

RESEARCH

Open Access



Robust error estimation based on factor-graph models for non-line-of-sight localization

O. Arda Vanli^{1*}  and Clark N. Taylor²

*Correspondence:

oavanli@eng.fsu.edu

¹ Department of Industrial and Manufacturing

Engineering, Florida A&M

University, Florida State

University, Tallahassee, FL, USA

Full list of author information is available at the end of the article

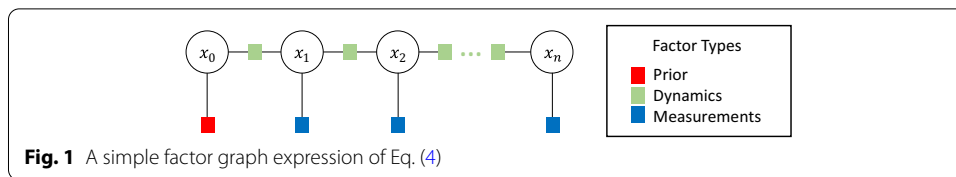
Abstract

This paper presents a method to estimate the covariances of the inputs in a factor-graph formulation for localization under non-line-of-sight conditions. A general solution based on covariance estimation and M-estimators in linear regression problems, is presented that is shown to give unbiased estimators of multiple variances and are robust against outliers. An iteratively re-weighted least squares algorithm is proposed to jointly compute the proposed variance estimators and the state estimates for the nonlinear factor graph optimization. The efficacy of the method is illustrated in a simulation study using a robot localization problem under various process and measurement models and measurement outlier scenarios. A case study involving a Global Positioning System based localization in an urban environment and data containing multipath problems demonstrates the application of the proposed technique.

Keywords: Bayesian estimation, Robust estimation, Multipath, Factor graphs

1 Introduction

An important challenge in Global Positioning System (GPS)-based localization is the non-line-of-sight or multipath problem, commonly encountered in urban environments. When the direct line of sight to a satellite is blocked by a building, its signal may reach the receiver on the ground via reflections from buildings, resulting in large errors or outliers in the pseudorange measurements [1]. Least squares-based localization methods that typically rely on Gaussian model errors can be severely distorted due to outliers, resulting in poor localization accuracy. Some robust estimation approaches in the recent literature to remedy this issue are based on data weighting: M-estimators that rely on downweighting of outlying observations [2], mixture distributions that explicitly model the outlying observations in the sensor model [3], switchable constraints that utilize switch variables to down-weight individual pose constraints [1, 4], dynamic covariance scaling [5] and Receiver Autonomous Integrity Monitoring (RAIM) that monitors the integrity of satellites [6]. However, none of these robust estimation approaches considered the unbiasedness of the variance estimators.



Many data fusion applications require combining measurements from heterogeneous sensors or matching features of multiple measurements, both tasks requiring an accurate characterization of the noise covariance matrices of measurements. In particular, GPS-based localization in urban or contested environments relies on fusion of measurements from inertial measurement unit (IMU) or light detection and ranging (LiDAR) systems in the navigation solution [7] to mitigate multipath problems. In vision-based localization, combining sequential measurements requires landmarks identified in these measurements to be matched, which assumes accurate knowledge of covariances [8, 9]. One of the common assumptions made in Bayesian state estimation of dynamic processes is that the covariance matrices for the noise sources are known or some reliable estimates exist a priori. Unfortunately, accurate a priori knowledge or estimates of the covariance matrices in practice may not be available and inaccurate covariance estimates can lead to a significant degradation in the estimation quality of the system states. In addition, as the applications of Bayesian estimation move from high-quality, expensive systems (e.g., the Apollo mission) to lower-cost systems with lower-quality sensors (e.g., cell phones), the ability to accurately characterize the uncertainty of all inputs to the system become both less repeatable across a class of sensors and less economically feasible.

In this paper, we propose new unbiased estimators of the noise covariances in a factor graph formulation when sensor data is contaminated with outliers and Gaussianity of data is not satisfied, particularly applicable to the GPS-based localization problem under non-line-of-sight conditions. The contributions of the paper are the extending of the unbiased variance estimators for factor-graph problem proposed in [10] to the multipath problem, the study of both linear and nonlinear vehicle motion models, and the investigation of the performance of the approach with real GPS data. A general solution, based on nonlinear regression and robust estimation is proposed that is shown to give unbiased estimators of the multiple variances in factor graph formulation. To jointly compute the proposed variance estimators and the state estimates, an iteratively re-weighted least squares (IRLS) algorithm is presented. By contrast to the existing approaches that rely on the maximum likelihood principle and use sample variances of the residuals to estimate noise variances, the primary contribution of the paper is to incorporate unbiased estimate of the noise variances, which, as will be illustrated, can achieve significant improvement in localization accuracy over the existing approaches.¹

2 Review of relevant literature

Factor graphs have become a popular smoothing-based alternative to sequential estimators of dynamic systems. Figure 1 illustrates the factor graph representation of a robot localization problem, in which states x_1, \dots, x_n form a sequence of n unknown robot

¹ The software for this study can be obtained at <https://github.com/avanli/Factor-Graph-Covar>.

locations. The factor graph provides a convenient way to represent the multiplicative relation of the “factors” (shown as boxes with different colors), or the probabilistic representations of states and measurements based on prior beliefs, process dynamics and measurement equations, and construct the optimization problem from which to solve for the robot locations.

In the simultaneous localization and mapping (SLAM) problem, where the factor graph formulation is commonly applied, it was shown that sequential state estimators may provide inconsistent solutions due to the inherent model nonlinearities and the accumulating effects of the linearization choices [11]. There has been a recent interest in smoothing as a fast alternative to sequential estimation, with its advantage of retaining the entire robot trajectory that significantly helps the performance [12]. Based off prior smoothing work [13, 14], the factor graph formulation of the estimation problem enables batch optimization of the system state estimates leading to better performance with nonlinear systems. In addition, the factor graph formulation easily handles situations where the type, frequency or quantity of measurements may change over time. In the SLAM problem, the subset of landmarks being tracked changes over time, meaning that the portions of the state that affect the measurement differ. This complex interaction of the state with the measurements is something easily handled in the factor graph formulation. Despite factor graph implementations being batch approaches, the sparsity of the matrix used to represent the factor graph leads to highly efficient implementations, including many that can be performed in real-time (e.g., [15]).

To estimate the states of the factor graph formulation, several Bayesian estimation algorithms have been introduced, including the sequential approaches of (extended, unscented, ensemble) Kalman filter and particle filter and smoothing-based techniques [16]. For non-line-of-sight localization, robust estimation and weighted least squares methods were investigated to mitigate multipath effects [17, 18]. Recently, methods based on Gaussian mixture models (GMM) and expectation–maximization (EM) framework have been proposed for robust estimation of both the states and noise covariances in a factor graph framework [19–22]. Pfeifer et al. [20] formulated the problem as a maximum likelihood problem and used an EM to achieve robustness against non-Gaussianity. By contrast [19, 21, 22] formulated the problem as a Bayesian estimation using maximum a posteriori (MAP) estimator and non-informative priors. As it is well known, a Bayesian MAP estimation with non-informative priors is equivalent to maximum likelihood inference of the joint conditional probability of the measurements.

As the review outlined, most of the existing robust estimation methods relied on sample covariances of the residuals, which maximizes the likelihood given the state estimates, as the direct estimators of the noise covariance matrices. As we show in the paper, these estimators are biased [23, p. 103] and may lead to a significant under-estimation of the covariance matrices, adversely impacting the localization accuracy. Other prior efforts focused on estimating the *state covariance* for factor graphs (see e.g., [8]) all assume accurate *input noise covariances* are known a-priori. Our focus in this paper, by contrast, is on the estimation of *input noise* (measurement and process) covariance [24] for the factor graph.

3 Review of factor graph formulation

We first review the existing factor graph formulation of the state estimation (localization) and the noise variance estimation problems. Consider a discrete-time dynamic system:

$$x_t = f(x_{t-1}, u_t) + v_t, \quad v_t \sim N(0, Q) \quad (1)$$

$$z_t = g(x_t) + w_t, \quad w_t \sim N(0, R) \quad (2)$$

where $x_t \in \mathbb{R}^{d_x}$ and $u_t \in \mathbb{R}^{d_u}$ are the state and control input vectors at time t , respectively, $z_t \in \mathbb{R}^{d_z}$ is the sensor measurement vector, with d_x, d_u and d_z representing the dimensions of the vectors, $f(x, u)$ is the process dynamics model representing what the state is at the next timestep given the current state and inputs, $g(x)$ is the measurement model giving the sensor measurement at the current time given the current state, and v_t and w_t are white, zero-mean Gaussian noise sources with covariance matrices Q and R , respectively. For the GPS-based localization problem, the states x_t may comprise the location coordinates of the vehicle and the measurements z_t may comprise vehicle odometry data or pseudorange measurements acquired from visible satellites. The measurements are conditionally independent. Localization problem consists of estimating the states, that is, solving for the sequence of vehicle locations x_1, \dots, x_n on the basis of a sequence of measurements z_1, \dots, z_n where n is the number of measurements available.

3.1 State estimation

To find the estimators of the system states, the following optimization problem is solved [15]:

$$X^* = \arg \max_X p(x_0) \prod_{t=1}^n p(x_t | x_{t-1}, u_t) \prod_{t=1}^n p(z_t | x_t), \quad (3)$$

where x_0 is the initial pose, and $X = \{x_t\}_{t=0}^n$ is the sequence of system states to be solved. This optimization problem can also be expressed as a factor graph as shown in Fig. 1, where the factors (shown as squares) connect only with the state (or portion of the state) that are required for that measurement/dynamic equation to be evaluated.

Based on the Gaussian process dynamics and measurement models, the probability densities shown in the factorization are $p(x_t | x_{t-1}, u_t) \propto \exp -\frac{1}{2} \|f(x_{t-1}, u_t) - x_t\|_Q^2$ and $p(z_t | x_t) \propto \exp -\frac{1}{2} \|g(x_t) - z_t\|_R^2$, where the sign \propto means “proportional to” and implies the density functions are given up to a proportionality constant. Note that because Q and R are assumed constant across time, the proportionality constants can be safely ignored. Because $\log(x)$ is a strictly monotonic function in x , we can rewrite Eq. (3) as:

$$\arg \min_X \|x_0 - \mu_0\|_{\Sigma_0}^2 + \sum_{t=1}^n \|f(x_{t-1}, u_t) - x_t\|_Q^2 + \sum_{t=1}^n \|g(x_t) - z_t\|_R^2 \quad (4)$$

assuming $p(x_0) \sim N(\mu_0, \Sigma_0)$, where μ_0 and Σ_0 are the mean and covariance of the initial pose. Since the model equations can be nonlinear in general, the states are solved for by linearizing the process and measurement equations using a first-order Taylor series expansion around an operating point $\{x_t^0\}_{t=0}^n$ as

$$f(x_{t-1}, u_t) - x_t \cong f(x_{t-1}^0, u_t) + F(x_{t-1}^0) \Delta x_{t-1} - x_t^0 - \Delta x_t \quad (5)$$

$$g(x_t) - z_t \cong g(x_t^0) + G(x_t^0) \Delta x_t - z_t \quad (6)$$

where $F(x_{t-1}) = \partial f(x_{t-1}, u_t) / \partial x_{t-1}$ and $G(x_t) = \partial g(x_t) / \partial x_t$ are the Jacobian matrices of the process and measurement equations, respectively, and $\Delta x_t = x_t - x_t^0$, $\Delta x_{t-1} = x_{t-1} - x_{t-1}^0$. Defining the vectors $a_t^0 \triangleq x_t^0 - f(x_{t-1}^0, u_t)$ and $c_t^0 \triangleq z_t - g(x_t^0)$ and letting $\Delta x_0 = x_0 - x_0^0$, a linear least squares representation of the problem in (4) is obtained as:

$$\arg \min_{\Delta X} \|\Delta x_0 + x_0^0 - \mu_0\|_{\Sigma_0}^2 + \sum_{t=1}^n \|F_{t-1} \Delta x_{t-1} - a_t^0 - \Delta x_t\|_Q^2 + \sum_{t=1}^n \|G_t \Delta x_t - c_t^0\|_R^2. \quad (7)$$

where $\Delta X = \{\Delta x_t\}_{t=0}^n$, $F_{t-1} \triangleq F(x_{t-1}^0)$, and $G_t \triangleq G(x_t^0)$. Each iteration of the state estimation problem can be written as that of finding the solution of the following system of linear equations:

$$b = A \Delta X + \epsilon \quad (8)$$

where $A \in \mathbb{R}^{m \times p}$ is the coefficient matrix, $b \in \mathbb{R}^m$ is the standardized measurement vector, $\epsilon \in \mathbb{R}^m$ is the vector of least squares solution errors, $m = (n+1)d_x + nd_z$ is the number of observations and $p = (n+1)d_x$ is the number of states. If the initial pose x_0 is known and not to be included in the estimation problem then $X = \{x_t\}_{t=1}^n$ are the states to be solved for, $m = nd_x + nd_z$ and $p = nd_x$.

The usual practice in factor graph models is to follow sparse linear algebra and solve for ΔX using a matrix factorization algorithm such as QR or Cholesky factorization followed by forward or back substitution [12]. The state estimates for the next iteration are found as $X \leftarrow X + \Delta X$, and the new matrices A and b are found at the linearization point of X . These iterations are continued until convergence. However, to facilitate stating the problem and proving the unbiasedness of the variance estimates we will show in the proposed method (see Sect. 4.1) the solution of ΔX in terms of a generalized inverse, such as the Moore–Penrose pseudo-inverse (typically $(A^T A)^{-1} A^T$).

As an example, for $n = 4$ poses, the factor-graph model given in Fig. 1 is written as

$$A = \begin{bmatrix} -\Sigma_0^{-\frac{1}{2}} & 0 & 0 & 0 & 0 \\ Q^{-\frac{1}{2}} F_1 & -Q^{-\frac{1}{2}} & 0 & 0 & 0 \\ 0 & Q^{-\frac{1}{2}} F_2 & -Q^{-\frac{1}{2}} & 0 & 0 \\ 0 & 0 & Q^{-\frac{1}{2}} F_3 & -Q^{-\frac{1}{2}} & 0 \\ 0 & 0 & 0 & Q^{-\frac{1}{2}} F_4 & -Q^{-\frac{1}{2}} \\ 0 & R^{-\frac{1}{2}} G_1 & 0 & 0 & 0 \\ 0 & 0 & R^{-\frac{1}{2}} G_2 & 0 & 0 \\ 0 & 0 & 0 & R^{-\frac{1}{2}} G_3 & 0 \\ 0 & 0 & 0 & 0 & R^{-\frac{1}{2}} G_4 \end{bmatrix}, \quad b = \begin{bmatrix} \Sigma_0^{-\frac{1}{2}} (x_0 - \mu_0) \\ Q^{-\frac{1}{2}} a_1 \\ Q^{-\frac{1}{2}} a_2 \\ Q^{-\frac{1}{2}} a_3 \\ Q^{-\frac{1}{2}} a_4 \\ R^{-\frac{1}{2}} c_1 \\ R^{-\frac{1}{2}} c_2 \\ R^{-\frac{1}{2}} c_3 \\ R^{-\frac{1}{2}} c_4 \end{bmatrix}. \quad (9)$$

3.2 Noise variance estimation

In the existing smoothing-based factor graph methods [19–22], the measurement noise covariances are estimated iteratively in conjunction with the state estimates, following a Gaussian mixture model (GMM) and an expectation and maximization (EM) approach. The maximization step in EM maximizes the likelihood, which minimizes the negative log likelihood (4), given current state estimates, to find the variances. This maximum likelihood (ML) estimator of variances estimates the variances as the sample variance of the measured state estimation error sequences $e_t = x_t - \hat{x}_t$ [25, p. 79] as

$$\sigma^2 = \frac{1}{nd_x} \sum_{t=1}^n e_t^T e_t. \quad (10)$$

A separate maximum likelihood estimate with the residuals of the initial pose, process dynamics and measurements are calculated, respectively, and these estimates form the constant diagonal elements of the covariance matrices Σ_0 , Q and R . To make the estimates robust against outliers, Sunderhauf et al. [1, 4] introduced the concept of switchable constraints (SC), which utilizes an augmented optimization problem to concurrently solve for states and weighting values of loop closure or prior constraints. The SC approach, however, adopts ML estimates of the noise variances.

A drawback of the ML estimators, such as (10), in the context of localization is that they are asymptotically (as the data size grows) unbiased estimates of the true variances [24]. However, as we show in our derivations and results below, the ML estimators are *biased* [25] and lose their effectiveness when data size is relatively small compared to the size of the state vector, a common situation in factor graph models where the number of states grow with data size. In control theory, correlation methods, which analyze the innovation sequence of a linear estimator, have been used to provide unbiased estimates of the noise variances of state space models for small data sizes [26]. However, such methods do not make any assumptions about the noise probability density functions and do not consider robust estimation to handle outliers and hence are not applicable to non-line-of sight conditions. In the present paper, we develop unbiased estimators of the multiple noise variances of the factor graph representation of dynamic systems, with a particular focus on GPS-based localization under non-line of sight conditions. Another important contrast of the proposed method over the existing methods [1, 4, 19–22] is that, while these methods present methodology to estimate the measurement noise covariance R , none provide explicit estimates of the covariances of the prior Σ_0 or of the process dynamics Q . By contrast, the proposed method is that it provides estimates of all these parameters, in addition to ensuring the unbiasedness of the resulting estimators.

4 Methods

In this section, we present the proposed unbiased estimators for the noise variances of the factor graph problem and an iteratively reweighted least squares algorithm for jointly finding the variance and state estimates in the robust M-estimation framework.

4.1 Unbiased noise variance estimation in factor graph problem

In a linear regression problem, a set of parameters Θ for the model $y = B\Theta + \epsilon$ are estimated, where $y \in \mathbb{R}^m$ is the measurement vector containing m observations, $B \in \mathbb{R}^{m \times p}$ is the regressor matrix for p regressor variables and $\epsilon \in \mathbb{R}^m$ is the vector of Gaussian distributed errors with 0 mean and a common variance of σ^2 . It is clear that the linearized factor graph formulation (8) is a linear regression problem, where b , A and ΔX correspond to y , B and Θ , respectively.

In linear regression problems, the variance of the measurements y is estimated based on the residuals of the fitted regression model, expressed as [25]

$$r = y - B\hat{\Theta} = (I - \tilde{H})y \quad (11)$$

where $\hat{\Theta}$ is the estimated parameter vector and $\tilde{H} = B(B^T B)^{-1} B^T$ is an orthogonal projection matrix and the superscript T denotes the matrix transpose. From this equation, the sum of squared residuals $r^T r$ is

$$r^T r = y^T (I - \tilde{H})y. \quad (12)$$

The expectation of the random quantity $r^T r$ with respect to the Gaussian probability distribution of the measurements can be found as [25, pp. 554–555]

$$E[r^T r] = \sigma^2(m - p)$$

where $p = \text{rank}(\tilde{H})$. Following the “method of moments approach” for estimation [27], an unbiased estimate of the variance is [25, p. 77]

$$\sigma^2 = \frac{r^T r}{m - p}.$$

There are two assumptions in linear regression work, however, that differ from the factor graph formulation. First, linear regression assumes that there is a single covariance for all inputs. The factor graph formulation, however, has at least two different covariances (R and Q), and often many more. Second, the residuals computed in a factor graph are pre-weighted by the current covariance estimates. In our proposed method, we extend the linear regression based method for finding variances to the factor graph formulation of the system state estimation problem by incorporating these two important aspects.

Assume a reordering of the factor graph linear system (8) and divide the equations in to s partitions as

$$\begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_s \end{bmatrix} = \begin{bmatrix} A_1 \\ A_2 \\ \vdots \\ A_s \end{bmatrix} \Delta X + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_s \end{bmatrix} \quad (13)$$

where b_1, \dots, b_s correspond to the standardized measurements of the partitions containing n_1, \dots, n_s elements such that $\sum_{i=1}^s n_i = m$ and $\epsilon_1, \dots, \epsilon_s$ are mutually un-correlated random errors of the partitions with 0 mean and unit variances. Each partition has its own covariance matrix before standardization. For example, the system described in Eq. (9) can be split into $s = 3$ partitions corresponding to unstandardized measurement

noise covariances Σ_0 , Q and R , where σ_1^2 , σ_2^2 and σ_3^2 are the constant diagonal elements of the respective 3 matrices.

The proposed method will provide unbiased estimators of the (unstandardized) measurement variances $\sigma_1^2, \dots, \sigma_s^2$. The standardized measurement vector b is a random variable with an identity covariance matrix, that is, $\Sigma_b = \text{diag}(I_{n_1}, \dots, I_{n_s})$, where I_{n_i} is an n_i dimensional identity matrix. The unit variances of the standardized partitions are satisfied if the estimated noise variances $\sigma_1^2, \dots, \sigma_s^2$ are correct. However, before convergence is achieved (when noise variances are incorrect) the standardized measurements has the covariance matrix

$$\Sigma_b = \text{diag}(k_1 I_{n_1}, \dots, k_s I_{n_s}). \quad (14)$$

Therefore, if $k_i, i = 1, \dots, s$ converge to 1 then the standardized noise variances converge to unity and the unstandardized noise variances converge to the true values $\sigma_1^2, \dots, \sigma_s^2$. In our approach, to accomplish convergence, the variance for the next iteration is found by scaling the variance in the current iteration with k_i as

$$\sigma_i^2 \leftarrow k_i \sigma_i^2 \text{ for } i = 1, \dots, s. \quad (15)$$

To do this scaling, we must estimate k_i in each iteration. Let us define $H = I - \tilde{H}$ where $H \in \mathbb{R}^{m \times m}$ and $\tilde{H} = A(A^T A)^{-1} A^T$ is the projection matrix found by the coefficient matrix A obtained with the standardized measurements and the covariance estimates in the current iteration. Partition H as

$$H = \begin{bmatrix} H_{11} & H_{12} & \dots & H_{1s} \\ H_{21} & H_{22} & \dots & H_{2s} \\ \vdots & \vdots & \ddots & \vdots \\ H_{s1} & H_{s2} & \dots & H_{ss} \end{bmatrix} = \begin{bmatrix} H_1 \\ H_2 \\ \vdots \\ H_s \end{bmatrix}. \quad (16)$$

where $H_{ij} \in \mathbb{R}^{n_i \times n_j}$ and $i, j = 1, 2, \dots, s$. Note that because H is symmetric, $H_{ji} = H_{ij}^T$. The residual vector for the i th partition, $r_i \in \mathbb{R}^{n_i \times 1}$, is obtained as

$$\begin{aligned} r_i &= H_i b \\ &= [H_{i1} \ H_{i2} \ \dots \ H_{is}] b \end{aligned}$$

leading to the sum of squared residuals of the i -th partition:

$$r_i^T r_i = b^T H_i^T H_i b. \quad (17)$$

Because this is a scalar, we can take the “trace” of the right-hand side, leading to:

$$\begin{aligned} r_i^T r_i &= \text{trace}(b^T H_i^T H_i b) \\ &= \text{trace}(b b^T H_i^T H_i). \end{aligned}$$

Applying expectation operation, $E[\cdot]$, we obtain:

$$E[r_i^T r_i] = \text{trace}(E[b b^T] H_i^T H_i) \quad (18)$$

$$= \text{trace}(\Sigma_b H_i^T H_i) \quad (19)$$

$$= \sum_{j=1}^s \text{trace}(H_{ji}H_{ij})k_j \quad (20)$$

$$= \sum_{j=1}^s T_{ij}k_j. \quad (21)$$

Appendix shows the calculations to obtain Eq. (20) from Eq. (19). Equation (21) is obtained by defining $T_{ij} = \text{trace}(H_{ji}H_{ij})$.

Equation (19) shows how the expected sum of squared residuals $E[r_i^T r_i]$ is related to the covariance Σ_b of the standardized measurements and therefore can be used to find its unbiased estimator. Similarly, Eq. (21) can be used to find unbiased estimates of the scaling factors $k_j, j = 1, 2, \dots, s$ (diagonals of Σ_b before convergence) needed to drive the unstandardized variances to their true values via Eq. (15). Following the method of moments approach, the unbiased estimators of the scaling factors for all partitions $i = 1, 2, \dots, s$ can be solved by writing out a matrix equation and setting equal the observed values $r_i^T r_i$ to their expected values, i.e., the right hand side of (21):

$$\begin{bmatrix} r_1^T r_1 \\ r_2^T r_2 \\ \vdots \\ r_s^T r_s \end{bmatrix} = \begin{bmatrix} T_{11} & T_{12} & \cdots & T_{1s} \\ T_{21} & T_{22} & \cdots & T_{2s} \\ \vdots & & \ddots & \vdots \\ T_{s1} & T_{s2} & \cdots & T_{ss} \end{bmatrix} \begin{bmatrix} k_1 \\ k_2 \\ \vdots \\ k_s \end{bmatrix}. \quad (22)$$

The solution k_1, \dots, k_s of the system (22) is a set of unbiased estimates for the scaling factors in each iteration. The unbiased variance estimators for the unstandardized partitions $\sigma_1^2, \dots, \sigma_s^2$ are obtained by using Eq. (15), given the scaling factors and the state estimates ("maximization" step). The states are then estimated by solving (7), given the variance estimates ("expectation" step), and the iterations continue until convergence. The following two subsections discuss the state estimation ("expectation") procedure and the IRLS procedure that implements the expectation and maximization steps.

4.2 Robust state estimation

To make the state estimates robust to outliers that arise in non-line-of sight conditions, we present an approach to incorporate the proposed factor graph unbiased variance estimation method with M-estimators. M-estimators [25] are a class of popular robust estimators in linear regression that make use of a weight matrix W that assigns a weight of 1 to inlying observations (which are consistent with the assumed distribution) and smaller weights to outlying observations. We implement the M-estimator by incorporating a weight matrix W within the coefficient matrix A and the measurement vector b of the system of equations (8) as follows:

$$A^w \triangleq W^{1/2}A, \text{ and } b^w \triangleq W^{1/2}b \quad (23)$$

in which the weight matrix is defined as

$$W = \text{diag}(\tau_1, \dots, \tau_m) \quad (24)$$

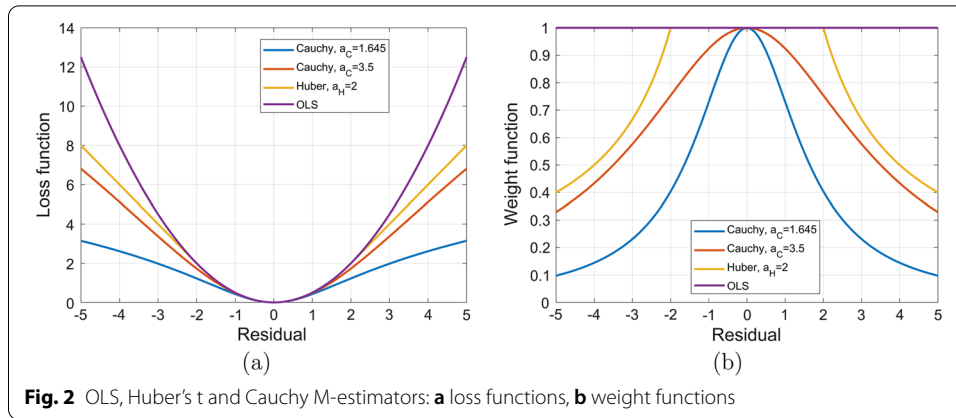


Table 1 Huber, Cauchy and OLS loss and weight functions

Method	Huber's t	Cauchy	OLS
Weight	$\tau_i = \begin{cases} 1.0 & \text{if } z_i \leq a_H \\ a_H/ z_i & \text{if } z_i > a_H \end{cases}$	$\tau_i = 1/(1 + (z_i/a_C)^2)$	$\tau_i = 1$
Loss fnc	$\rho_i = \begin{cases} 1/2z_i^2 & \text{if } z_i \leq a_H \\ z_i a_H - 1/2a_H^2 & \text{if } z_i > a_H \end{cases}$	$(a_C^2/2) \log(1 + (z_i/a_C)^2)$	$1/2z_i^2$
Parameter	a_H	a_C	NA

where τ_i is the weight found as a function of the i th scaled measurement defined as $z_i = b_i/\gamma$ and using the definitions given in Table 1, where γ is a nonparametric estimate of the standard deviation, defined as:

$$\gamma = \text{median}_{1 \leq i \leq n} |b_i - \text{median}_{1 \leq i \leq n} (b_i)| / 0.6745 \quad (25)$$

where the constant 0.6745 is chosen so that γ is an unbiased estimator based on Gaussian error distribution [25, pp. 373–374]. The proposed variance estimators are unbiased within the robust M-estimation framework. The Appendix proves this result, by extending the proof shown in Sect. 4.1 for the least squares estimation problem.

While there is a large number of weight functions for M-estimators available, we focus here on two of the most commonly utilized choices, Huber and Cauchy functions, to illustrate the main features of the weight and loss functions. We refer the reader to [28], for an extensive review of other M-estimators in regression. Huber and Cauchy weight and loss functions are plotted in Fig. 2 and defined in Table 1, along with the ordinary least squares (OLS) estimator (3), which is a quadratic function of the standardized residuals and does not provide any robustness. For a given loss function, the tuning parameter (a_C in Cauchy, a_H in Huber), to be specified by the user, determines the strength of downweighting applied to outliers. For the factor graph formulation, the weights are only applied to measurement residuals and weights for process residuals are unity, that is: τ_i for $i = nd_x + 1, \dots, m$ is found from Table 1 and $\tau_i = 1$ for $i = 1, \dots, nd_x$.

From Fig. 2b, the Huber weight function decreases for residuals larger than a_H , while Cauchy is smooth. Barron [29] proposed a family of robust loss functions to generalize

several popular robust kernels such as Huber, Cauchy, Geman–McClure and Welsh. The choice of the tuning parameter is typically made based on the measurement noise for the inlier observations or can be treated as an additional unknown parameter determined by minimizing a generalized loss function [2, 30]. In this paper, we illustrated the use of proposed variance estimators following the first approach of manually determining the tuning parameter based on the measurement residuals of the inlier data. However, the optimal tuning approaches [2, 30] can easily be adopted with the proposed estimators. The manual tuning is accomplished by overlaying the weight function (Fig. 2b) with the histogram of the residuals and adjusting the tuning parameter until the residuals greater than about 3 standard deviations are weighed down almost completely (i.e., τ_i is almost 0). For the Cauchy weight function, this corresponds to setting a_C equal to the standard deviation of the inlying data residuals, such that residuals larger than about $3a_C$ receive a very small weight.

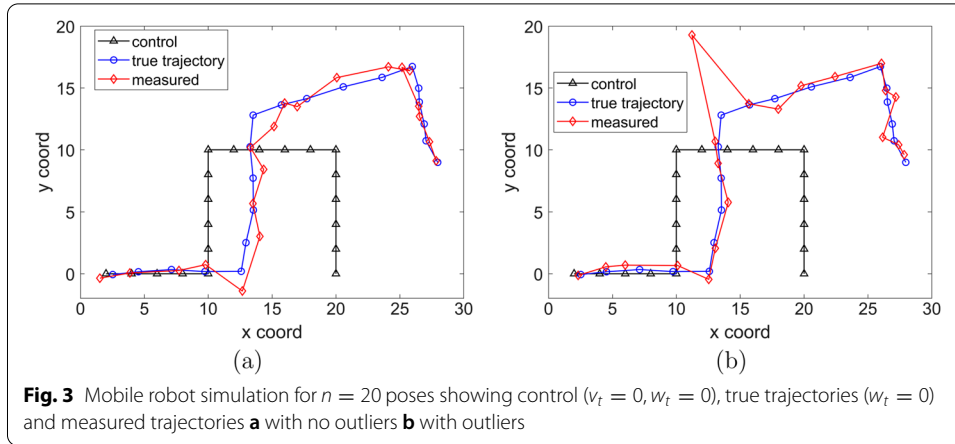
4.3 Iteratively reweighted least squares

In implementing the robust estimators in regression [31], an iteratively reweighted least squares (IRLS) algorithm is often used to solve iteratively, for the best state estimate for a given set of noise variance values and the best variance values for a given state solution until convergence. These steps are analogous to the expectation and maximization steps of the methods in [19–22] where the residuals are used to compute a covariance (the M-step), and the estimated covariances are used to compute new residuals (the E-step). The proposed IRLS algorithm to implement the proposed unbiased variance estimators of the factor graph formulation and the M-estimators for robust state estimation following this iterative framework is described in Algorithm 1.

Algorithm 1: IRLS algorithm for unbiased estimation of noise variances and robust estimation of states

Result: $X, \sigma_1^2, \sigma_2^2, \dots, \sigma_s^2$ and Q, R, Σ_0
Input: Z, U, μ_0, a_C (or a_H)
Initialization: $X, \sigma_1^2, \sigma_2^2, \dots, \sigma_s^2$
repeat
 Least Squares (E) step:
 repeat
 Based on $\sigma_1^2, \sigma_2^2, \dots, \sigma_s^2$ compute W and $A^w(X, W, \sigma_i^2)$ and $b^w(X, W, \sigma_i^2)$;
 Solve Equation (8) for ΔX ;
 $X \leftarrow X + \Delta X$
 until X converges;
 Reweighting (M) step:
 Use last $A^w(X, W, \sigma_i^2)$ to compute H ;
 Solve Equation (22) for multiplicative scale factors $\{k_1, k_2, \dots, k_s\}$;
 for $i = 1 \dots s$ **do**
 Update variance estimates using $\sigma_i^2 \leftarrow k_i \sigma_i^2$ (Equation (15))
 end
 Construct Q, R, Σ_0 based on $\sigma_2^2, \dots, \sigma_s^2$
until $X, \sigma_1^2, \sigma_2^2, \dots, \sigma_s^2$ converge;

Note that, the measurement vector b^w and matrix A^w used in the algorithm already incorporates the M-estimator weight matrix W , as shown in Eq. (23). The iterative approach solves for the best state estimate in the factor graph framework using the current unbiased variance estimates $\sigma_1^2, \sigma_2^2, \dots, \sigma_s^2$ and the scaling factors k_1, k_2, \dots, k_s of the partitions. This state estimate is then used to scale the variance estimates for the next iteration and the process is repeated until convergence.



5 Experiments

This section illustrates the efficacy of the proposed method using a simulation experiment involving a mobile robot and a case study experiment based on real GPS data.

5.1 Simulation experiment

The mobile robot simulation was based on a two-dimensional constant velocity model as shown in Fig. 3. Figure 3a, b shows, respectively, the simulations without and with outliers introduced due to non-line-of-sight conditions.

Two different motion models are considered. (i) A linear model where the state vector $x_t = (x_t, \dot{x}_t, y_t, \dot{y}_t)^T$ gives the position and velocity of the robot, the measurement vector $z_t = (x_t, y_t)$ gives the position of the robot in two dimensions and the control vector $u_t = (u_{1t}, u_{2t})^T$ affects the velocities \dot{x}_t and \dot{y}_t . (ii) A nonlinear model, with state vector $x_t = (x_t, y_t, \theta_t, \lambda_t)^T$ where the θ_t is rotational velocity and λ_t is forward velocity; measurement vector $z_t = (x_t, y_t)$ is the position of the robot, and control vector $u_t = (u_{1t}, u_{2t})^T$ affects the velocities θ_t and λ_t . Figure 3 shows simulations with $n = 20$ timesteps. The performance of robust estimators is studied by considering scenarios involving outliers for the linear robot motion model.

Table 2 shows the definitions of the terms of the dynamics and measurement equations (1) and (2) of the linear and nonlinear motion models. The sampling period was taken as $T = 1$ and the initial pose was $x_0 = (0, 2, 0, 0)^T$. According to these models, variances correspond to process dynamics and measurements as follows. For linear motion σ_{Q1}^2 is for the dynamics of (x_t, \dot{x}_t) ; σ_{Q2}^2 is for the dynamics of (y_t, \dot{y}_t) and σ_R^2 is for the measurements of (x_t, y_t) . For nonlinear motion σ_{Q1}^2 is for the dynamics of (x_t, y_t) ; σ_{Q2}^2 is for the dynamics of (θ_t, λ_t) and σ_R^2 is for the measurements of (x_t, y_t) .

The proposed variance estimation method was applied by performing the partitioning shown in Eq. (13) on the factor graph formulation with $s = 3$ partitions. The partitioned matrices A_1, A_2, A_3 are formed such that they correspond to the rows of A containing $\sigma_{Q1}^2, \sigma_{Q2}^2$ and σ_R^2 , respectively, and the partitions b_1, b_2, b_3 are similarly obtained from the rows of b . See Table 2 for the definitions of Q and R and Eq. (9) for how A and b are defined for an example with $n = 4$. In the simulations, the initial pose x_0 is assumed to be known and therefore, it treated as a constant in the factorization (7) and x_0 and Σ_0 are not estimated.

Table 2 Model terms of the mobile robot examples

Linear motion	Nonlinear motion
$x_t = f(x_{t-1}, u_t) + v_t$ $= Fx_{t-1} + Bu_t + v_t$	$x_t = f(x_{t-1}, u_t) + v_t$ $= x_{t-1} + \begin{bmatrix} \lambda_{t-1} T \cos \theta_{t-1} \\ \lambda_{t-1} T \sin \theta_{t-1} \\ 0 \\ 0 \end{bmatrix} + Bu_t + v_t$
$v_t \sim N(0, Q)$ $x_t = (x_t, \dot{x}_t, y_t, \dot{y}_t)$ $z_t = g(x_t) + w_t$ $w_t \sim N(0, R)$ $F = \partial f / \partial x_{t-1}$ $= \begin{bmatrix} 1 & T & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & T \\ 0 & 0 & 0 & 1 \end{bmatrix}$ $B = \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 0 \\ 0 & 1 \end{bmatrix}, G = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$ $Q = \begin{bmatrix} \sigma_{Q1}^2 & 0 & 0 & 0 \\ 0 & \sigma_{Q1}^2 & 0 & 0 \\ 0 & 0 & \sigma_{Q2}^2 & 0 \\ 0 & 0 & 0 & \sigma_{Q2}^2 \end{bmatrix}, R = \sigma_R^2 I_2$	$v_t \sim N(0, Q)$ $x_t = (x_t, y_t, \theta_t, \dot{\lambda}_t)$ $z_t = g(x_t) + w_t$ $w_t \sim N(0, R)$ $F_{t-1} = \partial f / \partial x_{t-1}$ $= \begin{bmatrix} 1 & 0 & -\lambda_{t-1} T \sin \theta_{t-1} & T \cos \theta_{t-1} \\ 0 & 1 & \lambda_{t-1} T \cos \theta_{t-1} & T \sin \theta_{t-1} \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$ $B = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}, G = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}$ $Q = \begin{bmatrix} \sigma_{Q1}^2 & 0 & 0 & 0 \\ 0 & \sigma_{Q1}^2 & 0 & 0 \\ 0 & 0 & \sigma_{Q2}^2 & 0 \\ 0 & 0 & 0 & \sigma_{Q2}^2 \end{bmatrix}, R = \sigma_R^2 I_2$

The control input for the linear motion case was defined as:

$$u_{1t} = \begin{cases} -2 & t = 5, 15, 25, \dots \\ +2 & t = 10, 20, 30, \dots \\ 0 & \text{otherwise} \end{cases} \quad (26)$$

$$u_{2t} = \begin{cases} +2 & t = 5, 20, 25, 40, 45, \dots \\ -2 & t = 10, 15, 30, 35, 50, 55, \dots \\ 0 & \text{otherwise} \end{cases} \quad (27)$$

which has the effect of (without process and measurement noises, i.e., $v_t = 0, w_t = 0$) making the robot move forward 10 units, then up 10 units, forward 10 units, then back down 10 units and as this pattern repeats, a “square wave” would be created. Figure 3a shows an example of running this system for 20 timesteps. The control input for the nonlinear motion case was defined as:

$$u_{1t} = \begin{cases} \pi/2 & t = 5, 15, 25, \dots \\ -\pi/2 & t = 10, 20, 30, \dots \\ 0 & \text{otherwise} \end{cases} \quad (28)$$

$$u_{2t} = 0 \text{ for all } t. \quad (29)$$

Similar to the linear motion model, this control input will make the robot to move forward 10 units, then up 10 units, forward 10 units, then back down 10 units. The control trajectory of the nonlinear motion is identical to that of linear motion shown in Fig. 3a.



The motion and measurement simulations were generated by using $\sigma_{Q1}^2 = 0.5$, $\sigma_{Q2}^2 = 0.2$ and $\sigma_R^2 = 1.5$ as the true variances of the models.

In order to realistically represent the non-line-of-sight conditions, we used a two component mixture process for the measurements of the linear motion model in Table 2

$$w_t \sim (1 - \alpha)N(0, \sigma_R^2 I_2) + \alpha N(0, 10^2) \quad (30)$$

where α is the mixture proportion and the second mixture component with variance 10^2 represents outliers with large variance (compared to the variance $\sigma_R^2 = 1.5^2$ of the inlying measurements). The mixture model injects unusually large or small deviations in randomly selected measurements, mimicking the effect of reflections of satellite signal from buildings. Figure 3b shows one simulation with the mixture parameter $\alpha = 0.1$ and $n = 20$ poses that resulted in only one outlier.

5.2 Case study experiment: GPS-based localization

In this section, we illustrate the application of the proposed method in a GPS-based localization under non-line-of-sight conditions using the real data set presented in Pfeifer and Protzel [19]. The data set was collected in the city center of Chemnitz, Germany, by driving an instrumented vehicle several times over a road network, as shown in Fig. 4. The urban setting and the road layout consist of several tall buildings that cause a large number of outliers due to satellite blockages or reflections from the buildings. The data set contains high-precision inertial measurements of the vehicle position with a precision of 2 cm used as ground truth and GPS pseudorange measurements that will be used in the estimation procedure. The data set contained

pseudorange measurements at 8570 timesteps. At each timestep pseudorange measurements are gathered from 7 to 12 satellites that are visible to the vehicle.

The proposed unbiased variance estimation method with M-estimators was applied to find the location of the vehicle while making the solution robust to outliers. Our results are compared to those of [1, 19, 20] who used the same data set.

To develop the factor graph model we define the state vector as $x_t = (x_t, y_t, z_t, b_t, d_t)$ where b_t is the satellite clock error and $d_t = \dot{b}_t$ is clock error drift. The process dynamics model, incorporating only the satellite clock error and drift, is $(b_t, d_t)^T = f(b_{t-1}, d_{t-1}) + v_t$ where

$$f(b_{t-1}, d_{t-1}) = \begin{bmatrix} b_{t-1} + Td_{t-1} \\ d_t \end{bmatrix} \quad (31)$$

where the sampling period in the data set was $T = 0.25$ s and $v_t \sim N(0, Q)$ is the process error, with covariance matrix

$$Q = \begin{bmatrix} \sigma_b^2 & 0 \\ 0 & \sigma_d^2 \end{bmatrix} \quad (32)$$

where σ_b^2 and σ_d^2 are the error variances of clock offset and clock drift. The Jacobian matrix $F = \partial f / \partial x_{t-1}$ of the process equation is

$$F = \begin{bmatrix} 0 & 0 & 0 & 1 & T \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

The measurement equation consists of the pseudorange from the vehicle to the i th satellite at time t

$$z_i(x_t) = g_i(x_t) + w_{it}, \quad i = 1, 2, \dots, ns_t \quad (33)$$

where ns_t is the number of satellites visible at time t , w_{it} is the measurement error of the i th satellite. The measurement errors of all satellites are Gaussian distributed $w_t \sim N(0, R)$ with covariance matrix is $R = \sigma_R^2 I_{ns_t}$. The noiseless measurement equation is (omitting the time index t in the state variables and satellite coordinates):

$$g_i(x) = \sqrt{(S_{xi} - x)^2 + (S_{yi} - y)^2 + (S_{zi} - z)^2} + c \times b + (\gamma/c)(S_{xi}y - S_{yi}x)$$

where (S_{xi}, S_{yi}, S_{zi}) are the coordinates of the i th satellite, $c = 3 \times 10^8$ m/s is the speed of light, and $\gamma = 7.3 \times 10^{-5}$ rad/s is the earth's rotation speed. The measurement equation for the complete set of satellites is

$$z_t = g(x_t) + w_t \quad (34)$$

where $g = (g_1, \dots, g_{ns_t})^T$. The Jacobian matrix $G_t = \partial g / \partial x_t$ of the measurement equations is

$$G_t = \begin{bmatrix} -(S_{x1,t} - x_t)/\tilde{g}_1 - (\gamma/c)S_{y1,t} & -(S_{y1,t} - y_t)/\tilde{g}_1 + (\gamma/c)S_{x1,t} & -(S_{z1,t} - z_t)/\tilde{g}_1 & c & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ -(S_{xns_t,t} - x_t)/\tilde{g}_{ns_t} - (\gamma/c)S_{y_{ns_t},t} & -(S_{y_{ns_t},t} - y_t)/\tilde{g}_{ns_t} + (\gamma/c)S_{x_{ns_t},t} & -(S_{z_{ns_t},t} - z_t)/\tilde{g}_{ns_t} & c & 0 \end{bmatrix}$$

where $(S_{xi,t}, S_{yi,t}, S_{zi,t})$ are the coordinates of the i th satellite at time t and $\tilde{g}_i = \sqrt{(S_{xi,t} - x_t)^2 + (S_{yi,t} - y_t)^2 + (S_{zi,t} - z_t)^2}$.

6 Results and discussion

6.1 Mobile robot experiments

Using the example mobile robot systems, we evaluate the performance of the factor graph noise variance and localization estimates. 1000 Monte Carlo replications of the robot simulation were generated for $n = 20$ timesteps and in each replication the location sequence of the robot and the noise variances are estimated. The proposed unbiased noise variance estimator [specified by Eqs. (22) and (15)] is compared to the existing ML estimator [specified by Eq. (10)]. Both estimators are implemented within the IRLS scheme given in Algorithm 1.

We compared the performance of the estimators based on (i) bias in the variance estimates and (ii) Mahalanobis distance of the location solution. Bias of the estimates (of all 3 variances) measures the difference between the computed variances and the true variances. Since we have a variance estimate in each simulation, the average bias from N simulations is calculated as

$$C = 1/(3N) \sum_{i=1}^3 \sum_{j=1}^N (\sigma_{ij}^2 - \sigma_{Ti}^2)^2 \quad (35)$$

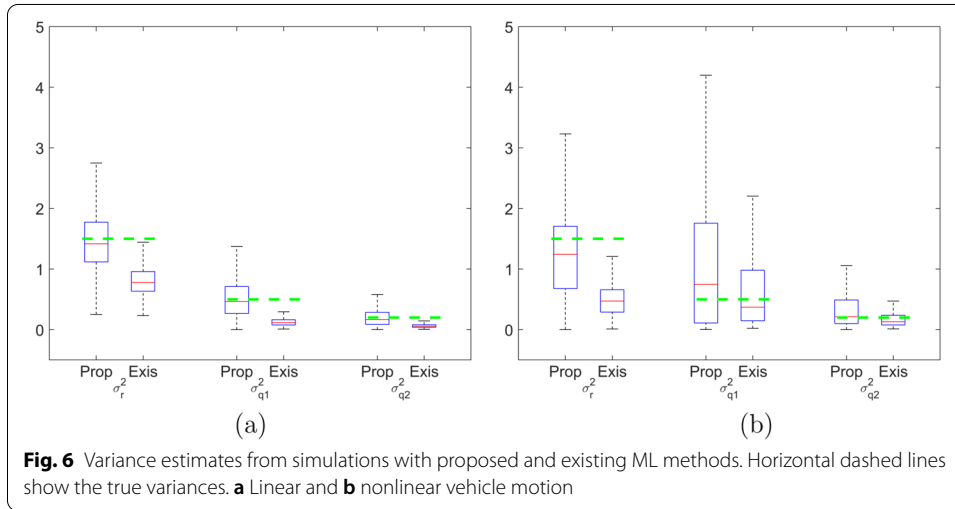
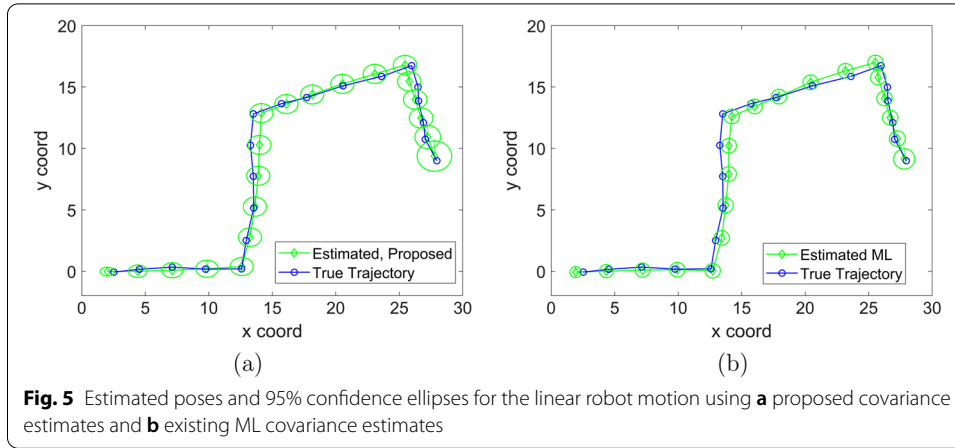
where σ_{Ti}^2 is the true value of i th variance, σ_{ij}^2 is its estimated value in j th simulation, $i = 1, 2, 3$ is the index of the variance component, $j = 1, \dots, N$ is the index of the simulation run and $N = 1000$.

Total squared Mahalanobis distance of the estimated locations from the truth over all timesteps is used to quantify the localization error. The total Mahalanobis localization error is found (in each simulation) as the sum of the errors over n poses as

$$G = \sum_{t=1}^n (x_t - \hat{x}_t)^T P_t^{-1} (x_t - \hat{x}_t) \quad , t = 1, 2, \dots, n \quad (36)$$

in which x_t is the state and \hat{x}_t is the state estimate of t th pose, $P_t \in \mathbb{R}^{d_x \times d_x}$ is the covariance matrix of the state estimate \hat{x}_t that can be obtained from the corresponding entries of the matrix $(A^T A)^{-1}$, which is an $m \times m$ matrix (see Sect. 3 for definition of m and d_x). The quantity (36) is also known as average normalized (estimation) error squared, or ANEES, in the tracking literature [32]. If the predicted uncertainty on the state estimate is correct, then G should follow a Chi-squared distribution with degree of freedom nd_x , with an expected value nd_x [23, 33].

The estimated poses and the 95% confidence ellipses using the two approaches superimposed with the true poses from a single simulation of the linear motion problem is shown in Fig. 5. The location estimates from both methods are very similar. However, the confidence ellipses, which are the direct result of the covariance estimates, from the methods result in two very different conclusions about the true (unknown) location of the robot. The localization with the proposed unbiased variance estimates has the true location within its confidence interval at every timestep, while the ML method variance estimates frequently has the true location either outside or very close to the confidence



ellipse. As the robot moves, the uncertainty in position estimates grows due to the increasing variance estimates, as indicated by uncertainty ellipses growing in diameter over time. Further, the variance estimates from the proposed method (Fig. 5a) are larger than those from ML method (Fig. 5b) as evidenced by the larger ellipses of the former, indicating larger uncertainty about the true location. This demonstrates some of the practical concerns of underestimation with the existing ML-based covariance estimation techniques.

The variance estimates σ_R^2 , σ_{Q1}^2 and σ_{Q2}^2 obtained across the Monte Carlo simulations (of robot motion with $n = 20$ steps) are shown in Fig. 6 (for linear and nonlinear models) and the distributions of the Mahalanobis distances (for linear motion model) are shown in Fig. 7. The cases of $n = 20$ and $n = 100$ timesteps are considered to study the influence of larger data sets on the quality of the noise variance estimates. The variance estimates obtained with $n = 20$ and 100 are shown in Fig. 8. The summaries of the distributions are given in Table 3 which reports the average variance estimate, the bias of the variance estimates C and the average Mahalanobis localization error G , defined in Eqs. (35) and (36), respectively, computed in the simulations of both linear and nonlinear models with $n = 20$.

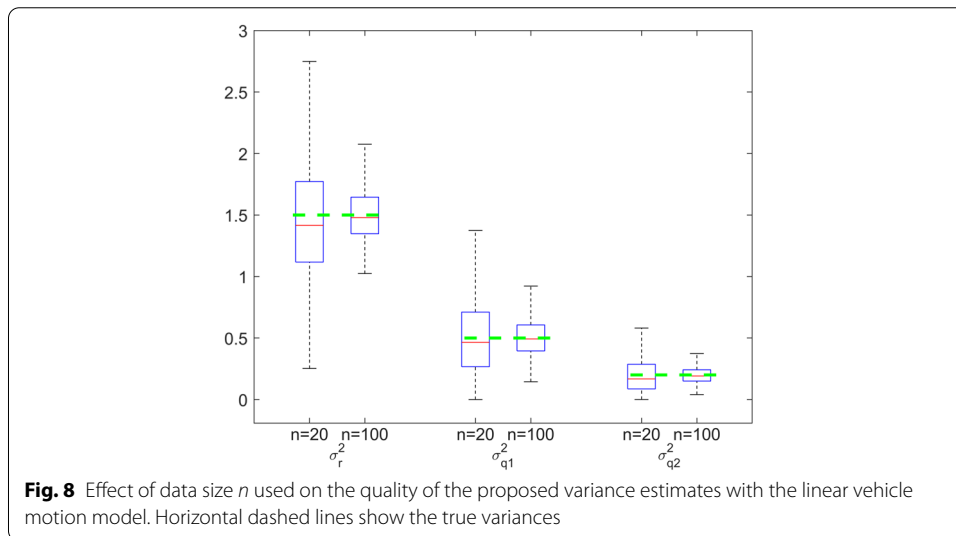
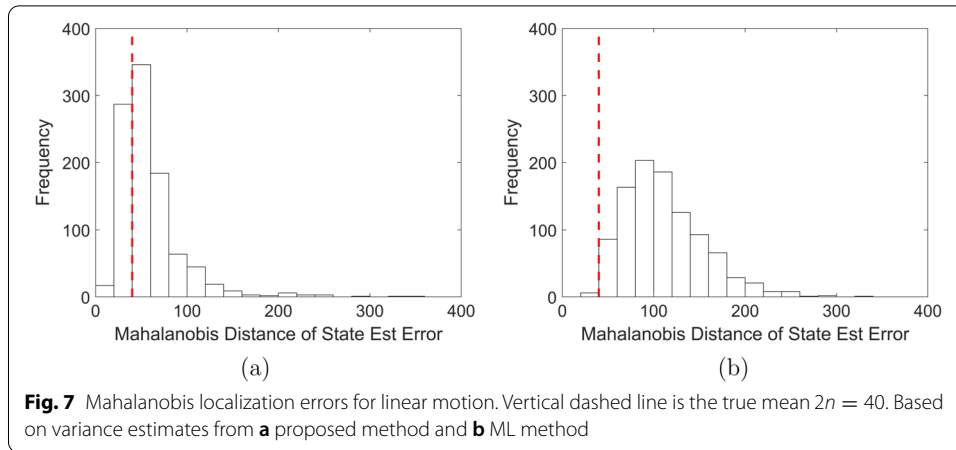


Table 3 Average variance estimates and localization errors, from simulations with $n = 20$

Motion model	Estimate	σ_R^2	σ_{Q1}^2	σ_{Q2}^2	C	G
Linear	Existing	0.81	0.13	0.06	0.628	103.93
	Proposed	1.46	0.52	0.17	0.003	51.13
	Truth	1.50	0.50	0.20	NA	NA
Nonlinear	Existing	0.49	0.77	0.18	1.097	95.81
	Proposed	1.20	1.22	0.38	0.638	59.57
	Truth	1.50	0.50	0.20	NA	NA

For all variances, the (existing) ML estimates are significantly smaller than the true value (underestimate), while the proposed variance estimates are centered around the true value (unbiased). See Fig. 6. The smaller ML variance estimates result in Mahalanobis errors being too large (Fig. 7). Further, the proposed variance estimates have improved precision with less variability while remaining unbiased with larger

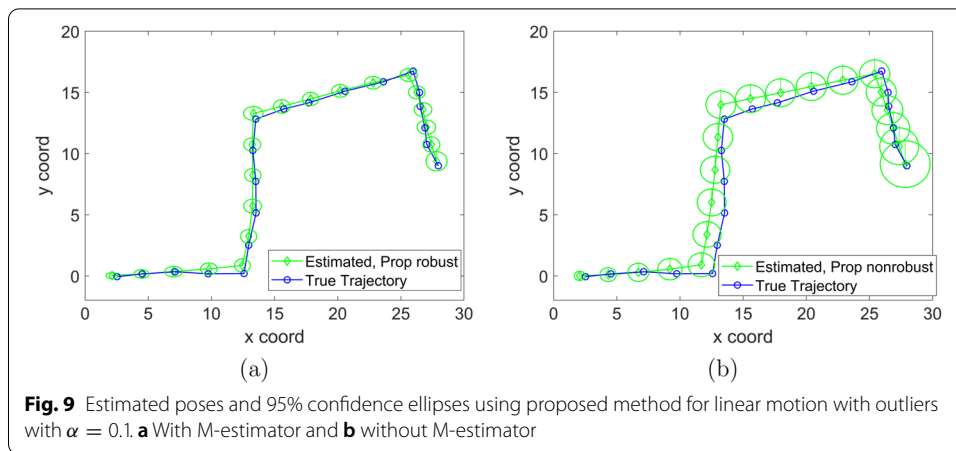


Table 4 Variance estimates and localization errors, linear model with outliers, $n = 20$

Method	α	σ_R^2	σ_{Q1}^2	σ_{Q2}^2	C	G
Proposed method	0.10	1.69	0.39	0.15	0.05	55.87
with M-estimator	0.25	3.57	0.39	0.15	4.31	47.98
	Truth	1.50	0.50	0.20	NA	NA
Proposed method	0.10	11.19	1.04	0.59	94.28	56.00
without M-estimator	0.25	25.37	0.96	0.73	570.28	59.37
	Truth	1.50	0.50	0.20	NA	NA

n (Fig. 8). For both motion models, the existing ML estimates have far more bias and the resulting average Mahalanobis error is significantly larger than it should be, which is $nd_x = 20 \times 2 = 40$ (Table 3). In comparison, the proposed variance estimates agree much more closely with true values (significantly smaller bias) resulting in about 50% smaller Mahalanobis localization errors.

Performance under non-line-of-sight conditions: Monte Carlo simulations of the linear robot motion model with the mixture measurement noise distribution equation (30) and $n = 20$ poses were conducted to study the performance of the proposed variance estimator with robust M-estimators. Outlier mixture proportions $\alpha = 0.10$ and 0.25 were considered, in which the latter case contains more outliers than the former. The proposed method was implemented with a Cauchy M-estimator with parameter $a_C = 1.645$. Figure 9 shows the estimates of robot location for one realization with $\alpha = 0.10$, with and without an M estimator. It can be seen that the localization estimates are similar with both methods; however, due to the presence of outliers, the confidence ellipses are significantly different depending on whether or not an M-estimator is utilized. The confidence ellipses of the position solution with the M-estimator (Fig. 9a), by virtue of downweighting the unusual observations, are smaller than those of the non-robust method (Fig. 9b) which are much larger (more uncertain).

Table 4 summarizes the means and biases of the variance estimates and the Mahalanobis distances from 1000 Monte Carlo simulations. The benefit of using an M-estimator in conjunction with the unbiased variance estimators on the quality of the estimates is evident, in particular when the outliers are more frequent (i.e., $\alpha = 0.25$ versus 0.10).

With the presence of more outliers, the variance estimates deviate more drastically from the true values if M-estimator is not employed (bias increases from 94.3 to 570.3 due to outliers). By contrast, when an M-estimator is used, in conjunction with the unbiased variance estimators, the bias in the estimates is much smaller (bias increases only from 0.05 to 4.31 due to outliers). The average Mahalanobis localization error is not drastically different from the expected value of $2n = 40$ due to the outliers, regardless of whether or not an M-estimator is used, since unbiased variance estimators are used in both results. However, when M-estimator is used, the localization error is smaller, due to the down-weighting of the unusual observations.

6.2 GPS experiments

The proposed unbiased variance estimation and the existing ML variance estimation were applied to the GPS data set of the vehicle driving experiment shown in Fig. 4. The methods were used to estimate the variances σ_b^2 , σ_d^2 and σ_R^2 of the clock error offset [seconds²], drift (unitless) and measurement error [meters²], respectively, and the vehicle trajectory by incorporating M-estimators following the procedure described in Sect. 4. The Cauchy M-estimator with parameter $a_C = 3.5$ was used to achieve robustness against outliers with both methods.

We compare the Mahalanobis localization errors of the proposed method to those reported by the existing approaches [1, 19, 20] that used the same data set. Reference [1] considered several combinations of the multiple factors of (i) pseudo range, (ii) prior, (iii) state transition and (iv) dependence between successive measurements. In our comparisons, we considered the results reported for the combinations (i & ii) and (i, ii, iii, & iv), which we refer to as SC method 1 and 2, respectively. Reference [19] used a Gaussian mixture model (GMM). Reference [20] improves the GMM by using an expectation–maximization (EM) algorithm. Both [19, 20] use a max-mixture approach as an efficient approximation of the Gaussian sum-mixture, as discussed in [3]. Self-tuning GMM approach, by contrast to the static GMM approach, includes the weights, means and variances of mixtures as additional variables in the optimization problem.

Figure 10 shows the localization results obtained with M-estimation and the proposed variance estimates and the ML variance estimates and Table 5 summarizes the variance estimates. It is clear that the localization with the proposed (unbiased) variance estimates is much closer to the ground truth than the ML (biased) estimates. From Table 5, it can be seen that, similar to the simulation study, the ML variance estimates are much smaller than the proposed variance estimates, which indicates that it is likely that the ML estimates are underestimating the actual variances.

Table 6 summarizes the Mahalanobis localization errors from the truth (the mean, median and max over all poses) computed using the proposed unbiased estimators, ML variance estimators and the existing approaches [1, 19, 20]. It can be seen that the proposed variance estimators, in conjunction with an M-estimator, achieve smaller median localization error than the existing SC method 1 and static GMM method, while the median error is slightly higher than those of SC method 2 and self-tuning GMM counterparts. The median and maximum errors are not reported in [20] and could not be included in Table 6. While the existing approaches with some more complex configurations have better performance in localization error, a significant benefit of the proposed

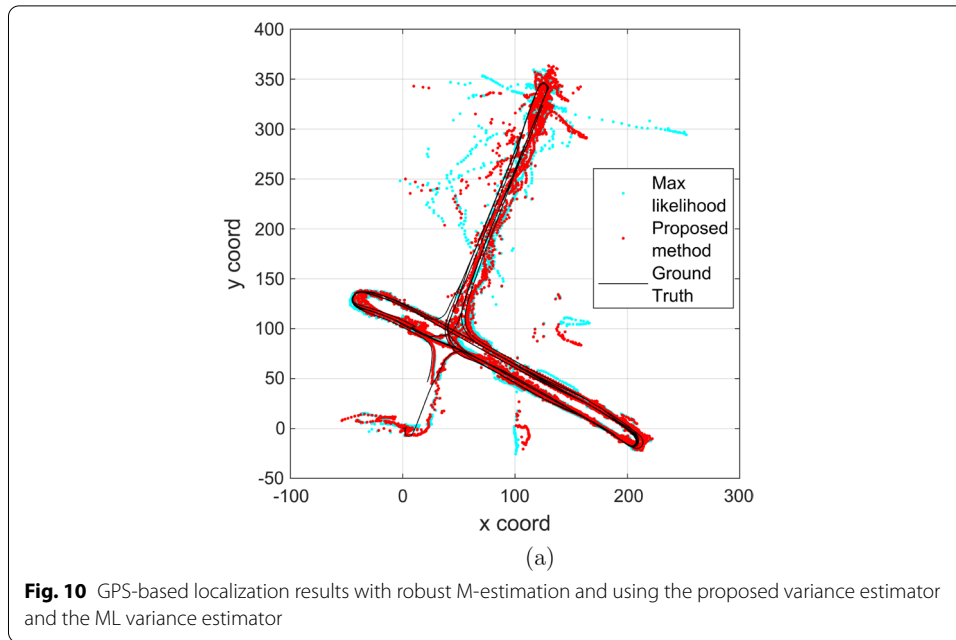


Table 5 GPS data variance estimates using the proposed method

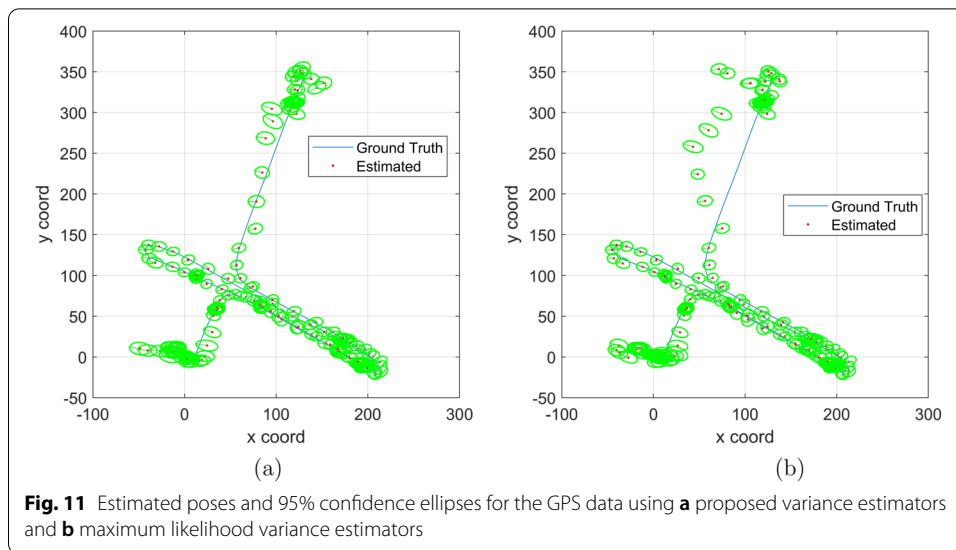
Method	$\sigma_a^2 [\text{m}^2]$	$\sigma_b^2 [\text{s}^2]$	σ_d^2
Proposed method	19.24	5.498	0.704
ML method	18.35	4.84E-9	9.35e-04

Table 6 Localization errors (m) of the proposed method and the existing SC [1], GMM [19] and EM-GMM [20] approaches using the GPS data

Method	Mean (m)	Median (m)	Max (m)
Proposed method	6.06	3.48	121.08
ML method	7.06	4.05	131.13
SC method 1 [1]	17.91	3.66	171.61
SC method 2 [1]	2.96	2.45	16.31
Static GMM [19]	4.52	3.57	31.75
Self-tuning GMM [19]	2.45	1.99	13.91
Static EM-GMM [20]	4.13	NA	NA
Self-tuning EM-GMM [20]	2.56	NA	NA

method is that of providing unbiased estimates of the prior and noise covariances. As discussed previously, and in literature [8], having good variance estimates is crucial for localization applications such as matching observations with landmarks and loop closure detection.

To get a better understanding of the quality and reasonableness of the variance estimates, the uncertainty of the localization results is visualized by plotting the 95% confidence ellipses of the true vehicle positions obtained with both the proposed and ML variance estimates in Fig. 11. In order to maintain clarity of the display, we showed only the initial 2870



poses out of the entire 8570 poses and plot the confidence ellipse of 1 pose for every 10 poses. The size and shape of the ellipse for each true position is proportional to the variance estimate and quantifies the uncertainty of the localization estimate. The estimated locations and their uncertainty ellipses with the proposed variance estimates more closely follow the ground truth than with the ML variance estimates. To summarize the performance of both methods, we compare the number of times the ground truth positions fall within the confidence ellipses for the entire data set. Using the proposed variance estimates, the ground truth is captured correctly for 7283 out of the 8570 of the vehicle poses (85.0% of the time), while using the ML variance estimates, the ground truth is captured correctly for the 6,832 poses (79.7% of the time). This demonstrates that the proposed variance estimates achieves a 6.6% improvement in quantifying the uncertainty for GPS-based localization over the ML variance estimates.

7 Conclusions

This paper proposed a new methodology to estimate multiple noise variances in a factor-graph based formulation of the GPS-based localization problem under non-line-of-sight conditions. Unbiased variance estimators of the factor graph formulation are developed following the method of moments approach in linear regression theory. An iteratively reweighted least squares (IRLS) approach was presented for jointly estimating the system state and the noise variances. The existing smoothing-based state estimation approaches have largely relied on the maximum likelihood (or residual sample variance) estimators of variances in an expectation–maximization framework. While the ML estimators are easy to compute and have good performance in many applications, they can result in significant underestimation of the true variances which may in turn impact the localization accuracy.

A simulated mobile robot motion analysis showed that using the proposed variance estimators, the localization errors can be significantly improved over the ML estimators, considering model nonlinearities and measurement outliers. In addition, a case study involving real GPS data compared the performance of the method to an existing switching

constraints method for a localization problem in an urban environment with significant multipath problems.

A major assumption in the current study was that the noise components are mutually uncorrelated and covariance matrix we estimate was diagonal. A future work of practical interest is to extend the presented approach to unbiased estimators of non-diagonal covariance matrices. Further, in our evaluations we demonstrated the benefits of proposed unbiased variance estimators over maximum likelihood variance estimators in an M-estimation framework. It is important to note that the approach is general and can be implemented with other weighting based robust estimators, such as switchable constraints and dynamic covariance scaling. As most of the existing robust estimation frameworks use ML-based variance estimators, if the unbiased variance estimators were to be implemented within these frameworks it is reasonable to expect a similar improvement in localization accuracy to that observed within the M-estimator framework.

Appendix: Unbiasedness of the Variance Estimators

In this appendix, we present the calculations needed to show unbiasedness of proposed variance estimators under both least squares (non-robust) and robust M-estimator based state estimation. First we prove how Eq. (20) follows from Eq. (19) for the least squares estimation problem.

$$E[r_i^T r_i] = \text{trace}(\Sigma_b H_i^T H_i) \quad (37)$$

$$= \text{trace} \begin{bmatrix} H_{i1}^T H_{i1} k_1 & H_{i1}^T H_{i2} k_1 & \cdots & H_{i1}^T H_{is} k_1 \\ H_{i2}^T H_{i1} k_2 & H_{i2}^T H_{i2} k_2 & \cdots & H_{i2}^T H_{is} k_2 \\ \vdots & \vdots & \ddots & \vdots \\ H_{ii}^T H_{i1} k_i & \cdots & H_{ii}^T H_{ii} k_i & \cdots & H_{ii}^T H_{is} k_i \\ \vdots & \vdots & \ddots & \vdots \\ H_{is}^T H_{i1} k_s & \cdots & \cdots & H_{is}^T H_{is} k_s \end{bmatrix} \quad (38)$$

$$= \sum_{j=1}^n \text{trace}(H_{ij}^T H_{ij}) k_j \quad (39)$$

To prove unbiasedness when robust M-estimators are used, we need to consider the weight matrix. When a weight matrix W is used, the system coefficient and observation matrices are

$$A^w = W^{1/2} A$$

$$b^w = W^{1/2} b$$

where A is an $m \times p$ matrix, W is an $m \times m$ diagonal matrix, with elements τ_i for $i = 1, 2, \dots, m$. Define $H^w = I - \tilde{H}^w$ where \tilde{H}^w is the projection matrix with the weighted system

$$\tilde{H}^w = A^w (A^{wT} A^w)^{-1} A^{wT} = W^{1/2} A (A^T W A)^{-1} A^T W^{1/2}.$$

Analogous to the way matrix H is partitioned in Eq. (16), partition H^w to find H_{ij}^w , which is an $n_i \times n_j$ matrix for $i, j = 1, 2, \dots, s$, and H_i^w , which is an $n_i \times m$ matrix for $i = 1, 2, \dots, s$. The sum of squared residuals with M-estimators for the i th partition is then written as

$$\begin{aligned} r_i^T r_i &= b^{wT} H_i^{wT} H_i^w b^w \\ &= \text{trace}(b^w b^{wT} H_i^{wT} H_i^w). \end{aligned}$$

Note that the sum of squared measurements is

$$b^w b^{wT} = W^{1/2} b b^T W^{1/2}.$$

Therefore, its expected value can be found as

$$E[b^w b^{wT}] = W^{1/2} E[b b^T] W^{1/2} = W^{1/2} \Sigma_b W^{1/2}.$$

Applying some matrix algebra, similar to Eqs. (19)–(21), we find

$$\begin{aligned} E[r_i^T r_i] &= \text{trace}(E[b^w b^{wT}] H_i^{wT} H_i^w) \\ &= \text{trace}(W^{1/2} \Sigma_b W^{1/2} H_i^{wT} H_i^w) \end{aligned} \quad (40)$$

$$\begin{aligned} &= \text{trace}(\Sigma_b W^{1/2} H_i^{wT} H_i^w W^{1/2}) \\ &= \text{trace}(\Sigma_b D_i^T D_i) \end{aligned} \quad (41)$$

$$= \sum_{j=1}^s \text{trace}(D_{ji} D_{ij}) k_j \quad (42)$$

where

$$D_i = H_i^w W^{1/2}$$

and D_{ij} is an $n_i \times n_j$ matrix obtained by partitioning the $n_i \times m$ matrix D_i as

$$D_i = [D_{i1} \ D_{i2} \ \cdots \ D_{is}].$$

Equation (40) proves that, when this system of equations, with the observed residuals replacing the expected values on the left hand side (based on the method of moments approach), is solved for $k_i, i = 1, 2, \dots, s$, one obtains unbiased estimators of the scaling factors needed for each linearizing iteration. Equation (42) follows from Eq. (41) using the steps shown in Eqs. (37)–(39) for the least squares case, but by replacing T_{ij} with D_{ij} . The system of equations for (42) is obtained, analogous to (22), but by replacing T_{ij} with D_{ij} . This completes the proof that the proposed variance estimators are unbiased within the M-estimation framework.

Abbreviations

GPS: Global Positioning System; RAIM: Receiver Autonomous Integrity Monitoring; LiDAR: Light detection and ranging; IMU: Inertial measurement unit; SLAM: Simultaneous localization and mapping; IRLS: Iteratively reweighted least squares; EM: Expectation maximization; ML: Maximum likelihood; SC: Switching constraints; CPU: Central processing unit.

Acknowledgements

The authors would like to thank the anonymous reviewers for their valuable suggestions. The first author (OAV) would like to acknowledge the support for this work from the AFOSR-SFFP program and the AFIT.

Author's contributions

OAV and CNT developed the core concepts of the methodology presented within this manuscript, OAV implemented the algorithm for the simulation and case studies, CNT provided debugging and refinements on the algorithm, OAV developed and refined the manuscript, CNT provided proofreadings, refinements and revisions of the manuscript. All authors read and approved the final manuscript.

Funding

This research was supported in part by the Air Force Research Laboratory, Air Force Institute of Technology (AFIT), through the Air Force Office of Scientific Research (AFOSR), Summer Faculty Fellowship Program (SFFP) [®], Contract Numbers FA8750-15-3-6003 and FA9550-15-0001.

Availability of data and materials

Data and computer software supporting our findings can be accessed at <https://github.com/avanli/Factor-Graph-Covar>.

Declaration

Competing interests

O. Arda Vanli and Clark N. Taylor declare that they have no conflict of interest.

Author details

¹Department of Industrial and Manufacturing Engineering, Florida A&M University, Florida State University, Tallahassee, FL, USA. ²Autonomy and Navigation Technology Center, Air Force Institute of Technology, WPAFB, OH, USA.

Received: 20 September 2021 Accepted: 11 January 2022

Published online: 25 January 2022

References

1. N. Sünderhauf, M. Obst, G. Wanielik, P. Protzel, Multipath mitigation in gnss-based localization using robust optimization, in: *IEEE Intelligent Vehicles Symposium*. IEEE 2012, pp. 784–789 (2012)
2. G. Agamennoni, P. Furgale, R. Siegwart, Self-tuning m-estimators, in: *2015 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, pp. 4628–4635 (2015)
3. E. Olson, P. Agarwal, Inference on networks of mixtures for robust robot mapping. *Int. J. Robot. Res.* **32**(7), 826–840 (2013)
4. N. Sünderhauf, P. Protzel, Switchable constraints for robust pose graph slam, in: *2012 IEEE/RSJ international conference on intelligent robots and systems*. IEEE, pp. 1879–1884 (2012)
5. P. Agarwal, G.D. Tipaldi, L. Spinello, C. Stachniss, W. Burgard, Robust map optimization using dynamic covariance scaling, in: *IEEE International Conference on Robotics and Automation*. IEEE 2013, pp. 62–69 (2013)
6. S. Hewitson, J. Wang, GNSS receiver autonomous integrity monitoring (RAIM) performance analysis. *Gps Solut.* **10**(3), 155–170 (2006)
7. A. Shetty, G. Xingxin Gao, Adaptive covariance estimation of LiDAR-based positioning errors for UAVs. *Navigation* **66**(2), 463–476 (2019)
8. M. Kaess, F. Dellaert, Covariance recovery from a square root information matrix for data association. *Robot. Auton. Syst.* **57**(12), 1198–1210 (2009)
9. J. Behley, C. Stachniss, Efficient surfel-based slam using 3d laser range data in urban environments, in: *Robotics: Science and Systems* (2018)
10. O.A. Vanli, C.N. Taylor, Covariance estimation for factor graph based Bayesian estimation, in: *2020 IEEE 23rd International Conference on Information Fusion (FUSION)*. IEEE, pp. 1–8 (2020)
11. S.J. Julier, J.K. Uhlmann, Simultaneous localisation and map building using split covariance intersection, in: *Proceedings 2001 IEEE/RSJ International Conference on Intelligent Robots and Systems*, vol. 3. IEEE, pp. 1257–1262 (2001)
12. F. Dellaert, M. Kaess, Square root SAM: simultaneous localization and mapping via square root information smoothing. *Int. J. Robot. Res.* **25**(12), 1181–1203 (2006)
13. H.E. Rauch, F. Tung, C.T. Striebel, Maximum likelihood estimates of linear dynamic systems. *AIAA J.* **3**(8), 1445–1450 (1965)
14. G.J. Bierman, *Factorization Methods for Discrete Sequential Estimation* (1977)
15. M. Kaess, A. Ranganathan, F. Dellaert, iSAM: incremental smoothing and mapping. *IEEE Trans. Robot.* **24**(6), 1365–1378 (2008)
16. Z. Chen, Bayesian filtering: from Kalman filters to particle filters, and beyond. *Statistics* **182**(1), 1–69 (2003)
17. R. Casas, A. Marco, J. Guerrero, J. Falco, Robust estimator for non-line-of-sight error mitigation in indoor localization. *EURASIP J. Adv. Signal Process.* **2006**, 1–8 (2006)
18. İ Güvenc, C.-C. Chong, F. Watanabe, H. Inamura, NLOS identification and weighted least-squares localization for UWB systems using multipath channel statistics. *EURASIP J. Adv. Signal Process.* **2008**, 1–14 (2007)
19. T. Pfeifer, P. Protzel, Robust sensor fusion with self-tuning mixture models, in: *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, pp. 3678–3685 (2018)
20. T. Pfeifer, P. Protzel, Expectation-maximization for adaptive mixture models in graph optimization, in: *2019 International Conference on Robotics and Automation (ICRA)*. IEEE, pp. 3151–3157 (2019)

21. R.M. Watson, J.N. Gross, C.N. Taylor, R.C. Leishman, Enabling robust state estimation through measurement error covariance adaptation. *IEEE Trans. Aerosp. Electron. Syst.* **56**, 2026–2040 (2019)
22. R.M. Watson, J.N. Gross, C.N. Taylor, R.C. Leishman, Robust incremental state estimation through covariance adaptation. *IEEE Robot. Autom. Lett.* **5**(2), 3737–3744 (2020)
23. Y. Bar-Shalom, X.-R. Li, *Estimation and Tracking-Principles, Techniques, and Software* (Artech House, Inc, Norwood, 1993)
24. J. Duník, O. Straka, O. Kost, J. Havlík, Noise covariance matrices in state-space models: a survey and comparison of estimation methods-Part I. *Int. J. Adapt. Control Signal Process.* **31**, 1505–1543 (2017)
25. D.C. Montgomery, E.A. Peck, G.G. Vining, *Introduction to Linear Regression Analysis*, vol. 821 (Wiley, Hoboken, 2012)
26. J. Duník, O. Kost, O. Straka, Design of measurement difference autocovariance method for estimation of process and measurement noise covariances. *Automatica* **90**, 16–24 (2018)
27. A.C. Rencher, G.B. Schaalje, *Linear Models in Statistics* (Wiley, Hoboken, 2008)
28. D. Q. De Menezes, D.M. Prata, A.R. Secchi, J.C. Pinto, A review on robust m-estimators for regression analysis. *Comput. Chem. Eng.* 107254 (2021)
29. J.T. Barron, A general and adaptive robust loss function, in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 4331–4339 (2019)
30. N. Chebrolu, T. Läbe, O. Vysotska, J. Behley, C. Stachniss, Adaptive robust kernels for non-linear least squares problems. *IEEE Robot. Autom. Lett.* **6**(2), 2240–2247 (2021)
31. C. Zach, Robust bundle adjustment revisited, in *European Conference on Computer Vision*. Springer, pp. 772–787 (2014)
32. D. Lerro, Y. Bar-Shalom, Tracking with debiased consistent converted measurements versus EKF. *IEEE Trans. Aerospace Electron. Syst.* **29**(3), 1015–1022 (1993)
33. C.N. Taylor, S. Lubold, Verifying the predicted uncertainty of Bayesian estimators, in *Geospatial Informatics, Motion Imagery, and Network Analytics VIII*, vol. 10645. International Society for Optics and Photonics, p. 106450E (2018)

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)