00_foundations

October 16, 2019

1 Applied Mathematics: an Introduction to Scientific Computing

Prof. Luca HELTAI

Prof. Gianluigi ROZZA

1.1 Basic Principles

Numerical Analysis is the *"art of approximating"*. Quoting Wikipedia:

> An approximation is an inexact representation of something that is still close enough to be useful. Although approximation is most often applied to numbers, it is also frequently applied to such things as mathematical functions, shapes, and physical laws.

Approximations should be used when incomplete information prevents use of exact representations. Many problems in physics are either too complex to solve analytically, or impossible to solve using the available analytical tools. Thus, even when the exact representation is known, an approximation may yield a sufficiently accurate solution while reducing the complexity of the problem significantly.

The type of approximation used depends on the available information, the degree of accuracy required, the sensitivity of the problem to this data, and the savings (usually in time and effort) that can be achieved by approximation.

In this course we focus on three main aspects:

- Methodologies (or approximation algorithms)
- Analysis (estimate errors and convergence properties)
- Implementation (through python and numpy notebooks)

Approximation is a matter of

- Representation (floating point values VS real numbers, finite dimensional spaces VS infinite dimensional ones, etc.)
- Measure of the Error (how do we know that we did a good job in approximating?)

In general, we will end up working with R^n. We recall here some basi principles:

1.1.1 Norms of vectors, matrices and functions

Given a vector space V over the field of real $\mathbb{R}$ or complex numbers $\mathbb{C}$ (V might be infinite dimensional), a *semi-norm* on V is a function $|\cdot| : V \to \mathbb{C}$ satisfying:

1. $|cf| = |c||f|$, for all $c \in \mathbb{C}$;
2. $|f+g| \le |f| + |g|$ (often known as triangle inequality).

As it can be easily seen (1)–(2) imply that the norm is always non-negative:

$$0 = 0 \cdot |f| = |0 \cdot f| = |(1-1)f| = |f - f| \le |f| + |(-1)f| = 2\ |f|.$$

The semi-norm becomes a *norm* if in addition to (1)–(2) we have also that for all $f \in V$

1. $|f| = 0$ if and only if $f = 0$,

A complete vector space with a norm is called a *Banach space*.

An inner product is any sesquilinear function $(\cdot,\cdot) : V \times V \mapsto \mathbb{C}$ satisfying the following conditions:

1. $(f,g) = \overline{(g,f)}$;
2. $(f,f) \ge 0$; $(f,f) = 0$ if and only if $f = 0$;
3. $(\alpha f, g) = \alpha(f,g)$ for all $\alpha \in \mathbb{C}$;
4. $(f+g,h) = (f,h) + (g,h)$.

The norm is then defined as $\|f\|^2 = (f,f)$. That this is a norm (i.e. satisfies the triangle inequality) is proved by first proving the *Cauchy-Schwarz* inequality

$$|(f,g)| \le \sqrt{(f,f)(g,g)}.$$

The proof of the latter is as follows: For any $\alpha \in \mathbb{C}$ we have that

$$0 \le (f - \alpha g, f - \alpha g) = (f,f) - \alpha(g,f) - \overline{\alpha}(f,g) + |\alpha|^2(g,g).$$

If $g = 0$ then the inequality is obviously true. If $g \neq 0$ we choose $\alpha = \frac{(f,g)}{(g,g)}$ to obtain the following

$$0 \le (f,f) - \frac{|(f,g)|^2}{(g,g)}.$$

The triangle inequality then follows from the Schwarz inequality

$$(f+g,f+g) = \|f+g\|^2 = \|f\|^2 + \|g\|^2 + \mathrm{Re}[(f,g)] \le (\|f\| + \|g\|)^2.$$

A Banach space with inner product and a norm induced by this product is called a *Hilbert space.*

Here are some examples of norms and semi-norms:

lp $V = \mathbb{R}^n$, and for any $x \in V$, its ℓ_p norm is defined as

$$\begin{aligned} \|x\|_p^p = & \sum_{i=1}^n |x_i|^p, \quad 1 \le p < \infty, \\ \|x\|_\infty = & \sup_{1 \le i \le n} |x_i|. \end{aligned}$$

Replacing n by ∞ in the above equations leads to the definition of a Banach space denoted usually with ℓ_p, with elements finite or infinite sequences for which

$$\begin{aligned} \|x\|_p^p = & \sum_i |x_i|^p, \quad 1 \le p < \infty, \\ \|x\|_\infty = & \sup_i |x_i|. \end{aligned}$$

are finite.

LP Let $I=(a,b)$. Then $L_p(I)$ is defined as the vector space of measurable functions f, for which

$$\begin{aligned}\|f\|_p :=& \quad \left(\int_I |f|^p \, dx\right)^{1/p} < \infty, \quad 1 \le p < \infty,\\ \|f\|_\infty :=& \qquad \operatorname*{ess\,sup}_{x\in I} |f(x)| < \infty, \quad p=\infty.\end{aligned}$$

That the quantity defined in Example [ex:LP] is a norm one can show by using Hölder's and Minkowski's inequalities, which are as follows: Let $1 \le p \le \infty$, and q is the conjugate exponent to p (i.e. $p^{-1}+q^{-1}=1$, with $q=\infty$, when $p=1$). Then

$$\begin{aligned}\|fg\|_1 \le \|f\|_p \|g\|_q, &\quad \text{(Hölder's inequality)}\\ \|f+g\|_p \le \|f\|_p + \|g\|_p, &\quad \text{(Minkowski's inequality)}.\end{aligned}$$

Note that for $p=q=2$ the Hölder's inequality is same as Schwarz inequality, by defining the inner product in L_2 as

$$(f,g) := \int_I f(x)g(x) \, dx.$$

Also note that the Minkowski's inequality is the triangle inequality in L_p. Similar inequalities hold true if the integrals are replaced by finite or infinite sums, and we also have Hölder's and Minkowski's inequalities for the spaces considered in example [ex:lp].

CK Let $I=[a,b]$ and $C^k(I)$ $k \in \mathbb{N}$ be the vector space of functions, whose derivatives of order $\le k$ are continuous. A (semi)-norm in $C^k(I)$ are then defined as

$$|f|_{k,\infty} = \|f^{(k)}\|_\infty = \sup_{x\in I} |f^{(k)}(x)|, \quad \|f\|_{W^k_\infty(I)} = \sup_{0\le i\le k} |f|_{i,\infty,I}.$$

WKP For $1 \le p < \infty$, and $k \in \mathbb{N}$ we define the *Sobolev (semi)-norm*, as follows:

$$|f|_{k,p,I} := \|f^{(k)}\|_{0,p,I} = \|f^{(k)}\|_p, \quad \|f\|_{k,p,I} := \left(\sum_{0\le i\le k} |f|^p_{i,p,I}\right)^{1/p}$$

The functions with finite Sobolev norm form the Banach space $W^k_p(I)$.

The examples above introduce several classes of Banach spaces, $C^k(I)$, $L_p(I)$ and W^k_p. They have a strightforward generalizations to higher dimensions.

For the normed finite dimensional spaces, the following result holds.

Let V be a finite dimensional vector space. Then all norms in V are equivalent.

and

Every finite dimensional space is closed.

1.1.2 Stability

In an abstract setting, we describe a generic problem as

$$F(x, d) = 0$$

where x is the unknown (generally a real number, a vector or a function) and $d \in D$ is the data. For each of the elements above we use an appropriate norm (see Lecture norms for a short introduction on norms and vector spaces), which will enable us to measure quantities of interests from a numerical point of view, such as errors, stability, and dependency of the solution from the data. In particular we will use the symbols $\|\cdot\|_F$, $\|\cdot\|_x$ and $\|\cdot\|_d$ to indicate the various norms.

In general, not all problems can be approximated. If we write a problem as above, then its approximation is useful only if the continuous problem has a unique solution which depends continuously on the data. We call these problems *well posed* or *stable*:

[def:cont-stability] A mathematical problem is *well posed* or *stable* if the following properties are satisfied:

- Uniqueness of solutions:
$$\forall d \in D, \exists!\, x \text{ s.t. } F(x, d) = 0.$$

- Continuous dependence on data:

Let δd be a perturbation of the data, such that $d + \delta d \in D$, and let $x + \delta x$ be the corresponding perturbed solution, i.e., $F(x + \delta x, d + \delta d) = 0$, then

$$\begin{aligned}
&\forall d \in D, \quad \exists \eta_0(d), K_0 \text{ s.t.} \\
&\|\delta d\|_d < \eta_0 \in D \quad \Longrightarrow \quad \|\delta x\|_x < K_0 \|\delta d\|_d.
\end{aligned}$$

1.1.3 Explanation for Dummies

The problem is considered stable is

Stable

$$\frac{||f(x - S_x) - f(x)||}{||S_x||} <= k$$

- *Where k is fixed and not depends on data*

ill-condition - problem is stable but k is **large**

Relative Stable

$$\frac{\frac{||f(x-S_x)-f(x)||}{||f(x)||}}{\frac{||S_x||}{||x||}} <= k$$

1.1.4 Condition numbers

A measure of how accurately we can approximate the problem at hand, is then given by the *Condition Number*:

Relative condition number:

$$K := \sup_{\delta d \text{ s.t. } d+\delta d \in D} \frac{\|\delta x\|_x / \|x\|_x}{\|\delta d\|_d / \|d\|_d}.$$

Absolute condition number (to be used when either $\|x\|_x = 0$ or $\|d\|_d = 0$):

$$K_{abs} := \sup_{\delta d \text{ s.t. } d+\delta d \in D} \frac{\|\delta x\|_x}{\|\delta d\|_d}.$$

If there exist a unique solution x to each data d, then we can construct a *resolvent map* G such that $G(d) = x$ and $F(G(d), d) = 0$. Assuming that G is differentiable, then a Taylor expansion of G around d allows us to express the condition numbers as

$$K \simeq \|G'(d)\| \frac{\|d\|_d}{\|G(d)\|_x}.$$

and

$$K_{abs} \simeq \|G'(d)\|.$$

A *stable* problem is *well conditioned* when its condition number is "small", where the meaning of "small" depends on the problem at hand.

1.1.5 Numerical stability

Once we have a *stable* problem, its approximation is usually given by a sequence of approximating problems

$$F_n(x_n, d_n) = 0, \qquad n \geq 1$$

such that

$$\lim_{n\to\infty} \|F_n - F\|_F = 0$$
$$\lim_{n\to\infty} \|x_n - x\|_x = 0$$
$$\lim_{n\to\infty} \|d_n - d\|_d = 0,$$

for some appropriate norms.

Equivalently to what happens in the continuous case, we can establish the stability of the approximate n-th problem.

A mathematical approximation of a stable problem is itself *stable* if the following properties are satisfied:

- Uniqueness of solutions:

$$\forall n \geq 1, \forall d_n \in D_n, \exists!\, x_n \text{ s.t. } F_n(x_n, d_n) = 0.$$

- Continuous dependence on data:

Let δd_n be a perturbation of the data, such that $d_n + \delta d_n \in D_n$, and let $x_n + \delta x_n$ be the corresponding perturbed solution, i.e., $F_n(x_n + \delta x_n, d_n + \delta d_n) = 0$, then

$$\begin{aligned} &\forall d_n \in D_n, \quad \exists \eta_n(d), K_n \text{ s.t.} \\ &\|\delta d_n\|_d < \eta_n \in D_n \quad \implies \quad \|\delta x_n\|_x < K_n \|\delta d_n\|_d. \end{aligned}$$

1.1.6 Consistency

Whenever the data d is admissible for F_n, then further properties of the approximations can be devised. In particular,

A numerical problem is *consistent*, when, assuming $d \in D_n \quad \forall n$,

$$\lim_{n\to\infty} F_n(x, d) = \lim_{n\to\infty} F_n(x, d) - F(x, d) = 0.$$

Moreover,

A numerical approximation is *strongly consistent* when

$$F_n(x, d) = 0, \qquad \forall n.$$

1.1.7 Convergence

A numerical method is *convergent* when

$$\begin{aligned} &\forall \varepsilon > 0, \qquad \exists\, n_0(\varepsilon), \exists\, \delta(\varepsilon, n_0) \text{ s.t.} \\ &\forall n > n_0, \quad \forall \delta d_n : \|\delta d_n\|_d < \delta \quad \implies \quad \|x(d) - x_n(d + \delta d_n)\| < \varepsilon, \end{aligned}$$

where $x(d)$ is the solution to $F(x, d) = 0$ and $x_n(d + \delta d_n)$ is the solution to $F_n(x_n, d + \delta d_n)$.

A convergent approximation is always stable.

1.1.8 Lax-Richtmyer theorem

One of the fundamental theorem of numerical analysis is the so called Lax-Richtmyer theorem:

> If a problem is consistent, then stability and convergence are equivalent.

1.1.9 Examples of Stable Problems

The problem of finding the solution $x \in R^n$ to the linear system of equations $Ax = d$, where $d \in R^n$ and $A \in R^{n\times n}$ can be written in the form $F(x, d) = 0$ simply by defining $F(x, d) := Ax - d$.

These problems are well defined if and only if the matrix A is invertible. In these cases, the resolvant $G(x)$ is the multiplication with the inverse of the matrix itself, i.e., if $F(G(d), d) = 0$ then it must be $G(d) = A^{-1}d$, and in general we have:

$$A(x + \delta x) = d + \delta d$$

$$\begin{aligned} A\delta x &= \delta d \\ \|A\delta x\| &= \|\delta d\| \end{aligned}$$

$$\begin{aligned} \delta x &= A^{-1}\delta d \\ \|\delta x\| &= \|A^{-1}\delta d\| \end{aligned}$$

$$\begin{aligned} \|\delta x\| &\leq \|A^{-1}\| \quad \|\delta d\| \\ \|d\| &\leq \|A\| \quad \|x\| \end{aligned}$$

$$\|\delta x\|/\|x\| \leq \|A^{-1}\|\|A\| \quad \|\delta d\|/\|d\|$$

We see that the **absolute condition number** of this problem is then equal to $\|A^{-1}\|$, and the relative condition number is instead equal to \$| A^{-1} || A | \$. This is what is usually called the **condition number** of a matrix (if nothing is said, it is intended that we are talking about the relative condition number).

[1]:
```python
import numpy as np
from numpy.linalg import solve, norm, cond

# Construct a random matrix, and check its condition number (fix random seed, so we have always the same number)
np.random.seed(101)
A = np.random.rand(100,100)

K = cond(A) # Linear algebra package of numpy

print(K)
```

```
1933.14691697
```

The condition number expresses the worst case scenario in relative error we should expect when solving linear systems. If we perturb the data with a unit vector, we should expect the new solution to be at distance K from the original solution...

[2]:

```
# Construct an artificial solution
x = np.random.rand(100)

d = A.dot(x); # We could use A*x only if we created a Matrix! This is an ndarray

# Verify that we got the right thing...
x_test = solve(A,d)
error = norm(x_test-x)
print("Initial Error", error)

# Now perturb d with a unit perturbation, and check the norm of the new solution
delta_d = np.random.rand(100)
delta_d /= np.linalg.norm(delta_d)

xnew = np.linalg.solve(A,d+delta_d)
delta_x = xnew-x

deviation = np.linalg.norm(delta_x)
print("Absolute deviation: ", deviation)

relative_deviation = (norm(delta_x)/norm(x))/(norm(delta_d)/norm(d))
print("Relative deviation: ", relative_deviation)
```

```
Initial Error 2.37596167042e-13
Absolute deviation:  0.676705032761
Relative deviation:  29.8201041662
```

We see in this case that, upon a unit norm perturbation in the data, we obtained a relative perturbation of about 50. The upper limit of this perturbation is given by the relative condition number. The bigger the condition number, the more sensitive to perturbations in the data will be your solution!