Using the Update of Conditional BFGS in Constrained Optimization

Abbas Y. Al-Bayati profabbasalbayati@yahoo.com
Ban Ahmed Mitras dr.banah.mitras@gmail.com
College of Computer Sciences and Mathematics
University of Mosul, Iraq

Received on: 12/03/2003 **Accepted on: 26/01/2004**

ABSTRACT

In this paper, we have used one of the preconditioned conjugate gradient algorithm with the Quasi – Newton approximation; namely the BFGS preconditioned algorithm which was suggested by (AL-Bayati and Aref, 2001). In this paper we have suggested a new algorithm for constrained optimization with robust numerical results, for solving constrained optimization problems.

Keyword: Constrained Optimization, QN-Condition, BFGS Update.

استخدام تحديث BFGS المشروط في الامثلية المقيدة

عباس يونس البياتي **بان احمد متراس**
كلية علوم الحاسبات والرياضيات، جامعة الموصل

تاريخ استلام البحث:2003/3/12 **تاريخ قبول البحث:2004/1/26**

الملخص

في هذا البحث تم استخدام إحدى خوارزميات التدرج المترافق المشروطة مع تقريبات اشباه نيوتن وهي خوارزمية BFGS المشروطة والمتوازية والمقترحة من قبل (Al-Bayati and Aref, 2001) في حل مسائل الامثلية المقيدة واقتراح خوارزمية جديدة في هذا المجال. تم اختبار الخوارزمية المقترحة على بعض المسائل المقيدة وأثبتت النتائج العددية كفاءة الخوارزمية الجديدة مقارنة" مع الخوارزميات الأصلية التي استخدمت لحل مسائل الامثلية المقيدة.

الكلمات المفتاحية: الأمثلية المقيدة، شرط QN و تحديث BFGS.

1.المقدمة: Introduction

لدينا مسألة الامثلية المقيدة الآتية:

$$\begin{aligned} &\min \quad f(x) \\ &s.t. \quad c_k(x) \geq 0, \quad \forall k \in I \end{aligned} \qquad \ldots(1)$$

حيث ان f و c_k هي دوال من E^n الى E. ولتكن I مجموعة منتهية من الأعداد الصحيحة الموجبة، بفرض $f(x)$ و $c_k(x)$ هي دوال قابلة للاشتقاق مرتين على فترة مفتوحة F حيث :

$$F = \{ x \in E^n : c_k(x) \geq 0, \quad \forall k \in I \} \qquad \ldots(2)$$

من الممكن إيجاد حل لمسألة مقيدة عامة باستخدام دالة هدف محورة

$$\theta(x,r) = f(x) + rB(x) \qquad \ldots(3)$$

حيث أن r تعرف على انها معلمة سيطرة (control parameter) والحد الثاني B(x) هو دالة Barrier . القيمة الابتدائية r تعتبر ذات اهمية في اختزال عدد التكرارات لتقليل الدالة $\theta(x,r)$. القيمة الابتدائية r_0 اقترحها العالمان (McCormick), (Fiacco) في عام (1968) وكانت كالاتي:

$$r_0 = \frac{\nabla f(x)^T \nabla B(x)}{\nabla B(x)^T \nabla B(x)} \qquad \ldots(4)$$

وتم ايجاد عدة بدائل لقيمة r_0 راجع (Al-Bayati and Hamed , 2000) و (Al-Bayati and Hamed , 1999).

في هذا البحث تم استخدام خوارزمية BFGS المتوازية والمشروطة المقترحة من قبل (Al-Bayati and Aref , 2001) في حل مسائل الامثلة المقيدة.

2. خوارزميات التدرج المترافق المشروط مسبقا":

Preconditioned Conjugate Gradient Methods (PCG)

خوارزميات أشباه نيوتن–(Quasi-Newton (QN والتدرج المترافــق Conjugate Gradient (CG) في تطبيقات الدوال العامة لها مساوئ ومحاسن عملية, حيث ان خوارزمية QN أسرع (وتتطلب حسابات اقل لقيمة الدالة المعينة) من خوارزميات التدرج المترافق كذلك خوارزميات QN تكون متتابعة من مصفوفات متناظرة، مربعة وموجبة قطعاً ، خوارزميات المتري المتغير المحدود لا يكون استعمالها فعالاً بسبب الخزن في حالة تزايد عدد المتغيرات بشكل اطرادي، ولهذا تم التطرق الى صنف جديد من خوارزميات التدرج المترافق CG المطور تسمى خوارزميات التدرج المترافق المشروط (PCG). فكرة الاشتراط هي لتحويل المسألة، إذ ان مصفوفة Hessian للمسألة المتحولة تمتلك القيم المميزة المتجمعة وتكون حسنة الشرط well – conditioned. الغرض من خوارزمية (PCG) هو لحفظ متطلبات الخزن من رتبة n مع تحسين تقارب الخوارزمية ولمزيد من المعلومات في هذا الصدد يمكن الرجوع الى (Buckely and Lenier, 1983 ; Nazareth, 1979)كذلك يمكن الرجوع الى (Buckely ,1978; AL-

Bayati and Mohammed Ali, 2002) في خوارزميات PCG وحيث ان اتجاه التدرج المترافق يتولد بالاعتماد على تحويلات المصفوفة.

1.2 فكرة خوارزمية التدرج المترافق المشروط مسبقا" PCG:

The Idea of the PCG Algorithm:

طورت خوارزميات PCG خوارزمية CG القياسية والتي اقترحها في البداية (Axelsson, 1974) واستعملت في حل انظمة المعادلات الخطية، فكرة الاشتراط امتدت مباشرة لحل المسائل غير الخطية.

هذا النوع من خوارزميات التدرج المترافق يكون ملائماً للمسائل التي تكون متغيراتها كبيرة. هذا بسبب الخزن فقط لعدد قليل من المتجهات المطلوبة لحل المسألة التي تتكون من n من المتغيرات.

إن خوارزمية CG الأصلية ليست فعالة دائماً لكن الاشتراط يستعمل المصفوفة المناسبة التي تعجل التقارب في خوارزمية CG عن طريق تحويل المتغيرات ، بينما يحافظ على الخواص الأساسية لهذه الخوارزمية. أسست المصفوفات المشروطة على تقريبات معكوس Hessian المتولدة في خوارزمية QN. لذلك فان الفكرة الأساسية وراء هذه الخوارزمية هي استخدامها تحسينات QN, المتري المتغير (VM) لإسراع خوارزمية CG.

هناك تقاربان واضحان لإسراع خوارزميات مع مصفوفات أشباه نيوتن (QN)، **اولاً:** تتكون من تطبيق خوارزميات CG مع المتري المتغير المطور وباستعمال صيغة QN. هذه الخوارزمية يطلق عليها خوارزمية التدرج المترافق المشروط. بمصفوفة التي اقترحها لأول مرة العالم Nazareth في (1979) ومن ثم طورها آخرون أمثال(Al-Bayati and Sharef, 2001).

التقارب الثاني– هو صف من الخوارزميات المعرفة بـ CG- QN التداخلية. هذه الخوارزمية تسمى CG-QN المتداخلة (Interleaved) حيث اقترحها لأول مرة العالم الكندي (Buckely,1987) وطورها (Shanno, 1981) و (Nazareth and Nocedal, 1982) و (Buckely and Lenir, 1983) و (Liu and Nocedal, 1988) و (Al-Bayati and Mohammed Ali, 2002) .

2.2 تقريبات أشباه نيوتن بوجود الاشتراط المسبق:

Preconditioning with Quasi – Newton Approximations

الاشتراط في خوارزميات التدرج المترافق يعني تحسين أدائها . وتتطلب فعالية الاشتراط معلومات عن المشتقة الثانية التي تكون غير متوافرة أحيانا ويمكن جمعها في طريقة خوارزمية QN. مسببات الخزن ويمكن إهمالها مؤقتاً لهذا فإن n×n من تقريبات QN يمكن استخدامها مع

بعض الشروط الموضوعة عن خط البحث والدوال غير الخطية ، المصفوفات H_k لتقريب معكوس Hessain تكون موجبة التعريف ومتناظرة. عموماً فإن مصفوفة الشرط H يجب ان تستخدم في كل مكان للحصول على التوقف التربيعي مع خوارزمية التدرج المترافق.

هذا بالنسبة للنماذج التربيعية, اما في النماذج غير التربيعية فيمكن الرجوع الى (Al-Bayati and Al-Assady, 1997).

2.3 تقريبات QN محدودة الذاكرة :

Limited – Memory Quasi – Newton Approximations

أثبت تحديد الذاكرة لتقريبات QN معنى اخر للاشتراط مع تقريبات QN ، ويمكن اعتبار تقريب معكوس مصفوفة Hessian الابتدائي مصفوفة متناظرة وموجبة التعريف يمكن خزنها في الذاكرة المتوافرة حتى وان ضربت في متجه سعته O(n) من العمليات، عادة ما تكون مصفوفة قطرية او شريطية (diagonal or banded) بدلاً من العناصر التطويرية الواضحة التقريب، ويكون التطوير تماماً بواسطة خزن المتجهات مع المعلمات التي تعرف بتطوير (Shanno, 1982) ولهذا يتطلب التقريب الابتدائي خزناً قليلاً، وتخزن مجموعة الازواج من المتجهات والمعلمات بدلاً من مصفوفة سعة *nxn*. والخزن ذو التقريب الذي يخزن في هذه الطريقة هو تقريب QN محدود الذاكرة. وهناك الخزن ذو الخطوة الواحدة والخزن ذو الخطوتين لمزيد من المعلومات انظر (AL-Bayati and Ahmed, 1997).

3. خوارزمية BFGS المطورة: Improved BFGS Algorithm

في خوارزميات المتري المتغير المشروط المعروفة نبدأ من نقطة اختيارية لاتجاه البحث s^k عند التكرار $-k$ وتحسب

$$s^k = -H_k \Delta f(x^k) \qquad \ldots(5)$$

حيث ان الدالة f المصغرة (على فرض انها مستمرة وتملك مشتقة ثانية) و x_k نقطة التكرار المتداولة ، و H_k التقريب الى معكوس مصفوفة Hessian عند x^k نقطة التكرار التالي x^{k+1} نحصل عليها من :

$$x^{k+1} = x^k + \lambda_k s^k \qquad \ldots(6)$$

حيث ان D_k تقوم مقام التقريب في تصغير f على طول اتجاه البحث s^k من x^k. أخيرا" ، نطور H_k بواسطة إضافتها الى المصفوفة D_k من الرتبة الأولى او الثانية التصحيحية لكي نحصل على

$$H_{k+1}=H_k+D_k \qquad \ldots(7)$$

حيث ان D_k تعتمد على

$$\sigma^k = x^{k+1} - x^k \qquad \ldots(8)$$

$$y^k = g(x^{k+1}) - g(x^k) \qquad \ldots(9)$$

حيث $g(x) = \Delta f(x)$.

هنالك العديد من الاختيارات الممكنة للمصفوفة التصحيحية D_k من أهمها (Fletcher and Powell, 1963 ; Broyden, 1970 ; Davidon , 1959 ; Al-Bayati, 1991).

نختار اولاً مصفوفة الرتبة الثانية المطورة BFGS

حيث ان D_k تعرف بالشكل الآتي:

$$D_k = \frac{H_k y^k (y^k)^T H_k}{(y^k)^T H_k y^k} + \frac{\sigma^k (\sigma^k)^T}{(\sigma^k)^T y^k} + w^k (w^k)^T \qquad \ldots(10)$$

عندما

$$w^k = ((y^k)^T H_k y^k)^{\frac{1}{2}} \left(\frac{\sigma^k}{(\sigma^k)^T y^k} - \frac{H_k (y^k)^T}{(y^k)^T H_k y^k} \right)$$

ويمكن تطوير H_k مع طريقة التدرج المترافق عن طريق اتجاه البحث s^{k+1} لكي نحصل على

$$s^{k+1} = -H_{k+1} g^{k+1} + \frac{(g^{k+1})^T H_{k+1} g^{k+1}}{(g^k)^T g^K} s^k \qquad \ldots(11)$$

حيث ان H_{k+1} هي مصفوفة متناظرة وموجبة التعريف:

لمزيد من المعلومات عن خوارزمية BFGS انظر (Al-Bayati and Ahmed, 1996).

4. خطوات خوارزمية BFGS المتوازية والمشروطة مسبقا":

(Al-Bayati and Aref, 2001)

خوارزمية (1)

الخطوة الاولى:– لتكن x^0 نقطة اختيارية مناسبة, H_0 مصفوفة ابتدائية وتكون عادة مصفوفة الوحدة $\delta^1, \delta^2, \ldots, \delta^n$ تمثل n من الازاحات المستقلة $\delta^j = h_j e^j$ حيث h_j هو حجم الخطوة المناسبة.

الخطوة الثانية :– $s^0 = -H_1 g(x^0)$

الخطوة الثالثة :– عندما k=1 . احسب $x^{k+1} = x^k + \lambda_k s^k$,

حيث λ_k تحسب من $\underset{\lambda_k}{arg\ \ min}\ \ f(x^k + \lambda_k s^k) = ELS$

الخطوة الرابعة :–راجع إذا كان $\left\| g(x^{k+1}) \right\| < \varepsilon$ توقف حيث ε عدد صحيح موجب .

الخطوة الخامسة :– احسب في التوازي $g(x^k), g(x^{k1}), \ldots, g(x^{kn}), f(x^k)$,

حيث $x^{kj} = x^k + \delta^j$, $y^{kj} = g(x^{kj}) - g(x^k)$

واحسب $\gamma^{kj},\ \ j = 1,2,\ldots,n$ من خلال

$$\gamma^{kj}=\frac{\delta^{j}}{(y^{kj})^{T}\delta^{j}}-\frac{H_{k}y^{kj}}{(y^{kj})^{T}H_{k}y^{kj}}$$

وتحسن أن من خلال

$$H_{k+1}=H_{k}-\frac{H_{k}y^{kj}(y^{kj})^{T}H_{k}}{(y^{kj})^{T}H_{k}y^{kj}}+\frac{\delta^{k}(\delta^{k})^{T}}{(\delta^{k})^{T}y^{kj}}+\left[(y^{kj})^{T}H_{k}y^{kj}\right]^{\frac{1}{2}}(\gamma^{kj})(\gamma^{kj})^{T}$$

الخطوة السادسة :احسب الاتجاه المطور

$$s^{k+1}=-H_{k+1}g^{k+1}+\frac{(g^{k+1})^{T}H_{k+1}g^{k+1}}{(g^{k})^{T}g^{K}}s^{k}$$

الخطوة السابعة : أجريت عملية الاسترجاع من البداية بوضع $k=1$ او استرجاع Powell او $(g^{k+1})^{T}H_{k}g^{k+1}>0$ حيث ان الشرط الأخير يضمن H_k مصفوفة موجبة التعريف.

لمزيد من المعلومات انظر (Al-Bayati and Aref, 2001) حول موضوع التوازي .

5. الخوارزمية المقترحة الى BFGS المتوازية والمشروطة المقيدة :

خوارزمية (2)

الخطوة الأولى والثانية / كما في خوارزمية (1)

الخطوة الثالثة : احسب

$$\phi(x,r)=f(x)+\sum_{k=1}^{m}r_{k}B\{c_{i}(x)\}$$

حيث ان r قياسي موجب و B{c} دالة رتبية الى c .

الخطوة الرابعة :- احسب $x^{k+1}=x^{k}+\lambda_{k}s^{k}$, حيث λ_k تحسب من

$$\underset{\lambda_k}{arg\ \ min}\ \ f(x^{k}+\lambda_{k}s^{k})=ELS$$

الخطوة الخامسة :-افحص إذا كان $\|g(x^{k+1})\|<\varepsilon$ توقف (حيث ε عدد صحيح موجب) .

والا اذهب الى الخطوة السادسة.

الخطوة السادسة :- احسب في التوازي $f(x^{k}), g(x^{kn}),\ldots,g(x^{k1}),g(x^{k})$,

حيث $x^{kj}=x^{k}+\delta^{j}$, $y^{kj}=g(x^{kj})-g(x^{k})$

واحسب $j=1,2,\ldots,n$, γ^{kj} من خلال

$$\gamma^{kj}=\frac{\delta^{j}}{(y^{kj})^{T}\delta^{j}}-\frac{H_{k}y^{kj}}{(y^{kj})^{T}H_{k}y^{kj}}$$

وتحسن H_k من خلال

$$H_{k+1}=H_k-\frac{H_k y^{kj}(y^{kj})^T H_k}{(y^{kj})^T H_k y^{kj}}+\frac{\delta^k(\delta^k)^T}{(\delta^k)^T y^{kj}}+\left[(y^{kj})^T H_k y^{kj}\right]^{\frac{1}{2}}(\gamma^{kj})(\gamma^{kj})^T$$

الخطوة السابعة : احسب الاتجاه:

$$s^{k+1}=-H_{k+1}g^{k+1}+\frac{(g^{k+1})^T H_{k+1}g^{k+1}}{(g^k)^T g^K}s^k$$

الخطوة الثامنة : افحص التقارب

اذا تحقق $(\varepsilon=0.01)$ $\quad r\sum_{k=1}^{m}\frac{1}{c_k(x)}<\varepsilon,$, توقف. وغير ذلك, إعتبر

$$x=x^*,\quad r_{k+1}=\frac{r_k}{10}$$

بوضع k=k+1 , واذهب الى الخطوة الاولى.

5. النتائج والحسابات: Results and Conclusions

تضمنت إختبارات المقارنة (5) دوال اختبار مقيدة معروفة بابعاد تتراوح بين $1\le n\le 3$ و $1\le c_k\le 7$. تمت مقارنة هذه الخوارزميات بواسطة حساب العدد الكلي للتكرارات number of iteration (NOI) مع العدد الكلي لقيم الدوال number of function evaluation (NOF) , ثم اخذ مقياس التوقف $\left\|g(x^{k+1})\right\|<\varepsilon$.

هناك نوعان من الخوارزميات التي تم اختبارها:

(1) خوارزمية SUMT.

(2) الخوارزمية المقترحة.

الجدول رقم (1) تضمن مقارنة بين الخوارزميتين (1), (2). حيث يبين هذا الجدول ان نتائج الخوارزمية المقترحة أفضل من نتائج الخوارزمية الأصلية.

إن فكرة الاشتراط المتوازي تم استخدامها هنا لاول مرة في الامثلية المقيدة حيث حصلنا على نتائج مشجعة في التجارب العملية.

يشير الجدول (2) الى انه لو أخذنا طريقة SUMT بنسبة 100% لكل من NOI,NOF , نلاحظ ان الخوارزمية المقترحة تحتاج الى 21% من NOI و18% من NOF .

Table (1)

مقارنة الخوارزمية الجديدة مع خوارزمية SUMT

Test function	SUMT algorithm NOI (NOF)	New algorithm NOI (NOF)
1.	32 (129)	26 (106)
2.	24 (81)	21 (73)
3.	37 (137)	27 (99)
4.	13 (50)	10 (42)
5.	39 (122)	31 (100)
Total	145 (519)	115 (420)

Table (2)

	SUMT Algorithm	New Algorithm
NOI	100%	79.3
NOF	100%	82.3

Appendix   7.الملحق

Test functions

(1) $min\, f(x) = (x_1 - 2)^2 + (x_2 - 1)^2$ **s.p(2,7)**

s.t

$x_1 - 2x_2 = -1$

$\frac{-x_1^2}{4} + x_2^2 + 1 \geq 0$

(2) $min \quad f(x) = -x_1^2 x_2$

s.t

$x_1 x_2 + (x_1^2)/2 = 6$

$x_1 + x_2 \geq 0$

(3) $min\, f(x) = x_1^2 + x_2^2$ **s.p(0.9,2)**

s.t.

$x_1 + 2x_2 = 4$

$x_1^2 + x_2^2 \leq 5$

$x_i \geq 0$

(4) $min\, f(x) = (x_1 - x_2)^2 + (x_2 - 2x_2)^2$ **s.p (0,1)**

s.t.

$x_1^2 - x_2 \leq 0$

(5) $min\, f(x) = x_1 x_4 (x_1 + x_2 + x_3) + x_3$ **x_0=(4,3,3,3)**

s.t.

$x_1^2 + x_2^2 + x_3^2 + x_4^2 = 40$

$x_1 x_2 x_3 \geq 25$

$5 \geq x_i \geq 1$

المصادر

[1] Al-Bayati A.Y. and Aref, S. (2001), " Modified parallel Self-Scaling variable metric algorithms", M.Sc. Thesis, Dep. Of Math., University of Mosul.

[2] Al-Bayati, A.Y. and Hamed, E.T. (1998), "New self-scaling sequential algorithm for the minimization of constrained nonlinear function", J. of Dirasat, Jordan, Vol (25), PP. 339-351.

[3] Al-Bayati, A.Y. and Hamed, E.T. (2000), "New Parameter for the inverse Barrier method in constrained nonlinear optimization" J. of Educ. And Sci. Mosul, Vol. (44), PP. 93-110.

[4] Al-Bayati, A.Y. and Shareef, S.G. (2001), " A modification PCG-method for non-linear optimization", Al-Rafidain J.of Sci. Mosul, Iraq.

[5] Al-Bayati A.Y. and Al-Assady N.H (1997), " PCG-Methods for nonlinear optimization J. of Mo'tah, Jordan, Vol (12).

[6] Al-Bayati, A.Y. and Ahmed, H.I. (1997), " Investigation on double update variable storage CG-method", J. of Al-Yarmouk, Jordan, Vol (6), pp. 23-41.

[7] Al-Bayati, A.Y. and Ahmed, H.I. (1996), " Investigation on single update memoryless CG-method", Qatar University Sci. J. Vol (16), pp. 183-192.

[8] Al-Bayati, A.Y. and Mahammed Ali, M.M. (2002), "Multi-Step hybrid CG-algorithm for Unconstrained optimization, Al-Rafidain J. of Sci., Mosul, Iraq, Vol. (13), pp. 94-101.

[9] Al-Bayati, A.Y. (1991), " A New Family of Self-Scaling Variable Metric Algorithms for Unconstrained Optimization", J of Education on Science, Mosul University.

[10] Axelsson, O. (1974), " On Preconditioning and Convergence Acceleration in Sparse Matrix Problems", CERN data Hunding Division Report, pp. 74-100.

[11] Buckley, A.G. and Lenir, A. (1983), "Quasi-Like Variable Storage Conjugate Gradients", Math. Prog. 27. Pp. 155-175.

[12] Buckley, A.G. (1978a), " A Combined CG-QN Minimum Algorithm, Math. Prog. 15, pp. 200-210.

[13] Broyden, C. G. (1970), "The Convergence of a Class of Double-Rank Minimization Algorithm", J. Inst. Apl. 6. pp. 66-90 and 222-23.

[14] Davidon, W.C. (1959), "Variable Metric Method for Minimization". AERC and D. Report A. NI-5990. Argonne II.

[15] Fletcher, R. and Powell, M.J.D. (1963), "A Rapidly Convergent Descent Method for Minimization", Comp. J. 6, pp. 163-168.

[16] Fiacco, A.V. and McCormik, G.P. (1968), " Non-Linear Programming, Sequential Unconstrained Minimization Technique". Wiley, New York.

[17] Liu, D.C. and Nocedal, J. (1988), "On the Limited Memory BFGS Method for Large-Scale Optimization:, Rep. NAM 03, Dept. of Electrical Engineering and Computer Science, North Wastrel University, Envonstor II.

[18] Nazareth, L. and Nocedal, J. (1982). " Conjugate Direction Methods with Variable Storage", Math. Prog. 23, pp. 341-348.

[19] Nazareth, L. (1979), " A Relationship Between the BFGS and Conjugate Algorithms and Implications for New Algorithms", SIAM J. Number, Anal. 16. Pp. 794-800.

[20] Shanno, D.F. (1981), On PCG-Methods" In: Non-Linear Programming 4. Eds. O. L. Mangasarian, R.R. Mayer and S.M. Robinson, Academic Press, New York, of QN Methods for function Minimization. Math. Comp. 24, pp. 647-656.