UNIVERSITY OF STRASBOURG
UFR DE MATHÉMATIQUES ET INFORMATIQUE

Problem Set #2: RB for Linear Affine Elliptic

Student name: *Roussel Desmond Nzoyem*

Course: *Calcul Scientifique 3* – Professor: *Pr. Christophe Prud'homme*
Due date: *November 20, 2020*

Question a)

Show that the operation count for the on-line stage of your code is independent of N.

For each step of the online stage, let's count the number of floating point operations (multiplications and additions):

1. Form $A_N(\mu)$ from $A_N(\mu) = \sum_{q=1}^{Q} \Theta^q(\mu) A_N^q$: requires QN^2 multiplications, and $(Q-1)N^2$ additions, hence $(2Q-1)N^2$ flops in total
2. Solve the $N \times N$ linear system $A_N(\mu)u_N(\mu) = F_N$: requires at most N^3 flops. In the case of the LU decomposition (where L has 1's along its diagonal), the cost of the decomposition is $\frac{2N^3}{3} - \frac{N^2}{2} - \frac{N}{6}$ flops, the cost for the descent algorithm is $N(N+1)$, and the cost for the ascent is $N(N+1) + N$
3. Evaluate the output $T_{rootN}(\mu) = L_N^T u_N(\mu)$: requires $2N-1$ flops as a dot product

In conclusion, the operation count for the on-line stage is independent of $\mathcal{N}$, it is equal to

$$(2Q-1)N^2 + \frac{2N^3}{3} - \frac{N^2}{2} - \frac{N}{6} + 2N(N+1) + N + 2N - 1$$

which yields

$$\frac{2}{3}N^3 + (2Q + \frac{1}{2})N^2 + \frac{29}{6}N - 1$$

roughly equal to

$$c_1 N^{\gamma_1} + c_2 N^{\gamma_2} + c_3 N^{\gamma_3}$$

with:

$$\begin{cases} c_1 = \frac{29}{6}, & \gamma_1 = 1 \\ c_2 = 2Q + \frac{1}{2}, & \gamma_2 = 2 \quad (Q = 6) \\ c_3 = \frac{2}{3}, & \gamma_3 = 3 \end{cases}$$

Question b)

1.

Generate the reduced basis matrix Z and all necessary reduced basis quantities.

For $N = 8$, $\mu = 1$ and $\mu = 10$, let's compare the condition number of $A_N(\mu)$, noted $\text{Cond}(A_N(\mu))$, when the Z matrix is taken directly from the "snapshots" (No G-S), and when when Z is orthonormalized using the Gram-Schmidt (G-S) process. γ and α are respectively the continuity (Lipschitz constant) and the coercivity constants for $A_{\mathcal{N}}$.

- $\mu = 1$:
  - **No G-S**: $\text{Cond}(A_N(\mu)) = 33575189921.6$
  - **with G-S**: $\text{Cond}(A_N(\mu)) = 1.0000000000959741$
  - **Upper bound**: $\frac{\gamma(\mu)}{\alpha(\mu)} = 1.0000000000000868$
- $\mu = 10$:
  - **No G-S**: $\text{Cond}(A_N(\mu)) = 23018985185.0$
  - **with G-S**: $\text{Cond}(A_N(\mu)) = 9.9286$
  - **Upper bound**: $\frac{\gamma(\mu)}{\alpha(\mu)} = 10.000$

It can be observed that the condition number is bounded by a function of μ when the Gram-Schmidt orthonormalization is applied to Z (specifically $\mu \mapsto \frac{1}{\mu}$ if $\mu < 1$ and $\mu \mapsto \mu$ if $\mu \geq 1$). This fits the results from the previous Problem Set, indicating that $\frac{\gamma(\mu)}{\alpha(\mu)}$ should be an upper bound for A_N's condition number (the upper bound is computed by solving a generalized eigen value problem $(A_{\mathcal{N}}(\mu) - \lambda A_{\mathcal{N}}(\bar{\mu}))\, x = 0$, for $x \in X^e$).

Without orthonormalization: Since μ_1, μ_2 were taken as snapshots to build the RB basis, one component of $u_N(\mu_1)$ and $u_N(\mu_2)$ must be equal to 1 while all the others are close to 0. This is what is observed. As for $\mu_3 = 1.0975$ which is not in 'sample1', the resulting $u_N(\mu_3)$ doesn't contain 0's nor 1's. Instead, the result is a combination of all the snapshots in Z, inducing a huge loss in accuracy.

With orthonormalization: Here the components in Z that do not count as part of the solution are less pronounced (closer to 0). For example, since $u_N(\mu_1)$ is the first snapshot, its first component is 1 and the rest are 0 (close to the floating point precision). The same result is observed for $u_N(\mu_2)$, which is a combination of the first and second snapshots.

2.

Let's verify the output.

For $\mu = 1.5$, $\text{Bi} = 0.1$, the computed value on a coarse grid is effectively

$$T_{rootN}(\mu) = 1.531074970789645$$

3.

Convergence study over a test sample.

Figure fig. 1 shows the convergence of the maximum relative error in the energy norm and the maximum relative output error as a function of N. As expected, it proves the errors decrease as our reduced basis' dimension N increases.

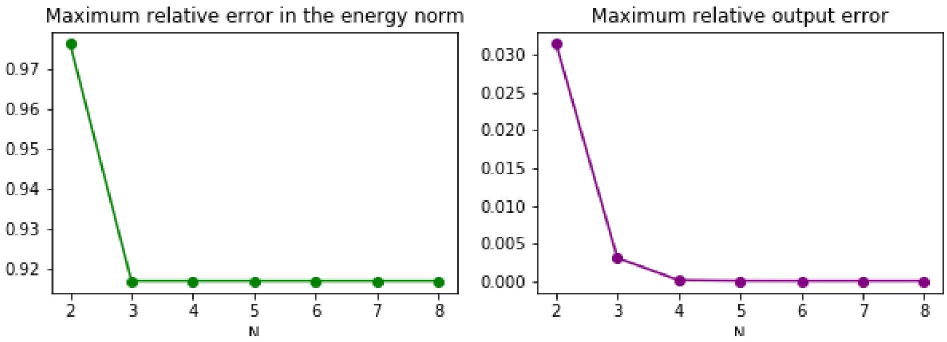
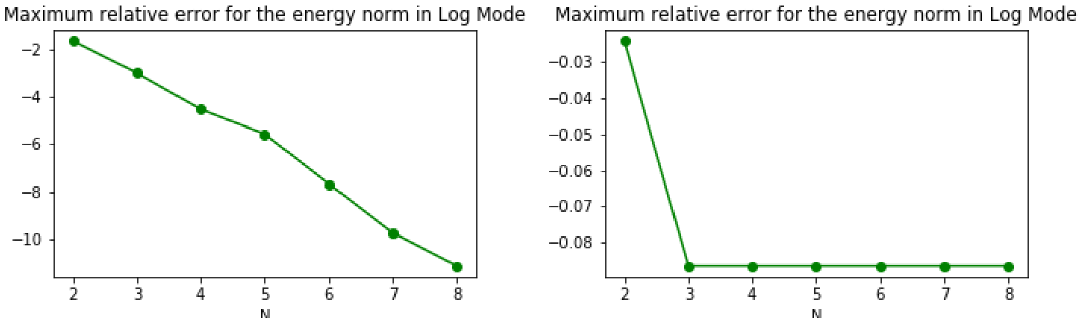


Figure 1: Maximum relative error in the energy norm and the maximum relative output error as a function of N, applied to sample 1. This plot is identical for a coarse, a medium, and fine FE triangulation

Now let's compare the convergence in energy norm (in log mode) when the matrix Z is (not) orthonormalized. fig. 2 shows a much faster convergence (and a considerably lower error) when the matrix Z is orthonormalized. This justifies the use of orthonormalization since question $b)1$.



(a) Without orthonormalization

(b) With orthonormalization

Figure 2: Comparison of energy norm error convergence in log mode when the RB matrix Z is orthonormalized, and when it is not.

4.

Average CPU time comparison.

From fig. 3, the relation between the computation and N cannot easily be deduced. However, it clearly indicates how much faster the reduced basis' online stage is, compared to the finite element's computation of the exact solution.

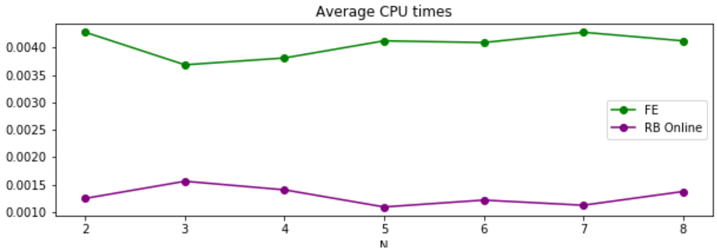


Figure 3: Average CPU time over test sample1 required to solve the reduced basis online stage with direct solution of the FE approximation as a function of N. This comparison is only valid on a coarse FE triangulation

5.

Required value of N for a 1% accuracy.

On a coarse triangulation, as long as N is greater or equal to 2, we have a relative accuracy of less than 1%. Moreover, the average time saving in terms on CPU time (compared to the FE method's computation) is about 0.0028125 seconds.

6.

Dependence of $\mathcal{N}$ on the results.

Repeating steps 3. to 5. on medium and fine triangulations we get the following results.

- COARSE:
  - Maximum relative errors: see fig. 1
  - CPU Time comparison: see fig. 3
  - Achieved accuracy: 0.3056515255907647 %
  - Required N: 3
  - CPU time savings: 0.0028645833333333327 sec
- MEDIUM:
  - Maximum relative errors: see fig. 1
  - CPU Time comparison: see fig. 4
  - Achieved accuracy: 0.3022017037569751 %
  - Required N: 3
  - CPU time savings: 0.012812499999999998 sec
- FINE:
  - Maximum relative errors: see fig. 1
  - CPU Time comparison: see fig. 5
  - Achieved accuracy: 0.300010213328605 %
  - Required N: 3
  - CPU time savings: 0.067875 sec

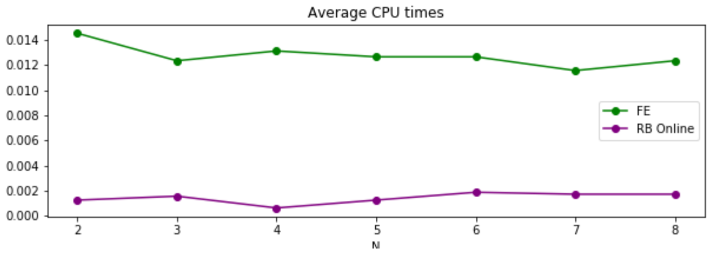


Figure 4: Average CPU time comparison on a medium FE triangulation

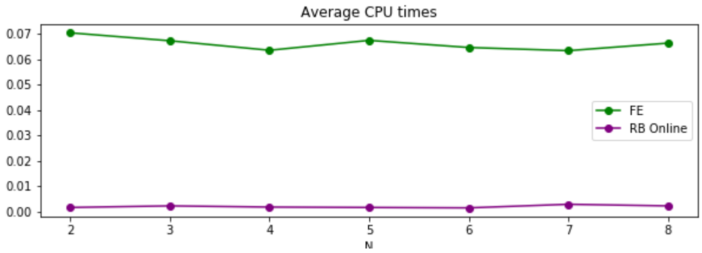


Figure 5: Average CPU time comparison on a fine FE triangulation

Unsurprisingly, the time required to compute the FE direct solution increases considerably (see figs. 3 to 5) while the time required by for the RB approximation is fairly constant around 0.002 second. This apparent non-dependence on $\mathcal{N}$ (on the triangulation) is once again observed on the maximum and relative errors in the RB approximation (the same figure (fig. 1) for all 3 triangulations).

As for the accuracy, we notice that while the required minimal N most still equal 3 to have a 1% output accuracy , the achieved accuracy clearly decreases with $\mathcal{N}$ for the wanted $N = 3$ (a 0% accuracy means the RB approximation is identical to the exact FE computation). With this output precision increase, the CPU time saved increases. In summary, **the finer the FE triangulation, the greater the need for a reduced basis approximation**. This is exactly what was expected.

Question c)

1.

Let's verify the output on Γ_{root}

For $\mu = 0.4, 0.6, 0.8, 1.2, 0.15$, the computed value on a medium grid is effectively

$$T_{rootN}(\mu) = 1.51561$$

2.

Convergence study over a test sample.

Figure fig. 6 shows the convergence of the maximum relative error in the energy norm and the maximum relative output error as a function of N.

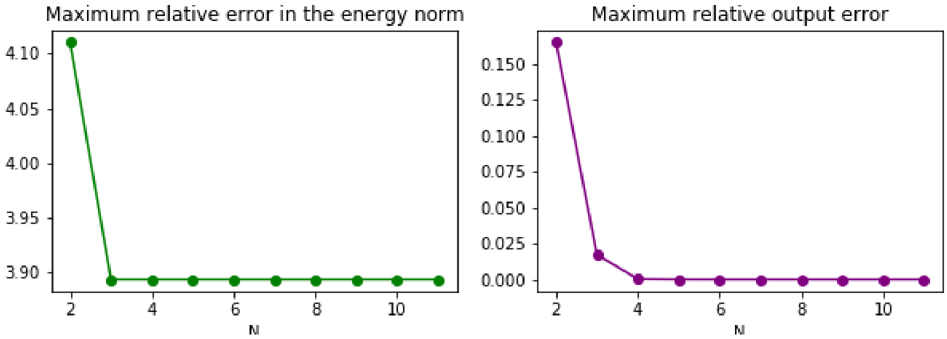


Figure 6: Maximum relative error in the energy norm and the maximum relative output error as a function of N, applied to sample 2.

We can see that the maximum relative errors in the energy norm and output error condirebably decrease as N gets bigger.

3.

Cost minimisation.

Applying the bisection method with a tolerance of 10^{-16}, we find that the optimal cost is $C = 1.4655$, obtained for the Biot number $\mathrm{Bi} = 0.40295$.

Question d)

1.

Convergence study over a test sample.

Figure fig. 7 shows the convergence of the maximum relative error in the energy norm and the maximum relative output error as a function of N.

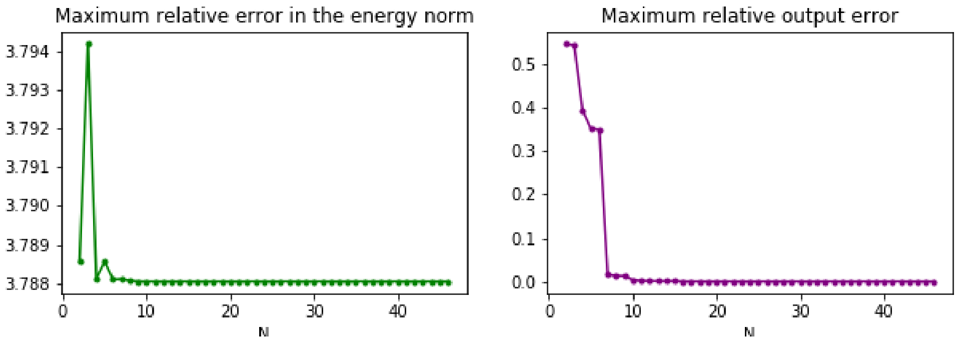


Figure 7: Maximum relative error in the energy norm and the maximum relative output error as a function of N, applied to sample 3.

As we noticed before, the maximum relative errors in the energy norm and output error decrease as N gets bigger (although not by much). However, when comparing it to figs. 1 and 6, it it clear that the errors reach quite higher values.