Science & Technology Facilities Council

Linear regression models and stopping criteria for Krylov methods

Mario Arioli and Serge Gratton (CERFACS,N7)

STFC Rutherford Appleton Laboratory

Birmingham, BAMC, 11 April 2011

Overview of talk

- Motivations and Modelling
- Perturbation Theory
- Statistical tests
- Least-squares and deterministic approach
- Stopping criteria
- Numerical examples

Linear regression problem

Let $A \in \mathrm{R}^{m\times n}$, $m \geq n$, with $\mathrm{Rank}(A) = n$. We consider the linear regression model

$$\mathbf{y} = A\mathfrak{x} + \mathbf{e}, \tag{1}$$

where $E[\mathbf{e}] = 0$ and $V[\mathbf{e}] = \sigma^2 I_m$. We point out that A defines a given model and $\mathfrak{x}$ is an unknown deterministic value. The minimum-variance unbiased (MVU) estimator of $\mathfrak{x}$ is related to $\mathbf{y}$ by the Gauss-Markov theorem.

Gauss-Markov Theorem

For the linear model (1) the minimum-variance unbiased estimator of $\mathfrak{x}$ is given by

$$\mathbf{x}^* = (A^T A)^{-1} A^T \mathbf{y}.$$

The variance $V[\mathbf{x}^*] = \sigma^2 (A^T A)^{-1}$. If $\mathbf{e} \sim \mathcal{N}\left(0, \sigma^2 I_m\right)$, and if we set

$$\mathbf{s^2} = \frac{1}{m-n} ||\mathbf{r}||_2^2, \qquad \mathbf{r} = \mathbf{y} - A\mathbf{x}^*$$

we have for our estimator of $\mathfrak{x}$ and for $\mathbf{s^2}$, our estimator for σ^2,

$$\mathbf{x}^* \sim \mathcal{N}\left(\mathfrak{x}, \sigma^2 (A^T A)^{-1}\right), \quad \mathbf{s^2} \sim \frac{\sigma^2}{m-n} \chi^2(m-n).$$

Moreover, the predicted value $\hat{\mathbf{y}} = A\mathbf{x}^*$ and the residual $\mathbf{r}$ are

$\hat{\mathbf{y}} \sim \mathcal{N}\left(A\mathfrak{x}, \sigma^2 A (A^T A)^{-1} A^T\right)$ and $\mathbf{r} \sim \mathcal{N}\left(0, \sigma^2 (I - A(A^T A)^{-1} A^T)\right)$.

Perturbation theory

What we mean with PERTURBATION ?

Perturbation theory

What we mean with PERTURBATION ?

Let $\delta\hat{\mathbf{y}}$ be a stochastic variable such that

$$\delta\hat{\mathbf{y}} \sim \mathcal{N}\left(0, \tau^2 A(A^T A)^{-1} A^T\right).$$

Under the Hypotheses of Gauss-Markov, and assuming that $\hat{\mathbf{y}}$ and $\delta\hat{\mathbf{y}}$ are independently distributed, we have

$$\hat{\mathbf{y}} + \delta\hat{\mathbf{y}} \sim \mathcal{N}\left(A\mathfrak{x}, (\tau^2 + \sigma^2) A(A^T A)^{-1} A^T\right).$$

Moreover, we have that

$$||\delta\hat{\mathbf{y}}||_2^2 \sim \tau^2 \chi^2(n).$$

Perturbation theory

Let $\delta\hat{\mathbf{y}} \sim \mathcal{N}\left(0, \tau^2 A(A^T A)^{-1} A^T\right)$.
Under the hypotheses of Gauss-Markov and assuming that $\hat{\mathbf{y}}$ *and* $\delta\hat{\mathbf{y}}$ *be uncorrelated, there exists*

$$\delta\mathbf{x}^* \sim \mathcal{N}(0, \tau^2 (A^T A)^{-1}), \quad \delta\mathbf{y} \sim \mathcal{N}(0, \tau^2 I_m),$$

such that

1. $\hat{\mathbf{y}} + \delta\hat{\mathbf{y}} = A(\mathbf{x}^* + \delta\mathbf{x}^*)$,
2. $\mathbf{x}^* + \delta\mathbf{x}^*$ *is the minimum-variance unbiased estimator of* $\mathfrak{x}$ *for the linear regression problem:*
   $\mathbf{y} + \delta\mathbf{y} = A\mathfrak{x} + \bar{\mathbf{e}}, \quad \bar{\mathbf{e}} \sim \mathcal{N}(0, (\sigma^2 + \tau^2) I_m)$,
3. *and* $\bar{\mathbf{s}}^2 = \frac{1}{m-n}\|\mathbf{y} + \delta\mathbf{y} - A(\mathbf{x}^* + \delta\mathbf{x}^*)\|_2^2$, *is the estimator for* $\rho^2 = \sigma^2 + \tau^2$ *with* $\bar{\mathbf{s}}^2 \sim \frac{\sigma^2+\tau^2}{m-n}\chi^2(m-n)$.

Least-squares problem

The minimum-variance unbiased (MVU) estimators of $\mathfrak{x}$ and σ^2 are closely related to the solution of the least-squares problem (LSP),

$$\min_x ||y - Ax||_2^2 \tag{2}$$

where y is a realization of $\mathbf{y}$. The least-squares problem (LSP) has the unique solution

$$x^* = (A^T A)^{-1} A^T y,$$

and the corresponding minimum value is achieved by the $||r||_2$

$$r = y - Ax^* = (I - P)y, \qquad \left(I - P = I - A(A^T A)^{-1} A^T\right).$$

Science & Technology Facilities Council

Least-squares problem

The minimum-variance unbiased (MVU) estimators of $\mathfrak{x}$ and σ^2 are closely related to the solution of the least-squares problem (LSP),

$$\min_x ||y - Ax||_2^2 \tag{2}$$

where y is a realization of $\mathbf{y}$. The least-squares problem (LSP) has the unique solution

$$x^* = (A^T A)^{-1} A^T y,$$

and the corresponding minimum value is achieved by the $||r||_2$

$$r = y - Ax^* = (I - P)y, \qquad \Big(I - P = I - A(A^T A)^{-1} A^T\Big).$$

We remark here that the solution of LSP is deterministic and, therefore, supplies only a realization of the MVU $\mathbf{x}^*$ and of $\mathbf{s^2}$ the corresponding estimator of σ^2.

Science & Technology Facilities Council

Least-squares problem

The vector x^* is also the solution of the normal equations, i.e. it is the unique stationary point of $||y - Ax||_2^2$:

$$A^T A x^* = A^T y. \tag{3}$$

We will denote in the following by

$$R(x) = A^T (y - Ax)$$

the residual of (3).

Least-squares problem

The vector x^* is also the solution of the normal equations, i.e. it is the unique stationary point of $||y - Ax||_2^2$:

$$A^T A x^* = A^T y. \tag{3}$$

We will denote in the following by

$$R(x) = A^T(y - Ax)$$

the residual of (3). Moreover, we have

$$||y||_2^2 - ||x^*||_{A^T A}^2 = ||\big(I - P\big)y||_2^2 = ||y - Ax^*||_2^2.$$

Least-squares problem

Given $\tilde{x}$ as an approximation of x^*,

$$\delta y = -A(A^T A)^{-1} R(\tilde{x})$$

is the minimum norm solution of

$$\min_w ||w||_2^2 \qquad \text{such that} \qquad A^T A \tilde{x} = A^T (y + w).$$

Moreover, using $R(\tilde{x}) = A^T(y - A\tilde{x}) = A^T A(x^* - \tilde{x})$, we have

$$||\delta y||_2^2 = ||R(\tilde{x})||^2_{(A^T A)^{-1}} = ||x^* - \tilde{x}||^2_{A^T A}.$$

A statistical point of view

How, we can link the deterministic theory for the least-squares perturbation to the Stochastic Perturbation theory?

A statistical point of view

If δy is a realization of a stochastic variable $\delta \mathbf{y}$ then $\|\delta y\|_2^2$ is a realization of $\|\delta \hat{\mathbf{y}}\|_2^2 \sim \tau^2\chi^2(n)$.
Therefore, we consider that δy is a sample of the stochastic variable $\delta \hat{\mathbf{y}}$ if for some small enough η,

$$\mathrm{Probability}(\|\delta \hat{\mathbf{y}}\|_2^2 \geq \|\delta y\|_2^2) \geq 1-\eta,$$

where we assume that the random variable $\frac{\|\delta \hat{\mathbf{y}}\|_2^2}{\tau^2}$ follows a centered χ^2 distribution with n degrees of freedom. Thus, we can formulate our criterion as

$$p_\chi\left(\frac{\|\delta y\|_2^2}{\tau^2}, n\right) \equiv \mathrm{Probability}\left(\frac{\|\delta \hat{\mathbf{y}}\|_2^2}{\tau^2} \leq \frac{\|\delta y\|_2^2}{\tau^2}\right) \leq \eta,$$

where, $p_\chi(., n)$ is the cumulative distribution function of the χ^2 distribution.

Preconditioned Conjugate Gradient for Normal equations 1

At each step k the conjugate gradient method minimizes the energy norm of the error $\delta x^{(k)} = x^* - x^{(k)}$ on a Krylov space $x^{(0)} + \mathcal{K}_k$:

$$\min_{x^{(k)} \in \, x^{(0)} + \mathcal{K}_k} \delta x^{(k)T} A^T A \delta x^{(k)},$$

where $\mathcal{K}_k = \mathrm{Span}(A^T y, (A^T A) A^T y, \ldots, (A^T A)^{k-1} A^T y)$. Let $R^{(k)} = A^T \left(y - A x^{(k)}\right)$ denote the normal equations residual at step k. Moreover, using $A^T A \delta x^{(k)} = A^T (y - A x^{(k)}) = R^{(k)}$, the value $\|\delta x^{(k)}\|_{A^T A}$ is equal to the dual norm of the residual $\|R^{(k)}\|_{(A^T A)^{-1}}$.

Preconditioned Conjugate Gradient for Normal equations 2

$$
\begin{aligned}
x^{(k)} &= x^{(k-1)} + \alpha_{k-1} q^{(k-1)}, \quad \alpha_{k-1} = \frac{R^{(k-1)T} R^{(k-1)}}{q^{(k-1)T} A^T A q^{(k-1)}}, \\
R^{(k)} &= R^{(k-1)} - \alpha_{k-1} A^T A q^{(k-1)}, \\
q^{(k)} &= R^{(k)} + \beta_{k-1} q^{(k-1)}, \quad \beta_{k-1} = \frac{R^{(k)T} R^{(k)}}{R^{(k-1)T} R^{(k-1)}}.
\end{aligned}
$$

The quantity α_{k-1} gives the step-size on the direction $q^{(k-1)}$ during the conjugate gradient algorithm.

Preconditioned Conjugate Gradient for Normal equations 3

In exact arithmetic, we have that the energy norm of the error $\delta x^{(k)} = x^* - x^{(k)}$ is

$$\|\delta x^{(k)}\|^2_{A^T A} = e^{(k)}_{A^T A} = \sum_{j=k}^{n-1} \alpha_j R^{(j)\,T} R^{(j)},$$

and the energy norm of $x^* - x^{(0)}$ is

$$\|x^* - x^{(0)}\|^2_{A^T A} = \sum_{j=0}^{n-1} \alpha_j R^{(j)\,T} R^{(j)}.$$

Science & Technology Facilities Council

Preconditioned Conjugate Gradient for Normal equations 3

In exact arithmetic, we have that the energy norm of the error $\delta x^{(k)} = x^* - x^{(k)}$ is

$$\|\delta x^{(k)}\|^2_{A^TA} = e^{(k)}_{A^TA} = \sum_{j=k}^{n-1} \alpha_j R^{(j)T} R^{(j)},$$

and the energy norm of $x^* - x^{(0)}$ is

$$\|x^* - x^{(0)}\|^2_{A^TA} = \sum_{j=0}^{n-1} \alpha_j R^{(j)T} R^{(j)}.$$

Under the assumption that $e^{(k+d)}_{A^TA} << e^{(k)}_{A^TA}$, where the integer d denotes a suitable delay, the Hestenes and Stiefel estimate ξ_k of the energy-norm of the error will then be computed by the formula

$$\xi_k = \sum_{j=k}^{k+d-1} \alpha_j R^{(j)T} R^{(j)}.$$

Science & Technology Facilities Council

Preconditioned Conjugate Gradient for Normal equations 4

Finally,

$$\|x^* - x^{(0)}\|^2_{A^TA} \geq \sum_{j=0}^{k-1} \alpha_j R^{(j)T} R^{(j)} = \nu_k.$$

Therefore,

$$\|y - Ax^*\|_2^2 = \|y - Ax^{(0)} - A(x^* - x^{(0)}\|_2^2 \leq \|y - Ax^{(0)}\|_2^2 - \nu_k.$$

Introducing a preconditioner, we want to speed up the convergence rate of the conjugate gradient method but this will change the matrix and, therefore, the energy norm. However, we still want to estimate $e^{(k)}_{A^TA}$. It is proved that the energy norm of the preconditioned problem is equal to $e^{(k)}_{A^TA}$(Arioli, Meurant). Moreover, when finite precision arithmetic is used, the Hestenes and Stiefel method for computing the error norm is numerically reliable (Strakos Tichy).

Preconditioned Conjugate Gradient for Normal equations 5

PCGLS algorithm Given an initial guess $x^{(0)}$, compute $r^{(0)} = (y - Ax^{(0)})$, $R^{(0)} = A^T r^{(0)}$, and solve $Mz^{(0)} = R^{(0)}$. Set $q^{(0)} = z^{(0)}$, $\beta_0 = 0$, $\nu_0 = 0$, $\chi_1 = R^{(0)T} z^{(0)}$, and $\xi_{-d} = \infty$. Set $k = 0$

while $\mathbf{z}(\xi_{k-d}, \|r^{(0)}\|_2, \nu_k, \tau^2, \sigma^2)) > \eta$ **do**
- $k = k + 1$;
- $p = Aq^{(k-1)}$;
- $\alpha_{k-1} = \chi_k / ||p||_2^2$;
- $\psi_k = \alpha_{k-1}\chi_k$; $\nu_k = \nu_{k-1} + \psi_k$;
- $x^{(k)} = x^{(k-1)} + \alpha_{k-1} q^{(k-1)}$;
- $R^{(k)} = R^{(k-1)} - \alpha_{k-1} A^T q^{(k-1)}$;
- Solve $Mz^{(k)} = R^{(k)}$;
- $\chi_{k+1} = R^{(k)T} z^{(k)}$;
- $\beta_k = \chi_{k+1}/\chi_k$;
- $q^{(k)} = z^{(k)} + \beta_k q^{(k-1)}$;
- **if** $k > d$ **then**

$$\xi_{k-d} = \sum_{j=k-d+1}^{k} \psi_j;$$

- **else**
  - $\xi_{k-d} = \infty$;
- **endif**

end while.

Science & Technology Facilities Council

Stopping criteria η and τ^2

We stop when $\mathbf{z}(\xi_{k-d}, \|r^{(0)}\|_2, \nu_k, \tau^2, \sigma^2)) \leq \eta$

Stopping criteria η and τ^2

We stop when $\mathbf{z}(\xi_{k-d}, \|r^{(0)}\|_2, \nu_k, \tau^2, \sigma^2)) \leq \eta$

1. η is a probability (in our experiments $\eta = 10^{-8}$ but $\eta = 10^{-3}$ can realistic)

Stopping criteria η and τ^2

We stop when $\mathbf{z}(\xi_{k-d}, \|r^{(0)}\|_2, \nu_k, \tau^2, \sigma^2)) \leq \eta$

1. η is a probability (in our experiments $\eta = 10^{-8}$ but $\eta = 10^{-3}$ can realistic)
2. $\tau^2 = \sigma^2$ can be fragile and it is better to impose $\tau^2 = 0.01\sigma^2$

Stopping criteria η and τ^2

We stop when $\mathbf{z}(\xi_{k-d}, \|r^{(0)}\|_2, \nu_k, \tau^2, \sigma^2)) \leq \eta$

1. η is a probability (in our experiments $\eta = 10^{-8}$ but $\eta = 10^{-3}$ can realistic)
2. $\tau^2 = \sigma^2$ can be fragile and it is better to impose $\tau^2 = 0.01\sigma^2$
3. ξ_k depend on d the delay. When the preconditioned problem is still ill conditioned we must choose $d = 20$ for reliability.
4. p_χ is cheap to compute (in our experiments we use the Matlab functions `fcdf`, and chis_cdf)

Stopping criteria summary

1. $\mu_1 = p_{FS}\left(\left(\frac{m-n}{n-k}\right)\frac{\xi_k}{||y||_2^2-\nu_k}, n-k, m-n\right)$ (F-test)
2. $\mu_2 = p_\chi\left(\frac{\xi_k}{\tau^2}, n\right)\ \tau^2 = 0.1^p\sigma^2,\ p = 0, 2,$
3. $\mu_3 = p_\chi\left(\frac{(m-n)\xi_k}{\tau_k^2}, n\right),\ \tau_k^2 = 0.1^p(\|y\|_2^2 - \nu_k),\ p = 0, 2.$

Science & Technology Facilities Council

Test problems (Paige-Saunders)

id	m	n	nnz	$\kappa(A)$
`well1033`	1033	320	4732	$1.6\ 10^2$
`illc1033`	10033	320	4719	$1.8\ 10^4$
`well1850`	1850	712	8755	$1.1\ 10^2$
`illc1850`	1850	712	8636	$1.4\ 10^4$

Dimensions, number of non zeros and $\kappa(A) = ||A^+||_2||A||_2$ for Paige-Saunders tests.

In all the problems the standard deviation has been normalized

Science & Technology Facilities Council

Numerical tests ($\tau^2 = \sigma^2$)

id	d	IC(1e-2)		
		μ_1	μ_2	μ_3
Well1033	10	28 (1.1e-03 , 1.061)	28 (1.1e-03 , 1.061)	28 (1.1e-03 , 1.061)
	20	28 (1.1e-03 , 1.061)	28 (1.1e-03 , 1.061)	28 (1.1e-03 , 1.061)
	30	28 (1.1e-03 , 1.061)	28 (1.1e-03 , 1.061)	28 (1.1e-03 , 1.061)
Well1850	10	22 (2.0e-03 , 1.048)	22 (2.0e-03 , 1.048)	22 (2.0e-03 , 1.048)
	20	22 (2.0e-03 , 1.048)	22 (2.0e-03 , 1.048)	22 (2.0e-03 , 1.048)
	30	22 (2.0e-03 , 1.048)	22 (2.0e-03 , 1.048)	22 (2.0e-03 , 1.048)
illc1033	10	124 (4.9e-03 , 1.886)	124 (4.9e-03 , 1.886)	59 (2.0e-02 , 6.742)
	20	157 (6.3e-04 , 1.021)	156 (1.3e-03 , 1.089)	153 (1.8e-03 , 1.160)
	30	157 (6.3e-04 , 1.021)	156 (1.3e-03 , 1.089)	153 (1.8e-03 , 1.160)
illc1850	10	62 (3.1e-02 , 5.010)	120 (7.9e-03 , 1.591)	59 (3.6e-02 , 5.770)
	20	140 (4.2e-03 , 1.198)	143 (3.5e-03 , 1.144)	93 (1.3e-02 , 2.283)
	30	143 (3.5e-03 , 1.144)	145 (3.0e-03 , 1.106)	132 (5.7e-03 , 1.343)

Number of iterations for each stopping criterion $\eta = 10^{-8}$ and $\tau^2 = \sigma^2$. In bracket, the value of $||x^* - \tilde{x}||_{A^T A}/||x^*||_{A^T A}$ and of the standard deviation of $r = b - A\tilde{x}$ at the corresponding iteration.

Numerical tests ($\tau^2 = \sigma^2$)

		IC(1e-3)		
Well1033	10	9 (7.8e-04 , 1.032)	9 (7.8e-04 , 1.032)	9 (7.8e-04 , 1.032)
	20	9 (7.8e-04 , 1.032)	9 (7.8e-04 , 1.032)	9 (7.8e-04 , 1.032)
	30	9 (7.8e-04 , 1.032)	9 (7.8e-04 , 1.032)	9 (7.8e-04 , 1.032)
Well1850	10	7 (1.2e-03 , 1.017)	7 (1.2e-03 , 1.017)	7 (1.2e-03 , 1.017)
	20	7 (1.2e-03 , 1.017)	7 (1.2e-03 , 1.017)	7 (1.2e-03 , 1.017)
	30	7 (1.2e-03 , 1.017)	7 (1.2e-03 , 1.017)	7 (1.2e-03 , 1.017)
illc1033	10	59 (1.4e-03 , 1.106)	59 (1.4e-03 , 1.106)	57 (1.8e-03 , 1.166)
	20	63 (1.3e-03 , 1.082)	63 (1.3e-03 , 1.082)	58 (1.5e-03 , 1.122)
	30	63 (1.3e-03 , 1.082)	63 (1.3e-03 , 1.082)	58 (1.5e-03 , 1.122)
illc1850	10	36 (4.1e-03 , 1.186)	37 (2.7e-03 , 1.085)	36 (4.1e-03 , 1.186)
	20	37 (2.7e-03 , 1.085)	37 (2.7e-03 , 1.085)	36 (4.1e-03 , 1.186)
	30	37 (2.7e-03 , 1.085)	37 (2.7e-03 , 1.085)	36 (4.1e-03 , 1.186)

Number of iterations for each stopping criterion $\eta = 10^{-8}$ and $\tau^2 = \sigma^2$. In bracket, the value of $||x^* - \tilde{x}||_{A^TA}/||x^*||_{A^TA}$ and of the standard deviation of $r = b - A\tilde{x}$ at the corresponding iteration.

Numerical tests ($\tau^2 = \sigma^2/100$)

id	d	IC(1e-2)		
		μ_1	μ_2	μ_3
Well1033	10	28 (1.1e-03 , 1.061)	33 (5.0e-05 , 1.000)	33 (5.0e-05 , 1.000)
	20	28 (1.1e-03 , 1.061)	33 (5.0e-05 , 1.000)	33 (5.0e-05 , 1.000)
	30	28 (1.1e-03 , 1.061)	33 (5.0e-05 , 1.000)	33 (5.0e-05 , 1.000)
Well1850	10	22 (2.0e-03 , 1.048)	25 (1.7e-04 , 1.000)	24 (3.8e-04 , 1.002)
	20	22 (2.0e-03 , 1.048)	25 (1.7e-04 , 1.000)	24 (3.8e-04 , 1.002)
	30	22 (2.0e-03 , 1.048)	25 (1.7e-04 , 1.000)	24 (3.8e-04 , 1.002)
illc1033	10	124 (4.9e-03 , 1.886)	164 (6.6e-05 , 1.000)	138 (2.1e-03 , 1.209)
	20	157 (6.3e-04 , 1.021)	164 (6.6e-05 , 1.000)	163 (1.5e-04 , 1.001)
	30	157 (6.3e-04 , 1.021)	164 (6.6e-05 , 1.000)	163 (1.5e-04 , 1.001)
illc1850	10	62 (3.1e-02 , 5.010)	178 (6.1e-04 , 1.005)	178 (6.1e-04 , 1.005)
	20	140 (4.2e-03 , 1.198)	196 (2.6e-04 , 1.001)	195 (3.6e-04 , 1.002)
	30	143 (3.5e-03 , 1.144)	196 (2.6e-04 , 1.001)	195 (3.6e-04 , 1.002)

Number of iterations for each stopping criterion $\eta = 10^{-8}$ and $\tau^2 = \sigma^2/100$. In bracket, the value of $||x^* - \tilde{x}||_{A^TA}/||x^*||_{A^TA}$ and of the standard deviation of $r = b - A\tilde{x}$ at the corresponding iteration.

Science & Technology Facilities Council

Numerical tests ($\tau^2 = \sigma^2/100$)

		IC(1e-3)		
Well1033	10	9 (7.8e-04 , 1.032)	13 (6.3e-05 , 1.000)	13 (6.3e-05 , 1.000)
	20	9 (7.8e-04 , 1.032)	13 (6.3e-05 , 1.000)	13 (6.3e-05 , 1.000)
	30	9 (7.8e-04 , 1.032)	13 (6.3e-05 , 1.000)	13 (6.3e-05 , 1.000)
Well1850	10	7 (1.2e-03 , 1.017)	8 (2.2e-04 , 1.001)	8 (2.2e-04 , 1.001)
	20	7 (1.2e-03 , 1.017)	8 (2.2e-04 , 1.001)	8 (2.2e-04 , 1.001)
	30	7 (1.2e-03 , 1.017)	8 (2.2e-04 , 1.001)	8 (2.2e-04 , 1.001)
illc1033	10	59 (1.4e-03 , 1.106)	83 (8.3e-05 , 1.000)	83 (8.3e-05 , 1.000)
	20	63 (1.3e-03 , 1.082)	83 (8.3e-05 , 1.000)	83 (8.3e-05 , 1.000)
	30	63 (1.3e-03 , 1.082)	83 (8.3e-05 , 1.000)	83 (8.3e-05 , 1.000)
illc1850	10	36 (4.1e-03 , 1.186)	55 (3.2e-04 , 1.001)	51 (4.1e-04 , 1.002)
	20	37 (2.7e-03 , 1.085)	55 (3.2e-04 , 1.001)	51 (4.1e-04 , 1.002)
	30	37 (2.7e-03 , 1.085)	55 (3.2e-04 , 1.001)	51 (4.1e-04 , 1.002)

Number of iterations for each stopping criterion $\eta = 10^{-8}$ and $\tau^2 = \sigma^2/100$. In bracket, the value of $||x^* - \tilde{x}||_{A^TA}/||x^*||_{A^TA}$ and of the standard deviation of $r = b - A\tilde{x}$ at the corresponding iteration.

Science & Technology Facilities Council

Data Assimilation

Data assimilation purpose is to reconstruct the initial conditions at $t = 0$ of a dynamical system based on knowledge of the system's evolution laws and on observations of the state at times t_i.

$$\dot{u} = f(t, u) \quad u(t) = M(t)u_0$$

Assume that the system state is observed (possibly only in parts) at times $\{t_i\}_{i=0}^{N}$, yielding observation vectors $\{y_i\}_{i=0}^{N}$, whose model is given by

$$y_i = Hu(t_i) + \epsilon, \quad \epsilon \sim \mathcal{N}\big(0, R_i = \sigma^2 I\big).$$

Find u_0 which minimizes

$$\frac{1}{2}\sum_{i=0}^{N} \|HM(t_i)u_0 - y_i\|^2_{R_i^{-1}}.$$

Science & Technology Facilities Council

Data Assimilation

Heat Equation model

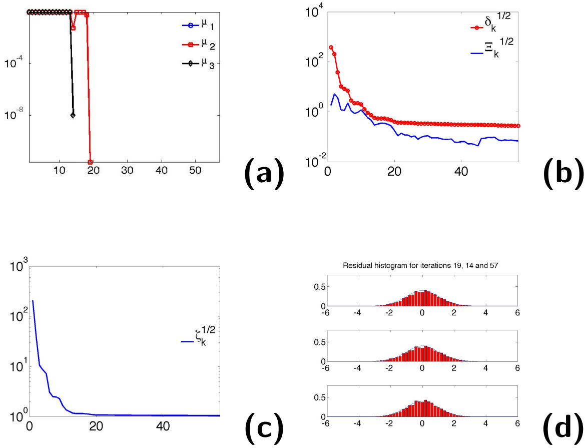
$$\frac{\partial u}{\partial t} = -\Delta u \quad in \quad S_2 = [0,1] \times [0,1],$$
$$u = 0 \quad on \quad \partial S_2, \quad u(.,0) = u_0 \quad in \quad S_2$$

Science & Technology Facilities Council

Data Assimilation

The system is integrated with time-step dt, using an implicit Euler scheme. In the physical domain, a regular finite difference scheme is taken for the Laplace operator, with same spacing h in the two spatial dimensions. The data of our problem is computed by imposing a solution $u_0(x, y, 0)$ computing the exact system trajectory and observing Hu at every point in the spatial domain and at every time step. In our application, $m = 8100$, $n = 900 = 30^2$, $dt = 1$, $h = 1/31$, $N = 8$ and $H = \mathrm{diag}(1^{1.5}, 2^{1.5}, \ldots, n^{1.5})$. The observation vector y is obtained by imposing $u_0(x, y, 0) = \frac{1}{4}\sin(\frac{1}{4}x)(x-1)\sin(5y)(y-1)$, and by adding a random measurement error with Gaussian distribution with zero mean and covariance matrix $R_i = \sigma^2 I_n$, where $\sigma = 10^{-3}$. In our numerical experiments, we use PCGLS without preconditioner.

Data Assimilation: results (d=5)



(a) Stopping criteria, (b) Energy norm of the error, (c) $\zeta_k = \frac{\|y\|_2^2 - \nu_k}{m-n}$, (d) Residual histogram.

Science & Technology Facilities Council

Conclusion

- **F**-test stopping rule for PCGLS

Science & Technology Facilities Council

Conclusion

- **F**-test stopping rule for PCGLS
- LSQR has similar behaviour

Conclusion

- **F**-test stopping rule for PCGLS
- LSQR has similar behaviour
- Orthodir (or re-orthogonalization) can be a valuable alternative for LSP where A is implicit

Conclusion

- **F**-test stopping rule for PCGLS
- LSQR has similar behaviour
- Orthodir (or re-orthogonalization) can be a valuable alternative for LSP where A is implicit
- How to generalize to other non linear regression problems or to NON-NORMAL distribution?