Revisiting the Variable Projection Method for Separable Nonlinear Least Squares Problems: Supplementary Document

Je Hyeong Hong
University of Cambridge
jhh37@cam.ac.uk

Christopher Zach
Toshiba Research Europe
christopher.m.zach@gmail.com

Andrew Fitzgibbon
Microsoft
awf@microsoft.com

1. Video for 3D reconstruction and camera localization per iteration

We provide a video (https://youtu.be/f5zy-suk5EI), which shows the evolution of 3D structure and affine camera directions at every successful iteration of VarPro and Joint+EPI on the *trimmed dinosaur* [2] and *skeleton Dubrovnik* [1] datasets (see [3] for details). Since we use affine cameras, which are theoretically infinite distance away from the 3D points, we plot their directions rather than their positions.

As affine bundle adjustment (BA) works in the projective (non-metric) frame, we visualize the upgraded solution at each iteration by finding the ambiguity matrix which minimizes the L^2-Euclidean distance between the quasi-ground truth 3D structure and its estimate. For the *trimmed dinosaur* dataset, the quasi-ground truth 3D structure is obtained by triangulating points with the ground truth camera poses, which can be downloaded from the University of Oxford VGG website (http://www.robots.ox.ac.uk/ vgg/data/data-mview.html). For the *skeleton Dubrovnik* dataset, we take an optimal 3D structure, which is obtained by performing robust bundle adjustment with good initialization, as semi-ground truth values.

Figure 1: An image from the *trimmed dinosaur* sequence

2. Deriving $d\mathbf{v}^*(\mathbf{u})/d\mathbf{u}$ for VarPro

In §3 of [3], we have

$$\frac{d\mathbf{v}^*(\mathbf{u})}{d\mathbf{u}} = \frac{d\left[\mathtt{G}(\mathbf{u})^\dagger\right]\mathbf{z}(\mathbf{u})}{d\mathbf{u}} + \mathtt{G}(\mathbf{u})^\dagger \frac{d\mathbf{z}(\mathbf{u})}{d\mathbf{u}}. \tag{1}$$

The following rule is used to differentiate a pseudo-inverse matrix $\mathtt{A}^\dagger$:

$$\partial \mathtt{A}^\dagger = -\mathtt{A}^\dagger \partial[\mathtt{A}]\mathtt{A}^\dagger + (\mathtt{A}^\top \mathtt{A})^{-1}\partial[\mathtt{A}]^\top(\mathtt{I} - \mathtt{A}\mathtt{A}^\dagger). \tag{2}$$

Applying (2) to the first term of (1) yields

$$\begin{aligned}\partial\left[\mathtt{G}(\mathbf{u})^\dagger\right]\mathbf{z}(\mathbf{u}) &= -\mathtt{G}(\mathbf{u})^\dagger\partial[\mathtt{G}(\mathbf{u})]\mathtt{G}(\mathbf{u})^\dagger\mathbf{z}(\mathbf{u}) + (\mathtt{G}(\mathbf{u})^\top\mathtt{G}(\mathbf{u}))^{-1}\partial[\mathtt{G}(\mathbf{u})]^\top(\mathtt{I} - \mathtt{G}(\mathbf{u})\mathtt{G}(\mathbf{u})^\dagger)\mathbf{z}(\mathbf{u}) \\ &= -\mathtt{G}(\mathbf{u})^\dagger\partial[\mathtt{G}(\mathbf{u})]\mathbf{v}^*(\mathbf{u}) - (\mathtt{G}(\mathbf{u})^\top\mathtt{G}(\mathbf{u}))^{-1}\partial[\mathtt{G}(\mathbf{u})]^\top\boldsymbol{\varepsilon}^*(\mathbf{u}),\end{aligned} \tag{3}$$

leading to

$$\begin{aligned}\partial\mathbf{v}^*(\mathbf{u}) &= \partial\left[\mathtt{G}(\mathbf{u})^\dagger\right]\mathbf{z}(\mathbf{u}) + \mathtt{G}(\mathbf{u})^\dagger\partial\mathbf{z}(\mathbf{u}) \\ &= -\mathtt{G}(\mathbf{u})^\dagger\partial[\mathtt{G}(\mathbf{u})]\mathbf{v}^*(\mathbf{u}) + \mathtt{G}(\mathbf{u})^\dagger\partial\mathbf{z}(\mathbf{u}) - (\mathtt{G}(\mathbf{u})^\top\mathtt{G}(\mathbf{u}))^{-1}\partial[\mathtt{G}(\mathbf{u})]^\top\boldsymbol{\varepsilon}^*(\mathbf{u}).\end{aligned} \tag{4}$$

By noting that

$$\mathtt{J}_{\mathbf{u}}(\mathbf{u},\mathbf{v}) = \frac{\partial \boldsymbol{\varepsilon}(\mathbf{u},\mathbf{v})}{\partial \mathbf{u}} = \frac{\partial [\mathtt{G}(\mathbf{u})]}{\partial \mathbf{u}} \mathbf{v}^*(\mathbf{u}) - \frac{\partial \mathbf{z}(\mathbf{u})}{\partial \mathbf{u}}, \tag{5}$$

(4) becomes

$$\partial \mathbf{v}^*(\mathbf{u}) = -\mathtt{G}(\mathbf{u})^{\dagger} \partial \boldsymbol{\varepsilon}(\mathbf{u}, \mathbf{v}^*(\mathbf{u})) - (\mathtt{G}(\mathbf{u})^{\top}\mathtt{G}(\mathbf{u}))^{-1} \partial [\mathtt{G}(\mathbf{u})]^{\top} \boldsymbol{\varepsilon}^*(\mathbf{u}). \tag{6}$$

Finally, using (5) and the equality $\mathtt{G}(\mathbf{u}) = \mathtt{J}_{\mathbf{v}}(\mathbf{u})$ yields

$$\begin{aligned} \frac{d\mathbf{v}^*(\mathbf{u})}{d\mathbf{u}} &= -\mathtt{G}(\mathbf{u})^{\dagger} \frac{\partial \boldsymbol{\varepsilon}(\mathbf{u}, \mathbf{v}^*(\mathbf{u}))}{\partial \mathbf{u}} - (\mathtt{G}(\mathbf{u})^{\top}\mathtt{G}(\mathbf{u}))^{-1} \frac{d[\mathtt{G}(\mathbf{u})]^{\top} \boldsymbol{\varepsilon}^*(\mathbf{u})}{d\mathbf{u}} \\ &= -\mathtt{J}_{\mathbf{v}}(\mathbf{u})^{\dagger} \mathtt{J}_{\mathbf{u}}(\mathbf{u}, \mathbf{v}^*(\mathbf{u})) - (\mathtt{J}_{\mathbf{v}}(\mathbf{u})^{\top}\mathtt{J}_{\mathbf{v}}(\mathbf{u}))^{-1} \frac{d[\mathtt{J}_{\mathbf{v}}(\mathbf{u})]^{\top} \boldsymbol{\varepsilon}^*(\mathbf{u})}{d\mathbf{u}}. \end{aligned} \tag{7}$$

3. A unified pseudocode table of algorithms

The algorithm chart derived from §4 is shown in Table 1.

Joint	Joint+EPI	VarPro	
			inputs: $\mathbf{u} \in \mathbb{R}^p, \mathbf{v} = \arg\min_{\mathbf{v}} \\|\boldsymbol{\varepsilon}(\mathbf{u},\mathbf{v})\\|_2^2 \in \mathbb{R}^q$
•	•	•	$[\lambda_{\mathbf{u}}\ ;\ \lambda_{\mathbf{v}}] \leftarrow [10^{-4}\ ;\ 10^{-4}]$
		•	$\lambda_{\mathbf{v}} \leftarrow 0$
•	•	•	**repeat**
•	•	•	**repeat**
•	•	•	$\mathbf{g} \leftarrow \mathtt{J}_{\mathbf{u}}(\mathbf{u},\mathbf{v})^{\top}(\mathtt{I} - \mathtt{J}_{\mathbf{v}}(\mathbf{u})\mathtt{J}_{\mathbf{v}}(\mathbf{u})^{-\lambda_{\mathbf{v}}})\boldsymbol{\varepsilon}(\mathbf{u},\mathbf{v})$
•	•	•	$\boldsymbol{\Delta}\mathbf{u} \leftarrow -\left(\mathtt{J}_{\mathbf{u}}(\mathbf{u},\mathbf{v})^{\top}(\mathtt{I} - \mathtt{J}_{\mathbf{v}}(\mathbf{u})\mathtt{J}_{\mathbf{v}}(\mathbf{u})^{-\lambda_{\mathbf{v}}})\mathtt{J}_{\mathbf{u}}(\mathbf{u},\mathbf{v}) + \lambda_{\mathbf{u}}\mathtt{I}\right)^{-1}\mathbf{g}$
•			$\boldsymbol{\Delta}\mathbf{v} \leftarrow \arg\min_{\boldsymbol{\Delta}\mathbf{v}} \\|\boldsymbol{\varepsilon}(\mathbf{u},\mathbf{v}) + \mathtt{J}_{\mathbf{u}}(\mathbf{u},\mathbf{v})\boldsymbol{\Delta}\mathbf{u} + \mathtt{J}_{\mathbf{v}}(\mathbf{u})\boldsymbol{\Delta}\mathbf{v}\\|_2^2 + \lambda_{\mathbf{v}}\\|\boldsymbol{\Delta}\mathbf{v}\\|_2^2$
	•	•	$\boldsymbol{\Delta}\mathbf{v} \leftarrow \arg\min_{\boldsymbol{\Delta}\mathbf{v}} \\|\boldsymbol{\varepsilon}(\mathbf{u}+\boldsymbol{\Delta}\mathbf{u},\mathbf{v}) + \mathtt{J}_{\mathbf{v}}(\mathbf{u}+\boldsymbol{\Delta}\mathbf{u})\boldsymbol{\Delta}\mathbf{v}\\|_2^2$
•	•	•	$[\lambda_{\mathbf{u}}\ ;\ \lambda_{\mathbf{v}}] \leftarrow 10\,[\lambda_{\mathbf{u}}\ ;\ \lambda_{\mathbf{v}}]$
•	•	•	**until** $\\|\boldsymbol{\varepsilon}(\mathbf{u}+\boldsymbol{\Delta}\mathbf{u},\mathbf{v}+\boldsymbol{\Delta}\mathbf{v})\\|_2^2 < \\|\boldsymbol{\varepsilon}(\mathbf{u},\mathbf{v})\\|_2^2$
•	•	•	$[\mathbf{u}\ ;\ \mathbf{v}] \leftarrow [\mathbf{u}\ ;\ \mathbf{v}] + [\boldsymbol{\Delta}\mathbf{u}\ ;\ \boldsymbol{\Delta}\mathbf{v}]$
•	•	•	$[\lambda_{\mathbf{u}}\ ;\ \lambda_{\mathbf{v}}] \leftarrow 0.01\,[\lambda_{\mathbf{u}}\ ;\ \lambda_{\mathbf{v}}]$
•	•	•	**until** convergence
			output: $\mathbf{u} \in \mathbb{R}^p, \mathbf{v} \in \mathbb{R}^q$

Table 1: A unified pseudocode table for the methods illustrated in [3]. Joint = Joint Optimization, Joint+EPI = Joint Optimization with Embedded Point Iterations, VarPro = Variable Projection with RW2 approximation. All methods are based on the Levenberg-Marquardt (LM) algorithm. For simplicity, we set initial $\mathbf{v}$ to be optimal for the initial value of $\mathbf{u}$.

References

[1] S. Agarwal, N. Snavely, S. M. Seitz, and R. Szeliski. Bundle adjustment in the large. In *11th European Conference on Computer Vision (ECCV): Part II*, pages 29–42, 2010. 1

[2] A. M. Buchanan and A. W. Fitzgibbon. Damped Newton algorithms for matrix factorization with missing data. In *2005 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 2, pages 316–322, 2005. 1

[3] J. H. Hong, C. Zach, and A. W. Fitzgibbon. Revisiting the variable projection method for separable nonlinear least squares problems. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017. 1, 2