# Linear state updater

March 7, 2010

We give a method for integrating numerically a linear differential system with arbitrary precision. Let $\dot{X} = AX + b$ be a system of linear differential equations, with $A$ an $D \times D$ matrix and $b$ an $D$-long vector. We want to integrate the system over $[0, \mathrm{d}t]$ with $X(0) = X_0$ and precision $\varepsilon$, that is, such that $\left\| \widetilde{X}(\mathrm{d}t) - X(\mathrm{d}t) \right\|_\infty \leq \varepsilon$ where $X(t)$ is the exact solution, and $\widetilde{X}(t)$ is the numerical solution. $\mathrm{d}t$ is supposed to be small, typically $\mathrm{d}t = 10^{-4}$.

We now find the matrix $\widetilde{A}$ and the vector $\widetilde{b}$ such that :

$$X(\mathrm{d}t) = \widetilde{A}X_0 + \widetilde{b} \tag{1}$$

We get:

$$\widetilde{A} = \exp(A\,\mathrm{d}t) \tag{2}$$

$$\widetilde{b} = \exp(A\,\mathrm{d}t) \cdot \int_0^{\mathrm{d}t} \exp(-Au) \cdot b\,\mathrm{d}u \tag{3}$$

We can calculate $\exp(A\,\mathrm{d}t)$ with high accuracy in the scipy library. The integral is more difficult to calculate with high precision. We propose to use the Simpson integration scheme with $n$ steps within $[0, \mathrm{d}t]$.

The Simpson method consists in using the following expression for calculating the integral of a function $f$ between $a$ and $b$ with $n$ subintervals ($n$ is even):

$$\int_a^b f(x)\,\mathrm{d}x \approx \left(\frac{h}{3}\right)\left[f(a)+2\sum_{j=1}^{n/2-1} f(a+2jh)+4\sum_{j=1}^{n/2} f(a+(2j-1)h)+f(b)\right] \tag{4}$$

Here, we use the Simpson method with the vector-valued function $f(x) = \exp(-Ax) \cdot b$. This requires to calculate the matrix $\exp(-Ax)$ for $n$ values of $x$ between 0 and $\mathrm{d}t$.

The question is now: how to chose $n$ such that the precision of $\widetilde{X}(\mathrm{d}t)$ is $\varepsilon$ ? We now answer to this question.

In general, the Simpson method yields an absolute error with respect to the exact integral given by:

$$\left| \int_a^b f(x)\,\mathrm{d}x - I_{\text{simpson}} \right| \leq \frac{(b-a)^5}{180n^4} \max_{[a,b]} |f^{(4)}(x)| \tag{5}$$

Here, we have:

$$\frac{\mathrm{d}}{\mathrm{d}t} \exp(-Au)\mathrm{d}u = -A\exp(-Au) \tag{6}$$

hence:

$$\frac{\mathrm{d}^4}{\mathrm{d}t^4}\exp(-Au)\mathrm{d}u = A^4\exp(-Au) \tag{7}$$

Therefore, we find that:

$$\left\|\widetilde{X}(\mathrm{d}t) - X(\mathrm{d}t)\right\|_2 = \left\|\widetilde{b} - b\right\|_2 \tag{8}$$

$$\leq \|\exp(A\,\mathrm{d}t)\|_2 \cdot \left\| I_{\text{simpson}} - \int_0^{\mathrm{d}t}\exp(-Au)\mathrm{d}u \right\|_2 \tag{9}$$

$$\leq \|\exp(A\,\mathrm{d}t)\|_2 \cdot \frac{\mathrm{d}t^5}{180n^4} \cdot \sqrt{D} \cdot \max_{[0,\mathrm{d}t]} \left\|A^4 \cdot \exp(-Au) \cdot b\right\|_\infty \tag{10}$$

$$\leq \|\exp(A\,\mathrm{d}t)\|_2 \cdot \frac{\mathrm{d}t^5}{180n^4} \cdot \sqrt{D} \cdot \|A\|_2^4 \cdot \|b\|_2 \cdot e^{\|A\|_2 \mathrm{d}t} \tag{11}$$

$$\leq \frac{\mathrm{d}t^5}{180n^4} \cdot \sqrt{D}\,\|A\|_2^4\, e^{2\|A\|_2 \mathrm{d}t}\,\|b\|_2 \tag{12}$$

Finally, we find that:

$$\left\|\widetilde{X}(\mathrm{d}t) - X(\mathrm{d}t)\right\|_\infty \leq \frac{\mathrm{d}t^5}{180n^4}\sqrt{D}\,\|A\|_F^4\, e^{2\|A\|_F \mathrm{d}t}\,\|b\|_2 \tag{13}$$

where $\|A\|_F$ is the Frobenius norm of $A$ (the euclidean norm of $A$ seen as a $D^2$-long vector), which is simpler to compute than $\|A\|_2$.

In order to have $\left\|\widetilde{X}(\mathrm{d}t) - X(\mathrm{d}t)\right\|_\infty \leq \varepsilon$, one must choose $n$ such that:

$$n \geq \left(\frac{\mathrm{d}t^5}{180\varepsilon}\sqrt{D}\,\|A\|_2^4\, e^{2\|A\|_2 \mathrm{d}t}\,\|b\|_2\right)^{1/4} \tag{14}$$

With this formula, we find that $n = 100$ is sufficient to achieve machine precision for $\widetilde{b}$ with $\mathrm{d}t = 10^{-4}$, $D = 10$, and $\|A\|_F, \|b\|_2 \leq 10^3$.