An interpolation scheme for precision reflection coefficient measurements at intermediate frequencies. Part 1: theoretical development

M G Cox, M P Dainton and N M Ridler
National Physical Laboratory
Teddingdon, Middlesex TW11 0LW, UK
Phone: +44 20 8943 6096, Fax: +44 20 8977 7091
Email: maurice.cox@npl.co.uk, URL: http://www.npl.co.uk

Abstract – The theoretical basis of an objective approach for impedance characteristic interpolation in the development of RF coaxial impedance standards is presented. Variants of the scheme apply to three impedances commonly used to calibrate vector reflectometers and automatic network analysers (ANAs). The scheme permits impedance standards to be characterised at intermediate frequencies in terms of their values and associated measurement uncertainties. A related paper is concerned with the practical implementation of the approach.

Keywords – constrained polynomial regression, data modelling, derivative information, empirical models, families of fitting functions, impedance characteristic interpolation, intermediate frequencies, measurement uncertainty, model fitting, model validation, physical models, reflection coefficients, RF coaxial frequency standards.

I. INTRODUCTION AND PROBLEM STATEMENT

In recent years, much work has been done to provide electrical impedance traceability at DC, low frequencies (LF), radio frequencies (RF) and microwave frequencies. However, a problem persists concerning suitable precision impedance measurement standards and methods for the frequency region between LF and RF, the so-called intermediate frequency (IF) region. The problem relates to the lack of a fully satisfactory physical model for an impedance at IF: the discrete circuit-element models used at DC and LF, and the transmission line models used at RF and microwave frequencies, both fail to describe adequately the electrical/electromagnetic interactions occurring at IF.

The recognition of the above difficulties has prompted the development of an alternative approach for characterising impedance standards at IF based on interpolating the impedance characteristics between “known” values at DC and RF. The physical justification for such an approach is based on the assumption that a device’s impedance characteristics, albeit unknown at any given frequency, can vary only smoothly over the IF region. Such an assumption is consistent with predictions based on Foster’s reactance theorem [1].

This paper describes the theoretical aspects of such an approach for this particular electromagnetic problem, concentrating on three impedances commonly used to calibrate vector reflectometers and automatic network analysers (ANAs). A related paper [2] presents the practical realisation of the interpolation scheme developed here, and presents results (in terms of reflection coefficient measurements) obtained using these schemes applied to an ANA calibration.

Previous work in the area is concerned with improving the traceability in the UK of coaxial impedance measurements at lower RF [3], with providing an uncertainty analysis for capacitance and dissipation factor characterisation at 1 MHz and 10 MHz for four terminal-pair devices (a different connector mechanism from the coaxial ones used at RF and microwave frequencies) [4], and with the sensitivity of a method to predict a frequency characteristic of a capacitor [5]. Related work is the Koffman analysis of four terminal-pair devices [6]. None of this work, however, led directly to an objective practical solution to the interpolation problem.

The approach that has been adopted is based on the objective analysis of calibration data and additional measurements (Section II). The theory of the approach (Section III) is concerned with modelling the data using a family of empirical functions. Physical information is also incorporated to help ensure that the result is physically feasible. The members of the family of models are obtained using enhancements (Section IV) of established least-squares fitting procedures. Model validation techniques are used (Section V) to select an appropriate model from the family and check its appropriateness. The required interpolated values are obtained from the selected model. Their uncertainties are evaluated (Section VI), taking into account both random and systematic error effects.

II. THE MEASUREMENT DATA

The raw or calibration data consists of:
- Frequencies \( f_i, i = 1, 2, ..., m \), spanning a range 100 MHz to (typically) 1000 MHz in steps of 50 MHz or 100 MHz,
- Corresponding measured complex-valued impedance-characteristic values \( \Gamma_i, i = 1, 2, ..., m \),
- Complex-valued quantities \( u_0, v_0, i = 1, 2, ..., m \), where \( 9u_0 \) and \( 3u_0 \) are respectively the standard uncertainty in the real and imaginary parts of \( \Gamma_i \) obtained from a Type A evaluation of uncertainty, and similarly for \( v_0 \) that from a Type B evaluation [7].
The nature of the measurements is such that for each value of \( i, R_{u_i} = 5u_i \) and \( R_{v_i} = 3v_i \). Fig. 1 and Fig. 2 show a calibration data set and the Type A evaluations of uncertainty for a matched load.

Additional data, the nature of which depends on the circuit type, is available as follows:

1. **Matched load.** The DC resistance \( r_L \) and its uncertainty \( u(r_L) \), and the characteristic impedance \( Z_0 \) (taken by definition as 50 \( \Omega \)). The value of \( \Gamma \) at DC, i.e., \( f = 0 \), and the corresponding uncertainty are formed from these values. This value of \( \Gamma \) is real.

2. **Open-circuit.** The shunt capacitance \( C_L \) and the offset length \( L_{o} \), and the characteristic admittance \( Y_0 \) (taken as \( 1/50 \) S).

3. **Short-circuit.** The series inductance \( L_{s} \) and the offset length \( L_s \), and the characteristic impedance \( Z_0 \) (taken as 50 \( \Omega \)).

Values for the constants in 2 and 3 are available from manufacturers' data sheets.

Finally, further, somewhat less reliable data is available, typically in the range 45 MHz to 100 MHz, with associated uncertainties. This data is unsuitable to be used as calibration data. It is useful, however, as secondary (validation) data in order to check the interpolated values provided by analysing the calibration data.

III. DATA MODELLING THEORY

The interpolation problem is to provide values of \( \Gamma \) at frequencies other than at which measurements are made. In other words, an interpolation procedure is required that delivers values at any point within the range of the data. This range is typically DC to 1000 MHz. In particular, interpolated values are required in the LF and IF region, i.e., up to 100 MHz, where measurement is difficult or possesses large uncertainties or both. It is reasonable to impose certain properties on the interpolation procedure:

1. Interpolated values obtained at the frequencies at which measurements were made are to reproduce the measured values of \( \Gamma \) to an extent consistent with their uncertainties.
2. Interpolated values are to be consistent with the calibration data and their uncertainties.
3. As far as possible, consistent with Property 2, the set of interpolated values corresponding to a fine subdivision of the range of the data lie on a smooth curve.

An approach to interpolation has been developed that is based on validated modelling of the data. It consists of the following steps:

1. Construct a family of models of the data.
2. Compare the deviations of the models from the calibration data with the input data uncertainties.
3. If possible, select a valid model from the family, i.e., one that satisfies the above properties. Otherwise, select a model that is satisfactory on some other basis, or investigate the reasons for failure.\(^1\)
4. Further validate the model by comparing the interpolated values produced from it with the secondary data.

\(^1\)The determination of a valid model is not always possible. For instance, examination of Fig 2 indicates the difficulty of constructing a smooth curve that provides a reasonable representation of the data in that the curve intersects some 95% of the 95% coverage intervals associated with the Type A evaluation of uncertainty for each calibration point. Such a property is expected if the errors characterised by the Type A evaluation of uncertainty can genuinely be regarded as random variables in the manner considered. The inability to obtain a valid model can be attributed to inconsistencies within the calibration data that ideally require investigation.
The family of models selected is part-physical, part-empirical. In particular, it is based as far as possible on physical information available about the system. This physical information is augmented by empirical terms that attempt to account for those aspects of the data that the current state of physical knowledge in the area is unable to explain. The choice of the number of these terms helps to achieve a model that constitutes a balance between closeness to the data and smoothness.

For the open- and short-circuits, a physical model has been developed [8] that provides information on the behaviour of the impedance characteristic $I$ in the neighbourhood of the origin ($f = 0$). By differentiating this physical model, expressions are obtained for the derivatives of $I$ at $f = 0$. The substitution into these expressions of the prescribed numerical values of $Y_0$, $C_0$, and $L$, in the case of an open-circuit, and $Z_0$, $L_0$, and $L$, for a short-circuit, provides numerical values that $I$ and its derivatives can be expected to assume at $f = 0$. The expression for these derivatives for the open-circuit (that for the short-circuit being similar) is [8]

$$\frac{d^r I}{df^r}|_{f=0} = \left(-\frac{4\pi}{c}\right)^r \left(\frac{r}{2\xi_0 (\xi - k)}\right)^{1/2} \left(C_0 \xi\right)^{1/2} \left(\frac{r}{2\xi_0}\right)^{-r} \left(1 - \frac{r}{\xi_0}\right)^{-r} = \alpha_r, \ldots (1)$$

where $c$ is equal to the speed of light in air at 23°C and 50% relative humidity. For both circuits, the derivatives of even order are real and those of odd order imaginary.

To motivate the subsequent discussion, consider a reduced problem in which the only data available is a value for $I$ at $f = 0$ and a further value at $f = 100$ MHz. The best estimate of $I$ at some frequency between zero and 100 MHz in light of this information alone is given by linear interpolation (a polynomial of degree one). However, if the true behaviour in the interval is nonlinear, this estimate will be biased. Suppose now that additionally the first derivative is given at $f = 0$. An improved estimate of $I$ at an intermediate point is given by a function that satisfies the value and first-derivative conditions at $f = 0$ and the value at $f = 100$ MHz. One simple form for this function is a polynomial of degree two. If all derivatives up to order $r$ are available at $f = 0$, a polynomial of degree $r + 1$ satisfying the information and the condition at $f = 100$ MHz can be constructed. If the information from which these derivatives is obtained is reliable, it can be expected that the greater the value of $r$ the better the interpolated value. This physical model from which these derivates are obtained is not perfect and hence, since differentiation exaggerates imperfection, it is appropriate to be able to select the number of derivatives that is to be satisfied.

For a matched load no suitable physical model is available. The only information available at $f = 0$ is the value of $I$ there (and its uncertainty).

For all circuit types, values of $I$ at frequencies higher than 100 MHz are also available, as described. It is appropriate to derive a function that satisfies a prescribed number of derivatives at $f = 0$ and that approximates in some best-fitting sense all other values of $I$. A family of models is therefore constructed that satisfies these derivative conditions whilst giving flexibility to represent the data at frequencies of 100 MHz and higher to various degrees of approximation.

For an open- or a short-circuit the above approach has two levels of flexibility, viz., the number $p$ of derivative conditions to be used at $f = 0$ and the degree $n$ of the polynomial. For a matched load, only $n$ is at choice.

Recommendations concerning the choice of $p$ are available [2]. Also see Section V. For any value of $p$, the derivative conditions to be imposed are calculated from the provided physical constants (Section IV). Model-validation techniques permit an appropriate value of $n$ to be selected (Section V).

In carrying out the modelling process, models based on polynomials of all degrees up to and including a specified degree are determined. This family of models is used as the basis for selecting a sensible degree. For each degree in the range, the model is constructed such that the polynomial coefficients are the best in the sense of least squares subject to the model’s satisfying the conditions at $f = 0$. A weighted least-squares fit is used to account properly for the calibration data uncertainties. Since the impedance data is complex-valued, the model will be complex-valued and hence the polynomial coefficients within it also.

IV. FORMULATION AND SOLUTION OF THE MODEL-FITTING PROBLEM

The algorithm used for constructing the family of models is an enhancement of conventional least-squares polynomial regression based on the following considerations:

1. Complex rather than real values of the dependent variable are permitted.
2. Weights are incorporated to reflect the data uncertainties [9].
3. The derivative conditions are imposed as constraints [9].
4. The independent variable is transformed linearly (i.e., "coded") for greater numerical stability [10].
5. Chebyshev polynomials rather than monomials ("powers of $f$") are used, also for greater numerical stability [9].
6. Models corresponding to all polynomial degrees are determined "simultaneously" [11].

Model fits of all degrees up to a prescribed maximum $n_{max}$ (less than the number of distinct frequency values $f_d$) are determined, to form a model family, all fits being constrained to have specified values and leading-derivative values at the origin. The computed polynomials are presented in a mixed monomial/Chebyshev-series form. Specifically, each member in the family of model functions is represented as

$$F(f) = \mu(f) + \nu(f)P(f),$$

where $\mu(f)$ is a complex-valued "constraint" polynomial that satisfies the specified conditions at
\( f = 0 \), \( \nu(f) \) is a real-valued "zeroising" polynomial, and \( P(f) \) is a polynomial with complex-valued adjustable coefficients.

For \( p \) leading-derivative values, including the value itself (the value of the derivative of order zero) specified at the origin, \( \mu(f) \) is a polynomial in monomial form which satisfies these conditions, \( \nu(f) \) is the zeroising polynomial \( f^p \), and \( P(f) \) is a polynomial in Chebyshev-series form in the normalised variable \( 2f/f_{\text{max}} - 1 \), where \( f_{\text{max}} \) is the largest data value of \( f \).

The monomial (Taylor) representation of the constraint polynomial is simply

\[
\mu(f) = \sum_{j=0}^{p} \frac{d_j}{j!} f^j, \quad \text{(2)}
\]

as can be verified immediately, since, for \( k = 0, 1, \ldots, p-1 \),

\[
\frac{d^k \mu}{df^k} \bigg|_{f=0} = \sum_{j=k}^{p} \frac{d_j}{j!} \binom{j}{k} f^{j-k} = d_k, \quad \text{(3)}
\]

The (Chebyshev) coefficients of \( P(f) \) are determined to minimise, in the least-squares sense, the departure of \( F(f) \) from the calibration data \( I \)-values.

Denote by \( P_n(f) \), the polynomial \( P(f) \) of degree \( n \), with complex-valued coefficients \( a_n = [a_{0n}, a_{1n}, \ldots, a_{nn}]^T \), i.e.,

\[
P_n(f) = \sum_{j=0}^{n} a_{jn} T_j(2f - f_{\text{max}} - 1), \quad \text{(4)}
\]

where \( T_j(x) \) is the Chebyshev polynomial of the first kind of degree \( j \) with argument \( x \).

The problem of determining these coefficients can be formulated as

\[
\min_{a_n} S = S(a_n) = \sum_{i=1}^{m} w_i^2 |e_i|^2, \quad \text{(5)}
\]

where the \( i \)th residual \( e_i = I_i - P_n(f) \), with weights \( w_i \) for \( i = 1, 2, \ldots, m \). The weights are a necessary part of the fitting process to ensure that the relative accuracies of the data values are properly reflected in the solution, and that the measurement uncertainties subsequently calculated are valid.

The coefficients of \( P_n(f) \) are thus given by fitting the function \( \nu(f)P_n(f) \) to the modified data points \( (f, I_i) \), where \( \Gamma_i = I_i - \nu(f)P_n(f) \), with weights \( w_i \), for \( i = 1, 2, \ldots, m \). The problem is therefore identical to conventional least-squares polynomial fitting except that (i) the data is weighted, (ii) the data ordinates and the polynomial coefficients are complex-valued, and (iii) the fit is constrained, viz., that in place of \( T_j(2f - f_{\text{max}} - 1), j = 0, 1, \ldots, n \), as basis functions, the functions \( \nu(f)T_j(2f - f_{\text{max}} - 1), j = 0, 1, \ldots, n \), are used.

Full details of the fitting algorithm are available [8].

V. MODEL VALIDATION

Model-validation techniques are applied in order that as far as possible the fitting function provides an appropriate model for the data. These techniques are based on fitting a family of candidate models to the data, and the selection of the most suitable model from the family. The family constitutes the models \( P_n(f) = \mu(f) + \nu(f)P_n(f) \), for \( n = 0, 1, \ldots, n_{\text{max}} \). For each \( n \), the residuals \( e_i \) are formed, as above, and the corresponding value of the weighted root-mean-square residual calculated from \( s = \{S(m - n - 1)^{1/2} \} \).

The values of \( s \) tend to decrease as \( n \) is increased until they essentially "saturate" at an approximately constant value. If this constant value is (close to) unity, a valid solution has been obtained and the smallest value of \( n \) corresponding to the saturation level is chosen. (If \( s \leq 1 \), the fit to the data is poor (under-fitting). Conversely, if \( s \geq 1 \), the fit to the data is "too good" (over-fitting), and follows the noise in the data too closely.)

Any choice of degree based on this criterion should also take into account the fact that the model might not be faithful to the data and vice versa. The model might be physically or empirically inadequate in providing an appropriate description of the data. The data might possess features that are inconsistent with any reasonable choice of (smooth) model. In any case the visual appearance of the sequence of fitted model curves in relation to the calibration data is taken into account.

A candidate model should also be validated by comparing the values it produces at the frequencies at which the calibration data was obtained. In particular, the model residuals should be compared with the provided Type A evaluations of uncertainty at each point.

Experiments were made with the number \( p \) of derivative conditions incorporated in the model in respect of the influence on the solution. The results obtained have been summarised [8]. Together with other tests of the approach, they demonstrate that there is advantage in utilising some derivative information, but a distinct disadvantage in using too much. Typically, taking \( p = 1 \) or \( p = 2 \) proves to be best.

Further validation is carried out by evaluating the model at the complete set of points comprising the calibration points and the secondary points. In particular, the model residuals for this composite set are contrasted with the expanded uncertainties based on the Type A and Type B evaluations. 95% of the fitted values are expected to fall within these uncertainty bars.

A final form of validation was considered for representative calibration data sets. Instead of a family of "pure" polynomials, polynomials in a transformed variable and other model
families were studied. It was found that these alternatives conferred no significant advantage.

The numerical software used to implement the method was itself validated using an existing methodology [12].

Results obtained using the approach described here are included in Part 2 of this paper [2].

VI. UNCERTAINTY EVALUATION

For the interpolated value provided at each prescribed frequency a statement of the uncertainty in this value is required. Corresponding to each point on the fitting function a standard uncertainty can be evaluated from the results of the fitting process. This standard uncertainty, termed colloquially a random uncertainty, corresponds to a Type A evaluation of uncertainty [7]. The data also contains the systematic uncertainties indicated in Section II, characterised by Type B evaluations of uncertainty. At each frequency corresponding to a specified calibration data value the combination in quadrature of the given systematic uncertainty at that point and the computed random uncertainty at that point can be expected to yield an improved combined standard uncertainty at that point. By taking the product of this combined standard uncertainty and an appropriate coverage factor [7], an expanded uncertainty that can be regarded as, say, a 95% coverage interval for the point, can be obtained. The locus of the lower endpoints of such coverage intervals define a lower coverage envelope for the fitted function and that of the upper endpoints an upper coverage envelope.

The standard uncertainties of values of the fitted function can be calculated using standard material [11]. Formally, their calculation involves deriving the covariance matrix of the parameters of the fitting function, followed by the propagation of this information to provide a standard uncertainty at each point (which will usually be different from point to point).

Consider that an acceptable determination of a model for the data has been made. That is, a validated mathematical representation of the calibration data has been obtained. Suppose that the model has been used to calculate interpolated values at a required set of frequency values. The procedure adopted for evaluating the uncertainties of those values is consistent with the Guide to the Expression of Uncertainty in Measurement [7], and in outline is:

1. Form the covariance matrix of the model parameters from the information provided by the fitting algorithm.
2. For any required frequency:
   2.1 Use this covariance matrix to provide the Type A evaluation of uncertainty of the model for that frequency.
   2.2 Provide the Type B evaluation of uncertainty at this frequency by interpolation in the Type B evaluations of uncertainty provided at the calibration points.
   2.3 Combine in quadrature the Type A and Type B evaluations of uncertainty to obtain the combined standard uncertainty in the interpolated value and hence obtain the expanded uncertainty at the specified frequency.

Full details of this procedure are available [8].

VII. SUMMARY AND FURTHER COMMENTS

The work described here provides an objective theoretical basis for an interpolation scheme for precision reflection coefficient measurements at intermediate frequencies. The approach, which applies to three impedances (matched load, open-circuit and short-circuit), employs calibration data, additional data and physical knowledge. A family of physically empirical fitting models, having smooth behaviour, is employed from which, in any particular case, the most suitable member is chosen on a statistical and visual basis. The means by which the fitting functions are obtained constitutes an enhancement of established polynomial regression procedures. Further validation checks are used to help ensure that the choice of model is appropriate. The required interpolated values are obtained by evaluating this model at the specified frequencies. The Type A evaluations of uncertainty in the calibration data are propagated through the modelling process in order to establish the corresponding contributions to the interpolated values. These contributions are aggregated with the Type B evaluations of uncertainty from the calibration in order to provide combined standard uncertainties of the interpolated values and hence the uncertainties corresponding to a 95% coverage probability. The sequel [2] to this paper addresses the practical implementation of the scheme.

The uncertainty structure (Section II) of the calibration data implies that the real and imaginary parts of the values of \( \Gamma \) are uncorrelated. As a consequence the real parts of the polynomial coefficients in the model representation of the reflection coefficient are uncorrelated with the imaginary parts. This fact is taken into account when providing the uncertainties of the interpolated values and means that the errors in the real and imaginary parts of the impedance standards so realised are mutually independent, a fact that is relevant to further processing of this information. An extension of the scheme permits more general input uncertainty structure, which
would lead to correlations between the real and imaginary parts of the realised impedance standards.

ACKNOWLEDGEMENT

The work described in this paper was funded by the National Measurement System Policy Unit of the UK Government's Department of Trade and Industry.

REFERENCES


