Experimental Analysis of Dynamic Covariance Scaling for Robust Map Optimization Under Bad Initial Estimates

Pratik Agarwal Giorgio Grisetti Gian Diego Tipaldi Luciano Spinello Wolfram Burgard Cyrill Stachniss

***Abstract*— Non-linear error minimization methods became widespread approaches for solving the simultaneous localization and mapping problem. If the initial guess is far away from the global minimum, converging to the correct solution and not to a local one can be challenging and sometimes even impossible. This paper presents an experimental analysis of dynamic covariance scaling, a recently proposed method for robust optimization of SLAM graphs, in the context of a poor initialization. Our evaluation shows that dynamic covariance scaling is able to mitigate the effects of poor initializations. In contrast to other methods that first aim at finding a good initial guess to seed the optimization, our method is more elegant because it does not require an additional method for initialization. Furthermore, it can robustly handle data association outliers. Experiments performed with real world and simulated datasets show that dynamic covariance scaling outperforms existing methods, both in the presence and absence of data association outliers.**

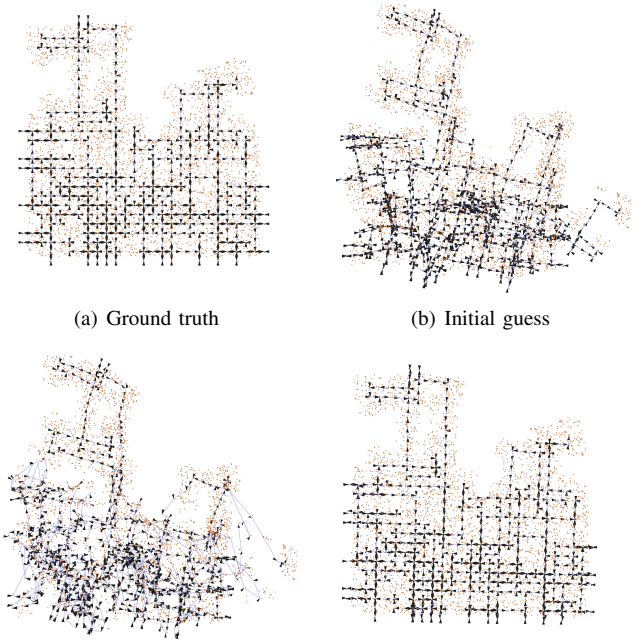
I. Introduction

State estimation and environment modeling are core capabilities of modern robots and in many state estimation and environment modeling problems, non-linear optimization plays a major role. This is, for example, the case in SLAM and bundle adjustment. Approaches to non-linear optimization such as Gauss-Newton, Levenberg-Marquardt, or Dog-Leg typically seek to find the minimum of the given error function. However, due to the non-convexity of the error surface, they cannot guarantee to find the global minimum. In practice, the initial guess has a strong impact on the quality of the computed solution. Finding the right solution can be challenging and sometimes even impossible if the initial guess is far away from the correct solution.

One of the reasons is that most approaches use linearizations of the error function around an initial guess to form the linear system used in the optimization. In case the initial guess if far away from the global minium, this approximation is likely to result in a poor solution. An example for that is illustrated in Figure 1.

There exists few approaches in the context of SLAM that explicitly address the problem of computing a good initial guess before optimization. The problem of good initialization is implicitly addressed by submapping and hierarchical techniques proposed over the last years in the context of EKF SLAM and graph-based techniques, for example [2, 5, 9, 18]. Although the motivation for most submapping techniques was bounding the computational complexity or online optimization, these techniques often also increase the robustness of the mapping system. Computing local solutions and combining them to a global solution can be seen as computing an improved initial alignment for parts of the problem. As a result, standard approaches often perform well when combining the partial solutions into a global one. Incremental optimization approaches [9, 14, 20] that optimize the graph at each step can have a similar effect.

Recently, Grisetti et al. [7] addressed the problem of computing a good initial guess for SLAM and bundle adjustment explicitly. Their approach is also related to submapping



(a) Ground truth (b) Initial guess

(c) Solution of Levenberg-Marquardt (100 iterations) (d) Solution of dynamic covariance scaling (15 iterations)

Fig. 1. A simulated robot (black triangles) equipped with a stereo camera moves in a grid world and observes features (orange squares). The top row shows the ground truth and the initialization. Levenberg-Marquardt fails to compute the optimal solution even after 100 iterations, while dynamic covariance scaling is able to obtain a close-to-optimal solution within 15 iterations.

All authors except Giorgio Grisetti are with the University of Freiburg, Institue of Computer Science, 79110 Freiburg, Germany. Giorgio Grisetti is with the La Sapienza University of Rome, Dept. of Systems and Computer Science, 00185 Rome, Italy. Cyrill Stachniss is also with the University of Bonn, Inst. of Geodesy and Geoinformation, 53115 Bonn, Germany.

This work has partly been supported by the European Commission under FP7-600890-ROVINA, ERC-AG-PE7-267686-LIFENAV, by the BMBF under contract number 13EZ1129B-iView and by the MWK for the project ZAFL-AAL.

as it partitions the factor graph and solves it locally, but the key contribution is to compute so-called condensed measurements (CM) from local solutions. This strategy leads to good initial configurations so that the original problem can then be solved with standard optimization approaches.

The contribution of this paper is an analysis of a recently proposed error minimization approach called dynamic covariance scaling (DCS) [1] under poor initial configurations that can occur in the context of SLAM. DCS was originally designed for problems that suffer from data association outliers and its key idea is to automatically re-weight constraints that yield a high error. Our analysis suggests that this approach has also a positive impact on the optimization in case of poor initial configurations. This positive impact is caused by the re-weighting as this reduces the influence of the Jacobians computed far away from the zero-error configuration. The experimental results presented in this paper show that DCS solves the optimization problems that previously required the computation of condensed measurements [7]. The advantage of DCS consists in not relying on any sophisticated initialization method and at the same time it naturally handles outliers. The code and the datasets[1] used in this evaluation are publicly available.

II. Related Work

SLAM is an active field of research in robotics and a large number of approaches has been presented in the past. There are different ways to address the SLAM problem and the graph-based formulation, initially proposed by Lu and Milios [17], is a popular solution these days.

Graph-based approaches model the poses of the robot and potential features in the environment by nodes in a graph and encode spatial relations resulting from observations and controls in the edges. Often, a least squares formulation is used and different methods have been proposed to minimize the error introduced by constraints. This includes relaxation [11, 6], stochastic gradient descent methods [19, 10], smoothing [4, 13, 14] and hierarchical techniques [2, 9, 7].

Most approaches assume Gaussian errors in the constraints of the graph, which are not resilient to outliers, i.e., wrong constraints between physically different locations. In the last two years, new approaches have been proposed that can deal with a substantial number of outliers among the constraints. The approach of Sünderhauf and Protzel [22, 23] has the ability to scale down the effect of potential outlier constraints. The scaling variable in their approach, is computed within the least squares formulation. Related to that, Olson and Agarwal [21] propose an approach that can consider multi-modal constraints efficiently. Outliers constraints can be captured by a Gaussian with a large variance. The RRR approach by Latif et al. [16] handles outliers by finding the maximum set of clustered edges that are consistent with each other. Compared to [23, 21], RRR fully rejects potentially wrong constraints while the two other approaches always maintain outlier constraints with a low weight. More recently, Agarwal et al. [1] proposed dynamic covariance scaling (DCS) as an alterative way of handling constraints.

The initial configuration of the graph to be optimized can have a strong impact on the final result as the error minimization procedure may get stuck in local minima. This holds for pose-graph SLAM as well as for graphs that contains robot poses and features. The sensor properties and the choice of the observation function has a strong impact on the convergence properties. As recently shown by Grisetti et al. [7], bad initializations quickly lead to divergence especially in the context of non-linear models. They propose to employ an approximation of the original problem that partitions the factor graph with a divide-and-conquer approach to exploit local estimates. As shown in their work [7], this offers a larger convergence basin than Levenberg-Marquardt and yields convergence to the true solution in real world and simulated scenes where other state-of-the-art methods fail. Hu et. al. [12] use the Cauchy M-estimator as a bootstrapper for optimizing datasets with high-noise but no outliers.

For pose-graphs, Carlone et al. [3] proposed a solution for finding a linear approximations. These results, however, do not generalize to graphs with features or with constraints in 3D.

This paper evaluates dynamic covariance scaling (DCS) in the situations considered by Grisetti et al. [7] and shows that DCS offers similar convergence properties, without requiring any condensed measurements, partitioning of the graph, re-initialization, or similar. We show that DCS is capable of optimizing non-linear problems with bad initial guesses and without the need of a special initialization method. We illustrate that DCS manages to reach the global minimum in challenging cases where only special initialization mechanism are successful. It is important to note that a better initialization, which is closer to the correct solution, will also result in a faster convergence with DCS. Note that the integration of DCS into existing optimization frameworks can be done with minimal modifications. This allows for optimizing non-linear problems starting from bad-initial configurations, which was not easily possible before.

III. Optimization with DCS

Graph-based SLAM systems aim at finding the configuration of the nodes that minimizes the error induced by observations. Let $X = (x_1, \dots, x_n)^T$ be the state vector where x_i describes the pose of node i. We can describe the error function $\mathbf{e}_{ij}(X)$ for a single constraint between the nodes i and j as the difference between the obtained measurement z_{ij} and the expected measurement $f(x_i, x_j)$

$$\mathbf{e}_{ij}(X) \quad = \quad f(x_i, x_j) - z_{ij}. \tag{1}$$

[1]The datasets can be downloaded from `http://www.informatik.uni-freiburg.de/~agarwal/resources/datasets-icra14-dcs.tar.gz`. DCS has been integrated into g2o and is available through its latest stable at `https://github.com/RainerKuemmerle/g2o`.

The error minimization can be written as

$$X^* = \underset{X}{\operatorname{argmin}} \sum_{ij} \mathbf{e}_{ij}(X)^T \Omega_{ij} \mathbf{e}_{ij}(X), \quad (2)$$

where Ω_{ij} is the information matrix associated to a constraint. Eq. 2 is typically solved using Gauss-Newton or Levenberg-Marquardt and requires to compute a linearization of the error function $\mathbf{e}_{ij}(X)$ in each step.

Dynamic covariance scaling or DCS [1] is a recently proposed method, which was developed to optimize pose graphs in the presence of outliers. DCS handles outlier constraints by scaling their information matrix and reducing their effect during optimization. To achieve this, it replaces Eq. 2 by:

$$X^* = \underset{X}{\operatorname{argmin}} \sum_{i} \mathbf{e}_{i,i+1}(X)^T \Omega_{i,i+1} \mathbf{e}_{i,i+1}(X) + \sum_{ij} s_{ij}^2 \underbrace{\mathbf{e}_{ij}(X)^T \Omega_{ij} \mathbf{e}_{ij}(X)}_{\chi^2_{ij}} \quad (3)$$

The first summand in Eq. 3 refers to the constraints from odometry or incremental scan-matching and the second one to the loop closing constraints. It obtains increased robustness by scaling each error term $\mathbf{e}_{ij}$ with s_{ij} or by scaling the information matrix Ω_{ij} with the squared of the scalar s_{ij}^2

$$\mathbf{e}_{ij}^{DCS} = \mathbf{e}_{ij}(X)^T (s_{ij}^2 \Omega_{ij}) \mathbf{e}_{ij}(X) \quad (4)$$

This reduces the confidence of outlier measurements. The scaling variable s_{ij} is computed as

$$s_{ij} = \min\left(1, \frac{2\Phi}{\Phi + \chi^2_{l_{ij}}}\right), \quad (5)$$

where Φ is a free parameter. A derivation of the scaling function and an analysis of the impact of Φ can be found in our previous work [1].

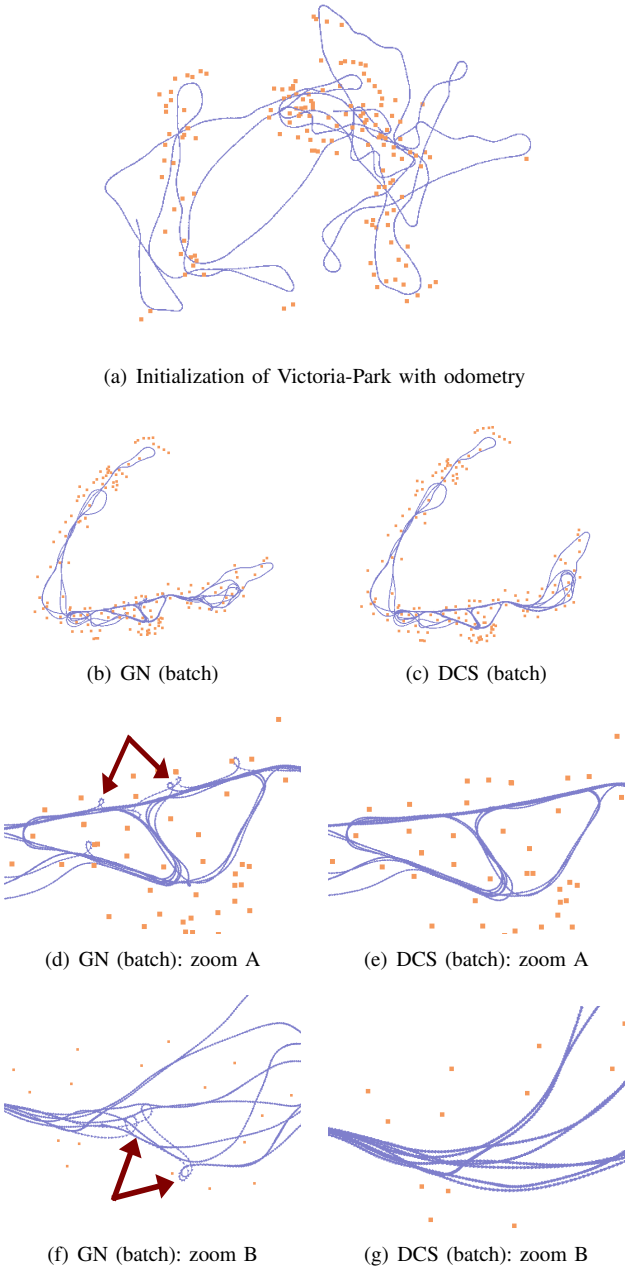
In practice, DCS has the effect of down-weighting constraints with large errors. Close to the zero-error configuration, DCS behaves like a normal squared kernel without any scaling. As the error increases, DCS scales the information matrix gradually.

With non-linear problems such as those involving orientations, the linear approximation of the error function $\mathbf{e}_{ij}(X)$ is poor if the initial estimate is far away from the correct solution. DCS mitigates the impact of a poor initial guess as it optimizes the problem while down-weighting constraints with large errors. The down-weighted constraints are those whose estimates are far away from the predicted measurements.

The final error minimization is carried out using the Levenberg-Marquardt approach. It leads to a quadratic form, which is minimized by solving the linear system

$$(H + \lambda I)\Delta X^* = -b, \quad (6)$$

where $H = \sum_{ij} J_{ij}^T (s_{ij}^2 \Omega_{ij}) J_{ij}$ and $b = \sum_{ij} J_{ij}^T (s_{ij}^2 \Omega_{ij}) e_{ij}$ are the elements of the quadratic form and J_{ij} is the Jacobian of the corresponding error function. The term λ is the damping factor of Levenberg-Marquardt and ΔX^* is the increment to the graph configuration that minimizes the error in the current iteration. The solution ΔX^*, which is here computed using g2o framework [15], is then used to update the current estimate. More details can be found in the graph-based SLAM tutorial [8].



(a) Initialization of Victoria-Park with odometry

(b) GN (batch)

(c) DCS (batch)

(d) GN (batch): zoom A

(e) DCS (batch): zoom A

(f) GN (batch): zoom B

(g) DCS (batch): zoom B

Fig. 2. Optimization of the Victoria-Park dataset with range-bearing measurements. The batch solution without DCS converges to the wrong solution. The errors in the robot poses are indicated by small loops in the odometry chain. These are not present when we used DCS. The batch solutions have a total error of $30,607.16$ compared to an error of 389.78 with DCS. Best results were obtained with $\Phi = 1$.

IV. Experimental Evaluation

We have evaluated DCS on both real and simulated datasets, which were originally evaluated with CM [7]. These

TABLE I

THE RESULTS OBTAINED FOR THE VICTORIA PARK DATASET USING DIFFERENT OPTIMIZATION METHODS (BATCH MODE).

Method	Resulting Error
CM	389
DCS	389.78
Gauss-Newton	30,607.16
Gauss-Newton with Dog-Leg	13,319.25
Levenberg-Marquardt	87,147.58

include Victoria-Park with range-bearing measurements and simulated 2D and 3D Manhattan world datasets with point features.

In all experiments, we used the odometry as the initial guess for the optimization. Figure 1(b) shows the initialization for a simulated dataset. To obtain a baseline comparison, we used ground truth initialization followed by the method under investigation. All errors reported with DCS were computed without the scaling function. Otherwise, DCS would report errors less than the global minimum after convergence

DCS has one free parameter Φ, which influences the scaling variable s, In all experiments, we set $\Phi = 10$ unless otherwise stated, but the optimization works on a wide range of values for Φ.

This experimental evaluation is designed to show the positive effect that DCS has on the computed solution in case of bad initial estimates. We show both, quantitative and qualitative benefits. We also illustrate the effect of the parameter Φ on the optimization process.

A. Victoria Park Dataset

The original Victoria Park dataset contains range-bearing observations of trees, which are used as point landmarks. It contains a total of 151 landmarks observed from 6,969 poses. This high pose to landmark ratio makes the problem challenging to converge for batch methods as illustrated in Figure 2. The batch method with Gauss-Newton without DCS seems to converge to the correct solution as shown in Fig. 2(b), but a more detailed analysis reveals that this is not the case. Figures 2(d) and 2(f) show enlarged parts for the solution obtained by batch methods. Non-existing loops appears in the odometry chain, which corresponds to a local minima in the optimization process. Figures 2(e) and 2(g) show the correct results obtained with DCS. This correct result without the small loops can also be verified by incrementally optimizing the graph which typically comes at an increased overall computationcal cost.

As depicted in Table I, The total error of the solution with DCS is 389.78 compared to $30,607.16$ with Gauss-Newton, $13,319.25$ with Dog-Leg and $87,147.58$ with LM. The solution obtained with DCS is similar in quality and final error compared to the CM approach [7]. Table. II shows that DCS converges to the correct solution for a wide variety of Φ. $\mathrm{F}_{\text{robust-DCS}}$ is the χ^2 error computed with the robust kernel. Note that the $\mathrm{F}_{\text{robust-DCS}}$ cannot be directly compared to F_{CM} as they use different error function. Thus, we run a few iterations of DCS setting all $s_{ij} = 1$. This results in using

TABLE II

DCS CONVERGES TO THE CORRECT SOLUTION FOR A RANGE OF VALUES FOR THE PARAMETER $\Phi \in [0.1, 20]$.

Φ	F_{DCS}	$\mathrm{F}_{\text{robust-DCS}}$
0.1	389.78	37.01
1.0	389.78	79.97
2.0	389.78	86.97
3.0	389.78	94.10
4.0	389.78	128.17
5.0	389.78	135.34
6.0	1,581.72	513.13
7.0	1,009.15	145.09
8.0	1,009.15	148.74
9.0	1,009.15	152.87
10.0	1,009.15	157.51
20.0	1,009.15	804.95

the identical squared error function as none of the constraints are scaled. It can also be interpreted as running GN with the initialization computed by DCS. In our previous work [1], we also showed that DCS could reject significant number of data association outliers in the Victoria Park dataset.

B. Simulation Results without Outliers

These experiments are designed to show that our approach is more robust with respect to the inital guess compared to Levenberg-Marquardt (LM) and performs similarly to the condensed measurement approach. The 2D simulated datasets contain planar range-bearing measurements. The 3D simulated datasets contain depth, disparity, and range-bearing sensor modalities. These were simulated using `g2o_simulator2d` and `g2o_simulator3d` methods.

Table III summarizes the result of experiments on the simulated datasets showing the type of dataset (2D or 3D), its size in terms of robot poses, number of landmarks, number of constraints, as well as the measurement model and the final χ^2 errors. The measurement model "cartesian" describes a sensor that is capable of measuring the $(\Delta X, \Delta Y)$-position of a landmark in the reference frame of the observer. "Depth" refers to a sensor that measures the depth of points in an image plane. Finally, "disparity" refers to a stereo camera model. F_{init} represents the total χ^2 of the initialization, which is performed by composing the odometry measurements. The landmarks are then initialized using the first pose-landmark constraint. $\mathrm{F}_{\text{ideal}}$ is obtained by running Levenberg-Marquardt (LM) starting with the ground truth solution as the initial guess, i.e., running LM on the perfect initialization. $\mathrm{F}_{\text{ideal}}$ will form our baseline comparison for correctness of the solution. F_{LM} is the final error after running LM on the odometry based initialization. F_{CM} is the result optained by the method of Grisetti et al. [7] followed by LM. F_{DCS} represents the final error of the DCS solution. The last column displays the total number of iterations required by DCS.

Table III shows that by using DCS the optimization always converges to the correct solution. LM can solve the smaller 3D datasets but as the problem size increases, it is unable to reach the correct solution. The significant examples are those, where LM fails to converge to the correct solution but DCS

TABLE III

SUMMARY OF THE COMPARISON EXPERIMENTS BETWEEN LM, CM, AND DCS.

Dataset	#constraints	Sensor model	F_{init}	F_{ideal}	F_{LM}	F_{CM}	F_{DCS}	#DCS-iters
A (2D)	1229	cartesian	25137.90	1706.69	1706.69	1709.69	1706.76	5
B (2D)	10223	cartesian	366551.00	18079.25	18079.25	18079.25	18079.39	5
C (2D)	105399	cartesian	1.26742e+09	205207.54	205207.54	205206.32	205206.35	7
D (2D)	534688	cartesian	1.79237e+10	1056677.58	1056677.58	1056677.58	1056677.58	10
E (3D)	226	depth	4706.70	116.91	116.91	116.91	116.91	6
E (3D)	226	disparity	6300.35	115.77	115.77	115.77	115.77	6
F (3D)	1809	depth	4.22496e+06	2988.96	2988.96	2988.96	3275.75	20
F (3D)	1809	disparity	1.40376e+07	2936.47	8038.63	2936.47	4309.50	20
G (3D)	19267	depth	1.72095e+11	43531.55	**16418112.01**	43531.54	43628.92	10
G (3D)	19267	disparity	4.53128e+11	43499.34	**10181039.20**	43499.35	43968.83	15
H (3D)	96659	depth	3.67085e+13	260937.23	**4547959956.76**	260937.23	261210.85	17
H (3D)	96659	disparity	2.42777e+12	261054.82	**1051509415.61**	261008.57	3172216.34	39

TABLE IV

DCS PERFORMANCE WITH A DIFFERENT NUMBERS OF OUTLIERS (PERCENTAGE OF OUTLIERS W.R.T. THE TOTAL NUMBER OF OBSERVATIONS). Y=RIGHT SOLUTION; N=WRONG MINIMA.

Dataset	Sensor model	#Constraints	5%	10%	25%	30%
C (2D)	cartesian	105399	Y	Y	N	N
D (2D)	cartesian	534688	Y	Y	N	N
G (3D)	depth	19267	Y	Y	Y	N
G (3D)	disparity	19267	Y	Y	Y	N

does. These include the larger 3D depth and disparity-based datasets. Note that the CM approach of Grisetti et al. [7] followed by LM always converges to the correct solution as DCS does. DCS, however, has the advantage over CM to not require an initialization technique that is different from the optimization method. In addition, DCS can also deal with data association outliers while CM cannot (compare [7, 1]).

C. Simulation Results in the Presence of Outliers

With the final experiment we want to show that DCS can deal with a substantial number of outliers, even for the more difficult optimization problems. Although the focus of this paper is not about robustness with respect to outliers, we evaluated how DCS was able to reject outliers in cases where LM could not optimize problems, even with zero outliers. Table IV summarizes our results when adding up to 30% outliers to the simulated datasets. We created the outliers by adding wrong constrains randomly between robot and landmark positions. The last four columns represent error percentage of false constraints added. "Y" represents a success and "N" represents failure. Note than LM was unable to converge to the correct solutions for the depth and disparity 3D-datasets G, even without outliers. For these challenging datasets, DCS converges to the correct solution even with 25% outlier constraints.

D. Influence of Φ on the Optimization

The next experiment is designed to illustrate the effect of the parameter Φ on the optimization process. In all experiments before, we set Φ to 10. It turns out that DCS converges for a large range of values for Φ, but it has an impact in the number of iterations needed. Figure 3 illustrates this behavior. The number of iterations required to reach the global minimum decreases with an increase in Φ. This does not scale arbitrarily since as $\Phi \to \infty$, DCS behaves like the original squared kernel.

E. Computation Cost

The computation time per iteration of all optimization methods used here is dominated by the sparse Cholesky factorization. The only overhead that DCS creates over LM is computing the scaling coefficient in each error function. This does not lead to any measurable increase in runtime.

V. CONCLUSION

The initial guess can have a substantial impact on the solution found by non-linear error minimization methods such as Gauss-Newton or Levenberg-Marquardt. We evaluated the dynamic covariance scaling method, a recent technique for solving SLAM problems with data association outliers, on SLAM-graphs with poor initial estimates. Our experiments suggest that dynamic covariance scaling is more resilient and robust to bad initial configurations compared than the standard use of Levenberg-Marquardt and Gauss-Newton for optimization. Our method can solve complex non-linear problems without the need of additional initialization mechanisms and without increasing the computational cost per iteration.

REFERENCES

[1] P. Agarwal, G.D. Tipaldi, L. Spinello, C. Stachniss, and W. Burgard. Robust map optimization using dynamic covariance scaling. In *Proc. of the IEEE Int. Conf. on Robotics & Automation (ICRA)*, 2013.

[2] M. Bosse, P. M. Newman, J. J. Leonard, and S. Teller. An ATLAS framework for scalable mapping. In *Proc. of the IEEE Int. Conf. on Robotics & Automation (ICRA)*, 2003.

[3] L. Carlone, R. Aragues, J. Castellanos, and B. Bona. A linear approximation for graph-based simultaneous localization and mapping. In *Proc. of Robotics: Science and Systems (RSS)*, 2011.

[4] F. Dellaert and M. Kaess. Square Root SAM: Simultaneous localization and mapping via square root information smoothing. *Int. Journal of Robotics Research*, 25(12):1181–1204, 2006.

[5] U. Frese. Treemap: An $O(\log n)$ algorithm for indoor simultaneous localization and mapping. *Autonomous Robots*, 21(2):103–122, 2006.

[6] U. Frese, P. Larsson, and T. Duckett. A multilevel relaxation algorithm for simultaneous localisation and mapping. *IEEE Transactions on Robotics*, 21(2), 2005.

[7] G. Grisetti, R. Kümmerle, and K. Ni. Robust optimization of factor graphs by using condensed measurements. In *Proc. of the IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS)*, 2012.

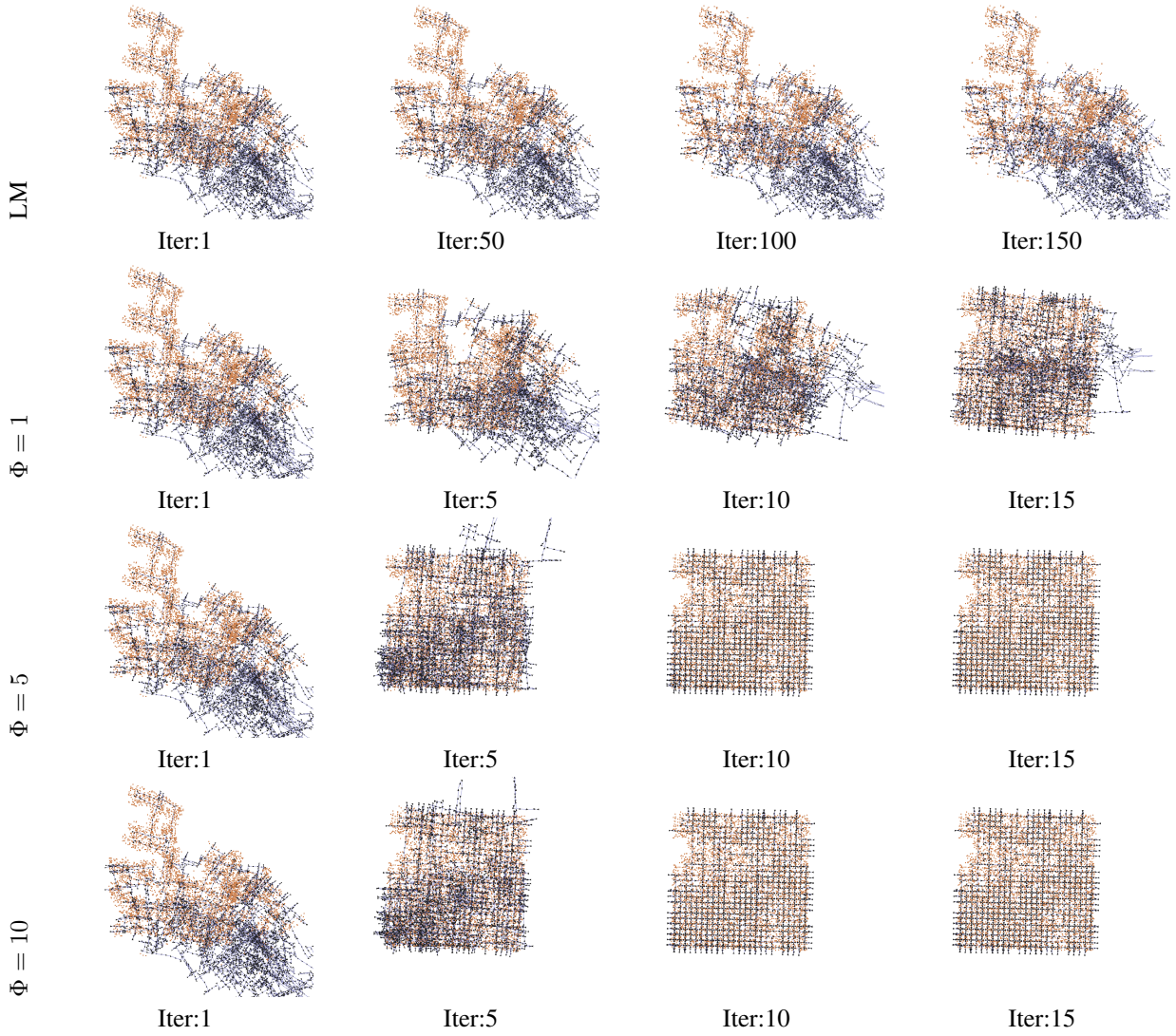


Fig. 3. Effect of Φ on the optimization process. All values of $\Phi \in 1, 5, 10$ are capable of optimizing the pose graph. By increasing the values of Φ from 1 to 10, lesser number of optimization steps are required. The above prose graph could not be solved using standard Levenberg-Marquardt. The time required for each iteration of LM is similar to that of DCS as the time is dominated by the sparse matrix factorization.

[8] G. Grisetti, R. Kümmerle, C. Stachniss, and W. Burgard. A tutorial on graph-based SLAM. *IEEE Transactions on Intelligent Transportation Systems Magazine*, 2010.

[9] G. Grisetti, R. Kümmerle, C. Stachniss, U. Frese, and C. Hertzberg. Hierarchical optimization on manifolds for online 2D and 3D mapping. In *Proc. of the IEEE Int. Conf. on Robotics & Automation (ICRA)*, 2010.

[10] G. Grisetti, C. Stachniss, and W. Burgard. Non-linear constraint network optimization for efficient map learning. *IEEE Transactions on Intelligent Transportation Systems*, 2009.

[11] A. Howard, M.J. Matarić, and G. Sukhatme. Relaxation on a mesh: a formalism for generalized localization. In *Proc. of the IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS)*, 2001.

[12] Gibson Hu, Kasra Khosoussi, and Shoudong Huang. Towards a reliable SLAM back-end. In *Proc. of the IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS)*, 2013.

[13] M. Kaess, A. Ranganathan, and F. Dellaert. Fast incremental square root information smoothing. In *Proc. of the Int. Conf. on Artificial Intelligence (IJCAI)*, 2007.

[14] M. Kaess, A. Ranganathan, and F. Dellaert. iSAM: Incremental smoothing and mapping. *IEEE Transactions on Robotics*, 26, 2008.

[15] R. Kümmerle, G. Grisetti, H. Strasdat, K. Konolige, and W. Burgard. g2o: A general framework for graph optimization. In *Proc. of the IEEE Int. Conf. on Robotics & Automation (ICRA)*, 2011.

[16] Y. Latif, C. Cadena, and J. Neira. Robust loop closing over time. *Proc. of Robotics: Science and Systems (RSS)*, 2012.

[17] F. Lu and E. Milios. Globally consistent range scan alignment for environment mapping. *Autonomous Robots*, 4, 1997.

[18] Kai Ni and Frank Dellaert. Multi-level submap based slam using nested dissection. In *Proc. of the IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS)*, 2010.

[19] E. Olson, J. Leonard, and S. Teller. Fast iterative optimization of pose graphs with poor initial estimates. In *Proc. of the IEEE Int. Conf. on Robotics & Automation (ICRA)*, 2006.

[20] E. Olson, J. Leonard, and S. Teller. Spatially-adaptive learning rates for online incremental SLAM. In *Proceedings of Robotics: Science and Systems*, Atlanta, GA, USA, 2007.

[21] Edwin Olson and Pratik Agarwal. Inference on networks of mixtures for robust robot mapping. *International Journal of Robotics Research*, July 2013.

[22] N. Sünderhauf and P. Protzel. BRIEF-Gist-closing the loop by simple means. In *Proc. of the IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS)*, 2011.

[23] N. Sünderhauf and P. Protzel. Switchable constraints for robust pose graph slam. In *Proc. of the IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS)*, 2012.