S-estimators in mapping applications

João Sequeira
Instituto Superior Técnico, Technical University of Lisbon
Portugal
Email: joao.silva.sequeira@ist.utl.pt
Antonios Tsourdos
Autonomous Systems Group, Department of Informatics and Sensors, Cranfield University
United Kingdom
Email: a.tsourdos@cranfield.ac.uk
Hyo-Sang Shin
Autonomous Systems Group, Department of Informatics and Sensors, Cranfield University
United Kingdom
h.shin@cranfield.ac.uk

***Abstract*—This paper discusses a class of S-estimators for covariance matrices applied to a mapping application. Two alternative solution methods are discussed, namely using an iterative formulation described in the literature, and an optimization formulation.**

Simulation results on a mapping example are presented to illustrate the discussion.

Keywords: *S-estimator*, *Mapping*.

I. Introduction

In recent years environment mapping and SLAM (Simultaneous Localization and Mapping) have been recurrent issues in robotics. Both problems deal with uncertain information on detected landmarks, namely position and eventually orientation. Uncertainty measures on the detected landmarks are thus very important.

Covariance of the estimated measures is general a preferred statistic to assess uncertainty as it includes spatial information (in the dimensions of the datasets). For example, some methods will discard points that are likely to be outliers while others do not make any discrimination among the dataset. Clearly, if the potential outliers are indeed outliers a robust method will yield an uncertainty measure better than a method that does not discard outliers (in the sense that the influence of outliers is reduced).

Common formulations for SLAM problems assume that sensors move. At the core of most approaches is the use of EKF estimation for the simultaneous estimation of position of robots and landmarks. The covariance of estimation error is obtained directly from the EKF equations. The use of the EKF implicitly assumes that some model of the motion of the sensor is available Covariance intersection has also been used as alternative for fusing covariances obtained from unrelated sources. When a motion model is not available, sensor noise is unknown, or outliers are likely to corrupt data, uncertainty must be estimated through robust strategies able to reduce the influence of disturbances.

II. S-estimators

Estimators that are simultaneously robust and high breakdown are computationally complex so that the best estimators currently available tend to focus on complexity (see for instance the discussion in [5]).

This paper aims at testing the performance of a S-estimator in mapping applications using two different solution strategies. The first strategy uses closed form expressions, available in literature, to estimate mean an covariance and an iterative procedure that is not guaranteed to converge. The second strategy addresses the estimation as a constrained optimization problem.

Given an n-variable dataset x, of length m, the class of S-estimators for a covariance matrix can be obtained by minimizing the volume of the error ellipsoid of the covariance estimate, $\det(\hat{\Sigma})$, subject to a constraint on the Mahalanobis distance of the form (see for instance [1], [9]), that is,

$$\min \det(\hat{\Sigma}) \tag{1}$$

subject to

$$\frac{1}{m}\sum_{i=1}^{m}\rho\left(\left(\left(x_i-\hat{\mu}\right)^T\hat{\Sigma}^{-1}\left(x_i-\hat{\mu}\right)\right)^{1/2}\right)=b_0 \tag{1b}$$

where $\hat{\mu}$ and $\hat{\Sigma}$ are the estimates of the mean and covariance for the given data, ρ is a symmetric C^1 function, and b_0 a constant (which can be given the interpretation of an average value of the ρ function).

The argument of $\rho(\cdot)$ is just the Mahalanobis distance of the i-th data point to the mean $\hat{\mu}$. In a sense, the ρ function imposes

the admissible shape for the stream of Mahalanobis distances. For example, this function weights specific regions in the range of Mahalanobis distances and hence constraining $\hat{\Sigma}$ to give relevancy only to the data in the right regions.

A. Iterative strategy

The typical behavior of the S-estimator is first assessed in this paper through a set of simulations using the iterative procedure described in [1] (see Table I). The results are shown in Figure 1 for two different datasets.

1.	Compute $\hat{\mu}$ and $\hat{\Sigma}$
2.	Scale $\hat{\Sigma}$ so that constraint (1b) is verified
3.	Iterate from step 1 until $\det(\hat{\Sigma})$ stops decreasing

Table I
THE ITERATIVE PROCEDURE FROM [1]

The core of this iterative algorithm is the scaling step 2. Choosing ρ as the Tukey function,

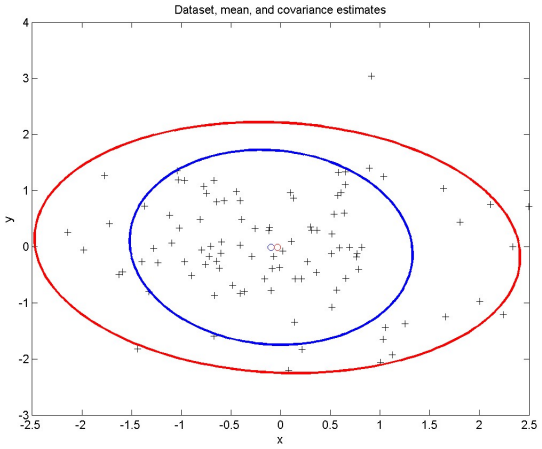
$$\rho(d_i) = d_i/2 - 2d_i^4/sC_0^2 + d_i^6/6C_0^4$$

monotonic over each of the half-spaces $d_i \geq 0$ and $d_i \leq 0$ (it is easy to check that the stationary points are at $\{-C_0, C_0, 0\}$, and using the simple scaling

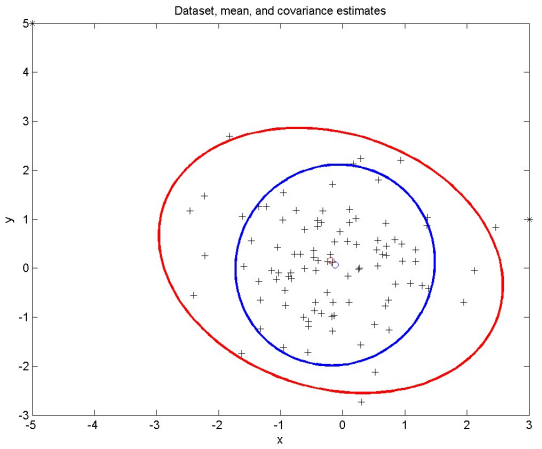
$$\text{scale factor} = \sum_{i=1}^{n} \rho(d_i)/(b_0 m)$$

achieve good results (though convergence is not guaranteed).

The computation of $\hat{\mu}$ and $\hat{\Sigma}$ follows the expressions in [1]. The results were obtained for normal populations with identical characteristics, with the exception of 2 samples in plot b which were replaced by outliers, show a good performance rejecting the influence of these outliers. Also, the volume of the error ellipsoid associated with the S-estimates (in blue) can be considerably smaller than that of the usual maximum-likelihood sample covariance estimation (in red). In both runs $b_0 = 1$ and $C_0 = 10$.



(a) Normally distributed data



(b) Normally distributed data with outliers

Figure 1. Typical performance of the iterative S-estimator

B. Solution as an optimization problem

Given that the unknowns are $\hat{\mu}$ and $\hat{\Sigma}^{-1}$ the solution is found in two steps. In the first step it is assumed that $\hat{\Sigma}^{-1}$ is known and solve for $\hat{\mu}$. Once a solution is found for $\hat{\mu}$ the problem is then solved for $\hat{\Sigma}^{-1}$.

If necessary, the procedure iterates until convergence of both values or some limit number of iterations is reached (see figure 2).

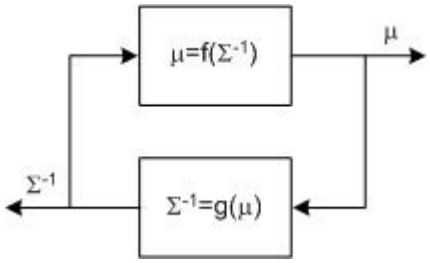


Figure 2. Solution strategy

For the procedure to converge it is necessary and sufficient that $f \circ g$ and $g \circ f$ both have fixed points, that is if both maps are non-expanding (see [2]). The proof that the maps involved are indeed non-expanding is given in the following sections.

1) Estimating $\hat{\mu}$: In this step (1b) is solved simply through the quadratic problem

$$\min_{\hat{\mu}} \left(\sum_{i=1}^{m} \rho \left(\left((x_i - \hat{\mu})^T \hat{\Sigma}^{-1} (x_i - \hat{\mu}) \right)^{1/2} \right) - m b_0 \right)^2$$

This is a convex problem, easily solved using standard optimization software. Figure 3 shows results obtained for different values of b_0 and C_0 for 100 runs using a normal dataset with 1000 points with $\mu = (0, 0)$ and $\Sigma = \mathbb{I}_2$. The plot shows two surfaces each interpolated from one coordinate of the $\hat{\mu}$ obtained.

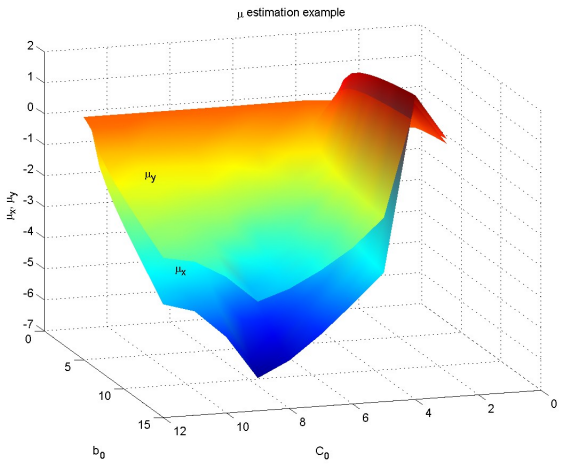


Figure 3. $\hat{\mu}$ estimates for $0 \leq b_0 \leq 10$ and $0 \leq C_0 \leq 10$

As $C_0, b_0 \to 0$ the estimate $\hat{\mu}$ tends to the real value ($C_0 = 0$ is a singularity of the Tukey function that yields no solution).

The plots show the convergence of the two coordinates to the correct value as b_0 goes to 0, that is, as the Mahalanobis distance is required to go to 0.

Let the constraint (1b) be written as,

$$\sum_{i=1}^{m} \phi \left(y_i^T \hat{\Sigma}^{-1} y_i \right) = \sum_{i=1}^{m} b_i \tag{2}$$

$$\sum_{i=1}^{m} b_i = b_0 \tag{2b}$$

where $\phi = \rho(\sqrt{(.)})/m$, $y_i = x_i - \hat{\mu}$, and the b_i are unknowns to be determined as part of the solution process (and constrained to (2b)).

Without loosing generality, Expression (2) can be written as a large system,

$$[y_1^T, \ldots, y_m^T] \begin{bmatrix} \hat{\Sigma}^{-1} & 0 & \cdots \\ 0 & \hat{\Sigma}^{-1} & \cdots \\ \vdots & & \vdots \\ & 0 & \hat{\Sigma}^{-1} \end{bmatrix} \begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix} = = y^T \hat{\Pi} y = p_0 \tag{3}$$

for some p_0 verifying

$$p_0 = \textstyle\sum_i \phi^{-1}(b_i)$$
$$\textstyle\sum_i b_i = b_0$$

The solution of (3) can be easily obtained as a function of p_0. Let P and Λ be the eigenvector matrix in the diagonalizing transformation of $\hat{\Pi}$ and the diagonal eigenvalues matrix, respectively. Then, the solution is of the form,

$$y = P \Lambda^{-1/2} \sqrt{p_0 \, \mathbf{1}_n} = P \Lambda^{-1/2} c$$

where $\mathbf{1}_n$ is a column vector corresponding to an element in the unit $(n-1)$-simplex. When spanning the $(n-1)$-simplex one obtains the whole range $\{y\}$ of solutions.

Using Cauchy-Schwarz inequality,

$$p_0 = |y^T \hat{\Pi} y| \leq \|y\|^2 \|\hat{\Pi}\| \tag{4}$$

and

$$\|y\| \leq \|P\| \|\Lambda^{-1/2}\| \|c\| \leq \|P\| \|\Lambda^{-1/2}\| C \tag{5}$$

for some constant $C \geq \|c\|$. Since P can be made a unitary matrix

$$\left(\frac{p_0}{\|\hat{\Pi}\|} \right)^{1/2} \leq \|y\| \leq \|\Lambda^{-1/2}\| C \tag{6}$$

that is, $\|y\|$ is bounded by the eigenvalues of $\hat{\Lambda}^{-1/2}$. Since Λ is diagonal, $\|\Lambda^{-1/2}\| = \|\Lambda^{-1}\|^{1/2}$ and hence

$$\|y\| \leq \|\Lambda^{-1}\|^{1/2} C \tag{7}$$

$\|\Lambda^{-1}\|$ depends, essentially, of the eigenvalues of Σ (the exact amount depending on the norm chosen) and hence a decrease in $\|\Sigma\|$ leads to a decrease in $\|\Lambda^{-1}\|$ and to a smaller decrease of $\|\Lambda^{-1}\|^{1/2}$ for $\|\Lambda^{-1}\| > 1$.

This means that the map $\|\Sigma\| \mapsto \|y\|$ is non-expanding for values of $\|\Sigma\|$ above 1 and expanding for values below 1 (in worst case as (7) is an upper bound).

2) Estimating $\hat{\Sigma}$: Each of the terms in the sum (2) can be represented in multiple forms, as for instance,

$$\phi\left(y_i^T \Sigma^{-1} y_i\right) = b_i \tag{8}$$

Since now Σ^{-1} is unknown, (1b) must be first converted to a form suitable to use the same strategy as before. Using theorem 7.5.2 in [3],

$$\Sigma^{-1} = v_1 v_1^T + \ldots + v_n v_n^T, \tag{9}$$

with the v_i orthogonal vectors and $n = \text{rank}(\Sigma)$, then the quadratic term $y_i^T \Sigma^{-1} y_i$ is decomposed as

$$\begin{aligned} y_i^T \Sigma^{-1} y_i = y_i^T \left(v_1 v_1^T + \ldots + v_n v_n^T\right) y_i = \\ = y_i^T v_1 v_1^T y_i + \ldots + y_i^T v_n v_n^T y_i = v_1^T y_i y_i^T v_1 + \ldots + v_n^T y_i y_i^T v_n = \\ = v_1^T \Upsilon_i v_1 + \ldots v_n^T \Upsilon_i v_n \end{aligned} \tag{10}$$

with $v_i^T v_j = 0,\ i \neq j$, leading to,

$$\phi(v_1^T \Upsilon_i v_1 + \ldots v_n^T \Upsilon_i v_n) = b_i \tag{11}$$

Applying the same idea over the complete set of samples, that is, $i = 1, \ldots, m$, yields,

$$\phi(v_1^T \Upsilon_i v_1 + \ldots v_n^T \Upsilon_i v_n) = b_i, \quad i = 1, \ldots, m \tag{12}$$

$$\sum_{i=1}^{m} b_i = b_0 \tag{12b}$$

$$v_i^T v_j = 0, \quad i \neq j, i = 1, \ldots, n \tag{12c}$$

with $m + 1 + n(n-1)/2$ equations and $n^n + m$ unknowns. Thus, in problems of practical interest ($n \geq 2$), the number of unknowns is bigger than the number of equations and the domain defined by (12-12c) is not empty (or even a singleton).

Following (1), the natural choice for the criterium to choose among a solution in this domain is,

$$\min\ \det(\hat{\Sigma}) \tag{13}$$

The cost function (13) is non convex but using a monotonic transformation such as

$$\min\ \log\left|\det(\hat{\Sigma})\right| \tag{14}$$

one gets a convex problem (see for instance [7], [8]).

The solution of (14) is obtained by exhaustive searching on the set of points obtained from (12a-c).

It is not difficult to see that points of the form

$$[v_1, \ldots, v_i, \ldots, v_n] = [0, \ldots, v, \ldots, 0]$$

can be made solutions of the above optimization problem. Since the matrices of the form $v_i v_i^T$ are singular one concludes that (9) becomes singular at these solutions.

A straightforward approach to tackle this problem is to add the following constraints,

$$\|v_i\| \geq v_{\min} \tag{15}$$

$$\forall i, j = 1, \ldots, n,\ \|v_i\| = \|v_j\| \tag{16}$$

For the purpose of the paper, $v_{\min} = 1e^{-3}$.

Figure 4 shows a consistency test for the Σ estimator. A random normal dataset, with 100 points of null mean and identity covariance, is fed into the estimator. Figure 4b shows the aggregate results for $\det(\hat{\Sigma})$ in a test with 100 runs. Moreover, $C_0 = 20$ and b_0 was chosen as the result of the computation of the lefthand side of (1b). The average for $\det(\hat{\Sigma})$ is 1.0235, close to the unit value of the true estimator.

In case the dataset includes outliers the results the S-estimator shows an interesting robustness property. Figure 5 shows a run using a dataset with identical statistical characteristics to the previous experiments but where two outliers were added, at positions $(4, 2)$ and $(5, 6)$. In this test C_0 and b_0 were chosen identical to the previous tests.

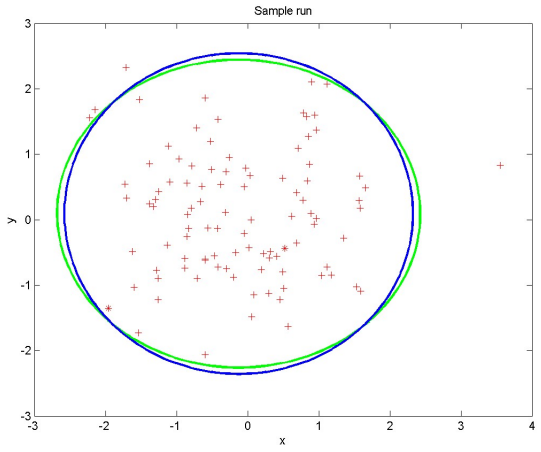
The true covariance (without outliers) error ellipse is shown in green. The influence of the outliers in the ML estimate including the outliers is clearly visible as the corresponding error ellipse (in red) has a dominant eigenvector pointing towards the outliers. The S-estimate (in blue) is closer to the true covariance than the ML estimate (in red).

The major drawback in this straight optimization approach is that the dimension of the problem increases with the dimension of the dataset which in turn increases the computational effort.

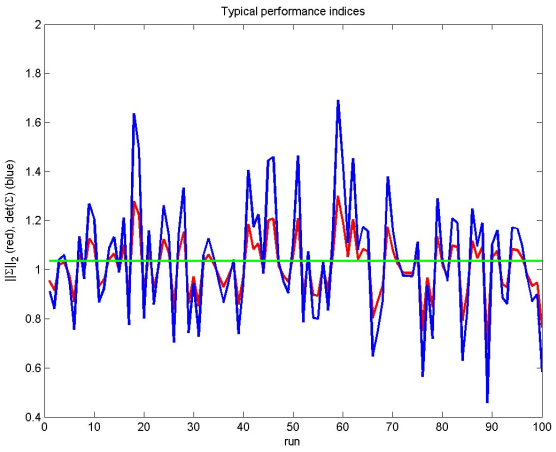
The analysis of the map $\hat{\mu} \mapsto \hat{\Sigma}$ follows the same general ideas of its reverse map in section II-B1. Expanding (12) as,

$$\begin{aligned} v_1^T \Upsilon_1 v_1 + \ldots + v_n^T \Upsilon_1 v_n + \ldots + v_1^T \Upsilon_m v_1 + \ldots + v_n^T \Upsilon_m v_n = \\ = v_1^T \left(\Upsilon_1 + \ldots + \Upsilon_m\right) v_1 + \ldots + v_n^T \left(\Upsilon_1 + \ldots + \Upsilon_m\right) v_n = \\ = p_0 \end{aligned} \tag{17}$$

for some $p_0 = \sum_{i=1}^{m} \phi^{-1}(b_i)$, one can write the more compact matrix form,



(a) Sample



(b) Aggregate

Figure 4. Consistency assessment for the Σ estimator; 100 runs

$$\left[v_1^T,\ldots,v_n^T\right]\begin{bmatrix}\Gamma & 0 & \ldots & 0\\ 0 & \Gamma & \ldots & 0\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \ldots & \Gamma\end{bmatrix}\begin{bmatrix}v_1\\ \vdots\\ v_n\end{bmatrix} = p_0 \quad (18)$$

with $\Gamma=\sum_{i=1}^{m}\Upsilon_i$.

Let $v^T = \left[v_1^T, v_2^T, \ldots, v_n^T\right]$. Expressing the orthogonality between pairs of vectors v_i^T, v_j can be written in matrix form as,

$$v^T\begin{bmatrix}0 & & \ldots & 0\\ & 0 & I_{ij} & \\ \vdots & & \ddots & \vdots\\ & & & 0\end{bmatrix}v = v^T P_{ij} v = v_i^T v_j = 0 \quad (19)$$

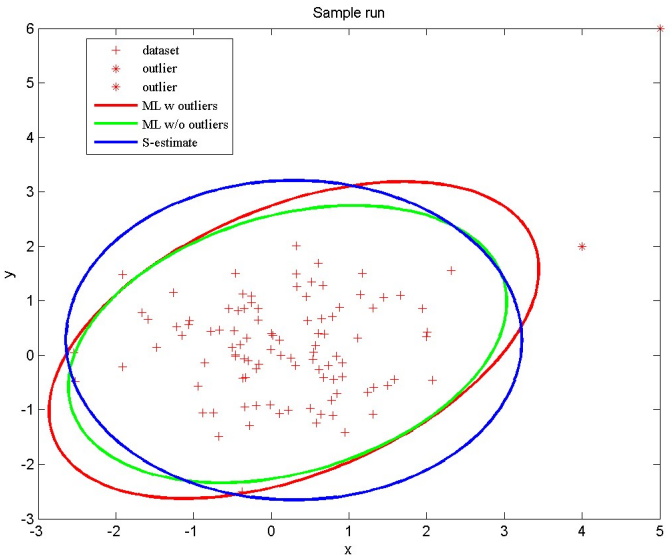


Figure 5. Sample run with 2 outliers

where I_{ij} is an $n\times n$ identity matrix located at block position i,j.

Let V be a $mn\times m$ block diagonal matrix of the form,

$$V=\begin{bmatrix}v & 0 & & \ldots\\ 0 & v & 0 & \ldots\\ \vdots & & \ddots & \vdots\\ 0 & & \ldots & v\end{bmatrix}$$

Merging expressions (18), and (19), yields

$$V^T\begin{bmatrix}\mathbf{\Gamma} & 0 & \ldots & 0\\ 0 & P_{12} & & 0\\ \vdots & & \ddots & \\ 0 & & \ldots & P_{ij}\end{bmatrix}V=\begin{bmatrix}p_0 & 0 & \ldots & \\ 0 & 0 & & \\ \vdots & & \ddots & \\ 0 & & & 0\end{bmatrix} \quad (20)$$

Expression (20) can also be written in a more compact form as

$$V^T\left(AV\right)=P_0 \quad (21)$$

where, by construction, A is a singular matrix.

The solution of (21) can be obtained using a least squares approach as

$$\mathcal{V}=\left(\mathcal{A}^T\mathcal{A}\right)^{-1}\mathcal{A}^T\mathcal{P}_0$$

where $\mathcal{A}$ is constructed from A and the constraints that define the equality of the v in the columns of V, $\mathcal{P}_0$ is the vector formed by stacking the columns of P_0, and $\mathcal{V}$ is the solution

vector with structure corresponding to the stacking of n vectors v. Then, using Cauchy-Schwartz inequality,

$$\|\mathcal{V}\| \leq \| \left(\mathcal{A}^T\mathcal{A}\right)^{-1} \| \|\mathcal{A}\| \|\mathcal{P}_0\|$$

and, using property $\| \left(\mathcal{A}^T\mathcal{A}\right)^{-1} \| \geq \|I\|/\|\mathcal{A}^T\mathcal{A}\|$, and theorem 5.6.9 in [3],

$$\|\mathcal{V}\| \leq \frac{\|\mathcal{A}\| \|\mathcal{P}_0\|}{\rho(\mathcal{A}^T\mathcal{A})}$$

with $\rho(\mathcal{A}^T\mathcal{A})$ the spectral radius of $\mathcal{A}^T\mathcal{A}$.

Given that $\|\mathcal{V}\| \leq n\|v\|$, $n > 1$, and the equality holding for $n = 1$, one can write

$$\|v\| \leq \frac{\|\mathcal{A}\| \|\mathcal{P}_0\|}{\rho(\mathcal{A}^T\mathcal{A})} \tag{22}$$

When $\|y\|$ decreases the eigenvalues of the Υ_i, and those of $\Upsilon_1 + \ldots + \Upsilon_m$, also decrease and hence also the eigenvalues of Γ. This means that $\|A\|$, and hence $\|\mathcal{A}\|$, decreases However, for $\rho(\mathcal{A}^T\mathcal{A}) > 1$ expression (22) decreases faster than $\|\mathcal{A}\|$ this meaning that the map $\|y\| \mapsto \|\hat{\Sigma}\|$ is non-expanding. The region defined by $\rho(\mathcal{A}^T\mathcal{A}) < 1$ corresponds to scenarios where $\|y\|$ is small.

Let $f_1 : \|\hat{\Sigma}\| \mapsto \|y\|$ and $f_2 : \|y\| \mapsto \|\hat{\Sigma}\|$ be the maps defined by bounds (7) and (22). If the variations of f_1 are big enough then map f_2 enters a non-expanding region, and similarly about big enough variations of f_2. This amounts to have variations in $\|y\|$ from the loop formed by the composition of f_2 and f_1 bounded by a negative semi-definite function, and similarly for variations in $\|\hat{\Sigma}\|$ generated by loop with the composition of f_1 and f_2. From the generalized version of the Lyapunov stability theorem (see for example theorem 8.2 in [6]), this means that the overall loop is stable (though eventually not asymptotically stable).

III. Experiments

This section presents a series of simulations using the estimator in (1)-(1b) under the two versions described in the previous sections, that is, based in [1], and using a mixture of an iterative technique with standard optimization.

The motivating application is the estimation of position of a set of targets and of the corresponding uncertainty, by a single vehicle equipped with a range sensor. Sensing data is acquired at a non uniform rate as the range is limited. At each measurement event the sensor sweeps the 2π with a beam of 2° aperture and if there is a target in range the corresponding measurement is returned.

While moving the vehicle keeps a local map with the targets detected. Once a measurement is acquired a data association procedure defines if the measurement belongs (i) to a newly detected target, or (ii) to a target already in the map.

The association procedure encompasses three stages and is based on grouping neighbor measurements. In the first stage a measurement is fused with the existing map if it falls inside the error ellipse of some target already in the map. If this condition is not verified the measurement is added to the map as a new target. At the second stage, after the end of the run, any two targets with overlapping error ellipses are fused together. The final stage groups targets lying in a small common neighborhood and selects only one of them to represent the real target. No fusion occurs at this stage. The covariance estimate is thus computed only at the first and second stages.

A measurement detecting a target has associated a position and a covariance matrix. The covariance estimate is computed by first joining the target position estimate corresponding to the measurement to the complete set of measurements taken by the robot along the path that were associated to this target, and applying the algorithm in Table I afterwards.

In all experiments $b_0 = 1$ and $C_0 = 10$.

A. Iterative strategy

Figure 6 illustrates the performance of the S-estimator in [1], in the mapping application, and in the absence of outliers and for two different gaussian noise levels. The real and estimated locations of the targets are marked $\circ$ (blue) and $+$ (green), respectively. For each estimated target the corresponding error ellipse, obtained from the estimated covariance is also shown (red). The red/black straight line are just the actual and reference flight paths.

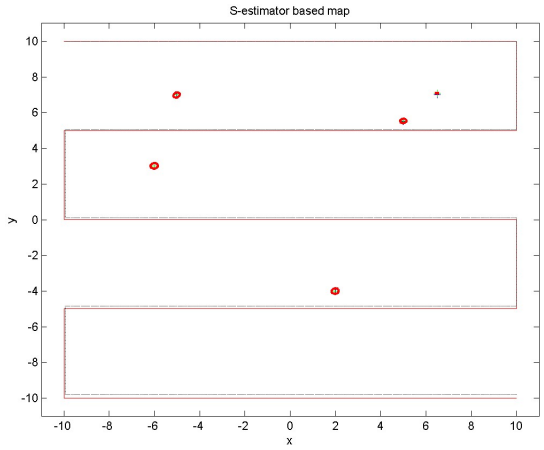
Figure 7 shows the results for the same mapping application when outliers are generated such that the time between them is exponentially distributed.

The fact that the real targets are detected to high accuracy whilst the phantom targets tend to show a much higher uncertainty is a consequence of the fact the phantom targets are constructed using only a small number of measurements (hence the uncertainty)

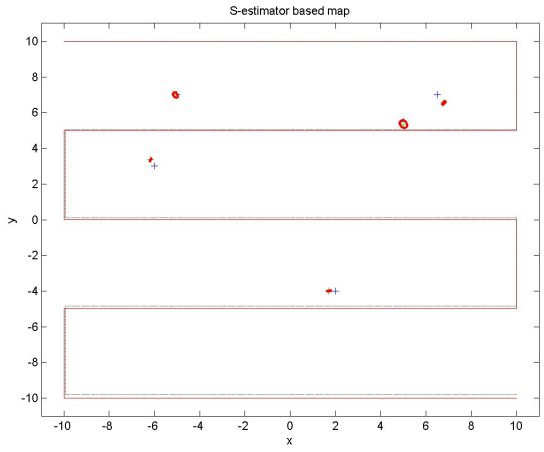
These results suggest that an additional filtering step can be used to remove some of the phantom targets, namely by sorting the targets according the 2-norm of the corresponding uncertainty. The idea is suggested by the fact that targets estimated from a bigger number of measurements will have smaller uncertainty and hence smaller 2-norm for the respective covariance matrix.

B. Optimization based strategy

As in the previous section, Figure 8 illustrates the performance of the S-estimator in the mapping application in the absence of outliers and for different gaussian noise levels. The exper-

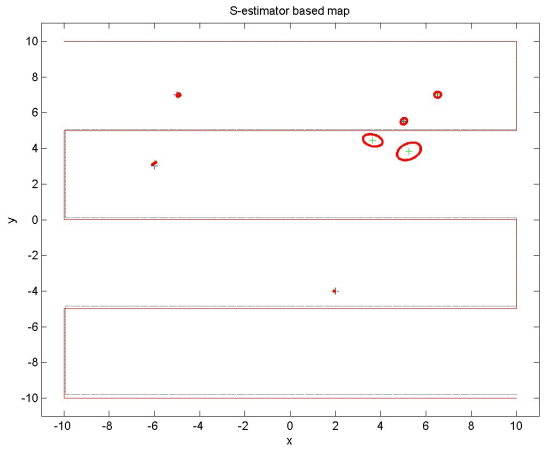


(a) noise 0.6

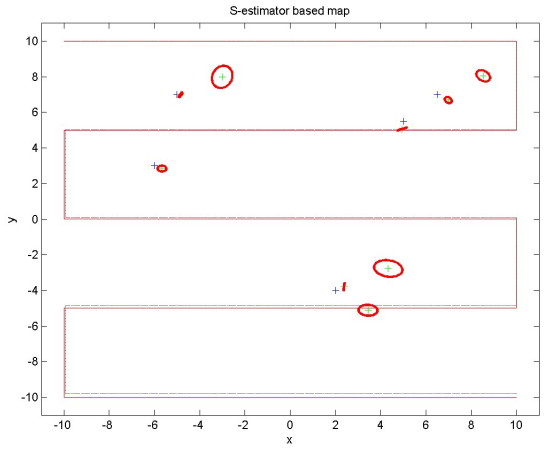


(b) noise 1

Figure 6. Mapping results for different noise levels and no outliers



(a) noise 0.6



(b) noise 1

Figure 7. Mapping results for different noise levels and with 6 outliers

iments are similar to those in the previous section. To reduce the computational burden, the number of measurements was reduced to 10% of those in the previous section.

Figure 9 shows the results for the same mapping application when outliers are generated such that the time between them has the same exponential distribution of the similar experiment in the previous section.

The results in this section seem to be slightly better than those obtained using the iterative approach described in [1]. Moreover, they also suggest improved performance over the hybrid OGK-CI estimator reported in [4].

IV. Conclusions

The paper addresses the use of a S-estimator in mapping applications. Two solution approaches are described and their performance assessed through simulation experiments and the results suggest a good performance by both strategies. The optimization based approach proposed in the paper was shown to converge.

Acknowledgments

This work was partially supported by FCT project PEst-OE/EEI/LA0009/2011.

References

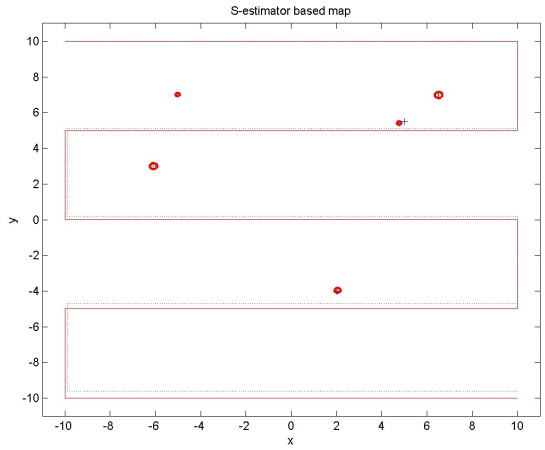
[1] N.A. Campbell, H.P. Lopuhaä, and P.J. Rousseeuw. On the Calculation of a Robust S-Estimator of a Covariance Matrix. *Statistics in Medicine*, 17:2685–2695, 1998.

[2] Kung-Ching Chang. *Methods in Nonlinear Analysis*. Springer Monographs in Mathematics. Springer, 2005.

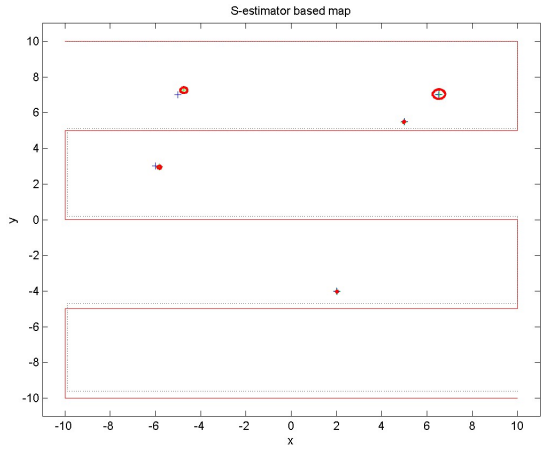
[3] Roger A. Horn and Charles R. Johnson. *Matrix Analysis*. Cambridge University Press, 21 edition, 2007. First published, 1985.

[4] A. Tsourdos J. Sequeira, S.B. Lazarus. Robust covariance estimation for data fusion from multiple sensors. *IEEE Transactions on Instrumentation and Measurement*, 2011. To appear.

[5] David J. Olive and Douglas M. Hawkins. Robust Multivariate Location and Dispersion. 2010.



(a) noise 0.6



(b) noise 1

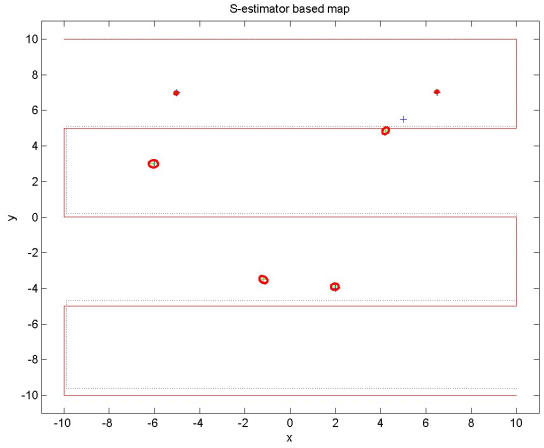
Figure 8. Mapping results for different noise levels and no outliers

[6] G. Smirnov. *Introduction to the Theory of Differential Inclusions*. Graduate Studies in Mathematics, vol 41. American Mathematical Society, 2002.

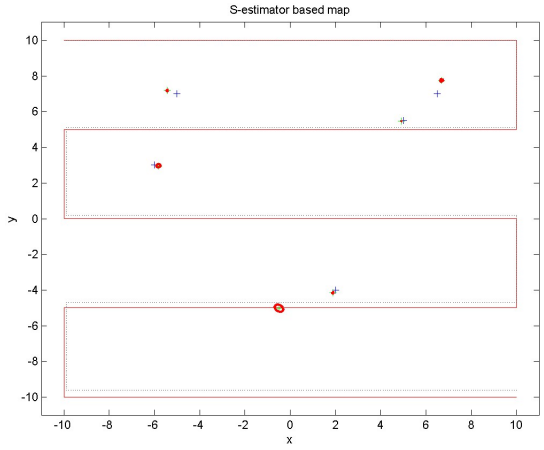
[7] Gilbert Strang. Inverse problems and derivatives of determinants. *Archive for Rational Mechanics and Analysis*, 114(3):255–265, 1991.

[8] Chengjing Wang, Defeng Sun, and Kim-Chuan Toh. Solving log-determinant optimization problems by a Newton-CG primal proximal point algorithm. *SIAM Journal on Optimization*, (20), 2010.

[9] Rand R. Wilcox. *Introduction to robust estimation and hypothesis testing*. Elsevier, 2005.



(a) noise 0.6, 13 outliers



(b) noise 1

Figure 9. Mapping results for different noise levels and outliers