Optimizing over SO(3)

Julian Panetta

October 31, 2018

To optimize an objective function involving rotational degrees of freedom, we first need to choose a representation for rotations. Our goal is to select a parametrization of $SO(3)$ that avoids singularities to the extent possible and that makes our optimizer's job easier.

We could use unit quaternions, which would solve the problem of singularities, but would require us to impose unit norm constraints during the optimization. We could use Euler angles, but these will run into singularities (gimbal lock) after just a $\frac{\pi}{2}$ rotation. Instead, we use the tangent space to $SO(3)$ (infinitesimal rotations) at some "reference" rotation R_0. In this representation, the additional rotation to be applied after R_0 is encoded as a vector pointing along the rotation axis with length equal to the rotation angle. The rotation is then obtained by the exponential map (more precisely, we construct the skew-symmetric cross-product matrix "X" for this vector and calculate $e^X R_0$).

This representation is nice because it allows rotations of up to π before running into singularities. We can avoid singularities entirely by setting bound constraints on our infinitesimal rotation components and then updating the parametrization (changing R_0 to the current rotation) if the optimizer terminates with one of these bounds active. We could even update R_0 at every step of the optimization, which would greatly simplify the gradient and Hessian formulas as we'll see in Section 4 (and as exploited in [2] and [3]). However, we derive the full formulas for the gradient and Hessian away from the identity, since updating the parametrization—changing the optimization variables—at every step isn't supported in off-the-shelf optimization libraries (e.g. Knitro or IPOPT). Note that [1] proposes using the same parametrization, though they only provide gradient formulas, not Hessian formulas (and work with quaternions instead of Rodrigues' rotation formula).

1 Representation and Exponential Map

We denote our infinitesimal rotation by vector $\mathbf{w}$, which encodes the rotation axis $\frac{\mathbf{w}}{\|\mathbf{w}\|}$ and angle $\|\mathbf{w}\|$. We can apply the rotation computed by the exponential map to a vector $\mathbf{v}$ using Rodrigues' rotation formula. For simplicity, we assume $R_0 = I$; this simplification can be applied in practice by first rotating $\mathbf{v}$ by R_0.

$$\tilde{\mathbf{v}} = R(\mathbf{w})\mathbf{v} = \mathbf{v}\cos(\|\mathbf{w}\|) + \mathbf{w}\mathbf{w}^T\mathbf{v}\frac{1-\cos(\|\mathbf{w}\|)}{\|\mathbf{w}\|^2} + (\mathbf{w}\times\mathbf{v})\frac{\sin(\|\mathbf{w}\|)}{\|\mathbf{w}\|}.$$

(We could obtain the entire rotation matrix by substituting the canonical basis vectors $\mathbf{e}^0$, $\mathbf{e}^1$, $\mathbf{e}^2$ in for $\mathbf{v}$.)

2 Gradients and Hessians

Now we compute derivatives of the rotated vector with respect to $\mathbf{w}$:

$$\frac{\partial\tilde{\mathbf{v}}}{\partial\mathbf{w}} = -(\mathbf{v}\otimes\mathbf{w})\frac{\sin(\|\mathbf{w}\|)}{\|\mathbf{w}\|} + \left[(\mathbf{w}\cdot\mathbf{v})I + \mathbf{w}\otimes\mathbf{v}\right]\left(\frac{1-\cos(\|\mathbf{w}\|)}{\|\mathbf{w}\|^2}\right) + (\mathbf{w}\otimes\mathbf{w})\left((\mathbf{w}\cdot\mathbf{v})\frac{2\cos(\|\mathbf{w}\|) - 2 + \|\mathbf{w}\|\sin(\|\mathbf{w}\|)}{\|\mathbf{w}\|^4}\right)$$
$$- [\mathbf{v}]_\times\frac{\sin(\|\mathbf{w}\|)}{\|\mathbf{w}\|} + [(\mathbf{w}\times\mathbf{v})\otimes\mathbf{w}]\frac{\|\mathbf{w}\|\cos(\|\mathbf{w}\|) - \sin(\|\mathbf{w}\|)}{\|\mathbf{w}\|^3}$$

$$\boxed{\begin{aligned} &= -(\mathbf{v} \otimes \mathbf{w} + [\mathbf{v}]_\times)\frac{\sin(\|\mathbf{w}\|)}{\|\mathbf{w}\|} + \big[(\mathbf{w}\cdot\mathbf{v})I + \mathbf{w}\otimes\mathbf{v}\big]\left(\frac{1-\cos(\|\mathbf{w}\|)}{\|\mathbf{w}\|^2}\right) \\ &\quad + (\mathbf{w}\otimes\mathbf{w})\left((\mathbf{w}\cdot\mathbf{v})\frac{2\cos(\|\mathbf{w}\|) - 2 + \|\mathbf{w}\|\sin(\|\mathbf{w}\|)}{\|\mathbf{w}\|^4}\right) + [(\mathbf{w}\times\mathbf{v})\otimes\mathbf{w}]\frac{\|\mathbf{w}\|\cos(\|\mathbf{w}\|) - \sin(\|\mathbf{w}\|)}{\|\mathbf{w}\|^3}, \end{aligned}}$$

where $[\mathbf{v}]_\times$ is the cross product matrix for $\mathbf{v}$. Next, we differentiate again to get the Hessian (a third order tensor whose two "rightmost" slots correspond to the differentiation variables):

$$\begin{aligned} \frac{\partial^2 \tilde{\mathbf{v}}}{\partial \mathbf{w}^2} &= -(\mathbf{v}\otimes I)\frac{\sin(\|\mathbf{w}\|)}{\|\mathbf{w}\|} - \big[(\mathbf{v}\otimes\mathbf{w} + [\mathbf{v}]_\times)\otimes\mathbf{w}\big]\left(\frac{\|\mathbf{w}\|\cos(\|\mathbf{w}\|) - \sin(\|\mathbf{w}\|)}{\|\mathbf{w}\|^3}\right) \\ &\quad + \big[I\otimes\mathbf{v} + \mathbf{e}^i\otimes\mathbf{v}\otimes\mathbf{e}^i\big]\left(\frac{1-\cos(\|\mathbf{w}\|)}{\|\mathbf{w}\|^2}\right) + \big[(\mathbf{w}\cdot\mathbf{v})I + \mathbf{w}\otimes\mathbf{v}\big]\otimes\mathbf{w}\left(\frac{2\cos(\|\mathbf{w}\|) - 2 + \|\mathbf{w}\|\sin(\|\mathbf{w}\|)}{\|\mathbf{w}\|^4}\right) \\ &\quad + \big[\mathbf{e}^i\otimes\mathbf{w}\otimes\mathbf{e}^i + \mathbf{w}\otimes\mathbf{e}^i\otimes\mathbf{e}^i\big]\left((\mathbf{w}\cdot\mathbf{v})\frac{2\cos(\|\mathbf{w}\|) - 2 + \|\mathbf{w}\|\sin(\|\mathbf{w}\|)}{\|\mathbf{w}\|^4}\right) \\ &\quad + \mathbf{w}\otimes\mathbf{w}\otimes\mathbf{v}\left(\frac{2\cos(\|\mathbf{w}\|) - 2 + \|\mathbf{w}\|\sin(\|\mathbf{w}\|)}{\|\mathbf{w}\|^4}\right) \\ &\quad + \mathbf{w}\otimes\mathbf{w}\otimes\mathbf{w}\left((\mathbf{w}\cdot\mathbf{v})\frac{8 + (\|\mathbf{w}\|^2 - 8)\cos(\|\mathbf{w}\|) - 5\|\mathbf{w}\|\sin(\|\mathbf{w}\|)}{\|\mathbf{w}\|^6}\right) \\ &\quad + \big[-\mathbf{e}^i\otimes\mathbf{w}\otimes[\mathbf{v}]_\times^i + (\mathbf{w}\times\mathbf{v})\otimes I\big]\left(\frac{\|\mathbf{w}\|\cos(\|\mathbf{w}\|) - \sin(\|\mathbf{w}\|)}{\|\mathbf{w}\|^3}\right) \\ &\quad + \big[(\mathbf{w}\times\mathbf{v})\otimes\mathbf{w}\otimes\mathbf{w}\big]\left(-\frac{3\|\mathbf{w}\|\cos(\|\mathbf{w}\|) + (\|\mathbf{w}\|^2 - 3)\sin(\|\mathbf{w}\|)}{\|\mathbf{w}\|^5}\right), \end{aligned}$$

were we sum over repeated superscripts ($i \in 0, 1, 2$) and defined $[\mathbf{v}]_\times^i$ to be the vector holding the i^{th} *row* of the cross product matrix $[\mathbf{v}]_\times$:

$$[\mathbf{v}]_\times = \begin{pmatrix} 0 & -v_2 & v_1 \\ v_2 & 0 & -v_0 \\ -v_1 & v_0 & 0 \end{pmatrix} \implies [\mathbf{v}]_\times^0 = \begin{pmatrix} 0 \\ -v_2 \\ v_1 \end{pmatrix},\ [\mathbf{v}]_\times^1 = \begin{pmatrix} v_2 \\ 0 \\ -v_0 \end{pmatrix},\ [\mathbf{v}]_\times^2 = \begin{pmatrix} -v_1 \\ v_0 \\ 0 \end{pmatrix}.$$

We can simplify this Hessian into a form that reveals the expected symmetry with respect to the two rightmost indices:

$$\boxed{\begin{aligned} \frac{\partial^2 \tilde{\mathbf{v}}}{\partial \mathbf{w}^2} &= -(\mathbf{v}\otimes I)\frac{\sin(\|\mathbf{w}\|)}{\|\mathbf{w}\|} - \big[(\mathbf{v}\otimes\mathbf{w}\otimes\mathbf{w} + \mathbf{e}^i\otimes([\mathbf{v}]_\times^i\otimes\mathbf{w} + \mathbf{w}\otimes[\mathbf{v}]_\times^i) + (\mathbf{v}\times\mathbf{w})\otimes I\big]\left(\frac{\|\mathbf{w}\|\cos(\|\mathbf{w}\|) - \sin(\|\mathbf{w}\|)}{\|\mathbf{w}\|^3}\right) \\ &\quad + \big[\mathbf{e}^i\otimes(\mathbf{e}^i\otimes\mathbf{v} + \mathbf{v}\otimes\mathbf{e}^i)\big]\left(\frac{1-\cos(\|\mathbf{w}\|)}{\|\mathbf{w}\|^2}\right) \\ &\quad + \Big[(\mathbf{w}\cdot\mathbf{v})\Big(\mathbf{e}^i\otimes(\mathbf{e}^i\otimes\mathbf{w} + \mathbf{w}\otimes\mathbf{e}^i) + \mathbf{w}\otimes I\Big) + \mathbf{w}\otimes(\mathbf{v}\otimes\mathbf{w} + \mathbf{w}\otimes\mathbf{v})\Big]\left(\frac{2\cos(\|\mathbf{w}\|) - 2 + \|\mathbf{w}\|\sin(\|\mathbf{w}\|)}{\|\mathbf{w}\|^4}\right) \\ &\quad + \mathbf{w}\otimes\mathbf{w}\otimes\mathbf{w}\left((\mathbf{w}\cdot\mathbf{v})\frac{8 + (\|\mathbf{w}\|^2 - 8)\cos(\|\mathbf{w}\|) - 5\|\mathbf{w}\|\sin(\|\mathbf{w}\|)}{\|\mathbf{w}\|^6}\right) \\ &\quad + \big[(\mathbf{v}\times\mathbf{w})\otimes\mathbf{w}\otimes\mathbf{w}\big]\left(\frac{3\|\mathbf{w}\|\cos(\|\mathbf{w}\|) + (\|\mathbf{w}\|^2 - 3)\sin(\|\mathbf{w}\|)}{\|\mathbf{w}\|^5}\right). \end{aligned}}$$

3 Numerically Robust Formulas

The rotation formula and its derivatives must be evaluated with care: around $\mathbf{w} = 0$, a naive implementation would attempt to calculate (approximately) $\frac{0}{0}$ for several of the expressions. In particular, we must use the following Taylor expansions to evaluate the problematic terms for $\|\mathbf{w}\| \ll 1$:

$$\begin{aligned}
\frac{\sin\|\mathbf{w}\|}{\|\mathbf{w}\|} &= 1 - \frac{\|\mathbf{w}\|^2}{6} + O(\|\mathbf{w}\|^4) \\
\frac{1-\cos(\|\mathbf{w}\|)}{\|\mathbf{w}\|^2} &= \frac{1}{2} - \frac{\|\mathbf{w}\|^2}{24} + O(\|\mathbf{w}\|^4) \\
\frac{\|\mathbf{w}\|\cos(\|\mathbf{w}\|) - \sin(\|\mathbf{w}\|)}{\|\mathbf{w}\|^3} &= -\frac{1}{3} + \frac{\|\mathbf{w}\|^2}{30} + O(\|\mathbf{w}\|^4) \\
\frac{2\cos(\|\mathbf{w}\|] - 2 + \|\mathbf{w}\|\sin(\|\mathbf{w}\|)}{\|\mathbf{w}\|^4} &= -\frac{1}{12} + \frac{\|\mathbf{w}\|^2}{180} + O(\|\mathbf{w}\|^4) \\
\frac{8 + (\|\mathbf{w}\|^2 - 8)\cos(\|\mathbf{w}\|) - 5\|\mathbf{w}\|\sin(\|\mathbf{w}\|)}{\|\mathbf{w}\|^6} &= \frac{1}{90} - \frac{\|\mathbf{w}\|^2}{1680} + O(\|\mathbf{w}\|^4) \\
\frac{3\|\mathbf{w}\|\cos(\|\mathbf{w}\|) + (\|\mathbf{w}\|^2 - 3)\sin(\|\mathbf{w}\|)}{\|\mathbf{w}\|^5} &= -\frac{1}{15} + \frac{\|\mathbf{w}\|^2}{210} + O(\|\mathbf{w}\|^4).
\end{aligned}$$

4 Variations around the Identity

Most of the terms in the gradient and Hessian formulas vanish when we evaluate at $\mathbf{w} = 0$. This means that if we update the parametrization at every iteration of Newton's method, we can use much simpler formulas:

$$\begin{aligned}
\left.\frac{\partial \tilde{\mathbf{v}}}{\partial \mathbf{w}}\right|_{\mathbf{w}=0} &= -[\mathbf{v}]_\times, \\
\left.\frac{\partial^2 \tilde{\mathbf{v}}}{\partial \mathbf{w}^2}\right|_{\mathbf{w}=0} &= -(\mathbf{v} \otimes I) + \frac{1}{2}\left[\mathbf{e}^i \otimes (\mathbf{e}^i \otimes \mathbf{v} + \mathbf{v} \otimes \mathbf{e}^i)\right].
\end{aligned}$$

References

[1] F Sebastian Grassia. Practical parameterization of rotations using the exponential map. *Journal of graphics tools*, 3(3):29–48, 1998.

[2] T Kugelstadt, D Koschier, and J Bender. Fast corotated fem using operator splitting. In *Computer Graphics Forum*, volume 37, pages 149–160. Wiley Online Library, 2018.

[3] Camillo J Taylor and David J Kriegman. Minimization on the lie group so (3) and related manifolds. *Yale University*, 16:155, 1994.