

Détermination du paramètre de régularisation pour la résolution de problèmes mal posés

Gérard MEURANT

CEA/DIF

February 17, 2007

- 1 Introduction
- 2 Régularisation
- 3 Bornes pour les formes quadratiques
- 4 Bidiagonalisation de Lanczos
- 5 Exemple : GCV
- 6 Améliorations
- 7 L-courbe
- 8 L-ruban
- 9 Conclusions

Introduction

On souhaite résoudre

$$Ax = y$$

où A (matrice $m \times n$) vient de la discrétisation d'un problème inverse

Ex : Problème de Baart (1982)

Equation intégrale de Fredholm de première espèce

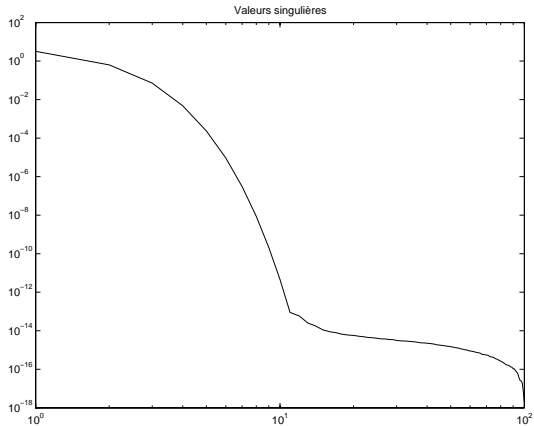
$$\int_0^{\pi} \exp(s \cos(t))x(t) dt = 2 \sinh(s)/s, \quad s \in [0, \pi/2]$$

$$\text{Solution : } x(t) = \sin(t)$$

En général, le second membre est entaché d'erreurs $y = \hat{y} + e$
 e bruit inconnu

Valeurs singulières de A

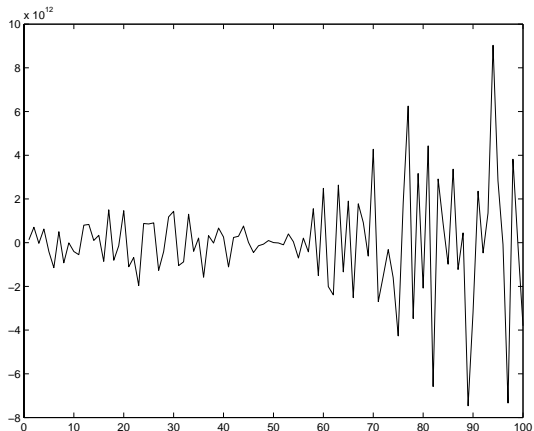
La matrice est très mal conditionnée $\sigma_{min} = 1.72 \cdot 10^{-18}$



- 1 Introduction
- 2 Régularisation**
- 3 Bornes pour les formes quadratiques
- 4 Bidiagonalisation de Lanczos
- 5 Exemple : GCV
- 6 Améliorations
- 7 L-courbe
- 8 L-ruban
- 9 Conclusions

Régularisation

On ne peut pas résoudre $Ax = y$ directement, même avec QR



Régularisation 2

Tikhonov : on résout

$$(A^T A + \nu I)x = A^T y$$

Problème : comment choisir ν ?

Régularisation 3

Il existe de nombreuses méthodes :

Discrepancy principle : $\|y - A(A^T A + \nu I)^{-1} A^T y\| = \|e\|$

Equivalent à $\nu^2 y^T (A A^T + \nu I)^{-2} y = \|e\|^2$

Nécessite la connaissance de $\|e\|$

Validation croisée généralisée (GCV) : minimum de

$$G(\nu) = \frac{\frac{1}{m} \|(I - A(A^T A + \nu I)^{-1} A^T) y\|^2}{\left(\frac{1}{m} \text{tr}(I - A(A^T A + \nu I)^{-1} A^T)\right)^2}$$

L-courbe : “coin” de $\log(\|x(\nu)\|)$ en fonction de $\log(\|y - Ax(\nu)\|)$

On peut résoudre facilement ces problèmes si l'on peut calculer la SVD de A

On s'intéresse au cas où A est de grande taille (donc pas de SVD)

Remarque :

Toutes ces méthodes demandent l'évaluation de

$$y^T A(A^T A + \nu I)^{-p} A^T y$$

et/ou

$$y^T (A A^T + \nu I)^{-p} y$$

$p = 1, 2, 3, 4$

- 1 Introduction
- 2 Régularisation
- 3 Bornes pour les formes quadratiques**
- 4 Bidiagonalisation de Lanczos
- 5 Exemple : GCV
- 6 Améliorations
- 7 L-courbe
- 8 L-ruban
- 9 Conclusions

Bornes pour les formes quadratiques

Golub & Meurant (1993)

Supposons B symétrique définie positive

$$B = Q\Lambda Q^T, \quad \Lambda \text{ diagonale}$$

On s'intéresse à

$$u^T f(B)u, \quad f(B) = Qf(\Lambda)Q^T$$

$$I[f] = u^T f(B)u = \int_a^b f(\lambda) d\alpha(\lambda)$$

Bornes pour les formes quadratiques 2

La mesure α est constante par morceaux, si $w = Q^T u$

$$\alpha(\lambda) = \begin{cases} 0 & \text{if } \lambda < a = \lambda_1, \\ \sum_{j=1}^i w_j^2 & \text{if } \lambda_i \leq \lambda < \lambda_{i+1}, \\ \sum_{j=1}^n w_j^2 & \text{if } b = \lambda_n \leq \lambda \end{cases}$$

On utilise des quadratures de [Gauss](#), [Gauss–Radau](#) pour calculer des bornes de $I[f]$

Bornes pour les formes quadratiques 3

Ces formules de quadrature sont intimement liées à l'algorithme de **Lanczos** en partant de $v^1 = u/\|u\|$

$$BV_k = V_k T_k + G_k, \quad G_k = (0 \quad \eta_{k+1} v^{k+1})$$

T_k is tridiagonal

Bornes pour les formes quadratiques 4

La borne donnée par **Gauss** à l'itération k de **Lanczos** est

$$(e^1)^T f(T_k) e^1$$

Pour nous

$$B = A^T A \text{ ou } B = AA^T$$

et

$$f(t) = (t + \nu)^{-p}$$

Pour **Gauss–Radau** on doit modifier T_k pour obtenir le noeud prescrit ($= 0$ dans notre cas)

- 1 Introduction
- 2 Régularisation
- 3 Bornes pour les formes quadratiques
- 4 Bidiagonalisation de Lanczos**
- 5 Exemple : GCV
- 6 Améliorations
- 7 L-courbe
- 8 L-ruban
- 9 Conclusions

Bidiagonalisation de Lanczos

On n'utilise pas B mais A et A^T

Golub–Kahan (1965)

$$q^0 = c/\|c\|, w = A^T q^0, \gamma_1 = \|w\|, p^0 = r^0/\gamma_1,$$

pour $k = 1, \dots$

$$w = Ap^{k-1} - \gamma_k q^{k-1}$$

$$\delta_k = \|w\|$$

$$q^k = w/\delta_k$$

$$w = A^T q^k - \delta_k p^{k-1}$$

$$\gamma_{k+1} = \|w\|$$

$$p^k = w/\gamma_{k+1}$$

Bidiagonalisation de Lanczos 2

$$C_k = \begin{pmatrix} \gamma_1 & & & & \\ \delta_1 & \ddots & & & \\ & \ddots & \gamma_{k-1} & & \\ & & \delta_{k-1} & \gamma_k & \\ & & & & \end{pmatrix}$$

Si

$$P_k = (p^0 \ \dots \ p^{k-1}), \quad Q_k = (q^0 \ \dots \ q^{k-1})$$

on a

$$\begin{aligned} A^T Q_k &= P_k C_k^T, \\ A P_k &= Q_k C_k + \delta_k q^k (e^k)^T \end{aligned}$$

Donc

$$A A^T Q_k = Q_k (C_k C_k^T) + \gamma_k \delta_k q^k (e^k)^T$$

et de même pour $A^T A$

Bidiagonalisation de Lanczos 3

On doit calculer des quantités du type

$$I[C_k] = (e^1)^T (C_k^T C_k + \nu I)^{-p} e^1$$

Ceci peut se faire en résolvant des pb de moindres carrés ou en calculant la SVD de $C_k (=U_k \Sigma_k V_k^T)$

$$I[C_k] = \sum_{i=1}^k \frac{d_i^2}{(\sigma_i^2 + \nu)^p}, \quad d = V_k^T e^1$$

Pb : les bornes s'améliorent quand $k \nearrow$, critère d'arrêt ?

Approximation de la trace

Hutchinson (1989)

Vecteurs aléatoires (distribution normale) u^i

$$\text{tr}[(AA^T + \nu I)^{-1}] \approx \frac{1}{q} \sum_{i=1}^q (u^i)^T (AA^T + \nu I)^{-1} u^i$$

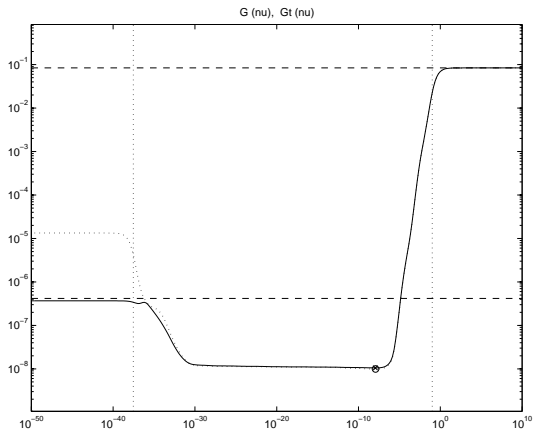
$$G(\nu) \rightarrow \tilde{G}(\nu)$$

En pratique, on prend un seul vecteur u et on calcule des bornes de $u^T (AA^T + \nu I)^{-1} u$

- 1 Introduction
- 2 Régularisation
- 3 Bornes pour les formes quadratiques
- 4 Bidiagonalisation de Lanczos
- 5 Exemple : GCV**
- 6 Améliorations
- 7 L-courbe
- 8 L-ruban
- 9 Conclusions

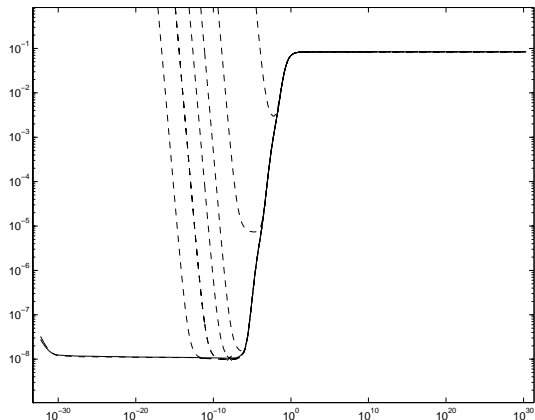
Exemple : GCV

Problème : Baart, Regutools (Hansen), $n = 100$



fonctions G et \tilde{G} , Baart, $m = n = 100$, $\|e\| = 10^{-3}$

Exemple : GCV 2



fonctions G et \tilde{G} et borne sup, Baart, $m = n = 100$, $\|e\| = 10^{-3}$

Critère d'arrêt

- ▶ **Golub et Von Matt** (1997) calculent aussi une borne inférieure et s'arrêtent lorsque le maximum de la borne inférieure est supérieure au minimum de la borne supérieure après $\lceil 3 \log \min(m, n) \rceil$ itérations

Critère d'arrêt

- ▶ **Golub et Von Matt** (1997) calculent aussi une borne inférieure et s'arrêtent lorsque le maximum de la borne inférieure est supérieure au minimum de la borne supérieure après $\lceil 3 \log \min(m, n) \rceil$ itérations
- ▶ On ne veut pas calculer les bornes en de trop nombreux points ν_i

Critère d'arrêt

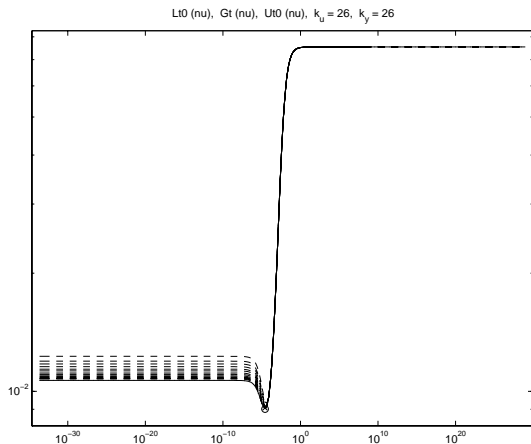
- ▶ Golub et Von Matt (1997) calculent aussi une borne inférieure et s'arrêtent lorsque le maximum de la borne inférieure est supérieure au minimum de la borne supérieure après $\lceil 3 \log \min(m, n) \rceil$ itérations
- ▶ On ne veut pas calculer les bornes en de trop nombreux points ν_i
- ▶ Il serait souhaitable de savoir si on a “convergé” avant de chercher les extrema

Critère d'arrêt

- ▶ Golub et Von Matt (1997) calculent aussi une borne inférieure et s'arrêtent lorsque le maximum de la borne inférieure est supérieure au minimum de la borne supérieure après $\lceil 3 \log \min(m, n) \rceil$ itérations
- ▶ On ne veut pas calculer les bornes en de trop nombreux points ν_i
- ▶ Il serait souhaitable de savoir si on a “convergé” avant de chercher les extrema
- ▶ Pb : la borne supérieure n'a pas le bon comportement quand $\nu \rightarrow 0$

Critère d'arrêt 2

Ce phénomène ne se produit pas lorsque $m > n$



fonctions G et \tilde{G} et borne sup, $m = 200$, $n = 100$, $\|e\| = 1$

- 1 Introduction
- 2 Régularisation
- 3 Bornes pour les formes quadratiques
- 4 Bidiagonalisation de Lanczos
- 5 Exemple : GCV
- 6 Améliorations**
- 7 L-courbe
- 8 L-ruban
- 9 Conclusions

Améliorations

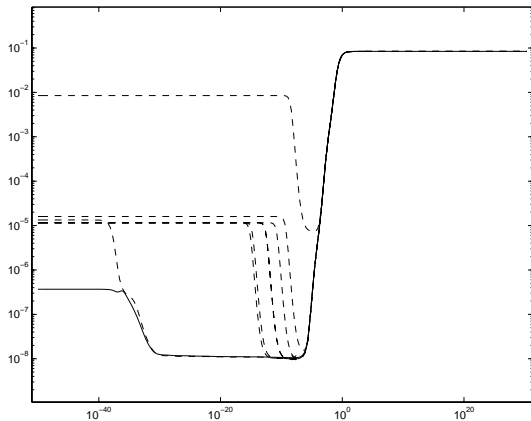
Pour améliorer la situation, on peut “tricher” un peu

Au lieu de considérer $p(\nu)/q(\nu)^2$, on s'intéresse à

$$\frac{p(\nu)}{q(\nu)^2 + \|y\|^2}$$

On obtient un comportement asymptotique plus intéressant et on peut tester la convergence pour une valeur petite de ν

Améliorations 2



fonctions G et \tilde{G} et borne sup modifiée, Baart,
 $m = n = 100, \|e\| = 10^{-3}$

Améliorations 3

- ▶ On teste la convergence pour une valeur petite de ν

Améliorations 3

- ▶ On teste la convergence pour une valeur petite de ν
- ▶ Ensuite on calcule uniquement un minimum des bornes supérieures

Améliorations 3

- ▶ On teste la convergence pour une valeur petite de ν
- ▶ Ensuite on calcule uniquement un minimum des bornes supérieures
- ▶ On en teste la convergence

Améliorations 3

- ▶ On teste la convergence pour une valeur petite de ν
- ▶ Ensuite on calcule uniquement un minimum des bornes supérieures
- ▶ On en teste la convergence
- ▶ Les valeurs des fonctions sont calculées à partir des SVD de C_k

Comparaisons

Bart

	$\ e\ $	nb it	ν/m	f min	f max	f total
GVM	10^{-7}	17	$9.6482 \cdot 10^{-15}$	494	613	1107
	10^{-5}	14	$9.7587 \cdot 10^{-12}$	125	132	257
	10^{-3}	14	$1.2018 \cdot 10^{-8}$	130	123	253
	10^{-1}	14	$1.0336 \cdot 10^{-7}$	128	126	254
	10	14	$8.8817 \cdot 10^{-8}$	127	119	246
GM	10^{-7}	12	$1.0706 \cdot 10^{-14}$	436	0	436
	10^{-5}	12	$1.0581 \cdot 10^{-11}$	437	0	437
	10^{-3}	8	$1.3077 \cdot 10^{-8}$	293	0	293
	10^{-1}	7	$1.1104 \cdot 10^{-7}$	294	0	294
	10	7	$9.1683 \cdot 10^{-8}$	294	0	294

Comparaisons 2

lLaplace

	$\ e\ $	nb it	ν/m	f min	f max	f total
GVM	10^{-7}	112	$2.1520 \cdot 10^{-15}$	12438	10216	22654
	10^{-5}	47	$5.2329 \cdot 10^{-12}$	4242	3428	7670
	10^{-3}	18	$2.2111 \cdot 10^{-8}$	620	541	1161
	10^{-1}	14	$1.9484 \cdot 10^{-5}$	120	125	245
	10	14	$6.5983 \cdot 10^{-3}$	124	126	250
GM	10^{-7}	58	$4.2396 \cdot 10^{-14}$	5239	0	5239
	10^{-5}	28	$5.4552 \cdot 10^{-11}$	1453	0	1453
	10^{-3}	17	$2.3046 \cdot 10^{-8}$	440	0	440
	10^{-1}	15	$2.0896 \cdot 10^{-5}$	293	0	293
	10	10	$6.8436 \cdot 10^{-3}$	296	0	296

Comparaisons 3

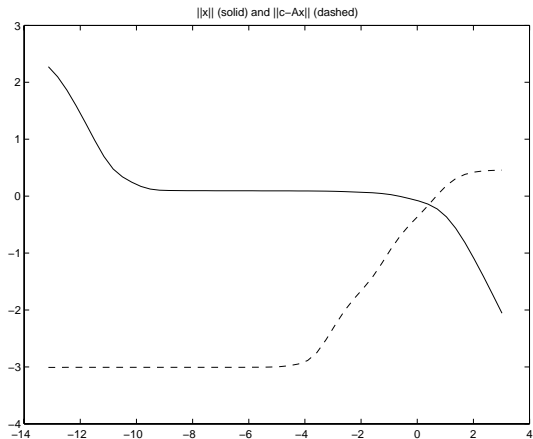
ILaplace

	$\ e\ $	ν/m	$\ y - Ax\ $	$\ x - x_0\ $	t(s)
GVM	10^{-7}	$2.1520 \cdot 10^{-15}$	$9.5132 \cdot 10^{-8}$	$1.4909 \cdot 10^{-2}$	10.06
	10^{-5}	$5.2329 \cdot 10^{-12}$	$9.6965 \cdot 10^{-6}$	$6.8646 \cdot 10^{-2}$	2.37
	10^{-3}	$2.2111 \cdot 10^{-8}$	$9.7215 \cdot 10^{-4}$	$1.9890 \cdot 10^{-1}$	0.35
	10^{-1}	$1.9484 \cdot 10^{-5}$	$9.8196 \cdot 10^{-2}$	$3.4627 \cdot 10^{-1}$	0.22
	10	$6.5983 \cdot 10^{-3}$	9.9095	$8.8165 \cdot 10^{-1}$	0.12
GM	10^{-7}	$4.2396 \cdot 10^{-14}$	$1.1004 \cdot 10^{-7}$	$2.7130 \cdot 10^{-2}$	2.03
	10^{-5}	$5.4552 \cdot 10^{-11}$	$1.0560 \cdot 10^{-5}$	$9.6771 \cdot 10^{-2}$	0.53
	10^{-3}	$2.3046 \cdot 10^{-8}$	$9.7243 \cdot 10^{-4}$	$1.9937 \cdot 10^{-1}$	0.29
	10^{-1}	$2.0896 \cdot 10^{-5}$	$9.8235 \cdot 10^{-2}$	$3.4634 \cdot 10^{-1}$	0.09
	10	$6.8436 \cdot 10^{-3}$	9.9115	$8.8791 \cdot 10^{-1}$	0.14

- 1 Introduction
- 2 Régularisation
- 3 Bornes pour les formes quadratiques
- 4 Bidiagonalisation de Lanczos
- 5 Exemple : GCV
- 6 Améliorations
- 7 L-courbe**
- 8 L-ruban
- 9 Conclusions

Normes

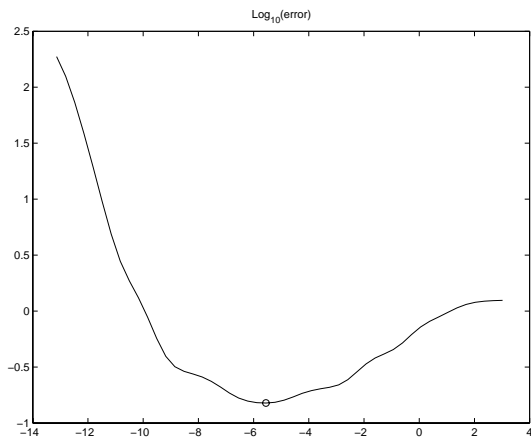
On trace $\log_{10} \|x(\nu)\|$ et $\log_{10} \|y - Ax(\nu)\|$ en fonction de $\log_{10} \nu$



Baart, $m = n = 100$, $\|e\| = 10^{-3}$

Normes 2

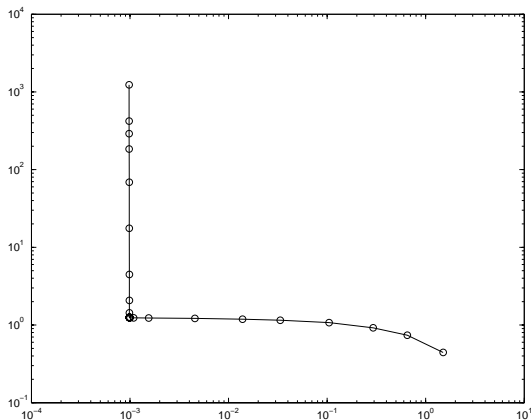
Ce qui est intéressant, c'est $\log_{10} \|x(\nu) - x_0\|$



$\log_{10} \|x(\nu) - x_0\|$ Baart, $m = n = 100$, $\|e\| = 10^{-3}$

L-courbe

En général, on trace $\log_{10} \|x(\nu)\|$ en fonction de $\log_{10} \|y - Ax(\nu)\|$



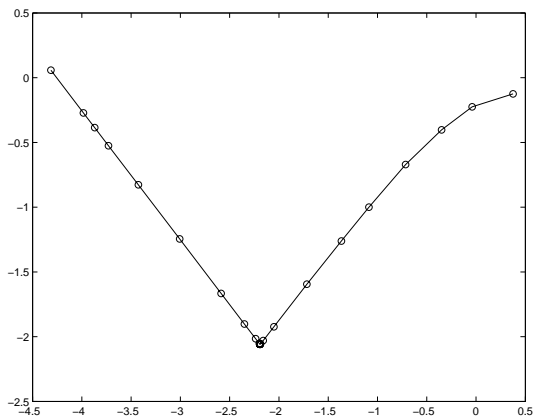
Baart, $m = n = 100$, $\|e\| = 10^{-3}$

L-courbe 2

Comment déterminer le “coin” quand on ne peut pas calculer la SVD ?

Hansen, Jensen et Rodriguez (2004) : pruning algorithm

Autre idée : faire pivoter la courbe



Baart, $m = n = 100$, $\|e\| = 10^{-3}$

L-courbe 3

- ▶ On calcule un petit nombre de points (25) ν_i

L-courbe 3

- ▶ On calcule un petit nombre de points (25) ν_i
- ▶ On détermine “l'origine” et on pivote de $-\pi/4$

L-courbe 3

- ▶ On calcule un petit nombre de points (25) ν_i
- ▶ On détermine “l’origine” et on pivote de $-\pi/4$
- ▶ On détermine le minimum d’index k

L-courbe 3

- ▶ On calcule un petit nombre de points (25) ν_i
- ▶ On détermine “l’origine” et on pivote de $-\pi/4$
- ▶ On détermine le minimum d’index k
- ▶ On recommence dans l’intervalle $[k - 1, k + 1]$

- 1 Introduction
- 2 Régularisation
- 3 Bornes pour les formes quadratiques
- 4 Bidiagonalisation de Lanczos
- 5 Exemple : GCV
- 6 Améliorations
- 7 L-courbe
- 8 L-ruban**
- 9 Conclusions

L-ruban

Calvetti, Golub, Reichel (1999)

Si on ne peut pas calculer des points sur la L-courbe, on s'intéresse aux bornes inf et sup de

$$y^T A(A^T A + \nu I)^{-2} A^T y$$

$$\nu^2 y^T (A A^T + \nu I)^{-2} y$$

L-ruban 2

- ▶ Pour chaque point ν_i , ça définit un rectangle

L-ruban 2

- ▶ Pour chaque point ν_i , ça définit un rectangle
- ▶ On considère les coins bas-gauche et haut-droit

L-ruban 2

- ▶ Pour chaque point ν_i , ça définit un rectangle
- ▶ On considère les coins bas-gauche et haut-droit
- ▶ A chaque itération, on cherche deux “coins” par pivotage des courbes “inf” et “sup”

L-ruban 2

- ▶ Pour chaque point ν_i , ça définit un rectangle
- ▶ On considère les coins bas-gauche et haut-droit
- ▶ A chaque itération, on cherche deux “coins” par pivotage des courbes “inf” et “sup”
- ▶ On s’arrête qd les valeurs de ν_c sont suffisamment proches

L-ruban 3

Baart

noise	nb it	ν/m	nb it sans reorth
10^{-7}	11	$6.0889 \cdot 10^{-17}$	40
10^{-5}	9	$6.1717 \cdot 10^{-13}$	19
10^{-3}	8	$6.3232 \cdot 10^{-9}$	10
10^{-1}	6	$7.2928 \cdot 10^{-5}$	6
10	5	$3.260 \cdot 10^{-2}$	5

L-ruban 4

lLaplace

noise	nb it	ν/m	nb it sans reorth
10^{-7}	23	$1.5871 \cdot 10^{-17}$	> 200
10^{-5}	20	$1.6951 \cdot 10^{-13}$	93
10^{-3}	15	$1.7535 \cdot 10^{-9}$	33
10^{-1}	10	$1.8139 \cdot 10^{-5}$	11
10	6	$1.3850 \cdot 10^{-2}$	6

- 1 Introduction
- 2 Régularisation
- 3 Bornes pour les formes quadratiques
- 4 Bidiagonalisation de Lanczos
- 5 Exemple : GCV
- 6 Améliorations
- 7 L-courbe
- 8 L-ruban
- 9 Conclusions

Conclusion

Les approximations par quadrature de **Gauss** et bidiagonalisation de **Lanczos** permettent de calculer le paramètre de régularisation sans avoir besoin de la SVD de **A**

Pbs :

Dans quel intervalle $[\nu_{min}, \nu_{max}]$ doit-on calculer ?

Préconditionnement pour les problèmes de grande taille ?