Norms of the error for the Conjugate Gradient

Gérard MEURANT
CEA/DIF
Bruyères le Chatel, France
(gerard.meurant@cea.fr)
Saint Girons, June 2003

$$Ax = b$$

A symmetric positive definite

The method of choice is the (preconditioned) Conjugate Gradient method

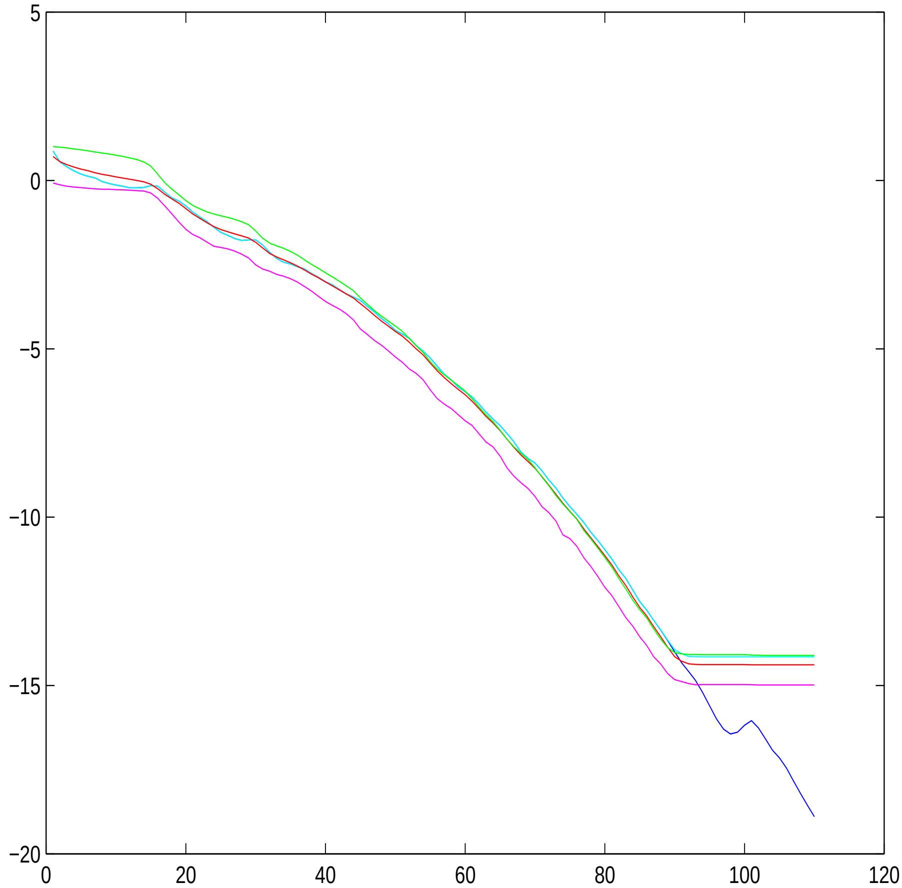
Measures of convergence:

- l_2 norm of residual $b - Ax^k$
- A-norm of the error $\epsilon^k = x - x^k$
- l_2 norm of the error
- Max norm of the error

Examples (CG in floating point arithmetic):

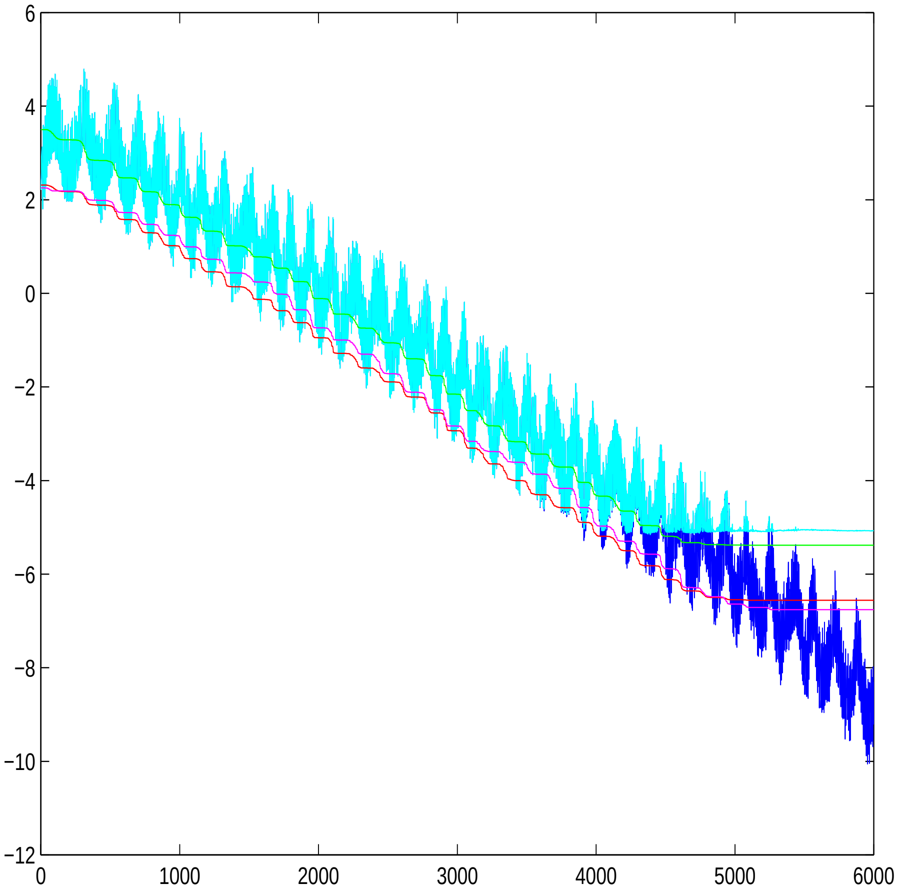
- Poisson equation with a 20×20 mesh, $n = 400$
- Nos7 from the Matrix Market, $n = 729, \lambda_{min} = 4.15\ 10^{-3}, \lambda_{max} = 9.86\ 10^6$
- Bcsstk01 from the Matrix Market, $n = 48, \lambda_{min} = 3.42\ 10^3, \lambda_{max} = 3.02\ 10^9$

The last two ones are "pathological"



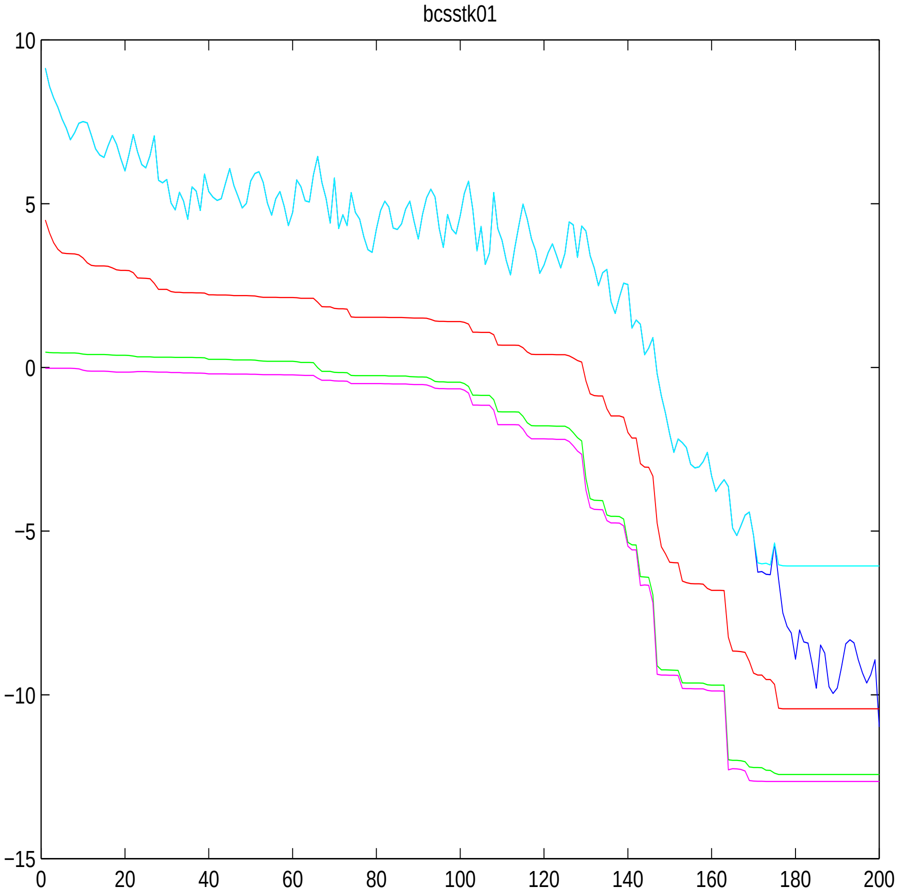
Norms for the Poisson problem, $m = 20$,

blue: residual, cyan: computed true residual, red: A norm of the error, green: l_2 norm of the error, magenta: l_∞ norm of the error



Norms for the Nos7 problem,

blue: residual, cyan: computed true residual, red: A norm of the error, green: l_2 norm of the error, magenta: l_∞ norm of the error



Norms for the Bcsstk01 problem,

blue: residual, cyan: computed true residual, red: A norm of the error, green: l_2 norm of the error, magenta: l_∞ norm of the error

Questions:

- How to estimate the norms of the error?
- Relations with the norm of the residual?
- When is the norm of the residual oscillating and being larger or smaller than the norms of the error?
- What is the maximum attainable accuracy?
- Why is there a difference between the computed and the "true" residuals?

Today, answers to the first question

CG

For Krylov methods:

$$r^k = b - Ax^k$$

Krylov subspace $\mathcal{K}_k(r^0, A)$ of order k

$$\mathcal{K}_k(r^0, A) = span\{r^0, Ar^0, \ldots, A^{k-1}r^0\}$$

$$x^k \in x^0 + \mathcal{K}_k(r^0, A), \quad x^k = x^0 + V_k z^k$$

V_k orthogonal basis of $\mathcal{K}_k(r^0, A)$

CG is an orthogonal residual (OR) method:

$$(r^k)^T V_k = 0$$

The basis vectors $V_k = [v^1, \ldots, v^k]$ are constructed by the Lanczos algorithm:

$$AV_k = V_k T_k + \eta_{k+1} v^{k+1} (e^k)^T,$$

$$V_k^T A V_k = T_k, \quad A V_n = V_n T_n$$

$$T_k = \begin{pmatrix} \alpha_1 & \eta_2 & & & \\ \eta_2 & \alpha_2 & \eta_3 & & \\ & \ddots & \ddots & \ddots & \\ & & \eta_{k-1} & \alpha_{k-1} & \eta_k \\ & & & \eta_k & \alpha_k \end{pmatrix}$$

The orthogonality relation gives

$$T_k z^k = \|r^0\| e^1$$

CG is obtained from Lanczos by considering the Cholesky factorization of $T_k = L_k \Delta_k^{-1} L_k^T$

$$\delta_1 = \alpha_1, \quad \delta_l = \alpha_l - \frac{\eta_l^2}{\delta_{l-1}}, \, l = 2, \ldots, k$$

CG

for $k = 0, 1, \ldots$ until convergence do,

$$\beta_k = \frac{(r^k, r^k)}{(r^{k-1}, r^{k-1})}, \; \beta_0 = 0,$$

$$p^k = r^k + \beta_k p^{k-1},$$

$$\gamma_k = \frac{(r^k, r^k)}{(Ap^k, p^k)},$$

$$x^{k+1} = x^k + \alpha_k p^k,$$

$$r^{k+1} = r^k - \alpha_k A p^k.$$

If $r^0 = \|r^0\| v^1$ CG=Lanczos

$$\alpha_k = \frac{1}{\gamma_{k-1}} + \frac{\beta_{k-1}}{\gamma_{k-2}}, \quad \beta_0 = 0, \quad \gamma_{-1} = 1,$$

$$\eta_{k+1} = \frac{\sqrt{\beta_k}}{\gamma_{k-1}}.$$

$$v^{k+1} = (-1)^k \frac{r^k}{\|r^k\|}.$$

$$\delta_{k+1} = \frac{1}{\gamma_k},$$

The norms of the error

$$\epsilon^k = x - x^k, \quad A\epsilon^k = r^k$$

Since $r^k = r^0 - AV_k z^k$ and $\epsilon^k = A^{-1} r^k$

$$\|\epsilon^k\|^2 = (A^{-1}r^0, A^{-1}r^0) - 2(A^{-1}r^0, V_k z^k) + (V_k z^k, V_k z^k)$$

$$\|\epsilon^k\|_A^2 = (A\epsilon^k, \epsilon^k) = (A^{-1}r^0, r^0) - 2(r^0, V_k z^k) + (AV_k z^k, V_k z^k)$$

We also notice that

$$\|\epsilon^k\|_{A^TA}^2 = (A^T A\epsilon^k, \epsilon^k) = (A\epsilon^k, A\epsilon^k) = \|r^k\|^2,$$

Also

$$\epsilon^k = A^{-1}r^k = A^{-1}r^0 - V_k z^k.$$

But $r^0 = \|r^0\| V_n e^1$ and $A^{-1}V_n = V_n T_n^{-1}$

Therefore

$$\epsilon^k = \|r^0\|(V_n T_n^{-1} e^1 - V_k T_k^{-1} e^1)$$

$$\epsilon_j^k = \|r^0\|(V_n T_n^{-1} e^1 - V_k T_k^{-1} e^1, e^j)$$

The A–norm of the error

Theorem

$$\|\epsilon^k\|_A^2 = \|r^0\|^2[(T_n^{-1}e^1, e^1) - (T_k^{-1}e^1, e^1)]$$

G. Dahlquist, S.C. Eisenstat and G.H. Golub. Bounds for the error of linear systems of equations using the theory of moments, J. Math. Anal. Appl. 37 (1972) pp. 151–166.

G. Dahlquist, G.H. Golub and S.G. Nash. Bounds for the error in linear systems. In Proc. of the Workshop on Semi–Infinite Programming, R. Hettich ed, Springer (1978), pp. 154–172.

Proof:

$$r^0 = \|r^0\|v^1 = \|r^0\|V_n e^1$$

$$A^{-1}r^0 = \|r^0\|A^{-1}V_n e^1 = \|r^0\|V_n T_n^{-1}e^1$$

$$(A^{-1}r^0, r^0) = \|r^0\|^2(V_n T_n^{-1}e^1, V_n e^1) = \|r^0\|^2(T_n^{-1}e^1, e^1).$$

$$(r^0, V_k z^k) = (V_k^T r^0, z^k) = \|r^0\|^2(e^1, T_k^{-1}e^1)$$

$$(AV_k z^k, V_k z_k) = (V_k^T A V_k z^k, z^k) = (T_k z^k, z^k) = \|r^0\|^2(T_k^{-1}e^1, e^1)$$

This formula is closely related to Gauss quadrature since the scalar product $(T_n^{-1}e^1, e^1)$ (or $(A^{-1}r^0, r^0)$) can be written as a Riemann–Stieltjes integral

In fact it is exactly Gauss quadrature since $(T_k^{-1}e^1, e^1)$ is nothing else than a Gauss quadrature approximation to this integral

G.H. Golub and G. Meurant, Matrices, moments and quadrature, in Numerical Analysis 1993, D.F. Griffiths & G.A. Watson, Eds. Pitman Research Notes in Mathematics, v 303, (1994), pp 105–156

G.H. Golub and G. Meurant, Matrices, moments and quadrature II or how to compute the norm of the error in iterative methods, BIT, v 37 no 3, (1997), pp 687–705

Other relation:

$$\|\epsilon^k\|_A^2 = (A^{-1}r^k, r^k) = (A^{-1}r^0, r^k) - (V_k z^k, r^k) = (A^{-1}r^0, r^k)$$

$V_n T_n^{-1} = A^{-1} V_n$ and $A^{-1}r^0 = \|r^0\| A^{-1} V_n e^1$

$$r^k = b - Ax^k = r^0 - AV_k z^k$$

$$r^k = r^0 - V_k T_k z^k - \eta_{k+1} v^{k+1} (e^k)^T z^k$$

$$r^k = -\eta_{k+1} \|r^0\| (e^k, T_k^{-1} e^1) v^{k+1}.$$

$$\|\epsilon^k\|_A^2 = -\eta_{k+1} \|r^0\|^2 (e^k, T_k^{-1} e^1)(V_n T_n^{-1} e^1, v^{k+1}).$$

But $V_n^T v^{k+1} = f^{k+1}$ where the components of the vector f^{k+1} are zero except for the $k+1$ st one which is 1

Other relations can be found by computing the first element of $x = T_k^{-1} e^1$

Let $T_k = UD^{-1}U^T$ be a UL factorization of T_k

$$d_k = \alpha_k, \quad d_l = \alpha_l - \frac{\eta_{l+1}^2}{d_{l+1}}, \, l = k-1, \dots, 1$$

$$x_1 = \frac{1}{d_1}, \quad x_l = -\frac{\eta_l}{d_l} x_{l-1} = (-1)^{l-1} \frac{\eta_l \cdots \eta_2}{d_l \cdots d_1}.$$

This shows that $x_1 > 0$ and the signs of the elements x_l alternate

Let $\bar{d}_l$ be the diagonal elements of the UL decomposition of T_{k+1}, then we have

$$d_k - \bar{d}_k = \frac{\eta_{k+1}^2}{\alpha_{k+1}}$$

and

$$d_{k-1} - \bar{d}_{k-1} = \eta_k^2 \left(\frac{d_k - \bar{d}_k}{d_k \bar{d}_k} \right)$$

Recursively

$$d_{j-1} - \bar{d}_{j-1} = \eta_j^2 \left(\frac{d_j - \bar{d}_j}{d_j \bar{d}_j} \right), \, j = k-1, \ldots, 2$$

Therefore $\bar{d}_l < d_l, l = 1, \ldots, k$

This shows that

$$|(T_k^{-1} e^1)_l| < |(T_{k+1}^{-1} e^1)_l|, l = 1, \ldots, k.$$

By induction

$$d_1 - \bar{d}_1 = \frac{\eta_2^2 \cdots \eta_k^2}{(d_2 \cdots d_k)(\bar{d}_2 \cdots \bar{d}_k)} \frac{\eta_{k+1}^2}{\alpha_{k+1}} = \frac{(\eta_2^2 \cdots \eta_{k+1}^2) d_1 \bar{d}_1}{\det(T_k) \det(T_{k+1})}$$

Therefore

$$\bar{x}_1 - x_1 = \frac{\eta_2^2 \cdots \eta_{k+1}^2}{\det(T_k) \det(T_{k+1})}$$

If $x_1^{(j)} = (T_j^{-1} e^1, e^1)$ and $\omega_j = \det(T_j)$,

$$x_1^{(n)} - x_1^{(k)} = \eta_2^2 \cdots \eta_{k+1}^2 \left[\frac{1}{\omega_k \omega_{k+1}} + \frac{\eta_{k+2}^2}{\omega_{k+1} \omega_{k+2}} + \cdots + \frac{\eta_{k+2}^2 \cdots \eta_n^2}{\omega_{n-1} \omega_n} \right]$$

$$\omega_k = \frac{1}{\gamma_0 \cdots \gamma_{k-1}}$$

and

$$\frac{\eta_2^2 \cdots \eta_{k+1}^2}{\omega_k^2} = \beta_1 \cdots \beta_k.$$

Hence

$$x_1^{(n)} - x_1^{(k)} =$$

$$= \frac{\|r^k\|^2}{\|r^0\|^2}[\gamma_k + \beta_{k+1}\gamma_{k+1} + \beta_{k+1}\beta_{k+2}\gamma_{k+2} + \cdots + \beta_{k+1} \cdots \beta_{n-1}\gamma_{n-1}]$$

This shows that

$$\|\epsilon_k\|_A^2 =$$

$$= \|r^k\|^2[\gamma_k + \beta_{k+1}\gamma_{k+1} + \beta_{k+1}\beta_{k+2}\gamma_{k+2} + \cdots + \beta_{k+1} \cdots \beta_{n-1}\gamma_{n-1}]$$

Theorem

$$\|\epsilon^k\|_A^2 = \sum_{j=k}^{n-1} \gamma_j \|r^j\|^2$$

In their 1952 paper, Hestenes and Stiefel (Th 6.1 p 416) proved that

$$\|\epsilon^k\|_A^2 - \|\epsilon^l\|_A^2 = \sum_{j=k}^{l-1} \gamma_j \|r^j\|^2, \ l > k$$

For an elementary proof of the HS result, see

Z. Strakos and P. Tichy, On error estimation in the conjugate gradient method and why it works in finite precision computation, ETNA v ?, (2002), pp ??

Proof:

$$\|\epsilon^k\|_A^2 - \|\epsilon^{k+1}\|_A^2 = \|x - x^{k+1} + x^{k+1} - x^k\|_A^2 - \|\epsilon^{k+1}\|_A^2$$

$$= \|x^{k+1} - x^k\|_A^2 + 2(x - x^{k+1})^T A(x^{k+1} - x^k)$$

$$= \gamma_k^2(p_k, Ap^k) + 2(r^{k+1}, x^{k+1} - x^k) = \gamma_k \|r^k\|^2$$

The l_2 norm of the error

Theorem

$$\|\epsilon^k\|^2 = \|r^0\|^2[(e^1, T_n^{-2}e^1) - (e^1, T_k^{-2}e^1)]$$

$$+(-1)^k 2\eta_{k+1}\frac{\|r^0\|}{\|r^k\|}(e^k, T_k^{-2}e^1)\|\epsilon^k\|_A^2$$

Proof:

$$\|\epsilon^k\|^2 = (A^{-1}r^0, A^{-1}r^0) - 2(A^{-1}r^0, V_k z^k) + (V_k z^k, V_k z^k)$$

the troubles come from the term $(A^{-1}r^0, V_k z^k) = (r^0, A^{-1}V_k z^k)$

Then, we use

$$V_k T_k^{-1} = A^{-1}V_k + \eta_{k+1}A^{-1}v^{k+1}(e^k)^T T_k^{-1},$$

$$(r^0, A^{-1}v^{k+1}) = \frac{(-1)^k}{\|r^k\|}(r^0, A^{-1}r^k).$$

But

$$(r^0, A^{-1}r^k) = (r^0, \epsilon^k) = (r^0, \epsilon^0) - (r^0, V_k z^k).$$

Therefore,

$$(r^0, A^{-1}r^k) = (r^0, A^{-1}r^0) - \|r^0\|^2(e^1, T_k^{-1}e^1) = \|\epsilon^k\|_A^2.$$

This can also be written in a different way by using

$$r^k = -\eta_{k+1} \|r^0\| (e^k, T_k^{-1} e^1) v^{k+1}.$$

Therefore

$$\frac{(-1)^{k+1} \|r^k\|}{\eta_{k+1}} = \|r^0\| (e^k, T_k^{-1} e^1).$$

and

$$\|\epsilon^k\|^2 = \|r^0\|^2 [(e^1, T_n^{-2} e^1) - (e^1, T_k^{-2} e^1)]$$

$$-2 \frac{(e^k, T_k^{-2} e^1)}{(e^k, T_k^{-1} e^1)} \|\epsilon^k\|_A^2$$

Other possibility:

$$\epsilon^k = \|r^0\|(V_n T_n^{-1} e^1 - V_k T_k^{-1} e^1)$$

Let $\rho_j = (T_n^{-1} e^1, e^j), j = 1, \ldots, n$ and $\sigma_j = (T_k^{-1} e^1, e^j), j = 1, \ldots, k$

$$\epsilon^k = \|r^0\| \sum_{j=1}^{n} \tau_j v^j$$

$$\|\epsilon^k\|^2 = \|r^0\|^2 \sum_{j=1}^{n} \tau_j^2$$

where

$$\tau_j = \rho_j - \sigma_j,\ 1 \le j \le k, \quad \tau_j = \rho_j,\ k+1 \le j \le n$$

Relations with the eigenvalues and Ritz values

Let

$$\lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_n$$

be the eigenvalues of A and

$$\theta_1^{(k)} < \theta_2^{(k)} < \ldots < \theta_k^{(k)}$$

be the eigenvalues of T_k, the so–called Ritz values and $z_{(k)}^j$ the corresponding eigenvectors

Using the spectral decomposition of T_n and T_k we have

$$\|\epsilon^k\|_A^2 = \|r^0\|^2 \left[\sum_{j=1}^{n} \frac{[(z_{(n)}^j)_1]^2}{\lambda_j} - \sum_{j=1}^{k} \frac{[(z_{(k)}^j)_1]^2}{\theta_j^{(k)}} \right].$$

Therefore, we need the first components of the eigenvectors of T_k

Results are known about this problem: Thompson and McEnteggert (1968), see also Paige, Parlett, Scott

Let

$$T_{j,k} = \begin{pmatrix} \alpha_j & \eta_{j+1} & & & \\ \eta_{j+1} & \alpha_{j+1} & \eta_{j+2} & & \\ & \ddots & \ddots & \ddots & \\ & & \eta_{k-1} & \alpha_{k-1} & \eta_k \\ & & & \eta_k & \alpha_k \end{pmatrix}$$

and $\chi_{j,k}(\lambda) = \det(T_{j,k} - \lambda I)$

$$(z^i_{(k)})^2_1 = \left| \frac{\chi_{2,k}(\theta_i^{(k)})}{\chi'_{1,k}(\theta_i^{(k)})} \right|$$

We can obtain more suitable results by introducing the "first pivot" function

$$d_k^{(k)}(\lambda) = \alpha_k - \lambda, \quad d_l^{(k)}(\lambda) = \alpha_l - \lambda - \frac{\eta_{l+1}^2}{d_{l+1}^{(k)}(\lambda)}, \; l = k-1, \ldots, 1$$

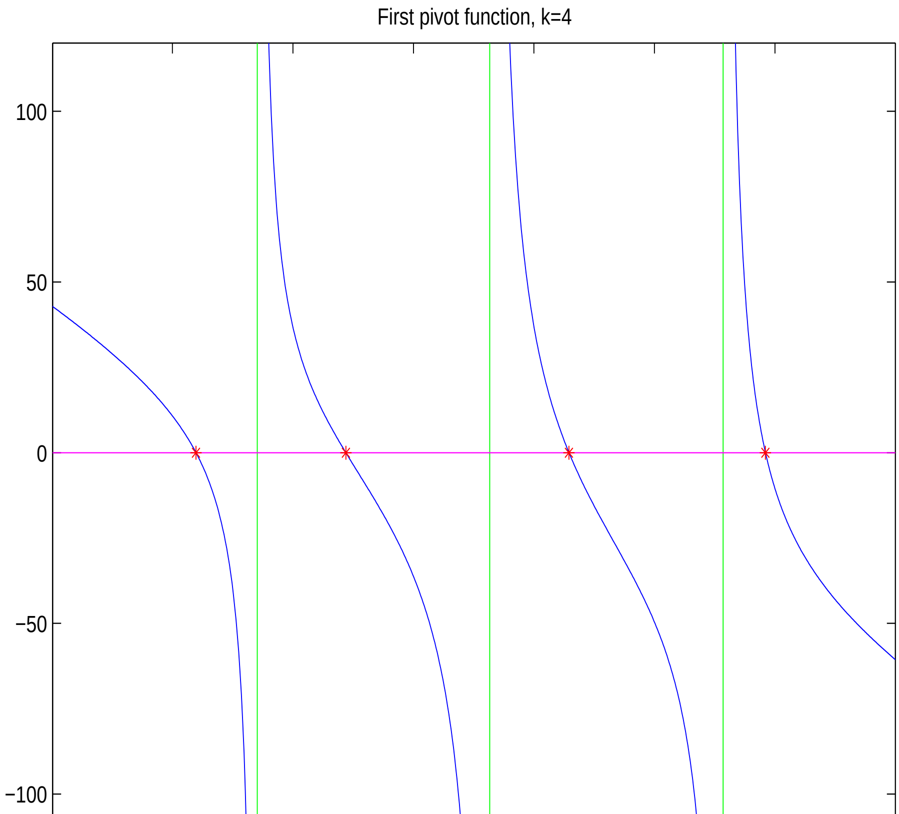
$$(z^i_{(k)})^2_1 = \left| \frac{1}{[d_1^{(k)}]'(\theta_i^{(k)})} \right|$$

Therefore

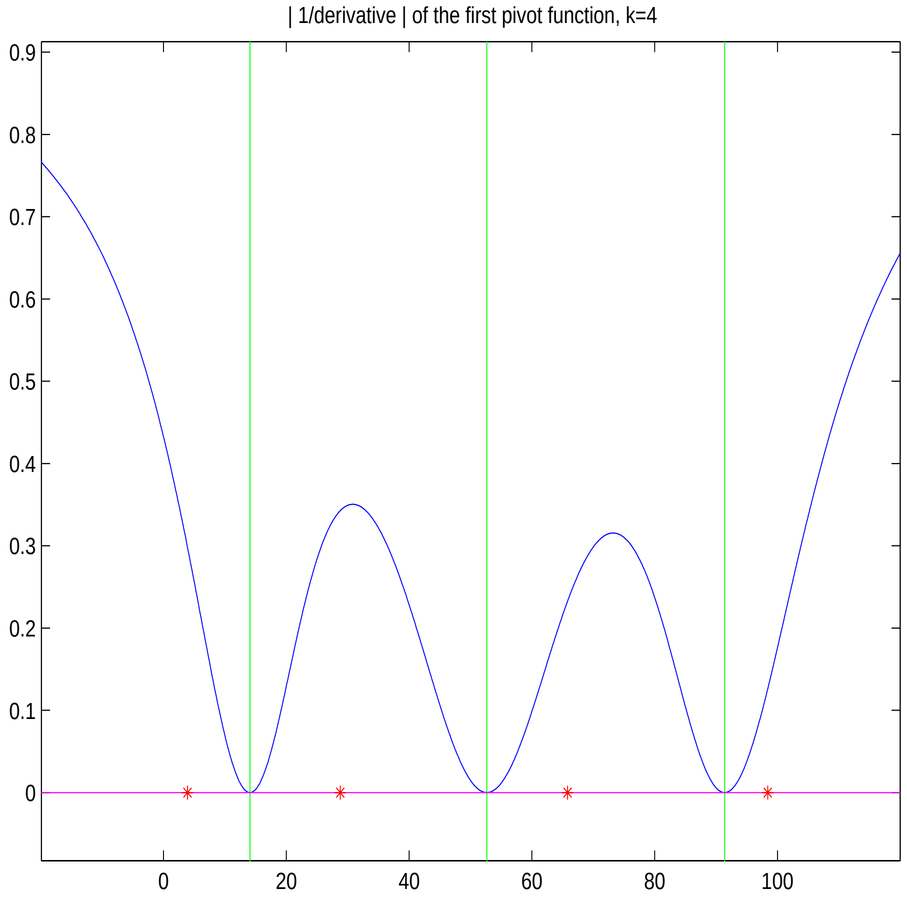
$$\|\epsilon^k\|_A^2 = \|r^0\|^2 \left[\sum_{j=1}^{n} \frac{1}{\lambda_j |[d_1^{(n)}]'(\lambda_j)|} - \sum_{j=1}^{k} \frac{1}{\theta_j^{(k)} |[d_1^{(k)}]'(\theta_j^{(k)})|} \right]$$

The first pivot function has zeros at the eigenvalues of T_k and poles at the eigenvalues of $T_{2,k}$

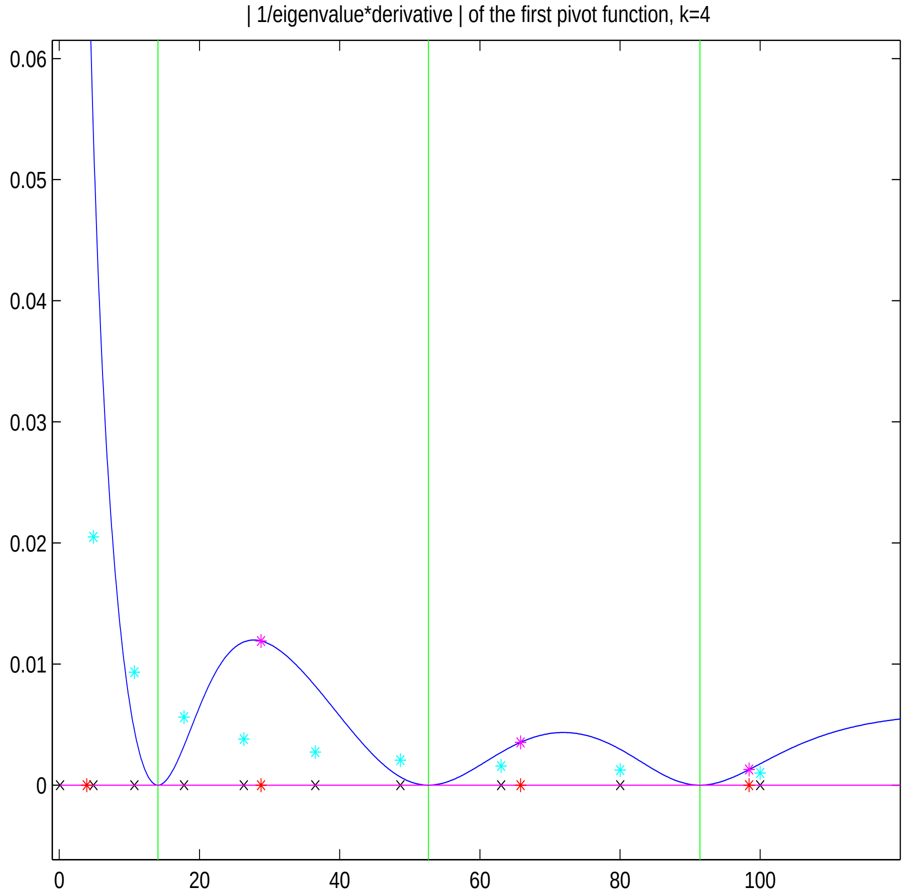
Example: Strakos matrix, $n = 10$, $\lambda_1 = 0.01$, $\lambda_n = 100$



First pivot function $d_1^{(k)}$



1 / | derivative of first pivot function $d_1^{(k)}$ |

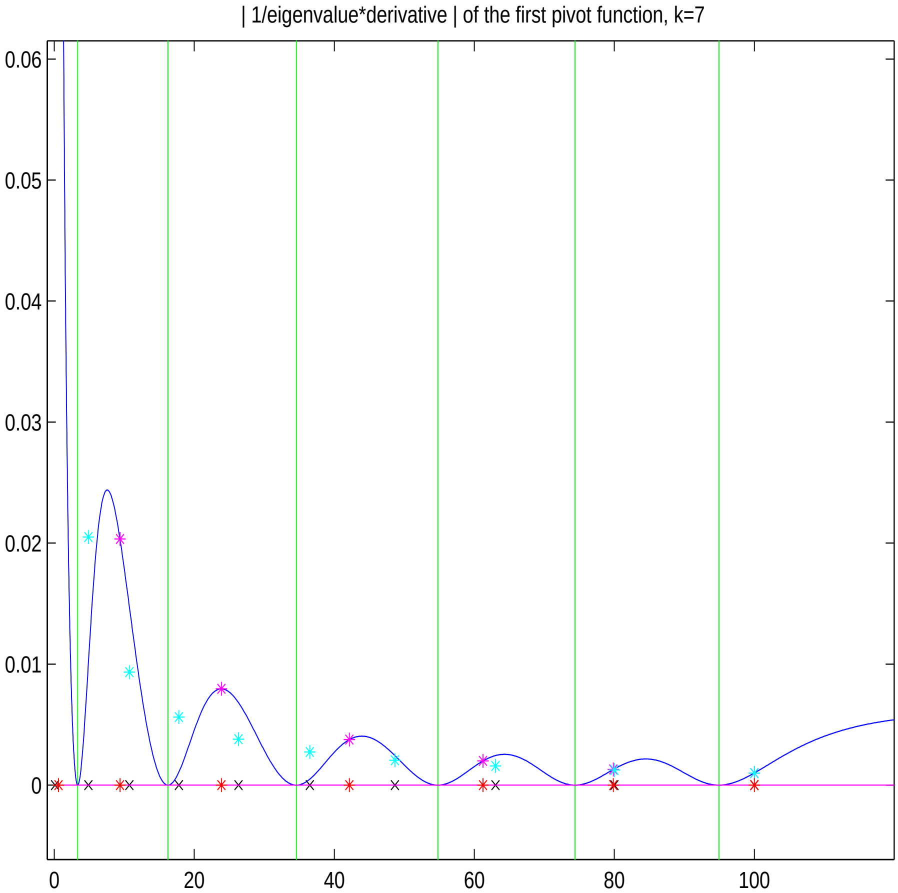


$1 \;/\; \lambda \;|$ derivative of first pivot function $d_1^{(k)} \;|$

light blue: values at λ_i, magenta: values at $\theta_i^{(4)}$

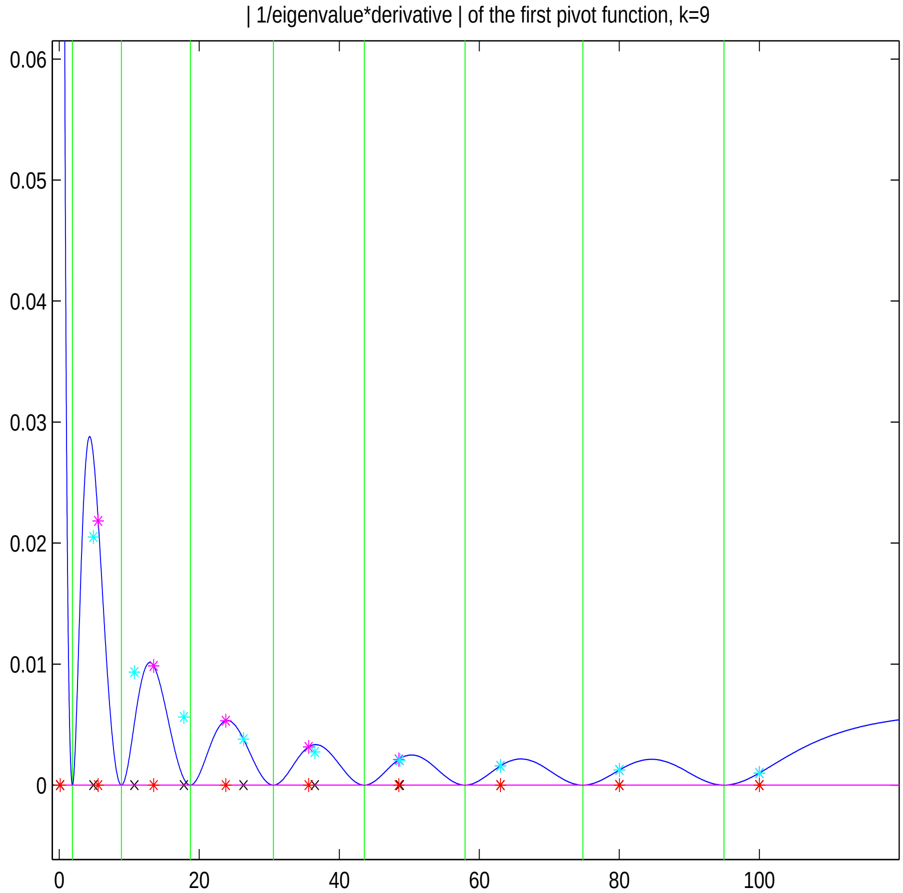
There is only one pole at 0

We see the largest eigenvalue does not contribute anymore to the norm of the error



$1 \ / \ \lambda \ |$ derivative of first pivot function $d_1^{(7)} \ |$,
iteration 7

light blue: values at λ_i, magenta: values at $\theta_i^{(k)}$



1 / λ | derivative of first pivot function $d_1^{(9)}$ |,
iteration 9

light blue: values at λ_i, magenta: values at $\theta_i^{(k)}$

How to compute bounds of the norms of the error?

Gauss quadrature gives a lower bound, to obtain an upper bound we use Gauss–Radau but we need an estimate of the smallest eigenvalue of A

This has been studied in

G. Meurant, The computation of bounds for the norm of the error in the conjugate gradient algorithm, Numerical Algorithms v 16, (1997), pp 77–87

G. Meurant, Numerical experiments in computing bounds for the norm of the error in the preconditioned conjugate gradient algorithm, Numerical Algorithms v 22, (1999), pp 353–365

One way is to use the Hestenes and Stiefel formula (which implicitly uses the UL decomposition of T_k)

Another possibility is to use the Sherman–Morrison formula

$$T_{k+1} = \begin{pmatrix} T_k & \eta_{k+1} e^k \\ \eta_{k+1} (e^k)^T & \alpha_{k+1} \end{pmatrix}$$

Let

$$\bar{T} = T_k - \frac{\eta_{k+1}^2}{\alpha_{k+1}} e^k (e^k)^T$$

The upper left block of T_{k+1}^{-1} is $\bar{T}^{-1}$ and

$$\bar{T}^{-1} = T_{k+1}^{-1} + \eta_{k+1}^2 \frac{(T_k^{-1} e^k)((e^k)^T T_k^{-1})}{\alpha_{k+1} - \eta_{k+1}^2 (e^k)^T T_k^{-1} e^k}$$

so we need the first and last element of $T_k^{-1} e^k$

Clearly

$$(T_k^{-1}e^k)_k = \frac{1}{\delta_k}$$

Therefore, the denominator is

$$\alpha_{k+1} - \frac{\eta_{k+1}^2}{\delta_k} = \delta_{k+1} = \frac{1}{\gamma_k}$$

The numerator is

$$\frac{(\eta_1 \dots \eta_{k+1})^2}{(\delta_1 \dots \delta_k)^2}$$

Starting from $c_1 = 1$ we can compute it by

$$c_k = c_{k-1}\frac{\eta_k}{\delta_{k-1}} = c_{k-1}\eta_k\gamma_{k-2}$$

Finally

$$(T_{k+1}e^1)_1 = (T_k e^1)_1 + c_{k+1}^2\gamma_k$$

It can be easily proved that $c_{k+1}^2 = \|r^k\|^2$

So, we are back to Hestenes and Stiefel!

However, using Sherman–Morrison is useful when using Gauss-Radau

All this is true in exact arithmetic. Things can be a little different in floating point

Calvetti and Reichel proposed to use Anti–Gauss quadrature rule (Laurie)

It is the same as Gauss quadrature except that the matrix is

$$\begin{pmatrix} \alpha_1 & \eta_2 & & & \\ \eta_2 & \alpha_2 & \eta_3 & & \\ & \ddots & \ddots & \ddots & \\ & & \eta_{k-1} & \alpha_{k-1} & \sqrt{2}\eta_k \\ & & & \sqrt{2}\eta_k & \alpha_k \end{pmatrix}$$

The last step of the Cholesky decomposition is modified to

$$\bar{\delta}_k = \delta_k - \frac{\eta_k^2}{\delta_{k-1}}$$

Here also Sherman–Morrison is useful!

Most of the time anti–Gauss gives an upper bound, specially when CG convergence is fast

Bounds and estimates

Of course, we do not know $(T_n^{-1}e^1, e^1)$ at iteration k!

We introduce a delay $d > 0$ and write

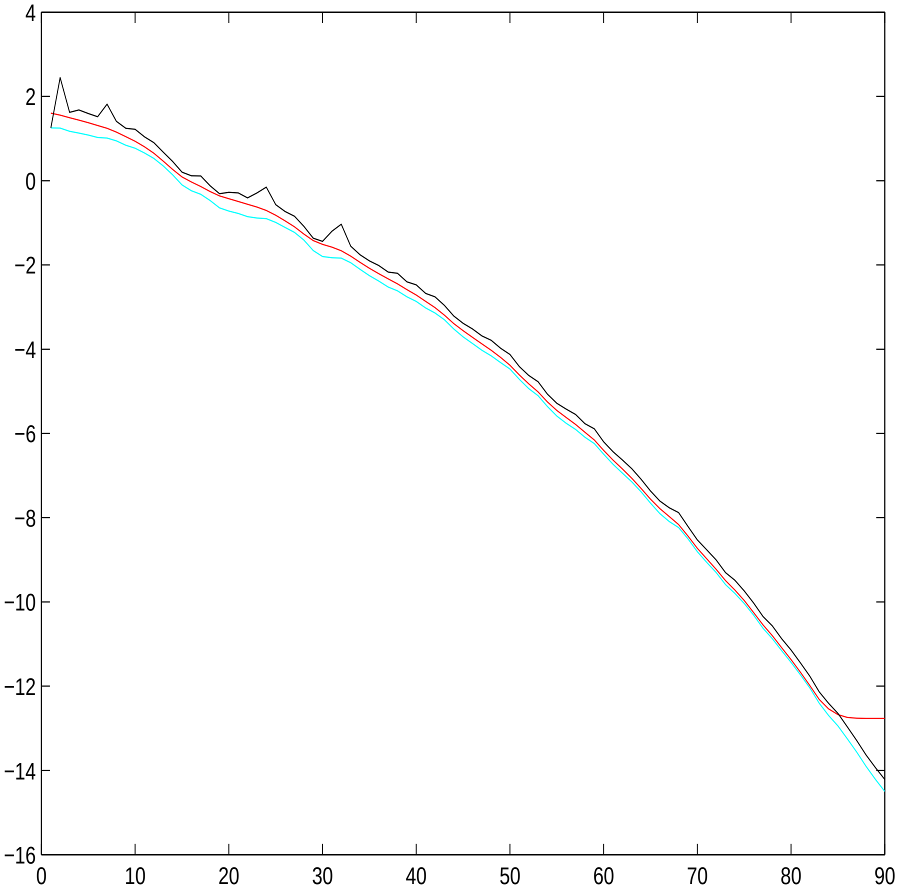
$$\|\epsilon^{k-d}\|_A^2 \simeq \|r^0\|^2[(T_k^{-1}e^1, e^1) - (T_{k-d}^{-1}e^1, e^1)]$$

This is easily computable during the CG iterations (for free!)

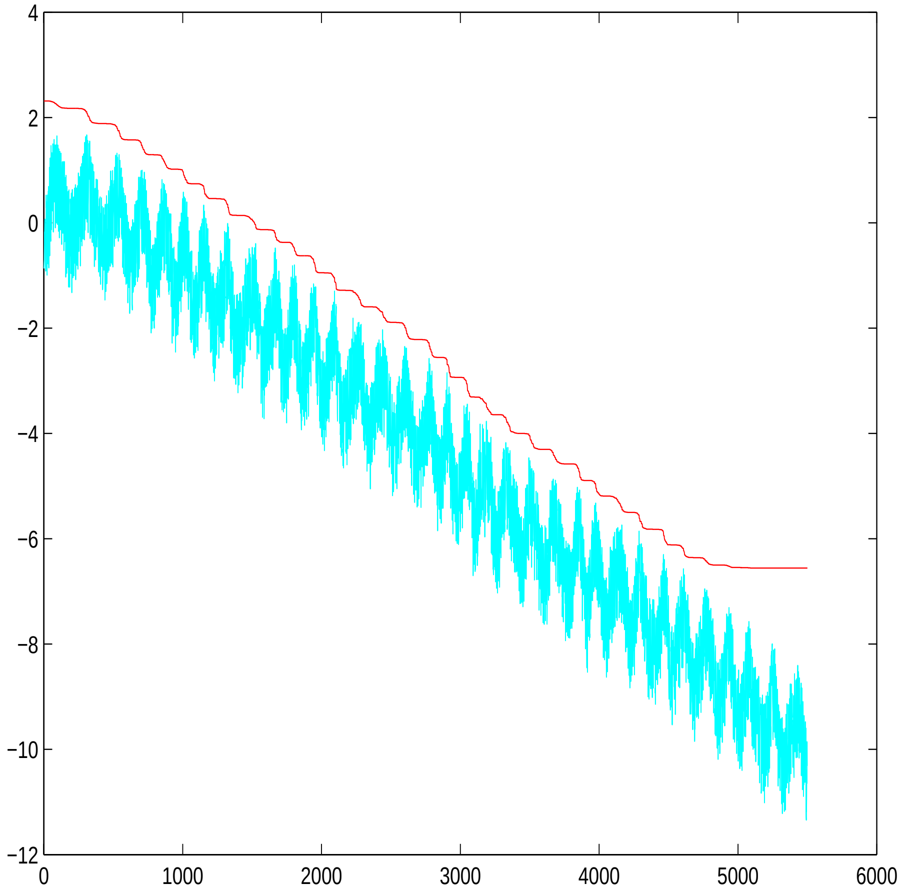
Gauss quadrature gives a lower bound (most of the time very close to the exact value), Gauss–Radau an upper bound and anti–Gauss sometimes an upper bound

Strakos and Tichy proved that these bounds work also in finite precision

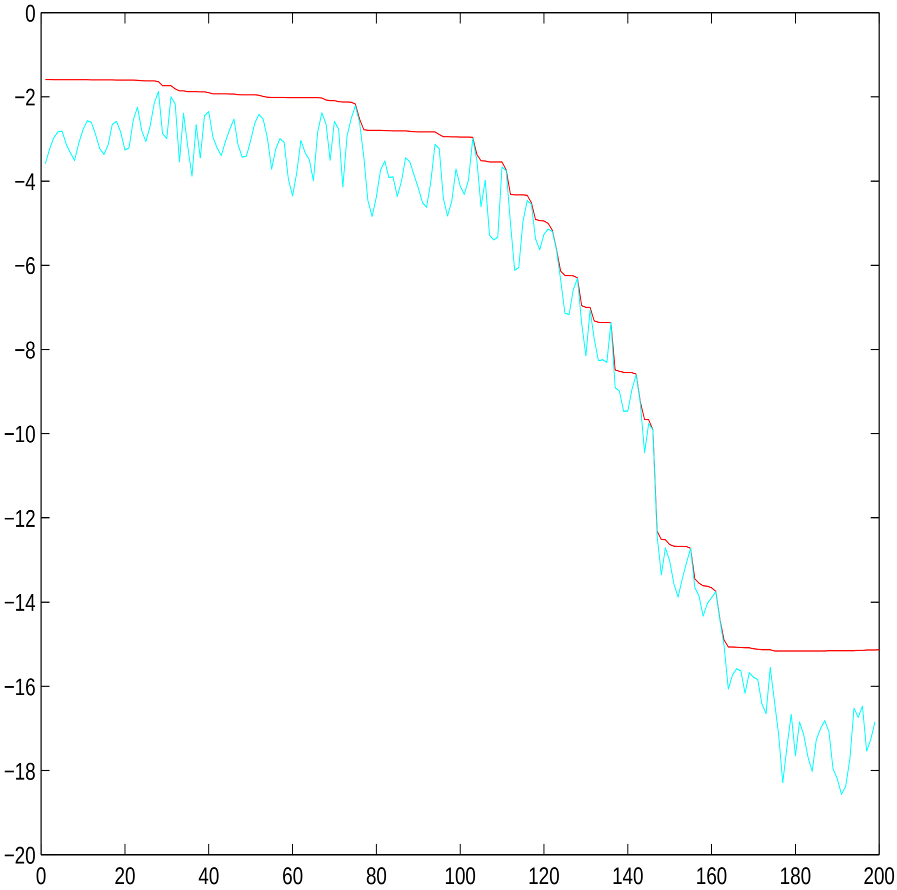
Examples:



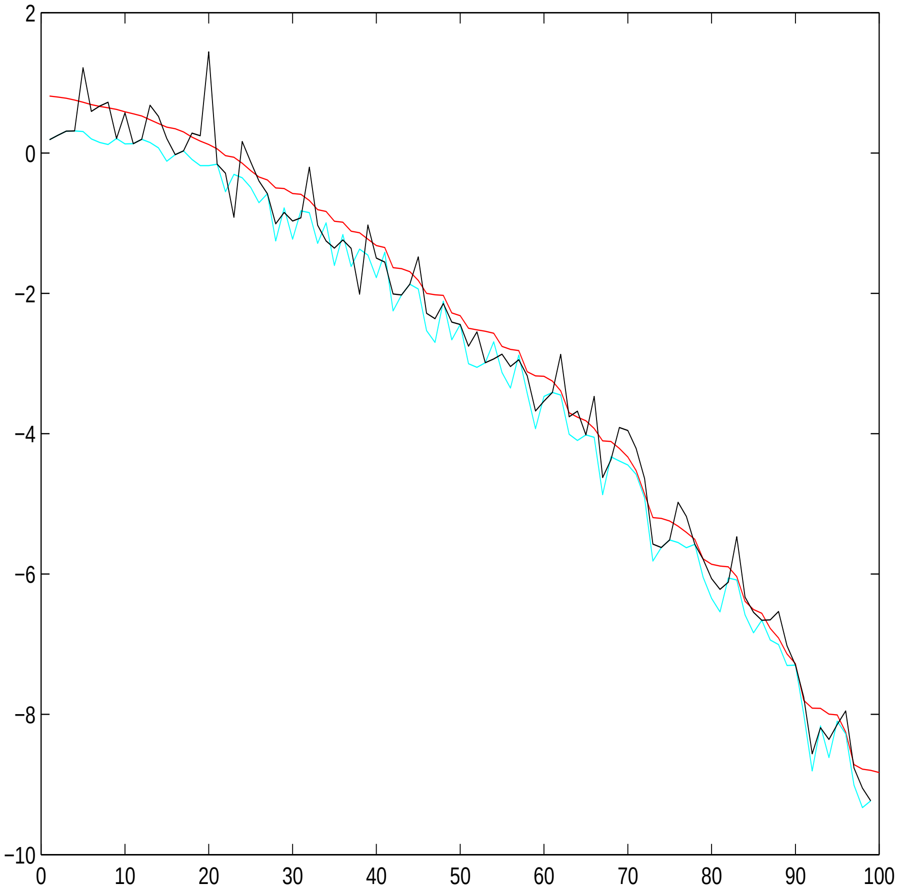
Poisson problem, $n = 400$, $d = 1$



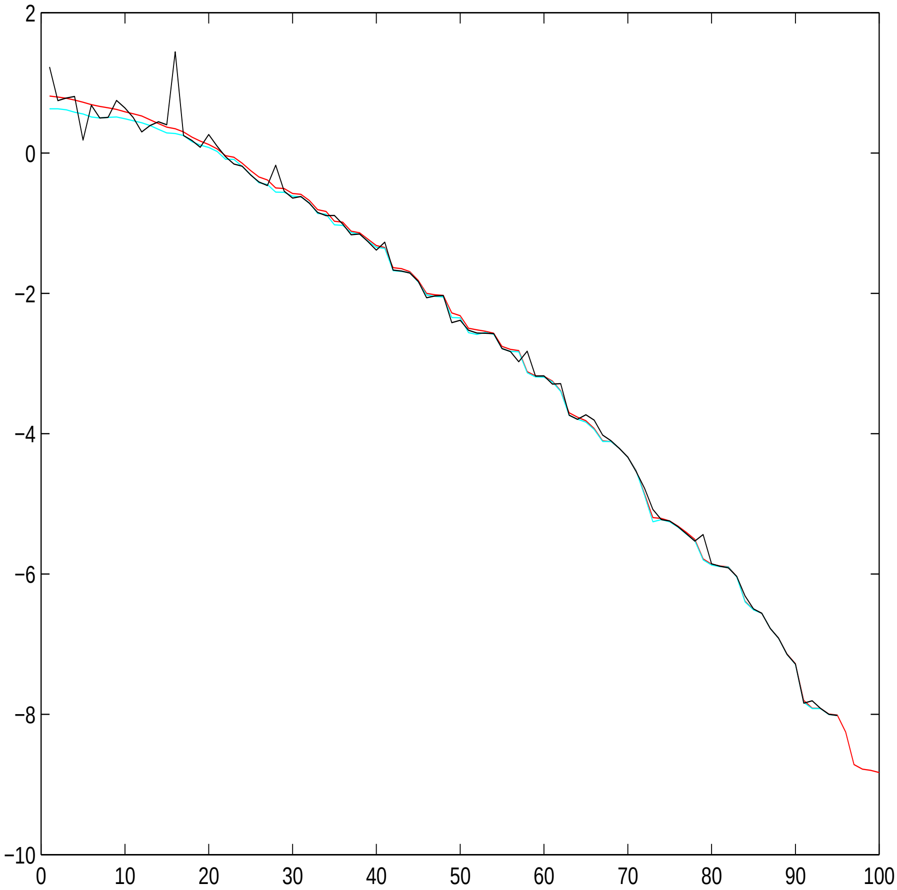
Nos7 problem, $d = 1$



Bcsstk01 problem $d = 1$



Strakos problem, $n = 48$, $d = 1$



Strakos problem, $n = 48$, $d = 5$

How to compute the l_2 norm estimates?

$$\|\epsilon^k\|^2 = \|r^0\|^2[(e^1, T_n^{-2}e^1) - (e^1, T_k^{-2}e^1)]$$

$$+(-1)^k 2\eta_{k+1}\frac{\|r^0\|}{\|r^k\|}(e^k, T_k^{-2}e^1)\|\epsilon^k\|_A^2$$

$$\|\epsilon^k\|^2 \simeq \|r^0\|^2[(e^1, T_k^{-2}e^1) - (e^1, T_{k-d}^{-2}e^1)]$$

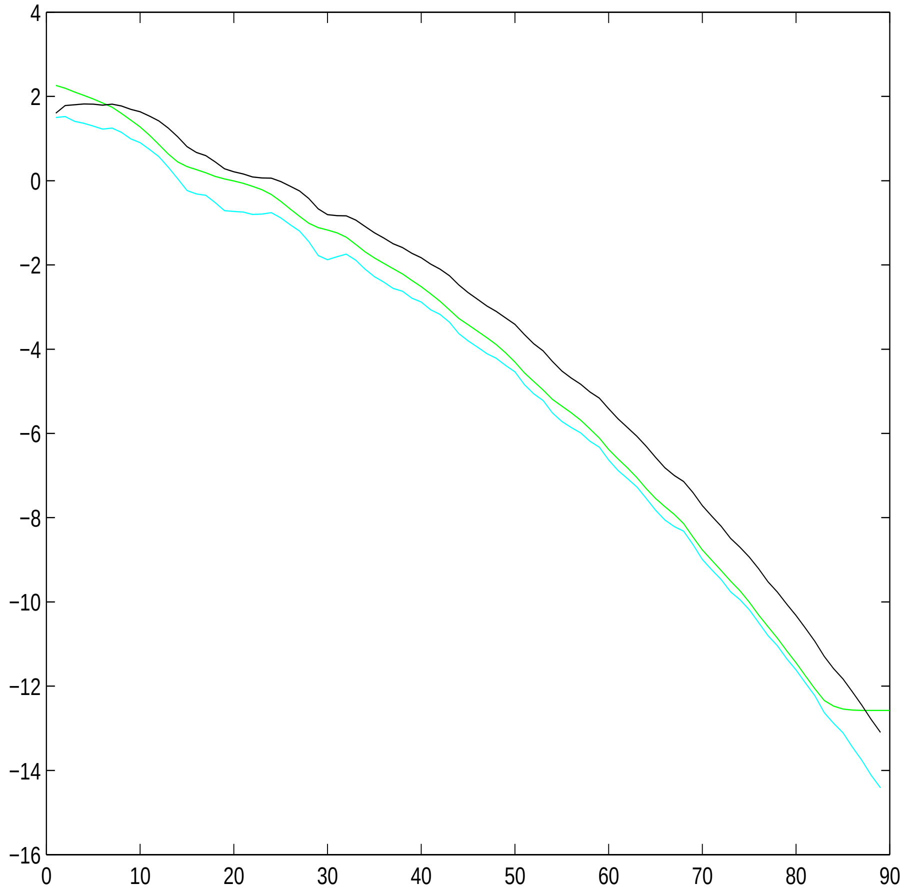
$$+(-1)^{k-d} 2\eta_{k+1-d}\frac{\|r^0\|}{\|r^{k-d}\|}(e^{k-d}, T_{k-d}^{-2}e^1)\|\epsilon^{k-d}\|_A^2$$

Use a QR decomposition of T_k (computed with Givens rotations)

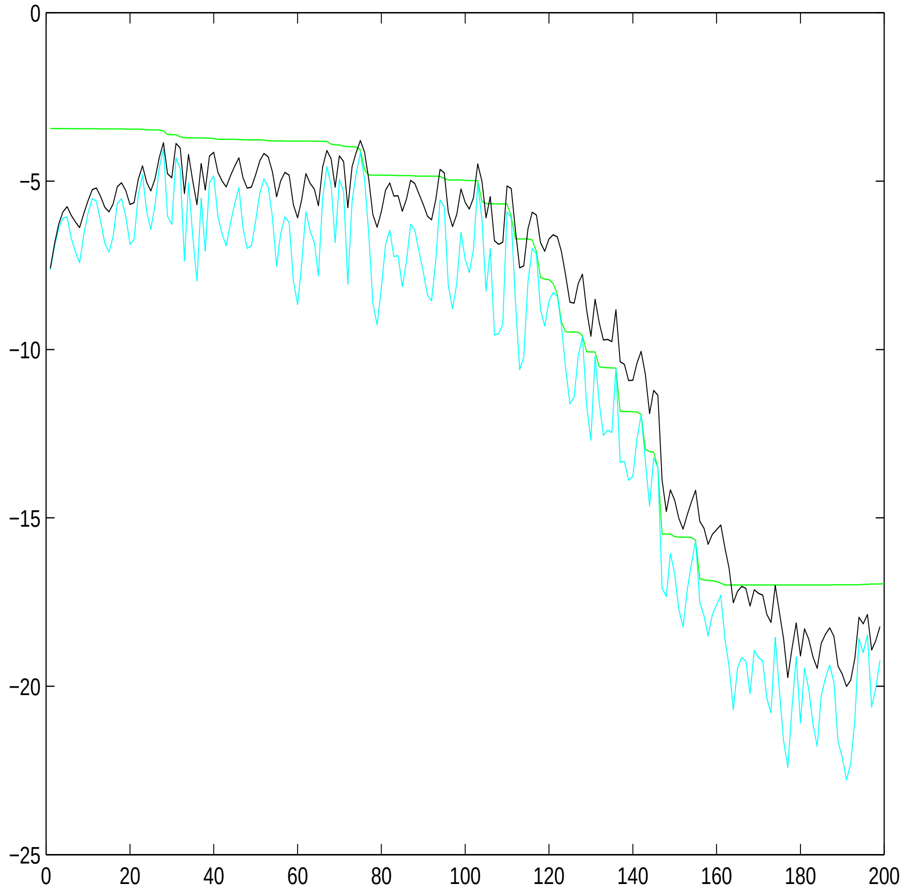
$$Q_k T_k = R_k$$

$$(e^1, T_k^{-2}e^1) = (R_k^{-T}e^1, R_k^{-T}e^1)$$

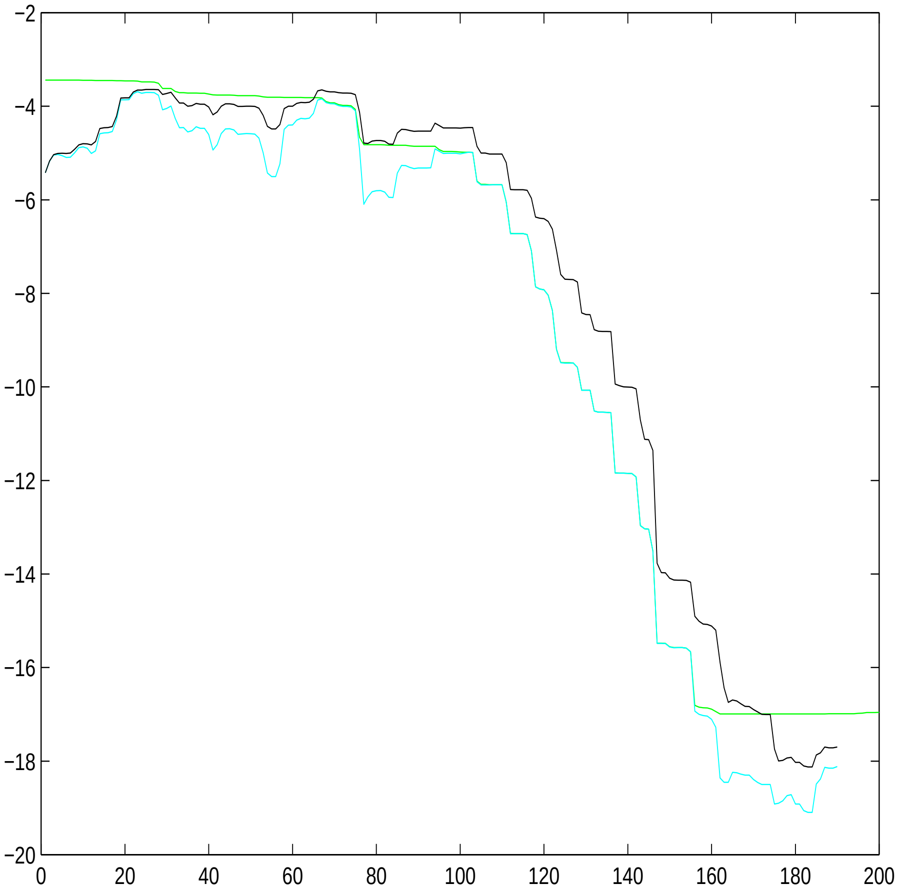
R_k has only 3 non zero diagonals



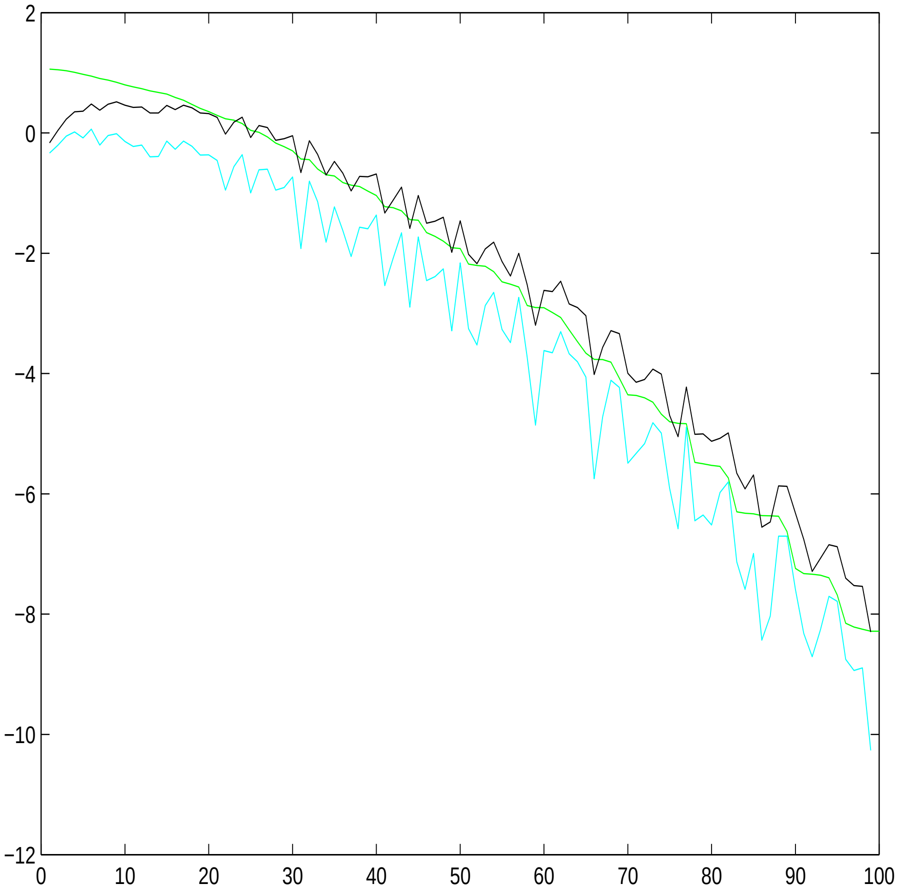
Poisson problem, $n = 400$, $d = 1$



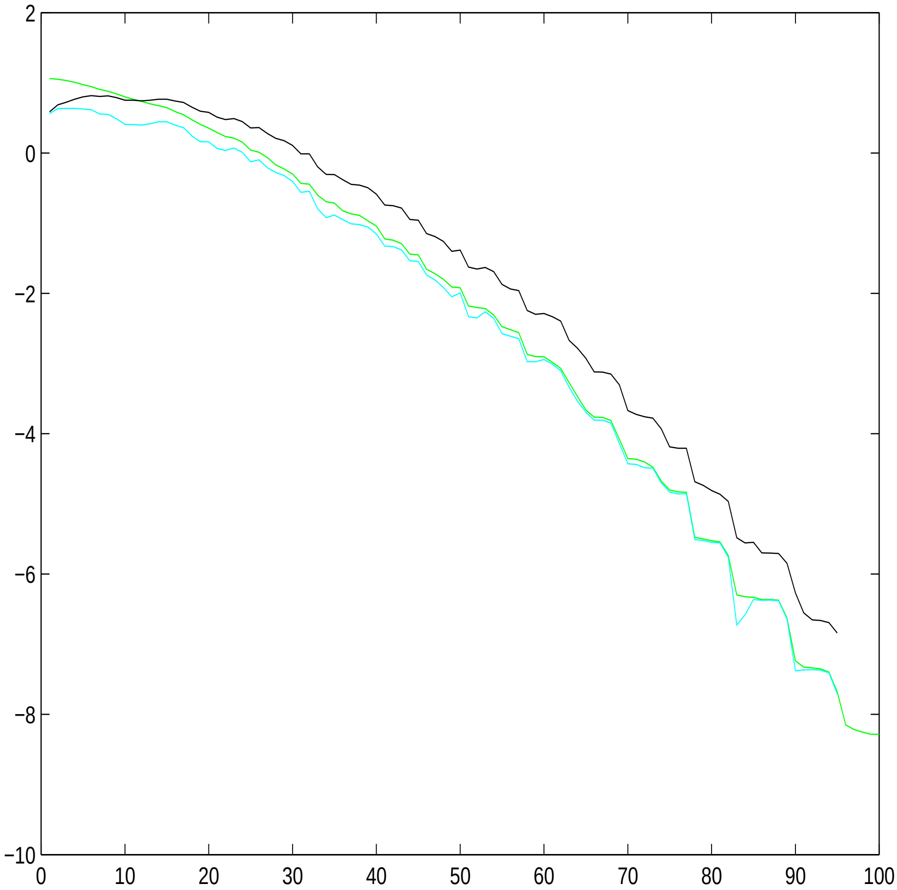
Bcsstk01 problem $d = 1$



Bcsstk01 problem $d = 10$



Strakos problem, $n = 48$, $d = 1$

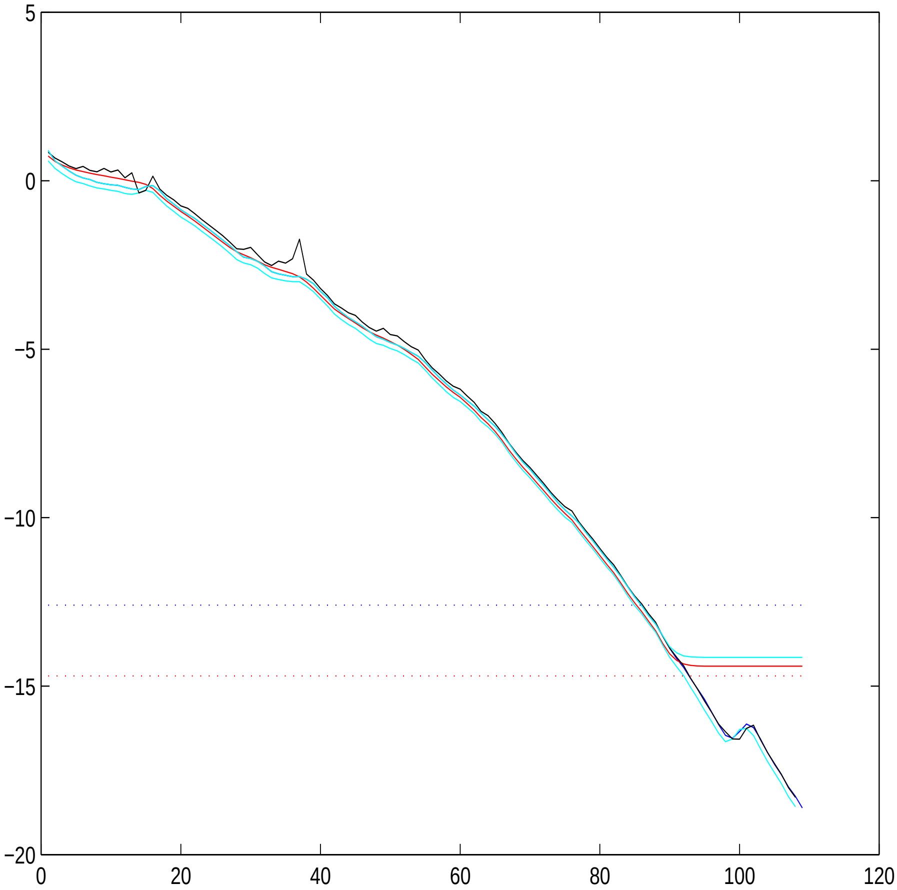


Strakos problem, $n = 48$, $d = 5$

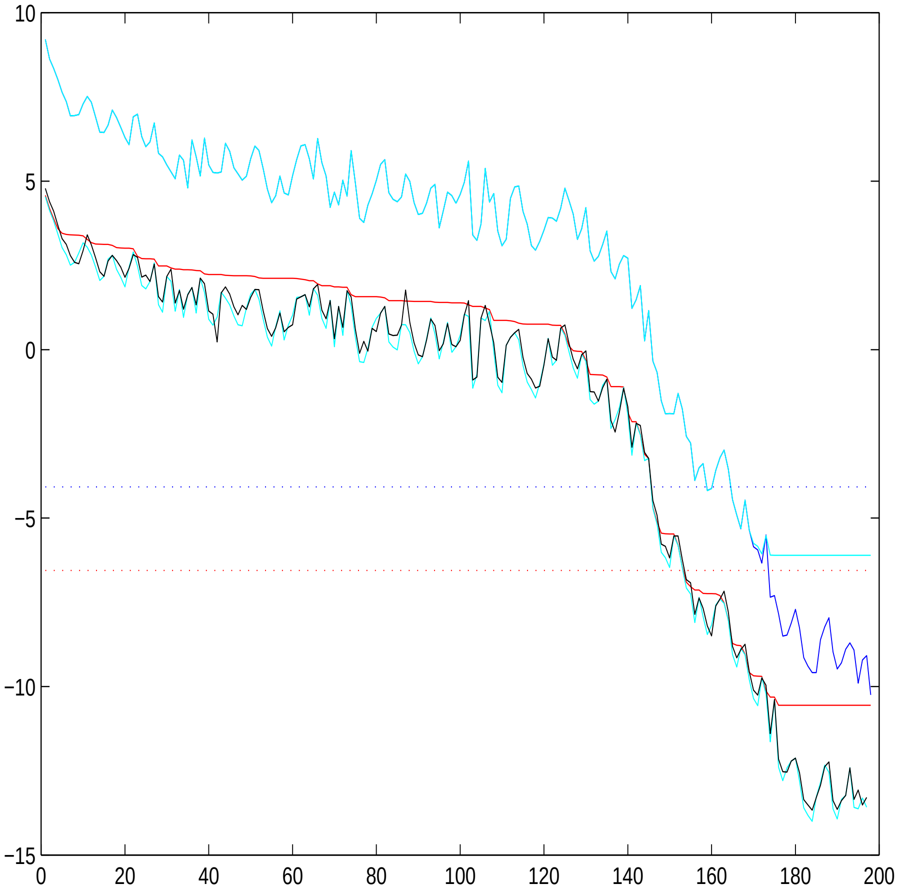
Conclusion

There is no reason not to use the estimates of the norms as stopping criterion for CG

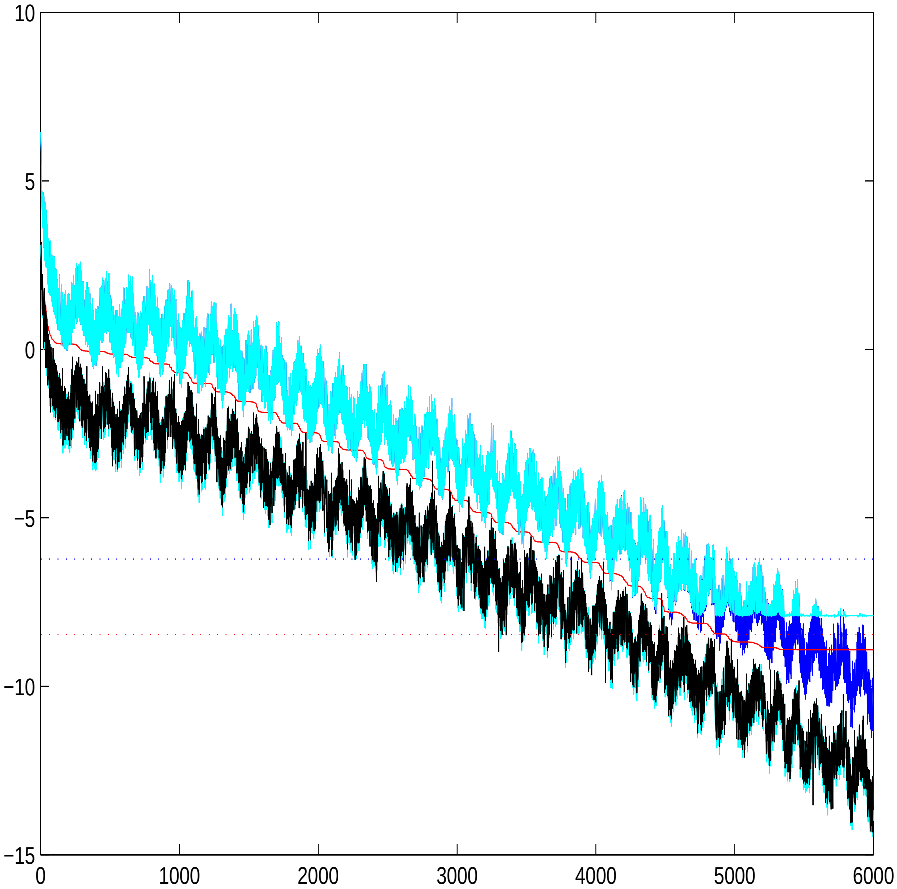
Everything is (almost) under control since there are (partial) answers to the other questions



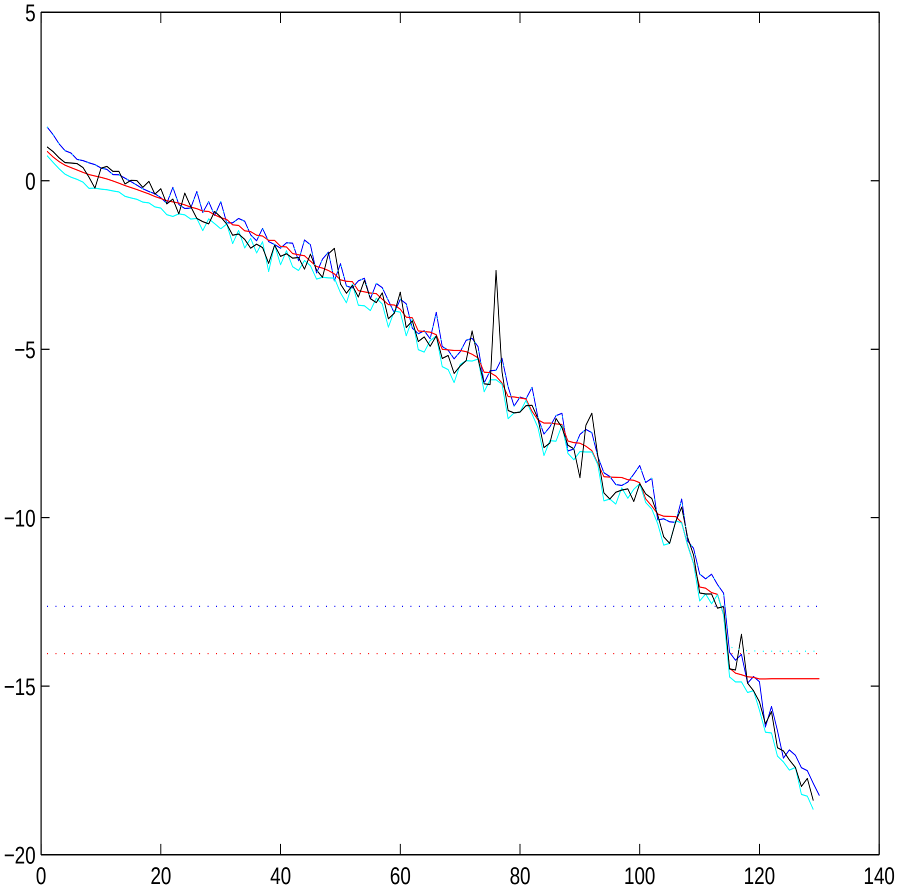
Poisson problem



Bcsstk01



Nos7



Strakos, $n = 48$