Block conjugate gradient algorithms for linear systems with multiple right-hand sides

Gérard MEURANT

https://gerard-meurant.fr

Porto, May 2026

Joint work with Petr Tichý (Charles University, Prague)

We would like to iteratively solve linear systems

$$AX = B$$

where A is real symmetric positive definite of order n and X and B are $n \times m$ real matrices

There are several ways to solve this system:

-if the m right-hand sides are not all available at the same time and not too different, one can use seed methods

-the so-called global methods using a matrix inner product $\langle X, Y \rangle = \text{trace}(X^T Y)$

-block methods

The block conjugate gradient (OL-BCG) algorithm was introduced by Dianne O'Leary in 1980:

1: **input** A, B, X_0
2: $R_0 = B - AX_0$
3: $P_0 = R_0$
4: **for** $k = 1, \ldots$ until convergence **do**
5: $\quad \Upsilon_{k-1} = (P_{k-1}^T A P_{k-1})^{-1} (R_{k-1}^T R_{k-1})$
6: $\quad X_k = X_{k-1} + P_{k-1} \Upsilon_{k-1}$
7: $\quad R_k = R_{k-1} - A P_{k-1} \Upsilon_{k-1}$
8: $\quad \Xi_k = (R_{k-1}^T R_{k-1})^{-1} (R_k^T R_k)$
9: $\quad P_k = R_k + P_{k-1} \Xi_k$
10: **end for**

The coefficients Υ_{k-1} and Ξ_k are $m \times m$ matrices

A known difficulty in this algorithm is that the columns of the block vectors R_k or P_k may become linearly dependent

In such cases, the block coefficients Υ_{k-1} or Ξ_k cannot be computed. This is generally called a *breakdown*

In finite precision arithmetic, the columns of R_{k-1} or P_{k-1} may become nearly linearly dependent and the inverses of $R_{k-1}^T R_{k-1}$ or $P_{k-1}^T A P_{k-1}$ cannot be computed reliably

We will refer to this as a *rank deficiency* or simply as a *deficiency*. This typically leads to delayed convergence or even to divergence of the algorithm

Without breakdown or deficiencies, it holds

$$R_i^T R_j = 0, \quad P_i^T A P_j = 0, \quad R_i^T P_j = 0 \quad i \neq j$$

and the algorithm converges. Moreover, it minimizes

$$\text{trace}((X - X_k)^T A (X - X_k))$$

see D.P. O'Leary, The block conjugate gradient algorithm and related methods, LAA, vol. 29 (1980), pp. 293-322

Let us look at an example:

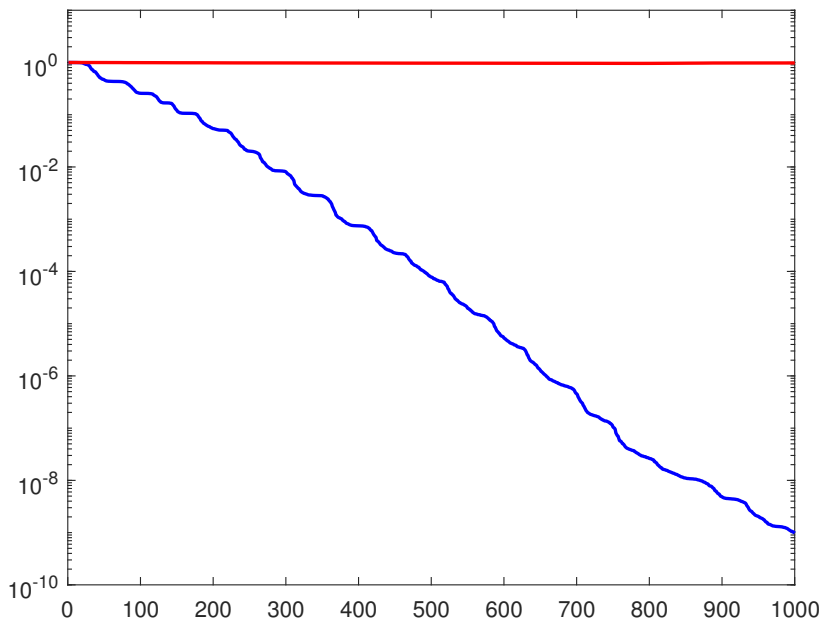
The symmetric positive definite matrix `nos7`[1] (a 3D problem in the unit cube) is of order 729 with a condition number equal to $2.3745\ 10^9$

We use right-hand sides with 10 columns and $X_0 = 0$

[1] Available at http://sparse.tamu.edu

Example with OL-BCG

$B_1 = randn(729, 10)$
B_2 such that $B_2(:, 2) = 5B_1(:, 1) + 10^{-6}\, B_1(:, 2)$ (an initial near-deficiency)



`nos7`, relative trace error norm, B_1 (blue) and B_2 (red)

A common way to solve these problems is to use *deflation*:

Monitor the rank of the blocks and reduce the block size if necessary, but it is not always easy to know which vectors to remove

A different and more efficient remedy to the rank deficiency problem was proposed by Augustin A. Dubrulle (1935-2016). He introduced a change of variables to address rank deficiency and described several variants

$$R_k = Q_k \Sigma_k, \quad P_k = S_k \Sigma_k$$

Dubrulle "cheated" a little bit. To derive the algorithm he assumed that Σ_k is nonsingular

However, the inverse of Σ_k is not involved in the final formulas

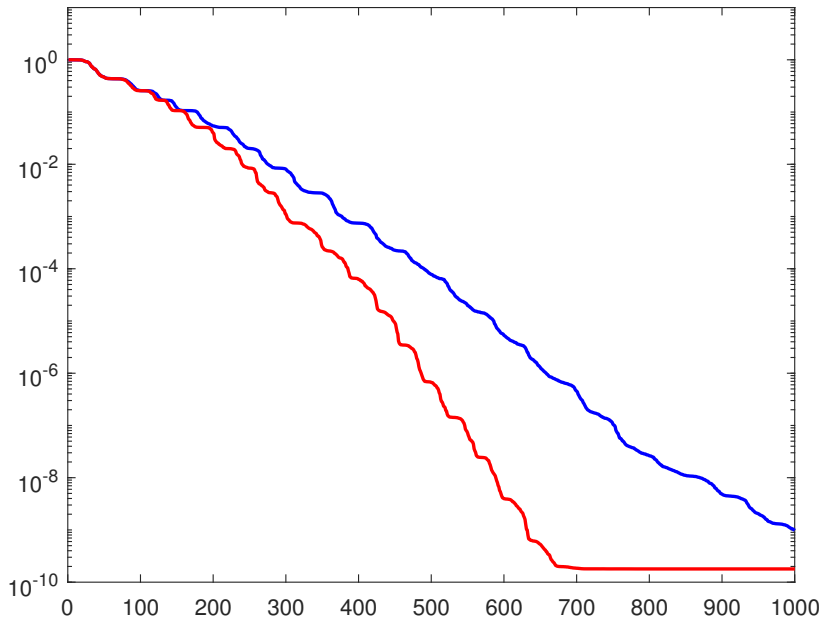
DR-BCG

1: **input** A, B, X_0
2: $R_0 = B - AX_0$
3: $[Q_0, \Psi_0] = \mathtt{qr}(R_0)$
4: $S_0 = Q_0$, $\Sigma_0 = \Psi_0$
5: **for** $k = 1, \ldots$ until convergence **do**
6: $\quad \Pi_{k-1} = (S_{k-1}^T A S_{k-1})^{-1}$
7: $\quad X_k = X_{k-1} + S_{k-1} \Pi_{k-1} \Sigma_{k-1}$
8: $\quad [Q_k, \Psi_k] = \mathtt{qr}(Q_{k-1} - A S_{k-1} \Pi_{k-1})$
9: $\quad S_k = Q_k + S_{k-1} \Psi_k^T$
10: $\quad \Sigma_k = \Psi_k \Sigma_{k-1}$
11: **end for**

An economy-size QR factorization of $Q_{k-1} - AS_{k-1}\Pi_{k-1}$ is performed (line 8) using Householder reflections in such a way that Q_k, which is $n \times m$, remains orthonormal ($Q_k^T Q_k = I$), and Ψ_k is an upper triangular $m \times m$ matrix with nonnegative diagonal entries

Example

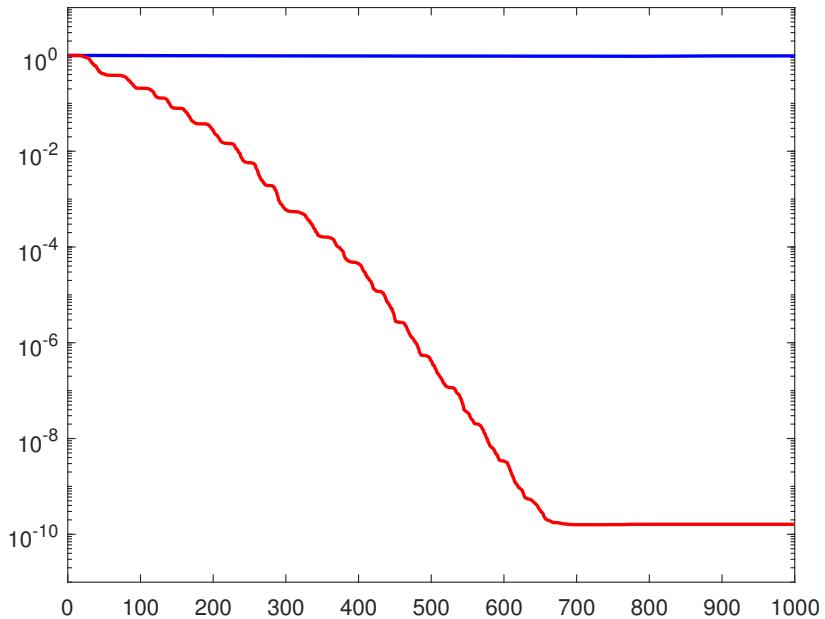
$B_1 = randn(729, 10)$



`nos7`, B_1, relative trace error norm, OL-BCG (blue) and DR-BCG (red)

Example

B_2 such that $B_2(:,2) = 5B_1(:,1) + 10^{-6}\, B_1(:,2)$

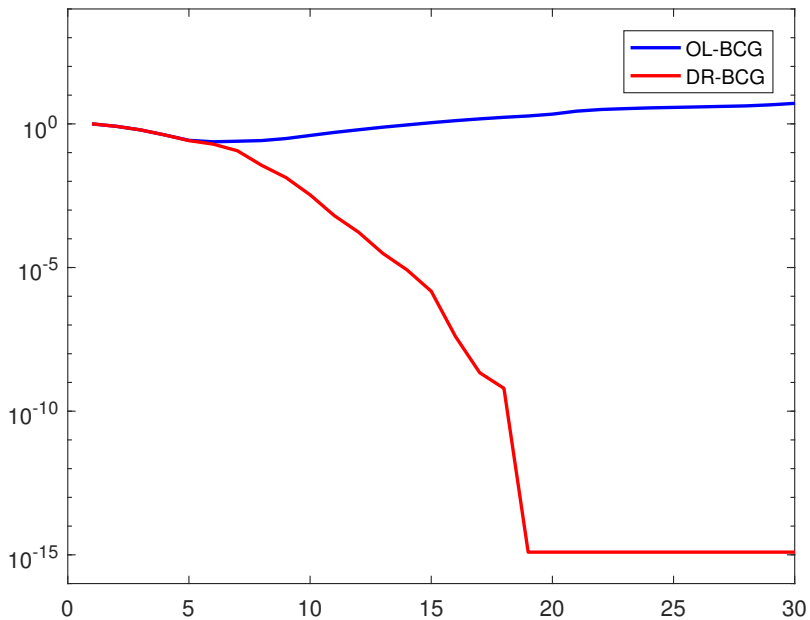


`nos7`, B_2, relative trace error norm, OL-BCG (blue) and DR-BCG (red)

Example with (near) deficiency

The matrix is of order 20 and its condition number is 90.128

The right-hand side with two columns is computed with an optimization code to have a small diagonal entry of Ψ_3 ($\approx 10^{-15}$) in DR-BCG



relative trace error norm, OL-BCG (blue) and DR-BCG (red)

Questions

Dubrulle did not do any mathematical analysis of his algorithms. But, we may ask:

1- Is $S_k^T A S_k$ always nonsingular?

Ψ_k may become singular or nearly singular with small diagonal entries

2- Is the algorithm converging in case of deficiencies?

The answers to these two questions is: Yes

First question

Lemma

Let $C \in \mathbb{R}^{m\times m}$ and $D \in \mathbb{R}^{m\times m}$. Then CD and DC have the same eigenvalues. Moreover, if at least one of C or D is nonsingular, then CD and DC are similar. (Horn and Johnson, Matrix Analysis)

Theorem

For all $k \geq 0$, the matrices $Q_k^T S_k$ are nonsingular, and consequently the block vectors S_k have full column rank. Moreover, all eigenvalues of $Q_k^T S_k$ are equal to one. For $k \geq 1$, we also have

$$S_k^T Q_k = I + \Psi_k S_{k-1}^T Q_k, \tag{1}$$

and both $\Psi_k S_{k-1}^T Q_k$ and $S_{k-1}^T Q_k \Psi_k$ are nilpotent.

Sketch of the proof

The proof is by induction. For $k = 0$, we have

$$Q_0^T S_0 = Q_0^T Q_0 = I,$$

which is nonsingular and has all eigenvalues equal to one. Now, assume that $Q_{k-1}^T S_{k-1}$ is nonsingular and that all its eigenvalues are equal to one. Since S_{k-1} has full column rank, Π_{k-1} is well defined, and

$$\begin{aligned} Q_{k-1}^T S_{k-1} &= Q_{k-1}^T S_{k-1} + I - \Pi_{k-1} S_{k-1}^T A S_{k-1} \\ &= I + (Q_{k-1} - A S_{k-1} \Pi_{k-1})^T S_{k-1} \\ &= I + \Psi_k^T Q_k^T S_{k-1} \end{aligned}$$

Sketch of the proof (2)

Next, it is easy to prove that relation (1) holds, since

$$Q_k^T S_k = Q_k^T \left(Q_k + S_{k-1} \Psi_k^T \right) = I + Q_k^T S_{k-1} \Psi_k^T .$$

By comparing the above relations for $Q_{k-1}^T S_{k-1}$ and $Q_k^T S_k$, and applying Lemma 1 with $C = \Psi_k^T$ and $D = Q_k^T S_{k-1}$, it follows that if $I + CD = Q_{k-1}^T S_{k-1}$ is nonsingular, then $I + DC = Q_k^T S_k$ is also nonsingular, and $Q_{k-1}^T S_{k-1}$ and $Q_k^T S_k$ have the same eigenvalues. Since $Q_k^T S_k$ is nonsingular, the block vectors S_k necessarily have full column rank

Second question: Convergence of DR-BCG

Lemma

The block vectors S_k defined by DR-BCG satisfy, for $k \geq 0$, the inequality

$$\left\| S_k^T A S_k \right\| \leq m \, \|A\| \, \kappa(A), \tag{2}$$

where $\kappa(A)$ is the condition number of A.

This holds because

$$\operatorname{trace}\left(S_k^T A S_k\right) = \operatorname{trace}\left(S_k^T A Q_k\right) = \sum_{j=1}^{m} (s_k^{(j)})^T A q_k^{(j)} \leq \|A\| \sum_{j=1}^{m} \|s_k^{(j)}\|$$

and, by the Cauchy-Schwarz inequality

$$\left(\sum_{j=1}^{m} \|s_k^{(j)}\| \right)^2 \leq m \sum_{j=1}^{m} \|s_k^{(j)}\|^2 \leq m \, \kappa(A) \sum_{j=1}^{m} \|s_k^{(j)}\|$$

To measure the convergence of DR-BCG, we use the trace of the matrix

$$\mathfrak{E}_k = (X - X_k)^T A (X - X_k).$$

The diagonal entries are the squares of the A-norms of the errors

Theorem

For every $k \geq 1$ we have

$$\mathfrak{E}_{k-1} - \mathfrak{E}_k = \Sigma_{k-1}^T (S_{k-1}^T A S_{k-1})^{-1} \Sigma_{k-1} \tag{3}$$

Theorem

The diagonal entries of $\mathfrak{E}_k$ decrease monotonically as long as the corresponding solutions have not been reached, and DR-BCG is always convergent.

Using relation (3), we have

$$[\mathfrak{E}_k]_{j,j} - [\mathfrak{E}_{k+1}]_{j,j} = e_j^T \Sigma_k^T (S_k^T A S_k)^{-1} \Sigma_k e_j. \tag{4}$$

If the j-th system has not yet been solved, then $\Sigma_k e_j \neq 0$ since $R_k = Q_k \Sigma_k$, which implies

$$[\mathfrak{E}_k]_{j,j} - [\mathfrak{E}_{k+1}]_{j,j} > 0$$

and the sequence $[\mathfrak{E}_k]_{j,j}$ converges

Therefore,

$$(\Sigma_k e_j)^T (S_k^T A S_k)^{-1} (\Sigma_k e_j) \to 0 \quad \text{as} \quad k \to \infty$$

Using inequality (2),

$$(\Sigma_k e_j)^T (S_k^T A S_k)^{-1} (\Sigma_k e_j) \geq \frac{\|\Sigma_k e_j\|^2}{\|S_k^T A S_k\|} \geq \frac{\|\Sigma_k e_j\|^2}{m \|A\| \kappa(A)}$$

which implies $\|\Sigma_k e_j\| \to 0$

Since this quantity is the norm of the jth column of R_k, DR-BCG converges for the jth system

Note that

$$[\mathfrak{E}_k]_{j,j} \geq (\Sigma_k e_j)^T (S_k^T A S_k)^{-1} (\Sigma_k e_j)$$

So, we have lower bounds for the squares of the A-norms of the errors and for the trace

Could we compute upper bounds?

To study that, let us go back to the single right-hand side case, that is, Lanczos and CG

Lanczos algorithm

A symmetric of order n

input A, v
$\beta_0 = 0$, $v_0 = 0$
$v_1 = v/\|v\|$
for $k = 1, \ldots$ **do**
    $w = Av_k - \beta_{k-1} v_{k-1}$
    $\alpha_k = v_k^T w$
    $w = w - \alpha_k v_k$
    $\beta_k = \|w\|$
    $v_{k+1} = w/\beta_k$
end for

It generates tridiagonal matrices T_k, $k = 1, \ldots, n$ with coefficients $(\beta_{i-1}, \alpha_i, \beta_i)$

CG algorithm

Solve $Ax = b$ with A symmetric positive definite

1: **input** A, b, x_0
2: $r_0 = b - Ax_0$
3: $p_0 = r_0$
4: **for** $k = 1, \ldots$ until convergence **do**
5: $\quad \gamma_{k-1} = \frac{r_{k-1}^T r_{k-1}}{p_{k-1}^T A p_{k-1}}$
6: $\quad x_k = x_{k-1} + \gamma_{k-1} p_{k-1}$
7: $\quad r_k = r_{k-1} - \gamma_{k-1} A p_{k-1}$
8: $\quad \delta_k = \frac{r_k^T r_k}{r_{k-1}^T r_{k-1}}$
9: $\quad p_k = r_k + \delta_k p_{k-1}$
10: **end for**

} `cgiter(k)`

$$\varepsilon_k = \|x - x_k\|_A^2 = r_k^T A^{-1} r_k = \min_{y \in x_0 + \mathcal{K}_k(A, r_0)} \|x - y\|_A^2$$

Relations between CG and Lanczos

$$T_k = L_k D_k L_k^T$$

with

$$L_k \equiv \begin{pmatrix} 1 & & \\ \sqrt{\delta_1} & \ddots & \\ & \ddots & \ddots \\ & & \sqrt{\delta_{k-1}} \quad 1 \end{pmatrix}, \qquad D_k \equiv \begin{pmatrix} \gamma_0^{-1} & & \\ & \ddots & \\ & & \gamma_{k-1}^{-1} \end{pmatrix}$$

$$\beta_k = \frac{\sqrt{\delta_k}}{\gamma_{k-1}}, \quad \alpha_k = \frac{1}{\gamma_{k-1}} + \frac{\delta_{k-1}}{\gamma_{k-2}}, \quad \delta_0 = 0, \quad \gamma_{-1} = 1$$

$$v_{j+1} = (-1)^j \frac{r_j}{\|r_j\|}, \qquad j = 0, \ldots, k$$

Let

$$A = U\Lambda U^T, \qquad UU^T = U^T U = I$$

with $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$, w be a given unit norm vector,

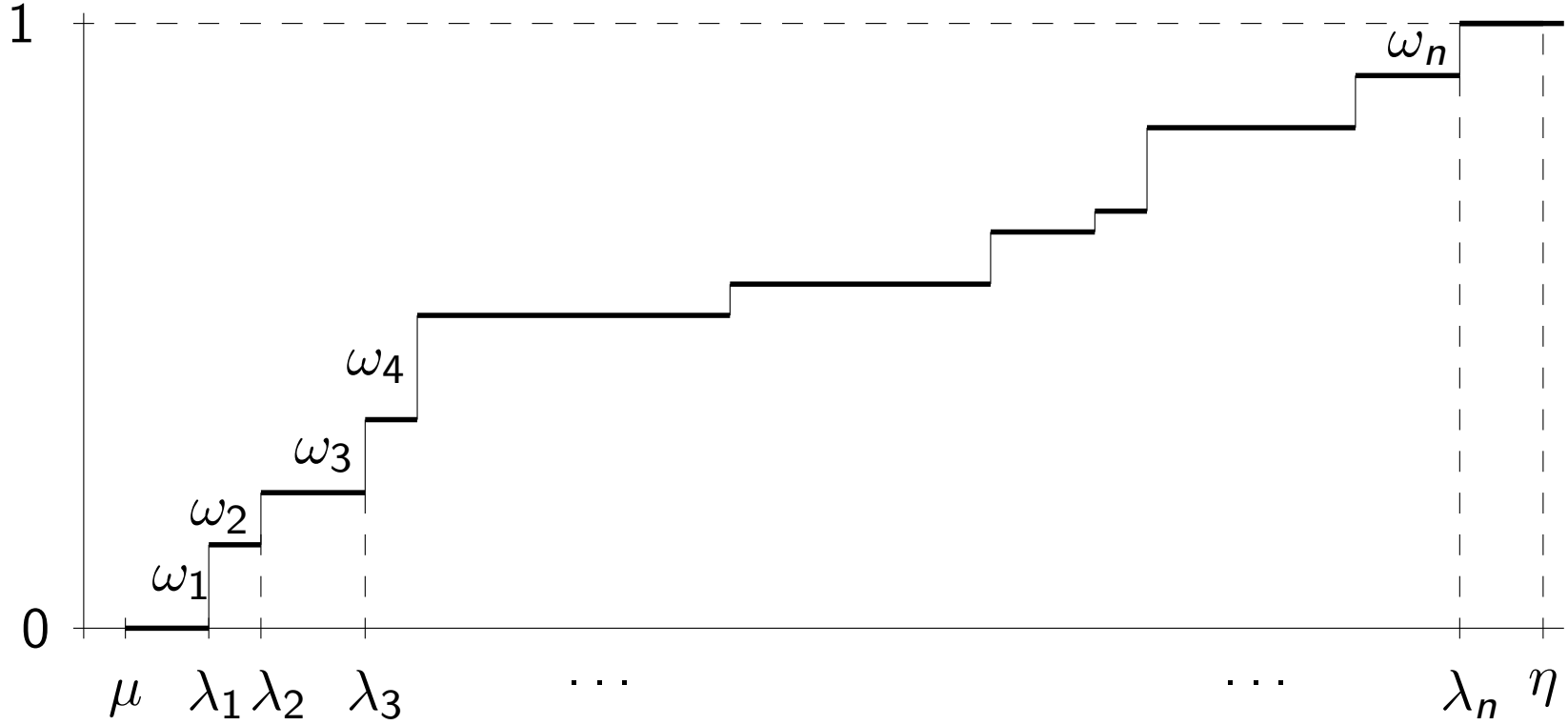
$$\omega_i \equiv (w, u_i)^2 \qquad \text{so that} \qquad \sum_{i=1}^{n} \omega_i = 1$$

and the stepwise constant distribution function

$$\omega(\lambda) \equiv \begin{cases} 0 & \text{for} \quad \lambda < \lambda_1\,, \\ \sum_{j=1}^{i} \omega_j & \text{for} \quad \lambda_i \le \lambda < \lambda_{i+1}\,, \quad 1 \le i \le n-1\,, \\ 1 & \text{for} \quad \lambda_n \le \lambda \end{cases}$$

$$\int_{\mu}^{\eta} f(\lambda)\, d\omega(\lambda) = \sum_{i=1}^{n} \omega_i f(\lambda_i) \;=\; w^T f(A) w$$

Choosing $w = r_k/\|r_k\|$ and $f(\lambda) = 1/\lambda$, it is clear that $r_k^T A^{-1} r_k$ can be written as a Riemann-Stieltjes integral


1
0
ω_1
ω_2
ω_3
ω_4
ω_n
μ
λ_1
λ_2
λ_3
$\cdots$
$\cdots$
λ_n
η

The Lanczos basis vectors can be written as polynomials in A applied to the initial vector v_1, as $v_k = p_k(A)v_1$

These polynomials satisfy a three-term recurrence and are orthogonal with respect to the dot product defined by the distribution function $\omega(\lambda)$

$$\int_{\mu}^{\eta} p_i(\lambda)\, p_j(\lambda)\, d\omega(\lambda) = 0, \quad i \neq j$$

For CG we used the relation

$$\varepsilon_k = \|x - x_k\|_A^2 = r_k^T A^{-1} r_k = \|r_0\|^2 \left[(T_n^{-1} e_1, e_1) - (T_k^{-1} e_1, e_1) \right]$$

It was shown by Z. Strakoš and P. Tichý that this relation holds in finite precision arithmetic up to a small perturbation term

That relation can also be written as

$$\varepsilon_0 = \sum_{j=0}^{k-1} \gamma_j \|r_j\|^2 + \varepsilon_k$$

or

$$\left(T_n^{-1}\right)_{1,1} = \left(T_k^{-1}\right)_{1,1} + \mathcal{R}_k^{(G)}[\lambda^{-1}]$$

$$\left(T_n^{-1}\right)_{1,1} = \frac{\varepsilon_0}{\|r_0\|^2} = \int_\mu^\eta \lambda^{-1}\, d\omega(\lambda)$$

$\left(T_k^{-1}\right)_{1,1}$ is the Gauss quadrature approximation of the integral and the remainder is

$$\mathcal{R}_k^{(G)}\left[\lambda^{-1}\right] = \frac{\varepsilon_k}{\|r_0\|^2}$$

The nodes of the Gauss quadrature rule are the eigenvalues of T_k

Since we don't know ε_0, we use

$$\varepsilon_{k-d} - \varepsilon_k = \sum_{j=k-d}^{k-1} \gamma_j \|r_j\|^2$$

where the delay d is a positive integer smaller than k

The right-hand side is a lower bound of the A-norm squared at iteration $\ell = k - d$

The simplest lower bound at iteration $k-1$ is $\gamma_{k-1}\|r_{k-1}\|^2$

To obtain an upper bound of the A-norm we use a Gauss-Radau quadrature rule with a fixed node $\mu \leq \lambda_1$

$$\varepsilon_0 = \|r_0\|^2 (\widehat{T}_{k+1}^{-1})_{1,1} + \widehat{\mathcal{R}}_{k+1}[\lambda^{-1}]$$

Subtracting

$$\varepsilon_0 = \|r_0\|^2 (T_{k-d}^{-1})_{1,1} + \varepsilon_{k-d}$$

we obtain

$$\varepsilon_{k-d} = \|r_0\|^2 [\, (\widehat{T}_{k+1}^{-1})_{1,1} - (T_{k-d}^{-1})_{1,1} \,] + \widehat{\mathcal{R}}_{k+1}[\lambda^{-1}]$$

The difference in the right-hand side can be written as

$$(\widehat{T}_{k+1}^{-1})_{1,1} - (T_{k-d}^{-1})_{1,1} = (\widehat{T}_{k+1}^{-1})_{1,1} - (T_k^{-1})_{1,1} + \Delta_{\ell:k-1}$$

with the Gauss lower bound

$$\Delta_{\ell:k-1} = \sum_{j=\ell}^{k-1} \gamma_j \|r_j\|^2$$

with $\ell = k - d$

The matrix $\widehat{T}_{k+1}$ is equal to T_{k+1}, except that α_{k+1} is replaced with $\alpha_{k+1}^{(\mu)}$ computed such that $\widehat{T}_{k+1}$ has μ as an eigenvalue (that is, it is a prescribed node of the quadrature rule)

Then, we can use the Sherman-Morrison formula for the difference $(\widehat{T}_{k+1}^{-1})_{1,1} - (T_k^{-1})_{1,1}$
This gives the CGQL algorithm of G.H. Golub and G. M., 1997

CG coeffs $\rightarrow$ Lanczos coeffs $\rightarrow$ Gauss-Radau upper bound

Can we compute the upper bound directly from the CG coefficients?

We look for a coefficient $\gamma_k^{(\mu)}$ such that

$$T_{k+1}^{(\mu)} = L_{k+1} \begin{pmatrix} D_k & \\ & \left(\gamma_k^{(\mu)}\right)^{-1} \end{pmatrix} L_{k+1}^T$$

such that μ is an eigenvalue

This problem was solved in

G. M. and P. Tichý, On computing quadrature-based bounds for the A-norm of the error in conjugate gradients,
Numer. Algorithms, 62(2), pp. 163-191, 2013

$$\gamma_{j+1}^{(\mu)} = \frac{\gamma_j^{(\mu)} - \gamma_j}{\mu(\gamma_j^{(\mu)} - \gamma_j) + \delta_{j+1}}, \qquad \gamma_0^{(\mu)} = \frac{1}{\mu}$$

This leads to the CGQ algorithm

CGQ algorithm

input A, b, x_0, μ, d

$r_0 = b - Ax_0$, $\quad p_0 = r_0$

$\gamma_0^{(\mu)} = 1/\mu$, $\quad \Delta_0^{(\mu)} = \gamma_0^{(\mu)} \|r_0\|^2$

$\varphi_0^{(\mu)} = \sqrt{\Delta_0^{(\mu)}}$

for $k = 1, \ldots$ until convergence **do**

  `cgiter`(k)

  $\Delta_{k-1} = \gamma_{k-1} \|r_{k-1}\|^2$

  $\gamma_k^{(\mu)} = \dfrac{\gamma_{k-1}^{(\mu)} - \gamma_{k-1}}{\mu(\gamma_{k-1}^{(\mu)} - \gamma_{k-1}) + \delta_k}$, $\quad \Delta_k^{(\mu)} = \gamma_k^{(\mu)} \|r_k\|^2$

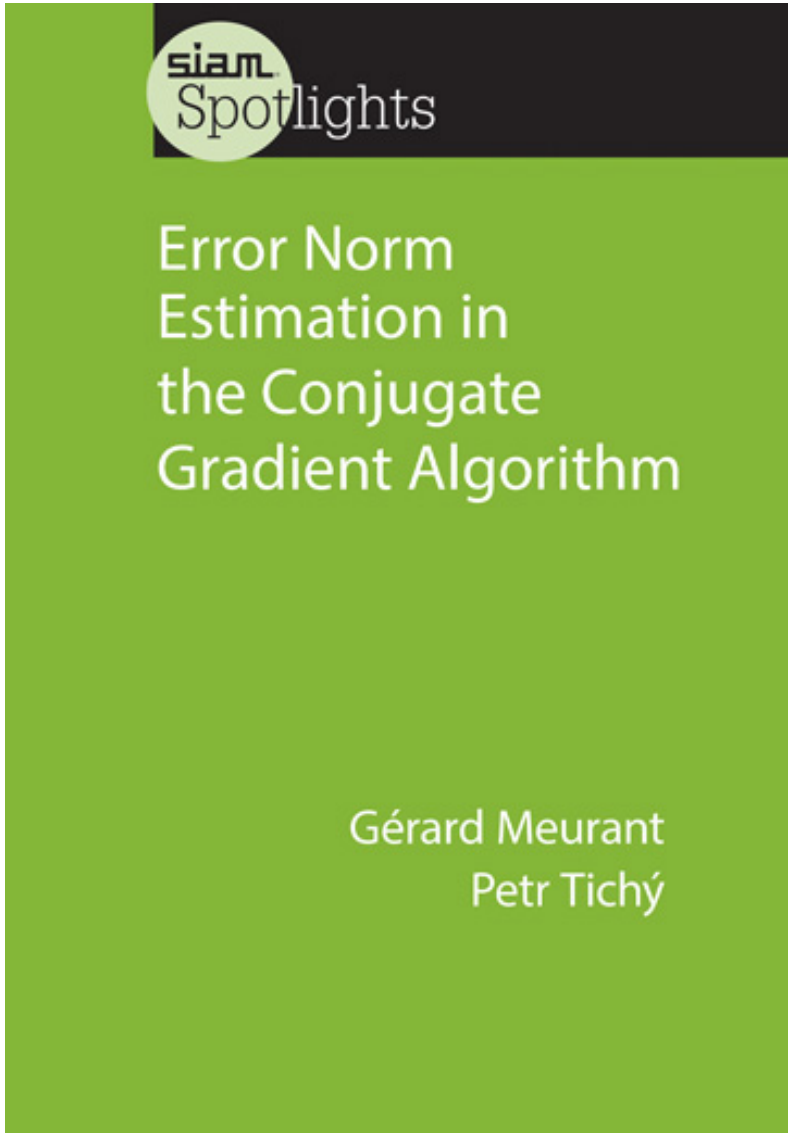
  $\ell = k - d$

  $\varphi_{k-1} = \sqrt{\Delta_{k-1}}$, $\quad \varphi_{\ell:k-1} = \sqrt{\Delta_{\ell:k-1}}$

  $\varphi_k^{(\mu)} = \sqrt{\Delta_k^{(\mu)}}$, $\quad \varphi_{\ell:k}^{(\mu)} = \sqrt{\Delta_{\ell:k-1} + \Delta_k^{(\mu)}}$,

end for

For details, see



The block case

Can we do the same things for the block case?

We have already seen how to compute lower bounds for DR-BCG. For OL-BCG, we have

Theorem

Assume that the block vectors R_k in OL-BCG are full rank. Then

$$\mathfrak{E}_{k-1} - \mathfrak{E}_k = \Theta_{k-1} \tag{5}$$

with $\Theta_{k-1} = (R_{k-1}^T R_{k-1})\, \Upsilon_{k-1}$

This gives lower bounds for the diagonal entries

Relation (5) can be considered as a block quadrature rule using the block tridiagonal matrix of the block Lanczos algorithm

To obtain upper bounds, we have to use a block Gauss-Radau quadrature rule with a prescribed node, that is, we have to modify the block tridiagonal matrix to have a given multiple eigenvalue

This is a little more tricky than for the scalar case. Finally, the diagonal entries of $\Theta_{k-1}^{(\mu)}$ give upper bounds of the squares of the A-norms of the errors where

$$\Theta_{k-1}^{(\mu)} = (R_{k-1}^T R_{k-1})\, \Upsilon_{k-1}^{(\mu)}$$

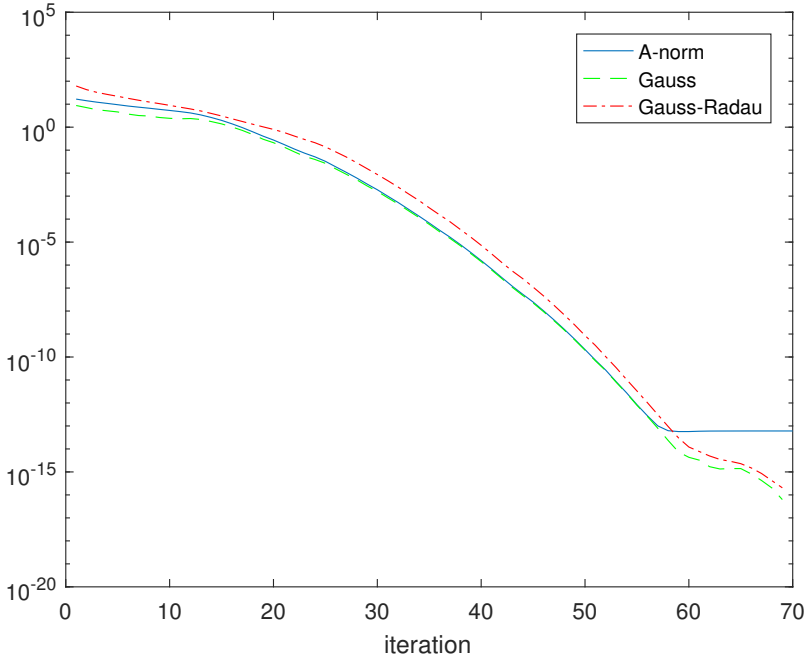
with

$$\Upsilon_{k0}^{(\mu)} = \mu^{-1} I, \quad \Upsilon_k^{(\mu)} = [\mu(\Upsilon_{k-1}^{(\mu)} - \Upsilon_{k-1}) + \Xi_k]^{-1}(\Upsilon_{k-1}^{(\mu)} - \Upsilon_{k-1})$$

For details and formulas for DR-BCG, see
G.M. and P.Tichý, Error norm estimates for the block conjugate gradient algorithm, SIAM J. Matrix Anal. Appl., v. 46 n. 4 (2025), pp. 2175-2196

Numerical experiments (1)

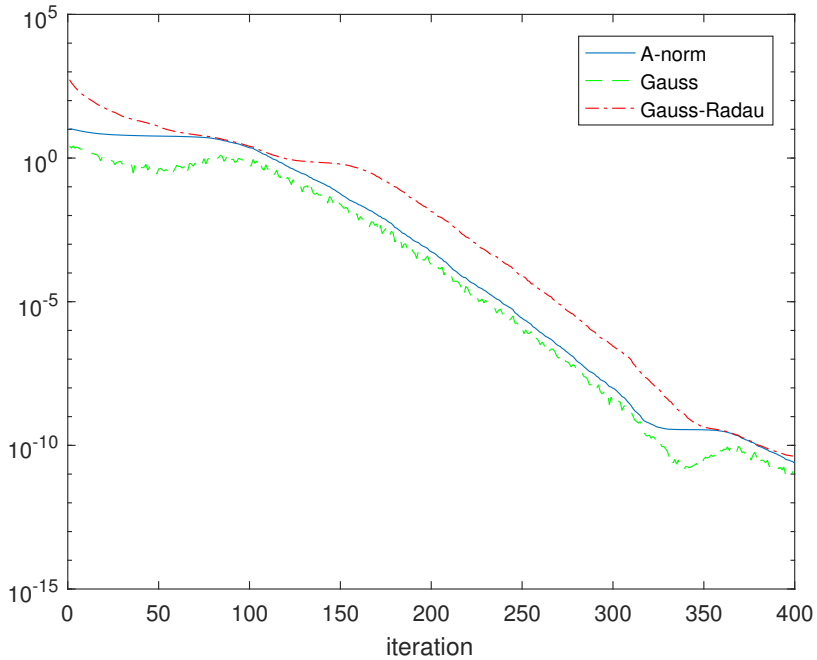
Poisson problem, $n = 900$, $\kappa(A) = 3.8881\ 10^2$, 10 random right-hand sides, OL-BCG, $X_0 = 0$, $d = 1$, $\lambda_1 = 0.0205227$, $\mu = 0.0205$



`Poisson`, first system, A-norm of the error and bounds

Numerical experiments (2)

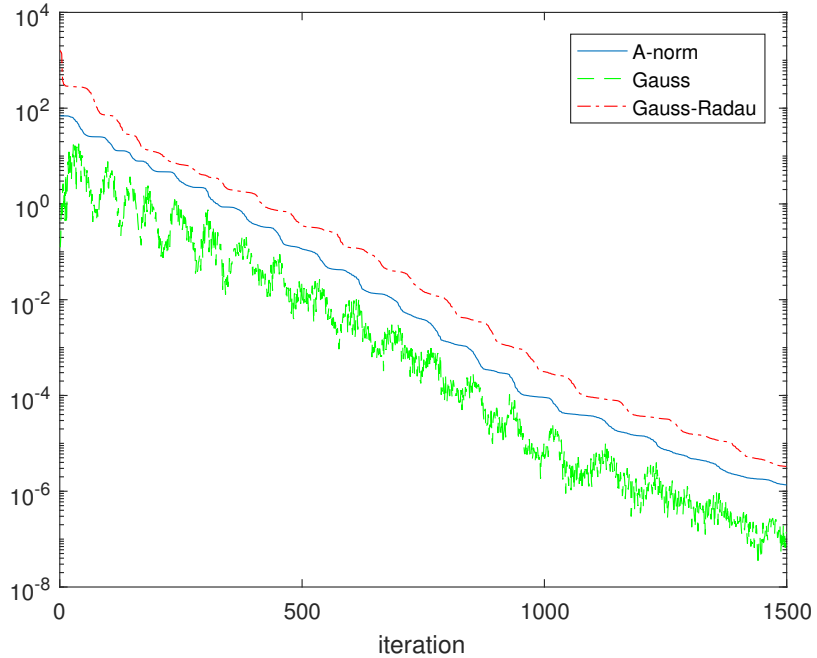
662_bus, $n = 662$, $\kappa(A) = 7.9413\ 10^5$, 5 random right-hand sides, OL-BCG, $X_0 = 0$, $d = 1$, $\lambda_1 = 5.0472\ 10^{-3}$, $\mu = 5\ 10^{-3}$



662_bus, first system, A-norm of the error and bounds

Numerical experiments (3)

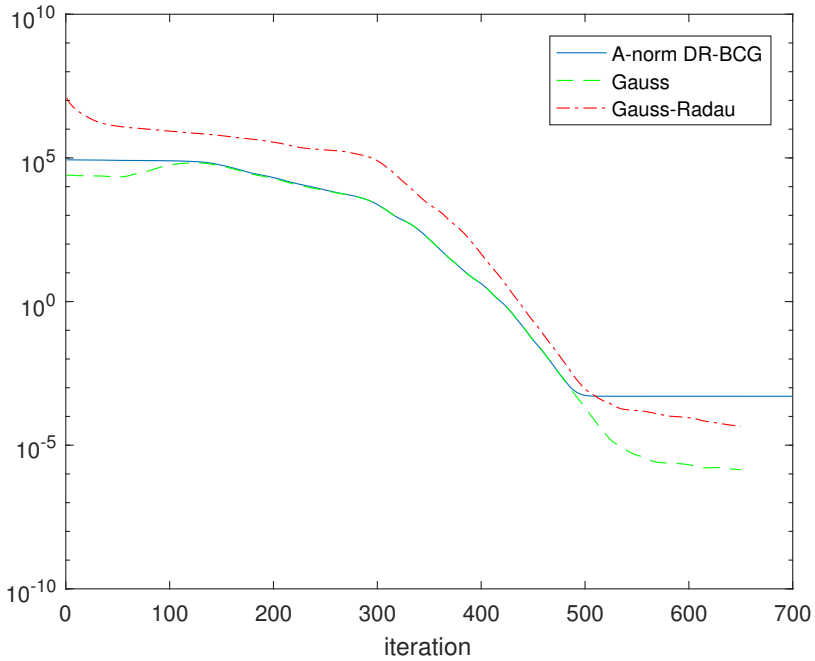
nos7, $n = 729$, $\kappa(A) = 2.3745\ 10^9$, 10 random right-hand sides,
OL-BCG, $X_0 = 0$, $d = 1$, $\lambda_1 = 4.1541\ 10^{-3}$, $\mu = 4.1540\ 10^{-3}$



nos7, first system, A-norm of the error and bounds

Numerical experiments (4)

s3dkt3m2, $n = 90448$, 10 random right-hand sides, DR-BCG, $X_0 = 0$, $d = 50$, $\mu = 1.589\ 10^{-8}$
Incomplete Cholesky preconditioner (with some fill-in)



`s3dkt3m2`, first system, A-norm of the error and bounds

Bibliography

D.P. O'Leary, The block conjugate gradient algorithm and related methods, Linear Algebra Appl., 29, pp. 293-322, 1980

A.A. Dubrulle, Retooling the method of block conjugate gradients, Electron. Trans. Numer. Anal., 12, pp. 216-233, 2001

P. Tichý, G. Meurant, and D. Simonova, Block CG algorithms revisited, Numer. Algorithms, 100(4), pp. 1527-1553, 2025

G. Meurant and P. Tichý, Error norm estimates for the block conjugate gradient algorithm. SIAM J. Matrix Anal. Appl., 46(4), pp. 2175-2196, 2025

G. Meurant, J. Papež, and P. Tichý, Block conjugate gradient methods with error norm estimates for least squares problems, BIT Numerical Mathematics, 66(2), 2026