A Derivative-free Multivariate Function Maximization Algorithm and Its Application to Maximum Likelihood

Tomoyuki AKITA*, Shizue IZUMI†, and Megu OHTAKI‡

*Department of Mathematics, Hiroshima University

†Department of Computer Science and Intelligent Systems, Oita University

‡Department of Environmetrics and Biometrics, Research Institute for Radiation Biology and Medicine, Hiroshima University

November 26, 2009

Abstract

We propose an algorithm for maximizing a multivariate functions that does not require derivatives, so it can be used more widely than traditional algorithms. We also present a method to obtain the maximum likelihood estimator and confidence interval for a parameter vector.

1 Introduction

In many cases of function maximization, no closed-form solution is available and the maximization has to be performed using numerical methods, often of an iterative nature. Newton type techniques are commonly used, although in general little is known about their global convergence properties (see, e.g. Spang (1962)). Brent (1973) proposed algorithms for optimization without derivatives. Our procedure also does not require derivatives, but uses a simpler algorithm based on univariate maximization which is similar to the one given by Hausman (1971). We first examine the case of a univariate function, where smoothness is not required; only continuity is assumed. The univariate maximization algorithm is called SPIDER1. In the case of maximizing a k-variate function, we use an approach similar to that proposed by Jensen et al. (1991), which is an alternating technique with cyclic fixing of groups of parameters, maximizing over the free remaining parameters. We call this algorithm SPIDER2, and apply it to maximization of the likelihood function for mortality and asbestos exposure, computing a confidence interval based on the likelihood ratio test.

2 Maximization of a Univariate Function

The following lemma is useful to develop our procedure in the case of a univariate function.

Lemma 2.1. *Suppose f is continuous on an interval $(\alpha, \beta) \subset (-\infty, \infty)$ and satisfies the following conditions:*

(*i*) *There exists $p \in (\alpha, \beta)$ s.t. $f(p) > f(y)$ for all $y \in (\alpha, \beta) \backslash \{p\}$*

(*ii*) *For all $y \in (\alpha, \beta) \backslash \{p\}$, there exists $\delta > 0$ s.t. f is s.m. (strictly monotone) on $(y - \delta, y + \delta)$.*

Then

$$\begin{aligned} p &> \frac{a+b}{2} \quad \text{if } f(a) < f\left(\frac{a+b}{2}\right) < f(b) \\ a < p &< b \quad \text{if } f\left(\frac{a+b}{2}\right) \geq \max\{f(a), f(b)\} \\ p &< \frac{a+b}{2} \quad \text{if } f(a) > f\left(\frac{a+b}{2}\right) > f(b) \end{aligned}$$

for all $(a, b) \subset (\alpha, \beta)$.

Proof. Let $I = \{y \in (\alpha, \beta) \mid f \text{ is s.m.i.(strictly monotone increasing) in a neighborhood of } y\}$, and $J = \{y \in (\alpha, \beta) \mid f \text{ is s.m.d.(strictly monotone decreasing) in a neighborhood of } y\}$. It follows from (*ii*), that $(\alpha, \beta) = I \cup J \cup \{p\}$. We now show that $I = (\alpha, p)$ and $J = (p, \beta)$. Suppose $(\alpha, p) \cap J \neq \phi$. Then pick an element $x_0 \in (\alpha, p) \cap J$, and let $\gamma = \sup\{z \in (x_0, p) \mid f \text{ is s.m.d. on}(x_0, z)\}$. If $\gamma \in J$, there exists $(u, v) \ni \gamma$ such that f is s.m.d. on (u, v). It follows that f is s.m.d. on (x_0, v), so that $\gamma < v \leq \gamma$. If $\gamma \in I$, there exists $(u, v) \ni \gamma$ such that f is s.m.i. on (u, v). Let $w_1 = (3u + \gamma)/4$ and $w_2 = (u + 3\gamma)/4$. Then $u < w_1 < w_2 < \gamma$, so that $f(w_1) < f(w_2) < f(w_1)$ which is a contradiction. If $\gamma = p$, then it follows from (*i*) that for all $\varepsilon > 0$, there exists $\delta > 0$ such that $f(p) - \varepsilon < f(z) < f(x_0)$ for all $z \in (p - \delta, p)$. Using a parallel argument we obtain $(p, \beta) \cap I = \phi$. Thus we deduce that $I = (\alpha, p)$ and $J = (p, \beta)$. The desired results follow immediately. □

We provide the following algorithm SPIDER1 for maximizing a univariate function which satisfying the conditions (*i*) and (*ii*) of Lemma 2.1.

[SPIDER1]

Step 1. Let τ be the tolerance the allowable length of the final subinterval containing the maximum. Given starting point t_0, compute the function values $f(t_0 - \tau/2)$, $f(t_0)$, and $f(t_0 + \tau/2)$.

Step 2. Let $g_0 := \max\{f(t_0 - \tau/2), f(t_0), f(t_0 + \tau/2)\}$. If $g_0 = f(t_0)$, then let $a := a_0 \leftarrow t_0 - \tau/4$ and $b := b_0 \leftarrow t_0 + \tau/4$. If $g_0 = f(t_0 - \tau/2)$, the maximizing value is in $(-\infty, t_0)$. In this case $t_n := t_0 - 2^{n-1}\tau$, $n = 1, 2, 3, \ldots$, and find n_0 s.t.

$$n_0 = \min\{n \mid f(t_n) \leq f(t_{n-1})\, n = 2, 3, \ldots\}.$$

Next let $a \leftarrow t_{n_0}$, and $b \leftarrow t_{n_0 - 1}$. If $g_0 = f(t_0 + \tau/2)$, the maximum is in (t_0, ∞). In this case $t_n := t_0 + 2^{n-1}\tau$, $n = 1, 2, 3, \ldots$, and find n_0 s.t.

$$n_0 = \min\{n \mid f(t_n) \geq f(t_{n-1})\, n = 2, 3, \ldots\}$$

Finally, let $a \leftarrow t_{n_0 - 1}$ and $b \leftarrow t_{n_0}$.

Step 3. At the m^{th} stage of the loop, let $t_{n_0+m} := c \leftarrow (a+b)/2$, $m = 1, 2, 3, \ldots$. Let $g := \max\{f(a), f(c), f(b)\}$. If $g = f(c)$, $a \leftarrow (a+c)/2$ and $b \leftarrow (b+c)/2$. If $g = f(a)$, $a \leftarrow (3a-c)/2$ and $b \leftarrow (a+c)/2$. If $g = f(b)$, $a \leftarrow (b+c)/2$ and $b \leftarrow (3b-c)/2$. Continue step 3 until convergence is achieved.

Lemma 2.2. *If a function f satisfies conditions (i) and (ii), then SPIDER1 converges in at most in $n_0 + n_1$ iterations, where $n_1 = [\log_2(b_0 - a_0)/\gamma] + 1$.*

3 Maximization of a Multivariate Function

In the case of a multivariate function, we use an approach similar to that proposed by Jensen et al. (1991). This alternating technique cyclically fixes groups of parameters, maximizing over the remaining free parameter. Our algorithm, which is called SPIDER2, is defined as follows.

[SPIDER2]

Step 1. Set an interval value $\boldsymbol{z}_0^{(0)}$ for finding the maximizing value $\boldsymbol{p}$ of a k-variate function f.

Step 2. (inner loop)

At the h^{th} stage of the outer loop, start with $\boldsymbol{z}_0^{(h)}$, $h = 0, 1, 2, 3, \ldots$. For $j = 1, \ldots, k$, define $f_j(t) := f(\boldsymbol{z}_{j-1}^{(h)} + t\boldsymbol{\delta}_j)$, where $\boldsymbol{\delta} = (\delta_{j1}, \ldots, \delta_{jk})'$ and δ_{jl} denotes Kronecher's delta. Using SPIDER1, optimize the function f_j and set

$$t_j := \arg\max_{t \in (-\infty, \infty)} f_j(t)$$
$$\boldsymbol{z}_j^{(h)} := \boldsymbol{z}_{j-1}^{(h)} + t_j \boldsymbol{\delta}_j.$$

(outer loop)

Set $\boldsymbol{x}_h := \boldsymbol{z}_k$ and calculate $\Delta_h := \|\boldsymbol{z}_0^{(h)} - \boldsymbol{z}_k^{(h)}\|$. Quit if Δ_h becomes small enough. Otherwise go back to the inner loop with $\boldsymbol{z}_0^{(h+1)} := \boldsymbol{z}_k^{(h)}$.

Step 3. Continue step 2 to convergence.

The following theorem is essential to establish for convergence of SPIDER2.

Theorem 3.1. *Suppose that Ω is open in $\mathbb{R}^k$ and that $\bar{\Omega}$ is compact. Suppose that f is continuous on Ω and satisfies the following conditions:*

(i) There exists $\boldsymbol{p} \in \Omega$ such that $f(\boldsymbol{y}) < f(\boldsymbol{p}) < \infty$ for any $\boldsymbol{y} \in \Omega \backslash \{\boldsymbol{p}\}$.

(ii) For any $\boldsymbol{y} \in \Omega \backslash \{\boldsymbol{p}\}$, there exists $j_0 \in \{1, \ldots, k\}$ and $\lambda > 0$ such that $f_{j_0}(\boldsymbol{t}|\boldsymbol{x}) = f(\boldsymbol{x} + t\boldsymbol{\delta}_{j_0})$ is strictly monotone in a neighborhood of 0 for any $\boldsymbol{x} \in B(\boldsymbol{y}, \lambda)$.

Here $B(\boldsymbol{y}, \lambda) = \{\boldsymbol{x} \in \mathbb{R}^k : \|\boldsymbol{x} - \boldsymbol{y}\| < \lambda\}$. Then $\boldsymbol{x}_n \to \boldsymbol{p}$ as $n \to \infty$ if $C(\{\boldsymbol{x}_n\}) \cap \partial\Omega = \phi$, where $C(\{\boldsymbol{x}_n\})$ denotes the set of cluster points of $\{\boldsymbol{x}_n\}$.

Proof. Let $g_n := f(\boldsymbol{x}_n)$, $n = 1, 2, 3, \ldots$. Then $g_1 \le g_2 \le \cdots \le g_n \le g_{n+1} \le \cdots \le f(\boldsymbol{p}) < \infty$, so there exists $\bar{g} = \lim_{n\to\infty} g_n \le f(\boldsymbol{p}) < \infty$. Since $C(\{\boldsymbol{x}_n\}) \cap \partial\Omega = \phi$, for all $\boldsymbol{y} \in C(\{\boldsymbol{x}_n\})$ and $\varepsilon > 0$, there exists $N_0 = N_0(\varepsilon)$ such that $g_n > \bar{g} - \varepsilon$ for all $n \ge N_0$, and there exists $\delta_0 > 0$ such that $|f(\boldsymbol{x}) - f(\boldsymbol{y})| < \varepsilon$ for all $\boldsymbol{x} \in B(\boldsymbol{y}, \delta_0)$. Hence, there exists $n_0 \ge N_0$ such that $\boldsymbol{x}_{n_0} \in B(\boldsymbol{y}, \delta_0)$ and $\bar{g} \ge g_{n_0} > \bar{g} - \varepsilon$. It follows that

$$|f(\boldsymbol{y}) - \bar{g}| \le |f(\boldsymbol{y}) - g_{n_0}| + |g_{n_0} - \bar{g}| < 2\varepsilon.$$

Therefore, $f(\boldsymbol{y}) = \bar{g}$.

Suppose that $\boldsymbol{y} \ne \boldsymbol{p}$. Then it follows from (ii) that there exists $j_0 \in \{1, \ldots, k\}$ and $\lambda_0 > 0$ such that $f_{j_0}(t|\boldsymbol{x})$ is s.m.i. in a neighborhood of 0 for all $\boldsymbol{x} \in B(\boldsymbol{y}, \lambda_0)$. Without loss of generality it is assumed that $j_0 = 1$. Let $\phi_{\lambda_0}(\boldsymbol{x}) = f_1(\lambda_0/2|\boldsymbol{x}) - f_1(0|\boldsymbol{x})$ for $\boldsymbol{x} \in \Omega$. Then, since ϕ_{λ_0} is continuous on Ω, there exists $\varepsilon_0 = \varepsilon_0(\lambda_0) > 0$ such that $\min_{x\in B(y,\lambda_0/2)} \phi_{\lambda_0}(\boldsymbol{x}) = \varepsilon_0 > 0$. Because f is continuous at $\boldsymbol{y} \in \Omega$, there exists $\delta_1 > 0$ such that $f(\boldsymbol{y}) - \varepsilon_0/2 < f(\boldsymbol{x})$ for all $\boldsymbol{x} \in B(\boldsymbol{y}, \delta_1)$. Let $\lambda_1 = \min\{\lambda_0/2, \delta_1\}$. Then, there exists n_1 such that $\boldsymbol{x}_{n_1} \in B(\boldsymbol{y}, \lambda_1)$ and, since $f_1(t|\boldsymbol{x}_{n_1})$ is s.m.i. on $(-\lambda_0/2, \lambda_0/2)$,

$$f(\boldsymbol{y}) \ge f(\boldsymbol{x}_{n_1+1}) \ge f_1\left(\frac{\lambda_0}{2}\middle|\,\boldsymbol{x}_{n_1}\right) \ge f_1(0|\boldsymbol{x}_{n_1}) + \varepsilon_0 = f(\boldsymbol{x}_{n_1}) + \varepsilon_0 > f(\boldsymbol{y}) + \frac{\varepsilon_0}{2},$$

which is a contradiction. Therefore $\boldsymbol{y} = \boldsymbol{p}$.

□

4 Application to Maximization of a Likelihood Function

In this section we illustrate an application of SPIDER. Consider the relationship between asbestos use and mortality from mesothelioma (see Nishikawa et al. (2008)). Data on asbestos use (1920–1999) and mortality (1979–1999) furnished by the WHO are given in Tables 1 and 2. First we assume:

- A1 The number of persons who are at risk of mesothelioma caused by asbestos increases in proportion to the use of asbestos. These persons will die after an interval of some years following their initial exposure.
- A2 The length of time between exposure to asbestos and death from mesothelioma is a random value following the Gamma distribution $G(k, \mu)$ with location parameter k and mean μ.

Then the mortality from mesothelioma in year t, $f(t)$ is described as follows:

$$f(t) \propto \int_{1920}^{t} \lambda(y) g(t-y|k,\mu) dy \propto \int_{1920}^{t} \lambda(y)(t-y)^{k-1} e^{-k(t-y)/\mu} =: h(t|k,\mu),$$

where $\lambda(y)$ is asbestos use and $g(w|k,\mu)$ is a probability density function of the Gamma distribution. The expectation of the cumulative mortality from mesothelioma in the population under the assumption of heterogeneity is $\xi h(t|k,\mu)$, where ξ is a constant. Given data $\{(n_y, m_y) \,|\, y = 1979, \ldots, 1999\}$, where n_y is total population size and m_y is number of deaths, if the parameters (k, μ) are known, the maximum likelihood estimator of ξ is

$$\hat{\xi}(k, \mu) = \frac{\sum_y m_y}{\sum_y n_y h(y|k, \mu)}.$$

We can use SPIDER2 to find the maximum likelihood estimator (MLE) of (k, μ) treating ξ as a nuisance parameter. MLEs of μ by gender are shown in Table 3.

5 Confidence Interval

We now consider how to construct confidence intervals for the parameters. Let $\boldsymbol{\mu} = (\mu_1, \mu_2, \ldots, \mu_k)$ and suppose we want to find a confidence interval for μ_i. Without loss of generality, we assume $\mu_1 = \mu_i$ and let $\boldsymbol{\mu}_2 = (\mu_2, \ldots, \mu_k)$. Let $f(\boldsymbol{\mu}) = -2 \log \ell(\boldsymbol{X} \,|\, \boldsymbol{\mu})$, where ℓ is the likelihood function, ε be small number, and H be some positive integer, then the confidence interval for μ_1 is obtained as follows:

Step 1. Find the maximum likelihood estimator of $\boldsymbol{\mu}$, $\hat{\boldsymbol{\mu}} = (\hat{\mu}_1, \hat{\boldsymbol{\mu}}_2)$ using SPIDER2.

Step 2. For $h = 1, 2, \ldots, H$, repeat the following loop:

For $k = 1, 2, \ldots$, calculate

$$\boldsymbol{\mu}_2^{(k)} = \arg\max_{\mu_2} f(\hat{\mu}_1 + k\varepsilon^{(h)}, \boldsymbol{\mu}_2)$$

using SPIDER2 until

$$(-1)^h \cdot \{f(\hat{\mu_1}^{(h)} + k\varepsilon^{(h)}, \boldsymbol{\mu}_2^{(k)}) - (f(\hat{\mu}_1, \hat{\boldsymbol{\mu}}_2) - \chi^2(1, \alpha/2))\} > 0,$$

where $\varepsilon^{(h)} = (-1)^{h-1} 10^{H-h+1} \varepsilon$ and $\chi^2(1, \alpha/2)$ is the upper $100 \cdot \alpha/2\%$ point of chi-square distribution with one degree of freedom.

Step 3. Output $\mu_1^{(H)} + k\varepsilon^{(H)}$ as the upper bound for the confidence interval for μ_1.

The lower bound is given by steps 1-3 substituting $\varepsilon^{(h)}$ for $-\varepsilon^{(h)}$. The algorithm is based on the asymptotic chi-square distribution of the likelihood ratio statistic. We use this method for the asbestos data in Tables 1 and 2, obtaining the confidence intervals shown in Table 3 in less than five minutes using an Intel Celeron 946 MHz processor.

6 Discussion

In this paper, we suggest an algorithm for derivative-free function maximization. Hausman(1971) and Brent(1973) suggested an algorithm for maximizing an univariate function without derivatives using golden section search. Our method is easier to deal with and can also be used with multivariate functions. On the other hand, Jensen et al. (1991) suggest an algorithm for maximizing an multivariate function employing the Newton method for univariate. We use a similar approach, but without derivatives, by modifying the univariate-function algorithm, SPIDER1 to a multivariate-function algorithm, SPIDER2.

Table 1: Asbestos use

JYEAR	CIN	JYEAR	CIN	JYEAR	CIN	JYEAR	CIN
1920	0.011410	1940	0.490513	1960	2.061435	1980	2.366669
1921	0.010201	1941	0.542695	1961	2.165136	1981	2.121210
1922	0.009008	1942	0.594130	1962	2.266579	1982	1.868984
1923	0.007830	1943	0.644836	1963	2.365837	1983	1.629376
1924	0.006668	1944	0.694826	1964	2.462980	1984	1.392677
1925	0.005520	1945	0.744117	1965	2.558075	1985	1.157740
1926	0.004388	1946	0.792722	1966	2.651186	1986	1.152337
1927	0.003270	1947	0.840657	1967	2.742374	1987	1.146726
1928	0.002166	1948	0.887934	1968	2.831698	1988	1.140805
1929	0.001076	1949	0.934568	1969	2.919215	1989	1.134474
1930	0.000000	1950	0.980571	1970	3.004979	1990	1.122268
1931	0.046265	1951	1.097542	1971	2.927804	1991	1.061751
1932	0.093117	1952	1.212395	1972	2.851788	1992	1.001945
1933	0.140569	1953	1.325189	1973	2.776906	1993	0.943411
1934	0.188631	1954	1.435977	1974	2.703133	1994	0.885990
1935	0.237315	1955	1.544812	1975	2.630444	1995	0.829132
1936	0.286635	1956	1.651747	1976	2.576884	1996	0.352481
1937	0.336601	1957	1.756829	1977	2.523731	1997	0.000938
1938	0.387227	1958	1.860108	1978	2.470980	1998	0.016635
1939	0.438527	1959	1.961628	1979	2.420106	1999	0.000256

Table 2: Mortality from mesothelioma

JYEAR	DMALE	DFEMALE	JYEAR	DMALE	DFEMALE
1979	347	193	1990	602	272
1980	365	196	1991	581	264
1981	363	228	1992	630	272
1982	416	203	1993	667	237
1983	434	213	1994	609	278
1984	469	198	1995	681	268
1985	492	217	1996	791	292
1986	541	254	1997	724	314
1987	544	239	1998	716	297
1988	564	235	1999	777	277
1989	584	225			

Table 3: MLE and 90%CI of μ

Male	19.136902	(17.406902,	21.216902)
Female	11.157669	(10.987669,	13.257669)

References

[1] Brent, R.P.(1973). Algorithms for Minimization Without Derivatives. Chapter 5, Prentice-Hall, Englewood Cliffs, New Jersey.

[2] Nishikawa, K. Takahashi, K. Karjalainen, A. Wen, C.P. Furuya, S. Hoshuyama, T. Todoroki, M. Kiyomoto, Y. Wilson, D. Higashi, T. Ohtaki, M. and Pan, G. (2008). Recent mortality from pleural mesothelioma, historical patterns of asbestos use, and adoption of bans: a global assessment. *Environmental Health Perspectives* **116**, 1675-1680.

[3] Hausman, Jr.,R.F.(1971). Function Optimization on a Line Segment by Golden Section. Lawrence Livermore National Laboratory, University of California.

[4] Jensen, S.T., Johansen, S. and Lauritzen, S.L.(1991). Globally convergent algorithms for maximizing a likelihood function. *Biometrika* **78**, 867-877.

[5] Spang, III,H.A.(1962). A review of minimization techniques for non-linear functions. *SIAM Review* **4**, 357-359.