Pareto Meets Huber: Efficiently Avoiding Poor Minima in Robust Estimation Supplementary Material

Christopher Zach
Chalmers University of Technology, Sweden
christopher.m.zach@gmail.com

Guillaume Bourmaud
University of Bordeaux, France
guillaume.Bourmaud@u-bordeaux.fr

1. Proofs of Propositions 1-3

For convenience we repeat the propositions.

Proposition 1. Let $\mu \in (0,1)$ and $\nu > 0$ be given, and let $\mathbf{v}^*$ be the update vector given by

$$\mathbf{v}^* = -\big((1-\mu)\mathtt{H} + \mu\tilde{\mathtt{H}} + \nu\mathtt{I}\big)^{-1}\big((1-\mu)\mathbf{g} + \mu\tilde{\mathbf{g}}\big). \quad (1)$$

If $m_\Psi(\mathbf{v}^*) < 0$ and $m_{\tilde\Psi}(\mathbf{v}^*) < 0$, then there exists a $\beta > 0$ and $\nu' > 0$ such that $\mathbf{v}^*$ is also the solution of

$$\min_{\mathbf{v}} \max\left\{m_\Psi(\mathbf{v}), \beta m_{\tilde\Psi}(\mathbf{v})\right\} + \frac{\nu'}{2}\|\mathbf{v}\|^2. \quad (2)$$

Proof. Eq. 2 can be restated as

$$\min_{\mathbf{v}} \max_{\lambda\in[0,1]} (1-\lambda) m_\Psi(\mathbf{v}) + \lambda\beta m_{\tilde\Psi}(\mathbf{v}) + \frac{\nu'}{2}\|\mathbf{v}\|^2, \quad (3)$$

from which we deduce the first-order optimality condition with respect to $\mathbf{v}$

$$\begin{aligned}\mathbf{v}^* &= -\big((1-\lambda)\mathtt{H} + \lambda\beta\tilde{\mathtt{H}} + \nu'\mathtt{I}\big)^{-1}\big((1-\lambda)\mathbf{g} + \lambda\beta\tilde{\mathbf{g}}\big)\\ &= -\left(\frac{(1-\lambda)\mathtt{H} + \lambda\beta\tilde{\mathtt{H}} + \nu'\mathtt{I}}{1-\lambda+\lambda\beta}\right)^{-1}\left(\frac{(1-\lambda)\mathbf{g} + \lambda\beta\tilde{\mathbf{g}}}{(1-\lambda)+\lambda\beta}\right)\\ &= -\big((1-\mu)\mathtt{H} + \mu\tilde{\mathtt{H}} + \nu\mathtt{I}\big)^{-1}\big((1-\mu)\mathbf{g} + \mu\tilde{\mathbf{g}}\big), \quad (4)\end{aligned}$$

where we identify $\mu = \lambda\beta/(1-\lambda+\lambda\beta)$ and $\nu = \nu'/(1-\lambda+\lambda\beta)$. Given two of the values μ, λ and β, the remaining one is determined by

$$\beta = \frac{\mu(1-\lambda)}{\lambda(1-\mu)} \qquad \lambda = \frac{1}{1+\beta(1/\mu-1)}. \quad (5)$$

From $\mu = \lambda\beta/(1-\lambda+\lambda\beta) \in (0,1)$ we deduce that $0 < \lambda\beta$ and $\lambda\beta < 1-\lambda+\lambda\beta$, therefore $\lambda < 1$. Since $\lambda \in [0,1]$ the inequality $\lambda\beta > 0$ implies $\lambda > 0$ and $\beta > 0$. Thus, $\lambda \in (0,1)$ and $\beta > 0$. Since λ is dual optimal and is strictly in the interior of $[0,1]$, we have generalized complementarity,

$$\begin{aligned}0 &\in \partial_\lambda\big((1-\lambda)m_\Psi(\mathbf{v}^*) + \lambda\beta m_{\tilde\Psi}(\mathbf{v}^*)\big) + \partial_\lambda \imath_{[0,1]}(\lambda)\\ &= \{\beta m_{\tilde\Psi}(\mathbf{v}^*)\big) - m_\Psi(\mathbf{v}^*)\}, \quad (6)\end{aligned}$$

where ∂_λ is the subgradient w.r.t. λ. Hence, $m_\Psi(\mathbf{v}^*) = \beta m_{\tilde\Psi}(\mathbf{v}^*)\big)$ and β is nothing else than the ratio $\beta = m_\Psi(\mathbf{v}^*)/m_{\tilde\Psi}(\mathbf{v}^*)$ (which implies $\beta > 0$ using the assumptions). Therefore $(\mathbf{v}^*, \lambda)$ is a primal-dual pair satisfying complementary slackness (or primal-dual optimality conditions). Further, $\nu' = \nu(1-\lambda+\lambda\beta) > 0$ is also strictly positive. □

Proposition 2. Let $\mathbf{g}$ and $\tilde{\mathbf{g}} \neq \mathbf{0}$. If $\mathbf{g} + \gamma\tilde{\mathbf{g}} = \mathbf{0}$ for some $\gamma \geq 0$, then $\mathbf{x}_0$ is locally Pareto critical for $(\Psi, \tilde\Psi)$. Otherwise there exists a $\nu > 0$ and a $\mu \in (0,1)$ such that $\mathbf{v}^*$ given by Eq. 1 satisfies $m_\Psi(\mathbf{v}^*) < 0$ and $m_{\tilde\Psi}(\mathbf{v}^*) < 0$. In particular, a universally admissible choice for μ is given by $\mu = \frac{\|\mathbf{g}\|}{\|\mathbf{g}\|+\|\tilde{\mathbf{g}}\|}$. This in turn implies $\Psi(\mathbf{x}_0 + \mathbf{v}^*) < \Psi(\mathbf{x}_0)$ and $\tilde\Psi(\mathbf{x}_0 + \mathbf{v}^*) < \tilde\Psi(\mathbf{x}_0)$.

Proof. The first part is clear. Now as assume $\mathbf{g} + \gamma\tilde{\mathbf{g}} \neq 0$ for all $\gamma > 0$. Choose $\mu = \|\mathbf{g}\|/(\|\mathbf{g}\| + \|\tilde{\mathbf{g}}\|)$ (and therefore $1-\mu = \|\tilde{\mathbf{g}}\|/(\|\mathbf{g}\| + \|\tilde{\mathbf{g}}\|)$) and define

$$\hat{\mathbf{g}} := (1-\mu)\mathbf{g} + \mu\tilde{\mathbf{g}} \neq 0. \quad (7)$$

We have

$$\begin{aligned}\hat{\mathbf{g}}^T\mathbf{g} &= (1-\mu)\mathbf{g}^T\mathbf{g} + \mu\tilde{\mathbf{g}}^T\mathbf{g}\\ &= \frac{1}{\|\mathbf{g}\| + \|\tilde{\mathbf{g}}\|}\left(\|\tilde{\mathbf{g}}\|\|\mathbf{g}\|^2 + \|\mathbf{g}\|\tilde{\mathbf{g}}^T\mathbf{g}\right)\\ &= \mu\left(\|\tilde{\mathbf{g}}\|\|\mathbf{g}\| + \tilde{\mathbf{g}}^T\mathbf{g}\right) \quad (8)\\ \hat{\mathbf{g}}^T\tilde{\mathbf{g}} &= (1-\mu)\mathbf{g}^T\tilde{\mathbf{g}} + \mu\tilde{\mathbf{g}}^T\tilde{\mathbf{g}}\\ &= \frac{1}{\|\mathbf{g}\| + \|\tilde{\mathbf{g}}\|}\left(\|\tilde{\mathbf{g}}\|\mathbf{g}^T\tilde{\mathbf{g}} + \|\mathbf{g}\|\|\tilde{\mathbf{g}}\|^2\right)\\ &= (1-\mu)\left(\mathbf{g}^T\tilde{\mathbf{g}} + \|\mathbf{g}\|\|\tilde{\mathbf{g}}\|\right). \quad (9)\end{aligned}$$

Since $\mathbf{g}$ and $\tilde{\mathbf{g}}$ are assumed to be non-zero and not of opposite direction, we have that $\|\tilde{\mathbf{g}}\|\|\mathbf{g}\| + \tilde{\mathbf{g}}^T\mathbf{g} > 0$. Together with $\mu \in (0,1)$ we deduce that $\hat{\mathbf{g}}^T\mathbf{g} > 0$ and $\hat{\mathbf{g}}^T\tilde{\mathbf{g}} > 0$. Hence, $-\hat{\mathbf{g}}$ is a descent direction for both Ψ and $\tilde\Psi$.

Setting $\mu = \|\mathbf{g}\|/(\|\mathbf{g}\| + \|\tilde{\mathbf{g}}\|)$ is a natural but not the only choice. E.g., if $\mathbf{g}^T\tilde{\mathbf{g}} > 0$ (i.e. the angle between $\mathbf{g}$ and $\tilde{\mathbf{g}}$ is less than $\pi/2$), then setting $\mu \in (0,1)$ arbitrarily

is admissible. In general, the range of allowed values for μ shrinks with increasing angles between $\mathbf{g}$ and $\tilde{\mathbf{g}}$.

Define $\hat{\mathtt{H}} := (1-\mu)\mathtt{H} + \mu\tilde{\mathtt{H}}$ and let $\mathtt{U}\Lambda\mathtt{U}^T$ be the eigenvalue decomposition of $\hat{\mathtt{H}}$, where Λ is a diagonal matrix with elements $\lambda_i \geq 0$. Further, let η be the largest eigenvalue in $\mathtt{H}$ and $\tilde{\mathtt{H}}$. Therefore $\lambda_i \in [0, \eta]$. Recall that $\mathbf{v}^* := -(\hat{\mathtt{H}} + \nu\mathtt{I})^{-1}\hat{\mathbf{g}}$ is the update direction. Now,

$$\begin{aligned} -(\mathbf{v}^*)^T\mathbf{g} &= \hat{\mathbf{g}}^T(\hat{\mathtt{H}} + \nu\mathtt{I})^{-1}\mathbf{g} = \hat{\mathbf{g}}^T\mathtt{U}(\Lambda + \nu\mathtt{I})^{-1}\mathtt{U}^T\mathbf{g} \\ &= \hat{\mathbf{h}}^T(\Lambda + \nu\mathtt{I})^{-1}\mathbf{h}, \end{aligned} \tag{10}$$

where we introduced $\mathbf{h} := \mathtt{U}^T\mathbf{g}$ and $\hat{\mathbf{h}} := \mathtt{U}^T\hat{\mathbf{h}}$. Further, let $\kappa_i := \mathbf{h}_i^T\hat{\mathbf{h}}_i$. We therefore read

$$\begin{aligned} -(\mathbf{v}^*)^T\mathbf{g} &= \sum_i \frac{\kappa_i}{\lambda_i + \nu} \\ &= \sum_{i:\kappa_i \geq 0} \frac{\kappa_i}{\lambda_i + \nu} + \sum_{i:\kappa_i < 0} \frac{\kappa_i}{\lambda_i + \nu} \\ &\geq \sum_{i:\kappa_i \geq 0} \frac{\kappa_i}{\eta + \nu} + \sum_{i:\kappa_i < 0} \frac{\kappa_i}{\nu}. \end{aligned} \tag{11}$$

Define $\kappa^+ := \sum_{i:\kappa_i \geq 0} \kappa_i \geq 0$ and $\kappa^- := \sum_{i:\kappa_i < 0} \kappa_i \leq 0$. Thus,

$$-(\mathbf{v}^*)^T\mathbf{g} = \hat{\mathbf{g}}^T(\hat{\mathtt{H}} + \nu\mathtt{I})^{-1}\mathbf{g} \geq \frac{\kappa^+}{\eta + \nu} + \frac{\kappa^-}{\nu}. \tag{12}$$

A sufficient condition for $(\mathbf{v}^*)^T\mathbf{g} < 0$ is therefore given by

$$-(\mathbf{v}^*)^T\mathbf{g} \geq \frac{\kappa^+}{\eta + \nu_0} + \frac{\kappa^-}{\nu_0} > 0 \tag{13}$$

or

$$(\kappa^+ + \kappa^-)\nu_0 > -\kappa^-\eta. \tag{14}$$

Recall that $\kappa^- \leq 0$, therefore the r.h.s. is non-negative, and that

$$\kappa^+ + \kappa^- = \sum_i \kappa_i = \hat{\mathbf{h}}^T\mathbf{h} = \hat{\mathbf{g}}^T\mathbf{g} > 0. \tag{15}$$

Therefore we read the sufficient condition

$$\nu_0 > \frac{|\kappa^-|\eta}{\hat{\mathbf{g}}^T\mathbf{g}}. \tag{16}$$

for $-(\mathbf{v}^*)^T\mathbf{g} > 0$ (or $(\mathbf{v}^*)^T\mathbf{g} < 0$) to hold. Analogously we obtain a 2nd constraint on ν_0 that implies $(\mathbf{v}^*)^T\tilde{\mathbf{g}} < 0$,

$$\nu_0 > \frac{|\tilde{\kappa}^-|\eta}{\hat{\mathbf{g}}^T\tilde{\mathbf{g}}}, \tag{17}$$

where $\tilde{\mathbf{h}} := \mathtt{U}^T\tilde{\mathbf{h}}$ and $\tilde{\kappa}^- = \sum_{i:\hat{\mathbf{h}}_i\tilde{\mathbf{h}}_i<0} \hat{\mathbf{h}}_i^T\tilde{\mathbf{h}}_i$.

This means that for sufficiently large

$$\nu > \max\left(\frac{|\kappa^-|\eta}{\hat{\mathbf{g}}^T\mathbf{g}}, \frac{|\tilde{\kappa}^-|\eta}{\hat{\mathbf{g}}^T\tilde{\mathbf{g}}}\right) \tag{18}$$

the update vector $\mathbf{v}^*$ is a descent direction for both Ψ and $\tilde{\Psi}$. The final requirement on ν is, that the induced step is not too large in order to obtain reductions in the quadratic surrogates m_Ψ and $m_{\tilde{\Psi}}$. If the quadratic model for Ψ and $\tilde{\Psi}$ are majorizers as assumed, then

$$\begin{aligned} \Psi(\mathbf{x}+\mathbf{v}) &\leq m_\Psi(\mathbf{v}) = \Psi(\mathbf{x}_0) + \frac{1}{2}\mathbf{v}^T\mathtt{H}\mathbf{v} + \mathbf{g}^T\mathbf{v} \\ &\leq \Psi(\mathbf{x}_0) + \frac{\eta}{2}\|\mathbf{v}\|^2 + \mathbf{g}^T\mathbf{v} \\ \tilde{\Psi}(\mathbf{x}+\mathbf{v}) &\leq m_{\tilde{\Psi}}(\mathbf{v}) = \tilde{\Psi}(\mathbf{x}_0) + \frac{1}{2}\mathbf{v}^T\tilde{\mathtt{H}}\mathbf{v} + \tilde{\mathbf{g}}^T\mathbf{v} \\ &\leq \tilde{\Psi}(\mathbf{x}_0) + \frac{\eta}{2}\|\mathbf{v}\|^2 + \tilde{\mathbf{g}}^T\mathbf{v} \end{aligned}$$

for all $\mathbf{v}$. Observe that

$$\|\mathbf{v}^*\|^2 = \hat{\mathbf{g}}^T(\hat{\mathtt{H}} + \mathtt{I})^{-2}\hat{\mathbf{g}} \leq \frac{\|\hat{\mathbf{g}}\|^2}{(\eta+\nu)^2}. \tag{19}$$

Thus,

$$\frac{\|\hat{\mathbf{g}}\|^2}{(\eta+\nu)^2} + \mathbf{g}^T\mathbf{v}^* = \frac{\|\hat{\mathbf{g}}\|^2}{(\eta+\nu)^2} - \hat{\mathbf{g}}^T(\hat{\mathtt{H}} + \nu\mathtt{I})^{-1}\mathbf{v}^* < 0 \tag{20}$$

is a sufficient condition for $\frac{1}{2}(\mathbf{v}^*)^T\mathtt{H}\mathbf{v} + \mathbf{g}^T\mathbf{v}^* < 0$ (analogously for the second constraint). Using Eq. 12 an even stronger sufficient condition for ν is

$$\frac{\|\hat{\mathbf{g}}\|^2}{(\eta+\nu)^2} - \frac{\nu(\kappa^+ + \kappa^-) + \eta\kappa^-}{\nu(\eta+\nu)} < 0 \tag{21}$$

or

$$\frac{\nu\|\hat{\mathbf{g}}\|^2}{\eta+\nu} < \nu(\kappa^+ + \kappa^-) + \eta\kappa^-, \tag{22}$$

since $\nu > 0$. The l.h.s. converges to $\|\hat{\mathbf{g}}\|^2$ and the r.h.s. approaches $\nu(\kappa^+ + \kappa^-) = \nu\hat{\mathbf{g}}^T\mathbf{g} \to \infty$ for $\nu \to \infty$ (since $\hat{\mathbf{g}}^T\mathbf{g} > 0$), hence there exists a $\nu_1 > 0$ such that Eq, 22 holds for all $\nu > \nu_1$. More specifically, consider the mapping

$$f(\nu) := \frac{\nu\|\hat{\mathbf{g}}\|^2}{\eta+\nu} - \nu(\kappa^+ + \kappa^-) - \kappa^-\eta.$$

f is monotonically decreasing for

$$\nu > \max\left\{\nu_0, \sqrt{\frac{\eta}{\kappa^+ + \kappa^-}}\|\hat{\mathbf{g}}\| - \eta\right\} \tag{23}$$

and $\lim_{\nu\to\infty} f(\nu) \to -\infty$. Hence there exists a $\nu_1 > \nu_0$ such that $f(\nu) < 0$ for all $\nu > \nu_1$. The symmetric reasoning applies to ensure that $\frac{1}{2}(\mathbf{v}^*)^T\tilde{\mathtt{H}}\mathbf{v}^* + \tilde{\mathbf{g}}^T\mathbf{v}^* < 0$.

Since m_Ψ and $m_{\tilde{\Psi}}$ are assumed to be majorizers of $\Psi(c) - \Psi(\mathbf{x}_0)$ and $\tilde{\Psi}(c) - \tilde{\Psi}(\mathbf{x}_0)$, respectively, $m_\Psi < 0$ and $m_{\tilde{\Psi}} < 0$ imply $\Psi(\mathbf{x}_0 + \mathbf{v}^*) < \Psi(\mathbf{x}_0)$ and $\tilde{\Psi}(\mathbf{x}_0 + \mathbf{v}^*) < \tilde{\Psi}(\mathbf{x}_0)$. □

The assumption that surrogate models are majorizers can be replaced by the assumption that Ψ and $\tilde{\Psi}$ have a Lipschitz gradient (or equivalently bounded second derivatives).

Proposition 3. If $\Psi(\mathbf{x}_0+\mathbf{v}^*) > \Psi(\mathbf{x}_0)$, then $m_\Psi(\mathbf{v}^*) > 0$. If $\tilde{\Psi}(\mathbf{x}_0 + \mathbf{v}^*) > \tilde{\Psi}(\mathbf{x}_0)$, then $m_{\tilde{\Psi}}(\mathbf{v}^*) > 0$.

This is a direct consequence of $\Psi(\mathbf{x}_0) + m_\Psi(\mathbf{v})$ and $\tilde{\Psi}(\mathbf{x}_0)+m_{\tilde{\Psi}}(\mathbf{v})$ being majorizers for Ψ and $\tilde{\Psi}$, respectively.

2. More results for bundle adjustment

The complete list of instances is: ladybug-73, ladybug-138, ladybug-318, ladybug-598, trafalgar-126, trafalgar-138, trafalgar-201, trafalgar-225, trafalgar-257, dubrovnik-150, dubrovnik-202, dubrovnik-253, dubrovnik-308, dubrovnik-356, venice-89, venice-245, venice-427, venice-744, final-93, final-394. The numerical results were obtained on an AMD Ryzen 2950X workstation with 64GB of memory. Our C++ implementation is single threaded.

Fig. 1 visualizes the performance profile for full bundle adjustment (optimizing over focal lengths and lens distortion parameters as well). The performance profiles mirror the results for linearized and metric bundle adjustment given in the main text (with the LM-MOO variants leading the other methods).

Figs. 2-7 depict the evolution of the best encountered objective w.r.t. wall clock time for all bundle adjustment datasets and configuration (linearized, metric, full).

3. Refining RANSAC solutions

In principle all considered methods for robust cost minimization are also suitable for low-dimensional parametric fitting problems such as homography and fundamental matrix estimation. In such settings random sampling scheme such as RANSAC can be used to determine initial estimates of the unknown model parameters, which are subsequently refined by any robust cost optimization method. We ran experiments on homography and fundamental matrix datasets [1] and observed, that the choice of optimization method has very little impact e.g. on the refined inlier ratio. Even IRLS works competitively for these problems. Detailed results are given below.

Table 1 depicts results for a slightly simplified fundamental matrix refinement task, where the initial solution is determined by a vanilla RANSAC method. The refinement objective is a robustified Sampson error (not solely an algebraic error), but the rank-2 constraint is dropped to avoid the highly non-linear constraint (or parametrization). The inlier threshold is 1.5 pixels (as suggested in [1]), and we limit the RANSAC iterations to 10000 (or less, according to the RANSAC formula at 99% confidence). The number of refinement iterations is 50. The numbers in Table 1 are averaged over 100 runs. Each method for robust cost minimization received the same set of initial RANSAC solutions.

Interestingly, GOM+ wins most of the time, with LM-MOO usually coming in second. Using 20 or 100 refinement iterations yields similar numbers and ranking. Overall, in all datasets the difference between the means is significantly smaller than the respective standard deviations. Hence, we do at this point not consider the ranking of methods statistically significant for the purpose of refining an initial RANSAC solution. Table 2 summarizes the corresponding results for robust homography estimation. The overall picture is similar to the one of fundamental matrix estimation.

References

[1] Rahul Raguram, Ondrej Chum, Marc Pollefeys, Jiri Matas, and Jan-Michael Frahm. Usac: a universal framework for random sample consensus. *IEEE Trans. Pattern Anal. Mach. Intell.*, 35(8):2022–2038, 2013. 3, 4

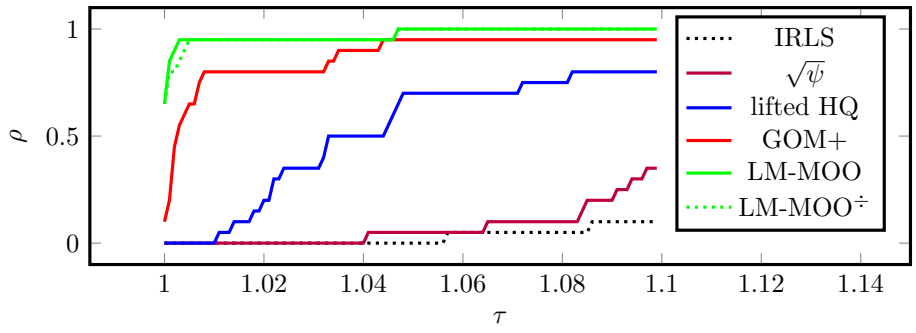

Figure 1: Performance profiles for full bundle adjusted computed from 20 bundle adjustment instances.

	RANSAC	IRLS	joint HQ	GOM+	LM-MOO
1	1396.68 (146.15)	1463.87 (145.04)	1457.17 (146.46)	**1472.36 (138.32)**	*1470.41 (143.31)*
2	298.72 (66.93)	312.29 (68.82)	*312.40 (68.34)*	**312.75 (66.82)**	312.33 (68.65)
3	365.32 (61.93)	378.45 (63.02)	380.28 (62.69)	**391.34 (61.36)**	*382.00 (63.57)*
4	170.63 (92.72)	207.79 (111.44)	208.28 (110.23)	**239.66 (117.09)**	*220.53 (116.82)*
5	145.95 (20.82)	151.80 (21.77)	153.65 (22.05)	**156.24 (23.37)**	*154.16 (22.14)*
6	103.76 (34.10)	114.90 (36.67)	116.55 (36.99)	**119.94 (35.98)**	*116.63 (37.17)*
7	387.81 (11.63)	397.64 (5.51)	*397.86 (5.57)*	393.15 (4.69)	**398.27 (6.14)**
8	684.15 (44.15)	701.10 (43.44)	703.21 (41.81)	**714.17 (28.71)**	*704.46 (40.45)*
9	2603.35 (19.10)	2614.35 (5.04)	2614.30 (5.21)	**2615.50 (6.13)**	*2614.75 (5.19)*
10	2240.34 (68.02)	2295.74 (44.98)	2295.01 (38.21)	**2303.56 (36.77)**	*2298.62 (46.63)*
11	209.01 (20.92)	214.56 (20.61)	*214.90 (20.66)*	**215.99 (20.57)**	214.79 (20.72)

Table 1: Inlier counts for fundamental matrix estimation for the 11 image pairs provided in [1]. The figures are the number of inliers at a 1.5 pixels threshold, and the respective standard deviation in parentheses. Bold corresponds to the winning (highest) number, and italics to the runner up. Overall, the differences between the means are significantly smaller than the respective standard deviations. Therefore we do at this point not consider the ranking of methods statistically significant.

	RANSAC	IRLS	joint HQ	GOM+	LM-MOO
1	1270.84 (93.62)	**1411.00 (0.00)**	**1411.00 (0.00)**	**1411.00 (0.00)**	**1411.00 (0.00)**
2	944.99 (60.13)	**998.00 (0.00)**	**998.00 (0.00)**	**998.00 (0.00)**	**998.00 (0.00)**
3	30.01 (18.71)	45.42 (27.53)	*46.70 (26.96)*	39.94 (18.98)	**48.51 (28.11)**
4	660.58 (45.80)	*705.00 (0.00)*	704.91 (0.38)	709.00 (0.00)	704.98 (0.14)
5	224.74 (63.35)	289.34 (46.85)	288.83 (46.70)	**297.09 (34.54)**	*293.76 (42.31)*
6	31.41 (15.00)	57.33 (23.72)	57.84 (22.32)	**68.49 (12.36)**	*62.48 (20.54)*
7	60.45 (5.11)	60.59 (5.78)	*61.40 (5.58)*	**61.69 (6.35)**	60.73 (5.79)
8	176.70 (4.52)	**179.00 (0.00)**	**179.00 (0.00)**	**179.00 (0.00)**	**179.00 (0.00)**
9	24.63 (27.19)	41.65 (47.28)	43.52 (48.26)	**71.09 (54.39)**	*47.42 (50.78)*
10	18.56 (1.63)	18.52 (1.72)	*19.38 (1.72)*	**19.55 (1.80)**	18.59 (1.76)

Table 2: Inlier counts for homography estimation for the 10 image pairs provided in [1]. The figures are the number of inliers at a 2 pixels threshold, and the respective standard deviation in parentheses. Bold corresponds to the winning (highest) number, and italics to the runner up. Again, the differences between the means are significantly smaller than the respective standard deviations.

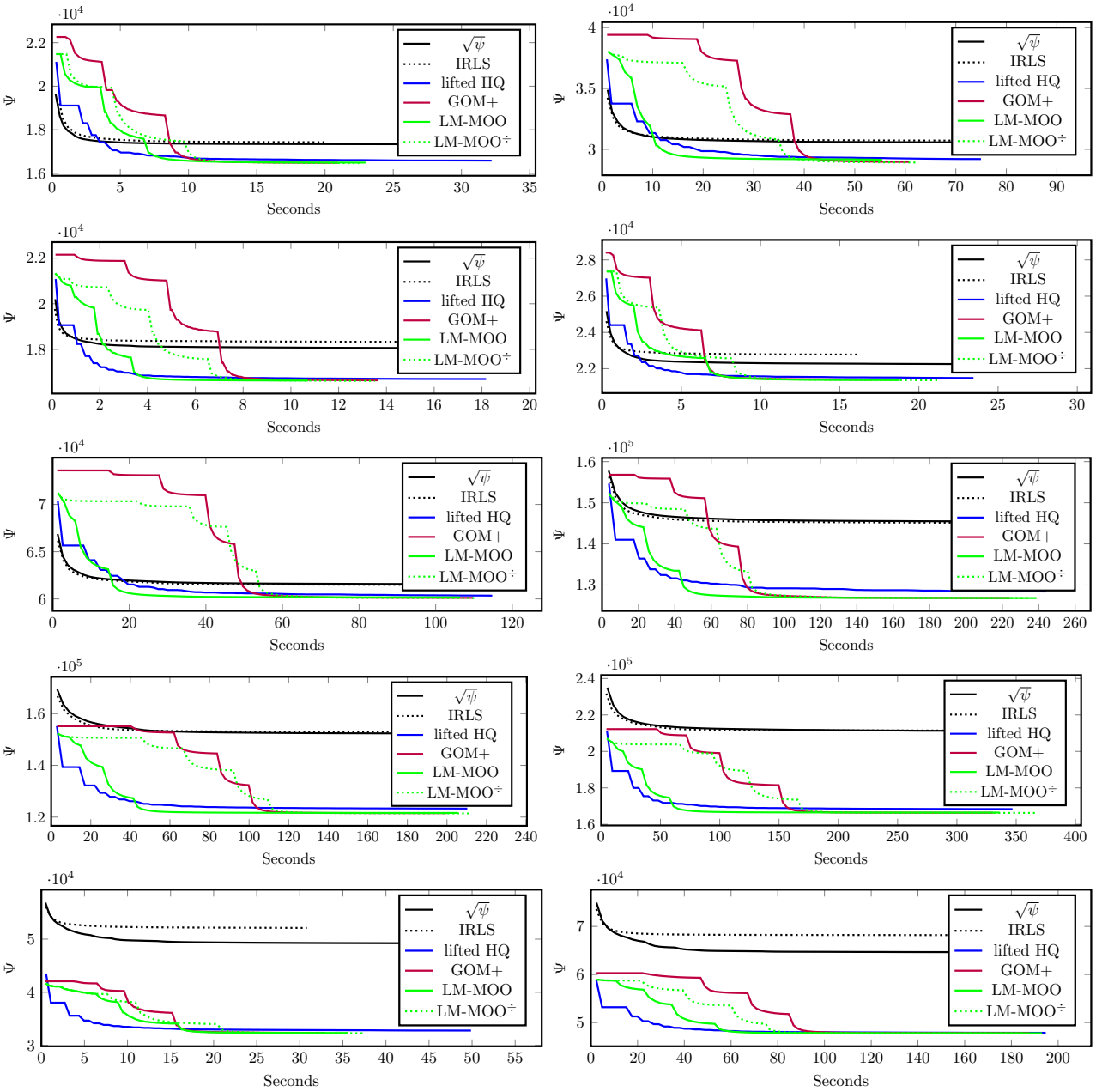

Figure 2: Evolution of the best cost for linearized bundle adjustment (datasets 1-10).

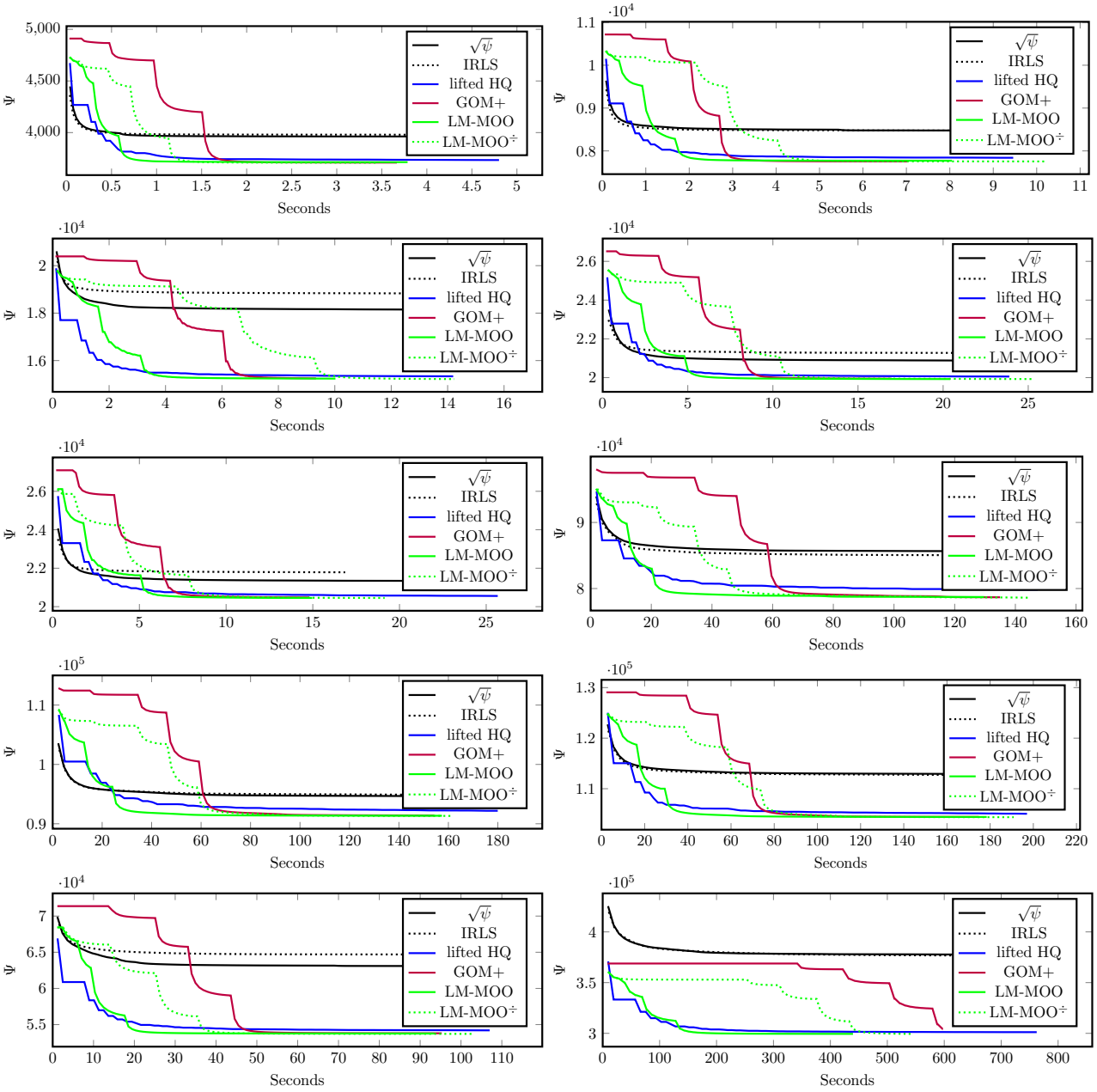

Figure 3: Evolution of the best cost for linearized bundle adjustment (datasets 11-20).

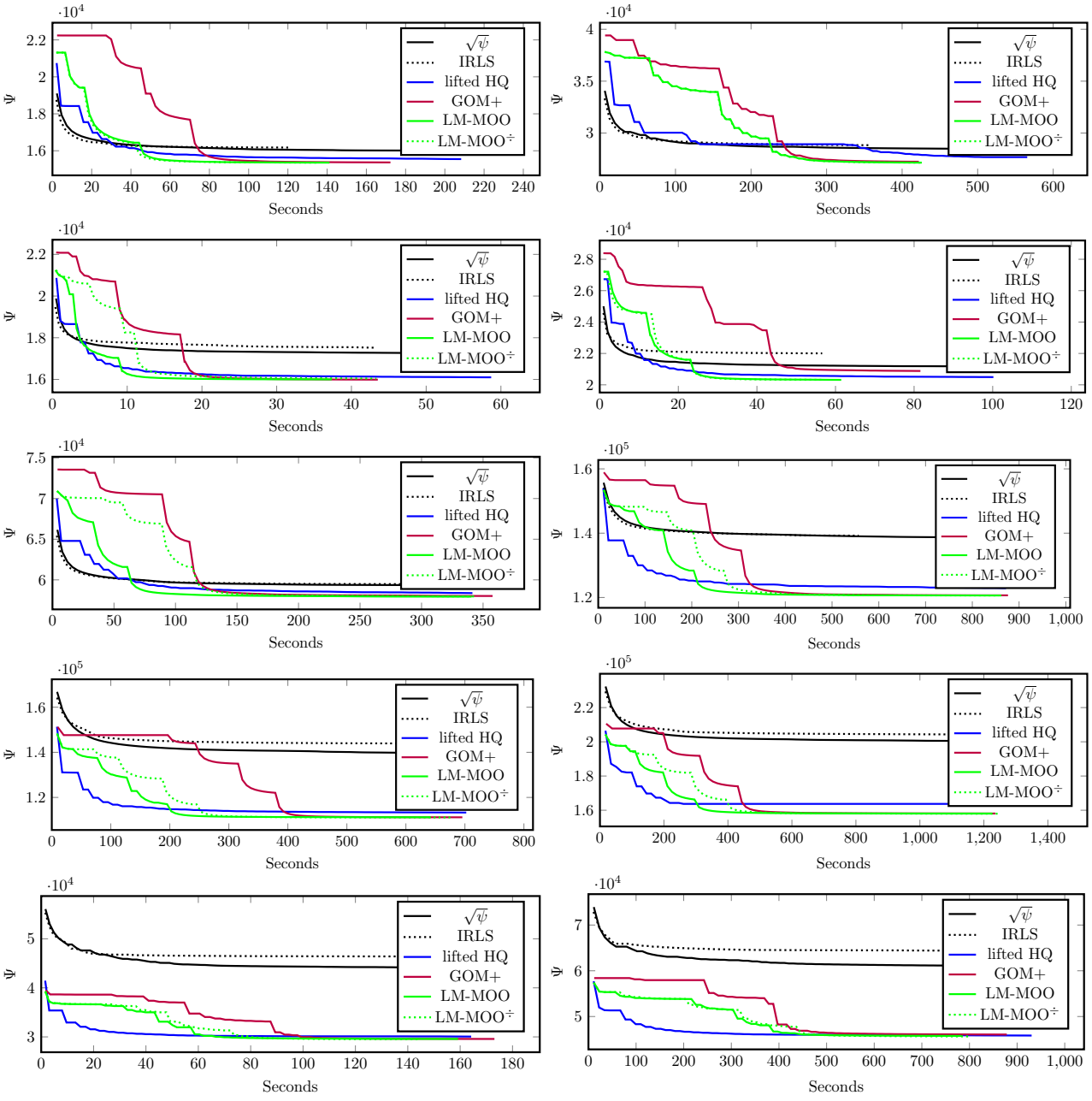

Figure 4: Evolution of the best cost for metric bundle adjustment (datasets 1-10).

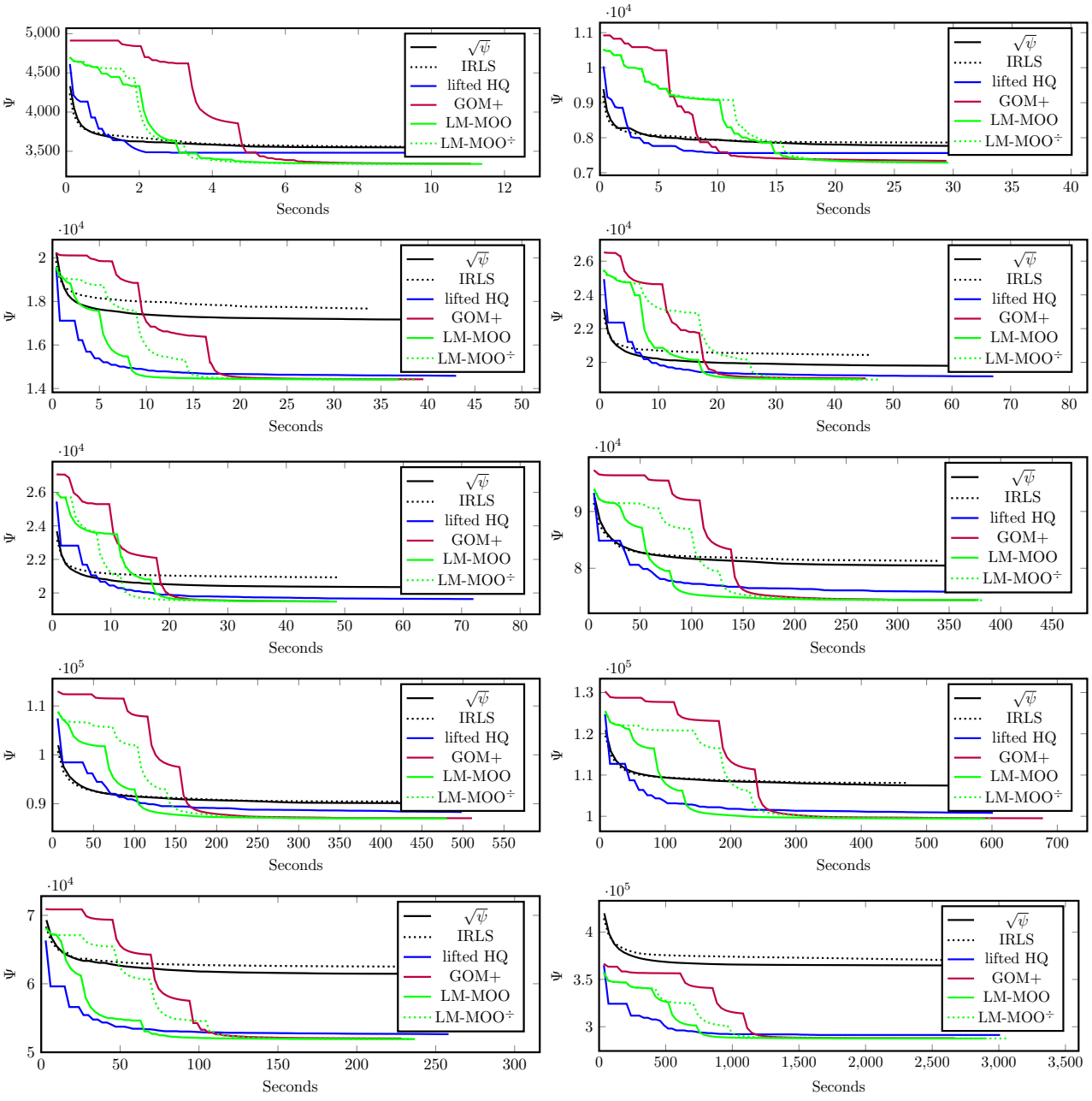

Figure 5: Evolution of the best cost for metric bundle adjustment (datasets 11-20).

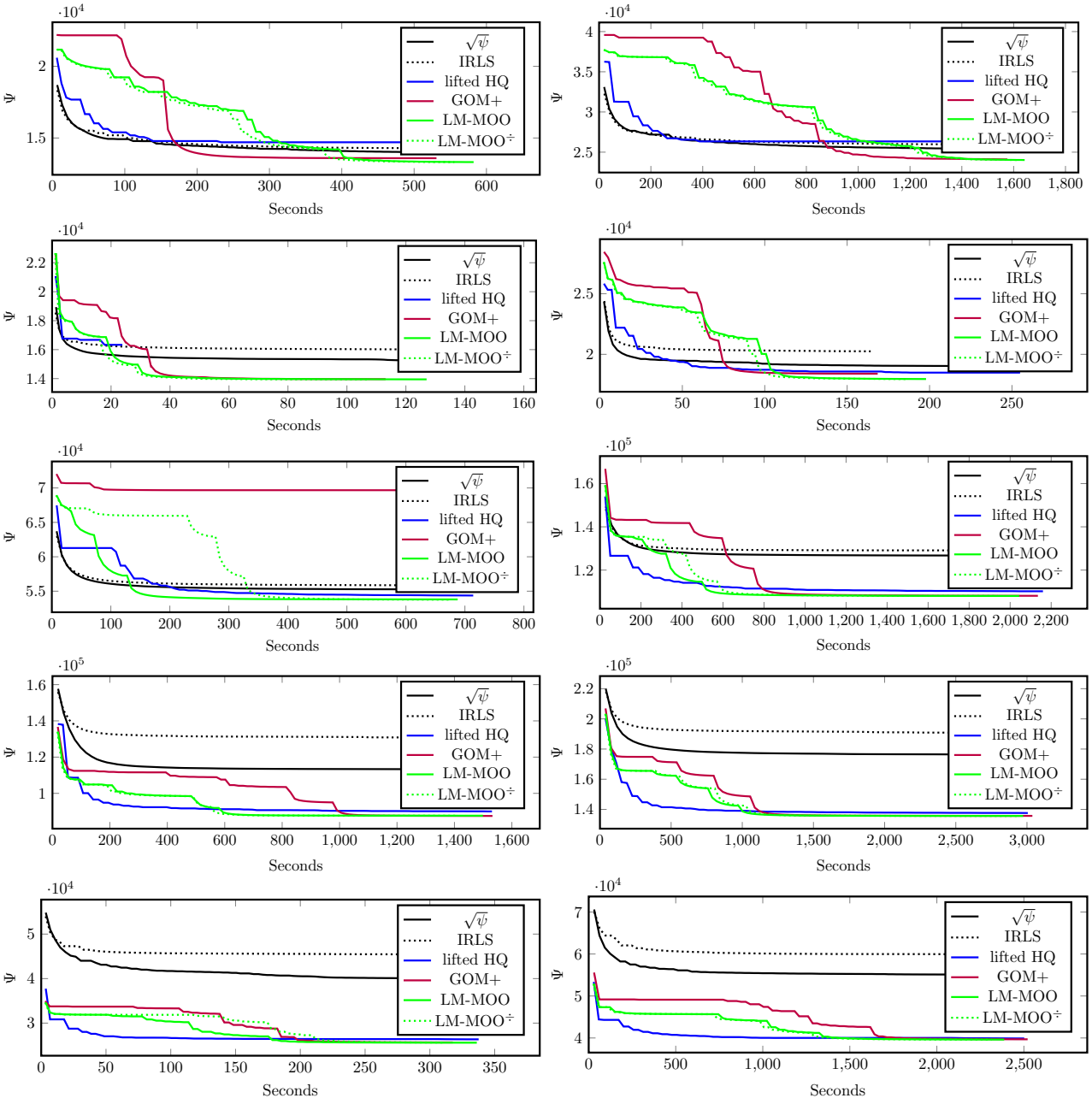

Figure 6: Evolution of the best cost for full bundle adjustment (datasets 1-10).

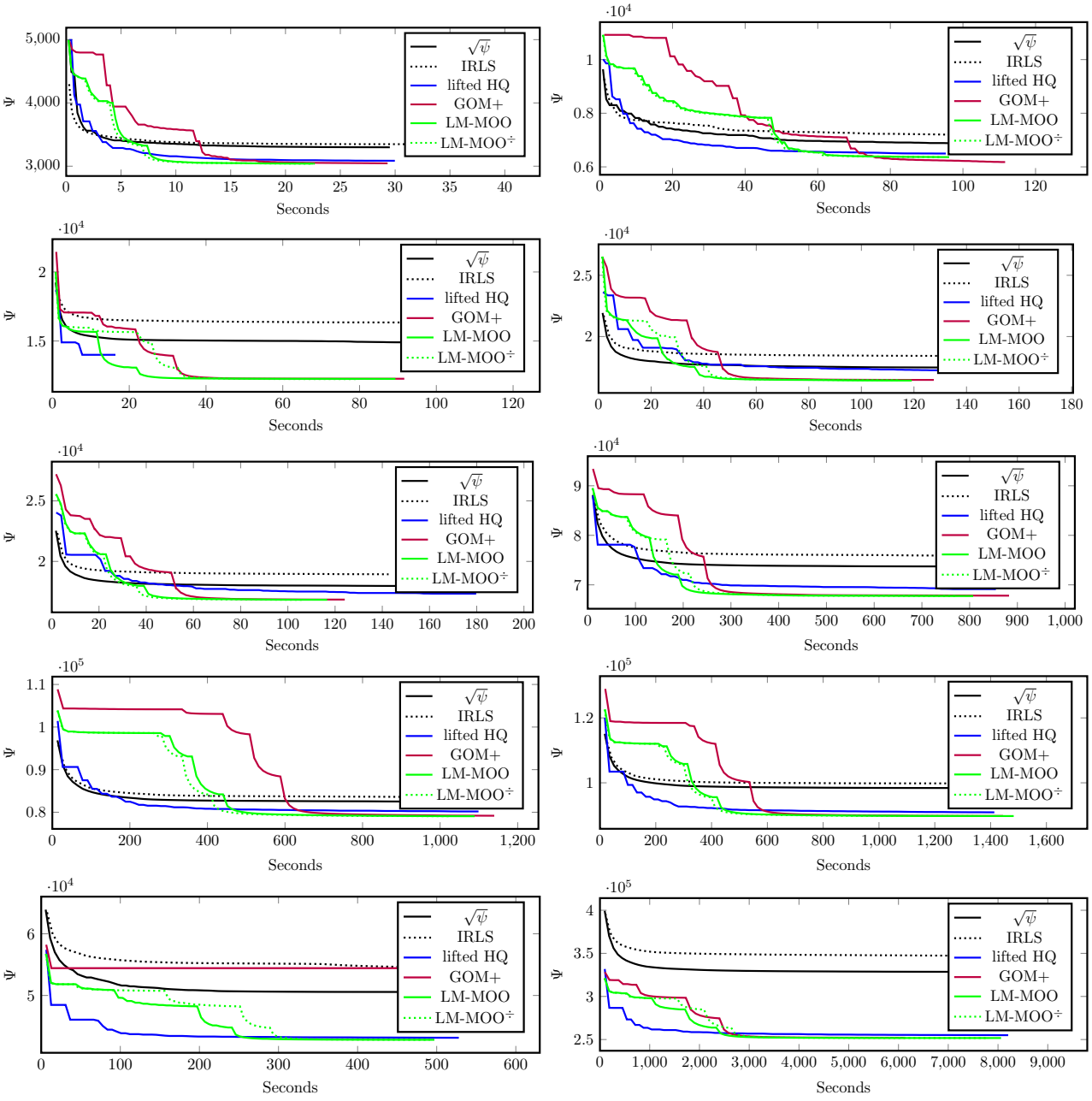

Figure 7: Evolution of the best cost for full bundle adjustment (datasets 11-20).