Algebraic Line Search for Bundle Adjustment

Julien Michot[1]
julien.michot@cea.fr
Adrien Bartoli[2]
adrien.bartoli@gmail.com
François Gaspard[1]
francois.gaspard@cea.fr

[1] CEA LIST,
Embedded Vision Systems Laboratory,
Point Courrier 94, Gif-sur-Yvette,
F-91191 France
[2] LASMEA,
UMR 6602 CNRS/UBP,
Clermont-Ferrand, France

The Computer Vision community has devoted a great interest to the SfM (Structure-from-Motion) problem for decades. Recovering the 3D scene structure and the camera motion is an important step in a wide range of applications, such as visual SLAM (Simultaneous Localization And Mapping, *e.g.* [1, 5]). This technique recovers the structure of the scene and simultaneously localizes the camera in the generated map. An optimization is performed over a set of parameters that represents the 3D structure (3D primitives) and the cameras. Classical minimization algorithms, such as Levenberg-Marquardt, are used to minimize the reprojection error. It calculates a direction in the parameter space, but does not necessarily provide the best magnitude, called the *step length*. The aim of Line Search techniques [2, 4, 6] is to find an efficient displacement length for a given direction.

We propose a new Line Search technique that we call ALS (Algebraic Line Search). It can easily be plugged in existing bundle adjustment implementations.

The 3D reconstruction problem aims at recovering a model of all camera poses and all the m 3D points $\mathbf{Q}_{j=1..m}$ of a scene from multiple views. Camera projections are written as 3×4 matrices $\mathsf{P}_{i=1..n}$ and can be decomposed in some metric coordinate frame, as $\mathsf{P}_i = \mathsf{K}_i(\mathsf{R}_i|\mathbf{t}_i)$, where K_i encodes the intrinsic parameters and $(\mathsf{R}_i, \mathbf{t}_i)$ represents the orientation and position of the camera in a world coordinate frame. Bundle Adjustment (BA) is based on nonlinear least squares minimization in order to refine the initial structure and camera motion. The objective function to refine is generally is the reprojection error $\varepsilon(\mathbf{x})$. This function is the sum of the squares of the distance between 2D observations (measurements in the images) and reprojections: $\varepsilon(\mathbf{x}) = \sum_{i,j} v_{ij}\, d^2\left(\mathbf{q}_{ij}, \mathsf{P}_i\mathbf{Q}_j\right)$, where $\mathbf{q}_{ij}$ is the observation of point $\mathbf{Q}_j$ in image (camera) P_i, and $v_{ij} = 1$ if the observation exists and 0 otherwise.

The algebraic distance $\tilde{d}$ is defined by: $\tilde{d}(\mathbf{q},\mathbf{q}') = \|\mathsf{S}[\mathbf{q}]_\times \mathbf{q}'\|$, with $\mathsf{S} = \begin{pmatrix}1 & 0 & 0\\ 0 & 1 & 0\end{pmatrix}$ and $[\mathbf{q}]_\times$ is the matrix representation of the vector cross product. It is agreed [3] that under an appropriate normalization, cost functions based on this distance give satisfying results, even if this distance is not geometrically or statistically meaningful. The algebraic cost function is defined as:

$$\tilde{\varepsilon}(\mathbf{x}) = \sum_{i,j} v_{ij} \left\|\mathsf{S}[\mathbf{q}_{ij}]_\times \mathsf{P}_i \mathbf{Q}_j\right\|^2. \qquad (1)$$

Our idea is **not** to use the algebraic distance as the cost function of a bundle adjustment: we keep the basic geometric distance for this. Our proposition is however to use the algebraic distance to find an efficient step length, in a Line Search manner. We call this technique ALS (Algebraic Line Search).

We investigate two different approaches for ALS. The Global ALS is a Line Search technique that aims at finding a global efficient step length for the whole parameter set (camera and scene structure). In a different way, the Two-way ALS determines two distinct step lengths, one for the cameras and the other for the scene structure.

<u>Global Algebraic Line Search (`G-ALS`)</u>

Once the optimization provides us with a step direction $\delta^\top = (\delta_{P_1}^\top, \cdots, \delta_{P_n}^\top, \delta_{Q_1}^\top, \cdots, \delta_{Q_m}^\top)$, with $\delta_{P_i} = vect(\Delta_{P_i})$, we want to determine the step length α. Considering the update rule, the algebraic reprojection error (1) becomes a function of α:

$$\tilde{\varepsilon}(\mathbf{x}+\alpha\delta) = \sum_{i,j} v_{ij} \left\|\mathsf{S}[\mathbf{q}_{ij}]_\times (\mathsf{P}_i + \alpha\Delta_{P_i})(\mathbf{Q}_j + \alpha\delta_{Q_j})\right\|^2. \qquad (2)$$

Since we search for the best step length α^* that minimizes $\tilde{\varepsilon}(\mathbf{x})$ in the previously computed direction δ, the optimums are the real positive solutions of $\frac{\partial\tilde{\varepsilon}}{\partial\alpha} = 0$, given by inspecting the roots of the polynomial:

$$\frac{\partial\tilde{\varepsilon}(\mathbf{x}+\alpha\delta)}{\partial\alpha} = a\alpha^3 + b\alpha^2 + c\alpha + d \qquad (3)$$

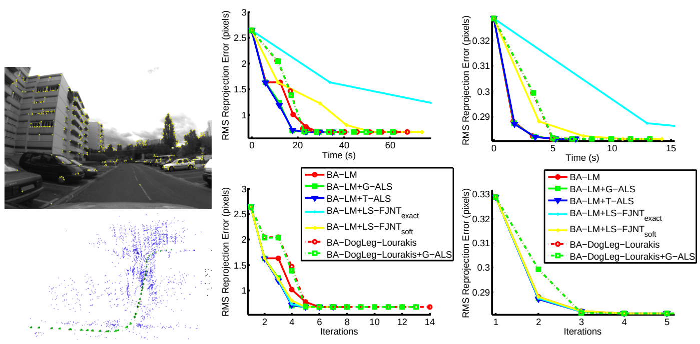
<u>Two-way Algebraic Line Search (`T-ALS`)</u>

Since there are two different types of parameters to refine in common bundle adjustment (the scene structure and the cameras), it sounds attractive to dissociate the step length for each kind of parameters since they do not share the same units. We propose the Two-way ALS that finds two step lengths α_P^* and α_Q^*, respectively for camera and scene structure displacements:

$$\min_{\alpha_P,\alpha_Q} \sum_{i,j} v_{ij} \left\|\mathsf{S}\left[\mathbf{q}_{ij}\right]_\times \left(\mathsf{P}_i + \alpha_P\Delta_{P_i}\right)\left(\mathbf{Q}_j + \alpha_Q\delta_{Q_j}\right)\right\|^2. \qquad (4)$$

The search for the global minimum (α_P^*, α_Q^*) can be performed in a fast way. Nullifying the partial derivatives of equation (4) with respect to α_P and α_Q gives a system of two polynomials in α_P and α_Q which can be solved using Gröbner basis so that α_P^* is the root of a degree 5 polynomial. α_Q^* is then deduced from α_P^*.

One drawback of the algebraic distance is that this error is only an approximation of the Euclidean distance since, so we use the Wolfe Conditions to select efficient step length.



(a) (b) (c)

Figure 1: Results of a video sequence (a). Figures (b) and (c) show the evolution of the RMS (pixels) over time (top) and iterations (bottom), for 30 cameras (b) and 6 cameras (c).

The method can be applied to any problem minimizing a geometric error, such as camera pose estimation and point triangulation. Results showed that the further away the initial solution is from the optimal one, the greater the improvement provided by ALS (typically for $RMS > 1$ pixel).

[1] A. Davison. Real-time simultaneous localisation and mapping with a single camera. In *ICCV*, pages 1403–1410 vol.2, Washington, DC, USA, October 2003. IEEE Computer Society. ISBN 0-7695-1950-4.

[2] P.E. Frandsen, K. Jonasson, H.B. Nielsen, and O. Tingleff. *Unconstrained Optimization*. 1999.

[3] R. Hartley and A. Zisserman. *Multiple view geometry in computer vision*. Cambridge University Press, Cambridge, UK, first edition, 2003.

[4] J. Moré and D. Thuente. Line search algorithms with guaranteed sufficient decrease. *ACM Trans. Math. Softw.*, 20(3):286–307, 1994. ISSN 0098-3500. doi: http://doi.acm.org/10.1145/192115.192132.

[5] E. Mouragnon, M. Lhuiller, M. Dhome, F. Dekeyser, and P. Sayd. Generic and real-time structure from motion. In *BMVC*, 2007.

[6] J. Nocedal and Y. Yuan. Combining trust region and line search techniques. Technical report, Advances in Nonlinear Programming, (Kluwer), 1992.