ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume IV-2/W3, 2017
International Conference on Unmanned Aerial Vehicles in Geomatics, 4–7 September 2017, Bonn, Germany

ON THE QUALITY AND EFFICIENCY OF APPROXIMATE SOLUTIONS TO BUNDLE ADJUSTMENT WITH EPIPOLAR AND TRIFOCAL CONSTRAINTS

Johannes Schneider, Cyrill Stachniss and Wolfgang Förstner

Institute of Geodesy and Geoinformation, University of Bonn
(johannes.schneider, cyrill.stachniss, wolfgang.foerstner)@igg.uni-bonn.de

KEY WORDS: Bundle Adjustment, Trifocal Constraint, Approximate Solution, Quality Evaluation

ABSTRACT:

Bundle adjustment is a central part of most visual SLAM and Structure from Motion systems and thus a relevant component of UAVs equipped with cameras. This paper makes two contributions to bundle adjustment. First, we present a novel approach which exploits trifocal constraints, i.e., constraints resulting from corresponding points observed in three camera images, which allows to estimate the camera pose parameters without 3D point estimation. Second, we analyze the quality loss compared to the optimal bundle adjustment solution when applying different types of approximations to the constrained optimization problem to increase efficiency. We implemented and thoroughly evaluated our approach using a UAV performing mapping tasks in outdoor environments. Our results indicate that the complexity of the constraint bundle adjustment can be decreased without loosing too much accuracy.

1. INTRODUCTION

Precise models of the environment are needed for several robotic applications and are central to several UAV-based services. Most SLAM and visual mapping systems use a form of bundle adjustment (BA) for simultaneously refining camera pose parameters and 3D point coordinates. Thus, effectively solving the BA or the underlying error minimization problem is essential for many approaches such as structure from motion (Agarwal et al., 2011) and online SLAM or visual odometry. BA has favorable properties: it is statistically optimal in case all statistical properties are modeled and considered correctly, it is efficient in case sparse matrix operations are used, and can be parallelized. A broad review is given by Triggs et al. (2000).

Steffen et al. (2010) have shown that rigorous BA can be formulated based only on epipolar and trifocal constraints. The epipolar constraint is a relation between two camera views, that enforces an image point to be on the epipolar line described by an corresponding image point in another image and the essential or fundamental matrix between the views (Hartley and Zisserman, 2004). A trifocal constraint between image points is necessary if the corresponding scene point lies on the trifocal plane, which practically always is true for neighbored images in an image sequence, where projection centers are collinear or nearly collinear. Epipolar and trifocal constraints lead to implicit functions that enforce the intersection of bundle of rays in 3D space without explicitly representing 3D point coordinates. This reduces the number of unknown parameters of the underlying optimization problem to the camera pose parameters. The obtained normal equations are equivalent to the normal equation system of classical BA when applying the Schur Complement to eliminate the unknown 3D point coordinates. Its solution is therefore in statistical terms as optimal as classical BA.

BA based only on epipolar and trifocal constraints has several advantages over classical BA:

- it allows to integrate image points, whose projections rays have small parallactic angles, which in classical bundle adjustments would lead to 3D points lying numerically at infinity. Including such observations increases the generality of BA and improves the estimated rotations of the camera pose parameters (Schneider et al., 2012);
- it leads directly to the normal equation system reduced to the camera pose parameters. Classical BA needs to apply the Schur Complement to eliminate the 3D points;
- it does not require an initial guess for the locations of the 3D points, which is required for classical BA;
- it allows to arrive at approximate solutions, e.g., by neglect correlations between multi-view constraints and the relinearization of observations, which significantly increases the efficiency without a substantial loss in accuracy.

Nevertheless, the trifocal BA formulation without simplifying approximations leads to higher computational complexity than classical BA. Because of the implicit epipolar and trifocal constraints one needs to employ the Gauss–Helmert Model for optimization, which requires the costly determination of corrections for all observations, which is not needed in the Gauss–Markov Model employed for classical BA. To substantially reduce the computational complexity Indelman et al. (2012) propose to neglect (1) correlations between the constraints and (2) corrections to the observations during optimization and (3) fix the weights for the individual constraints after the first iteration.

In this paper we investigate (1) the gain in efficiency and (2) the loose of quality of several assumptions, which lead to an approximate solution of BA which substantially reduces computational complexity. Additionally we propose a new formulation for trifocal constraints which can be employed in BA without structure estimation. Contrary to formulations of trifocal constraints in previous work it does not degenerate in specific situations.

2. RELATED WORK

Sparse BA is most efficient in case a sparse representation is used (Hartley and Zisserman, 2004). The publicly available software package SBA for generic sparse bundle adjustment by Lourakis and Argyros (2009) is used for example in a modified version in Bundler (Agarwal et al., 2011) to solve large-scale structure from motion. Konolige (2010) introduced Sparse SBA (sSBA)

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
doi:10.5194/isprs-annals-IV-2-W3-81-2017

which exploits the sparse secondary structure of sparse camera to camera relations, which increases computational and memory efficiency. More recently, the popular software package g2o (Kümmerle et al., 2011) shows a comparable efficiency as sSBA but uses a more generic formulation of the optimization problem using factor graphs.

BA without structure estimation has been proposed by Rodríguez et al. (2011), but their approach relies only on epipolar constraints, which are not able to transfer a consistent scale between cameras having parallel epipolar planes which occur on straight camera trajectories. Thus Steffen et al. (2010) propose to use epipolar and trifocal constraints in BA without structure estimation. But their trifocal constraints can not be computed in a closed form expression which is why a stable condition needs to be sampled, where the number of samples is not fixed. Indelman et al. (2012) propose simplifying approximations to the optimization problem by rewriting the implicit trifocal and epipolar constraints into explicit expressions. This way the authors obtained a pose graph formulation, which can be optimized with the computational efficient incremental smoothing and mapping (iSAM) algorithm by Kaess et al. (2012). But their approach can not handle all possible camera configurations.

3. CLASSICAL BUNDLE ADJUSTMENT

The general objective of BA is to optimally estimate camera rotations $\widehat{R}_t$, camera positions $\widehat{\boldsymbol{Z}}_t$ and 3D point coordinates $\widehat{\mathbf{X}}_i$ simultaneously. In the following, we assume that each observed 2D image point $\boldsymbol{x}_{it}$ in view t is associated to a certain 3D point i and that the intrinsic camera calibration is given by calibration matrix K_t. Given an initial guess, i.e., knowing approximate quantities $\widehat{R}_t^a$, $\widehat{\boldsymbol{Z}}_t^a$ and $\widehat{\mathbf{X}}_i^a$, the reprojection with projection matrix $\mathsf{P}_t = \mathsf{K}_t \widehat{R}_t^{a\mathsf{T}} \left[I_3 \,|\, -\widehat{\boldsymbol{Z}}_t^a \right]$ yields the homogeneous image point

$$\mathbf{x}_{it}^a = \mathsf{P}_t^a \mathbf{X}_i^a. \tag{1}$$

With the three rows $\mathsf{P}_{1,t}$, $\mathsf{P}_{2,t}$ and $\mathsf{P}_{3,t}$ of P_t we obtain the reprojected image point in Euclidean coordinates $\boldsymbol{x}_{it}^a = \left[\mathsf{P}_{1,t}\widehat{\mathbf{X}}_i^a / \mathsf{P}_{3,t}\widehat{\mathbf{X}}_i^a \,,\, \mathsf{P}_{2,t}\widehat{\mathbf{X}}_i^a / \mathsf{P}_{3,t}\widehat{\mathbf{X}}_i^a \right]^\mathsf{T}$ and the reprojection error $\boldsymbol{v}_{it} = \boldsymbol{x}_{it}^a - \boldsymbol{x}_{it}$ in the image plane. Assuming the image points to be corrupted with mutually uncorrelated Gaussian noise $\Sigma_{\boldsymbol{x}_{it}\boldsymbol{x}_{it}}$, maximum likelihood estimates are obtained by iteratively improving the unknown parameters by minimizing the squared Mahalanobis distance $\sum_{it} \boldsymbol{v}_{it}^\mathsf{T} \Sigma_{\boldsymbol{x}_{it}\boldsymbol{x}_{it}}^{-1} \boldsymbol{v}_{it}$ using the Gauss–Markov model, see (Förstner and Wrobel, 2016, Sect. 4.4).

With T camera poses and I observed 3D points, the total number of unknown parameters to be optimized counts $6T + 3I$. If one is only interested in estimating the camera poses, the normal equation system can be reduced to the $6T$ pose parameters by applying the Schur complement. However, as BA needs to be solved iteratively due to its non-linearity one is forced to compute the 3D point in each iteration, even if they are not of interest. In contrast to that, we can directly obtain the reduced normal equation system without applying the Schur complement or determining 3D points by employing epipolar and trifocal constraints, which are introduced in the next section.

4. EPIPOLAR AND TRIFOCAL CONSTRAINTS

The classical solution to BA seeks to minimize the reprojection error of corresponding points. Alternatively, we can formulate an error minimization problem that exploits constraints from the epipolar geometry of image pairs as well as constraints that result from observing the same point from three different camera images, so-called trifocal constraints. This formulation has the advantage, that only the camera extrinsics are unknown parameters and we do not need to estimate the 3D point parameters in the optimization. After describing how to formulate such constraints in this section, we explain in Sec. 5 how to consider them in BA.

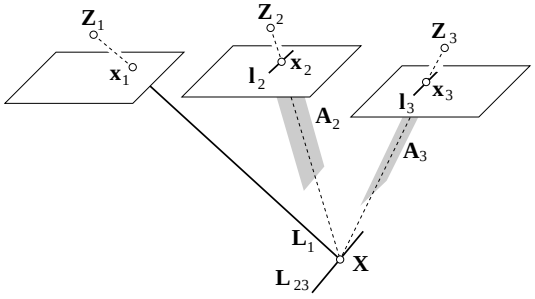

Figure 1. Trifocal constraint. We choose two lines $\mathbf{l}_2$ and $\mathbf{l}_3$ through the image points $\mathbf{x}_2$ and $\mathbf{x}_3$ of the second and third camera. The corresponding projection planes $\mathbf{A}_2$ and $\mathbf{A}_3$ raise the intersection line $\mathbf{L}_{23}$. The constraint requires that projection line $\mathbf{L}_1$ and intersection line $\mathbf{L}_{23}$ intersect in one single point.

Given three corresponding image points ($\boldsymbol{x}_1$,$\boldsymbol{x}_2$,$\boldsymbol{x}_3$) in three views $t = 1, 2, 3$, we need to formulate three independent constraints (g_1, g_2, g_3) for each correspondence related to a 3D point. Two-view epipolar constraints do not allow to transfer a consistent scale given straight trajectories with collinear projection centers (Rodríguez et al., 2011), which usually appear in image sequences. We always use one trifocal and two epipolar constraints, which are simpler and one trifocal constraint is sufficient.

The *first two constraints* are epipolar constraints and enforce the camera rays to be on their epipolar lines w.r.t. the first camera

$$g_1 = \mathbf{x}_1^\mathsf{T} R_1^\mathsf{T} S(\boldsymbol{Z}_2 - \boldsymbol{Z}_1) R_2 \mathbf{x}_2 \tag{2}$$

$$g_2 = \mathbf{x}_1^\mathsf{T} R_1^\mathsf{T} S(\boldsymbol{Z}_3 - \boldsymbol{Z}_1) R_3 \mathbf{x}_3 \,, \tag{3}$$

where $S(\cdot)$ is the skew symmetric matrix of the input vector. Note that we assume here, that the the camera calibration is given, such that we can convert an image point $\boldsymbol{x}$ into a ray direction $\mathbf{x}$, e.g.in case of a pinhole camera with calibration matrix K by $\mathbf{x} = \mathsf{K}^{-1} \left[\boldsymbol{x}^\mathsf{T} \,,\, 1 \right]^\mathsf{T}$.

The *third constraint* enforces the intersection of all ray directions in a single point, which we formulate in the following way. Consider two planes $\mathbf{A}_2$ and $\mathbf{A}_3$ that go along the ray direction $\mathbf{x}_2$ and $\mathbf{x}_3$ and are projected as 2D lines $\mathbf{l}_2$ and $\mathbf{l}_3$ such that we have $\mathbf{A}_2 = \mathsf{P}_2^\mathsf{T} \mathbf{l}_2$ and $\mathbf{A}_3 = \mathsf{P}_3^\mathsf{T} \mathbf{l}_3$, see Figure 1. The intersection of both planes in 3D yields the 3D line

$$\mathbf{L}_{23} = \overline{\mathsf{T}}(\mathbf{A}_2)\mathbf{A}_3 \quad \text{with} \quad \overline{\mathsf{T}}(\mathbf{A}) = \begin{bmatrix} S(\boldsymbol{A}_h) & \mathbf{0}_3 \\ A_0 I_3 & -\boldsymbol{A}_h \end{bmatrix} \tag{4}$$

and $\mathbf{A} = \left[\boldsymbol{A}_h^\mathsf{T} \,,\, A_0 \right]^\mathsf{T}$. The final constraint is that the 3D line $\mathbf{L}_1$ along the ray direction $\mathbf{x}_1$

$$\mathbf{L}_1 = \begin{bmatrix} R_1 \\ S(\boldsymbol{Z}_1) R_1 \end{bmatrix} \mathbf{x}_1 \tag{5}$$

has to intersect $\mathbf{L}_{23}$ in a single point. Using homogeneous coor-

dinates this leads to the constraint

$$g_3 = \mathbf{L}_1^\mathsf{T} D\, \mathbf{L}_{23} \qquad \text{with} \qquad D = \begin{bmatrix} 0 & I_3 \\ I_3 & 0 \end{bmatrix}. \tag{6}$$

where D is the dualizing matrix.

We need to guarantee that $\mathbf{L}_1$ intersects $\mathbf{L}_{23}$ in one single point. In order to achieve a numerically stable constraint we choose the two lines $\mathbf{l}_2$ and $\mathbf{l}_3$ and specify their direction $\mathbf{v}_2$ and $\mathbf{v}_3$ to be perpendicular to the epipolar lines in the second and third image, such that we have

$$\mathbf{l}_2 = S(\mathbf{v}_2)\mathbf{x}_2 \qquad \text{with} \qquad \mathbf{v}_2 = S(R_2^\mathsf{T}(\boldsymbol{Z}_2 - \boldsymbol{Z}_1))\mathbf{x}_2, \tag{7}$$

$$\mathbf{l}_3 = S(\mathbf{v}_3)\mathbf{x}_3 \qquad \text{with} \qquad \mathbf{v}_3 = S(R_3^\mathsf{T}(\boldsymbol{Z}_3 - \boldsymbol{Z}_1))\mathbf{x}_3. \tag{8}$$

When using the constraint g_3 in an estimation procedure, the vectors $\mathbf{v}_2$ and $\mathbf{v}_3$ can be treated as fixed entries.

These constraints work for all points if they are not close to an epipole, which would also not work in a classical BA. Then at least two projection planes are nearly parallel and the intersecting line is numerically unstable or, in case of observational noise, inaccurate. This especially holds for forward motion, for which image points close to the focus of expansion, i.e. the epipole, cannot be handled.

If none of the image points are close to the epipoles, then, following Figure 1, the two projection planes $\mathbf{A}_2$ and $\mathbf{A}_3$ of the second image intersect the projection line $\mathbf{L}_1$ of the first image in well defined points and are not parallel, thus have a well defined intersection line $\mathbf{L}_{23}$, which therefore needs to pass the projection line $\mathbf{L}_1$. Hence, the triplet constraint never has a singularity and fixes the image points $\mathbf{x}_2$ and $\mathbf{x}_3$ perpendicular to the epipolar lines w.r.t. the first image used in Eq. (2) and (3).

Indelman (2012) uses the trifocal constraint

$$\begin{aligned} z_3 = &(S(R_2\mathbf{x}_2)R_1\mathbf{x}_1)^\mathsf{T}\, S(R_3\mathbf{x}_3)(\boldsymbol{Z}_3 - \boldsymbol{Z}_2) \\ &- (S(R_1\mathbf{x}_2)(\boldsymbol{Z}_2 - \boldsymbol{Z}_1))^\mathsf{T}\, S(R_3\mathbf{x}_3)R_2\mathbf{x}_2\,, \end{aligned} \tag{9}$$

but this formulation degenerates in case the epipolar plane normals $\mathbf{n}_{12}$ and $\mathbf{n}_{23}$ of first and second camera and second and third camera are perpendicular. The constraint projects normal direction $\mathbf{n}_{12}$ given with different lengths in Eq. (9) by $S(R_2\mathbf{x}_2)R_1\mathbf{x}_1$ and $S(R_1\mathbf{x}_2)(\boldsymbol{Z}_2 - \boldsymbol{Z}_1)$ on normal direction $\mathbf{n}_{23}$ given with different lengths by $S(R_3\mathbf{x}_3)(\boldsymbol{Z}_3 - \boldsymbol{Z}_2)$ and $S(R_3\mathbf{x}_3)R_2\mathbf{x}_2$. In case of perpendicular normal directions, the constraint would be fulfilled under multiple solutions.

So far we have only considered three-view correspondences. In case of correspondences in less than three view we can only apply the epipolar constraint (2). In case of $N_i > 3$ correspondences, we need to avoid to use the same constraints twice, and use only independent constraints between the different views. Each corresponding image observation contributes with two constraints, therefore the total number of constraints between corresponding views counts $(2N_i - 3)$. Each correspondence needs to be involved in at least one epipolar and one trifocal constraint. Indelman et al. incorporate an new image of an image sequence by formulating epipolar constraints between the last two recent images and trifocal constraints between the last three recent images.

As in a classical BA the estimation of the poses of calibrated cameras is only possible up to a similarity transformation. To overcome the 7 DOF ambiguity of the overall translation, rotation and scale, we define either the gauge by imposing seven centroid constraints on the approximate values of the projection centers. This results in a free BA, where the trace of the covariance matrix of estimated camera poses is minimal. Or we estimate the camera poses relative to one camera, which fixes six DOF. To define the overall scale, we constrain two cameras to have a certain distance to each other, see (Förstner and Wrobel, 2016, Chapt. 4.5).

5. TRIFOCAL BUNDLE ADJUSTMENT

We sketch the maximum likelihood estimation with implicit functions, also called estimation with the Gauss–Helmert model, see (Förstner and Wrobel, 2016, Chapt. 4.8) and relate it to the classical regression model, also called Gauss–Markov model. This is the basis of four variants for simplifications. These lead to approximations which are then compared with the statistically optimal ones w.r.t. accuracy and speed of convergence.

5.1 The Estimation Model

5.1.1 Gauss–Helmert Model The Gauss–Helmert model starts from G constraints, $\boldsymbol{g} = [g_g]$, among the N observations $\boldsymbol{l} = [l_n]$, which are assumed to be a sample of a multivariate Gaussian distribution $\mathcal{N}(\mathbb{E}(\underline{\boldsymbol{l}}), \sigma_0^2\Sigma_{ll}^a)$, and U unknown parameters $\boldsymbol{x} = [x_u]$:

$$\boldsymbol{g}(\mathbb{E}(\underline{\boldsymbol{l}}), \boldsymbol{x}) = \mathbf{0} \qquad \text{and} \qquad \mathbb{D}(\underline{\boldsymbol{l}}) = \sigma_0^2\Sigma_{ll}^a\,. \tag{10}$$

We assume the covariance matrix of the observations is approximately Σ_{ll}^a, hence we assume $\sigma_0 = 1$; we will be able to estimate this factor later. Given observations $\boldsymbol{l} = [l_n]$ there are no parameters $\boldsymbol{x}$ for which $\boldsymbol{g}(\boldsymbol{l}, \boldsymbol{x}) = 0$ holds. Therefore the goal is to find corrections $\widehat{\boldsymbol{v}}$ of the observations and best estimates $\widehat{\boldsymbol{x}}$ such that the constraints

$$\boldsymbol{g}(\widehat{\boldsymbol{l}}, \widehat{\boldsymbol{x}}) = \boldsymbol{g}(\boldsymbol{l} + \widehat{\boldsymbol{v}}, \widehat{\boldsymbol{x}}) = \mathbf{0} \tag{11}$$

between the fitted observations $\widehat{\boldsymbol{l}} = \boldsymbol{l} + \widehat{\boldsymbol{v}}$ and the estimated parameters $\widehat{\boldsymbol{x}}$ hold and the weighted sum of the squared residuals

$$\Omega(\widehat{\boldsymbol{l}}, \widehat{\boldsymbol{x}}) = \widehat{\boldsymbol{v}}^\mathsf{T}\, \Sigma_{ll}^{-1}\, \widehat{\boldsymbol{v}} \tag{12}$$

is minimum.

5.1.2 Solution in the Gauss–Helmert model The solution is iterative. Starting from approximate values $\widehat{\boldsymbol{l}}^a$ and $\widehat{\boldsymbol{x}}^a$ for the fitted observations $\widehat{\boldsymbol{l}}$ and the estimated parameters $\widehat{\boldsymbol{x}}$ we determine corrections $\widehat{\Delta\boldsymbol{l}}$ and $\widehat{\Delta\boldsymbol{x}}$ to iteratively update the fitted observations and the unknown parameters

$$\widehat{\boldsymbol{l}} = \widehat{\boldsymbol{l}}^a + \widehat{\Delta\boldsymbol{l}} = \boldsymbol{l} + \widehat{\boldsymbol{v}}\,, \qquad \widehat{\boldsymbol{x}} = \widehat{\boldsymbol{x}}^a + \widehat{\Delta\boldsymbol{x}}\,. \tag{13}$$

Each iteration solves for the corrections $\widehat{\Delta\boldsymbol{l}}$ and $\widehat{\Delta\boldsymbol{x}}$ with the linearized substitute constraints

$$\boldsymbol{g}(\widehat{\boldsymbol{l}}, \widehat{\boldsymbol{x}}) = \boldsymbol{g}(\boldsymbol{l}, \widehat{\boldsymbol{x}}^a) + A\widehat{\Delta\boldsymbol{x}} + B^\mathsf{T}\widehat{\boldsymbol{v}} = \mathbf{0} \tag{14}$$

Observe, due to

$$\underline{\boldsymbol{g}} := \boldsymbol{g}(\underline{\boldsymbol{l}}, \boldsymbol{x}) \approx \boldsymbol{g}(\mathbb{E}(\underline{\boldsymbol{l}}), \boldsymbol{x}) + B^\mathsf{T}\underline{\boldsymbol{v}} \tag{15}$$

we introduce the covariance matrix of the constraints

$$\Sigma_{gg} = B^\mathsf{T}\Sigma_{ll}B = W_{gg}^{-1} \tag{16}$$

which we assume is regular, thus has the weight matrix of the constraints W_{gg} as its inverse.

We can determine the corrections in two steps. First, the corrections $\widehat{\Delta \boldsymbol{x}}$ are determined from the linear equation system

$$\boldsymbol{A}^{\mathsf{T}} \boldsymbol{W}_{gg} \boldsymbol{A} \widehat{\Delta \boldsymbol{x}} = \boldsymbol{A}^{\mathsf{T}} \boldsymbol{W}_{gg} \boldsymbol{c}_g \quad \text{with} \quad \boldsymbol{c}_g = -\boldsymbol{g}(\boldsymbol{l}, \widehat{\boldsymbol{x}}^a) \,. \tag{17}$$

Second, we determine the corrections $\widehat{\Delta \boldsymbol{l}}$ from

$$\widehat{\Delta \boldsymbol{l}} = \boldsymbol{\Sigma}_{ll} \boldsymbol{B} \boldsymbol{W}_{gg} (\boldsymbol{c}_g - \boldsymbol{A} \widehat{\Delta \boldsymbol{x}}) - (\widehat{\boldsymbol{l}}^a - \boldsymbol{l}) \,, \tag{18}$$

From Eq. (12) we can determine the estimated variance factor

$$\widehat{\sigma}_0^2 = \frac{\Omega(\widehat{\boldsymbol{x}}, \widehat{\boldsymbol{l}})}{R} \qquad \text{with} \qquad R = G + H - U \tag{19}$$

and the weighted sum of the residuals Eq. (12) evaluated at the estimated values

$$\Omega(\widehat{\boldsymbol{x}}, \widehat{\boldsymbol{l}}) = \widehat{\boldsymbol{v}}^{\mathsf{T}} \boldsymbol{W}_{ll} \widehat{\boldsymbol{v}} = \widehat{\boldsymbol{c}}_g^{\mathsf{T}} \boldsymbol{W}_{gg} \widehat{\boldsymbol{c}}_g \,. \tag{20}$$

5.1.3 On the Structure of the Weight Matrix $\boldsymbol{W}_{gg}$ If each constraint g_j depends on one observational group $\boldsymbol{l}_j$, whose observations are not part of another constraint $g_{j'}$, the covariance matrix $\boldsymbol{\Sigma}_{gg}$ is diagonal. The same holds, if the constraints can be partitioned into groups $\boldsymbol{g}_i$ which only depend on one observational group $\boldsymbol{l}_i$; then the covariance matrix $\boldsymbol{\Sigma}_{gg}$ is block diagonal. If the corresponding Jacobians for each group are $\boldsymbol{A}_i^{\mathsf{T}}$ and $\boldsymbol{B}_i^{\mathsf{T}}$, the matrix $\boldsymbol{N}$ of the normal equation system can be expressed as a sum over all constraints:

$$\boldsymbol{N} = \boldsymbol{A}^{\mathsf{T}} \boldsymbol{W}_{gg} \boldsymbol{A} = \sum_i \boldsymbol{A}_i \boldsymbol{\Sigma}_{g_i g_i} \boldsymbol{A}_i^{\mathsf{T}} = \sum_i \boldsymbol{A}_i (\boldsymbol{B}_i^{\mathsf{T}} \boldsymbol{\Sigma}_{l_i l_i} \boldsymbol{B}_i)^{-1} \boldsymbol{A}_i^{\mathsf{T}} \tag{21}$$

and accordingly $\boldsymbol{A}^{\mathsf{T}} \boldsymbol{W}_{gg} \boldsymbol{c}_g = \sum_i \boldsymbol{A}_i (\boldsymbol{B}_i^{\mathsf{T}} \boldsymbol{\Sigma}_{l_i l_i} \boldsymbol{B}_i)^{-1} \boldsymbol{c}_{g_i}$.

If a group of constraints $\boldsymbol{g}_i$ shares observations, as in our case, the covariance matrix $\boldsymbol{\Sigma}_{g_i g_i} = \boldsymbol{B}_i^{\mathsf{T}} \boldsymbol{\Sigma}_{l_i l_i} \boldsymbol{B}_i$ will not be diagonal or block diagonal any more. Then their inverse, i.e. weight matrix, will be full in general.

For example, in BA, all N_i observations referring to the same scene point $\mathcal{X}_i$ will have a sparse but not diagonal covariance matrix $\boldsymbol{\Sigma}_{g_i g_i}$, hence a full weight matrix of size $G_i \times G_i$. Let us consider three epipolar constraints $g = 1, 3, 5$ and two trifocal constraints $g = 2, 4$ between the image points of four consecutive images, $\mathcal{X}_{it}$, $t = 1, 2, 3, 4$. Then the structure of $\boldsymbol{B}_i^{\mathsf{T}}$ for this group of cameras will be as follows

$$\boldsymbol{B}_i^{\mathsf{T}} = \begin{bmatrix} \boldsymbol{B}_{11} & \boldsymbol{B}_{12} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{B}_{21} & \boldsymbol{B}_{22} & \boldsymbol{B}_{23} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{B}_{32} & \boldsymbol{B}_{33} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{B}_{42} & \boldsymbol{B}_{43} & \boldsymbol{B}_{44} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{B}_{53} & \boldsymbol{B}_{54} \end{bmatrix}_i \tag{22}$$

and the covariance matrix will be

$$\boldsymbol{\Sigma}_{g_i g_i} = \begin{bmatrix} \boldsymbol{\Sigma}_{g_1 g_1} & \boldsymbol{\Sigma}_{g_1 g_2} & \boldsymbol{\Sigma}_{g_1 g_3} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{\Sigma}_{g_2 g_1} & \boldsymbol{\Sigma}_{g_2 g_2} & \boldsymbol{\Sigma}_{g_2 g_3} & \boldsymbol{\Sigma}_{g_2 g_4} & \boldsymbol{\Sigma}_{g_2 g_5} \\ \boldsymbol{\Sigma}_{g_3 g_1} & \boldsymbol{\Sigma}_{g_3 g_2} & \boldsymbol{\Sigma}_{g_3 g_3} & \boldsymbol{\Sigma}_{g_3 g_4} & \boldsymbol{\Sigma}_{g_3 g_5} \\ \boldsymbol{0} & \boldsymbol{\Sigma}_{g_4 g_2} & \boldsymbol{\Sigma}_{g_4 g_3} & \boldsymbol{\Sigma}_{g_4 g_4} & \boldsymbol{\Sigma}_{g_4 g_5} \\ \boldsymbol{0} & \boldsymbol{\Sigma}_{g_5 g_2} & \boldsymbol{\Sigma}_{g_5 g_3} & \boldsymbol{\Sigma}_{g_5 g_4} & \boldsymbol{\Sigma}_{g_5 g_5} \end{bmatrix} \tag{23}$$

with $\boldsymbol{\Sigma}_{g_j g_k} = \boldsymbol{B}_j^{\mathsf{T}} \boldsymbol{\Sigma}_{ll} \boldsymbol{B}_k$. The covariance matrix has in general a full inverse. Hence, matrix $\boldsymbol{N}$ reads as

$$\begin{aligned} \boldsymbol{N} \;=\; \boldsymbol{A}^{\mathsf{T}} \boldsymbol{W}_{gg} \boldsymbol{A} \;&=\; \sum_j \sum_k \boldsymbol{A}_j \boldsymbol{W}_{g_j g_k} \boldsymbol{A}_k^{\mathsf{T}} \\ &=\; \sum_j \sum_k \boldsymbol{A}_j (\boldsymbol{B}^{\mathsf{T}} \boldsymbol{\Sigma}_{ll} \boldsymbol{B})_{jk}^{-1} \boldsymbol{A}_k^{\mathsf{T}} \,. \end{aligned} \tag{24}$$

The effort of inverting the generally sparse matrix $\boldsymbol{\Sigma}_{g_i g_i}$ can be significantly reduced, if the matrix product $\boldsymbol{F}_i = \boldsymbol{W}_{g_i g_i} \boldsymbol{A}_i$ is determined by solving the (generally sparse) equation system $\boldsymbol{\Sigma}_{g_i g_i} \boldsymbol{F}_i = \boldsymbol{A}_i$ for $\boldsymbol{F}_i$.

5.1.4 Solution in the Gauss–Markov Model The solution for the estimated parameters can also be obtained from a Gauss–Markov model when substituting

$$\boldsymbol{v}_g = -\boldsymbol{B}^{\mathsf{T}} \boldsymbol{v} \tag{25}$$

into Eq. (14). Using Eq. (15) and $\boldsymbol{c}_g$ from Eq. (17) we immediately obtain the linearized Gauss–Markov model

$$\boldsymbol{c}_g + \boldsymbol{v}_g = \boldsymbol{A} \widehat{\Delta \boldsymbol{x}} \qquad \text{and} \qquad \mathbb{D}(\underline{\boldsymbol{v}}_g) = \boldsymbol{\Sigma}_{gg} \,. \tag{26}$$

which leads to the same estimates as in Eq. (17).

The iterative solution of this model, however, has to take the linearization point for the Jacobians $\boldsymbol{A}$ and $\boldsymbol{B}$ into account, which are the fitted observations $\widehat{\boldsymbol{l}}$ and the the estimated parameters $\widehat{\boldsymbol{x}}$. Hence, the result of estimation in the Gauss–Markov model only is the same, if we in each iteration step determine $\widehat{\Delta \boldsymbol{l}}$ via Eq. (18) to obtain the fitted original observations $\widehat{\boldsymbol{l}}$ via Eq. (13). This is possible, but requires access to the Jacobian $\boldsymbol{B}$. Then there is no difference between the Gauss–Markov and the Gauss–Helmert model. In addition, we need the inverse of the covariance matrix $\boldsymbol{\Sigma}_{gg}$, which in general will not be a diagonal block matrix with small blocks referring to groups of two or three constraints.

These are reasons to investigate approximate solutions, which can be expected to be computationally more efficient.

5.2 Approximations of the Optimal Model

We address four cases of simplifications of the original estimation model. All are approximations of the original model and lead to suboptimal results.

CASE A: *Approximated Jacobians*

The Jacobians $\boldsymbol{A}$ and $\boldsymbol{B}$ are approximated, by *linearizing at the original observations* $\boldsymbol{l}$, instead of at the fitted observations $\widehat{\boldsymbol{l}}$. The approximation will increase if the standard deviations of the observations increases, or if there are outliers in the observations. The suboptimality of this approximation has already been discussed in (Stark and Mikhail, 1973).

CASE B: *Approximated Weights of the Constraints*

The matrix $\boldsymbol{W}_{gg}$ is approximated by *neglecting the correlations between the constraints*. Hence we use the inverse of the diagonalized covariance matrix,

$$\boldsymbol{\Sigma}_{gg}^{\text{CASE B}} = \text{Diag}\left(\boldsymbol{b}_g^{\mathsf{T}} \boldsymbol{\Sigma}_{ll} \boldsymbol{b}_g\right) \,, \tag{27}$$

with the rows $\boldsymbol{b}_g$ of $\boldsymbol{B}$. This significantly reduces the effort for determining $\boldsymbol{W}_{gg}$. For CASE B we assume the Jacobians $\boldsymbol{A}$ and $\boldsymbol{B}$ are taken at the estimated parameters and the estimated observations. This can only be realized within the Gauss–Helmert model, since otherwise the estimated observations $\widehat{\boldsymbol{l}}$ are not available.

CASE C: *Approximated Jacobians and Weights for the Constraints*

We approximate both the Jacobians, by *linearizing at the given observations*, and the weight matrix, by *neglecting the correlations between the constraints*. This approximation is useful when applying the Gauss–Markov model.

CASE D: *Approximated Weight Matrix W_{gg} of the First Iteration*

We approximate the weight matrix W_{gg} by that obtained in the first iteration. This reduces the computational burden in the further iterations.

The weight matrix W_{gg} then *depends on the approximate values for the observations and the parameters*. Since the approximate values for the parameters usually deviate more from the estimated parameters, than the observations deviate from their fitted values, the degree of approximating the weight matrix by the one of the first iteration mainly depends on the quality of the approximate values for the parameters.

This type of approximation may refer to the full weight matrix or to its diagonal version, as in CASE C. Here, we assume the constraints are treated as uncorrelated. Then we arrive at the same iteration scheme as Indelman et al. (2012).

We will investigate the effect of these four approximations onto the result as a function of the noise level, namely the assumed variance σ_{0l}^2 of the observations, and the variance σ_{0x}^2 of the approximate values $\boldsymbol{x}^a$. In order to be able to make the two standard deviations comparable, we assume they describe relative uncertainties with unit 1, for angles units radians. The standard deviation σ_{0l} is the directional uncertainty σ_l/c, where σ_l is the standard deviation of the image coordinates and c the focal length.

5.3 Generating Approximate Values with a Specified Relative Precision

We perform tests with simulated data by taking the final estimates of real datasets as true values and artificially generate noisy observations and noisy approximate values. In this section we describe how to generate approximate values for the rotation matrices R_t and the positions $\boldsymbol{Z}_t$ of the projection centers.

The relative precision of directions or angles is easily specified by their standard deviation measured in radians. Hence if we prespecify the *relative precision* of the approximate values with σ_{0x}, e.g. $\sigma_{0x} = 0.01 = 1\,\%$, we just need to deteriorate the rotation axes and rotation angles by zero-mean noise with standard deviation $\sigma_\alpha = \sigma_{0x}$.

The relative precision of the coordinates of a set of camera positions, say $\boldsymbol{Z}_t$, is less clear. We propose to use the standard deviation of the direction vectors $\boldsymbol{D}_{tt'} = (\boldsymbol{Z}_t - \boldsymbol{Z}_{t'})/d_{tt'}$, with the distance $d_{tt'} = |\boldsymbol{Z}_t - \boldsymbol{Z}_{t'}|$ between two neighbouring points $\mathcal{Z}_t$ and $\mathcal{Z}_{t'}$; here we assume isotropic uncertainty. Furthermore we take relative standard deviation of the distance $d_{tt'}$ between neighbouring points, i.e. $\sigma_{r_{tt'}} := \sigma_{d_{tt'}}/d_{tt'}$ as measure.

There is no obvious way to generate a set of points such that the average relative standard deviation of a given point set fulfills this measure. The following approximation appears sufficient for the experiments. We assume the true values of the camera positions are given by $\widetilde{\boldsymbol{Z}}_t$. We distort them by taking them as approximate values. Then we generate disturbing observations, namely the coordinate differences $\boldsymbol{D}_{tt'}$ with a covariance matrix of $\mathbb{D}(\underline{\boldsymbol{D}}_{tt'}) = \sigma_r\ d_{tt'} I_3$, with $\sigma_r = \sigma_{0x}$. We only use pairs $(tt') \in \mathcal{T}$ from a Delaunay triangulation. Since coordinate differences alone do not allow to estimate the coordinates, we fix the gauge by requiring the sum of all estimated coordinates to be zero. Therefore we have the following linear Gauss–Markov model with constraints

$$\boldsymbol{D}_{tt'} = \widehat{\boldsymbol{Z}}_{t'} - \widehat{\boldsymbol{Z}}_t\,,\ \ \mathbb{D}(\underline{\boldsymbol{D}}_{tt'}) = \sigma_r^2\ d_{tt'}^2\ I_3\,,\ \ (tt') \in \mathcal{T}\,, \tag{28}$$

$$\mathbf{0} = \sum_t \widehat{\boldsymbol{Z}}_t\,. \tag{29}$$

Minimizing $\sum_{tt'} |\boldsymbol{D}_{tt'}|^2$ under the constraint leads to estimates $\widehat{\boldsymbol{Z}}_t$. Due to the estimation process, their relative standard deviations will generally be smaller than $\sigma_r d_{tt'}$, for all t' in the neighbourhood of t. Hence, we need to increase the distance of the points $\widehat{\boldsymbol{Z}}_t$ from the true values adequately: By taking the average relative variance

$$\overline{\sigma_r^2} = \frac{\sum_{tt'} |\boldsymbol{D}_{tt'}|^2/d_{tt'}^2}{3\ \sum_{tt'} 1} \tag{30}$$

we can adapt all coordinates by

$$\widehat{\boldsymbol{Z}}_t := \widetilde{\boldsymbol{Z}}_t + \frac{\sigma_r}{\overline{\sigma_r}}(\widehat{\boldsymbol{Z}}_t - \widetilde{\boldsymbol{Z}}_t)\,. \tag{31}$$

and thus achieve $\overline{\sigma_r} = \sigma_r = \sigma_{0x}$.

5.4 Evaluating the Results of the Approximations

As quality measure we use the differences

$$\Delta\widehat{\boldsymbol{x}}_{\text{CASE}} = \widehat{\boldsymbol{x}}_{\text{CASE}} - \tilde{\boldsymbol{x}} \tag{32}$$

between estimated pose parameters $\widehat{\boldsymbol{x}}_{\text{CASE}}$ obtained using the the approximation of a certain case and the true values $\widetilde{\boldsymbol{x}}$. Due to the freedom of choosing seven gauge parameters for BA, we can only compare $U = 6T - 7$ parameters, if we have T unknown poses.

In order to illustrate the loss in accuracy we will employ the root-mean-square error (RMSE) of the of deviations of the coordinates

$$\text{RMSE}_Z = \sqrt{\frac{1}{3T}\sum_t |\widehat{\boldsymbol{Z}}_t - \widetilde{\boldsymbol{Z}}_t|^2} \tag{33}$$

and of the rotation angles

$$\text{RMSE}_R = \sqrt{\frac{1}{6T}\sum_t ||\widehat{\mathsf{R}}_t\widetilde{\mathsf{R}}_t^\mathsf{T} - I_3||^2}\,. \tag{34}$$

We will give the deviation

$$\Delta\text{RMSE}_{\text{CASE}} = \sqrt{\text{RMSE}_{\text{CASE}}^2 - \text{RMSE}_0^2} \tag{35}$$

of the $\text{RMSE}_{\text{CASE}}$ of each case from the RMSE_0 obtained with the rigorous estimation. In addition to the deviation of the RMSE we will report the *loss in accuracy*

$$L_{\text{CASE}} = \frac{\Delta\text{RMSE}_{\text{CASE}}}{\text{RMSE}_0} \tag{36}$$

for each case compared to the ideal solution. Observe, these measures do not take the inhomogeneous precision of the estimates into account and depend on the chosen gauge.

Therefore we also provide the squared normalized Mahalanobis distance

$$F_{\text{CASE}} = \frac{1}{U}\Delta\widehat{\boldsymbol{x}}_{\text{CASE}}^\mathsf{T}\,\Sigma_{\widehat{\boldsymbol{x}}\widehat{\boldsymbol{x}}}^{-1}\,\Delta\widehat{\boldsymbol{x}}_{\text{CASE}}\,|H_0 \backsim F(U,\infty) \tag{37}$$

with the covariance matrix $\Sigma_{\widehat{x}\widehat{x}}$ of the parameters obtained using the rigorous estimation. The squared normalized Mahalanobis

distance F is a test statistic which follows a Fisher distribution $F(U,\infty)$, if the model is statistically optimal, which is the zero hypothesis H_0. It is a sufficient test statistic and does not depend on the chosen gauge. The expected value for F is 1, the one-sided confidence interval for a significance level S is $[0, F(U,\infty,S)]$; we use $S = 0.99$ in the following.

In addition to the Fisher test statistic F we also give the *loss in accuracy related to the Fisher test statistic*

$$\Delta F_{\text{CASE}} = \sqrt{F_{\text{CASE}} - 1} = \frac{\sigma_{\widehat{x}_{b,\text{CASE}}}}{\sigma_{\widehat{x}}} \,. \tag{38}$$

Hence, we assume the loss in accuracy is induced due to a bias $\boldsymbol{x}_{b,\text{CASE}}$ caused by the approximation, so that $\widehat{\boldsymbol{x}}_{\text{CASE}} = \widehat{\boldsymbol{x}} + \widehat{\boldsymbol{x}}_{b,\text{CASE}}$.

6. EXPERIMENTS

Our experimental evaluation is designed to investigate the accuracy decrease of BA when applying the individual approximations proposed in Sec. 5.2, which are meant to increase efficiency. We illustrate the loss in accuracy as a function of the noise level, namely the standard deviation of observations, and on the relative precision of camera poses. We evaluate the individual simplifications on two image sequences recorded on different UAVs.

The first image sequence BUILDING contains 119 images taken with a 5 MPixel camera with a focal length of $c = 1587.87$ pixel on a 5 kg UAV platform, triggered each second. The flight was guiding the UAV along the facade of a house, the variation in position is around 60 m and 15 m in height, see Figure 2.

The second image sequence FIELD contains 24 images taken with a 12 MPixel camera of the *DJI Panthom 4* with a focal length of $c = 2347.1$ pixel. The camera was pointing downwards while the copter was flying a meandering pattern at 100 m height with three stripes, each stripe consists of eight images, see Figure 3. The images have an front and sidelap of 80 %, this way the 24 images cover an area of $100 \times 90\,\text{m}^2$.

For both datasets we match interest points in the images to obtain corresponding image points. In order to determine the simplification effects, we need ground truth for camera poses and corresponding image points, to incorporate deteriorations under controlled conditions. We use the observed image coordinates as input for the BA software *BACS* (Schneider et al., 2012) to obtain estimated pose parameters and fitted image points for two realistic UAV flight scenarios, which are consistent and are used as ground truth.

6.1 Checking the Rigorous Reference Solution

First we check how the rigorous trifocal BA reacts on different noise level σ_{0l} of observed image points on the BUILDING dataset. The estimated variance factor $\widehat{\sigma}_0^2$, see Eq. (19), needs to become one, in case the noise level σ_{0l} used to deteriorate the image points is used also in the covariance matrix Σ_{ll}^a in Eq. (10). We follow Sec. 5.3 to generate deteriorated approximate values by using a moderate relative precision of $\sigma_{0x} = 0.001$. We deteriorate the observations on each noise level 100 times with different random noise and apply the trifocal BA. Figure 4 shows the mean of the obtained standard deviation $\widehat{\sigma}_0$ of estimated variance factors using different noise levels σ_{0l} to disturb the observed image points. Having high noise we observe, that $\widehat{\sigma}_0$ deviates from one, as then second order effects, which are neglected in the estimation procedure of BA, become visible. The effects are negligibly small and within the tolerance bounds [0.9943, 1.0058] of

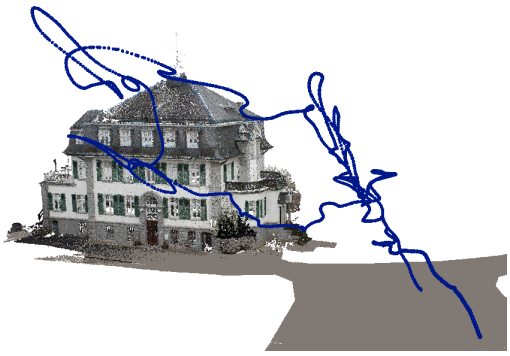

Figure 2. Trajectory of the UAV flight capturing the images of the BUILDING dataset overlayed with a 3D model of a near-by building.

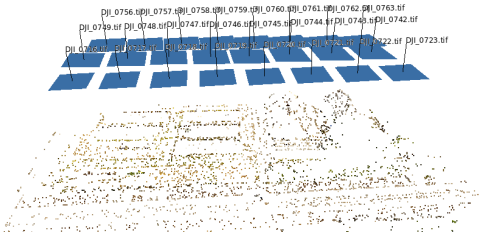

Figure 3. Ground truth camera poses and 3D points of the FIELD dataset.

the fisher test using a significance level of 1 %. For the following evaluation of the accuracy decrease of the individual approximations we will use a maximum noise level of $\sigma_{0l} = 0.003$.

Figure 5 shows the mean and standard deviation of the number of iterations until BA achieves convergence under different noise levels. The number of necessary iterations increases with the noise level as expected. Convergence is achieved if all corrections

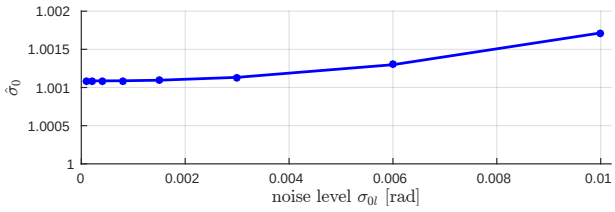

Figure 4. The standard deviation $\widehat{\sigma}_0$ of the estimated variance factor at different noise level σ_{0l}. With focal length $c = 1587.87$ pixel, 0.001 radian corresponds to an uncertainty of 1.5 pixel in the image points.

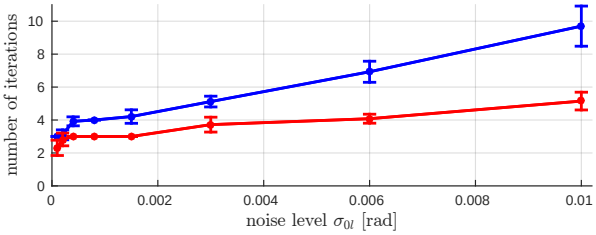

Figure 5. The number of iterations needed to achieve convergence at different noise level σ_{0l} and moderate relative precision of $\sigma_{0x} = 0.001$ when using $T_c = 0.001$ (blue line) or $T_c = 0.1$ (red line).

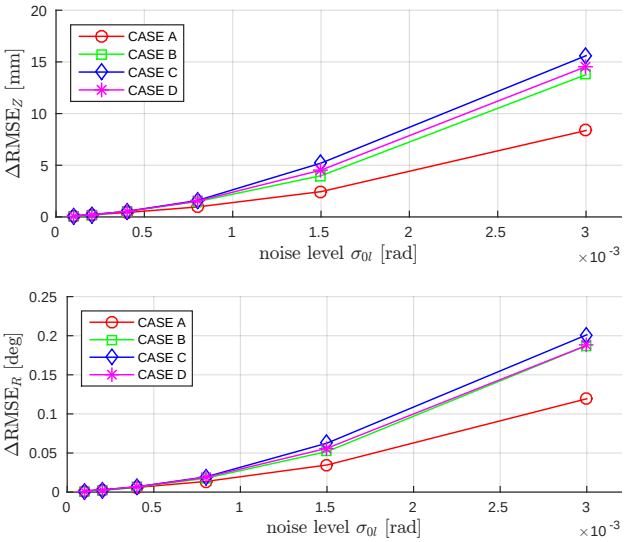

Figure 6. Deviations between the RMSE of estimated camera positions and rotations of ideal result and approximations CASE A-D at different noise levels σ_{0l} of dataset BUILDING.

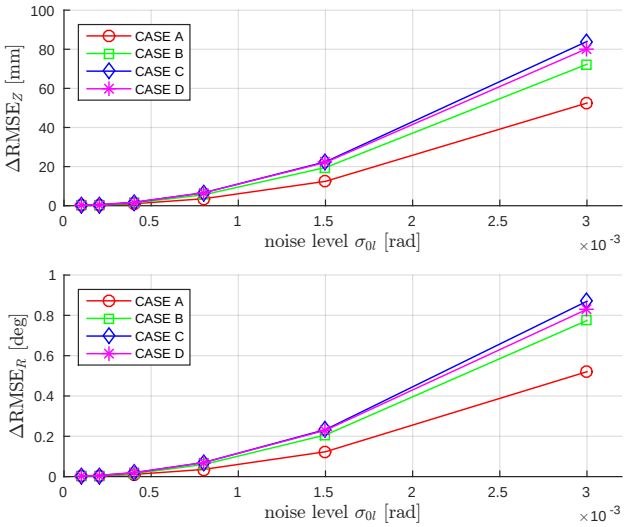

Figure 7. Deviations between the RMSE of estimated camera positions and rotations of ideal result and approximations CASE A-D at different noise levels σ_{0l} of dataset FIELD.

for observations $\widehat{\Delta l}_n$ and parameters $\widehat{\Delta x}_u$ are small compared to their standard deviation, $|\widehat{\Delta l}_n/\sigma_{0l}| < T_c$, $|\widehat{\Delta x}_u/\sigma_{x_u}| < T_c$, with a threshold $T_c = 0.001$, thus requiring the corrections to be less than 0.1 % of their standard deviation. This requires σ_{x_u} to be known, which is for this experiment derived in each iteration from the inverse normal equation matrix.

6.2 Effects of Approximations

We now experimentally evaluate the decrease of accuracy when applying the individual approximations for BA proposed in Sec. 5.2 on the two image sequences recorded by UAVs.

We add normal distributed noise to the true observation values with different magnitudes to obtain different noise levels. We use a moderate relative precision of $\sigma_{0x} = 0.001$ to deteriorate the approximate values of the camera poses. After that we optimize the pose parameters with the rigorous estimation, which is called CASE 0 in the following, and with the approximations of CASE A-D. With the estimated and true pose parameters we can determine the root mean square error of the estimated camera positions and rotations according to Eq. (33) and Eq. (34). For each noise level we randomly generate 100 times different noise for the observations and determine the RMSE for each case. Figure 6 and Figure 7 give the mean of the deviations of the RMSE of CASE A-D to the ideal result of CASE 0 obtained with Eq. (35) under different noise levels σ_{0l}.

In both datasets the approximation made in CASE A induces the smallest deviations to the optimally estimated coordinates and rotations in both datasets, the deviations induced by the approximation of CASE B are almost twice as big. CASE C, which contains the approximations of CASE A and CASE B, shows slightly higher deviations than CASE B, thus is mainly affected by the approximations of CASE B. CASE D shows smaller deviations than CASE C, even though it contains an additional approximation. The reason could be the high relative precision σ_{0x} used in this experiment. The convergence of CASE D is affected by the relative precision as the weight matrix $\boldsymbol{W}_{gg}$ is fixed after the first iteration, while CASE A-C are not affected. Thus we will investigate the decrease in precision of CASE D by varying σ_{0x} in a further experiment.

Figure 8 gives the loss in accuracy L_{CASE} in percent, which can be obtained with Eq. (36), under different noise level σ_{0l} for each approximation. Both datasets recorded on different UAVs, flight trajectories and cameras show nearly the same loss in the accuracy of the estimated poses due to the approximations.

The RMSE does not take the inhomogeneous uncertainty of the estimated positions and rotations of all images into account and depends on the chosen gauge. Thus we use the squared normalized Mahalanobis distance given in Eq. (37) which considers the covariance information of the parameters, which are obtained by the rigorous estimation. Table 1 lists the mean loss in accuracy ΔF_{CASE} of the estimated pose parameters in percent when applying the individual approximation cases under different noise levels σ_{0l}. The loss of accuracy ΔF_{CASE} is obtained with Eq. (38). The obtained values are similar to the values obtained by using the root mean square error.

CASE A, CASE B and therefore also CASE C are mainly affected by the noise level σ_{0l}, whereas CASE D is affected from both, noise level σ_{0l} and the relative precision σ_{0x} of the approximate values. Therefore we also investigate the average decrease in precision by varying σ_{0x} from a moderate relative precision of 0.1 % to an inferior relative precision of 10 % to deteriorate the approximate values of the camera poses. We use a moderate noise level

$\sigma_{0l} =$	0.0001	0.0002	0.0004	0.0008	0.0015	0.0030
BUILDING						
CASE A	5.49	6.05	6.79	12.29	18.21	25.67
CASE B	9.87	11.34	11.73	14.75	21.90	36.92
CASE C	9.82	11.40	14.10	14.32	32.96	41.83
CASE D	9.82	11.39	13.09	14.46	26.49	38.43
FIELD						
CASE A	4.52	5.24	6.31	10.96	16.09	27.86
CASE B	11.36	11.45	11.21	12.76	20.86	33.03
CASE C	11.50	11.76	11.74	16.32	24.95	43.86
CASE D	11.50	11.76	11.74	16.30	24.75	42.46

Table 1. The loss in accuracy ΔF_{CASE} in percent of estimated pose parameters induced by the individual approximations of CASE A-D at different noise levels σ_{0l} in radian.

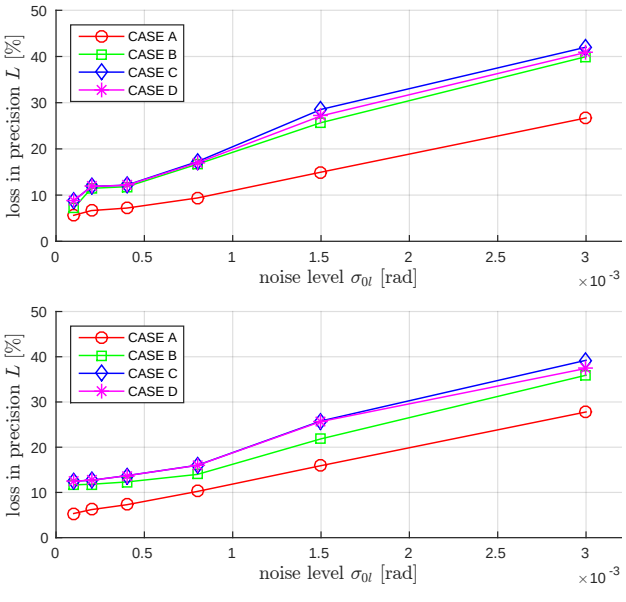

Figure 8. The loss in accuracy L_{CASE} in percent of approximations CASE A-D compared to ideal solution at different noise levels σ_{0l} in radian on dataset BUILDING (top) and FIELD (bottom).

of $\sigma_{0l} = 0.001$ for the observations by adding normal distributed noise to the true observation values with different magnitudes to obtain different noise levels. Note that the estimated parameters converge differently when applying randomly generated approximate values. Thus, we randomly generate 100 times different noise for the observations and approximate values and determine the mean loss of accuracy $\Delta F_{\text{CASE D}}$ for each relative precision level σ_{0x}. Table 2 shows the obtained ΔF_{CASE}, which increases with the relative precision σ_{0x} of the approximate values for the camera poses.

In both experiments the decrease in accuracy is less than a 1/3 of noise variance. This is a moderate loss, and – if computing time is essential — may be accepted.

7. CONCLUSION

In this paper, we presented an approach to bundle adjustment without structure estimation by employing epipolar and trifocal constraints between corresponding image points. We introduced a novel closed-form expression for the trifocal constraint, which does not degenerate at certain configurations. The proposed bundle adjustment is as optimal as classical bundle adjustment, but leads to more computational complexity as the Gauss-Helmert model needs to be employed for optimization.

We evaluated the quality decrease of simplifying approximations which allow to employ the Gauss-Markov model to increase the computational efficiency on two datasets acquired by UAVs. The empirically investigated loss in accuracy of the estimated camera pose parameters are shown to be small in case of small noise in the observations.

In spite of this favorable result w.r.t. the investigated approximations, the effect of the approximations onto outlier detection, which relies on the variances of the residuals needs to be investigated, in order to identify the loss in the power of outlier detection methods.

$\sigma_{0x} =$	0.001	0.003	0.01	0.03	0.1
BUILDING					
CASE D	9.82	10.52	17.16	20.97	31.44
FIELD					
CASE D	11.50	13.28	14.52	16.94	25.36

Table 2. The loss in accuracy $\Delta F_{\text{CASE D}}$ in percent of estimated pose parameters at noise level $\sigma_{0l} = 0.0001$ and different relative precision σ_{0x} of approximate values, both in radian.

ACKNOWLEDGMENTS

This work has partly been supported by the DFG under the grant number FOR 1505: Mapping on Demand.

References

Agarwal, S., Furukawa, Y., Snavely, N., Simon, I., Curless, B., Seitz, S. and Szeliski, R., 2011. Building rome in a day. *Communications of the ACM (CACM)*.

Förstner, W. and Wrobel, B., 2016. *Photogrammetric Computer Vision – Statistics, Geometry, Orientation and Reconstruction*. Springer.

Hartley, R. and Zisserman, A., 2004. *Multiple View Geometry in Computer Vision*. 2nd edn, Cambridge University Press.

Indelman, V., 2012. Bundle adjustment without iterative structure estimation and its application to navigation. In: *Proc. of the Position Location and Navigation Symposium*, pp. 748–756.

Indelman, V., Roberts, R., Beall, C. and Dellaert, F., 2012. Incremental light bundle adjustment. In: *Proc. of the British Machine Vision Conference*, pp. 134.1–134.11.

Kaess, M., Johannsson, H., Roberts, R., Ila, V., Leonard, J. and Dellaert, F., 2012. iSAM2: Incremental Smoothing and Mapping Using the Bayes Tree. *Intl. Journal of Robotics Research (IJRR)* 31(2), pp. 217–236.

Konolige, K., 2010. Sparse sparse bundle adjustment. In: *Proc. of the British Machine Vision Conference*, pp. 102.1–102.11.

Kümmerle, R., Grisetti, G., Strasdat, H., Konolige, K. and Burgard, W., 2011. G2o: A general framework for graph optimization. In: *Proc. of the IEEE Intl. Conf. on Robotics & Automation (ICRA)*, pp. 3607–3613.

Lourakis, M. and Argyros, A., 2009. Sba: A software package for generic sparse bundle adjustment. *ACM Trans. on Mathematical Software (TOMS)* 36(1), pp. 1–30.

Rodríguez, A., de Teruel, P. L. and Ruiz, A., 2011. Reduced epipolar cost for accelerated incremental sfm. In: *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, pp. 3097–3104.

Schneider, J., Schindler, F., Läbe, T. and Förstner, W., 2012. Bundle adjustment for multi-camera systems with points at infinity. In: *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. I-3, pp. 75–80.

Stark, E. and Mikhail, E., 1973. Least Squares and Non-Linear Functions. *Photogrammetric Engineering* 39, pp. 405–412.

Steffen, R., Frahm, J.-M. and Förstner, W., 2010. Relative bundle adjustment based on trifocal constraints. In: *Trends and Topics in Computer Vision*, Lecture Notes in Computer Science (LNCS), Vol. 6554, pp. 282–295.

Triggs, B., McLauchlan, P., Hartley, R. and Fitzgibbon, A., 2000. Bundle adjustment – a modern synthesis. In: *Vision Algorithms: Theory and Practice*, Lecture Notes in Computer Science (LNCS), Vol. 1883, pp. 298–372.