FINDING THE GLOBAL MINIMA OF OBJECTIVE FUNCTIONS USING MODIFIED STURM SEQUENCES

David G. Long

Electrical and Computer Engineering Department
Brigham Young University, Provo, UT USA 84602

ABSTRACT

The minimization of an objective function is a common requirement in signal processing. Minimization can be especially difficult in the presence of local minima and/or multiple global minima. This paper addresses the problem of finding *all* of the global minima of a real-valued objective function f which can be expressed as a polynomial over a finite interval. The approach used is based on generating a modified Sturm sequence from $f - z$ where z is the value of f at the global minima. The technique provides the number of the global minima of f, the value of z, and the locations of the global minima of f. A simple numerical example is given. While the paper concentrates on the one-dimensional case, the extension of the method to multidimensional objective functions is also discussed.

1. Introduction

Finding the global minima of an objective function in the presence of multiple global and/or local minima can be a very difficult task. In this paper a minimization technique based on modified Sturm sequences is proposed. The minimization problem to be solved may be stated as: given a real-valued objective function $f(x)$ defined on an interval $[a, b]$ where a and b are arbitrary, determine the set Y of $x \in [a, b]$ values for which $f(x)$ is minimized, i.e.,

$$Y = \left\{ y : f(y) \leq f(x) \; \forall x \in [a, b] \right\}. \qquad (1)$$

$f(x)$ is assumed to be one-dimensional polynomial. Note that $f(x)$ may be a Chebychev polynomial in $\cos x$.

To minimize $f(x)$, the value of $f(x)$ for $x \in Y$ is first determined. This is done by determining the smallest value of z for which $g(x) = f(x) - z$ has at least one zero. The number of zeros of $g(x)$ is determined using a modified Sturm sequence (described later). Once z is determined, the Sturm sequence provides a count of the number of global minima of $f(x)$. The global minima of $f(x)$ are then isolated by partitioning the interval $[a, b]$ into subintervals and counting the number of zeros of $g(x)$. If there are no zeros in a subinterval, the subinterval is discarded. Otherwise the subinterval is further divided and the process repeated until each subinterval contains a single zero of $g(x)$. By further subdividing these subintervals, the precise location of each zero (and, hence, the location of the global minimum of f) may be determined.

This paper first discusses some of the properties of Sturm sequences and then provides a description of the method. Computational considerations and extensions to multidimensional objective functions are then discussed.

2. Sturm Sequences

This section is provided to acquaint the reader with Sturm polynomial sequences (see also [1] and [2]). We begin with some useful concepts. An nth degree real polynomial over the field of real numbers is defined (x real),

$$P(x) = \sum_{n=0}^{N} p_n x^n \qquad (2)$$

where the coefficients p_n $n = 0, 1, \ldots, N$ are real and $p_n \neq 0$. The derivative of $P(x)$ with respect to x (x real) is,

$$P'(x) = \frac{d}{dx} P(x) = \sum_{n=0}^{N-1} (n+1) p_{n+1} x^n$$

From the fundamental theorem of algebra we know that the polynomial $P(x)$ evaluated at a and at b, $(b > a)$ must have the same sign over the entire interval if $P(x)$ has no real roots in the closed interval $[a, b]$. Further, between two real roots of $P(x)$ must be at least one real root of $P'(x)$ [2]. We will need the following definition:

Definition 1. Variations in Sign. *Let $\{P\} = \{P_0(x), P_1(x), \ldots, P_r(x)\}$ be an ordered set of real polynomials. The variations in sign of the sequence of polynomials evaluated at the point $x = a$ (a is a real number) is denoted as $\mathcal{V}_a$ and is the number of sign changes in the sequence $P_0(a), P_1(a), \ldots, P_r(a)$. Note that zeros do not count as a sign change. For example, there are four variations in sign in the sequence,* 5, −3, 2, 0, 2, −1, 0, 3.

A Sturm sequences is a sequence of polynomials $\{P\}$, corresponding to a given polynomial $P_0(x)$, from which it is possible to compute the exact number of roots of $P_0(x)$ between $x = a$ and $x = b$ $(b > a)$ as $\mathcal{V}_a - \mathcal{V}_b$ [2].

For the real polynomial defined in Eq. (2), a Sturm sequence $\{P\}$ may be constructed starting with $P_0(x) = P(x)$. This is the 0^{th} polynomial in the Sturm sequence. The 1^{st} polynomial is the derivative with respect to x of the 0^{th} polynomial, i.e.,

$$P_0(x) \triangleq P(x) = \sum_{n=0}^{N} p_0(n) x^n$$
$$P_1(x) \triangleq P'(x) = \sum_{n=0}^{N-1} p_1(n) x^n$$

where $p_0(n) = p_n$ and $p_1(n) = (n+1)\, p_{n+1}$.

The remainder of the Sturm polynomial sequence is generated using the "negative remainder" relationship,

$$P_{k+1}(x) = Q_k(x) P_k(x) - P_{k-1}(x) \qquad (3)$$

where $Q(x)$ is a polynomial chosen such that the order of the $(k+1)^{th}$ element of the Sturm sequence is less than the order of the k^{th} element. Methods for choosing $Q(x)$ will be discussed later. Eq. (3) is iterated until the S^{th} polynomial does not change sign over the interval of interest (e.g.,

CH2977-7/91/0000-1617 $1.00 © 1991 IEEE

$P_S(x)$ = constant). The resulting sequence of polynomials are called the *Sturm functions* for the polynomial $P(x)$. The Sturm theorem permits calculation of the number of roots (if any) in $P(x)$ using the Sturm sequence generated from $P(x)$.

Theorem 1. The Sturm theorem for real polynomials. *For a real number x let V_x be the number of variations in the sign of the Sturm functions $P_0(x), P_1(x), \ldots, P_S(x)$ for the real polynomial $P(x)$. If a and b are real number such that $a < b$, then $V_a - V_b$ is the exact number of roots of $P(x)$ in the interval $[a, b]$ if $P(a)P(b) \neq 0$.*

Proof of this theorem is found in [2].

3. Sturm Sequences for Function Minimization

To locate the global minima of $f(x)$ we assume that Y is a finite set and that the order of $f(x)$ be at least 1, i.e., $f(x)$ is not merely a constant. We further assume that $f(x)$ is not minimized at either a or b. The interval $[a, b]$ is problem dependent. Since $f(x)$ is a polynomial, $f(x)$ will be finite everywhere in the interval $[a, b]$.

Consider the function $g(x) = f(x) - z$ where z is arbitrary. If we choose z such that

$$z = \min_{x \in [a,b]} f(x) \tag{4}$$

then $g(x) \geq 0$ for all x. Equality occurs only for $x \in Y$, i.e., at the global minima of $f(x)$. Note that since $f(x)$ is a polynomial that $g(x)$ will also be a polynomial.

To determine the value of z for which Eq. (4) holds, we use a Sturm sequence generated from $g(x)$. Set $P_0(x) = g(x)$ and $P_1(x) = g'(x)$ with the rest of the Sturm sequence $\{P_i\}$ generated using the negative remainder relationship given in Eq. (3). Let $V_a(z)$ be the number of sign changes in the Sturm sequence evaluated at a for a particular value of z and let $V_b(z)$ be the number of sign changes in the Sturm sequence evaluated at b for a particular value of z. For convenience, we will define $\mathcal{V}(z) = V_b(z) - V_a(z)$. $\mathcal{V}(z)$ gives the number of zeros in the function $g(x) = f(x) - z$ in the interval $[a, b]$.

For $z \ll 0$, $\mathcal{V}(z)$ will be 0 since $g(x)$ will be positive for all x (refer to Figure 1). If we plot $\mathcal{V}(z)$ versus z, we see that as z is increased, at some z_0 where

$$z_0 = \min_{x \in [a,b]} f(x)$$

$g(x)$ will have at least one zero at the point(s) where $f(x) = z_0$. As z is increased further $\mathcal{V}(z)$ will increase to at least two. Depending on the structure of $f(x)$, $\mathcal{V}(z)$ may increase further or decrease. For sufficiently large z, $\mathcal{V}(z)$ will be zero (see Figure 2).

The value of $\mathcal{V}(z)$ at z_0 gives the number of global minima while is the minimum value of $f(x)$. To determine z_0, we must locate the smallest value of z which gives a non-zero value for the discrete function $\mathcal{V}(z)$. One approach is to use either a binary or golden section search. Starting with an initial value of z, say z_1, for which $\mathcal{V}(z_1)$ is zero and $g(x) > 0$ for all $x \in [a, b]$, and a second value of z, say $z_2 > z_1$, for which $\mathcal{V}(z_2) > 1$, we choose a new test z value $z_t = (z_1 + z_2)/2$ (for a binary search), and evaluate $\mathcal{V}(z_t)$. If $\mathcal{V}(z_t) > 1$ then z_2 is set to be z_t otherwise z_1 is set to z_t. This process is repeated until $z_2 - z_1 < T$ where T is a user-defined threshold. The threshold defines a range of values of $f(x)$ considered to be essentially identical. The value of z_0 is then z_2.

After z_0 is determined, $\mathcal{V}(z_0)$ gives the number of global minima. The locations of these global minima can be determined using the Sturm sequence generated from $g(x) = f(x) - z_0$. The interval $[a, b]$ is partitioned into subintervals $[a, y]$ and $[y, b]$. The number of sign reversals in the Sturm sequence evaluated at a, b, and y are then computed and the number of zeros of $g(x)$ in each interval determined using the Sturm theorem. If there are no zeros in the subinterval, the subinterval is discarded. Otherwise the subinterval is further divided and the process repeated until subintervals are determined which contain a single zero of $g(x)$. By further subdividing these subintervals, the location of all of the zeros of $g(x)$ may be determined to any desired accuracy. Since the locations of the zeros of $g(x)$ are the global minima of $f(x)$, this procedure will find all of the global minima of $f(x)$.

4. Computational Considerations

The objective function $f(x)$ may be a polynomial in x or a Chebychev polynomial in $\cos x$ [3]. In either case, $f(x)$ may be expressed as,

$$f(x) = \sum_{n=0}^{N} a_n x^n.$$

We compute the Sturm sequence of the polynomial $g(x) = f(x) - z$ where z will be treated as a constant independent of x. We define the first polynomial in the Sturm sequence $P_0(x)$ as,

$$P_0(x) \triangleq g(x) = f(x) - z = \sum_{n=0}^{N} p_0(n) x^n \tag{5}$$

where

$$p_0(n) = \begin{cases} a_0 - z, & n = 0, \\ a_n, & n > 0. \end{cases}$$

The second polynomial in the sequence is the derivative of $P_0(x)$ with respect to x,

$$P_1(x) = \frac{\partial}{\partial x} P_0(x) = \sum_{n=0}^{N-1} p_1(n) x^n \tag{6}$$

where $p_1(n) = (n+1)p_0(n+1)$.

Note that $P_1(x)$ does not depend on the value of z. The remainder of the Sturm sequence is computed using Eq. (3). In general $Q_k(x)$ may be any polynomial with order 1 or greater so that $P_{k+1}(x)$ has order less than $P_k(x)$. For simplicity let us assume that $Q_k(x)$ is first order polynomial of the form,

$$Q_k(x) = q_k x + r_k.$$

By writing the equations for the coefficients the powers of x, it can be easily shown that q_k and r_k must be

$$q_k = \frac{p_{k-1}(N-k+1)}{p_k(N-k)} \tag{7}$$

$$\begin{aligned} r_k &= \frac{p_{k-1}(N-k) - q_k p_k(N-k-1)}{p_k(N-k)} \\ &= \frac{p_{k-1}(N-k)}{p_k(N-k)} \\ &\quad - \frac{p_k(N-k-1)p_{k-1}(N-k+1)}{p_k^2(N-k)} \end{aligned} \tag{8}$$

in order for $P_{k+1}(x)$ to have order less than $P_k(x)$. Then,

$$p_{k+1}(n) = \begin{cases} r_k p_k(0) - p_{k-1}(0), & n = 0, \\ r_k p_k(n) + \\ \quad q_k p_k(n-1) - p_{k-1}(n), & 1 \leq n \leq N-k-2. \end{cases} \tag{9}$$

A consideration when directly implementing these equations is the case when $p_k(N-k) = 0$. This can occur when $P_{k-1} = (q_k x + r_k)P_k(x)$ exactly in the previous stage. When this occurs the remainder of the Sturm sequence is zero and no additional Sturm functions need be computed.

Note that the divisions in Eqs. (7) and (8) can lead to significant errors in computing the Sturm polynomial coefficients due to arithmetic truncation and rounding errors. To improve the accuracy of the procedure we modify the Sturm sequence to avoid the division in Eqs. (7) and (8) by scaling these equations by $p_k^2(N-k)$. This is equivalent to multiplying $P_{k+1}(x)$ by a positive constant. This does not effect the number of sign changes in the Sturm sequence. This modification of the Sturm sequence also avoids the need to detect the special case when $p_k(N-k)=0$ discussed above. Instead, the iteration is repeated for $1 \leq k \leq N-1$ with zero polynomials occurring if $p_k(N-k)=0$. This modified Sturm sequence has been used successfully for phase unwrapping [4].

After scaling, Eqs. (7) through (9) become

$$q_k = p_{k-1}(N-k+1)p_k(N-k) \quad (10)$$

$$r_k = p_{k-1}(N-k)p_k(N-k) - p_k(N-k-1)p_{k-1}(N-k+1) \quad (11)$$

$$p_{k+1}(n) = \begin{cases} r_k p_k(0) - p_{k-1}(0), & n=0, \\ r_k p_k(n) + q_k p_k(n-1) \\ \quad -p_{k-1}(n)p_k^2(N-k), & 1 \leq n \leq N-k-2 \end{cases} \quad (12)$$

While scaling eliminates the truncation and rounding errors occurring due division, the Sturm sequence polynomial coefficients may grow extremely large for large N. Careful numerical implementation using extensible number representation may be required (see [4]). Note that when the coefficients of $g(x)$ are integers that q_k, r_k and all of the coefficients of the Sturm sequence polynomials will be integers.

During the determination of z_0 each new z value requires recomputation of the Sturm sequence. However, once z_0 is determined, the Sturm sequence does not need to be recomputed for the subinterval searches. Only the evaluation of the Sturm sequence at the interval end points is required to count the sign changes.

5. A Numerical Example

In this section a simple numerical example illustrating the proposed minimization technique is provided. While this example is somewhat contrived (to give two global minima) it illustrates the technique. The technique will work for any arbitrary finite-order real polynomial. The example objective function is

$f(x) = 1+(x^2-x-1)^2 = 2+2x-x^2-2x^3+x^4$. The term in parenthesis has two real roots, $1/2(1 \pm \sqrt{5})$; hence, $f(x)$ has two global minima at these roots with a minimum value of 1. A plot of $f(x)$ is shown in Figure 3. Let us consider the interval $[-2,3]$.

Using the equations given in the previous section (Eqs. (5), (6), and (10) through (12)) the modified Sturm sequence for $g(x)=f(x)-z$ for $z=0$ is,

$$\begin{aligned} P_0(x) &= 2+2x-x^2-2x^3+x^4 \\ P_1(x) &= 2-2x-6x^2+4x^3 \\ P_2(x) &= -36-20x+20x^2 \quad q_1=4 \;\; r_1=-2 \\ P_3(x) &= 640-1280x \quad q_2=80 \;\; r_2=-40 \\ P_4(x) &= 67174400 \quad q_3=-25600 \;\; r_3=12800. \end{aligned}$$

Table 1 shows this Sturm sequence evaluated at $a=-2$

Table 1: Modified Sturm sequence generated from $g(x)=f(x)$ at $x=-2,3$

k	$P_k(x=-2)$	$P_k(x=3)$
0	26	26
1	-50	50
2	84	84
3	3200	-3200
4	67174400	67174400
$\mathcal{V}(z=0)$	2	2

and $b=3$. In each case the number of sign changes is 2. Hence, since $\mathcal{V}(z=0)=\mathcal{V}_a(0)-\mathcal{V}_b(0)=0$, there are no zeros of $g(x)$ for $z=0$ in $[-2,3]$. A plot of $\mathcal{V}(z)$ versus z for $z \in [0,3]$ is shown in Figure 4. Note that $\mathcal{V}(z)$ for $z<1$ is zero but is 2 at $z=1$. Hence, there are 2 global minima of $f(x)$ and the minimum value of $f(x)$ is $z_0=1$.

The modified Sturm sequence of $g(x)=f(x)-z_0$ is,

$$\begin{aligned} P_0(x) &= 1+2x-x^2-2x^3+x^4 \\ P_1(x) &= 2-2x-6x^2+4x^3 \\ P_2(x) &= -20-20x+20x^2 \quad q_1=4 \;\; r_1=-2 \\ P_3(x) &= 0 \quad q_2=80 \;\; r_2=-40 \\ P_4(x) &= 0 \quad q_3=0 \;\; r_3=0. \end{aligned}$$

Note that $P_3(x)=P_4(x)=0$ does not affect the number of sign changes for any x. Using the binary search technique described above, we determine the zeros of $g(x)$ to be at $x=1.618034 \approx (1+\sqrt{5})/2$ and $x=-0.618034 \approx (1-\sqrt{5})/2$ as desired.

6. Multidimensional Objective Functions

The method described above is a 1-d minimization procedure for finding the global minima of an arbitrarily complex real-valued polynomial objective function. While it is not possible to directly extend the technique to multiple dimensions due to the lack of a multidimensional extension of Sturm sequences, we can use a suboptimial approach in which the multidimensional space is mapped to a single dimensional space.

To minimize a multidimensional objective function $f(\vec{x})$ over a finite space $\vec{x} \in \Omega$, we define a parametric path $\vec{x}=\psi(\omega)$ which is dense in Ω. This permits expressing $f(\vec{x})$ as a one dimensional function of ω, $f(\omega)$ which can be minimized with respect to ω. The minima of $f(\vec{x})$ are the points along the path at which $f(\psi(\omega))$ is minimized. Selection of the path and its density is crucial to the success of this approach. Note that the objective function evaluated on the path function $\psi(\omega)$ must be a 1-d polynomial in ω. The Chebychev polynomials can be easily used as path functions.

To illustrate, consider minimizing a two dimension function $f(x,y)$ over the region $x^2+y^2 \leq 1$. Define the vector-valued function $\psi(\omega)$ as,

$$\psi(\omega) = (\omega \cos \alpha\omega, \omega \sin \alpha\omega),$$

where $\alpha \gg 1$ is a constant which determines the density of the path in the search space and $\omega \in [0,1]$. Note that $\psi(\omega)$ traces out a spiral path from the origin out to the unit circle. By choosing a sufficiently large α, the path will pass very close or through all of the minima of $f(x,y)$. In this two-dimensional example, α parameter determines the final polynomial order of $f(\omega)$. If the α is large, then $f(\omega)$ will be a very high order polynomial, requiring careful numerical implementation. The approach can be extended for additional dimensions.

The primary disadvantage of the parametric path approach is that if the path does not go through all of the global minima, some of the global minima may not be identified. This can be remedied by locating the near-global minima along the path using the modified Sturm sequence and using a gradient search routine to perform the final optimization, i.e., use the parametric path minimization to initialize a more conventional search algorithm.

7. Conclusion

A minimization technique for locating the global minima of a real-valued polynomial objective function has been described. While this modified Sturm sequence minimization technique is computationally intensive, it may be a useful technique when locating all of the global minima is required.

References

[1] M. Marden, *The Geometry of the Zeros of a Polynomial in a Complex Variable*, American Mathmatical Society, New York, New York, 1949.

[2] L. Weisner, *Introduction to the Theory of Equations*, The MacMillian Co., New York, New York, 1938.

[3] T. J. Rivlin, *The Chebychev Polynomials*, John Wiley and Sons, New York, New York, 1974.

[4] D. G. Long, "Exact Computation of the Unwrapped Phase of a Finite-Length Time Series," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, Vol. 36, No. 11, pp. 1787-1790, Nov. 1988.

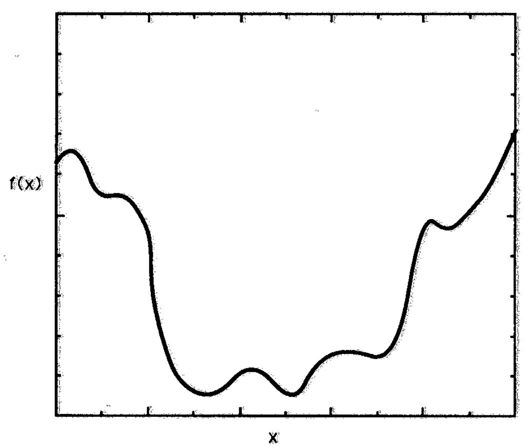

Figure 1: Multi-global minima $f(x)$ plotted versus x.

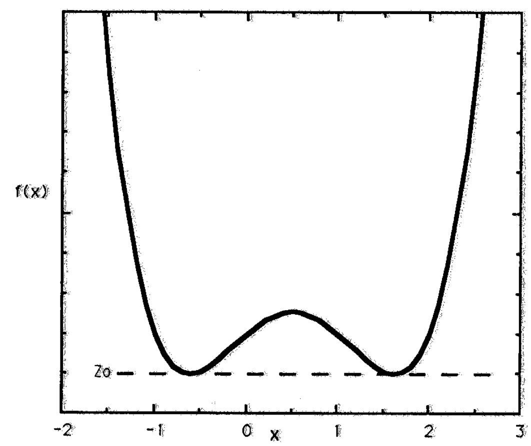

Figure 3: $f(x)$ over the interval $[-2, 3]$.

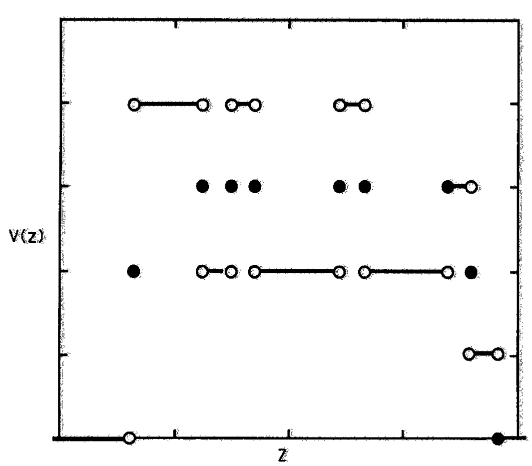

Figure 2: $V(z)$ versus z for the example in Fig. 1.

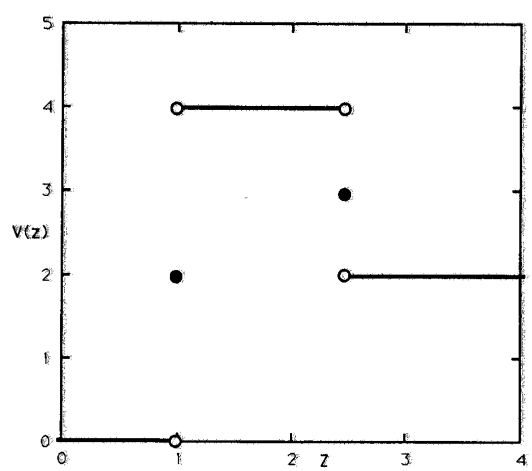

Figure 4: $V(z)$ versus z corresponding to Fig. 3.