Block conjugate gradient methods for least squares problems

Gérard MEURANT

https://gerard-meurant.fr

Saint Girons, July 2026

Joint work with
Petr Tichý (Charles University, Prague)
and
Jan Papež (Institute of Mathematics, Prague)

Let A be an $n \times m$ real matrix with $n \geq m$ and full column rank
We would like to solve

$$Ax^{(i)} = b^{(i)}, \quad i = 1, \ldots, s,$$

where $b^{(i)} \in \mathbb{R}^n$ in the least squares sense

The minimization of $\|b^{(i)} - Ax^{(i)}\|$ is equivalent to solving

$$A^T Ax^{(i)} = A^T b^{(i)}, \quad i = 1, \ldots, s.$$

In block form

$$A^T AX = A^T B, \quad X = \left[x^{(1)}, \ldots, x^{(s)}\right], \quad B = \left[b^{(1)}, \ldots, b^{(s)}\right].$$

A being full rank m, the matrix $A^T A$ is symmetric positive definite and there is a unique solution X

HS-BCGLS

Like in CGLS [Hestenes and Stiefel, 1952], block CG algorithms can be modified to solve the block linear system $A^T AX = A^T B$ without explicitly computing the matrix $A^T A$

1: **input** A, B, X_0
2: $R_0 = B - AX_0$
3: $\widetilde{R}_0 = A^T R_0$
4: $P_0 = \widetilde{R}_0$
5: **for** $k = 1, \ldots$ until convergence **do**
6: $\quad \Upsilon_{k-1} = [(AP_{k-1})^T AP_{k-1}]^{-1}(\widetilde{R}_{k-1}^T \widetilde{R}_{k-1})$
7: $\quad X_k = X_{k-1} + P_{k-1}\Upsilon_{k-1}$
8: $\quad R_k = R_{k-1} - AP_{k-1}\Upsilon_{k-1}$
9: $\quad \widetilde{R}_k = A^T R_k$
10: $\quad \Xi_k = (\widetilde{R}_{k-1}^T \widetilde{R}_{k-1})^{-1}(\widetilde{R}_k^T \widetilde{R}_k)$
11: $\quad P_k = \widetilde{R}_k + P_{k-1}\Xi_k$
12: **end for**

As in HS-BCG, R_k and P_k may become (almost) rank deficient

In that case, convergence is slowing down or the algorithm may diverge

Instead of using deflation to get rid of the dependent columns, let us "Dubrullize" HS-BCGLS $\rightarrow$ DR-BCGLS

DR-BCGLS

input A, B, X_0
$R_0 = B - AX_0$
$[Q_0, \Sigma_0] = \mathtt{qr}(A^T R_0)$
$S_0 = Q_0$
for $k = 1, \ldots$ until convergence **do**
    $Y_{k-1} = AS_{k-1}$
    $\Pi_{k-1} = (Y_{k-1}^T Y_{k-1})^{-1}$
    $X_k = X_{k-1} + S_{k-1}\Pi_{k-1}\Sigma_{k-1}$
    $[Q_k, \Psi_k] = \mathtt{qr}(Q_{k-1} - A^T Y_{k-1}\Pi_{k-1})$
    $S_k = Q_k + S_{k-1}\Psi_k^T$
    $\Sigma_k = \Psi_k \Sigma_{k-1}$
end for

To avoid using the matrix $A^T A$, we introduce new block vectors $Y_k = AS_k$, which are $n \times s$ matrices

As HS-BCGLS, the new algorithm DR-BCGLS requires two matrix-block vector multiplications per iteration: one with A and one with A^T

Although the block coefficients Ψ_k may become (nearly) singular, the columns within a block Q_k remain linearly independent

As for DR-BCG, DR-BCGLS cannot break down because $S_{k-1}^T A^T A S_{k-1} = Y_{k-1}^T Y_{k-1}$ is nonsingular

There is only one economy-size QR factorization of an $m \times s$ matrix at each iteration

Error norm estimates

Let us denote

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

and

$$\Theta_{k-1} = \mathfrak{E}_{k-1} - \mathfrak{E}_k, \qquad \mathfrak{R}_k = (B - AX_k)^T AA^T (B - AX_k),$$

Then

$$\begin{array}{lrcl} \text{(HS-BCGLS)} & \Theta_{k-1} & = & (\widetilde{R}_{k-1}^T \widetilde{R}_{k-1}) \Upsilon_{k-1}, \quad \mathfrak{R}_k = \widetilde{R}_k^T \widetilde{R}_k \\ \text{(DR-BCGLS)} & \Theta_{k-1} & = & \Sigma_{k-1}^T \Pi_{k-1} \Sigma_{k-1}, \quad \mathfrak{R}_k = \Sigma_k^T \Sigma_k \end{array}$$

These $s \times s$ matrices are computable during the iterations

Error norm estimates (2)

Θ_{k-1} is (similar to) a symmetric positive definite matrix

The diagonal entries of Θ_{k-1} are lower bounds of the diagonal entries of $\mathfrak{E}_{k-1}$ (squares of the error norms)

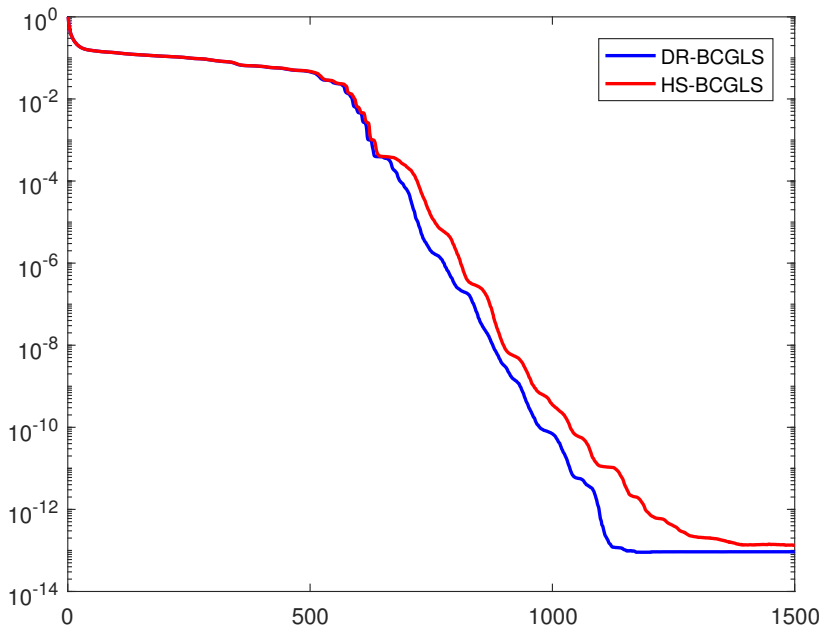
Like in CG and HS-BCG [M. and Tichý, 2025], better bounds can be obtained by using a delay d larger than one

Using block Gauss-Radau quadrature rules, upper bounds can be obtained if an estimate μ of the smallest eigenvalue of $A^T A$ is known

Heuristic techniques for computing "optimal" values of the delays are available

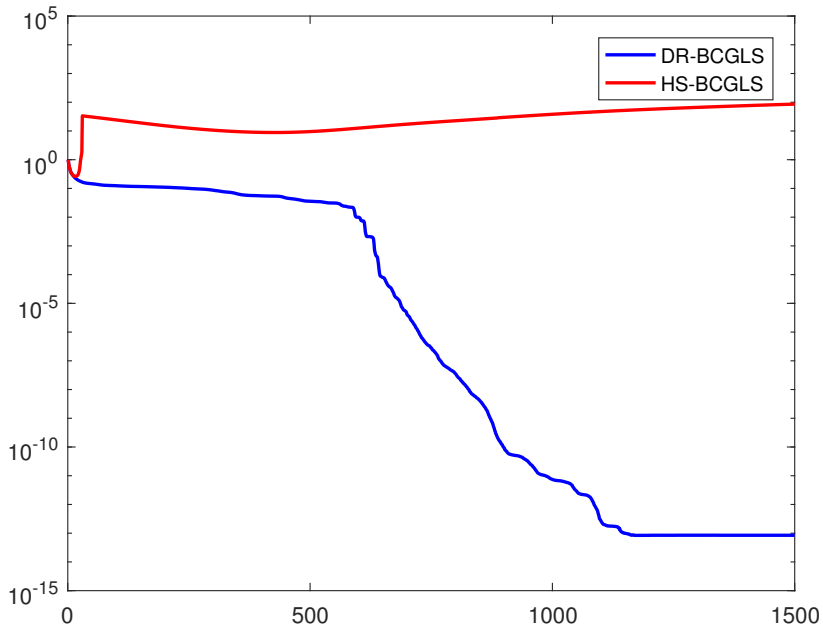
Numerical experiments

Matrix `illc1850`, 1850×712, $8,636$ nonzero entries,
$\kappa(A) \approx 1.4 \times 10^3$
B random with $s = 2$, $X_0 = 0$
Relative trace error
$\tau_k = \sqrt{\mathrm{tr}((X - X_k)^T A^T A (X - X_k))} / \sqrt{\mathrm{tr}(X^T A^T A X)}$



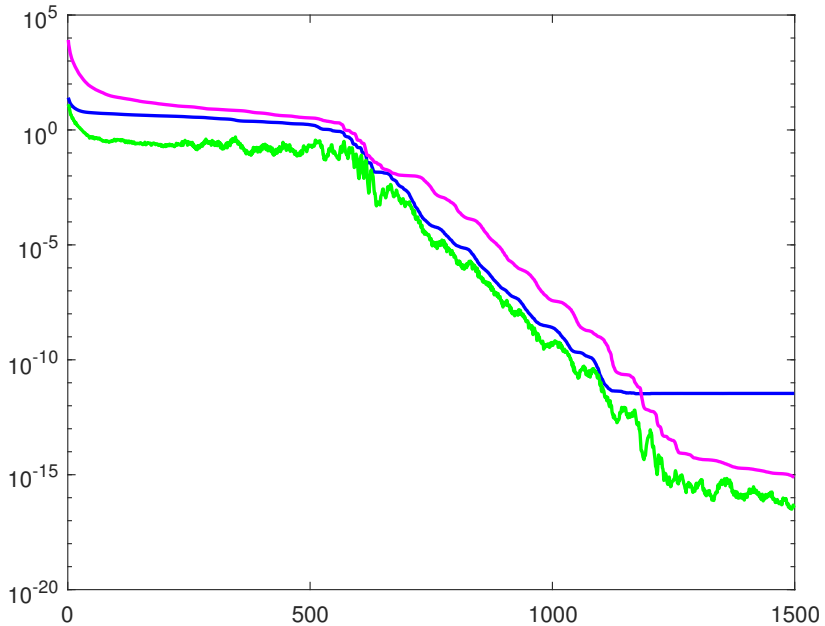
Numerical experiments (2)

Matrix `illc1850`, rank(B)=1 (initial deficiency), $s = 2$
Relative trace error τ_k



Numerical experiments (3)

Matrix `illc1850`, rank(B)=2, $s = 2$
DR-BCGLS
Trace error and bounds, $d = 1$, $\mu = 2.2 \times 10^{-6}$



Details are given in

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

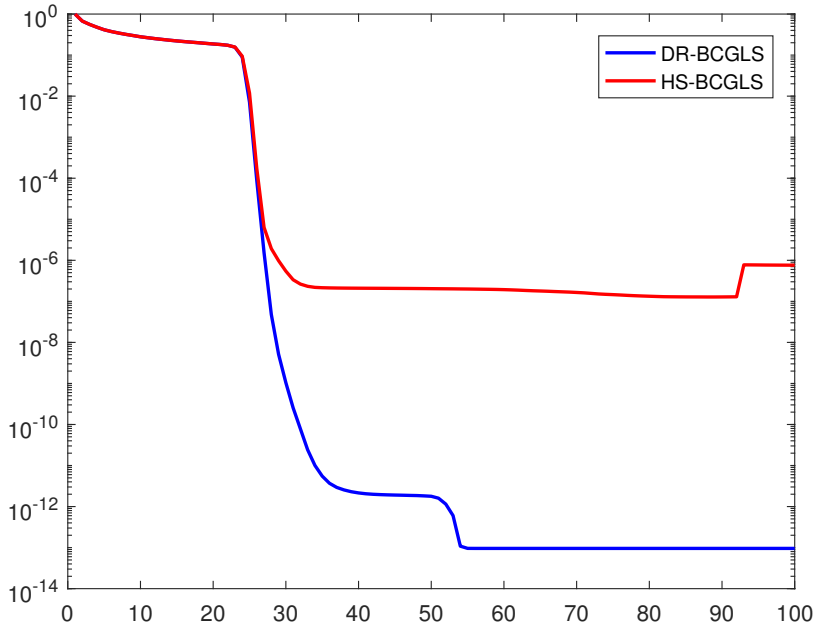
Special issue in celebration of the 90th birthday of Åke Björck

MATLAB codes are available at

`https://github.com/JanPapez/Block-CGlike-methods-with-error-estimate`

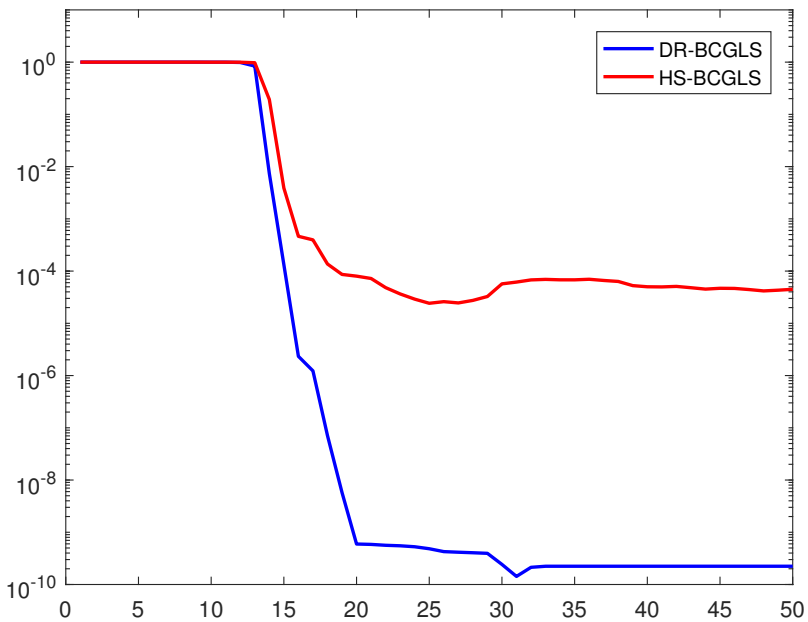
More numerical experiments

Matrix `illc1850`, B random, $s = 32$
Relative trace error τ_k



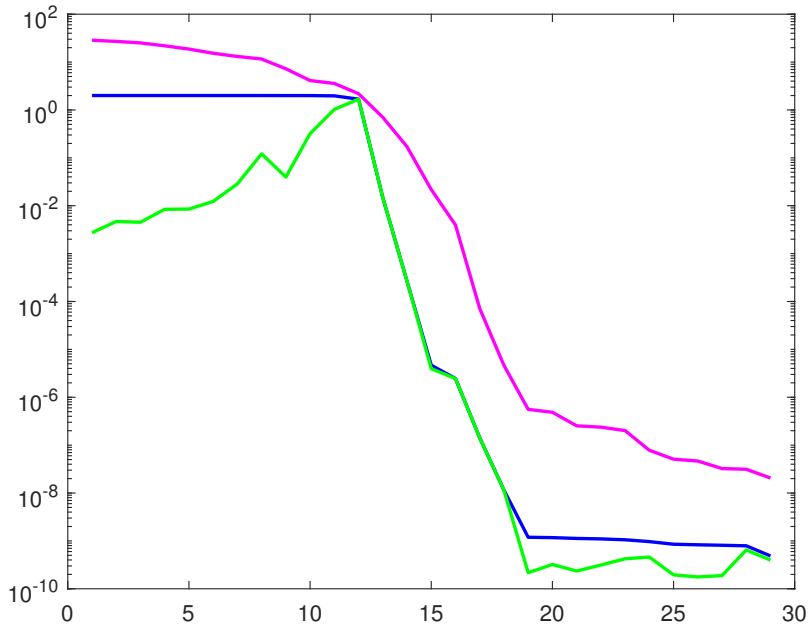
More numerical experiments (2)

Matrix `P(80,40,1,3)` 80×40, $\kappa(A) = 6.4 \times 10^4$ [Paige-Saunders, 1982], $s = 4$
Relative trace error τ_k



More numerical experiments (3)

Matrix `P(80,40,1,3)`, $s = 4$
Trace error and bounds $d = 1$, $\mu = 2.4 \times 10^{-10}$



More numerical experiments (4)

Matrix `P(80,40,1,3)`, $s = 4$
Trace error and bounds $d = 5$, $\mu = 2.4 \times 10^{-10}$