PhD meeting 2-Jan-16
Underpinning theory of the generalized scaling and squaring approximated exponential integrators.

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

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. The solution is given by

$$\phi_1(\mathbf{x}) \simeq \mathbf{x} + \mathbf{v}(\mathbf{x}) + \frac{1}{2}\mathrm{J}_{\mathbf{v}(\mathbf{x})}\mathbf{v}(\mathbf{x}) \qquad \forall \mathbf{x} \in \Omega \ . \tag{2}$$

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 $\phi_t \circ \phi_s = \phi_{t+s}$.

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

$$\dot{\mathbf{x}}(t) = \mathbf{v}(\mathbf{x}(t)) \qquad \mathbf{x}(0) = \mathbf{x} \ . \tag{3}$$

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

$$\dot{\mathbf{x}}(t) = A\mathbf{x}(t), \quad \mathbf{x}(0) = \mathbf{b} \qquad \Longrightarrow \qquad \mathbf{x}(t) = \varphi_0(tA)\mathbf{b} \quad \forall t \in \mathbb{R} \ . \tag{4}$$

$$\dot{\mathbf{x}}(t) = A\mathbf{x}(t) + \mathbf{b}, \quad \mathbf{x}(0) = \mathbf{0} \qquad \Longrightarrow \qquad \mathbf{x}(t) = t\varphi_1(tA)\mathbf{b} \quad \forall t \in \mathbb{R} \ . \tag{5}$$

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, $\varphi_k(Z) = \sum_{j=0}^{\infty} \frac{Z^k}{(j+k)!}$.

Explanation of the equation (2): 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 (3) can be written as:

$$\dot{\mathbf{y}}(t) = \mathbf{v}(\mathbf{y}(t) + \mathbf{x}(0)) \qquad \mathbf{y}(0) = \mathbf{0} \ . \tag{6}$$

The Taylor expansion of the SVF $\mathbf{v}$ around $\mathbf{x}(0)$ computed at $\mathbf{y}(t) + \mathbf{x}(0)$ for any real t provides

$$\dot{\mathbf{y}}(t) = \mathbf{v}(\mathbf{y}(t) + \mathbf{x}(0)) = \mathbf{v}(\mathbf{x}(0)) + \mathrm{J}_{\mathbf{v}(\mathbf{x}(0))}\mathbf{y}(t) + \mathcal{O}(\mathbf{y}(t)^2) \ .$$

This last expansion, gives us the hint ot follow the *exponential integrators approach* [1]. It consists in the strategy of separating the linear part (whose integration is provided by the exponential map expm) and the non-linear part of the SVF. Equation (6) can be written as

$$\dot{\mathbf{y}}(t) = \mathbf{v}_0 + \mathrm{J}_{\mathbf{v}_0}\mathbf{y}(t) + \mathcal{N}_{\mathbf{v}}(\mathbf{y}(t)) \ , \qquad \mathbf{y}(0) = \mathbf{0} \ , \tag{7}$$

where $\mathbf{v}(\mathbf{x}(0))$ is indicated with $\mathbf{v}_0$ for notation convenience and $\mathrm{J}_{\mathbf{v}_0}$ is the $d \times d$ 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 $\mathbf{x} = \mathbf{x}(0) \in \Omega$, 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 $\mathcal{N}_{\mathbf{v}}(\mathbf{y}(t)) = \mathbf{0}$ and [1] $\mathcal{N}_{\mathbf{v}}(\mathbf{y}(t)) \in \mathcal{O}\big((\mathbf{x}(t) - \mathbf{x}(0))^2\big)$ 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), at the initial ODE problem we can associate a linearised version:

$$\dot{\mathbf{y}}(t) = \mathbf{v}_0 + \mathrm{J}_{\mathbf{v}_0}\mathbf{y}(t) \ , \qquad \mathbf{y}(0) = \mathbf{0} \ . \tag{8}$$

[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 $\|x - x_0\| < \delta$.

Since $\mathcal{N}_{\mathbf{v}}(\mathbf{y}(t)) \in \mathcal{O}\big(\mathbf{y}(t)^2\big)$, the solution of the linearised problem 8 approximates the solution of the initial problem (6).

Passing in *homogeneous coordinates*, equation (8) 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 (4), we have that the solution, written again in homogeneous coordinates, is

$$\begin{bmatrix} \mathbf{y}(t) \\ 1 \end{bmatrix} = \mathrm{expm}\left(t \begin{bmatrix} \mathrm{J}_{\mathbf{v}_0} & \mathbf{v}_0 \\ 0 & 0 \end{bmatrix} \right) \begin{bmatrix} \mathbf{0} \\ 1 \end{bmatrix} , \tag{9}$$

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 (5) to the linearised problem (8):

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

and 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

$$\phi_t(\mathbf{x}) = \mathbf{x} + t\mathbf{v}_0 + \frac{t^2}{2}\mathrm{J}_{\mathbf{v}_0}\mathbf{v}_0 + \mathcal{O}\big(t^3\mathrm{J}^2_{\mathbf{v}_0}\mathbf{v}_0\big) + \mathcal{O}\big((\mathbf{x}(t) - \mathbf{x}(0))^2\big) \ , \tag{10}$$

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.

Since $\mathcal{O}\big((\mathbf{x}(t) - \mathbf{x}(0))^2\big) = m\mathcal{O}\big(\mathbf{v}^2(\mathbf{x}(t))\big)$ for $t \to 0$ and some positive real m, the last equation can be rewritten as a function defined over Ω as:

$$\phi_t = \mathrm{Id} + t\mathbf{v} + \frac{t^2}{2}\mathrm{J}_{\mathbf{v}}\mathbf{v} + \mathcal{O}\big(t^3\mathrm{J}^2_{\mathbf{v}}\mathbf{v} + \mathbf{v}^2(\mathbf{x}(t))\big) \ , \tag{11}$$

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:

$$\phi_1 \simeq \mathrm{Id} + \mathbf{v} + \frac{1}{2}\mathrm{J}_{\mathbf{v}}\mathbf{v} \ , \tag{12}$$

Using equation (12) 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.

Notes for the next steps: (for a possible paper that may follow the workshop WBIR)

1. Computational complexity analysis.
2. Error analysis.
3. Extension to TVVF approximating the time variation with polynomials. Working on where the following equation originates [2]:
$$\frac{\mathrm{d}\phi_t}{\mathrm{d}t} = A\phi_t + ct, \quad \phi_0 = \mathbf{0} \quad \Longrightarrow \quad \phi_t = t^2\varphi_2(tA)c \quad \forall t \in \mathbb{R} \ . \tag{13}$$
where the separation between the temporal and the spatial transformation in $A\phi_t + ct$, may be given by the fact that the dimension is stored in the fourth component of the SVF. This requires further investigations...!
4. Extension to TVVF through autonomisation of variables (see slides notes).
5. Autonomisation for scaling and squaring. See if there something is making sense.

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).