Underpinning theory of the generalized scaling and squaring approximated exponential integrators, and some other things. Meeting notes.

Partecipants: Sebastiano Ferraris, Marco Lorenzi, Pankaj Daga, Marc Modat, Tom Vercauteren

1 Generalized scaling and squaring with approximated exponential integrator

Details of the proof of the generalized scaling and squaring with approximated exponential integrator formula.

Thesis: Let $\Omega \subset \mathbb{R}^d$ be an image's domain ($d = 2, 3$ for bi and tri-dimensional images), $\phi : \Omega \to \Omega$ a diffeomorphism and $\mathbf{v}$ its tangent vector field that defines the transformation's velocities and directions at each point of Ω.
The relationship between $\mathbf{v}$ and ϕ is given by the stationary ODE

$$\frac{\mathrm{d}\phi_t}{\mathrm{d}t} = \mathbf{v}(\phi_t), \qquad \phi_0 = \mathrm{Id}\ , \tag{1}$$

where ϕ_0 coincides with the identity function Id defined on Ω, and the solution at the time point $t = 1$ coincides with the diffeomorphism ϕ. When the one-parameter subgroup is applied to a point $\mathbf{x}$ in Ω, the new point $\phi_t(\mathbf{x})$ can be *denoted* with $\mathbf{x}(t)$ and its time derivative with $\dot{\mathbf{x}}(t)$. Equation (1), when considered for one particular point, can be rewritten as

$$\dot{\mathbf{x}}(t) = \mathbf{v}(\mathbf{x}(t)) \qquad \mathbf{x}(0) = \mathbf{x}_0\ . \tag{2}$$

The *associated linearised* ODE is provided by the Jacobian, that linearises the vector field $\mathbf{v}$ around $\mathbf{x}_0$, computed in $\mathbf{x}(t)$, as

$$\dot{\mathbf{x}}(t) = \mathbf{v}(\mathbf{x}_0) + \mathrm{J}_{\mathbf{v}(\mathbf{x}_0)}\mathbf{v}(\mathbf{x}(t) - \mathbf{x}_0)\ . \tag{3}$$

Where the terms with asympthotic error convergence $\mathcal{O}((\mathbf{x}(t) - \mathbf{x}_0)^2)$ has been neglected. The analytic solution of the linearised ODE is given by

$$\mathbf{x}(t) = \mathbf{x}_0 + \varphi_0\left(t \begin{bmatrix} \mathrm{J}_{\mathbf{v}(\mathbf{x}_0)} & \mathbf{v}(\mathbf{x}_0) \\ 0 & 0 \end{bmatrix}\right) \begin{bmatrix} \mathbf{0} \\ 1 \end{bmatrix}, \tag{4}$$

where φ_0 is the matrix exponential. It can be approximated with:

$$\mathbf{x}(t) \approx \mathbf{x}_0 + \mathbf{v}(\mathbf{x}_0) + \frac{1}{2}\mathrm{J}_{\mathbf{v}(\mathbf{x}_0)}\mathbf{v}(\mathbf{x}_0) \ . \tag{5}$$

for any initial condition $\mathbf{x}_0 \in \Omega$.

Some introductory facts: At the core of the ODE (1) there is the concept of *flow of diffeomorphisms*. It is defined as the family of diffeomorphisms $\{\phi_t\}_{t\in\mathbf{R}}$ continuously parametrized by a time-parameter t, such that ϕ_0 equals the identity Id and that satisfy the one-parameter subgroup property $\phi_t \circ \phi_s = \phi_{t+s}$.

Some more introductory facts: For any real $d \times d$ matrix A and any d-dimensional column vector $\mathbf{b}$, it holds

$$\dot{\mathbf{x}}(t) = A\mathbf{x}(t), \quad \mathbf{x}(0) = \mathbf{x}_0 \quad \Longrightarrow \quad \mathbf{x}(t) = \varphi_0(tA)\mathbf{x}_0 \quad \forall t \in \mathbb{R} \ . \tag{6}$$

$$\dot{\mathbf{x}}(t) = A\mathbf{x}(t) + \mathbf{b}, \quad \mathbf{x}(0) = \mathbf{0} \quad \Longrightarrow \quad \mathbf{x}(t) = t\varphi_1(tA)\mathbf{b} \quad \forall t \in \mathbb{R} \ . \tag{7}$$

Where $\varphi_0(Z)$ is the matrix exponential of $e^Z = \sum_{j=0}^{\infty} \frac{Z^k}{j!}$, whose numerical computation is performed with expm, and $\varphi_1(Z)$ is the shifted Taylor expansion given by $(\varphi_0(Z) - I)Z^{-1} = \sum_{j=0}^{\infty} \frac{Z^k}{(j+1)!}$ [2]. For a positive integer k, $\varphi_k(Z) = \sum_{j=0}^{\infty} \frac{Z^k}{(j+k)!}$.

Proof: For the affine ODE, the solution is given by 7. Without any loss of generality it is always possible to *translate the coordinate frame* so that the initial position of $\mathbf{x}_0$ coincides with the origin of the axis $\mathbf{0}$. The translation is given by $\mathbf{y}(t) := \mathbf{x}(t) - \mathbf{x}_0$, and in this new frame we have $\mathbf{y}(0) = \mathbf{0}$, $\mathbf{x}(t) = \mathbf{y}(t) + \mathbf{x}_0$, and equation (2) can be written as:

$$\dot{\mathbf{y}}(t) = \mathbf{v}(\mathbf{y}(t) + \mathbf{x}_0) \qquad \mathbf{y}(0) = \mathbf{0} \ . \tag{8}$$

Following the *exponential integrators approach* [1] of separating the linear part (whose integration is provided by the exponential map expm) and the non-linear part of the SVF, we linearise the SVF via Taylor expansion around $\mathbf{x}_0$ at $\mathbf{y}(t)+\mathbf{x}_0$ for any real t provides[1] as

$$\dot{\mathbf{y}}(t) = \mathbf{v}(\mathbf{y}(t) + \mathbf{x}_0) = \mathbf{v}(\mathbf{x}_0) + \mathrm{J}_{\mathbf{v}(\mathbf{x}_0)}\mathbf{y}(t) + \mathcal{O}(\mathbf{y}(t)^2) \ .$$

Neglecting the non-linear part, the ODE in homogeneous coordinates, becomes:

$$\begin{bmatrix} \dot{\mathbf{y}}(t) \\ 0 \end{bmatrix} = \begin{bmatrix} \mathrm{J}_{\mathbf{v}(\mathbf{x}_0)} & \mathbf{v}(\mathbf{x}_0) \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \mathbf{y}(y) \\ 1 \end{bmatrix} . \tag{9}$$

1 Given f, g and h, vector valued function in an Euclidean space, with the notation $f(x) = g(x) + \mathcal{O}(h(x))$ for $x \to x_0$ we mean that exists a real positive M and a δ such that $\|f(x) - g(x)\| < M\|h(x)\|$ when $\|x - x_0\| < \delta$.

Equation (8) can be written as

$$\dot{\mathbf{y}}(t) = \mathbf{v}_0 + \mathrm{J}_{\mathbf{v}_0}\mathbf{y}(t) + \mathcal{N}_{\mathbf{v}}(\mathbf{y}(t)) \ , \qquad \mathbf{y}(0) = \mathbf{0} \ , \tag{10}$$

where $\mathbf{v}(\mathbf{x}_0)$ is indicated with $\mathbf{v}_0$ for notation convenience and $\mathrm{J}_{\mathbf{v}_0}$ is the $d \times d$ spatial Jacobian of $\mathbf{v}_0$. The non-linear part of the SVF, indicated with $\mathcal{N}_{\mathbf{v}}(\mathbf{y}(t))$, can be seen as an operator on the space of the SVF over Ω, that subtract the linear part of $\mathbf{v}$ computed with the Taylor expansion of $\mathbf{v}$ in $\mathbf{x}(t)$ around $\mathbf{x}_0$. For a fixed $\mathbf{x}_0 \in \Omega$, on which is acting a one-parameter subgroup of diffeomorphisms, it is defined by:

$$\begin{aligned}\mathcal{N}_{\mathbf{v}}(\mathbf{y}(t)) &= \mathbf{v}(\mathbf{y}(t) + \mathbf{x}_0) - \big(\mathbf{v}_0 + \mathrm{J}_{\mathbf{v}_0}\mathbf{y}(t)\big) \\ &= \mathbf{v}(\mathbf{x}(t)) - \mathbf{v}(\mathbf{x}_0) - \mathrm{J}_{\mathbf{v}(\mathbf{x}_0)}(\mathbf{x}(t) - \mathbf{x}_0) \ .\end{aligned}$$

It follows easily that $\mathcal{N}_{\mathbf{v}}(\mathbf{y}(0)) = \mathbf{0}$ and $\mathcal{N}_{\mathbf{v}}(\mathbf{y}(t)) \in \mathcal{O}\big((\mathbf{x}(t) - \mathbf{x}(0))^2\big)$ when $\mathbf{x}(t) \to \mathbf{x}(0)$.

When the time-parameter t is in a small neighbour of the origin (as it happens when scaling the SVF by an appropriate factor in the generalized scaling and squaring framework), the non linear part $\mathcal{N}_{\mathbf{v}}(\mathbf{y}(t)) \in \mathcal{O}\big(\mathbf{y}(t)^2\big)$ is small and the solution of the linearised problem 10 approximates the solution of the initial problem (8). Now we can use implications (6) and (7) to compute its numerical solution.

Passing in *homogeneous coordinates*, the linearised ODE can be written as

$$\begin{bmatrix}\dot{\mathbf{y}} \\ 0\end{bmatrix} = \begin{bmatrix}\mathrm{J}_{\mathbf{v}_0} & \mathbf{v}_0 \\ 0 & 0\end{bmatrix}\begin{bmatrix}\mathbf{y}(t) \\ 1\end{bmatrix}$$

And using the implication (6), we have that the solution, written again in homogeneous coordinates, is

$$\begin{bmatrix}\mathbf{y}(t) \\ 1\end{bmatrix} = \varphi_0\left(t\begin{bmatrix}\mathrm{J}_{\mathbf{v}_0} & \mathbf{v}_0 \\ 0 & 0\end{bmatrix}\right)\begin{bmatrix}\mathbf{0} \\ 1\end{bmatrix}, \tag{11}$$

which is the exact solution of the linearised ODE, that approximates the sought solution for any time parameter t close enough to 0.

In order to *avoid the computational cost of the exponential of a matrix*, and to have an approach that can be easily vectorized, we can apply (7) to the linearised ODE:

$$\mathbf{y}(t) = t\varphi_1(t\mathrm{J}_{\mathbf{v}_0})\mathbf{v}_0 \ .$$

and, by translating the coordinate frame to the initial one with $\mathbf{y}(t) = \mathbf{x}(t) - \mathbf{x}(0)$, it follows:

$$\mathbf{x}(t) = \mathbf{x}(0) + t\varphi_1(t\mathrm{J}_{\mathbf{v}_0})\mathbf{v}_0 \ , \qquad \mathbf{x}(1) = \mathbf{x}(0) + \varphi_1(\mathrm{J}_{\mathbf{v}_0})\mathbf{v}_0 \ ,$$

that is the solution of the linearised ODE associated to (1) at the point $\mathbf{x}$.

For its *numerical computation, we can approximate* φ_1 truncating it at its second order:

$$\begin{aligned}\mathbf{x}(t) &= \mathbf{x} + t\varphi_1(t\mathrm{J}_{\mathbf{v}_0})\mathbf{v}_0 \\ &= \mathbf{x} + t\big(\mathrm{I} + \frac{t\mathrm{J}_{\mathbf{v}_0}}{2} + \frac{t^2\mathrm{J}^2_{\mathbf{v}_0}}{6} + \cdots\big)\mathbf{v}_0 \\ &= \mathbf{x} + t\mathbf{v}_0 + \frac{t^2}{2}\mathrm{J}_{\mathbf{v}_0}\mathbf{v}_0 + \mathcal{O}\big(t^3\mathrm{J}^2_{\mathbf{v}_0}\mathbf{v}_0\big) \ .\end{aligned}$$

for $t \to 0$. Therefore the solution to the ODE 1 can be written as

$$\phi_t(\mathbf{x}) = \mathbf{x} + t\mathbf{v}_0 + \frac{t^2}{2}\mathrm{J}_{\mathbf{v}_0}\mathbf{v}_0 + \mathcal{O}\big(t^3\mathrm{J}^2_{\mathbf{v}_0}\mathbf{v}_0\big) + \mathcal{O}\big((\mathbf{x}(t) - \mathbf{x}(0))^2\big) \ , \tag{12}$$

where the first asymptotic error limit is a consequence of having truncated φ_1, and the second as a consequence of having linearised the problem.

For $t \to 0$, the last equation can be rewritten as a function defined over Ω as:

$$\phi_t = \mathrm{Id} + t\mathbf{v} + \frac{t^2}{2}\mathrm{J}_{\mathbf{v}}\mathbf{v} + \mathcal{O}\big(t^3\mathrm{J}^2_{\mathbf{v}}\mathbf{v} + (\mathbf{x}(t) - \mathbf{x}(0))^2\big) \ , \tag{13}$$

where $\mathrm{J}_{\mathbf{v}}$ is the Jacobian function that at each $\mathbf{x}$ provides the vector valued operators $\mathrm{J}_{\mathbf{v}(\mathbf{x})}$.

When $t = 1$ we have that the initial ODE system can have approximation solution as:

$$\phi_1 \simeq \mathrm{Id} + \mathbf{v} + \frac{1}{2}\mathrm{J}_{\mathbf{v}}\mathbf{v} \ . \tag{14}$$

Using equation (14) in the generalized scaling and squaring: The equation just derived can be applied to the SVF $\mathbf{w}$, after having it reduced by a multiplicative factor of 2^N in the scaling and squaring framework. The final algorithm that improves the currently used scaling and squaring is given by the following steps:

1. Scaling of $\mathbf{w}$ by a factor of 2^N: $\mathbf{v} = \mathbf{w}/2^N$.
2. The approximation of the Lie exponential, indicated with $\widetilde{\exp}(\mathbf{v})$ si computed as

$$\widetilde{\exp}(\mathbf{v}) = \mathbf{x} + \mathbf{v}(\mathbf{x}) + \frac{1}{2}\mathrm{J}_{\mathbf{v}(\mathbf{x})}\mathbf{v}(\mathbf{x}) \ .$$

3. The result $\widetilde{\exp}(\mathbf{v})$ is pair-wise composed by itself 2^N-times.

2 Linear insights: Linear SVFs and homographies

Considering $d = 2$ for simplicity, when A is an element of the group of the rigid body transformations it can be represented in homogeneous coordinates as

$$A = \begin{bmatrix} \cos(\theta) & -\sin(\theta) & t_1 \\ \sin(\theta) & \cos(\theta) & t_2 \\ 0 & 0 & 1 \end{bmatrix}$$

where θ is the rotational part and t_1, t_2 are the translational part. The ODE system

$$\dot{\mathbf{x}}(t) = \mathbf{v}(\mathbf{x}(t)) \qquad \mathbf{x}(0) = \mathbf{x} \ .$$

becomes:

$$\begin{bmatrix} \dot{x}_1(t) \\ \dot{x}_2(t) \\ 0 \end{bmatrix} = \begin{bmatrix} \cos(\theta) & -\sin(\theta) & t_1 \\ \sin(\theta) & \cos(\theta) & t_2 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1(t) \\ x_2(t) \\ 1 \end{bmatrix} \ .$$

The analytical solution is given by $\varphi_0(A)$, element of the Lie group $SE(2)$, multiplied by the initial condition. When $\theta \neq 0$ is given by

$$\varphi_0(A) = \begin{bmatrix} 0 & -\theta & \frac{\theta}{2}\left(\frac{\sin(\theta)}{1-\cos(\theta)} t_1 + t_2\right) \\ \theta & 0 & \frac{\theta}{2}\left(-t_1 + \frac{\sin(\theta)}{1-\cos(\theta)} t_2\right) \\ 0 & 0 & 0 \end{bmatrix} \ .$$

The real parts of the eigenvalues of A are zeros and the integral curves are circles in the d-dimensional space around the fixed point of the transformation.

Homography group. The second example that we propose came from considering $\mathbf{x}(t)$ in projective coordinates as $\mathbf{X}(t)$, and from considering H as element of the Lie algebra of homographies. Again for $d = 2$ equation (2) becomes:

$$\begin{bmatrix} \dot{X}_1(t) \\ \dot{X}_2(t) \\ \dot{X}_3(t) \end{bmatrix} = H \begin{bmatrix} X_1(t) \\ X_2(t) \\ X_3(t) \end{bmatrix} \ .$$

whose exponential is an element in the group of homographies $\varphi_0(H)$. Passing in homogeneous coordinates with $(x_1(t), x_2(t)) = (X_1(t)/X_3(t), X_2(t)/X_3(t))$ and taking the derivative of the change of coordinates we obtain the ODE

$$\begin{cases} \dot{x}_1(t) = \frac{\dot{X}_1(t)X_3(t) - X_1(t)\dot{X}_3(t)}{X_3(t)^2} \\ \dot{x}_2(t) = \frac{\dot{X}_2(t)X_3(t) - X_2(t)\dot{X}_3(t)}{X_3(t)^2} \end{cases}$$

which corresponds, for $X_3(0) = 1$, to the non-linear system

$$\begin{cases} \dot{x}_1(t) = h_{11}x_1 + h_{12}x_2 + h_{13} - x_1(h_{31}x_1 + h_{32}x_2 + h_{33}) \\ \dot{x}_2(t) = h_{21}x_1 + h_{22}x_2 + h_{23} - x_2(h_{31}x_1 + h_{32}x_2 + h_{33}) \end{cases}$$

where h_{ij} are the components of H.

The solution of the last non-linear system is liked with the homogeneous solution

$$S = \varphi_0(H)\mathbf{X}(0) \qquad \mathbf{X}(0) = [\mathbf{x}(0), 1]^T$$

by $\mathbf{x}(1) = S[1:2]/S[3]$, that still has an analytic formulation.

References

1. Hochbruck, Marlis, and Alexander Ostermann. "Exponential integrators." Acta Numerica 19 (2010): 209-286.
2. Higham, Nicholas J., and Lin Lijing. "Matrix Functions: A Short Course." (2013).