

Blok Krylov metotlarının yüksek başarımlı uygulanması

Ali DİNLER^{*1}, Kamil ORUÇOĞLU²

¹İTÜ Mühendislik Bilimleri Bölümü, 34469, Ayazağa, İstanbul

²İTÜ Fen Edebiyat Fakültesi, Matematik Bölümü, 34469, Ayazağa, İstanbul

Özet

Kısmi türevli bir denklemde parametreye bağlı değişen sınır koşulları, hareketli bir kaynak terimi ya da değişken kuvvet terimi ile özel veya optimize bir sayısal çözümün aranması durumunda denklemin bu parametrenin değişen değerlerine göre çözülmesi gerekir. Bu da sağ-taraf vektörü değişen bir lineer sistemin birçok defa aynı katsayı matrisi ile çözümünü gerektirir. Böyle bir problem üç-boyutta ise problemin sayısal çözümünde ortaya çıkan katsayı matrisi büyüktür ve bu durumda uygun iteratif metot kullanılması önemlidir. Bu gibi problemlerde katsayı matrisi değişmediği halde, m tane sağ taraf ile lineer sistemin m defa çözülmesi gerekir. Bunun yerine $AX = B$ blok lineer sistemi oluşturulur, burada A katsayı matrisi, X bilinmeyenler matrisi ve $B = [b_1, b_2, \dots, b_m]$ sağ taraf vektörlerinden oluşan matristir. Bu blok sistem daha verimli ve hızlı bir şekilde blok Krylov metotları ile bir defada çözülebilir. Geliştirmekte olduğumuz Blok İteratif Metotlar paketi (BİM++), simetrik ve simetrik-olmayan blok lineer sistemleri çözen blok Krylov metotlarının yüksek başarımlı uygulamasıdır. Bu çalışmada, blok Krylov metotları ve özellikle blok GMRES (generalized minimal residual) metodu kısaca tanıtıldıktan sonra bu metotlar için geliştirmekte olduğumuz BİM++ paketinin performansı gösterildi. Dahası bazı örnek matrisler üzerinde, çok sağ-taraflı sistemlerin çözümünde blok metotların blok-olmayan metotlara göre üstünlük gösterebildiği ve sağ taraf sayısının artması ile bu blok-olmayan metotlara olan üstünlüğün arttığı gösterildi.

Anahtar Kelimeler: İteratif metotlar, blok Krylov altuzayı metodu.

*Yazışmaların yapılacağı yazar: Ali DİNLER. dinlera@itu.edu.tr; Tel: (212) 285 69 91.

Makale metni 26.05.2009 tarihinde dergiye ulaşmış, 03.09.2008 tarihinde basım kararı alınmıştır. Makale ile ilgili tartışmalar 31.03.2010 tarihine kadar dergiye gönderilmelidir.

High performance implementation of block Krylov methods

Extended abstract

Multiple large linear systems with a same coefficient matrix but different right-hand side vectors arise in several problems of engineering such as in solution of Navier-Stokes equations in different directions (Dinler, 2007). Moreover, seeking an optimum or a special solution of a multi-dimensional PDE with a parameter-dependent moving source term or changing force term, or variable boundary conditions requires repeated solution of this PDE with changing values of the parameter. In that case, we need to solve multiple large linear systems with a same coefficient matrix but different right-hand side vectors. To solve these large block linear systems efficiently, block Krylov subspace methods (in short block Krylov methods) are an important tool.

Krylov subspace methods are projection methods that search the solution of a linear system iteratively in the Krylov subspace. Comparing with direct methods such as Gauss elimination or LU decomposition, direct methods are prohibitive to solve large linear systems however Krylov subspace methods are efficient. Moreover in several engineering problems, block Krylov methods are an important tool for solution of large linear systems with multiple right hand sides. In addition to that, block methods reduce memory-access cost and increase efficient data reuse because they allow multiple matrix-vector products with less memory access. However, their robustness problem and preconditioning are difficult to handle.

So, to come up with a robust implementation of block Krylov methods, we initiated BIM++ (Block Iterative Methods) project, which is devoted to research and development of block Krylov methods to make them easier to use. BIM++ package is a high performance implementation of block Krylov methods written in C/C++ under GPL license, for solution of both symmetric and nonsymmetrical large linear systems with multiple right hand sides. In addition to this BIM++ includes important non-block Krylov methods such as CG (conjugate gradient), BICGSTAB (bi-conjugate gradient stabilized), GMRES (generalized minimal residual) and GMRES(m) (GMRES restarted).

While developing BIM++, we stick to “keep-it-easy rule” and put performance first. Easy-to-use and easy-to-install are also important. While developing BIM++, our goals are 1) creating an environment for easy implementation and easy development of advanced numerical linear algebra algorithms, 2) creating a high performance and user friendly package, 3) enable to utilize C++ technologies, 4) enable to programming Krylov methods shorter and more readable, 5) enable to use block Krylov methods handy, 6) creating reusable, portable and easily upgradeable package, 7) enable to use BIM++ in Windows PCs, 8) creating a clear documentation and a webpage.

To achieve these goals, we borrowed many good ideas from LAPACK++ and IML++. We built BIM++ over LAPACK. We improved and used CPPLapack interface (Onishi, 2003) for FORTRAN LAPACK to make the package more user friendly. Also we tried to use generic programming to achieve performance.

In this technical note, we demonstrate test results of BIM++ package for different matrices and different number of right-hand side vectors. Our results are:

- *Block Krylov methods may be efficient for linear systems with multiple right-hand sides.*
- *Block Krylov methods decrease the memory access cost as a result convergence can be achieved in less time.*
- *Performance of block Krylov methods depends on performance of matrix-matrix multiplication performance.*
- *Using ATLAS library, which is hardware dependent optimal version of LAPACK, under BIM++ increases performance.*
- *Superiority of block Krylov methods over non-block Krylov methods increases in case there are more right-hand side vectors. Moreover, Convergence may be improved with suitable extra right-hand vectors.*
- *Comparing with non-block Krylov methods, convergence behavior of block Krylov methods are different so different and special preconditioning is necessary.*

Keywords: *Iterative methods, block Krylov subspace methods.*

Giriş

Blok Krylov metotlarının yüksek başarılı uygulanması için geliştirmekte olduğumuz blok iteratif metotlar paketi, BİM++, simetrik ve simetrik olmayan blok lineer sistemleri çözen blok Krylov metotlarının bir uygulamasıdır. Blok Krylov metotları son yıllarda mühendislik problemlerinin çözümünde çok önem kazanmıştır (Dinler, 2007). Blok Krylov metotları birden fazla sağ taraflı büyük lineer sistemlerin çözümünde çok önemlidir. Aynı katsayı matrisine ama farklı sağ taraflara sahip bir sistem, blok metotlar ile bir kerede çözülebilir. Blok Krylov metotları tek bir vektörle değil vektör blokları ya da alt-matrisler ile iş yaparlar, bu da bellek kullanımını iyileştirir. Bu nedenle tek sağ taraflı sistemler için de blok yaklaşım kullanılabilir (Baker vd. 2003, Li 1997). Bilimsel hesaplama da bellek yönetimi ve kullanımının çok önemli hale gelmesi, blok metotları ön plana getirmiştir (Dongarra vd., 1998; Lam vd., 1991).

LAPACK (Linear Algebra PACKage) kütüphanesi üzerine C++ programlama diliyle geliştirilmekte olan bu paketin amaçları, CPPLapack arayüzünü (Onishi, 2003) geliştirerek

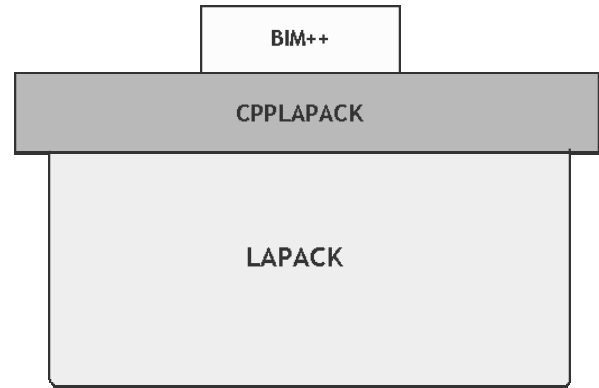
- Blok Krylov metotlarını -özellikle hesaplamalı mekanik problemlerinin çözümünde- kolay kullanabilmek,
- C++ programlama dilinin sahip olduğu teknolojilerden en kolay şekilde faydalanmak,
- İleri sayısal lineer cebir algoritmalarının daha kısa ve daha anlaşılır şekilde programlanmasını sağlamak,
- Tekrar kullanılabilir, taşınabilir ve kolayca geliştirilebilir bir paket oluşturmak,
- Windows işletim sisteminde de çalışabilir yüksek başarılı bir paket oluşturmaktır.

LAPACK kütüphanesi, kullanması dikkat isteyen ve kolayca karıştırılabilir çok parametrelili fonksiyonlardan oluşur. Bu yüksek başarılı kütüphane için C++ teknolojisi ile kullanmayı kolaylaştıracak birçok arayüzler ve kütüphaneler yazılmıştır. Bunlardan bazıları LAPACK++,

CvmLib, FLENS, CPPLapack, SparseLib++, MV++ ve IML++ 'dır. Bunların bazısının hala kullanımı kolay değildir, bazısının kurulumu özel derleyici ve/veya derleyici seçenekleri gerektirir, bazısının dokümantasyonu iyi değildir, bazısı ise yeterince geliştirilmemiş ya da geliştirilmesi bırakılmıştır.

Blok GMRES (Generalized Minimal Residual) metodu gibi yeni sayısal lineer cebir algoritmaları için kolay kullanılabilir bir ortam oluşturmak ve LAPACK Kütüphanesini kullanan yüksek başarılı bir paket oluşturmak için CPPLapack arayüzü seçildi. Sonra bu arayüze paketin amaçlarına uygun eklemeler yapıldı. Böylece kurulum gerektirmeyen başlık (header) dosyalarından oluşan BİM++ paketi oluşturuldu.

BİM++, blok Krylov metotlarını kolayca kullanabilmek için yazıldı. BİM++ 'nın kolayca kullanılması için CPPLapack arayüzüne eklemeler yapıldı ve LAPACK'e en kolay ulaşmayı sağlayan KEYcpplapack.h başlık dosyasını oluşturuldu. Dahası olabildiğince az programlama bilgisi ile bu paket kullanılır hale getirildi. Şekil 1'de BİM++ paketinin yapısı görülebilir.



Şekil 1. BİM++ paketinin yapısı

Bu çalışmada blok Krylov metotları kısaca tanııldıktan sonra bu metotlar için geliştirmekte olduğumuz BİM++ paketinin performansı gösterilecektir.

Blok Krylov metotları

Krylov metotları, genel anlamda, $Ax = b$ lineer sisteminin çözümünü Krylov altuzayında arayan

ve matris-vektör çarpımları ile $Ax = b$ lineer sistemini iteratif çözen metotlardır. Burada, A tekil olmayan katsayı matrisi, x bilinmeyenler vektörü, b de sağ taraf vektörüdür. Katsayı matrisi çok büyük olduğunda, Gauss yok-etme gibi direk lineer sistem çözücü yöntemler yerine iteratif bir yöntem kullanılması mecburidir. Dahası, bazı problemlerde katsayı matrisi A değişmediği halde, farklı sağ taraflar ile $Ax = b_i$, $i = 1, \dots, m$ lineer sisteminin m defa çözülmesi gerekir. Bunun yerine $AX = B$, $B = [b_1, b_2, \dots, b_m]$ blok sistemi oluşturulabilir ve çok daha verimli ve hızlı bir şekilde blok Krylov metotları ile bir defada çözülebilir, (Dinler, 2007). Ancak bu durumda artık matris-vektör çarpımları yerine, matris-matris çarpımları ile sonuca gidilir.

En bilinen blok Krylov metotları; blok Conjugate Gradient (CG), blok Bi-Conjugate Gradient (BICG), blok Bi-Conjugate Gradient STABILized (BICGSTAB), blok Generalized Minimal RESidual (GMRES) ve GMRES restarted (GMRES(m))'dır, (El Guannouni vd., 2003; Saad, 1996; Simoncini ve Szyld, 2007).

Blok GMRES Algoritması: $n \times n$ 'lik ve p sağ taraf sayısı ya da blok genişliği olmak üzere, $AX = B$ blok sistemi için algoritma şöyledir; $n = m.p$ ve m blok Krylov altuzayının boyutu olmak üzere

1. Başlangıçta bir $n \times p$ 'lik bir X_0 matrisi seçilir
2. $R_0 = B - AX_0$
3. R_0 QR ayrışımı ile $R_0 = V_1 R_{10}$ şeklinde ayrıştırılır.
4. For $j = 1, \dots, m$ Do :
5. $W_j = AV_j$
6. For $i = 1, \dots, j$ Do :
7. $H_{i,j} = V_i^T W_j$
8. $W_j = W_j - V_i H_{i,j}$
9. End Do
10. W_j QR ayrışımı ile $W_j = V_{j+1} H_{j,j+1}$ şeklinde ayrılır.
11. End Do
12. $V_{m+1} = [V_1, V_2, \dots, V_{m+1}]$
13. $V_m = [V_1, V_2, \dots, V_m]$
14. $\beta = V_{m+1}^T B$
15. H QR ayrışımı ile $H = Q_H R_H$ şeklinde ayrılır.

$$16. G_m = Q_H^T \beta$$

17. $R_H Y_m = G_m$, sistemi üst üçgensel $n \times n$ 'lik bir sistem buradan Y_m bulunur.

$$18. X_m = V_m Y_m$$

Burada V_{m+1} $n \times (m+1)p$ 'lik matris, V_m $n \times n$ 'lik matris, B sağ taraf matrisi, X ise sonuç matrisidir.

Blok Krylov metotları yukarıda görüldüğü gibi çok sayıda matris-matris çarpım işlemleri gerektirir. Bu nedenle BİM++ paketinin performansı doğrudan matris-matris çarpım işleminin performansına bağlıdır.

BİM++ paketinin performansı

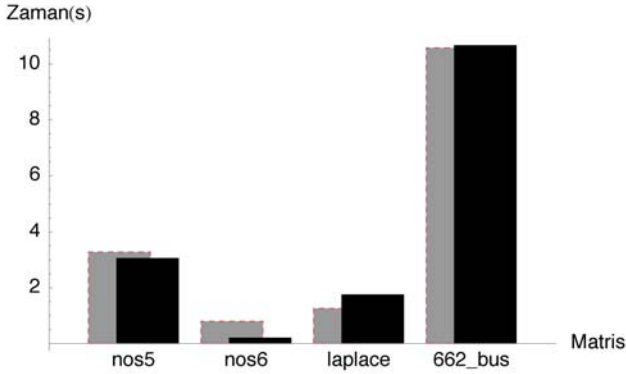
İki farklı bilgisayar sisteminde, 1) Mac OSX Powerbook 1.5 GHz, LAPACK kütüphanesi kullanılarak, 2) Windows Pentium 4 1.6 GHz, ATLAS (Automatically Tuned Linear Algebra Software)-LAPACK kütüphanesinin donanıma bağlı olarak optimize edilmiş sürümü- kütüphanesi kullanılarak, 2 ve 5 sağ taraf ile Tablo 1'deki 10 farklı reel katsayı matrisi ile testler yapıldı.

Tablo 1. Test yapılan katsayı matrisleri ve özellikler

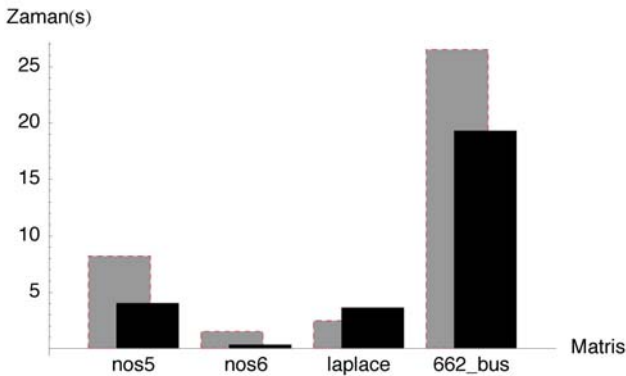
Matris İsmi	Boyutu	Özelliği
nos5	468	Simetrik pozitif belirli
nos6	675	Simetrik pozitif belirli
laplace	900	Simetrik pozitif belirli
662_bus	662	Simetrik pozitif belirli
steam1	240	Anti-simetrik
fs_541_1	541	Anti-simetrik
hor131	434	Anti-simetrik
west0655	655	Anti-simetrik
mfce	765	Anti-simetrik
impcol_e	225	Anti-simetrik

[Http://math.nist.gov/MatrixMarket/](http://math.nist.gov/MatrixMarket/) internet adresinden alınan yukarıdaki katsayı matrisleri ile çok sağ taraflı lineer sistemlerin çözümü için

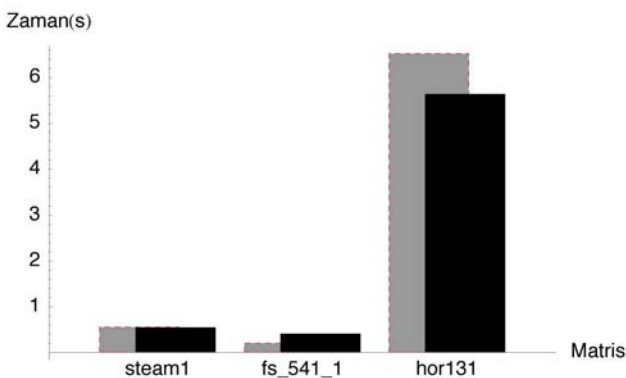
testler yapıldı. Bu testler sonucu elde edilen çözüm zamanları blok ve blok olmayan yöntemler için Şekil 2 – 7’de verilmiştir.



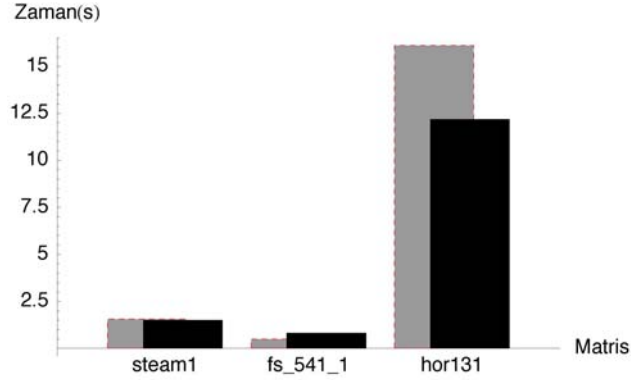
Şekil 2-a. İki sağ taraflı blok CG metodu (siyah) ile iki kere CG metodunun (gri) LAPACK ile çalıştırılması ile elde edilen sonuçlar



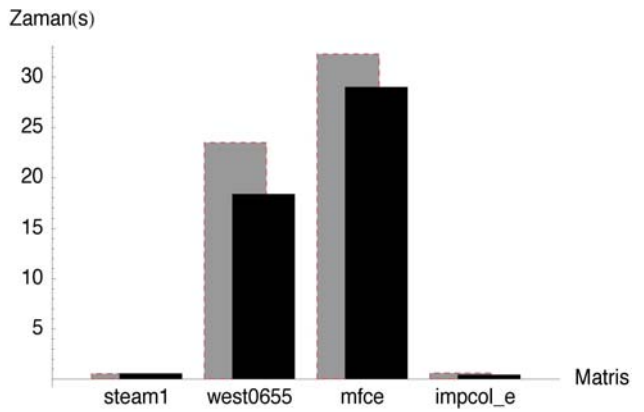
Şekil 2-b. Beş sağ taraflı blok CG metodu (siyah) ile beş kere CG metodunun (gri) LAPACK ile çalıştırılması ile elde edilen sonuçlar



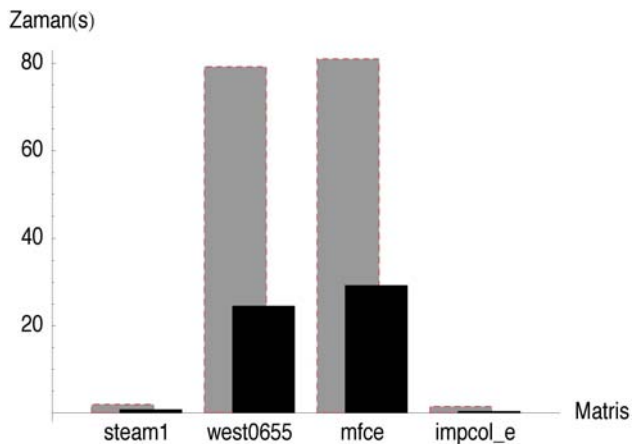
Şekil 3-a. İki sağ taraflı blok BICGSTAB metodu (siyah) ile iki kere BICGSTAB metodunun (gri) LAPACK ile çalıştırılması ile elde edilen sonuçlar



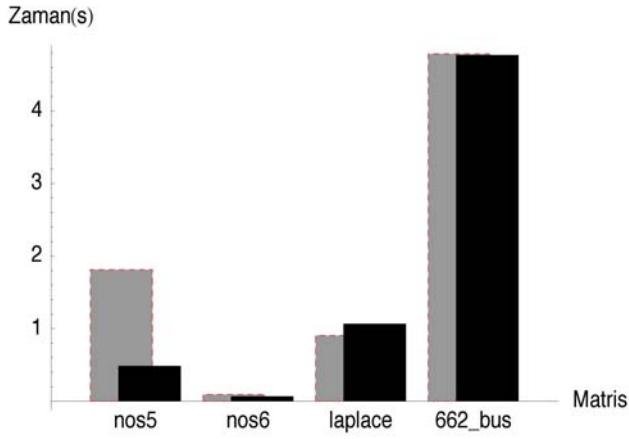
Şekil 3-b. Beş sağ taraflı blok BICGSTAB metodu (siyah) ile beş kere BICGSTAB metodunun (gri) LAPACK ile çalıştırılması ile elde edilen sonuçlar



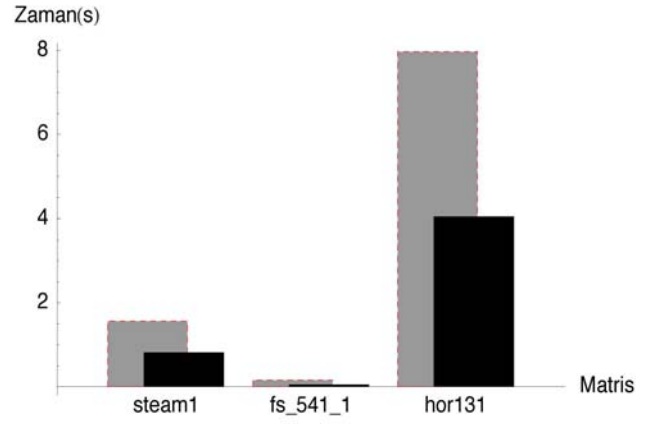
Şekil 4-a. İki sağ taraflı blok GMRES metodu (siyah) ile iki kere GMRES metodunun (gri) LAPACK ile çalıştırılması ile elde edilen sonuçlar



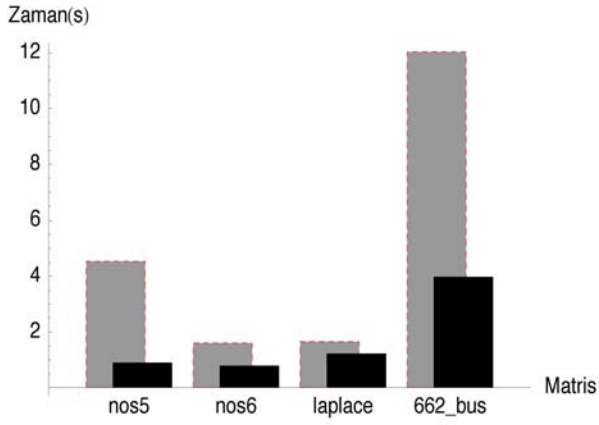
Şekil 4-b. Beş sağ taraflı blok GMRES metodu (siyah) ile beş kere GMRES metodunun (gri) LAPACK ile çalıştırılması ile elde edilen sonuçlar



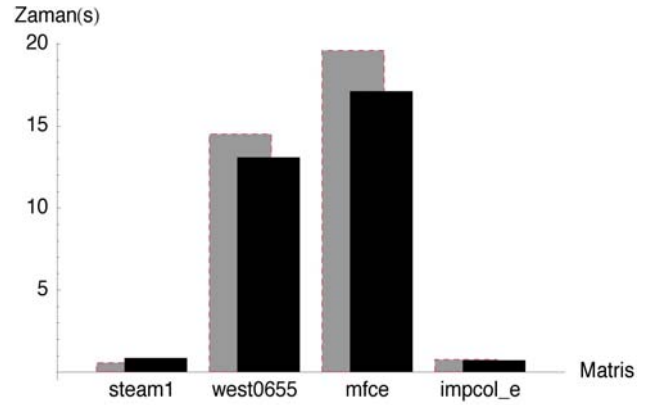
Şekil 5-a. İki sağ taraflı blok CG metodu (siyah) ile iki kere CG metodunun (gri) ATLAS ile çalıştırılması ile elde edilen sonuçlar



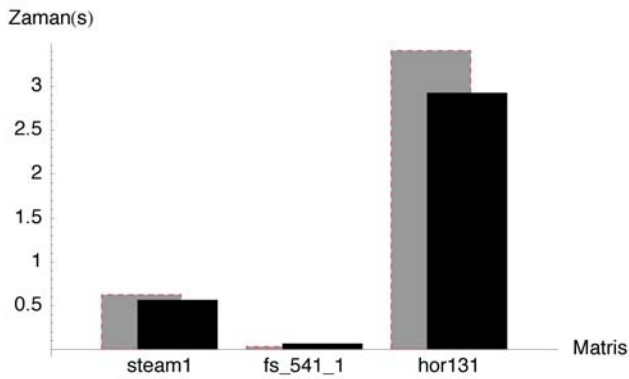
Şekil 6-b. Beş sağ taraflı blok BICGSTAB metodu (siyah) ile beş kere BICGSTAB metodunun (gri) ATLAS ile çalıştırılması ile elde edilen sonuçlar



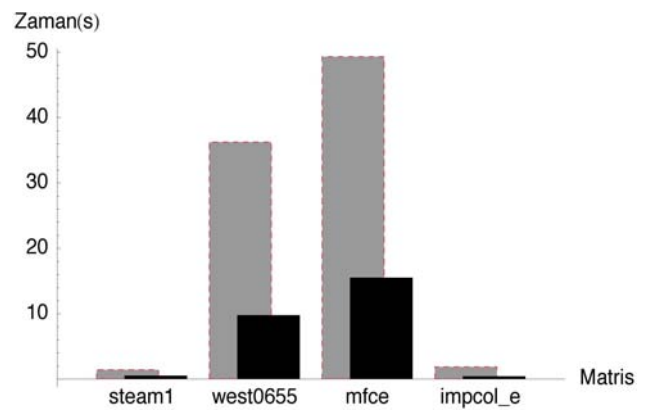
Şekil 5-b. Beş sağ taraflı blok CG metodu (siyah) ile beş kere CG metodunun (gri) ATLAS ile çalıştırılması ile elde edilen sonuçlar



Şekil 7-a. İki sağ taraflı blok GMRES metodu (siyah) ile iki kere GMRES metodunun (gri) ATLAS ile çalıştırılması ile elde edilen sonuçlar



Şekil 6-a. İki sağ taraflı blok BICGSTAB metodu (siyah) ile iki kere BICGSTAB metodunun (gri) ATLAS ile çalıştırılması ile elde edilen sonuçlar



Şekil 7-b. Beş sağ taraflı blok GMRES metodu (siyah) ile beş kere GMRES metodunun (gri) ATLAS ile çalıştırılması ile elde edilen sonuçlar

Sonuçlar

Elde edilen sonuçlar aşağıdaki gibi özetlenebilir:

- Çok sağ taraflı sistemlerin çözümünde blok metotlar blok olmayan metotlara göre ciddi üstünlük gösterebilmektedir.
- Blok algoritmalar bellek yönünden daha verimli çalıştığından aynı iterasyon sayısında bile daha az zamanda sonuca gidebilmektedir.
- Blok Krylov metotlarının performansı, verimli bellek kullanımına ve matris-matris çarpımının performansına bağlı olduğundan ATLAS kütüphanesinin kullanılması ile blok Krylov metotlarının performansı daha da iyileşmektedir.
- Sağ taraf sayısının artması blok olmayan metotlara göre blok metotların üstünlüğünü arttırmaktadır. Dahası bazı örneklerde yakınsamayı da hızlandırmıştır ki bu çok önemli bir sonuçtur.
- Blok iteratif metotların yakınsaklık özelliği blok olmayan iteratif metotlardan farklıdır. Yakınsamanın iyileştirilmesi için önkoşullandırıcılar (preconditioners) gerekmektedir.

Kaynaklar

Baker A. H., Dennis J. M., Jessup E. R., (2003). An Efficient Block Variant of GMRES, Report CU-CS-945-03, University of Colorado, Department of Computer Science.

- Dinler A., (2007). Navier-Stokes denklemlerinin Blok Krylov tekniği ile paralel çözümü, *Yüksek Lisans Tezi*, İTÜ Bilişim Enstitüsü, İstanbul.
- Dongarra J., Duff I. S., Sorensen D. C., Van Der Vorst H. A., (1998). Numerical Linear Algebra for High-Performance Computers, SIAM Software Environments Tools.
- El Guannouni A., Jbilou K., Sadok H. (2003). A block version of Bicgstab for linear systems with multiple right-hand sides. *Electronic Transactions on Numerical Analysis*, **16**, 129-142.
- Lam M. S., Rothberg E. E., Wolf M. E., (1991). Cache performance and optimizations of blocked algorithms, in *Proceedings of Sixth International Conference on Architectural Support for Languages and Operating System*.
- Li G., (1997). A block Variant of the GMRES Method on Massively Parallel Processors, *Parallel Computing*, **23**, 1005-1019.
- Onishi Y., (2003). CPPLapack C++ class wrapper for BLAS and LAPACK, yazılım paketi, <http://cpplapack.sourceforge.net/>.
- Saad Y., (1996). Iterative Methods for Sparse Linear Systems, PWS Publishing Company.
- Simoncini V., Szyld D. B., (2007). Recent developments in Krylov subspace methods for linear systems, *Numerical Linear Algebra with Applications*, **14**, 1-59.