Continuous methods for numerical linear algebra problems

Li-Zhi Liao
(http://www.math.hkbu.edu.hk/∼liliao)
Department of Mathematics
Hong Kong Baptist University

The First International Summer School on Numerical Linear Algebra, Guangzhou and Hong Kong, July 17 - August 5, 2006

Roadmap of my talk

I. What is the continuous method?

II. Applications in numerical linear algebra (**NLA**).

III. Case study in symmetric eigenvalue problems.

I.- a) What is the continuous method

Target Problem:

$$\min_{x \in \Omega \subset R^n} f(x). \tag{1}$$

Conventional methods: iterative, $\{x_k\}$, $x_k \to x^*$, where x^* is a local minimizer of (1).

Original idea of the continuous method: form a continuous path (trajectory) from x_0 to x^*.

I.- a) What is the continuous method (cont.)

A) Mathematician's route

- Started in 1950's – K. J. Arrow, L. Hurwicz and H. Uzawa, *Studies in Linear and Nonlinear Programming*, Stanford University Press, Stanford, CA, 1958.

- Bear many names, ODE method, dynamic method, etc.

- Mathematical model:

$$\frac{dx(t)}{dt} = \text{descent and feasible direction of } f(x).$$

  $+$: easy to construct the ODE;

  –: difficult to prove the convergence of $x(t)$;

  –: difficult to solve constrained problems.

I.- a) What is the continuous method (cont.)

B) Engineer's route (Hopfield neural network)

Mathematical model:

$$\begin{cases} \text{Energy (or merit) function } E(x) \\ \frac{dx(t)}{dt} = p(x(t)), \quad \text{and} \quad \frac{dE(x(t))}{dt} < 0. \end{cases}$$

$+$: introduce an energy function;

$+$: hardware implementation;

–: energy function $E(x)$ must be a Lyapunov function;

–: $E(x)$ and $p(x)$ must be simple functions.

Ref.: L.-Z. Liao, H.-D. Qi, L. Qi, "Neurodynamical optimization" *J. Global Optim.*, 28, pp. 175-195, 2004.

I.- a) What is the continuous method (cont.)

Continuous method

Idea: take all the $+$'s and overcome all the –'s.

Geometric meaning: If we view the conventional method as that we put a person somewhere in a mountain with both eyes covered and ask him to find the bottom of the mountain, then we can view the continuous method as that the person finds a large metal ball and squeeze himself into the ball and let the ball falls freely.

I.- b) Framework of the continuous method

A mathematical framework of the continuous method for (1)

i) Define an energy function $E(x)$;

ii) Construct an ODE in the form of

$$\frac{dx(t)}{dt} = p(x(t))$$

such that $\frac{dE(x(t))}{dt} < 0$ and $\frac{dE(x(t))}{dt} = 0 \iff p(x) = 0$.

iii) Any local minimizer of (1) is an equilibrium point of the ODE.

I.- c) Why is the continuous method attractive

Theoretical aspect:

i) strong convergence results;

ii) weak conditions or assumptions;

iii) suitable for many kinds of problems.

Computational aspect:

i) simple and neat ODE systems;

ii) relaxed ODE solvers;

iii) large-scale problems;

iv) parallelizable.

II. Applications in NLA

a) Symmetric eigenvalue problems

Let $A \in R^{n\times n}$ and $A^T = A$.

From the Real Schur Decomposition, we have

$$A = U\Lambda U^T,$$

where $\Lambda = diag(\lambda_1, \lambda_2, \cdots, \lambda_n)$ with $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$, and $U = (u_1, u_2, \cdots, u_n)$ is orthogonal.

<ins>Extreme eigenvalue problem:</ins> find λ_1 and u_1.

<ins>Interior eigenvalue problem:</ins> find a λ_i and u_i such that $\lambda_i \in [a, b]$, where a and b are predefined constants.

II. - b) Least squares problems

Let A be a given $m \times n$ matrix of rank r and let b a given vector.

- **Linear least squares:**

  Find x^* so that

  $$\|b - Ax^*\|_2 = min.$$

  (Trivial)

- **Least squares with linear constraints:**

  Find x^* so that

  $$\|b - Ax^*\|_2 = min$$

  subject to

  $$C^T x^* = 0.$$

  (Trivial)

II. - b) Least squares problems (Cont.)

- **Least squares with linear and quadratic constraints:**

  Find x^* so that

  $$\|b - Ax^*\|_2 = min$$

  subject to

  $$C^T x^* = 0 \quad \text{and} \quad \|x^*\|_2^2 \leq \alpha^2.$$

  (Trivial)

- **Total least squares:**

  Find x^*, a matrix $\hat{E}$, and a residual $\hat{r}$ so that

  $$\left(\|\hat{E}\|_F^2 + \|\hat{r}\|_2^2\right) = min$$

  subject to

  $$(A + \hat{E})x^* = b + \hat{r}.$$

II. - b) Least squares problems (Cont.)

- **Regularized total least squares:**

  Need to solve the following problem

$$\min \frac{\|b - Ax\|_2^2}{1 + x^T V x}$$

  subject to

$$x^T V x = \alpha^2,$$

  where V is a given symmetric positive definite matrix.

II. - c) Nonnegative matrix factorization

Let $A \in R^{m \times n}$ be nonnegative, i.e. $A_{ij} \geq 0, \ \forall i, \ j$. For any given $k \leq \min(m, n)$, find nonnegative matrices $W \in R^{m \times k}$ and $H \in R^{k \times n}$ such that

$$\|A - WH\|_F^2 = min.$$

The extension of the above problem is the nonnegative tensor factorization.

III. Case study in symmetric eigenvalue problems

a) Conversion to optimization problems

Extreme eigenvalue problem:

$$\begin{array}{ll} \min\limits_{(\lambda,x)} & \lambda \\ s.t. & Ax = \lambda x, \\ & x^T x = 1. \end{array}$$

$$\Updownarrow$$

$$\begin{array}{ll} \min\limits_{x \in R^n} & x^T A x \\ s.t. & x^T x = 1. \end{array}$$

$$\Updownarrow$$

$$\begin{array}{ll} \min\limits_{x \in R^n} & x^T A x - c x^T x \\ s.t. & x^T x \leq 1, \end{array} \qquad (2)$$

where $c \geq \lambda_n + 1$.

III. - a) Conversion ... (Cont.)

Problem (2) has:

- concave objective function; and
- convex set constraint.

III. - a) Conversion ... (Cont.)

Properties of problem (2):

$$x \text{ is a local minimizer of (2)}$$

$$\Updownarrow$$

$$x \text{ is a global minimizer of (2)}$$

$$\Updownarrow$$

$$x \text{ satisfies } Ax = \lambda_1 x,\ x^T x = 1.$$

See G. Golub and L.-Z. Liao, "Continuous methods for extreme and interior eigenvalue problems", LAA, 415, pp. 31-51, 2006.

III. - a) Conversion ... (Cont.)

Interior eigenvalue problem:

$$\min_{(\lambda,x)} 1 \;\; s.t. \; Ax = \lambda x, \; x^T x = 1, \; a \le \lambda \le b.$$

$$\Downarrow$$

$$\begin{aligned} \min_{x \in R^n} \quad & x^T (A - aI_n)(A - bI_n)x \\ s.t. \quad & x^T x = 1. \end{aligned}$$

$$\Updownarrow$$

$$\begin{aligned} \min_{x \in R^n} \quad & x^T (A - aI_n)(A - bI_n)x - cx^T x \\ s.t. \quad & x^T x \le 1, \end{aligned} \qquad (3)$$

where $c \ge \max_{1 \le i \le n} (\lambda_i - a)(\lambda_i - b) + 1.$

Problem (3) has:

- concave objective function; and
- convex set constraint.

III. - a) Conversion ... (Cont.)

Properties of problem (3):

1) x is a local minimizer of (3) $\iff$ x is a global minimizer of (3)

2) Let x^* be a global minimizer of (3) and $\eta = (x^*)^T(A - aI_n)(A - bI_n)x^*$, then

   2a) If $\eta > 0$, then there exists no eigenvalue of A in $[a, b]$.

   2b) If $\eta \leq 0$, then there exists at least one eigenvalue of A in $[a, b]$.

   2c) If $\eta = 0$, then one of the eigenvalues of A must be a or b.

III. - a) Conversion ... (Cont.)

If we combine problems (2) and (3), we have

$$\min_{x \in R^n} \quad x^T H x - c x^T x \qquad (4)$$
$$s.t. \quad x^T x \leq 1,$$

where $H = A$ for (2) and $H = (A - aI_n)(A - bI_n)$ for (3) and $c \geq \lambda_{max}(H) + 1$. Note: the objective function is a concave function, so the solution is always on the boundary.

III. - b) Continuous models for eigenvalue problems

Dynamic Model 1:

Merit function:

$$f(x) = x^T H x - c x^T x. \tag{5}$$

Dynamical system:

$$\frac{dx(t)}{dt} = -\big\{x - P_\Omega[x - \nabla f(x)]\big\} :\equiv -e(x), \tag{6}$$

where $\Omega = \{x \in R^n \mid x^T x \leq 1\}$ and $P_\Omega(\cdot)$ is a projection operation defined by

$$P_\Omega(y) = arg \min_{x \in \Omega} \|x - y\|_2, \quad \forall y \in R^n.$$

III. - b) Continuous models ...(Cont.)

Properties of Model 1:

1) For any $x_0 \in R^n$, there exists a unique solution $x(t)$ of the dynamical system (6) with $x(t = t_0) = x_0$ in $[t_0, +\infty)$.

2) If $e(x_0) = 0 \implies x(t) \equiv x_0, \ \forall t \geq t_0$. If $e(x_0) \neq 0 \implies \lim_{t \to +\infty} e(x(t)) = 0$.

3) $e(x) = 0$ with $x \neq 0 \iff x$ is an eigenvector of H with $\|x\|_2 = 1$.

4) If $\|x_0\|_2 > 1$, $\|x(t)\|_2$ is monotonically decreasing to 1. If $\|x_0\|_2 < 1$, $\|x(t)\|_2$ is monotonically increasing to 1. If $\|x_0\|_2 = 1$, $\|x(t)\|_2 \equiv 1$.

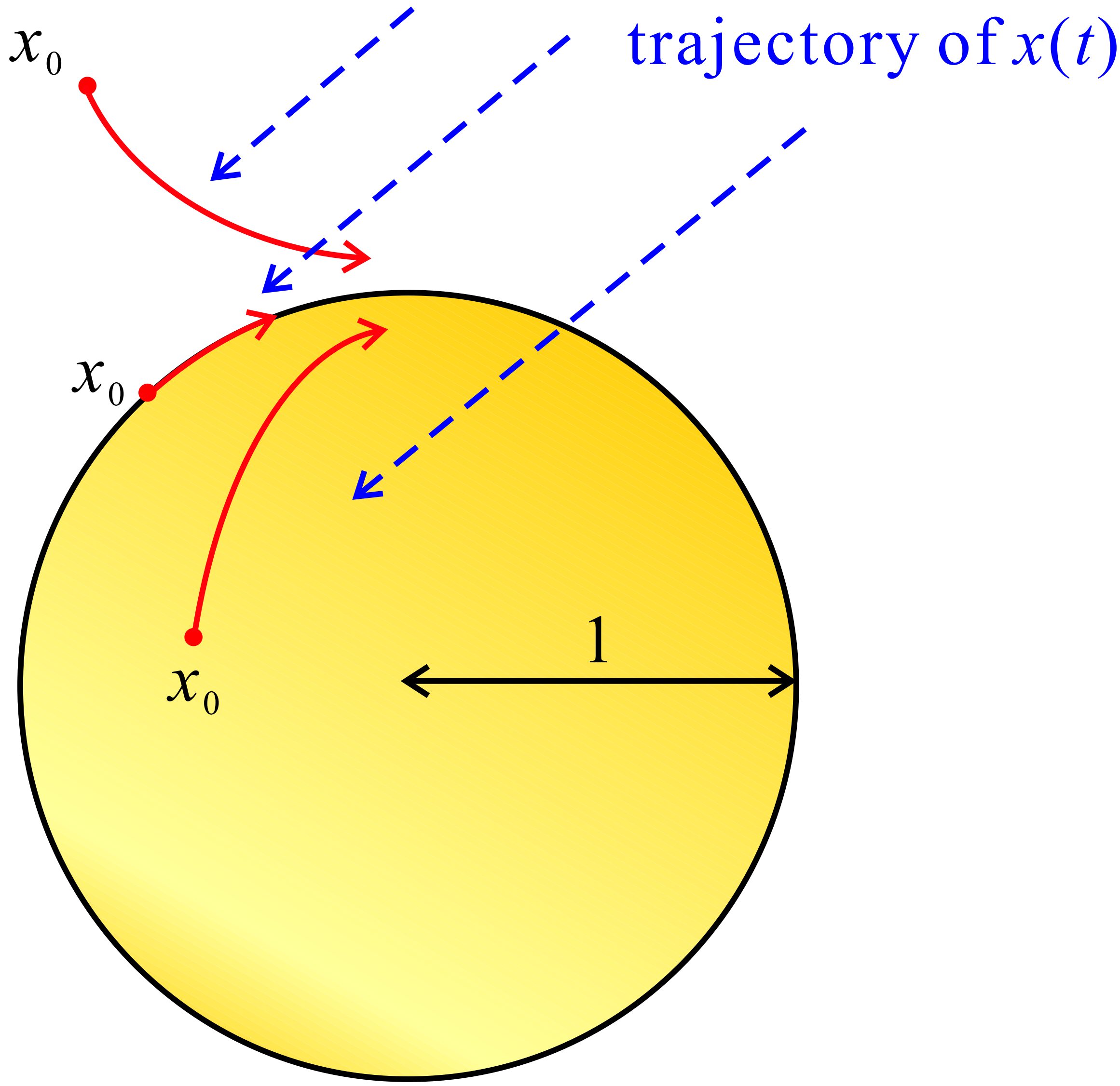

Figure 1: Dynamical trajectory of (6)

III. - b) Continuous models ...(Cont.)

5) If $x_0 \neq 0 \implies \exists\, x^*$ such that $\lim_{t \to +\infty} x(t) = x^*$ and $\|x^*\|_2 = 1$.

5.1) For $H = A$, we have

$$\lim_{t \to +\infty} x^T(t) A x(t) = \lambda_k,$$

where $k = \min\{i \mid x_0^T u_i \neq 0,\ i = 1, \cdots, n\}$.

Note: (λ_1, x^*) is what we want!

5.2) For $H = (A - aI_n)(A - bI_n)$, we have

$$\lim_{t \to +\infty} x^T(t)(A - aI_n)(A - bI_n)x(t) = \theta_k,$$

where $k = \min\{i \mid x_0^T v_i \neq 0,\ i = 1, \cdots, n\}$, $H = V \Theta V^T$, $\Theta = diag(\theta_1, \theta_2, \cdots, \theta_n)$, and $V^T V = I_n$ with $V = (v_1, v_2, \cdots, v_n)$.

Note: This θ_k may not be an eigenvalue of A.

III. - b) Continuous models ...(Cont.)

Steps to obtain an eigenvalue of A in $[a, b]$:

1) If $\theta_k = 0$, one of a or b is an eigenvalue of A. This can be verified by checking the values of $\|Ax^* - ax^*\|_2$ or $\|Ax^* - bx^*\|_2$.

2) If $\theta_k < 0$, solve

$$(\lambda_k - a)(\lambda_k - b) = \theta_k.$$

Two λ_k values can be obtained. Pick the one such that $\|Ax^* - \lambda_k x^*\|_2$ is very small. An eigenvalue of A in $[a, b]$ is found.

3) If $\theta_k > 0$, a new starting point has to be selected to start over. If after several tries, all θ_k's are positive, we may conclude that there is no eigenvalue of A in $[a, b]$.

III. - b) Continuous models ...(Cont.)

In numerical computation, we take

$$c = \|H\|_1 + 1.$$

But the numerical results seem to be sensitive to the value of c. The larger, the worse.

Reason:

Let θ_1 be the smallest eigenvalue of H, θ_s $(> \theta_1)$ be the second smallest eigenvalue of H. Then we have

$$\|x(t) - x^*\|_2 \le \epsilon \cdot e^{\frac{2(\theta_1 - \theta_s)}{1+2(c-\theta_1)}} \cdot \|x_0 - x^*\|_2.$$

Can we improve this?

III. - b) Continuous models ...(Cont.)

Dynamic Model 2:

Remember $\quad e(x) = x - P_\Omega[x - \nabla f(x)].$

Now, we define

$$e_\gamma = x - P_\Omega[x - \gamma \nabla f(x)], \quad 0 \leq \gamma \leq 1.$$

Merit function: (no change)

$$f(x) = x^T H x - c x^T x.$$

Dynamical system:

$$\frac{dx(t)}{dt} = -\alpha(x)\nabla f(x) - e_\gamma(x), \tag{7}$$

where

$$\alpha(x) = \begin{cases} -\eta\frac{|x^T e_\gamma(x)|}{x^T \nabla f(x)} & x \neq 0, \\ \gamma\eta & x = 0. \end{cases}$$

It can be shown that $\alpha(x)$ is locally Lipschitz continuous.

III. - b) Continuous models ...(Cont.)

All the theoretical results of Model 1 are also held for Model 2. In addition, Model 2 enjoys

$$\frac{df(x)}{dt} \leq -\alpha(x)\|\nabla f(x)\|_2^2 - \frac{1}{\gamma}\|e_\gamma(x)\|_2^2.$$

III. - c) Numerical results for eigenvalue problems

Example 1:

We construct Example 1 in the following steps:

1. Select $\Lambda = diag(-1e-4, -1e-4, 0, 0, 1, \cdots, 1) \in R^{n\times n}$.

2. Let $B = rand(n, n)$ and $[Q, R] = qr(B)$.

3. Define $A = Q^T \Lambda Q$.

Example 2:

Example 2 is similar to Example 1 except $\Lambda = diag(-1, -1, 0, 0, 1, \cdots, 1) \in R^{n\times n}$.

The two starting points used are $x_0 = (1, \cdots, 1)^T$ and $-x_0$.

III. - c) Numerical ... (Cont.)

Stopping criterion: $\|\frac{dx(t)}{dt}\|_\infty \leq 10^{-6}$.

Dynamical system solver: Matlab **ODE45** with **RelTol**$= 10^{-6}$ and **AbsTol**$=10^{-9}$.

III. - c) Numerical ... (Cont.)

Extreme eigenvalue problem: Model 1

<u>Test 1:</u> sensitivity to the initial point

We fix $n = 5,000$ and $c = \|A\|_1 + 1$ (=5.32 for Example 1 and =8.04 for Example 2).

Table 1 – Model 1

	Example 1		Example 2	
	CPU(s)	$\lambda^\dagger + 10^{-4}$	CPU(s)	$\lambda^\dagger + 1$
x_0	469.1	$3.0e-5$	475.6	$-6.5e-6$
$P_\Omega(x_0)$	298.6	$3.0e-5$	306.3	$-5.8e-6$
$-x_0$	469.0	$3.0e-5$	474.9	$-6.5e-6$
$P_\Omega(-x_0)$	301.3	$3.0e-5$	305.6	$-5.8e-6$

†: $\lambda = (x^*)^T A x^*$ is the computed eigenvalue.

III. - c) Numerical ... (Cont.)

Test 2: sensitivity to c

Table 2 – Model 1

		Example 1		Example 2	
	c	CPU(s)	$\lambda^{\dagger} + 10^{-4}$	CPU(s)	$\lambda^{\dagger} + 1$
$\frac{x_0}{\|x_0\|_2}$	10	423.5	$3.0e-5$	361.8	$3.8e-7$
	default	298.6	$3.0e-5$	306.3	$-5.8e-6$
	2	253.5	$3.0e-5$	258.3	$-4.0e-9$
$\frac{-x_0}{\|x_0\|_2}$	10	422.4	$3.0e-5$	362.0	$3.8e-7$
	default	301.3	$3.0e-5$	305.6	$-5.8e-6$
	2	252.6	$3.0e-5$	258.7	$-4.0e-9$

†: $\lambda = (x^*)^T A x^*$ is the computed eigenvalue.

III. - c) Numerical ... (Cont.)

Test 3: computational cost (we fix $c = 2$)

Table 3 – Model 1

		Example 1		Example 2	
n		CPU(s)	$\lambda^\dagger + 10^{-4}$	CPU(s)	$\lambda^\dagger + 1$
$1,000$	$\frac{x_0}{\|x_0\|_2}$	13253	$7.4e-6$	12.4	$-1.8e-9$
	$\frac{-x_0}{\|x_0\|_2}$	13290	$7.4e-6$	12.3	$-1.8e-9$
$2,500$	$\frac{x_0}{\|x_0\|_2}$	10694	$1.7e-5$	68.4	$1.7e-9$
	$\frac{-x_0}{\|x_0\|_2}$	10729	$1.7e-5$	69.3	$1.7e-9$
$5,000$	$\frac{x_0}{\|x_0\|_2}$	253.5	$3.0e-5$	258.3	$-4.0e-9$
	$\frac{-x_0}{\|x_0\|_2}$	252.6	$3.0e-5$	258.7	$-4.0e-9$
$7,500$	$\frac{x_0}{\|x_0\|_2}$	857.5	$7.1e-5$	951.5	$6.7e-9$
	$\frac{-x_0}{\|x_0\|_2}$	861.8	$7.1e-5$	953.8	$6.7e-9$

†: $\lambda = (x^*)^T A x^*$ is the computed eigenvalue.

The CPU time grows at a rate of $n^{2+\epsilon}$ where $\epsilon > 0$

III. - c) Numerical ... (Cont.)

Interior eigenvalue problem: Model 1

Test 4: no eigenvalue in the defined interval

We select $[a, b] = [-3 \times 10^{-4}, -2 \times 10^{-4}]$, and fix $n = 5,000$.

Table 4 – Model 1

	$P_\Omega(x_0)$		$P_\Omega(-x_0)$	
	c : def.	$c = 2$	c : def.	$c = 2$
Example 1				
CPU(s)	280.6	105.2	279.5	107.7
$\lambda = (x^*)^T A x^*$	-7e-5	-7e-5	-7e-5	-7e-5
$\\|Ax^* - \lambda x^*\\|_\infty$	$2e-5$	$4e-6$	$2e-5$	$4e-6$
θ_k	$1e-6$	$5e-8$	$1e-6$	$5e-8$
Example 2				
CPU(s)	538.3	111.0	534.4	106.0
$\lambda = (x^*)^T A x^*$	$1e-5$	$2e-8$	$1e-5$	$2e-8$
$\\|Ax^* - \lambda x^*\\|_\infty$	$6e-5$	$2e-6$	$6e-5$	$2e-6$
θ_k	$1e-5$	$8e-8$	$1e-5$	$8e-8$

III. - c) Numerical ... (Cont.)

Test 5: one eigenvalue in the defined interval

We select $[a, b] = [0.9, 1.1]$, and fix $n = 5,000$.

Table 5 – Model 1

	$P_\Omega(x_0)$		$P_\Omega(-x_0)$	
	c : def.	$c = 2$	c : def.	$c = 2$
Example 1				
CPU(s)	85.5	45.6	105.6	30.6
$\lambda = (x^*)^T A x^*$	1-1e-6	1+4e-9	1-1e-6	1+4e-9
$\|Ax^* - \lambda x^*\|_\infty$	2.9e-5	2.4e-6	2.9e-5	2.4e-6
θ_k	-0.01	-0.01	-0.01	-0.01
Example 2				
CPU(s)	147.0	65.7	142.4	66.3
$\lambda = (x^*)^T A x^*$	1-2e-6	1+6e-8	1-2e-6	1+6e-8
$\|Ax^* - \lambda x^*\|_\infty$	6.4e-5	2.3e-6	6.4e-5	2.3e-6
θ_k	-0.01	-0.01	-0.01	-0.01

III. - c) Numerical ... (Cont.)

Extreme eigenvalue problem: Model 2

We fixed the initial value at $P_\Omega(x_0)$ and select $c_i = \|A\|_1 + 10^i,\ i = 0, 1, 2.$

Table 6 – CPUs for Model 1 (Model 2) of Example 1

c	n		
	1,000	3,000	5,000
c_0	11.69 (8.172)	102.0 (63.73)	281.1 (175.5)
c_1	18.08 (7.828)	155.1 (61.34)	458.4 (173.1)
c_2	61.33 (24.89)	518.6 (98.13)	1515 (666.1)

The CPU time grows at a rate of n^2 for Model 2.

III. - c) Numerical ... (Cont.)

Interior eigenvalue problem: Model 2

We select $[a,b]^{(1)} = [-3 \times 10^{-4}, -2 \times 10^{-4}]$, $[a,b]^{(2)} = [0.9, 1.1]$, and $[a,b]^{(3)} = [0, 2]$. In addition, we fix $n = 5,000$, the initial value at $P_\Omega(x_0)$ and select $c_i = \|A\|_1 + 10^i,\ i = 0, 1, 2.$

Table 7 – CPUs for Model 1 (Model 2) of Example 1

c	$[a,b]$		
	$[a,b]^{(1)}$	$[a,b]^{(2)}$	$[a,b]^{(3)}$
c_0	505.6 (230.6)	91.03 (34.08)	139.9 (30.52)
c_1	603.2 (293.8)	104.5 (32.30)	157.9 (28.86)
c_2	1343 (577.6)	207.6 (26.34)	313.2 (22.92)

Why is Model 2 much better than Model 1?

Answer: Don’t know yet, in process

Concluding remarks

The continuous method is powerful, attractive, and still under development.