On the iterative solution of systems of the form $A^\top Ax = A^\top b + c$

E. Riccietti (IRIT-INP, Toulouse)

Joint work with: H. Calandra (TOTAL)
S. Gratton (IRIT-INP, Toulouse)
X. Vasseur (ISAE-SUPAERO, Toulouse)

Manchester University

30th July, 2019

Context

Given $A \in \mathbb{R}^{m \times n}$, $m \geq n$ with $\mathrm{rank}(A) = n$, $b \in \mathbb{R}^m$ and $x, c \in \mathbb{R}^n$, solve

$$A^\top A x = A^\top b + \mathbf{c} \tag{SYS}$$

or

$$\min_x \|Ax - b\|^2 - x^\top \mathbf{c}$$

Remarks

- This is a generalization of the normal equations for least-squares problems (case $c = 0$)

Motivating applications (I)

- Multilevel Levenberg-Marquardt method

Calandra, H., Gratton, S., Riccietti, E., Vasseur, X., *On the approximation of the solution of partial differential equations by artificial neural networks trained by a multilevel Levenberg-Marquardt method*, arXiv e-print, 2019

$$\min_{x\in\mathbb{R}^n} f(x) = \frac{1}{2}\|F(x)\|^2.$$

We have at disposal an approximation to the objective function:

$$f^H(x^H) = \frac{1}{2}\|F^H(x^H)\|^2, \quad x^H \in \mathbb{R}^{n_H},\ n_H < n$$

Coarse model:

$$m_k^H(x_k^H, s^H) = \frac{1}{2}\|F^H(x_k^H) + J^H(x_k^H)s^H\|^2 + \frac{\lambda_k}{2}\|s^H\|^2 + (R\nabla f(x_k) - \nabla f^H(x_0^H))^\top s^H,$$

with $J^H(x_k^H)$ the Jacobian matrix of F^H at x_k^H, R a full-rank linear restriction operator and $x_0^H = Rx_k$.

Motivating applications (II)

- Penalty function method

Fletcher, R., *A class of methods for nonlinear programming: III. Rates of convergence*, Numerical Methods for Nonlinear Optimization, 1973

Estrin, R. and Orban, D. and Saunders, M. A., *LNLQ: An iterative method for least-norm problems with an error minimization property*, technical report, 2018

$$\min_x f(x)$$
$$\text{s.t. } g(x) = 0,$$

Penalty function :

$$\Phi_\sigma(x) = f(x) - g(x)^\top y_\sigma(x),$$

where $y_\sigma(x) \in \mathbb{R}^m$ is defined as the solution of the following minimization problem:

$$\min_y \|A(x)^\top y - \nabla f(x)\|^2 + \sigma g(x)^\top y,$$

with $A(x)$ the Jacobian matrix of $g(x)$ at x and $\sigma > 0$, a given real-valued penalty parameter.

Interesting questions

- What is the conditioning of $A^\top Ax = A^\top b + c$?
  - Standard theory for linear systems do no take into account structured perturbations and gives underwhelming results
  - Structured conditioning analysis is necessary. Presence of c results in a different mapping from data to solution
- What is the backward error?
  - Different set of admissible perturbations on the matrix
- How to numerically solve it by an iterative method?
  - Methods for normal equations such as CGLS cannot be used.

THEORETICAL RESULTS

Conditioning, case $c = 0$

Let $\delta x = x - \hat{x}$, $\hat{x}$ a perturbed solution.

Forward error bound

From standard theory on linear systems:

$$\frac{\|\delta x\|}{\|x\|} \leq \kappa(A)^2 u$$

For least squares problems:

$$\frac{\|\delta x\|}{\|x\|} \leq \gamma_m \kappa_{LS} u, \quad \kappa_{LS} = \kappa(A)\left(1 + \frac{\|A^\dagger\|\|r\|}{\|x\|}\right), \quad r = b - Ax$$

Underwhelming result!

The conditioning of the problem depends on $\kappa(A)^2$ only if $\|r\|$ is large!

Conditioning

Definition

If F is a continuously differentiable function

$$\begin{aligned} F : \mathcal{X} &\to \mathcal{Y} \\ x &\longmapsto F(x), \end{aligned}$$

the absolute condition number of F at x is the scalar $\|F'(x)\|_{\mathrm{op}}$. The relative condition number of F at x is

$$\frac{\|F'(x)\|_{\mathrm{op}} \, \|x\|_{\mathcal{X}}}{\|F(x)\|_{\mathcal{Y}}}.$$

J . R . Rice, *A theory of condition*, SIAM J . Numer . Anal ., 1966

Conditioning, case $c = 0$

Definition of F

We consider F as the function that maps A, b to the solution x of a least squares problem:

$$F : \mathbb{R}^{m\times n} \times \mathbb{R}^m \to \mathbb{R}^n$$
$$(A, b) \longmapsto F(A, b) = A^\dagger b.$$

Explicit formula for the conditioning

The absolute condition number of a least-squares problem, with Euclidean norm on the solution and Frobenius norm on the data[a], is given by

$$\kappa_{NE} = \|A^\dagger\|\sqrt{1 + \|x\|^2 + \|A^\dagger\|^2\|r\|^2}$$

Gratton, S., *On the condition number of linear least squares problems in a weighted Frobenius norm*, BIT Numerical Mathematics, 1996

[a] $\|[A, b]\|_F^2 := \|A\|_F^2 + \|b\|^2$

A formula for the condition number, $c \neq 0$

Lemma

The absolute condition number of the problem SYS is given by

$$\|F'(A,b,c)\|_{\mathrm{op}} = \|[(r^{\top} \otimes (A^T A)^{-1})L_T + x^{\top} \otimes A^{\dagger}, A^{\dagger}, (A^T A)^{-1}]\|,$$

where L_T is the linear operator such that $\mathrm{vec}(A^{\top}) = L_T \mathrm{vec}(A)$ and $r = b - Ax$.

Case $c = 0$

$$\|F'(A,b,c)\|_{\mathrm{op}} = \|[(r^{\top} \otimes (A^T A)^{-1})L_T + x^{\top} \otimes A^{\dagger}, A^{\dagger}]\|.$$

An explicit formula for the condition number, $c \neq 0$

We consider F as the function that maps A, b, c to the solution x of SYS

$$F : \mathbb{R}^{m\times n} \times \mathbb{R}^m \times \mathbb{R}^n \to \mathbb{R}^n$$
$$(A, b, c) \longmapsto F(A, b, c) = A^\dagger b + A^\dagger (A^\dagger)^\top c.$$

Theorem

The absolute condition number of problem SYS, with Euclidean norm on the solution and Frobenius norm on the data[a], is $\sqrt{\|\bar{M}\|}$, with $\bar{M} \in \mathbb{R}^{n\times n}$ given by

$$\bar{M} = (1 + \|r\|^2)(A^\top A)^{-2} + (1 + \|x\|^2)(A^\top A)^{-1} - 2\,\mathrm{sym}(B),$$

with $B = A^\dagger r x^\top (A^T A)^{-1}$, $\mathrm{sym}(B) = \frac{1}{2}(B + B^\top)$ and x the exact solution of SYS.

[a] $\|[A, b, c]\|_F^2 := \|A\|_F^2 + \|b\|^2 + \|c\|^2$

Upper bound for the condition number

$$\sqrt{\|\bar{M}\|} \leq (1 + \|r\| + 2\sqrt{\|c\|\|x\|})\|A^\dagger\|^2 + (1 + \|x\|)\|A^\dagger\|.$$

Backward error analysis

Let $A \in \mathbb{R}^{m\times n}$, $b \in \mathbb{R}^m$, $c \in \mathbb{R}^n$ and $\tilde{x}$ a perturbed solution to SYS. Find the smallest perturbation E of A such that the vector $\tilde{x}$ exactly solves

$$(A+E)^\top (A+E)x = (A+E)^\top b + c,$$

i.e. given

$$\mathcal{G} := \{E \in \mathbb{R}^{m\times n} : (A+E)^\top (A+E)\tilde{x} = (A+E)^\top b + c\},$$

we want to compute the quantity:

$$\eta(\tilde{x}) = \min_{E\in\mathcal{G}} \|E\|_F.$$

Set of admissible perturbations on the matrix

Theorem

Let $A \in \mathbb{R}^{m\times n}$, $b \in \mathbb{R}^m$, $c, \tilde{x} \in \mathbb{R}^n$ and assume that $\tilde{x} \neq 0$. Let $\tilde{r} = b - A\tilde{x}$ and define two sets $\mathcal{E}, \mathcal{M}$ by

$$\begin{aligned}
\mathcal{E} =& \{E \in \mathbb{R}^{m\times n} : \ (A+E)^\top (b-(A+E)\tilde{x}) = -c\,\},\\
\mathcal{M} =& \{v\left(\alpha c^\top - v^\dagger A\right) + (I_m - vv^\dagger)(\tilde{r}\tilde{x}^\dagger + Z(I_n - \tilde{x}\tilde{x}^\dagger)) \ :\\
& \qquad v \in \mathbb{R}^m,\ Z \in \mathbb{R}^{m\times n}, \alpha \in \mathbb{R},\ s.t.\ \alpha\|v\|^2(v^\dagger b - \alpha c^\top \tilde{x}) = -1\}.
\end{aligned}$$

Then $\mathcal{E} = \mathcal{M}$.

Case $c = 0$

$$\begin{aligned}
\mathcal{E} =& \{E \in \mathbb{R}^{m\times n} : \ (A+E)^\top (b-(A+E)\tilde{x}) = 0\,\},\\
\mathcal{M} =& \{-vv^\dagger A + (I_m - vv^\dagger)(\tilde{r}\tilde{x}^\dagger + Z(I_n - \tilde{x}\tilde{x}^\dagger)) \ : v \in \mathbb{R}^m,\ Z \in \mathbb{R}^{m\times n}\}.
\end{aligned}$$

Lower bound on the backward error

Lemma

The set of admissible perturbations $\mathcal{E}$ defined in Theorem is such that $\mathcal{E} \subseteq \mathcal{M}_2$, with

$$\mathcal{M}_2 = \{v\left(\alpha c^\top - v^\dagger A\right) + (I_m - vv^\dagger)(\tilde{r}\tilde{x}^\dagger + Z(I_n - \tilde{x}\tilde{x}^\dagger)) \; : \\ v \in \mathbb{R}^m,\, Z \in \mathbb{R}^{m\times n},\, \alpha \in \mathbb{R}\}.$$

Then,

$$\min_{\mathcal{E}} \|E\|_F^2 \geq \min_{\mathcal{M}_2} \|E\|_F^2 = \frac{\|\tilde{r}\|^2}{\|\tilde{x}\|^2} + \min\{\lambda_*, 0\},$$

for $\lambda_* = \lambda_{\min}\left(A(I_n - cc^\top)A^\top - \dfrac{\tilde{r}\tilde{r}^\top}{\|\tilde{x}\|^2}\right)$, with $\lambda_{\min}(M)$ denoting the smallest eigenvalue of the matrix M.

Case $c = 0$

$$\min_{\mathcal{E}} \|E\|_F^2 = \frac{\|\tilde{r}\|^2}{\|\tilde{x}\|^2} + \min\{\lambda_*, 0\}, \quad \lambda_* = \lambda_{\min}\left(AA^\top - \frac{\tilde{r}\tilde{r}^\top}{\|\tilde{x}\|^2}\right).$$

NUMERICAL SOLUTION OF THE SYSTEM

CG vs CGLS for normal equations

Same method in exact arithmetic, different performance in finite precision for some problems:

- in CGLS $d_k = b - Ax_k$ is recurred and $r_k = A^\top d_k$.

Algorithm 1 CG for $A^\top Ax = A^\top b$

Input: A, b, x_0.
Define $r_0 = A^\top(b - Ax_0)$, $p_1 = r_0$.
for $k = 1, 2, \ldots$ **do**
$$\alpha_k = \frac{r_{k-1}^\top r_{k-1}}{\|Ap_k\|^2},$$
$$x_k = x_{k-1} + \alpha_k p_k,$$
$$r_k = r_{k-1} - \alpha_k A^\top(Ap_k),$$
$$\beta_k = \frac{r_k^\top r_k}{r_{k-1}^\top r_{k-1}},$$
$$p_{k+1} = r_k + \beta_k p_k.$$
end for

Algorithm 2 CGLS for $A^\top Ax = A^\top b$

Input: A, b, x_0.
Define $d_0 = b - Ax_0$, $r_0 = A^\top d_0$, $p_1 = r_0$.
for $k = 1, 2, \ldots$ **do**
$$t_k = Ap_k,$$
$$\alpha_k = \frac{r_{k-1}^\top r_{k-1}}{\|t_k\|^2},$$
$$x_k = x_{k-1} + \alpha_k p_k,$$
$$d_k = d_{k-1} - \alpha_k t_k,$$
$$r_k = A^\top d_k,$$
$$\beta_k = \frac{r_k^\top r_k}{r_{k-1}^\top r_{k-1}},$$
$$p_{k+1} = r_k + \beta_k p_k.$$
end for


Paige, C. C. and Saunders, M. A., *LSQR: An Algorithm for Sparse Linear Equations and Sparse Least Squares*, ACM Trans. Math. Softw., 1982


Björck, A. and Elfving, T. and Strakos, Z. , *Stability of conjugate gradient and Lanczos methods for linear least squares problems*, SIMAX, 1998

CG for $A^T Ax = A^T b + c$

Initial rounding error due to the product $r_0 = A^T b + c - A^T Ax_0$:

$$\|\delta x\| \leq \kappa(A)^2 u \left(\frac{\|b\|}{\|A\|} + \frac{\|c\|}{\|A\|^2} \right).$$

This initial error cannot be canceled, and the best error bound we can hope for will include the term given above.
Optimal bound:

$$\|\delta x\| \leq \sqrt{\|\bar{M}\|} \|[A, b, c]\|_F u$$

If

$$\|b\|\|A\| + \|c\| >> \left[1 + \|r\| + 2\sqrt{\|c\|\|x\|} + \frac{1 + \|x\|}{\|A^\dagger\|} \right] \sqrt{\|A\|_F^2 + \|b\|^2 + \|c\|^2}$$

CG can be expected to produce less than optimal accuracy.

IDEA to design a stable method

- Extend the successful algorithmic procedures to the case $c \neq 0$
- Need to factorize matrix A in both the left and right hand sides

$$A^T(A^T x - b)$$

Two solution methods

We propose two iterative methods based on two different reformulations of the problem

Proposed methods (I) CGLSϵ

Given $\epsilon > 0$, let us then define

$$A_\epsilon = \begin{bmatrix} A \\ \epsilon c^\top \end{bmatrix}, \quad b_\epsilon = \begin{bmatrix} b \\ 1/\epsilon \end{bmatrix}.$$

We then consider the following linear least squares problem:

$$\min_x \|A_\epsilon x - b_\epsilon\|^2,$$

with normal equations

$$(A^\top A + \epsilon^2 cc^\top)x = A^\top b + c. \qquad (\text{SYS}\epsilon)$$

CGLSϵ solves SYSϵ with CGLS method

Lemma

Let x_ϵ be the solution of SYSϵ and x be the solution of SYS. Then, $\lim_{\epsilon \to 0} x_\epsilon = x$ and the relative norm of the error satisfies

$$\frac{\|x_\epsilon - x\|}{\|x\|} \leq \epsilon^2 \frac{\|c\|\|w\|}{1 + \epsilon^2 c^\top w}, \quad w = (A^\top A)^{-1} c.$$

Remarks (I)

- Will a really small ϵ may cause large errors in finite arithmetic?
- A perturbed solution $\tilde{x}_\epsilon = x_\epsilon + \delta x_\epsilon$ will be such that:

$$(A_\epsilon^\top A_\epsilon)(\delta x_\epsilon) = \delta(A_\epsilon^\top b_\epsilon). \quad |\delta(A_\epsilon^\top b_\epsilon)| \leq \gamma_{m+1}|A_\epsilon^\top||b_\epsilon|$$

This overestimates the error!

$$\mathrm{fl}(A_\epsilon^\top b_\epsilon) = \mathrm{fl}(A^\top b) + \mathrm{fl}\left(\epsilon c \frac{1}{\epsilon}\right) + \delta_s,$$

with δ_s error due to the summation.

- If $\epsilon = 2^i$ for $i \in \mathbb{Z}$, then $\mathrm{fl}\left(\epsilon\ c\ \frac{1}{\epsilon}\right) = c$. Then,

$$\mathrm{fl}\left(A_\epsilon^\top b_\epsilon\right) = A^\top b + c + \delta_p + \delta_s, \quad \text{with} \quad |\delta_s| \leq u|\mathrm{fl}(A^\top b) + c|,\ |\delta_p| \leq \gamma_m|A||b|,$$

and the bound does not depend on ϵ.

Remarks (II)

- What about the conditioning of the problem?
- Due to the presence of small ϵ in the right-hand side the residual will generally be really large.
- Standard conditioning analysis of least squares problems is not well-suited in this case
- We can show that the conditioning does not depend on $\|b_\epsilon - A_\epsilon x_\epsilon\|$, that will be really large, but rather on $\|r_\epsilon\| = \|b - Ax_\epsilon\|$, that will be indeed much smaller

Conditioning

- Let F_ϵ be the function that maps A, b, c to the solution x_ϵ of SYSϵ

$$F_\epsilon : \mathbb{R}^{m\times n} \times \mathbb{R}^m \times \mathbb{R}^n \to \mathbb{R}^n$$
$$(A, b, c) \longmapsto F_\epsilon(A, b, c) = (A_\epsilon^\top A_\epsilon)^{-1}(A^\top b + c),$$

  and let $r_\epsilon = b - Ax_\epsilon$.
- The absolute condition number of problem SYSϵ, with Euclidean norm on the solution and Frobenius norm on the data, is then given by:

$$\|F'_\epsilon(A, b, c)\|_{\text{op}} = \|[(r_\epsilon^\top \otimes (A_\epsilon^\top A_\epsilon)^{-1})L_T + x_\epsilon^\top \otimes (A_\epsilon^\top A_\epsilon)^{-1}A^\top, (A_\epsilon^\top A_\epsilon)^{-1}A^\top, (1 - 2\epsilon c^\top x_\epsilon)(A_\epsilon^\top A_\epsilon)^{-1}]\|.$$

- Computable formula: $\sqrt{\|\bar{M}_\epsilon\|}$, with

$$\bar{M}_\epsilon = ((1 - 2\epsilon c^\top x_\epsilon)^2 + \|r_\epsilon\|^2)(A_\epsilon^\top A_\epsilon)^{-2} + (1 + \|x_\epsilon\|^2)(A_\epsilon^\top A_\epsilon)^{-1}A^\top A(A_\epsilon^\top A_\epsilon)^{-1} - 2\,\text{sym}(B_\epsilon)$$

  with $B_\epsilon = (A_\epsilon^\top A_\epsilon)^{-1}A^\top r_\epsilon x_\epsilon^\top (A_\epsilon^\top A_\epsilon)^{-1}$ and $\text{sym}(B_\epsilon) = \frac{1}{2}(B_\epsilon + B_\epsilon^\top)$.

Proposed method (II) CGLSI

Given $\hat{I} \in \mathbb{R}^{(m+1)\times(m+1)}$, we define $\hat{A} \in \mathbb{R}^{(m+1)\times n}$ and $\hat{b} \in \mathbb{R}^{m+1}$ as:

$$\hat{A} = \begin{bmatrix} A \\ c^\top \end{bmatrix}, \qquad \hat{I} = \begin{bmatrix} I_m & 0 \\ 0 & 0 \end{bmatrix}, \qquad \hat{b} = \begin{bmatrix} b \\ 1 \end{bmatrix}.$$

We then reformulate SYS as:

$$\hat{A}^\top \hat{I} \hat{A} x = \hat{A}^\top \hat{b}$$

- Possible to factorize $\hat{A}^\top$ in both the right and the left-hand sides:
  - no need of recurring the residual $r = \hat{A}^\top(\hat{I}\hat{A}x - \hat{b})$ (simply update $\hat{d} = \hat{I}\hat{A}x - \hat{b}$ along the iterations and form r by multiplication with $\hat{A}^\top$)
  - computation of $p_k^\top A^\top A p_k$ as $\|\hat{I}\hat{A}p_k\|^2$

  We can therefore expect the same benefits of CGLS as compared to CG.

Algorithm

Algorithm 3 CGLSI for $A^\top Ax = A^\top b + c$

Input: $\hat{A}$, $\hat{b}$, x_0

Define $\hat{d}_0 = \hat{b} - \hat{A}x_0$, $r_0 = \hat{A}^\top(\hat{b} - \hat{A}x_0)$, $p_1 = r_0$.

for $k = 1, 2, \ldots$ **do**

$\hat{t}_k = \hat{I}\hat{A}p_k$,

$$\alpha_k = \frac{r_{k-1}^\top r_{k-1}}{\hat{t}_k^\top \hat{t}_k},$$

$x_k = x_{k-1} + \alpha_k p_k$,

$\hat{d}_k = \hat{d}_{k-1} - \alpha_k \hat{t}_k$,

$r_k = \hat{A}^\top \hat{d}_k$,

$$\beta_k = \frac{r_k^\top r_k}{r_{k-1}^\top r_{k-1}},$$

$p_{k+1} = r_k + \beta_k p_k$.

end for

First order approximation for the forward error

First order approximation for the forward error can be obtained as

$$\frac{\|x-\hat{x}\|}{\|x\|} \sim \frac{\kappa_{SYS}\|[A,b,c]\|_F}{\|x\|}u, \quad u \text{ machine precision}$$

We define the following error estimates:

$$\hat{E}_{CGLSI} := \frac{\sqrt{\|\bar{M}\|}\|[A,b,c]\|_F}{\|x\|}u,$$

$$\hat{E}_{CGLS_\epsilon} := \epsilon^2\frac{\|c\|\|w\|}{1+\epsilon^2 c^\top w} + \frac{\sqrt{\|\bar{M}_\epsilon\|}\|[A,b,c]\|_F}{\|x\|}u\left\|I_n - \frac{\epsilon^2 wc^\top}{1+\epsilon^2 c^\top w}\right\|,$$

u being the machine precision.

- CGLSϵ: the error on the computed solution $\hat{x}_\epsilon$ depends on two terms:

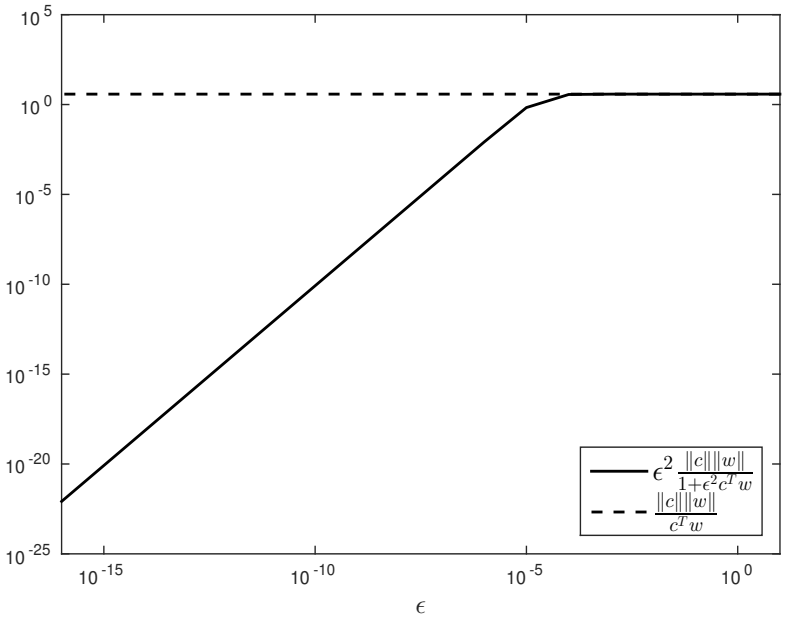
$$\frac{\|x-\hat{x}_\epsilon\|}{\|x\|} \leq \frac{\|x-x_\epsilon\|}{\|x\|} + \frac{\|x_\epsilon-\hat{x}_\epsilon\|}{\|x\|} = \frac{\|x-x_\epsilon\|}{\|x\|} + \frac{\|x_\epsilon-\hat{x}_\epsilon\|}{\|x_\epsilon\|}\frac{\|x_\epsilon\|}{\|x\|}.$$

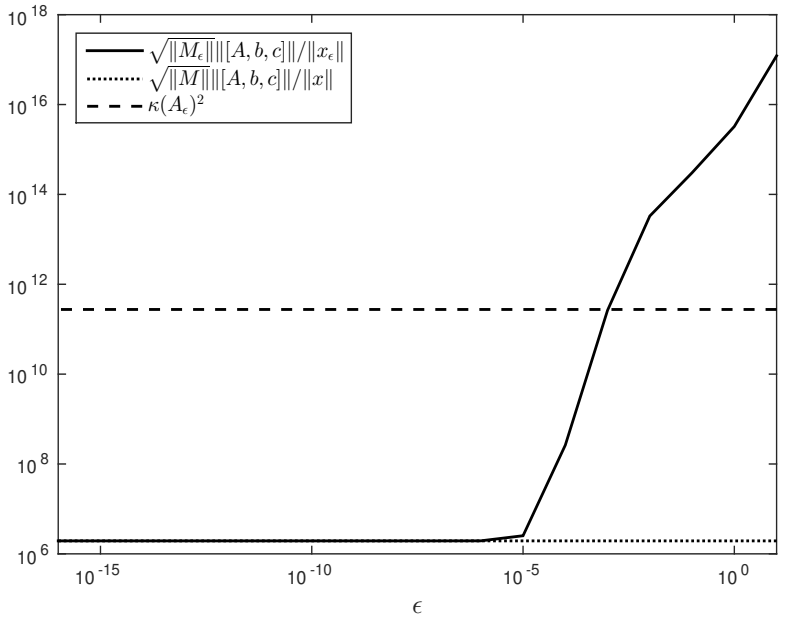
NUMERICAL TESTS

Numerical tests: setting

- All the numerical methods have been implemented in Matlab
- $A \in \mathbb{R}^{m\times n}$, $A = U\Sigma V^\top$, where U and V from *gallery('orthog',m/n,j)*, $j = 1, \ldots, 6$.
- C1 : $\Sigma_{ii} = a^{-i}$, for $a > 0$,
  C2 : $\Sigma_{ii} = u_i$, $u = \mathrm{linspace}(dw, up, n)$, with $\mathrm{dw}, \mathrm{up} > 0$,
  for $i = 1, \ldots, n$.
- Matrix dimensions: $m = 40$ and $n = 20$ for the tests and $m = 100$, $n = 50$ for performance profiles
- Performance profiles: 40 matrices, with condition number between 1 and 10^{10}. The optimality measure is $\dfrac{\|x - \hat{x}\|}{\|x\|}$, with x the exact solution $(x = (n-1:-1:0))$. A simulation is considered unsuccessful if the relative solution accuracy is larger than 10^{-2}.

How to choose ϵ?





$$\frac{\|x-\hat{x}_\epsilon\|}{\|x\|} \leq \frac{\|x-x_\epsilon\|}{\|x\|} + \frac{\|x_\epsilon-\hat{x}_\epsilon\|}{\|x\|} = \frac{\|x-x_\epsilon\|}{\|x\|} + \frac{\|x_\epsilon-\hat{x}_\epsilon\|}{\|x_\epsilon\|}\frac{\|x_\epsilon\|}{\|x\|}.$$

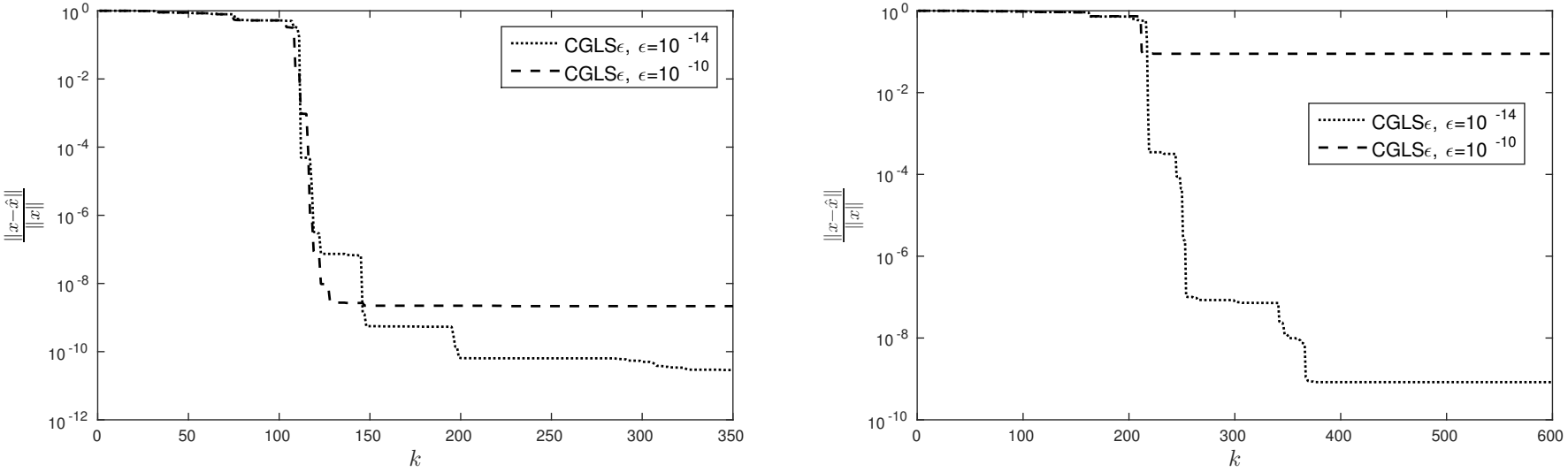
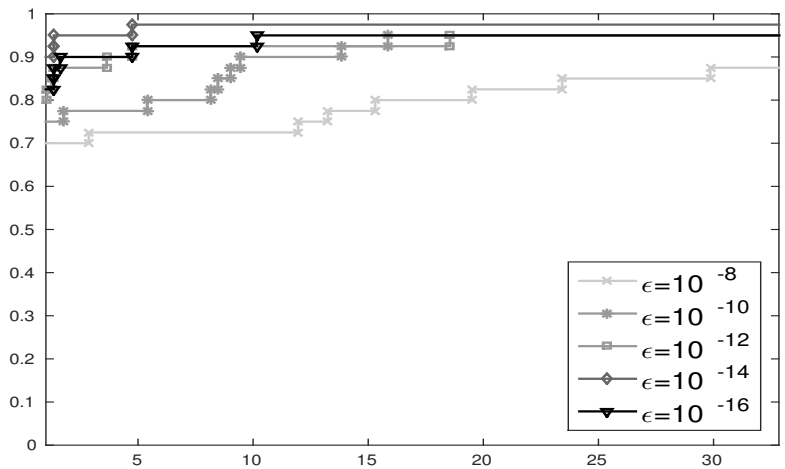


Figure: Left: right hand side of small norm, Right: right hand side of large norm



Comparison with CG

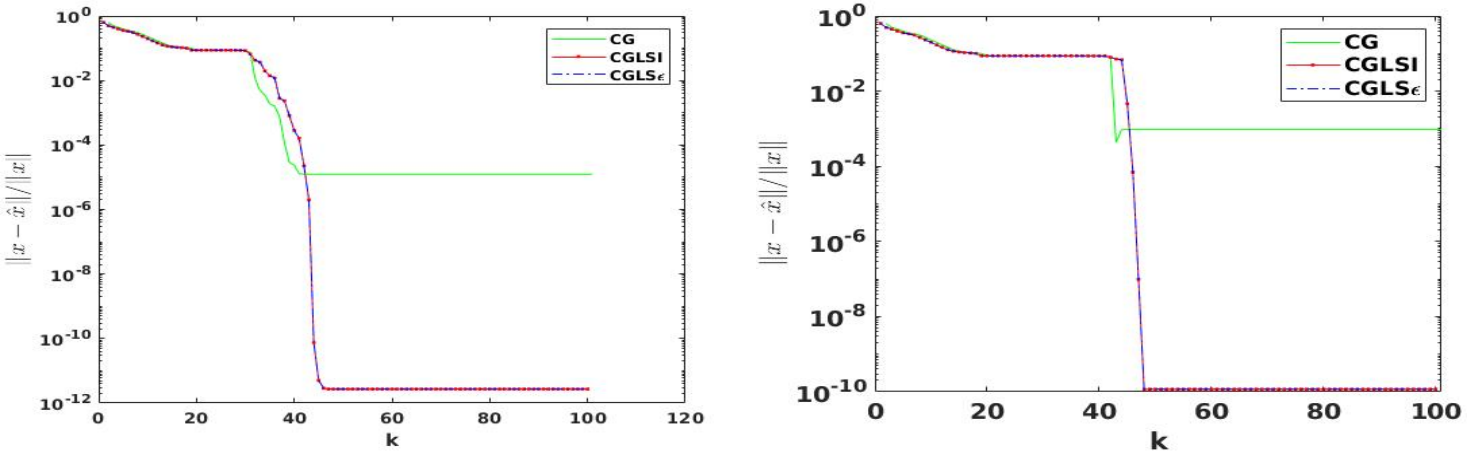
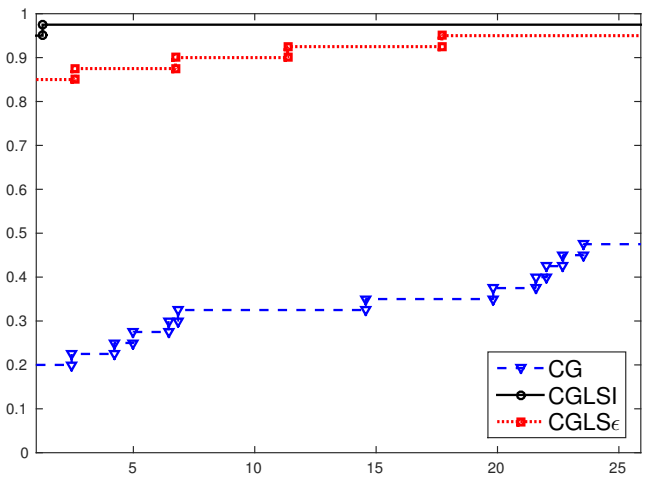


Figure: Left: $\kappa(A) = 10^5$, $\kappa(\hat{A}) = 10^5$. Right: $\kappa(A) = 10^7$, $\kappa(\hat{A}) = 10^{10}$.



Performance of CGLS*I* and CGLSϵ is comparable but

- CGLS*I* is parameter free
- CGLS*I* is less sensible to the right hand side

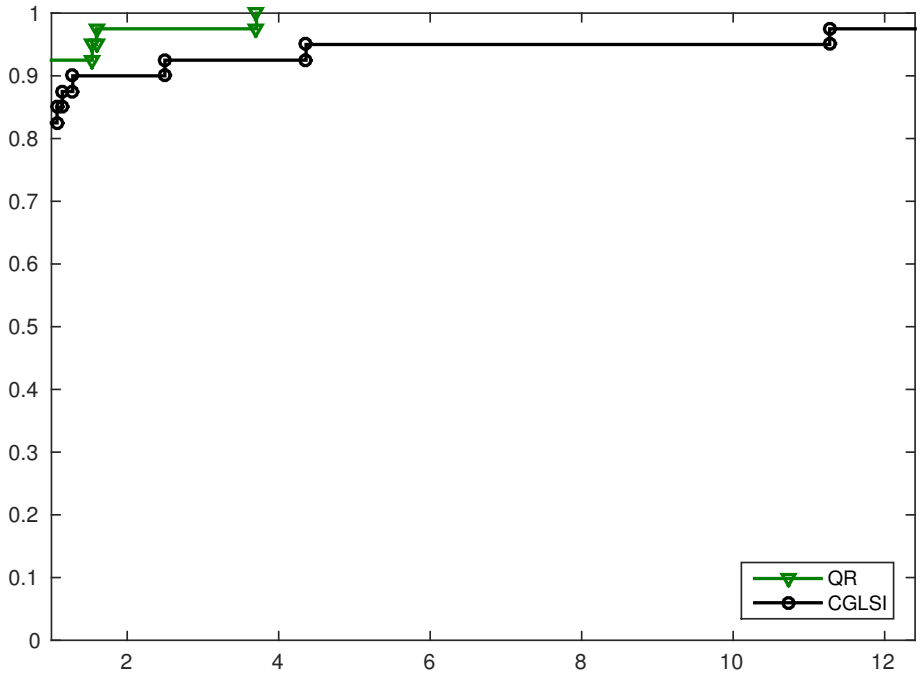
Much better performance than CG

Validation of error bounds

Problem	$\kappa(A)^2 u$	E_{CGLSI}	$\hat{E}_{CGLSI}$	$E_{CGLS\epsilon}$	$\hat{E}_{CGLS\epsilon}$
$a = 2.0$	10^{-10}	10^{-11}	10^{-8}	10^{-9}	10^{-9}
$a = 2.5$	10^{-2}	10^{-9}	10^{-6}	10^{-9}	10^{-6}
$a = 1.5$	10^{-10}	10^{-13}	10^{-11}	10^{-13}	10^{-11}
$a = 1.3$	10^{-12}	10^{-13}	10^{-13}	10^{-14}	10^{-13}
$a = 1.1$	10^{-14}	10^{-14}	10^{-13}	10^{-14}	10^{-14}
$a = 0.7$	10^{-10}	10^{-12}	10^{-12}	10^{-12}	10^{-12}
$\mathrm{up} = 1$	10^{-2}	10^{-9}	10^{-7}	10^{-8}	10^{-8}
$\mathrm{up} = 1$	10^{-2}	10^{-8}	10^{-4}	10^{-4}	10^{-6}
$a = 1.5$	10^{-10}	10^{-13}	10^{-12}	10^{-10}	10^{-10}

- Better performance that standard CG, both in terms of accuracy and of rate of convergence.
- The error bounds much better predict forward errors than classical bounds.

Comparison with QR method



- Propose method can compare with direct methods in terms of solution accuracy

THANK YOU FOR YOUR ATTENTION

Calandra, H., Gratton, S., Riccietti, E., Vasseur, X., On the solution of systems of the form $A^T Ax = A^T b + c$, In preparation

Effect of large right-hand sides

Let us assume to apply CG to SYS and CGLS to SYSϵ. We would respectively compute:

$$\alpha_1 = \frac{\|r_0\|^2}{p_1^\top A^\top A p_1} = \frac{\|A^\top b + c\|^2}{\|A(A^\top b + c)\|^2}, \qquad x_1 = \alpha_1 (A^\top b + c) = \alpha_1 p_1,$$

and

$$\alpha_1(\epsilon) = \frac{\|A^\top b + c\|^2}{\|A(A^\top b + c)\|^2 + \epsilon \|c^\top (A^\top b + c)\|}, \qquad x_1(\epsilon) = \alpha_1(\epsilon) p_1(\epsilon) = \alpha_1(\epsilon) p_1.$$

Notice that if ϵ tends to zero, so does the term $\epsilon\|c^\top(A^\top b + c)\|$ in the denominator of $\alpha_1(\epsilon)$. Consequently $\alpha_1(\epsilon)$ tends toward α_1 and $x_1(\epsilon)$ tends toward x_1. If ϵ has to be fixed, its value should be small enough to let $\epsilon\|c^\top(A^\top b + c)\|$ be small compared to $\|A(A^\top b + c)\|^2$, otherwise the found approximation will be close to a solution of SYSϵ rather than to one of SYS. This choice is then particularly difficult when $\|A^\top b + c\|$ is large.