lambda)F = 0 \tag{81}$$

where $\Delta\lambda F$ is the load increment, and Δu is the change in the displacement field in moving from solution u at load-level λF to solution $u + \Delta u$ at load $(\lambda + \Delta\lambda)F$. Expanding $P(u + \Delta u)$ in a multivariate Taylor series and retaining only first-order terms in Δu, along with the definition of $\mathbf{K}_T = \mathbf{K}_T(u)$ in (27) leads to

$$P(u) + \mathbf{K}_T(u)\Delta u + \lambda F + \Delta\lambda F = 0 \tag{82}$$

If u is the exact solution at the previous load-level or u has been obtained by a previous step of incremental loading, $P(u) + \lambda F = 0[(\Delta\lambda)^2]$. This gives an algorithm with $0(\Delta\lambda)$ accumulated error:

$$
\begin{cases}
u^{(0)} = 0, \qquad \lambda_0 = 0 \\[2ex]
\lambda_i = \lambda_{i-1} + \Delta\lambda \\[2ex]
\Delta u = -\Delta\lambda \mathbf{K}_T^{-1}(u^{(i-1)})F \\[2ex]
u^{(i)} = u^{(i-1)} + \Delta u
\end{cases}
\tag{83}
$$

The iterations in (83) can be carried out until $\lambda = 1$, or interrupted before then so that the current $u^{(i)}$ can be replaced by a refined $u^{(i)}$. The algorithms which do the refinement (such as the Newton-Raphson method described in the next section) tend to be more costly than one step of incremental loading. Therefore in the attempt to remain in the region of convexity, a judicious balance between load-increment, $\Delta\lambda$, and number of intermediate refinements may be called for. In our experience with the Mooney-Rivlin constitutive law, convexity can be maintained with very large $\Delta\lambda$, so that no intermediate refinement is needed, in the 2-D plane-stress model. On the other hand, efficient computations in the 3-D case require a careful balance between choice of $\Delta\lambda$ and the number and frequency of intermediate refinements. This is due to the sensitivity of the incompressibility constraint equations to small volume changes [9,10].

4.2. Newton-Raphson iterations

Let λF denote the current load-level, for $0 < \lambda \leq 1$. If u is an initial guess at this load-level, obtained either by incremental loading or other means, we wish to compute a correction, Δu, so that

$$P(u + \Delta u) + \lambda F = 0 \tag{84}$$

Again making a Taylor series expansion for $P(u + \Delta u)$ and neglecting second order terms in Δu yields the Newton-Raphson method:

$$\begin{cases} u^{(0)} = u \\ \Delta u = - \mathbf{K}_T^{-1}(u^{(i-1)})[P(u^{(i-1)}) + \lambda F] \\ u^{(i)} = u^{(i-1)} + \Delta u \end{cases} \tag{85}$$

The neglect of second order terms in (85) means that several iterations may be needed. It can be shown that in any vector norm $\|u^{(i+1)} - u^{(i)}\| = 0(\|u^{(i)} - u^{(i-1)}\|^2)$ [27], if $u^{(i-1)}$ is "sufficiently close" to the exact solution. This means that convergence can be very rapid. But it also implies that no prior guarantees can be made if $u^{(i-1)}$ is far from the exact solution. The algorithm can get lost in a nonconvex region, even if it begins with u in the desired region of convexity. We have found that with the 2-D problems, almost any u obtained by incremental loading (even with $\Delta\lambda$ large) is sufficiently close, right from the start of (85), to initiate quadratic or near quadratic convergence. Furthermore, even if u is far enough from the solution so that the ultimate quadratic convergence rate is not achieved from the start, the iterates of (85) converge more slowly until $u^{(i-1)}$ is close enough to give quadratic convergence.

In 3-D problems, we find that care must be taken in choosing $\Delta\lambda$ so that the initial guess, u, is accurate enough for (85) not to get lost. If this is done successfully then the ultimate quadratic convergence rate is guaranteed. The dominant computational cost in (85) is the assembly of $\mathbf{K}_T$. 2-D problems are so cheap and Newton's method so robust that there is little to be gained over the employment of (83) and (85) as described here. But for 3-D problems with penalties, our experience indicates that the sensitivity of the iterative schemes and the dominant computational cost of assembling $\mathbf{K}_T$ warrant the investigation of other iteration schemes. A likely candidate may be "modified Newton's method" in which $\mathbf{K}_T$ is reassembled less often than once every step [21,22].

4.3. Interpolatory mesh refinement

Another way to generate an initial guess either for the next load increment in (83) or to initiate (85) is to take a solution to (24) or (25) which is essentially fully converged on a given mesh and interpolate that solution to a finer mesh. This is done by taking the values given by (12) at the new nodal locations to define nodal values of a new trial function in the form (12). Of course, the new trial function has more degrees of freedom, so in general it will not represent an equilibrium solution to (24) or (25). The hope is that it will be a good initial guess to such a solution. A strategy which has proved successful in our problems is to combine load increments (83) and any required refinements by Newton's method (85) on a crude mesh until $\lambda = 1$. At the final load-level we refine the mesh in several steps—each generating an initial guess for the next which is refined by Newton's method. We have not performed a critical comparison of this strategy with others which are possible. The interpolatory refinements could be mixed with iterations of (83) or (85) in a variety of ways, and it would take careful study to determine the most efficient strategy. The reason we have used this particular strategy is that it provides a sequence of refinements at the final load-level which can give some idea of convergence of the model as the mesh is refined.

Again we find that the 3-D model is much more sensitive to the mesh refinement scheme than the 2-D model, because of the penalty. The sensitivity is reflected in the fact that the interpolated initial guess is often outside of the region of convexity. As described in [10], this can be traced to the fact that volumetric integration points, ζ_r, of (25) and (29) on the refined mesh do not overlay the ζ_r on the cruder mesh. This leads to inaccurate computation of the pressure, h, via (41). A smoothing technique for h is described in [10],

which is based on the equivalence theorem of [8]. It is employed as a standard feature of the 3-D code and appears to work as long as the refinement is not too drastic.

4.4. The in-core solvers

The first priority in the choice of linear equation-solving procedures for the systems in (83) or (85) is that they take advantage of the symmetric, band-matrix structure of K_T, described in (30). We store K_T as $d(\omega + 1)$ vectors of length dN. Since it is hoped that the application of the procedures outlined earlier in this section will lead to positive-definite matrices, we employ elimination procedures with no interchanges [23,27–29]. There are several algorithms which can reduce the matrix to a compact triangular form which overlays the original matrix bands and require no additional storage space [23,27–29]. We employ two such algorithms. The first is the Cholesky method [23,27–29]; this will only work for positive-definite matrices. If any iterate of (83) or (85) departs from the convex region, the Cholesky method will fail, the iterations will cease, and our program will produce a suitable error message.

The Cholesky method is used unless it is determined that there is no reasonable load-refinement strategy to avoid departure from the convex region. In that case we employ the symmetric column elimination procedure described in [23]. This will usually provide the desired triangular decomposition, though, strictly speaking, the existence of the decomposition is not guaranteed without the use of interchanges [23,28]. Interchanges are not allowed if the band structure of the decomposition is to be preserved, and violation of band structure is incompatible with the goal of maximizing the size of K_T which can be decomposed in-core. Even when the decomposition without interchanges exists for an indefinite K_T, the decomposition process is potentially unstable in a numerical sense [28]. When this difficulty is combined with the tendency of Newton's method to become lost if convexity is lost, it means that the symmetric elimination must be used with caution and as a last resort, when convexity cannot be maintained. If the iteration procedures (83) and (85) survive the loss of convexity, and the iterates converge, eventually convexity must be regained [26]. When this happens, then there is no essential difference numerically or theoretically between the symmetric elimination and the Cholesky decomposition, as long as convexity is maintained. We find that in 3-D problems, we often cannot avoid loss of convexity in the iteration scheme. On the other hand, if the load increments, $\Delta\lambda$, are not too large and the mesh refinements not too drastic, then the loss of convexity can be survived and solutions obtained. Therefore the symmetric elimination provides an indispensible way of recovering from loss of convexity, which would be impossible with the Cholesky method.

5. The stretching of rubber sheets

To illustrate the type of problem the model described here is designed to solve, we give an example of the stretching of a rubber sheet. This is similar to a problem described in [21], except that we have used the geometry of an A.S.T.M. standard tensile-test specimen for rubber sheet [30]. Our major concern was to investigate the application of penalty methods to impose the incompressibility constraint; therefore the emphasis in the results reported here is on the success or failure of the modelling process. However, we have been able to make some observations—largely qualitative—about the nature of the strain field in such specimens.

5.1. The A.S.T.M. tensile-test specimen

Figure 7 illustrates a top view of the A.S.T.M. tensile-test specimen described in [30] for the testing of the failure strength of rubber sheets. The dimensions are given in the lower portion of the figure. These specimens are referred to as "dumbbells" by the A.S.T.M., but this does not imply axisymmetry. Rather, the specimens are cut from a sheet which is typically one to several millimeters thick. Therefore, the plane-stress assumption described earlier seems appropriate. The 3-D model assumes a thickness of 2.5 mm.

In the upper portion of figure 7, a fairly crude irregular mesh is depicted. Experimentation with various mesh subdivisions allows concentration of subdivision in areas where spatial variations of strain is the largest. The large elements are placed in regions where the deformation is close to simple extension, which can be represented exactly, even by large elements of the type described in section 3.3.
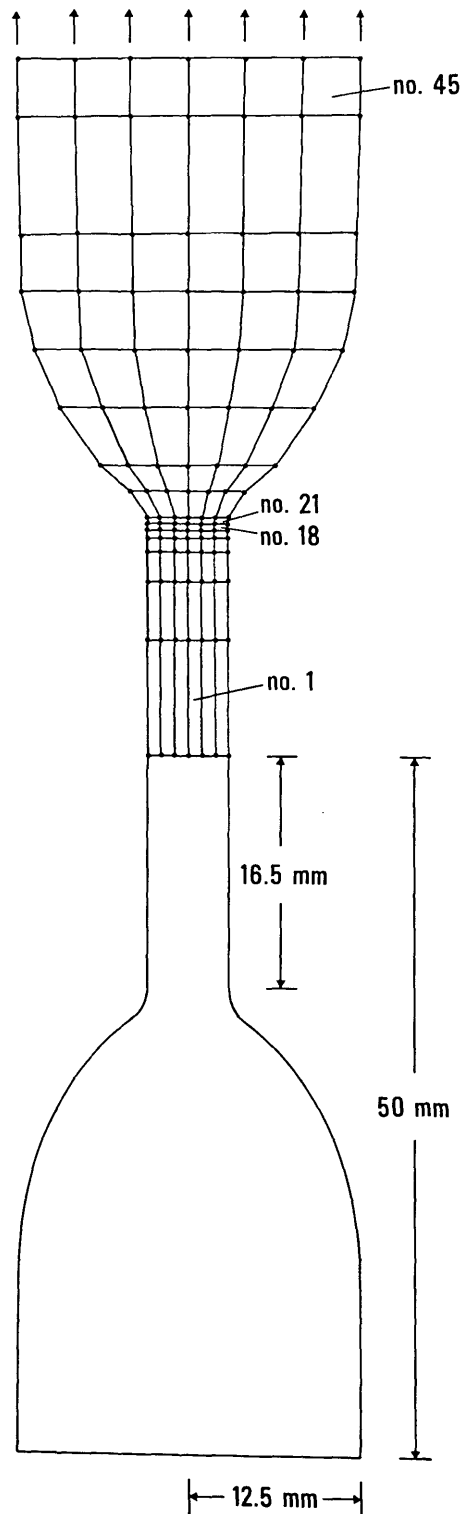
FIGURE 7. The A.S.T.M. tensile test specimen, lower
half showing Ω with exact dimensions. Upper half
shows a typical non-uniform finite element mesh giving Ω_σ.

5.2. Two- and three-dimensional models

a. Symmetry

As has been illustrated in figures 5 and 6, there are certain symmetry assumptions in the current implementation of our model. Both 2- and 3-D versions assume symmetry of the deformation between the upper and lower portions of figure 7, as implied by figures 5 and 6. Also as implied by figures 5 and 6, symmetry of solutions between right and left halves of figure 7 is assumed, so that in the 2-D case, only the upper right quadrant is discretized. In the 3-D model symmetry is assumed along the midplane as implied by figure 6, so that only one octant of the body is actually discretized.

b. Boundary conditions

The boundary conditions along symmetry lines and planes are enforced as they are in linear analysis [22,23]. Displacements into, and certain shears along the symmetry lines and planes are assumed to be zero. The physical boundary conditions are a fairly simplified idealization of possible attachments to physical gripping devices (for example, see [31]). The conditions amount to a simple specification of end displacements to give various total extensions of the specimen. In 3-D, this roughly corresponds to gluing the whole end of the specimen to a relatively rigid surface. In the 2-D model, since the only end displacement available is on the midplane, the 2-D boundary condition is roughly like gluing the centerline of the end of the specimen to a rigid surface. The rest of the end is free to slip on the support. The 2- and 3-D physical boundary conditions are illustrated in figure 8.

More realistic boundary conditions and interfacing with gripping devices can be envisaged, and could be incorporated in the model. In [32] methods for dealing with more realistic grips are described. However, for the purposes of testing penalty methods we did not feel that such considerations were crucially important.
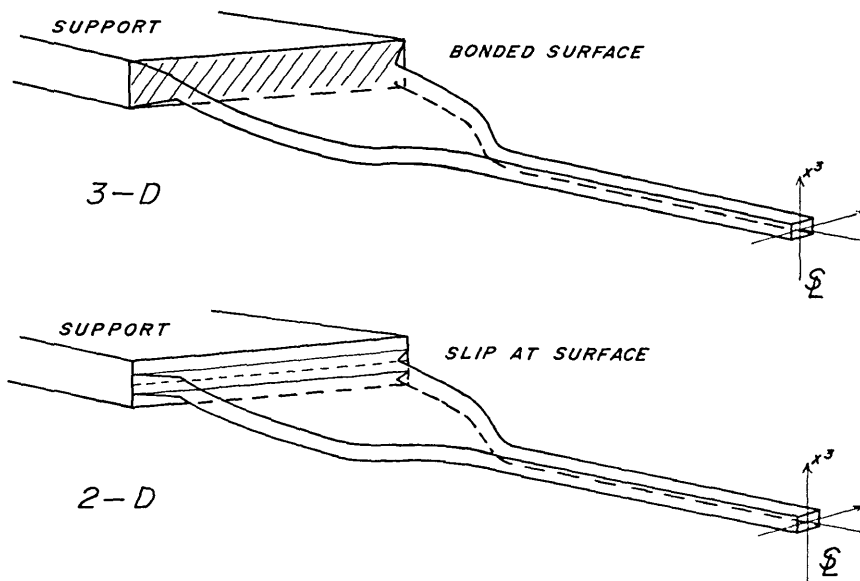
FIGURE 8. Boundary conditions for 3-D (upper)—bonding to a rigid support, and for 2-D (lower)—bonding of centerline only.

c. The use of penalties

The penalty method of section 2.2 was employed to impose near incompressibility. The Mooney-Rivlin law (52) was used with $c_2/c_1 = 0.2$ and $z = 50\,c_1$. The actual value of c_1 serves only to scale the stresses, so c_1 can be used to give appropriate dimensional units. This value of the penalty leads to a typical compressibility error of about 2-4 percent. For example, the most highly refined 3-D mesh of figure 7 has three elements

across the half-width and fifteen along the length and one through the thickness $(3 \times 15 \times 1)$. For the element in this mesh with the highest strain component, III = 1.0344 instead of III = 1.

A penalty is also used to enforce the displacement boundary condition on the end of the specimen [14]. If Γ_d denotes the surface of the specimen which is bonded to the support, then the following term is appended to the variational principle (6):

$$z_d \int_{\Gamma_d} (u_i - u_i^*)(u_i - u_i^*) ds \quad \text{(sum on } i\text{)} \tag{86}$$

where z_d is the boundary penalty and u_i^* is the specified displacement vector. In practice it has been found that z_d can be taken quite large. In the current examples we have taken $z_d = 10^4$. The boundary integral in (86) is easy to evaluate [22,23]. Employment of this boundary penalty means that all essential b.c. are homogenous [22,23].

5.3. Code verification procedures

Because exact solutions to the problem we describe here are not available, special attention must be paid to insuring that the computer code is functioning correctly. Of course, it is not really possible to give an ironclad guarantee of the correctness of the code, but we have used three tests which assure us, beyond reasonable doubt, that the results we present here are artifacts of the model and not coding blunders.

a. The small-strain limit

As described in section 3.4, our code has the capability to produce linearly elastic solutions with the same geometry as for the nonlinear results. A simple asymptotic analysis of the Mooney-Rivlin law and the variational principle (6) using that law shows that when squares of the components of $G_{ij} - \delta_{ij}$ are small enough to be neglected, linear elasticity and nonlinear elasticity should give nearly the same results. We found that our nonlinear code reproduced the 2- and 3-D linear displacements to 4 digits when the total extension of the specimen was on the order of 0.5 percent. The differences between the two solutions were typically in the second digit for extensions in the range of 1-2 percent and progressively more noticeable for higher extensions. These results are in total accordance with what is to be expected from the asymptotic analysis.

b. Quadratic convergence of Newton's method

The fact that the small-strain limit is obtained from the nonlinear code does not guarantee that the code is working correctly. If an error is made in the coding of energy-density terms of higher order in displacement-gradients, this might not affect the small-strain limit. In the course of debugging this code and other nonlinear codes, it has been our experience that the quadratic convergence of Newton's method described in section 4.2 can only be obtained if $\mathbf{K}_T(u)$ is derived and coded correctly. What the obtaining of quadratic convergence means is that the derivatives computed according to section 2.3 are the correct second derivatives (29) of the first derivatives (24) of the variational principle (6). It does not completely assure that the first derivatives are derived correctly. The fact that, when solutions are obtained, $\mathbf{K}_T(u)$ is invariably positive definite suggests that the first derivatives are also correct.

Table 1 shows the obtaining of quadratic convergence in a 2-D mesh (pt. A.) and a 3-D mesh (pt. B.). Part C. of table 1 shows an example of a sequence of iterations for a 3-D mesh becoming "lost." All values in the table are the square roots of the sum of the squares of the components of F_i^r. The 2-D results are from a 7×26 element mesh discretizing the first quadrant of the specimen, extended to a total extension of 200 percent. Ten load increments (83) of 20 percent were performed on a regular 3×5 mesh with no intermediate refinement. Then the 3×5 fully extended displacements were refined by Newton's method and interpolated to a 4×7 regular mesh. The same procedure was followed to proceed to a 5×10 mesh, a 5×12 mesh, then a 6×15 mesh, and finally these results were interpolated to the irregular 7×26 mesh. It is the refinement by Newton's method of these last interpolated results which is given in table 1A.

102

TABLE 1. *Quadratic Convergence of Newton's method as reflected in the decline of the rms residual of the equilibrium equations A and B. C shows iterations becoming lost.*

	A.	B.	C.
iter. no.	7×26 2-D	$3 \times 15 \times 1$ 3-D	$3 \times 8 \times 1$ 3-D
1	0.422×10^1	0.551×10^3	0.900×10^3
2	0.588×10^0	0.271×10^2	0.794×10^2
3	0.262×10^{-1}	0.665×10^1	0.529×10^5
4	0.652×10^{-4}	0.883×10^0	0.370×10^8
5	0.627×10^{-9}	0.108×10^0	0.130×10^8
6	0.445×10^{-14}	0.585×10^{-3}	0.176×10^8

Table 1B is obtained from the performance of 10 load increments of 10 percent on a $3 \times 8 \times 1$ element 3-D mesh, to give a total extension of 100 percent. Newton's method was used to refine the results after every two load increments. Interpolatory refinement with Newton iterations was used to go to a $3 \times 10 \times 1$ and then a $3 \times 12 \times 1$ regular mesh. Finally, the results were interpolated to the irregular $3 \times 15 \times 1$ mesh whose top view is given in figure 7. It is the refinement by Newton's method of these last interpolated results which is given in table 1B. The **LDL**^{T} column solver is required for the procedures involved in table 1B, since convexity is lost after the first pair of load increments. It is always regained by the Newton iteration, but is lost at the next load increments, and just after each interpolatory refinement.

Table 1C shows what happens when an attempt is made to carry out the load increments described for table 1B on a $3 \times 5 \times 1$ mesh instead of a $3 \times 8 \times 1$ mesh. The loss of convexity is so severe that when an attempt is made to interpolate the displacements to a $3 \times 8 \times 1$ mesh and refine by Newton's method the results of table 1C are obtained. The ability to obtain quadratic convergence with a careful choice of load increments leads us to believe that the code is correct, but that remaining close enough to the convex region is a delicate matter in the 3-D model.

c. Convergence with mesh refinement

Figure 9 shows values of the extensional strain J_{22} and shear strain J_{21} for a specimen extended to a total extension of 100 percent. The results are taken from a 2-D calculation with 26 elements along the length, with smaller elements concentrated at the top of the neck. The smallest elements are on the order of a quarter the size of the smallest elements in figure 7 and the larger elements of the refined 7×26 mesh are in the simple extension portion of the neck and are half the length of the corresponding elements in figure 7. Figure 9 plots values of J_{22} and J_{21} at a position indicated by the arrow on the cruder 3×15 mesh (both halves shown) of figure 10. This is where the strain-gradients are largest. With the same lengthwise subdivision of the specimen into 26 pieces, results are compared for meshes which have 3, 4, 6 and 7 elements across the neck. The 7×26 mesh is obtained from the 6×26 by subdividing the outer element in two. Also plotted on figure 9 are the results of the $3 \times 15 \times 1$ 3-D mesh, whose top view is given in figure 7.

The results show that to a large extent, the results from all the meshes overlay the same curve, with the largest deviations coming—as expected-from the $3 \times 15 \times 1$ and 3×26 meshes. Therefore we are led to conclude that convergence is taking place and that, qualitatively at least, even the cruder meshes give good results to graphical accuracy. A similar conclusion can be drawn from a comparison of tables 2 and 3 in which values of the pressure H of (70) for a 7×26 and 3×15 mesh are compared—likewise for a total extension of 100 percent.

We believe that the circumstantial evidence gathered from the obtaining of the small-strain limit, quadratic convergence of Newton's method, and convergence with mesh refinement make it highly unlikely that our numerical results are an artifact of coding blunders. We have run exactly parallel two- and three-dimensional versions of the 3×15 mesh of figure 7 ($3 \times 15 \times 1$ in the 3-D case). While we have not run as many 3-D convergence tests as 2-D convergence tests, we find that the agreement in displacements and strain is in general as good as the agreement illustrated in figure 9. We are convinced that convergence of displacements with mesh refinement is taking place in the 3-D model. What is happening to the pressures in table 4 is another—and very interesting—matter, which will be discussed below.
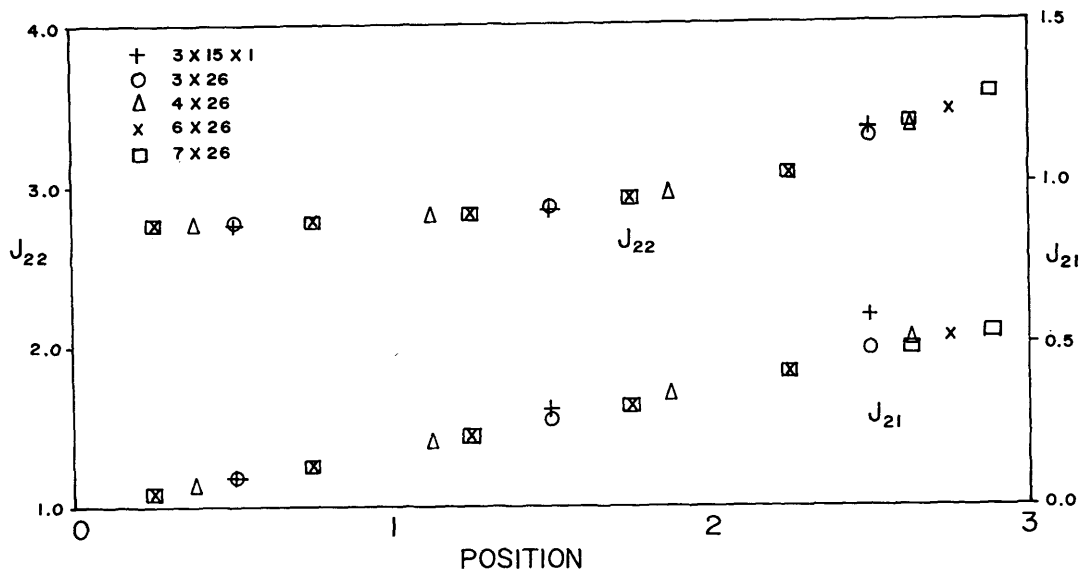
103

FIGURE 9. Extensional strain (upper) and shear-strain (lower) sampled at element centroids in stress-concentration region for various meshes. Values given on cross-section of neck.
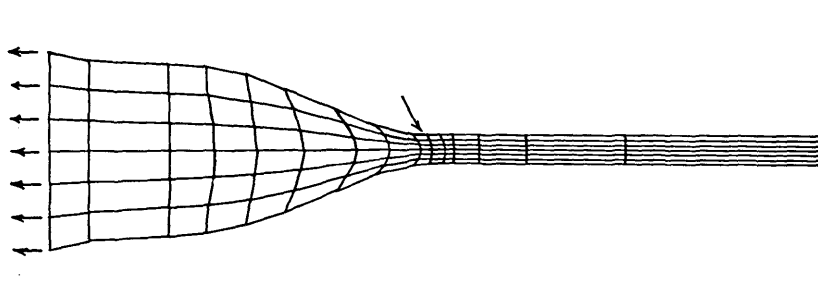

FIGURE10. Mesh of figure 7 deformed to total extension of 100 percent. Arrow marks stress-concentration.

TABLE 2. *Piecewise constant pressures along a cross-section of the neck in the stress-concentration region. Values taken from a refined 2-D mesh (7 × 26). Positions taken at center point of individual elements.*

Position	0.25	0.75	1.25	1.75	2.25	2.625	2.875
Pressure	0.537	0.533	0.523	0.508	0.480	0.452	0.438

TABLE 3. *Similar values to table 2, taken from a crude 2–D mesh.*

Position	0.50	1.50	2.50	Average value
Pressure	0.534	0.514	0.498	0.515

TABLE 4. *Values of the pressure taken from a 3–D mesh along cross-section near the stress-concentration region.*

Position	0.50	1.50	2.50	Average value
Pressure	0.621	0.454	0.563	0.546

104

5.4. Discussion of results

a. The plane-stress results

Our experience shows that the use of Newton's method with incremental loading described in section 4., when combined with the isoparametric plane-stress model described in this paper, produces a robust and computationally inexpensive algorithm. However, we must point out that the geometric assumptions and boundary conditions used in the current study advise a cautious interpretation of our results. We do think that we can draw some broad conclusions about the behavior of the A.S.T.M. speciments based on the qualitative picture which emerges from our 2-D results.

The picture which emerges is that the A.S.T.M. specimen geometry does what it is intended to do: namely concentrate the extension of the specimen in the neck region and away from the grips. The geometry does produce a region of concentrated deformation at the position of the arrow in figure 10, but our results indicate that the concentration is not severe. The data we cite to support these points, some of which are plotted in figure 9, are taken from our 7×26 mesh—not the cruder mesh of figure 10.

To illustrate the concentration of extension in the neck, we look at the value of the extension ratio, J_{22}, in the element nearest the center of the neck as compared to the total extension ratio of the specimen (based on grip separation). For 10 percent total extension of the specimen, the neck value of J_{22} is 7.2 percent higher than the total extension ratio of 1.10. For 100 percent total extension, the neck has a J_{22} 51 percent higher than the total extension ratio 2.00, and for total extension of 200 percent, J_{22} is 67 percent higher on the neck than the total extension ratio of 3.00.

Turning to the region of concentrated deformation, we find that there is a maximum of J_{22} at the point indicated by the arrow in figure 10. For 10 percent total extension, the maximum extension ratio is 5.5 percent higher than the extension ratio in the neck. For 100 percent total extension, the maximum extension ratio is 10 percent higher than in the neck, and for 200 percent total extension, the maximum is 20 percent higher than in the neck.

The extension ratio in the region of concentrated deformation seems to approach a value of 20 percent above the neck value in the limit as total extension is increased. The extension ratio J_{22} is not a rotationally invariant quantity, and one might raise the question as to whether the concentration of deformation grows unboundedly with total extension ratio by increasing shear-strain concentration. We look at the energy-density, which from (50), (51) and (64) can be seen to be $W(\mathrm{I}, \mathrm{II}) = c_1 (\mathrm{I}-3) + c_2 (\mathrm{II}-3)$, if one requires that the energy-density of the undeformed state is zero. Again we compare neck values of W with the maximum value. For 10 percent total extension W is 92 percent higher at its maximum than in the neck; for 100 percent total extension, it is 57 percent higher; and for 200 percent total extension, it is only 53 percent higher. So the deformation-energy concentration actually decreases in intensity with increasing extension ratio. Thus the shear-strain intensity cannot be growing unboundedly. In fact, if we look at the quantity $J_{21}/(J_{22}-1)$, we find that for 10 percent total extension, the maximum value is 27 percent above the value in the neck. For 100 percent total extension, the intensity is 21 percent over the neck value, and for 200 percent total extension, the intensity is 19 percent over the neck value. So in fact shear-deformation appears to become relatively less important as the total extension increases.

The A.S.T.M. standard [30] specifies that the total extension of the neck be measured visually, using marks on the neck. If the results we present here represent an adequate qualitative picture of the specimen, we would expect the specimen to fail in the neck, just below the wide portion. The A.S.T.M. standard requires that neck extension be recorded at the moment of failure. Our results indicate that there is a built-in safety factor of about 20 percent, in that the actual extension ratio near the point of faliure could have been about 20 percent higher than recorded. Shear deformation near the point of such a failure would not appear to be a major factor.

b. The three-dimensional results

For reasons discussed in detail earlier, the 3-D incremental loading/Newton algorithm combined with the isoparametric finite element/penalty model described here cannot be classified as a robust or computationally inexpensive algorithm. One would expect 3-D nonlinear finite element models to be dramatically more expensive than 2-D models in general. But one would also hope that the nonlinear iteration scheme could be made more reliable, so that situations like the one illustrated in table 1C do not lead to the waste

of expensive iterations. But there is a more basic theoretical question raised by the 3-D results, which should be answered before much effort is spent on the optimization of iteration schemes. This is a question of the stability and accuracy of the penalty or equivalent mixed formulation [8, 13–16].

Table 5 shows values of the pressure in a cross-section of elements near the stress-concentration region of a 3 × 10 × 1 mesh. The elements are rather large, and the centroids from which the pressures are taken are further down the neck towards the middle of the specimen than the sample points of tables 2 and 3, therefore the pressure is lower. But note that there is a smooth variation of pressure. Ths contrasts sharply with the oscillation of pressure across the cross-section in the stress-concentration region shown in table 4. These results are taken from the irregular 3 × 15 × 1 mesh whose top view is given in figure 7. We have already seen that J_{22} and J_{21} vary smoothly and are reasonably accurate in these elements from figure 9. In general, a detailed comparison of the 3 × 15 × 1 results shows that u_i and J_{ij} compare well with those from the 2-D 7×26 mesh—particularly in comparison to the contrast of the results in tables 2 and 4. What we observe there is that the 3-D pressures alone are affected by a "checkerboard" pattern of high and low values in the stress-concentration region. This is illustrated in figures 11 and 12. The pattern of high-low-high on one cross-section is countered by a reverse pattern on adjacent cross-sections. The average pressure value along a checkerboarded cross-section seems to agree much better with average values taken from the 2-D model, as is illustrated in figure 12 and tables 3 and 4. This leads us to speculate that averaging techniques may be able to remove the checkerboard.

TABLE 5. *Values of the pressure taken from a cross-section near the stress concentration for a cruder 3–D mesh.*

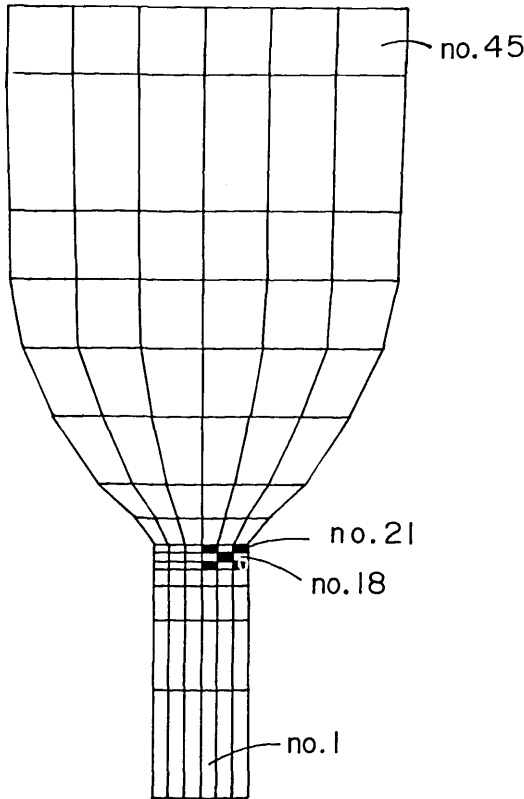| Position	0.50	1.50	2.50
Pressure	0.446	0.443	0.421

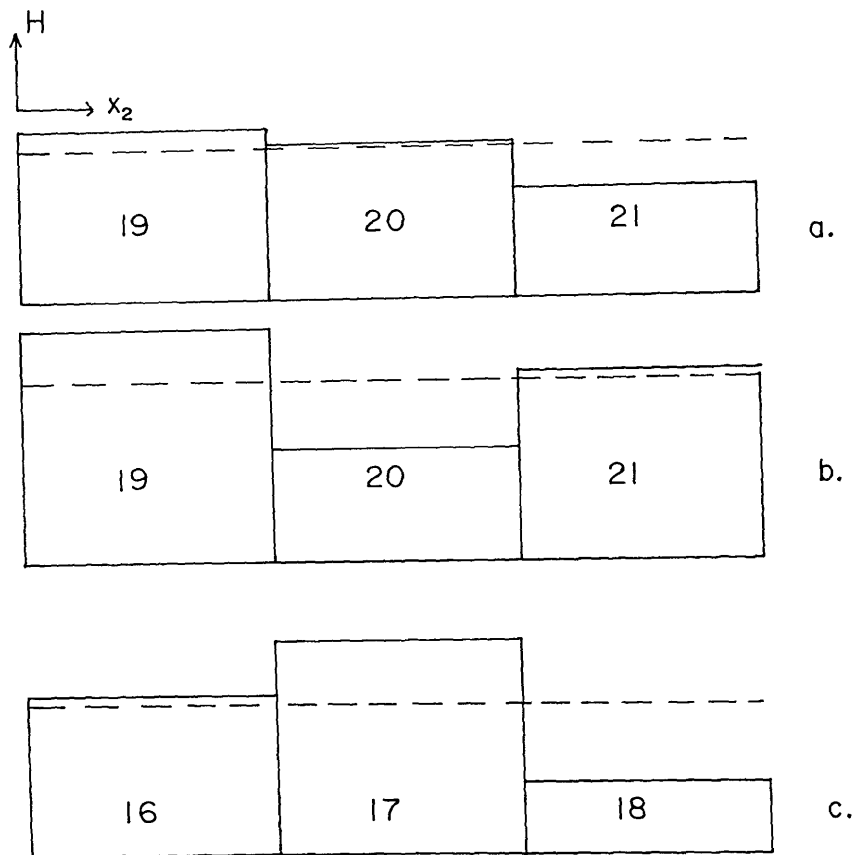FIGURE 11. Top view of a 3-D mesh—qualitative picture of checkerboard pressures.

106

FIGURE 12. Piecewise-constant pressures on a cross-section of the neck: (a) Accurate pressures; (b) checkerboard pressures; (c) checkerboard variation on adjacent cross-section to (b).

In view of the success obtained using this trilinear isoparametric element in a variety of penalty or mixed formulations [2, 7, 8, 11], we are led to conclude that the checkerboard pressure problem is a relatively rare occurrence. But, particularly since this phenomenon occurs when the mesh is refined, we are led to doubt that our element can satisfy the basic stability requirement of [13–16]. This matter is under investigation by the first author and others [14, 15, 34]. Techniques to overcome these difficulties using modified penalty procedures are also being investigated [18, 19, 33, 34]. Though the appearance of checkerboard pressure keeps us from drawing any physical conclusions based on our 3-D results, we are convinced that it will prove extremely useful to future theoretical investigations to be able to reliably reproduce this checkerboard phenomenon in such a basically simple problem.

6. Conclusions

The purpose in the development of the finite element model described here is to generate test problems for research on penalty methods in finite elasticity. As such it has proved very successful. In spite of the simplicity of the model, it has been able to isolate two important areas which will require further investigation if finite element penalty methods are to become productive tools in the analysis of incompressible materials under finite strain. The first area is that of the sensitivity of iteration schemes to small volume changes. Progress has been made in this direction, but still more is required [9, 10]. Perhaps more important is the question of element stability raised by the appearance of checkerboard pressure modes. The stability condition is known [13–17], but few elements seem to satisfy it—whether in penalty or equivalent mixed formulations [8–10]. On the other hand, even for elements which apparently do not satisfy it, the

107

checkerboard modes may be removable and often do not occur. More precise mathematical predictions need to be provided for these phenomena. It would be very useful to be able to aid such theoretical investigations with a code with more general element capabilities than the current one. Clearly the model admits to a variety of element types. Beyond that, new developments indicate that there are other penalty formulations which may be more stable than the reduced/selective integration techniques incorporated in the current model [19, 33, 34]. Therefore the model should be enhanced to include these possibilities.

Finally it is tempting to think of ways in which to make the code itself more sophisticated. The mesh generator could be based on the unions of several reference bricks, making the variety of bodies describable much wider. Load increment/refinement strategies could be made adaptive, in order to maintain convexity as much as possible without human intervention. More physically realistic boundary conditions and non-conservative loads could be incorporated. Many other constitutive laws fit within the framework of the model and would require only minor coding to implement. However, these enhancements belong further down the road, after some of the challenging questions raised by the current model are resolved.

---

7. References

[1] Herrmann, L. R., Elasticity equations for incompressible and nearly incompressible materials by a variational theorem, AIAA J. **3** 1896–1900 (1965).

[2] Hughes, T. J. R. and Allik, H., Finite elements for compressible and incompressible continua, Proc. Symp. Civil Engng., Vanderbilt Univ., Nashville, TN, 27–62 (1962).

[3] Fried, I., Finite element analysis of incompressible material by residual energy balancing. Int. J. Solids Structs. **10**, 993–1002 (1974).

[4] Malkus, D. S., Finite Element Analysis of Incompressible Solids, Ph.D. Thesis, Boston University, 1975.

[5] Malkus, D. S., A finite element displacement model valid for any value of the compressibility, Int. J. Solids Structs., **12**, 731–738 (1976).

[6] Malkus, D. S., Calculation of hole error by finite element methods, Proc. 8th Int. Cong. on Rheology, 618–619 (1977).

[7] Hughes, T. J. R., Equivalence of finite elements for nearly incompressible elasticity, J. Appl. Mech. **44**, 181–183 (1977).

[8] Malkus, D. S. and Hughes, T. J. R., Mixed finite element methods—reduced and selective integration: A unification of concepts, Comp. Meth. Appl. Mech. Engng., **15**, 63–81 (1978).

[9] Malkus, D. S., Penalty methods in the finite element analysis of fluids and structures, Nuc. Engng. Design, **57**, 441–448 (1980).

[10] Malkus, D. S., Finite elements with penalties in nonlinear elasticity, Int. J. Num. Meth. Engng., to appear.

[11] Hughes, T. J. R., Liu, W. K. and Brooks, A., Finite element analysis of incompressible viscous flows by the penalty function formulation, J. Comp. Phys., **30**, 1–60 (1979).

[12] Bernstein, B., Kadivar, M. K. and Malkus, D. S., Steady flow of memory fluids with finite elements: Two test problems, J. Comp. Meth. Appl. Mech. Engng., to appear.

[13] Babuska, I., Oden, J. T. and Lee, J. K., Mixed-hybrid finite element approximations of second-order elliptic boundary-value problems, Comp. Meth. Appl. Mech. Engng., **11**, 175–206 (1977).

[14] Bercovier, M., Perturbation of mixed variational problems. Application to mixed finite element methods, R.A.I.R.O. Analyse numerique/Numerical Analysis, **12**, 211–236 (1978).

[15] Oden, J. T., A theory of penalty methods for finite element approximations of highly nonlinear problems in continuum mechanics, Comp. Structs. **8**, 445–449 (1978).

[16] Malkus, D. S., Incompressible elements: The LBB condition and the discrete eigenstructure, Proc. U.S.-Europe Workshop on Finite Elements in Nonlinear Structural Mechanics, Springer-Verlag (1980).

[17] Hughes, T. J. R., Current trends in finite element research, Appl. Mech. Rev., to appear.

[18] Nagtegaal, J. C. and de Jong, J. E., Some computational aspects of elastic-plastic large strain analysis, research report M.S. 10.910, Technical University, Eindhoven, The Netherlands (1979).

[19] Cescotto, S. and Fonder, G., A finite element approach for large strains of nearly incompressible rubber-like materials, Int. J. Solids Structs. **15**, 589–605 (1979).

[20] Oden, J. T., Existence theory for a class of problems in nonlinear elasticity, J. Math. Anal. Applics. **69**, 51–83 (1979).

[21] Oden, J. T., *Finite Elements of Nonlinear Continua*, McGraw-Hill, New York (1972).

[22] Zienkiewicz, O. C., *The Finite Element Method*, Third Edition McGraw-Hill, London (1977).

[23] Bathe, K. J. and Wilson, E. L., *Numerical Methods in Finite Element Analysis,* Prentice-Hall, Englewood Cliffs, NJ (1976).

[24] de Boor, C., *A Practical Guide to Splines,* Springer-Verlag, New York (1978).

[25] Fung, Y, C., *Foundations of Solid Mechanics,* Prentice-Hall, Englewood Cliffs, NJ (1965).

[26] Gel'Fand, I. M. and Fomin, S. V., *Calculus of Variations,* Prentice-Hall, Englewood Cliffs, NJ (1963).

[27] Dahlquist, G. and Bjorck, A, *Numerical Methods,* Prentice-Hall, Englewood Cliffs, NJ (1974).

[28] Wilkinson, J. H., *The Algebraic Eigenvalue Problem,* Clarendon Press, Oxford (1965).

[29] Schwarz, H. R., Rutishauser, H., and Stiefel, E., *Numerical Analysis of Symmetric Matrices,* Prentice-Hall, Englewood Cliffs, NJ (1973).

[30] Standard test methods for rubber properties in tension (A.S.T.M. Standard Test Method D412–75), in 1976 Annual Book of ASTM Standards, Part 37, Rubber, Natural and Synthetic—General Test Methods; Carbon Black, American Society for Testing and Materials, Philadelphia, 79–80 (1976).

[31] Dewey, B. R., Finite element analysis of creep and plasticity tensile-test specimens, Experimental Mech. **16**, 16–20 (1976).

[32] Mitchell, R. A., Woolley, R. M., and Halsey, N., High-strength end fittings for FRP rod and rope, J. Eng. Mech. Div., ASCE, **100**, 687–706 (1974).

[33] Hughes, T. J. R., Generalization of selective integration procedures to anisotropic and nonlinear media, Int. J. Num. Meth. Engng., to appear.

[34] Hughes, T. J. R. and Malkus, D. S., manuscript in preparation on a generalized mixed method/penalty method equivalence theorem.

109