IV. THE GENERAL VARIATIONAL APPROACH

The general problem that will be considered here is the estimation of a function $P(x)$ from an accurate but incomplete set of facts concerning various properties of $P(x)$. The estimation procedure involves choosing an extremal principle and then finding the $P(x)$ that satisfies the principle under the constraint that $P(x)$ agrees fully with all knowledge about $P(x)$.

A. The General Equations

Let the variational principle be the desire to achieve an extremal for the integral

$$\int V\,[\,P(x),\,x\,]\,dx\,. \tag{IV-1}$$

The information that is directly known about $P(x)$ is contained in the N equations

$$\int G_n[\,P(x),\,x\,]\,dx = \gamma_n\,, \quad n = 1 \text{ to } N\,. \tag{IV-2}$$

In both cases, x may be multidimensional and the integrals are over the same specified space.

To solve this problem we shall use Lagrange multipliers, λ_n. Thus, with $A = \int dx$, we need

$$\delta \int \left\{ V[\,P(x),\,x\,] - \sum_{n=1}^{N} \lambda_n [G_n[\,P(x),\,x\,] - \lambda_n/A\,] \right\} dx = 0\,, \text{ or}$$

$$\int \left\{ V'[\,P(x),\,x\,] - \sum_{n=1}^{N} \lambda_n\, G'_n\,[\,P(x),\,x\,] \right\} \delta P(x)\, dx = 0\,.$$

Thus an extremal is attained when

$$V' [P(x), x] = \sum_{n=1}^{N} \lambda_n G'_n [P(x), x] . \qquad \text{(IV-3)}$$

The prime on V and G_n indicate the derivatives with respect to P(x) holding the explicit dependence of V and G_n on x constant. The Lagrange multipliers, λ_n , are to be chosen so that equations (IV-2) are satisfied.

B. A Solution Procedure

The solving of (IV-2) and (IV-3) will in general require an iterative solution technique. An iterative technique which will always solve the problem in many important cases is given below.

Step 1: Starting with some set of values for λ_n , n=1 to N , solve (IV-3) for P(x) . This step in itself may be a difficult iteration problem since (IV-3) is an implicit function of P(x) . Also, the chosen values for λ_n may not allow a reasonable solution for P(x) . One method of alleviating these problems is to start by imposing only one of the constraints on P(x) and going through the procedure steps until a solution is obtained. Then, starting with that solution, add another constraint to the problem and go through the procedure steps again, etc.

Step 2: Using the derived P(x) , calculate the values of the N integrals (IV-2) , i.e.,

$$g_n = \int G_n [P(x), x] dx . \qquad \text{(IV-4)}$$

If $g_n = \gamma_n$ for $n=1$ to N, then we have an exact solution. Normally, however, the g_n will not be equal to the γ_n and we will need to change the λ_n so that the g_n become closer to the γ_n. To do this, we can write the differential of g_n with respect to the λ_n as

$$dg_n = \sum_{s=1}^{N} \frac{\partial g_n}{\partial \lambda_s} d\lambda_s , \quad n=1 \text{ to } N ,$$

or in matrix form

$$\begin{Bmatrix} dg_1 \\ dg_2 \\ \vdots \\ dg_N \end{Bmatrix} = \begin{bmatrix} H_{11} & H_{12} & & H_{1N} \\ H_{21} & H_{22} & & H_{2N} \\ - & - & - & - \\ H_{N1} & H_{N2} & & H_{NN} \end{bmatrix} \begin{Bmatrix} d\lambda_1 \\ d\lambda_2 \\ \vdots \\ d\lambda_N \end{Bmatrix} , \qquad \text{(IV-5)}$$

where $H_{ns} = \partial g_n / \partial \lambda_s$. Letting $\varepsilon_n = \gamma_n - g_n$, one can estimate the change $\Delta\lambda_n$ in λ_n necessary to make $g_n = \gamma_n$ by solving the matrix equation

$$\underline{\Delta\lambda} = H^{-1} \underline{\varepsilon} \qquad \text{(IV-6)}$$

where $\underline{\Delta\lambda}$ and $\underline{\varepsilon}$ are column vectors and H is the square matrix in (IV-5).

Step 3: Replace λ_n by $\lambda_n + \alpha \Delta \lambda_n$, where $0 \le \alpha \le 1$, and go to steps 1 and 2 again, etc., until the length of the error vector $\underline{\varepsilon}$ is small enough. Here, the vector $\underline{\Delta\lambda}$ gives us a direction to move in to reduce the length of the error vector $\underline{\varepsilon}$. However, if α is too large, we may overshoot and the length of $\underline{\varepsilon}$ may increase. During the early steps in the iteration, small values for α may be necessary for convergence. During the later iterations, α can usually be set to unity.

To solve for $\partial g_n / \partial \lambda_s$, one takes the partial derivatives of (IV-3) and (IV-4) with respect to λ_s to get

$$V''[\,P(x),\,x\,]\,\frac{\partial P(x)}{\partial \lambda_s} = G_s'[\,P(x),\,x\,] + \sum_{n=1}^{N} \lambda_n G_n''[\,P(x),x\,]\,\frac{\partial P(x)}{\partial \lambda_s}$$

or

$$\frac{\partial P(x)}{\partial \lambda_s} = \frac{G_s'[\,P(x),\,x\,]}{V''[\,P(x),\,x\,] - \sum_{n=1}^{N} \lambda_n G_n''[\,P(x),\,x\,]} \qquad \text{(IV-7)}$$

and

$$\frac{\partial g_n}{\partial \lambda_s} = \int G_n'[\,P(x),\,x\,]\,\frac{\partial P(x)}{\partial \lambda_s}\,dx\,. \qquad \text{(IV-8)}$$

Putting (IV-7) into (IV-8) we have

$$\frac{\partial g_n}{\partial \lambda_s} = \int \frac{G_n'[\,P(x),\,x\,]\,G_s'[\,P(x),\,x\,]}{V''[\,P(x),\,x\,] - \sum_{n=1}^{N} \lambda_n G_n''[\,P(x),\,x\,]}\,dx\,. \qquad \text{(IV-9)}$$

If we let

$$Q[P(x)] = V[\,P(x),\,x\,] - \sum_{n=1}^{N} \lambda_n\,[\,G_n[\,P(x),x\,] - \lambda_n/A\,]\,, \qquad \text{(IV-10)}$$

we see from (IV-3) that our desired solution occurs when $Q'[P(x)] = 0$. From (IV-9), we see that $Q''[P(x)]$ is the denominator in the integral. Now if, for all values of x that are being integrated over,

$Q''[P(x)] > 0$ and no non-zero linear combination of the G_n functions is a constant, then we can make the powerful assertion that the matrix H is positive definite. This is easily proved by remembering that $H_{ns} = \partial g_n / \partial \lambda_s$ and that H is positive definite if and only if $\underline{a}^T H \underline{a} > 0$ when $\underline{a} \neq \underline{0}$, where the superscript T indicates the transpose. Using (IV-9), we see that

$$\underline{a}^T H \underline{a} = \int \frac{\sum_{n=1}^{N} a_n G_n'[P(x), x] \sum_{s=1}^{N} a_s G_s'[P(x), x]}{V''[P(x), x] - \sum_{n=1}^{N} \lambda_n G_n''[P(x), x]} dx$$

or

$$\underline{a}^T H \underline{a} = \int \frac{\left\{ \sum_{n=1}^{N} a_n G_n'[P(x), x] \right\}^2}{Q'[P(x)]} dx > 0 ,$$

if $Q''[P(x)] > 0$ and $a \neq 0$ since the integrand will be non-negative but positive for some x. Likewise, if $Q''[P(x)] < 0$ for all x, then H will be negative definite.

If H is a definite matrix for all of our iterations, then we will always be able to solve (IV-6) under rather general conditions. Furthermore, if H is a definite matrix throughout the iteration procedure, then if there is a solution to the problem, the specified iteration procedure will converge to that solution and that solution will be unique. To prove this, we shall reformulate the problem to make it clear that if H is positive definite, then we are actually maximizing a particular concave function.

C. Global Convergence to a Unique Solution

Let us consider maximizing the function $F(\lambda) = F(\lambda_1, \lambda_2, \ldots, \lambda_N)$ over a convex region, Ω , of N dimensional Euclidean λ space, where

$$F(\lambda) = \int \{ V[P(x),x] - \sum_{n=1}^{N} \lambda_n [G_n[P(x),x] - \gamma_n / A] \} \; dx \, , \qquad \text{(IV-11)}$$

$A = \int dx$ and $P(x)$ is defined as an implicit function of the λ's by

$$V'[\, P(x),x \,] = \sum_{n=1}^{N} \lambda_n \, G_n' [\, P(x),x \,] \quad . \qquad \text{(IV-12)}$$

Taking the partial derivative of (IV-11) with respect to λ_s , we have

$$\frac{\partial F}{\partial \lambda_s} = \int \{ [V'[P(x),x] - \sum_{n=1}^{N} \lambda_n \, G_n'[P(x),x]] \frac{\partial P(x)}{\partial \lambda_s} - [G_s[P(x),x] - \gamma_s / A] \} \; dx \qquad \text{(IV-13)}$$

If a maximum exists inside our convex region, Ω , the partial derivatives must vanish. Thus, from (IV-12), we see that a requirement for a maximum is

$$\int G_s \, [\, P(x),x \,] \; dx = \gamma_s \, , \quad (\, s=1 \text{ to } N \,) \, . \qquad \text{(IV-14)}$$

We note that under this formulation of the problem, our constraint equations (IV-2) will be automatically satisfied at a maximum of $F(\lambda)$. Taking the partial of (IV-13) with respect to λ_r , we have

$$\frac{\partial^2 F}{\partial \lambda_s \partial \lambda_r} = \int \{ [\, V''[P(x),x] - \sum_{n=1}^{N} \lambda_n G_n''[P(x),x] \,] \frac{\partial P(x)}{\partial \lambda_r} \frac{\partial P(x)}{\partial \lambda_s}$$

$$- G_r[P(x),x] \frac{\partial P(x)}{\partial \lambda_s} - G_s'[P(x),x] \frac{\partial P(x)}{\partial \lambda_r} \} \; dx \quad . \qquad \text{(IV-15)}$$

Taking the partial of (IV-12) with respect to λ_s , we get (IV-7), which allows us to write (IV-15) as

$$\frac{\partial^2 F}{\partial\lambda_s \partial\lambda_r} = -\int \frac{G_s'\,[P(x),x]\; G_r'[P(x),x]}{V''[P(x),x] - \sum_{n=1}^{N} \lambda_n G_n''[P(x),x]}\, dx \quad . \qquad \text{(IV-16)}$$

If no non-zero linear combination of the $G_n'[P(x),x]$ is zero and if $V''[P(x),x] - \sum_{n=1}^{N} \lambda_n G_n''[P(x),x] > 0$ over Ω , then the N by N matrix formed from $\frac{\partial^2 F}{\partial\lambda_s \partial\lambda_r}$ will be negative definite (the H matrix will be positive definite) and the function $F(\lambda)$ will be a strictly concave function in Ω . In this situation, if a maximum exists inside Ω , it is unique and will be found by following the procedure discussed in the last section (see Luenberger, David G. ref 9). Finally, we note that when F is a maximum with respect to the λ's , then all of the conditions for F being a minimum with respect to $P(x)$ are met. As can be seen, this procedure is the multidimensional case of Newton's method. Luenberger[9] discusses many other descent algorithms which may be more practical for solving the variational equation for specific problems.

D. Variational Principles

In maximum entropy spectral analysis, the variational principle is to find a maximum for

$$\int \ln [\, P(f) \,] \;\; df \; . \qquad \text{(IV-17)}$$

As for most density functions, the usual constraint equations are linear functionals in $P(f)$. That is, we know the values of integrals of the form

$$\int G_n(f)\; P(f)\; df \;=\; \gamma_n \;,\; n=1 \text{ to } N \; . \qquad \text{(IV-18)}$$

In this case, equation (IV-3) is

$$\ln'[P(f)] = 1 / P(f) = \sum_{n=1}^{N} \lambda_n G_n(f) . \qquad \text{(IV-19)}$$

This is of course easily solved explicitly for $P(f)$ as

$$P(f) = \frac{1}{\sum_{n=1}^{N} \lambda_n G_n(f)} \qquad \text{(IV-20)}$$

Equation (IV-10) for $Q''[P(f)]$ becomes $-1 / P^2(f)$ and thus $Q'' \leqslant 0$ for all f and H is thus negative definite.

In looking at the variational principle (IV-17), we see that $P(f)$ should be positive for all f since the logarithm of a negative number is complex. However, a better argument is perhaps given by (IV-19) in which we see that if $P(f)$ is close to zero, then $\ln'[P(f)]$ is large and positive and a small increase in $P(f)$ will make an appreciable increase in the value of the integral. Thus $P(f)$ is driven away from zero by the variational principle. Thus, the variational principle itself is used to impose the constraint of positiveness on the solution.

For the same set of constraint equations, (IV-18), there are an infinite number of variational principles which would make $P(f)$ positive for all frequencies. One example is

$$\int P(f) \ \ln[P(f)] \ df .$$

Then

$$Q'[P(f)] = \ln[P(f)] + 1 - \sum_{n=1}^{N} \lambda_n G_n(f) = 0 , \quad \text{which gives}$$

$$P(f) = \exp\left\{\sum_{n=1}^{N} \lambda_n G_n(f) - 1\right\} \quad \text{and} \quad Q''(f) = 1 / P(f) .$$

Here we would wish to minimize the integral and we see that when $P(f)$ is small, the $\ell n[P(f)]$ is a large negative number, and a small increase in $P(f)$ produces a sizable decrease in the integral. We also note that H is positive definite.

Another example is

$$\int P^a(f)\,df \quad , \quad \text{where} \quad a < 1 \quad \text{but} \quad a \neq 0 \ .$$

Then

$$Q'[\,P(f)\,] = a\,P^{a-1}(f) - \sum_{n=1}^{N} \lambda_n G_n(f) = 0$$

or

$$P(f) = \frac{1}{\left[\frac{1}{a}\sum_{n=1}^{N} \lambda_n G_n(f)\right]^{\frac{1}{1-a}}}$$

and $Q'' = a(a-1)\,P^{a-2}(f)$. Again we have a definite H matrix and the variational principle repells $P(f)$ away from zero.

A common variational principle is to minimize the average square value of a function, i.e., to minimize

$$\int P^2(f)\,df \ .$$

In this case,

$$Q'(f) = 2\,P(f) - \sum_{n=1}^{N} \lambda_n G_n(f) = 0 \quad \text{and}$$

$$Q''(f) = 2 \ .$$

Thus, H is positive definite and the functional form of our solution is simply

$$P(f) = \frac{1}{2}\sum_{n=1}^{N} \lambda_n G_n(f) \ .$$

However, this variational principle does not require $P(f) > 0$ and thus it should not be (but is) used for spectral analysis.

A final example is a variational principle that involves f explicitly.

$$\int \left[\frac{1}{P(f) - L(f)} + \frac{1}{U(f) - P(f)} \right] df ,$$

where $U(f) \geqslant L(f)$ for all f . Here if we start with $P(f)$ between $L(f)$ and $U(f)$ and try to minimize the integral under constraints, $P(f)$ will be repelled by both the lower and upper boundaries. This example shows how complex constraints can be introduced into a problem by the appropriate variational principle.

E. Consistency and Usefulness of Measurements

Aside from problems which arise from statistical uncertainties, there are two fundamental questions that can be raised about a set of measurements. One question is concerned with the results of the measurements and the other with the measurements themselves. It should be remarked here that the following observations are somewhat philosophical and argumentative.

To illustrate the first question, suppose we measure the zero and first lags of the autocorrelation function of some spectrum and find that $\Phi(0) = 1$ and $\Phi(1) = 2$. Since we know that $\Phi(0) \geqslant \Phi(1)$, these measurement results must be inconsistent. That is, it is clearly impossible to find a power spectrum which will be in agreement with these measurements. In another case, suppose $\Phi(0) = 1$, $\Phi(1) = 0.5$, $\Phi(2) = -0.1$ and $\Phi(3) = 0.3$. Are these measurements consistent?

This is a less trivial but still straightforward question to answer since one only needs to check the corresponding 4 by 4 Toeplitz matrix for semi-positive definiteness. However, if we also threw in the information that the power out of a filter with a complex frequency response of $Y(f)$ was 3 , i.e.,

$$\int Y(f)\, Y^{*}(f)\, P(f)\, df = 3 ,$$

then the question of consistency for the complete set of measurements becomes quite difficult.

The solution to this question of measurement consistency can be found by use of a variational principle approach. The reason is that if the data are consistent, then there is at least one spectrum which agrees with the measurements. If that spectrum is unique, then it is the extremal spectrum for all variational principles. If there is a set of spectra agreeing with the data, then, loosely speaking, if the value of a variational principle is bounded over this set, a particular member of this set will be selected by the variation principle. If suitable mathematical conditions (compactness, convexity) are imposed, then we could conclude that

1) If an extremal solution cannot be found that maximizes a particular variational principle, but the constraints bound the maximum value of the integral, then the data must be inconsistent. This result is independent of the particular variational principle if it is bounded.

2) If one variational principle has a solution, then any bounded variational principle has a solution.

The second question is about the measurements themselves, i.e., the properties or characteristics for which numerical values can be found. This question is inter-related to the boundedness of the variational principle. To give an example, suppose that we know the values of $\Phi(1)$ through $\Phi(10)$ but do not know the value of $\Phi(0)$. What is the maximum entropy solution for this set of measurements? The answer is that there is no maximum entropy extremal since the set of measurements cannot bound the entropy integral. One can always add more white noise to the spectrum, i.e., make $\Phi(0)$ larger and larger, without changing $\Phi(1)$ through $\Phi(10)$. In this case, one may object to the problem on the grounds that knowing $\Phi(1)$ through $\Phi(10)$ really doesn't tell us much about the spectrum. In fact, any set of numbers for $\Phi(1)$ through $\Phi(10)$ are consistent if we make $\Phi(0)$ large enough.

This example produces two observations:

1) Some sets of measurements may be missing a key characteristic without which the measurements are incomplete.

2) A variational principle must be bounded by the measurement set before it can be useful.