

# Power Variable Projection for Initialization-Free Large-Scale Bundle Adjustment

Simon Weber<sup>1,2</sup>    Je Hyeong Hong<sup>3</sup>    Daniel Cremers<sup>1,2</sup>

<sup>1</sup> Technical University of Munich

<sup>2</sup> Munich Center for Machine Learning

<sup>3</sup> Department of Electronic Engineering, Hanyang University

**Abstract.** Most Bundle Adjustment (BA) solvers like the Levenberg-Marquard algorithm require a good initialization. Instead, initialization-free BA remains a largely uncharted territory. The under-explored Variable Projection algorithm (VarPro) exhibits a wide convergence basin even without initialization. Coupled with object space error formulation, recent works have shown its ability to solve small-scale initialization-free bundle adjustment problem. To make such initialization-free BA approaches scalable, we introduce Power Variable Projection (PoVar), extending a recent inverse expansion method based on power series. Importantly, we link the power series expansion to Riemannian manifold optimization. This projective framework is crucial to solve large-scale bundle adjustment problems without initialization. Using the real-world BAL dataset, we experimentally demonstrate that our solver achieves state-of-the-art results in terms of speed and accuracy. To our knowledge, this work is the first to address the scalability of BA without initialization opening new venues for initialization-free structure-from-motion.

**Keywords:** Bundle Adjustment · Initialization-Free · Schur Complement · Riemannian Manifold Optimization

## 1 Introduction

Bundle adjustment (BA) is the key component of many structure-from-motion and 3D reconstruction algorithms. With the recent emergence of large-scale internet photo collections [3] and new applications (mixed reality, autonomous driving, digital twins), the need to solve large-scale BA has become an important challenge. Traditional BA addresses the following question: *Given image measurements and approximate landmark positions and camera parameters, can we derive the exact positions and parameters?* The gold standard is to use the Levenberg-Marquardt algorithm [25] coupled with the Schur complement trick and a scalable solver for the reduced camera system, which is often the preconditioned conjugate gradient algorithm. Recent work [23] achieves outstanding speed for large-scale BA by using a power series expansion of the inverse Schur complement.

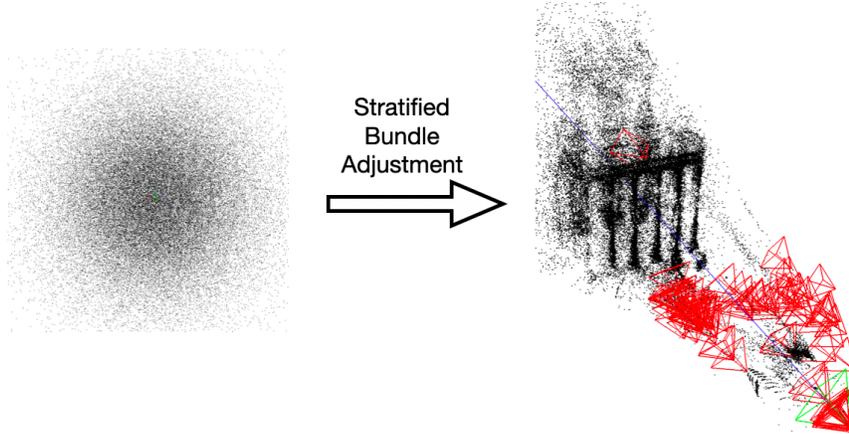
Recently, a new line of works [11, 13, 14] has attempted to solve the BA problem *without* careful initialization: *Given only image measurements, how do we derive pose parameters and 3D landmark positions?* This challenge is largely uncharted, and its scalability a blind spot. In particular, most existing works aim to formalize the problem into a stratified BA formulation, and none of them try to design effective solvers. It is noteworthy that even the most recent works only use direct factorization which becomes impractical for large-scale problems with several hundreds of cameras. In contrast to the traditional BA problem, the deficiency of competitive solvers for initialization-free BA can be broadly explained by the difference of convergence behaviour between a well-initialized problem and an initialization-free problem.

Following up on the recent findings concerning inverse expansion methods, we address the scalability of initialization-free BA. Our new solver based on the Variable Projection algorithm overcomes the issues of convergence of the scalable preconditioned conjugate gradients algorithm, while being efficient for thousand of camera viewpoints. In summary, we make the following contributions:

- We introduce Power Variable Projection (*PoVar*) for efficient large-scale bundle adjustment *without* good initialization of camera poses and 3D landmarks. To the best of our knowledge, we are the first to address the scalability of initialization-free bundle adjustment formulation.
- We provide theoretical proofs that justify the extension of recent *inverse expansion* method to the variable projection algorithm. While sharing a close algorithmic structure, the proposed extension and the existing *power-series-based* method largely differ in the theory, in the applications and in the convergence behaviour.
- We theoretically extend the power series expansion for bundle adjustment to Riemannian manifold optimization. We take advantage of the matrix-specific structure to propose an efficient storage and memory-efficient computation for such optimization.
- We perform extensive evaluation of the proposed approach on the real-world BAL dataset. We emphasize the benefits of *PoVar* in terms of scalability, speed and accuracy. In contrast to state-of-the-art solvers, our work is the first that solves large-scale bundle adjustment without initialization.
- We release our solver as open source to facilitate further research: <https://github.com/tum-vision/povar>.

## 2 Related Work

As we address the scalability of the variable projection (VarPro) algorithm for initialization-free bundle adjustment (BA), we review works on VarPro and on BA from arbitrary initialization. We also provide some background on the inverse expansion methods. A more general description of BA can be found in [22].



**Fig. 1:** In contrast to traditional bundle adjustment problem, initialization-free BA is a largely under-explored problem. It does not assume any approximation of pose and landmark parameters, making the problem much harder to solve. From a random initialization (*left figure*), and given only image measurements, we aim to recover pose and landmark parameters. Our approach, that extends inverse expansion method, is motivated by the lack of scalability of existing solvers. On the real-world BAL problems (e.g. *Venice-89*, *right figure*), we demonstrate the efficiency of the proposed combination of our novel solver Power Variable Projection (PoVar) and Riemannian manifold optimization framework for expansion method to solve the stratified BA problem.

**Variable projection (VarPro) algorithm.** VarPro is an optimization approach for solving bivariate problems that can be formulated as minimizing a cost function  $f(u, v)$  with  $f : \mathbb{R}^m \times \mathbb{R}^n \rightarrow \mathbb{R}$  over two sets of variables  $u \in \mathbb{R}^m$  and  $v \in \mathbb{R}^n$ . Unlike alternation, which fixes  $u$  and optimizes over  $v$  and vice versa, or joint optimization, which optimizes the stack of  $u$  and  $v$  simultaneously, variable projection replaces  $v$  with  $v^*(u) := \arg \min_v f(u, v)$  ( $v$  always optimal over  $u$ ) such that the optimized cost function  $f(u, v^*(u)) =: f^*(u)$  becomes a function of  $u$  only. The original VarPro algorithm by Golub and Pereyra [9] and its approximations by Ruhe and Wedin [19] assume solving a separable nonlinear least squares (SNLS) problem where  $v^*(u)$  can be obtained in closed form. VarPro was consistently ignored in the computer vision community and even misidentified as a form of alternating optimization [5]. It first receives proper attention when Okatani et al. [17] demonstrates VarPro equipped with a trust-region approach such as Levenberg-Marquardt can yield a wide basin of convergence for several toy problems that can be formulated as a SNLS problem such as affine structure-from-motion and factorization-based non-rigid structure-from-motion. Shortly after, Strelow [20, 21] extends VarPro to the nonlinear case where  $v^*(u)$  is not in closed form such as bundle adjustment. Nevertheless, these works do not address the issue of increased algorithmic complexity. Later, it has been shown

by Hong et al. [12] that VarPro can be efficiently implemented by performing inner iterations [2] (also known as embedded point iterations [15]) over the set of eliminated variables  $u$  followed by performing a joint optimization step with no damping on  $u$ . While this enables faster runtime compared to previous studies, it has only been tested up to small-medium sized problems with around 300 camera views. To this date, no work to the best of our knowledge has improved the scalability of the VarPro algorithm beyond [12].

**Initialization-free bundle adjustment.** While traditional (large-scale) bundle adjustment is regularly studied ([3, 4, 6, 7, 18, 24, 28]), initialization-free BA is a recent research topic. In a seminal work, Hong et al. [13] propose to solve projective bundle adjustment from arbitrary initialization with the Variable Projection algorithm. Notably, they propose a stratified bundle adjustment formulation with increasing difficulty. In a follow-up work, pOSE [11] incorporates the equivalence between nonlinear VarPro and the Schur complement recently identified in [12]. Additionally, they define an objective in-between affine and projective models that leads to a wide convergence basin. Iglesias et al. [14] complements this *pseudo object space error* formulation with an exponential regularization term, and show interesting results on very-small-scale problems. Nevertheless, to our knowledge, none of these studies look closely into the scalability of the proposed frameworks for initialization-free bundle adjustment.

**Inverse expansion method.** Despite its short recent appearance in the literature, inverse expansion method is a highly efficient and competitive approach for solving the linearized system of equations for bundle adjustment that is already challenging established factorization and iterative methods. It links the Schur complement [26] to the power series expansion of its inverse. Using the *power Schur complement* as a preconditioner for normal equations leads to good results for physics problems such as convection-diffusion [27]. PoBA [23] proposes to directly apply the power Schur complement to the right-hand side of the reduced camera system. Their solver results in improved speed and accuracy for traditional bundle adjustment formulation with respect to the existing methods.

### 3 Problem Statement and Motivation

We consider a typical form of bundle adjustment. Given observations  $m_{ij}$  for pose  $i$  and landmark  $j$ ,  $K_i, R_i \in SO(3)$ ,  $t_i \in \mathbb{R}^3$  the intrinsics, rotation and translation of pose  $i$ , and  $x_j \in \mathbb{R}^3, \tilde{x}_j \in S^4$  the inhomogeneous and homogeneous landmark 3D positions (where  $S^n$  denotes the set of all vectors on the unit  $n$ -sphere), we aim to solve:

$$\min_{\{R_i\}_i \in SO(3), \{t_i\}, \{\tilde{x}_j\}} \sum_{(i,j) \in \Omega} \|\pi(K_i[R_i|t_i]\tilde{x}_j) - m_{ij}\|_2^2, \quad (1)$$

where  $\pi$  is the perspective projection  $\pi([x, y, z]^T) := [x/z, y/z]^T$ .

Recently, Eq. 1 with good initialization has been solved efficiently with a novel inverse expansion method.

### 3.1 Inverse expansion method

Inverse expansion method [23, 27] relates on the expansion of the inverse of a matrix into a power series, as stated in the following proposition:

**Proposition 1.** *Let  $M$  be an  $n \times n$  matrix. If the spectral radius of  $M$  satisfies  $\|M\| < 1$ , then*

$$(I - M)^{-1} = \sum_{i=0}^m M^i + R, \quad (2)$$

where the error matrix

$$R = \sum_{i=m+1}^{\infty} M^i, \quad (3)$$

satisfies

$$\|R\| \leq \frac{\|M\|^{m+1}}{1 - \|M\|}. \quad (4)$$

Given a linear approximation of Eq. 1 followed by the Schur complement trick, Weber et al. [23] relates the inverse Schur complement of the Levenberg-Marquardt algorithm to its power series. They show significant improvement in terms of speed and accuracy to solve BA problem with good initialization. In addition to this empirical insights, some convergence behaviour concerning the approximated results for the BA problem are theoretically proved.

However, it is well-known in the literature (see e.g. [13]) that solving Eq. 1 from arbitrary initialization is non-feasible. Hong et al. [11] override this challenge by proposing a stratified bundle adjustment problem. Let us revisit the formulation of this approach.

### 3.2 Initialization-free bundle adjustment

The stratified BA problem is decomposed in two minimization problems, followed by a metric upgrade. Although we consider a pinhole camera model, the first stage assumes a projective model.

#### First stage: separable nonlinear optimization with projective camera.

Given  $n_p$  poses and  $n_l$  landmarks,  $x = (x_p, x_l)$  contains all the optimization variables. We model the camera as a projective one. For pose  $i$ , we consider the camera parameters  $x_p^i \in \mathbb{R}^{3 \times 4}$  and solve the following generic nonlinear separable problem:

$$\min_{x_p, \tilde{x}_l} F(x_p, \tilde{x}_l) = \|r(x_p, \tilde{x}_l)\|_2^2 = \|G(x_p)\tilde{x}_l - z(x_p)\|_2^2, \quad (5)$$

with  $G(\cdot)$  and  $z(\cdot)$  some linear operators, and the last coefficient of  $\tilde{x}_l^j$  fixed to 1. For instance, pOSE [11] – extensively used in our analysis (see Supplemental), proposes the following minimization problem:

$$F_{pOSE}(x_p, x_l) = \sum_{(i,j) \in \Omega} \left\| \frac{\sqrt{1-\eta}(x_p^{i,1:2}\tilde{x}_l^j - (x_p^{i,3}\tilde{x}_l^j)m_{ij})}{\sqrt{\eta}(x_p^{i,1:2}\tilde{x}_l^j - m_{ij})} \right\|_2^2, \quad (6)$$

with  $x_p^{i,1:2}$  and  $x_p^{i,3}$  respectively the first two rows and the third row of  $x_p^i$ , and  $\eta \in [0, 1]$ .

Hong et al. [12] argue the superiority of VarPro over joint optimization to solve the previous equation, due to the random initialization of this stage.

**Second stage: projective refinement.** The cameras and landmarks parameters obtained by solving Eq. 5 are refined by minimizing the projective standard objective [13] over the projective camera models in homogeneous form ( $\{\tilde{x}_p^i \mid \text{vec}(\tilde{x}_p^i) \in S^{12}\}$ ) and the 3D landmarks in homogeneous coordinates ( $\{\tilde{x}_l^j \mid \tilde{x}_l^j \in S^4\}$ ):

$$\sum_{(i,j) \in \Omega} \|\pi(\tilde{x}_p^i \tilde{x}_l^j) - m_{ij}\|_2^2. \quad (7)$$

Importantly, the optimization is performed in homogeneous coordinates. Especially, Riemannian manifold optimization [1] has to be incorporated.

**Metric upgrade.** The last stage is a minimization problem to enforce the projective camera matrices to satisfy  $SE(3)$  properties. Note that our work mostly focuses on the first two stages. The proposed implementation for this third stage has illustrative purpose, see Fig. 1. We refer the reader to Supplemental for further details.

### 3.3 Limitations and proposed method

In contrast to the Levenberg-Marquardt algorithm, few solvers have been designed to efficiently solve VarPro. In practice, even the most recent works [11, 14] use a direct factorization (e.g. Cholesky decomposition, QR factorization), that is well-known to be poorly scalable (see e.g. [3]), to solve Eq. 5. In particular, we note that these works consider bundle adjustment problems with only few tens of cameras – sometimes less than ten. On the other hand, Hong and Fitzgibbon [10] show that coupling VarPro with the popular preconditioned conjugate gradients algorithm may not converge efficiently.

We propose to build on recent inverse expansion method. We first adapt the power series expansion to the VarPro algorithm (Section 4.2). We show that the so-called *PoVar* efficiently solves Eq. 5. Moving forward, we extend the power series expansion to Riemannian manifold optimization (Section 4.3). This new Riemannian framework, that we call *RiPoBA*, is necessary to use expansion method for solving Eq. 7. We demonstrate that the combination of this two solvers is highly competitive (Section 5).

## 4 Power Variable Projection

We start by revisiting the VarPro algorithm. We refer to [9] for further details.

### 4.1 Revisited variable projection

In contrast to joint optimization, VarPro optimizes over landmark parameters and camera parameters, but in a way different from the standard alternating least squares such that landmarks are not assumed to be fixed when updating the camera parameters. It first considers Eq. 5 as a nonlinear minimization problem over  $\tilde{x}_l$  only. Due to the separability of the equation, a closed-form solution for optimal  $\tilde{x}_l^*(x_p)$  is straightforward:

$$\tilde{x}_l^*(x_p) = \arg \min_{\tilde{x}_l} \|G(x_p)\tilde{x}_l - z(x_p)\|_2^2 = G(x_p)^\dagger z(x_p), \quad (8)$$

with  $G(x_p)^\dagger$  the pseudo-inverse of  $G(x_p)$ . As we set the last coordinate of  $\tilde{x}_l$  to 1, we normalize  $\tilde{x}_l^*(x_p)$  by its last coordinate and its substitution in Eq. 5 leads to the following reduced problem:

$$\min_{x_p} r^*(x_p) = \min_{x_p} \|(G(x_p)G^\dagger(x_p) - I)z(x_p)\|_2^2. \quad (9)$$

Following [12], Eq. 9 can be solved with LM algorithm over  $x_p$ . The Jacobian of  $r^*$  is approximated with the so-called RW2 approximation [16], that leads to the normal equation:

$$\begin{pmatrix} U_\lambda & W \\ W^\top & V_0 \end{pmatrix} \begin{pmatrix} \Delta x_p \\ \Delta x_l \end{pmatrix} = - \begin{pmatrix} b_p \\ b_l \end{pmatrix}, \quad (10)$$

where

$$U_\lambda = J_p^\top J_p + \lambda D_p^\top D_p, \quad (11)$$

$$V_0 = J_l^\top J_l, \quad W = J_p^\top J_l, \quad (12)$$

$$b_p = J_p^\top r^0, \quad b_l = J_l^\top r^0, \quad (13)$$

with  $J_l$  and  $J_p$  respectively the landmark and pose Jacobians of the original residual (Eq. 5) around an equilibrium  $r^0$ , and  $D_p$  diagonal damping matrix for pose variables. In contrast to joint optimization, only the pose Jacobian is damped in the associated Hessian. It follows that  $U_\lambda$  is symmetric positive-definite [22], whereas  $V_0$  is only guaranteed to be symmetric positive-semidefinite. Nevertheless, we observe  $V_0$  is usually of full rank unless a 3D landmark is observed by few cameras with narrow baselines. By using the Schur complement trick, the update equation for VarPro becomes:

$$(U_\lambda - W V_0^{-1} W^\top) \Delta x_p = b_p - W V_0^{-1} b_l. \quad (14)$$

Note that the Schur complement associated to VarPro:

$$S^V = U_\lambda - W V_0^{-1} W^\top, \quad (15)$$

while sharing a close structure to the Schur complement of the traditional BA problem, has a different convergence behaviour, due to the undamped landmark Jacobian  $V_0$ .

#### 4.2 Power series for VarPro

Inspired by Weber et al. [23], the following lemma holds, even if  $V_0$  is only symmetric positive-semidefinite:

**Lemma 1.** *Let  $\mu$  be an eigenvalue of  $U_\lambda^{-1}WV_0^\dagger W^\top$ . Then*

$$0 \leq \mu < 1. \quad (16)$$

*Proof.* We refer the reader to the Supplemental.

It follows that the pose update  $\Delta x_p$  in Eq. 14 can be directly approximated<sup>4</sup> by  $x(m)$  with Proposition 1 applied to the Schur complement  $S^V$ :

$$x(m) = - \sum_{i=0}^m (U_\lambda^{-1}WV_0^{-1}W^\top)^i U_\lambda^{-1}(b_p - WV_0^{-1}b_l). \quad (17)$$

Once the pose update is estimated, the landmark update is derived following closed-form Eq. 8. That extends the inverse expansion method to Variable Projection algorithm. The difference between both solvers is the role of the damped parameters. While it seems slight, this is enough to offer a dissimilar convergence behaviour, as we will see in the experiments.

Before that, let us investigate Riemannian manifold optimization, that is necessary to solve the second stage of the stratified BA problem. In particular, we show that we can link this framework to expansion method.

#### 4.3 Power Riemannian manifold optimization

As the projective refinement step in Eq. 7 involves both camera matrices and 3D landmarks in homogeneous forms, we exhibit local scale freedom for both camera and landmark parameters. This necessitates incorporation of the Riemannian manifold optimization framework without which the linearized system of equations is always rank-deficient and unsolvable. While a complete theoretical overview of such framework can be found in [1], we formalize Eq. 7 as:

$$\arg \min_{\Delta \tilde{x}_p, \Delta \tilde{x}_l} \|f(\tilde{x}_p + \Delta \tilde{x}_p, \tilde{x}_l + \Delta \tilde{x}_l)\|_2^2, \quad (18)$$

where  $\tilde{x}_p \in \mathbb{R}^{12n_p}$  denotes the stack of vectorized homogeneous camera parameters,  $\tilde{x}_l \in \mathbb{R}^{4n_l}$  denotes the stack of homogeneous 3D landmarks and  $\Delta \tilde{x}_p \in \mathbb{R}^{12n_p}$  and  $\Delta \tilde{x}_l \in \mathbb{R}^{4n_l}$  are the updates in homogeneous camera parameters and 3D

<sup>4</sup> As in practice  $V_0$  is full-rank, and to avoid encumbering notations, we write  $V_0^{-1}$  instead of the pseudo-inverse in the rest of the paper.

landmarks respectively. The unknowns are searched in the tangent space of current  $\tilde{x} = [\tilde{x}_p^\top, \tilde{x}_l^\top]^\top$ , that we note  $\tilde{x}^\perp \in \mathbb{R}^{(12n_p+4n_l) \times (11n_p+3n_l)}$  such that  $(\tilde{x}^\perp)^\top \tilde{x} = 0$ . To simplify notations,  $\tilde{x}_p$  is considered as a vector in this section, and  $\tilde{x}_p^\perp \in \mathbb{R}^{12n_p \times 11n_p}$ ,  $\tilde{x}_l^\perp \in \mathbb{R}^{4n_l \times 3n_l}$  are block-diagonal, each block corresponding to the associated pose  $i$  and landmark  $j$ , respectively. The projective refinement becomes:

$$\arg \min_{\Delta \tilde{x} \perp \tilde{x}} \|f(\tilde{x}_p + \Delta \tilde{x}_p, \tilde{x}_l + \Delta \tilde{x}_l)\|_2^2. \quad (19)$$

By coupling Riemannian manifold optimization and LM algorithm, and according to [1], we get the following normal equation, projected onto the tangent space of  $\tilde{x}$ :

$$\begin{pmatrix} (\tilde{x}_p^\perp)^\top U_\lambda \tilde{x}_p^\perp & (\tilde{x}_p^\perp)^\top W \tilde{x}_l^\perp \\ (\tilde{x}_l^\perp)^\top W^\top \tilde{x}_p^\perp & (\tilde{x}_l^\perp)^\top V_\lambda \tilde{x}_l^\perp \end{pmatrix} \begin{pmatrix} \Delta x_p \\ \Delta x_l \end{pmatrix} = - \begin{pmatrix} (\tilde{x}_p^\perp)^\top b_p \\ (\tilde{x}_l^\perp)^\top b_l \end{pmatrix}. \quad (20)$$

where  $\Delta x_p \in \mathbb{R}^{11n_p}$  and  $\Delta x_l \in \mathbb{R}^{3n_l}$  are the camera update and the landmark update respectively made on the tangent space of  $x$ . By keeping coherent notations we can note the projected Jacobians and the projected damping parameters onto the tangent space of  $x$  as:

$$\tilde{J}_p = J_p \tilde{x}_p^\perp, \quad \tilde{J}_l = J_l \tilde{x}_l^\perp, \quad \tilde{\lambda}_p = (\tilde{x}_p^\perp)^\top \lambda \tilde{x}_p^\perp, \quad \tilde{\lambda}_l = (\tilde{x}_l^\perp)^\top \lambda \tilde{x}_l^\perp, \quad (21)$$

and then Eq. 20 becomes:

$$\begin{pmatrix} \tilde{U}_{\tilde{\lambda}} & \tilde{W} \\ \tilde{W}^\top & \tilde{V}_{\tilde{\lambda}} \end{pmatrix} \begin{pmatrix} \Delta x_p \\ \Delta x_l \end{pmatrix} = - \begin{pmatrix} \tilde{b}_p \\ \tilde{b}_l \end{pmatrix}, \quad (22)$$

where

$$\tilde{U}_{\tilde{\lambda}} = \tilde{J}_p^\top \tilde{J}_p + D_p^\top \tilde{\lambda}_p D_p, \quad (23)$$

$$\tilde{V}_{\tilde{\lambda}} = \tilde{J}_l^\top \tilde{J}_l + D_l^\top \tilde{\lambda}_l D_l, \quad (24)$$

$$W = \tilde{J}_p^\top \tilde{J}_l, \quad (25)$$

$$\tilde{b}_p = \tilde{J}_p^\top r^0, \quad \tilde{b}_l = \tilde{J}_l^\top r^0, \quad (26)$$

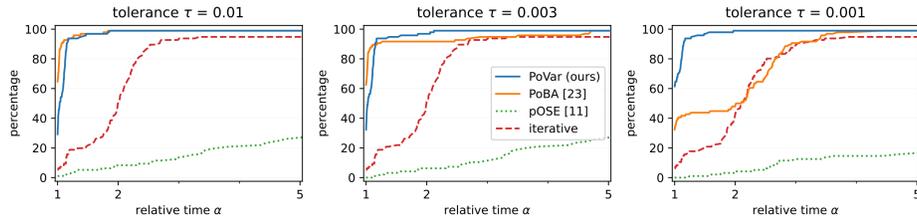
We have unified the notations of bundle adjustment with Riemannian manifold optimization. As the projection  $x^\perp$  is full-rank, it follows that  $\tilde{U}_{\tilde{\lambda}}$  and  $\tilde{V}_{\tilde{\lambda}}$  are symmetric positive-definite (see Supplemental), and then the associated Riemannian Schur complement:

$$\tilde{S} = \tilde{U}_{\tilde{\lambda}} - \tilde{W} \tilde{V}_{\tilde{\lambda}}^{-1} \tilde{W}^\top \quad (27)$$

satisfies the assumption of Proposition 1:

**Lemma 2.** *Let  $\tilde{\mu}$  be an eigenvalue of  $\tilde{U}_{\tilde{\lambda}}^{-1} \tilde{W} \tilde{V}_{\tilde{\lambda}}^{-1} \tilde{W}^\top$ . Then*

$$0 \leq \tilde{\mu} < 1. \quad (28)$$



**Fig. 2:** Average performance profiles across all BAL problems for solving the first stage Eq. 5. Given a tolerance  $\tau \in \{0.01, 0.003, 0.001\}$ , it represents the percentage of solved problems ( $y$ -axis) with relative runtime  $\alpha$  ( $x$ -axis). Expansion methods *PoVar* and *PoBA* show outstanding speed-accuracy results. Our solver *PoVar* is competitive, and most notably for the highest accuracy  $\tau = 0.001$ .

The power series expansion can be applied to the inverse Riemannian Schur complement:

$$\tilde{S}^{-1} \approx \sum_{i=0}^m (\tilde{U}_{\tilde{\lambda}}^{-1} \tilde{W} \tilde{V}_{\tilde{\lambda}}^{-1} \tilde{W}^{\top})^i \tilde{U}_{\tilde{\lambda}}^{-1}, \quad (29)$$

to get pose updates, and then landmark updates by back-substitution.

Finally, the homogeneous pose and landmark updates  $\Delta \tilde{x}_p$  and  $\Delta \tilde{x}_l$  are retrieved by back-projecting pose and landmark updates in the tangent space to the original vector space dimension as follows:

$$\Delta \tilde{x}_p = \tilde{x}_p^{\perp} \Delta x_p, \quad \Delta \tilde{x}_l = \tilde{x}_l^{\perp} \Delta x_l.$$

After above updates are added to the pose and landmark parameters  $x$ , we carry out manifold retraction by normalizing individual vectors of camera parameters and 3D landmarks to maintain their normalized homogeneous forms.

We call this projective framework *Riemannian PoBA* (RiPoBA), that extends PoBA [23] to the Riemannian manifold optimization framework.

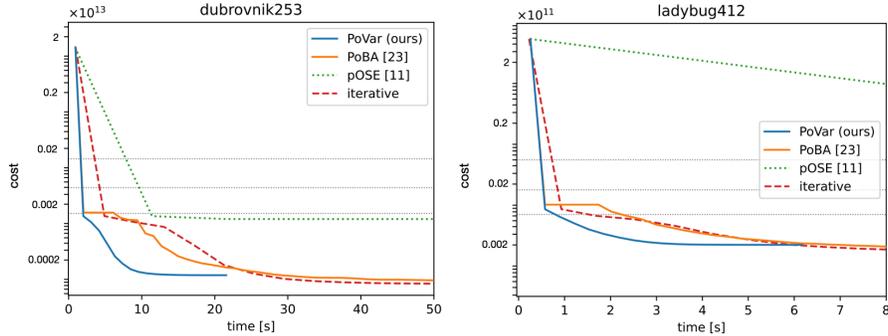
## 5 Experiments

### 5.1 Implementation

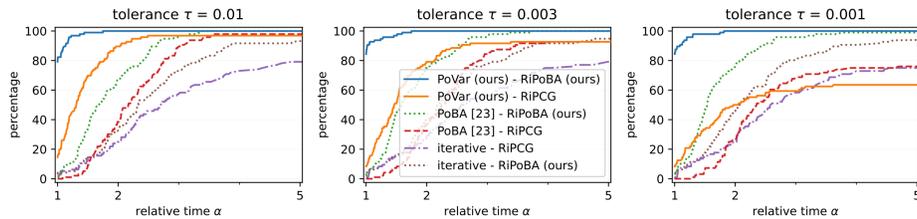
We implement *pOSE*<sup>5</sup> [11], *PoVar* and *RiPoBA* framework in C++, directly on the publicly available implementation of PoBA<sup>6</sup> [23]. That leads to fair comparisons with this recent and challenging solver. *pOSE* differs from *PoVar* by the use of a direct sparse Cholesky factorization. We also compare to VarPro with the conjugate gradients algorithm, preconditioned with Schur-Jacobi preconditioner, called *iterative* in our experiments. For the second stage, we compare *RiPoBA* to

<sup>5</sup> We use our custom implementation as the official code of [11] is not publicly available.

<sup>6</sup> <https://github.com/simonwebertum/poba>



**Fig. 3:** Convergence plots of *Dubrovnik-253* (left) from BAL datasets with 253 poses and *Ladybug-412* with 412 poses, for solving the first stage (Eq. 5). The dotted lines correspond to cost thresholds for tolerance  $\tau \in \{0.01, 0.003, 0.001\}$ .

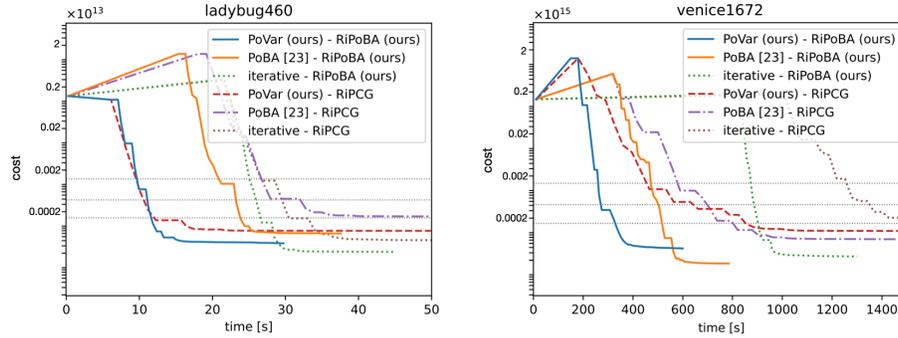


**Fig. 4:** Performance profiles for all BAL problems for solving the first two stages (Eq. 5 and Eq. 7). In each combination, the first solver is used to solve the first stage, and the second solver is used for the second stage. Our proposed combination *PoVar* followed by a Riemannian expansion method outperforms by a large margin compared to other competitors. Also, our proposed Riemannian-expansion method *RiPoBA* outperforms the iterative baseline in all cases, given a same solver for the first stage.

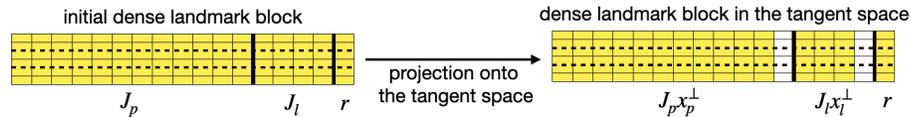
the conjugate gradients algorithm preconditioned by Schur-Jacobi preconditioners with Riemannian manifold optimization framework, called *RiPCG*. Except the solver itself, all implementations share much of the code with [23]. We run experiments on MacOS 14.2.1 with an Intel Core i5 (4 cores at 2GHz).

## 5.2 Experimental settings

**Setup.** For each stage, we set the maximum number of iterations to 50, stopping earlier if a relative function tolerance of  $10^{-6}$  is achieved. The damping factor  $\lambda$  starts for each stage at  $10^{-4}$  and is updated accordingly to the success or failure of the iteration. For expansion methods, we set the maximal order of power series to 20 and a threshold to 0.01. For iterative methods, we set the maximum number of inner iterations to 500. We set  $\eta$  for pOSE (Eq. 6) to 0.1.



**Fig. 5:** Convergence plots of *Ladybug-460* (left) from BAL datasets with 460 poses and *Venice-1672* with 1672 poses, for solving the second stage (Eq. 7). The dotted lines correspond to cost thresholds for tolerance  $\tau \in \{0.01, 0.003, 0.001\}$ . Note that for fair comparison, the initial cost is derived before the first stage. The second cost – that may be higher than the initial one, is the initial cost of the second stage, after the first stage has been run. The runtime includes the time spent to solve the first stage.



**Fig. 6:** Example of dense landmark block in the tangent space with 3 observations from a single landmark. We project the initial dense landmark block (*left*) onto the tangent space by applying to each  $i$ -th pose Jacobian  $J_{p_i} \in \mathbb{R}^{2 \times 12}$  the projection  $\tilde{x}_{p_i}^\perp$ , and to each  $j$ -th landmark Jacobian  $J_{l_j} \in \mathbb{R}^{2 \times 4}$  the projection  $\tilde{x}_{l_j}^\perp$ . The resulting pose and landmark Jacobians in the tangent space (*right*) belong to  $\mathbb{R}^{2 \times 11}$  and  $\mathbb{R}^{2 \times 3}$  respectively.

**Efficient storage for Riemannian manifold optimization framework.** As in [23], we leverage the special structure of BA problem and propose a memory-efficient storage. We organize the landmarks into dense blocks. In particular, we apply on each row associated to a landmark the block-matrices of the projection  $\tilde{x}_{l_j}^\perp, \tilde{x}_{p_i}^\perp$  corresponding to the landmark and to the cameras stored in the considered dense landmark block (see Fig. 6).

**Dataset.** We extensively evaluate our solver and the baselines on the 97 real-world bundle adjustment problems from the BAL project page [3]. The number of poses goes from 16 to 13682. We refer the reader to Supplemental for further details about these problems. For each problem, we only keep the observation measurements. Pose parameters are randomly drawn from an isotropic Gaussian distribution with mean 0 and variance 1, and landmark parameters are deduced

with Eq. 8. Notably and contrary to previous works on initialization-free BA, each solver is ran on the same randomized problem, for fair comparisons.

### 5.3 Performance profile

We jointly evaluate both runtime and accuracy with performance profiles [8]. Given a solver, the performance profile maps the relative runtime  $\alpha$  to the percentage of problems solved with accuracy  $\tau$ . Graphically, the performance profile of a given solver is the percentage of problems solved faster than the relative runtime  $\alpha$  on the  $x$ -axis. Let be  $S$  and  $P$  the sets of solvers and problems, respectively. In practice, we can define the objective threshold for a problem  $p$  by:

$$f_\tau(p) = f^*(p) + \tau(f^0(p) - f^*(p)), \quad (30)$$

with  $f^0(p)$  the initial objective and  $f^*(p)$  the smallest error reached by the family of solvers. The runtime a solver  $s$  needs to reach this threshold is noted  $T_\tau(p, s)$ . The performance profile of a solver for a relative runtime  $\alpha$  is defined as:

$$\rho(s, \alpha) = \frac{100}{|P|} |\{p \in P | T_\tau(p, s) \leq \alpha \min_{s \in S} T_\tau(p, s)\}|. \quad (31)$$

Graphically, a curve on the left of the performance profile is linked to better runtime, whereas a curve on the right is related to better accuracy. Note that for meaningful comparison, all solvers should have the same initial objective.

### 5.4 Analysis

**First stage.** Fig. 2 shows the performance profiles for all BAL datasets with tolerances  $\tau \in \{0.01, 0.003, 0.001\}$  to solve Eq. 5. As expected, the direct factorization solver (dashed green) used in Hong et al. [11] shows poor performance. Our solver *PoVar* (blue) challenges *PoBA* for the largest tolerance 0.01, and is by far the most competitive solver for the smallest tolerance  $\tau = 0.001$ , that is for the highest accuracy. For  $\tau = 0.003$ , *PoVar* slightly outperforms *PoBA*. We also highlight that our solver outperforms the two competitors associated to the VarPro algorithm, iterative (dashed red) and direct factorization.

Fig. 3 shows two examples of the cost decrease during the first stage. Notably, *PoVar* (blue) converges much more smoothly than its main challenger *PoBA* which gets stuck in early iterations. By considering the intersections of the solvers with the cost thresholds (dashed grey lines), it also demonstrates the slowness of the iterative method compared to expansion methods in terms of runtime.

**First and second stage.** Fig. 4 shows the performance profiles for all BAL datasets with tolerances  $\tau \in \{0.01, 0.003, 0.001\}$  to solve the first two stages, that are Eq. 5 followed by Eq. 7. Note that we compare in this experiment the cost of the second stage only, as the first stage is only used to get an approximated initialization for the projective formulation. As direct factorization shows poor

performance during the first stage, we only take into account the most promising combinations of solvers. *PoVar* followed by *RiPoBA* (blue) outperforms all the competitors both in terms of runtime and accuracy. The combinations with *RiPoBA* outperform *RiPCG* for the highest accuracy  $\tau = 0.001$  for all relative time greater than  $\alpha = 2$ , that reflects the better convergence of Riemannian expansion method compared to *RiPCG*. We also note, given a same solver for first stage, *RiPoBA* outperforms *RiPCG* during the second stage across all tolerances.

Fig. 5 illustrates on two examples the cost decrease during the second stage. On the left figure (*Ladybug-460*), the best two solvers in terms of final convergence are built on our framework *RiPoBA*. On the right figure (*Venice-1672*), all the solvers using *RiPoBA* converge to a smaller error than their iterative competitor *RiPCG*. By considering the intersection between the solvers and the cost thresholds (dashed grey lines), the combination *PoVar-RiPoBA* outperforms all other combinations, and most notably for the largest problem with 1672 poses.

**Conclusive remark.** The experiments emphasize the high efficacy of our solvers *PoVar* and *RiPoBA*, during both first and second stages of the stratified BA problem. Concerning the first stage, the convergence of *PoVar* is much smoother than its competitors, that explains its larger speed with respect to *PoBA* to reach the cost thresholds, albeit both are built with power series. Regarding the second stage, for a same given solver in the first stage, our *RiPoBA* outperforms the preconditioned conjugate gradients with Riemannian manifold optimization framework in terms of speed and accuracy, especially when coupled with *PoVar*.

## 6 Conclusion

We have introduced a novel approach to address the scalability challenge for initialization-free bundle adjustment. Our proposed Power Variable Projection (PoVar) algorithm, theoretically justified, offers new insights to this uncharted problem. By extending recent inverse expansion techniques to the VarPro algorithm on one hand, and to Riemannian manifold optimization on the other hand, we have demonstrated the capability to efficiently solve large-scale stratified BA problem with thousands of cameras. Notably, we achieve state-of-the-art results in terms of speed and accuracy on the real-world BAL dataset. While initialization-free BA is still in its nascent stage, we hope that our method will pave the way for further exploration of this difficult optimization problem, and will generate further steps towards initialization-free structure-from-motion.

**Limitations and future work.** First, our analysis applies to the BA problem in which excess outlier point tracks are assumed to have been filtered out as in previous formulations [11–14]. Second, the 3D reconstruction in Fig. 1 assumes that the intrinsics are known during the metric upgrade stage. In practice, we use the approximated focal lengths given in the BAL dataset, that results in imperfectly accurate illustration. However, some formulations based on pOSE handle unknown intrinsics and could be easily adapted to *PoVar*.

## Acknowledgements

This work was supported in part by the NRF grants funded by the Korea government (MSIT) (No. 2022R1C1C1004907) and in part by the Institute of Information & communications Technology Planning & Evaluation (IITP) under the artificial intelligence semiconductor support program to nurture the best talents (IITP-(2024)-RS-2023-00253914) grant funded by the Korea government (MSIT).

## References

1. Absil, P.A., Mahony, R., Sepulchre, R.: Optimization algorithms on matrix manifolds. Princeton University Press (2008)
2. Agarwal, S., Mierle, K., Others: Ceres solver. <http://ceres-solver.org> (2024)
3. Agarwal, S., Snavely, N., Seitz, S.M., Szeliski, R.: Bundle adjustment in the large. In: Computer Vision—ECCV 2010: 11th European Conference on Computer Vision, Heraklion, Crete, Greece, September 5–11, 2010, Proceedings, Part II 11. pp. 29–42. Springer (2010)
4. Belder, A., Vivanti, R., Tal, A.: A game of bundle adjustment—learning efficient convergence. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 8428–8437 (2023)
5. Buchanan, A.M., Fitzgibbon, A.W.: Damped Newton algorithms for matrix factorization with missing data. In: 2005 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). vol. 2, pp. 316–322 (2005). <https://doi.org/10.1109/CVPR.2005.118>
6. Demmel, N., Gao, M., Laude, E., Wu, T., Cremers, D.: Distributed photometric bundle adjustment. In: 2020 International Conference on 3D Vision (3DV). pp. 140–149. IEEE (2020)
7. Demmel, N., Sommer, C., Cremers, D., Usenko, V.: Square root bundle adjustment for large-scale reconstruction. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 11723–11732 (2021)
8. Dolan, E.D., Moré, J.J.: Benchmarking optimization software with performance profiles. *Mathematical programming* **91**, 201–213 (2002)
9. Golub, G.H., Pereyra, V.: The differentiation of pseudo-inverses and nonlinear least squares problems whose variables separate. *SIAM Journal on numerical analysis* **10**(2), 413–432 (1973)
10. Hong, J.H., Fitzgibbon, A.: Secrets of matrix factorization: Approximations, numerics, manifold optimization and random restarts. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 4130–4138 (2015)
11. Hong, J.H., Zach, C.: pose: Pseudo object space error for initialization-free bundle adjustment. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 1876–1885 (2018)
12. Hong, J.H., Zach, C., Fitzgibbon, A.: Revisiting the variable projection method for separable nonlinear least squares problems. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 5939–5947. IEEE (2017)
13. Hong, J.H., Zach, C., Fitzgibbon, A., Cipolla, R.: Projective bundle adjustment from arbitrary initialization using the variable projection method. In: Computer Vision—ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I 14. pp. 477–493. Springer (2016)

14. Iglesias, J.P., Nilsson, A., Olsson, C.: expose: Accurate initialization-free projective factorization using exponential regularization. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 8959–8968 (2023)
15. Jeong, Y., Nister, D., Steedly, D., Szeliski, R., Kweon, I.S.: Pushing the envelope of modern methods for bundle adjustment. In: 2010 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 1474–1481 (2010). <https://doi.org/10.1109/CVPR.2010.5539795>
16. Kaufman, L.: A variable projection method for solving separable nonlinear least squares problems. *BIT Numerical Mathematics* **15**, 49–57 (1975)
17. Okatani, T., Yoshida, T., Deguchi, K.: Efficient algorithm for low-rank matrix factorization with missing components and performance comparison of latest algorithms. In: 2011 IEEE International Conference on Computer Vision (ICCV). pp. 842–849 (2011). <https://doi.org/10.1109/ICCV.2011.6126324>
18. Ren, J., Liang, W., Yan, R., Mai, L., Liu, S., Liu, X.: Megba: A gpu-based distributed library for large-scale bundle adjustment. In: European Conference on Computer Vision. pp. 715–731. Springer (2022)
19. Ruhe, A., Wedin, P.Å.: Algorithms for separable nonlinear least squares problems. *SIAM Review (SIREV)* **22**(3), 318–337 (1980). <https://doi.org/10.1137/1022057>
20. Strelow, D.: General and nested Wiberg minimization. In: 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 1584–1591 (2012). <https://doi.org/10.1109/CVPR.2012.6247850>
21. Strelow, D.: General and nested Wiberg minimization: L2 and maximum likelihood. In: 12th European Conference on Computer Vision (ECCV). pp. 195–207 (2012). [https://doi.org/10.1007/978-3-642-33786-4\\_15](https://doi.org/10.1007/978-3-642-33786-4_15)
22. Triggs, B., McLauchlan, P.F., Hartley, R.I., Fitzgibbon, A.W.: Bundle adjustment—a modern synthesis. In: Vision Algorithms: Theory and Practice: International Workshop on Vision Algorithms Corfu, Greece, September 21–22, 1999 Proceedings. pp. 298–372. Springer (2000)
23. Weber, S., Demmel, N., Chan, T.C., Cremers, D.: Power bundle adjustment for large-scale 3d reconstruction. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 281–289 (2023)
24. Weber, S., Demmel, N., Cremers, D.: Multidirectional conjugate gradients for scalable bundle adjustment. In: DAGM German Conference on Pattern Recognition. pp. 712–724. Springer (2021)
25. Wright, S.J.: Numerical optimization. Springer (2006)
26. Zhang, F.: The Schur complement and its applications, vol. 4. Springer Science & Business Media (2006)
27. Zheng, Q., Xi, Y., Saad, Y.: A power schur complement low-rank correction preconditioner for general sparse linear systems. *SIAM Journal on Matrix Analysis and Applications* **42**(2), 659–682 (2021)
28. Zhou, L., Luo, Z., Zhen, M., Shen, T., Li, S., Huang, Z., Fang, T., Quan, L.: Stochastic bundle adjustment for efficient and scalable 3d reconstruction. In: European Conference on Computer Vision. pp. 364–379. Springer (2020)