A Hybrid Line Search Technique with Modified Goldstein and Wolfe Conditions

Sawsan S. Ismael
Department of Math / College of Education
University of Mosul

Received 16 / 02 / 2011 **Accepted** 10 / 07 / 2011

الملخص

شروط Wolfe و Goldstein المطورة يمكن استخدامها لغرض دراسة خاصية التقارب لخوارزمية تكرارية غير خطية، بهذه الشروط تمكنا من إيجاد خط بحث هجين ملائم لتحقيق خاصية التقارب الشامل للخوارزمية المقترحة. التطوير يجعل خط البحث ملائم للخوارزمية المقترحة. خط البحث المقترح تفوق على خط البحث الاعتيادي بأخذ تسع دوال لاخطية.

ABSTRACT

Modified Goldstein or Wolfe conditions can be imposed on a hybrid line search to ensure the convergence property of an iterative nonlinear optimization algorithm to a stationary point. Modifying these conditions can make it significantly easier to find an acceptable step size. Our numerical results indicate that the new proposed line search beats the standard one ,for a selected nine test functions.

Keywords: Truncated Newton, Line-search (Modified) Wolfe conditions, (Modified) Goldstein conditions

1. Introduction:

Optimization might be defined as the science of determining the best solutions to certain mathematically defined problems, which are often models of physical reality, It involves the study of optimality criteria for problems, the determination of algorithms methods of solution, the study of the structure of such methods and computer experimentation with methods both under trial conditions and on real life problems. There is extremely diverse range of practical applications. The applicability of optimization methods is widespread, reaching into almost every activity in which numerical information is processed (science, engineering, mathematics, economics, commerce etc.) [3].

1.1 The Goldstein Conditions:

The Goldstein conditions for a step size α are:

$(G1) f(x_k+\propto d_k) < f(x_k) + m_1 \propto (d_k g_k)$
$(G2) f(x_k+\propto d_k) > f(x_k) + m_2 \propto (d_k g_k)$ [5]

Here f is the function being minimized, x_k is the current search position, d_k is a search direction at x_k, $\propto$ is a step size found by the line search, $g_k = f'(x_k), (d_k, g_k)$ denotes inner production and $m_2 = 1 - m_1$ where $(d_k\, g_k) < 0$ iff d_k is a descent direction.

1.2 The Wolfe Conditions:

The Wolfe conditions may be used with the line search instead of the Goldstein conditions. The Wolfe conditions are:

$(W1) f(x_k + \alpha d_k) < f(x_{k6}) + m_1 \alpha (d_k g_k)$
$(W_2) d_k f'^{(x_k+\alpha d_k)} > m_2 (d_k g_k)$ [6]

As with the Goldstein conditions, m_1, m_2 are constants with $0 < m_1 < m_2 < 1$ and usually $m_1 < 0.5, m_2 = 1 - m_1$

2. Algorithms:

2.1 Algorithm (1) (Goldstein) [1]

Let α be the initial step size (usually α =l) and let R>1 be a positive constant. Now perform $\beta := \propto;$
while not $G_1(\alpha)$ do $\beta := \alpha;\ \alpha := \beta / R$ enddo; while not $G_1(\beta)$ do $\alpha := \beta;\ \beta; = R * \alpha$ enddo
while not $G_2(\beta)$ do $\propto :;\ \beta := R_{*\propto}$ enddo ;

The first iteration must terminate because f is differentiate at x_k.

The second iteration must terminate because f is bounded below.

Since if G_1 is false then G_2 is true.

We have the postcondition that α satisfies G_1 and β satisfies G_2, and either $\beta = \alpha$ or $\beta = R\alpha$

Note that the two while loops can be placed in either order and that at most one of them will be performed.

2.2 Algorithm (2) (Wolfe) [1]

Let R>1 and $0 < r \leq 0.5$ be positive constants.
Let t be the initial step size (usually t=1) and perform

$\alpha := 0;\ \beta := \infty;$
while not ($\alpha = \beta$) do
if $w_1(t)$ then $\alpha := t$ else $\beta := t$ end if;
if $w_1(t)$ and $w_2(t)$ then $\beta := t$
else
choose a new t with $\alpha < t < \beta$;
if $\beta = \infty$ then t:=max ($t, R * \alpha$)
else t:=max ($(1-r) * \alpha + r * \beta, \min(t, r * \alpha + (1-r) * \beta)$)

end if
end if
end do

For the first iteration, t is chosen to be initial step size. In later iterations, t may be chosen by polynomial extrapolation or interpolation, but must be adjusted to ensure first that β eventually becomes finite and subsequently that $\beta-\alpha$ decreases by a factor of $1-r$ in the worst case but even in the best case the factor is at least r, which can be unfortunate when t is close to α .

3. Modified Algorithms:

3.1 Modified Goldstein:

The two modified Goldstein conditions become as follows:

$$(MG_1) f(x_k+\alpha d_k) \le f(x_k)+m_1 \alpha g_k^T dk - m_2\alpha^2 \| d_k \|^2$$
$$(MG_2) f(x_k+\alpha d_k) > f(x_k)+m_2 \propto d_k^T g_k - m_2\alpha^2 \| d_k \|^2 \qquad (3.1.1)$$

where m_1= 0.0001
and m_2= 0.9999
d_k is a search direction
α is a step size
f is the function being minimized
x_k is the current search position
MG_1 and MG_2 are modified forms of the standard Goldstein line search procedure with strictly descent property.

3.2 Modified Wolfe:

$$(Mw_1) f(x_K+\alpha d_k) < (x_k)+m_1 \alpha d_K^T gk - m_2\alpha^2 \| d_k \|^2$$
$$(Mw_2) dk.f'(x_K+\alpha d_k) > m_2 d_K^T g_k - m_2\alpha^2 \| d_k \|^2 \qquad (3.2.1)$$

where m_1 = 0.0001
and m_2= 0.9999
d_k is a search direction
α is a step size
$g_k = f'(x_k)$
f is the function being minimized
x_k is the current search position
Mw_1 and Mw_2 are modified forms of the standard Goldstein line search procedure with strictly descent property.

3.2 Theorem 1 (on modified Goldstein):

Suppose that f is bounded below and has Lipshitz continuous derivative f' on the *basin* $y: f(y) \le f(x_0)$. Let $R>1$ be a positive constant. Let $d_k^T g_k < 0$ and $\propto_k$ β_k are chosen with $0<\beta_k<R\alpha_k$ and such that the step size $\propto_k$ satisfies *(MG_1)* and the step size β_k satisfies *(MG_2)* set $x_{k+1}=x_k+\alpha_k d_k$ then either $g_k=0$ for some k or else

$$\sum_{k=0}^{\infty}[cos^2\theta_k \parallel g_k \parallel^2] < \infty$$

where θ_k is the angle between d_k *and* g_k

Proof:

Let *m* be min $[m_1, 1-m_2]$ then by *(MG2)* we have

$$f(x_k+\beta_k d_k)-f(x_k) > (1-m)\beta_k d_k^T g_k - m\beta_k^2 \parallel d_k \parallel^2$$

Hence by (MVT) we have for some c with $0 < c < \beta_k$

$$d_k.f'(x_k+cd_x) > (1-m)d_k^T g_k - m \parallel d_k \parallel^2$$

$$d_k.f'(x_k+cd_x) > d_k^T g_k - m d_k^T g_k \ - \ m \parallel d_k \parallel^2 > d_k^T g_k - m(d_k^T g_k - \parallel d_k \parallel^2)$$

so

$$d_k[f'(x_k+cdk)-f'(x_k)] > \parallel d_k \parallel \parallel g_k \parallel cos\theta - m(\parallel d_k \parallel \parallel g_k \parallel cos\theta - \parallel d_k \parallel^2)$$

Let *k* be a Lipshitz constant for f' then

$\|f'(x_k+cd_k)-f'(x_k)\| < c_k\|d_k\|$ and so

$$c_k \parallel d_k \parallel^2 > \parallel d_k \parallel \parallel g_k \parallel cos\theta - m(\parallel d_k \parallel \parallel g_k \parallel cos\theta - \parallel d_k \parallel^2)$$

Also $c < \beta_k < R_{\alpha k}$

So

$$\alpha_1 k \parallel d_1 k \parallel > -(m/R_1 k) \parallel g_1 k \parallel cos\theta - (m/R_1 k)(\parallel g_1 k \parallel cos\theta - \parallel d_1 k \parallel$$

We note that both sides are positive, because d_k is a descent direction so $cos\theta_k < 0$ from (MG1) we have

$$f(x_k)-f(x_k+\alpha_k d_k) > -m \propto_k d_k^T g_k + m \propto_k^2 \parallel d_k \parallel^2 > \left(\frac{m^2}{R_k}\right) \parallel g_k \parallel^2 cos^2\theta_k > 0$$

So, the $f(x_k)$ are monotone decreasing and bounded below by L

Hence,

$$f(x_0) - L \geq \sum_{k=0}^{\infty}[f(x_k) - f(x_k+1))] > (m^2/R_k)\sum_{k=0}^{\infty}[cos^2\theta \parallel g_k \parallel^2]$$

3.3 Notes about Modified Wolfe Condition:

Corollary:

Under the conditions of theorem 1 if we choose α_k, β_k to satisfy $Mw1, Mw2$ instead of MG_1, MG_2 respectively, then the same conclusion holds.

Proof: Let *m* be *min[m_1,1-m_2]* and proceed as in the proof of theorem, instead of applying the MVT to MG_2 use MW_2 directly to get

$$dk.f'(x_k+b_k d_k) > (1-m)d_k^T g_k - (1-m) \parallel d_k \parallel^2$$

Then proceed as before with β_k in place of c.

Note that although in practice β_k is usually chosen so that $\beta k \geq \alpha k$ this condition is not required by either proof while the original corollary poured in [1].

4. Modified Goldstein and Wolfe Line Searches:

Use of the de-linked modified Goldstein conditions allows a very simple and quick line search to be implemented.

4.1 Algorithm (Modified Goldstein) :

Let α be the initial step size (usually $\alpha = 1$) and let $R > 1$ be positive constant. Now:

$\beta \coloneqq \alpha;$

while not $MG_1(\alpha)$ do $\beta \coloneqq \alpha$; $\alpha \coloneqq \beta / R$ enddo

while not $MG_2(\beta)$ do $\alpha \coloneqq \beta; \beta \coloneqq R * \alpha$ enddo

The first iteration must terminate because *f* is differential at x_k

The second iteration must terminate because *f* is bounded below. Since if MG_1 is false then MG_2 is true.

We have the post condition that α satisfies MG_1 and β satisfies MG_2, and either $\beta = \alpha$ or $\beta = R\alpha$, we note that the two while loops can be placed in either order, and that at most one of them will ever be performed, while the original property found in [1].

4.2 Algorithm (Modified Wolfe):

Let $R > 1$ and $0 < r \leq 0.5$ to positive constants.

Let *t* be the initial stepsize (usually $t = 1$) and perform

$\alpha \coloneqq 0;\ \beta \coloneqq \infty$

while not $(\alpha = \beta)$ do

if MW_1 then $\alpha :=t$ else $\beta :=t$ end if;

if MW_2 *(t)* and W_2 *(t)* then $\beta := t$

else

choose a new *t* with $\alpha < t < \beta;$

if $\beta = \infty$ then $t:= max\ (t, R * \alpha)$

else $t:= max\ ((1-r)*\alpha + r*\beta,\ min\ (t, r*\alpha + (1-r)*\beta))$

end if

end if

end do

for the first iteration, *t* is chosen to be the initial stepsize. In later iterations *t* may be chosen by polynomial extrapolation or interpolation, but must be adjusted to ensure first that β eventually becomes finite, and subsequently that $\beta - \alpha$ decreases geometrically with each iteration. The length $\beta - \alpha$ decreased by a factor of *1-r* in the worst case, but even in the best case the fact is that least *r*, which can be unfortunate when *t* is close to α .

5. Numerical Results

To ensure the effectness of the two considered line search techniques we have to employ one of them namely [modified wolfe-line search] inanstandard CG-method (say FRCG method) and as follows:

5-1 Algorithm (out lines)

Step 1 :set $x_1, \in, k = 1, d_1 = -g_1$

Step 2 :Find $x_{k+1} = x_k + \lambda_k d_k$

Where λ_k must be founded by algorithm (4-2)

Step 3 :Compute d_{k+1}=-g_{k+1}+$\beta_k d_k$

$$\beta_k = \| g_{k+1} \|^2 / \| g_k \|^2$$

Step 4 :If k=n or Powell restarting condition satisfied go to step 1 else set k=k+1 go to step 2

Table (5-1)

Comparison between standard FR and wolf and the proposed modified method .

Test Function	N	FR and wolf standard (Nof)	(NoI)	FR and wolf Modified (Nof)	(NoI)
Powll	100	256	(126)	245	(114)
	500	1023	(505)	1009	(502)
	1000	2025	(1005)	2007	(1001)
Wood	100	828	(276)	872	(323)
	500	3702	(1007)	3338	(1000)
	1000	6797	(1296)	7739	(2006)
Rosen	100	296	(103)	296	(103)
	500	453	(182)	394	(152)
	1000	453	(182)	408	(159)

In this comparison we have compared three well-known standard test function with high dimensions as asample for a comparison between the standard and modified line search criterion used in the standard FRCG method. Table (5-1) utilizes the numerical results for this comparisons ,for nine case of (NoI) and (NoF) together, we have found that. The new proposed line search beats the standard one, in six cases out of nine while it has a comparable results in one of these cases.

The rate of improvement was about 34% in both (NoI) and (NoF) according to our selected group of test Functions.

6. Conclusions:

This paper investigates an optimization algorithm with two modified line search ceiteria which are satisfy both modified Goldstein conditions MG_1 and MG_2 (or both modified Wolfe conditions Mw_1 and Mw_2)

Although line search algorithms ensures that the step size is not too small or too long , this means that backtracking line search technique cannot be combined with any arbitrary search direction using algorithms (4-1) and (4-2).

Appendix

1- Generalized Powell Function

$$F(x)=\sum_{i=2}^{n/4}\left[(x_{4i-3}-10x_{4i-2})^2+5(x_{4i-1}-x_{4i})^2+(x_{4i-1}-2x_{4i-1})^2+10(x_{4i-3}-x_{4i})^2\right]$$

$$x_0=(3,-1,0,1,\ldots\ldots)^T$$

2- Generalized Wood Function

$$F(x)=\sum_{i=2}^{n/4}\left[100(x_{4i-3}-10x_{4i-2})^2+(1-x_{4i-3})^2+90(x_{4i}-x^2{}_{4i-1})^4+(1-x^2{}_{4i-1})^2+1.0\right]$$

$$x_0=(-3,-1,-3,-1,\ldots\ldots)^T$$

3- Generalized Rosenbrock Function

$$F(x)=\sum_{i=2}^{n/4}\left[100(x_{2i}-x^2{}_{2i-1})^2+(1-x_{2i-1})^2\right]$$

$$x_0=(-1,2,1,\ldots\ldots)^T$$

References:

1) Bruce Christianson, Global convergence using de-linked Goldstein or Wolfe line search conditions, University of Hertfordshire, Hatfield A110 9AB, England.

2) L. Armijo, Minimization of functions having Lipschitz continuous first partial derivatives, Pacific Journal of Mathematics, 16(1), 1-3, 1966.

3) R. Fletcher, Practical methods of optimization. Second edition.

4) L.C.W. Dixon and R.C. Price, Numerical experience with the truncated Newton method for unconstrained optimization, Journal of Optimization Theory and Application, 56, 245-255, 1988.

5) A.A. Goldstein, Constructive Real Analysis, Harper and Row, London, 1967.

6) P. Wolfe, Convergence conditions for ascent method, SIAM Review, 17, 226-235, 1969 and 13, 185-188, 1971.