A novel estimation procedure for robust CANDECOMP/PARAFAC model fitting — Supplementary materials —

1. Simulation results for 100 × 20 × 20 tensors

1.1. Comparing the FIT diagnostic of R-INT2 and R-ALS.

	$F = R = 3$					
	BL points		GL points		Residual outl.	
	A	B	A	B	A	B
0%	1.4	53.5	1.3	55.4	1.2	53.7
10%	1.3	53.8	0.7	57.3	1.2	54.7
20%	1.1	54.0	0.5	62.1	1.2	56.2
	$F = R + 1 = 4$					
	BL points		GL points		Residual outl.	
	A	B	A	B	A	B
0%	35.0	50.2	31.8	49.9	30.8	50.4
10%	28.2	50.2	61.8	65.1	28.2	49.7
20%	27.0	49.6	86.2	79.6	24.8	50.4

Table 1: Compare the FIT diagnostic and verify that the variability explained by R-INT2 is not significantly different from that explained by R-ALS. **Column A:** Percent cases in which the difference of R-ALS FIT and R-INT2 FIT is higher than $1e^{-4}$. **Column B:** Percent cases in which the fit of R-INT2 is better than that of R-ALS which demonstrates that R-INT2 is capable of identifying the best low rank approximation as good as R-ALS.

1.2. Mean squared error (MSE) and subspaces B and C

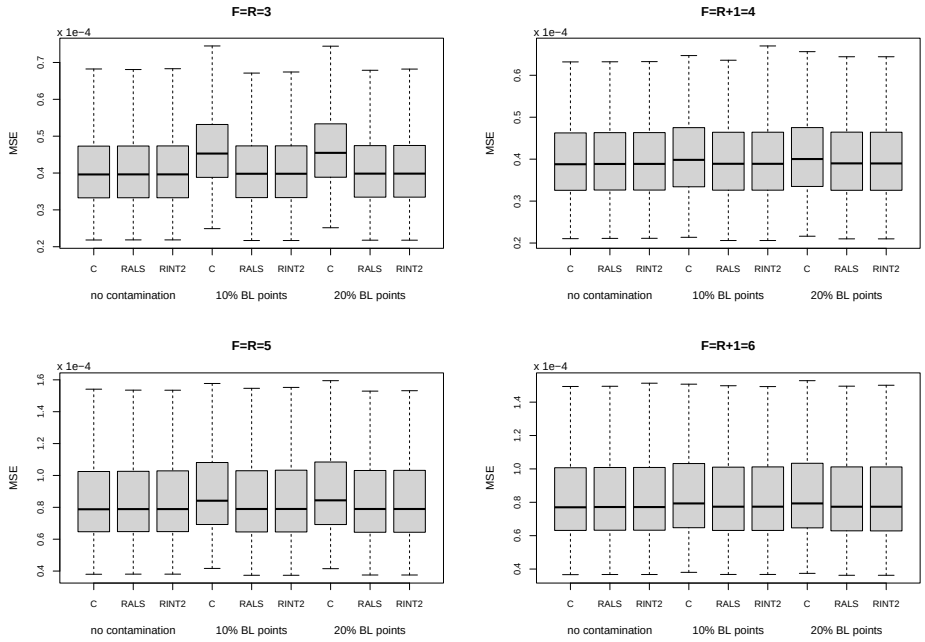

Figure 1: MSE values for classical CP (C), robust CP with ALS estimation (R-ALS) and robust CP with INT-2 estimation (R-INT2) on data sets without contamination, and with 10% and 20% **bad leverage (BL) points** respectively. The results are aggregated over all considered CONG and Noise levels.

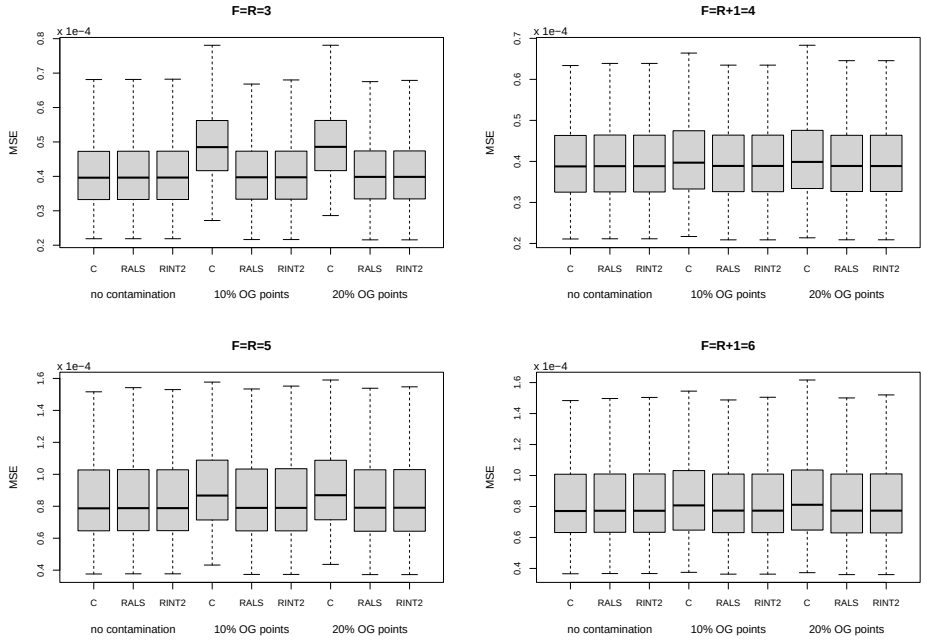

Figure 2: MSE values for classical CP (C), robust CP with ALS estimation (R-ALS) and robust CP with INT-2 estimation (R-INT2) on data sets without contamination, and with 10% and 20% **residual outliers** respectively. The results are aggregated over all considered CONG and Noise levels.

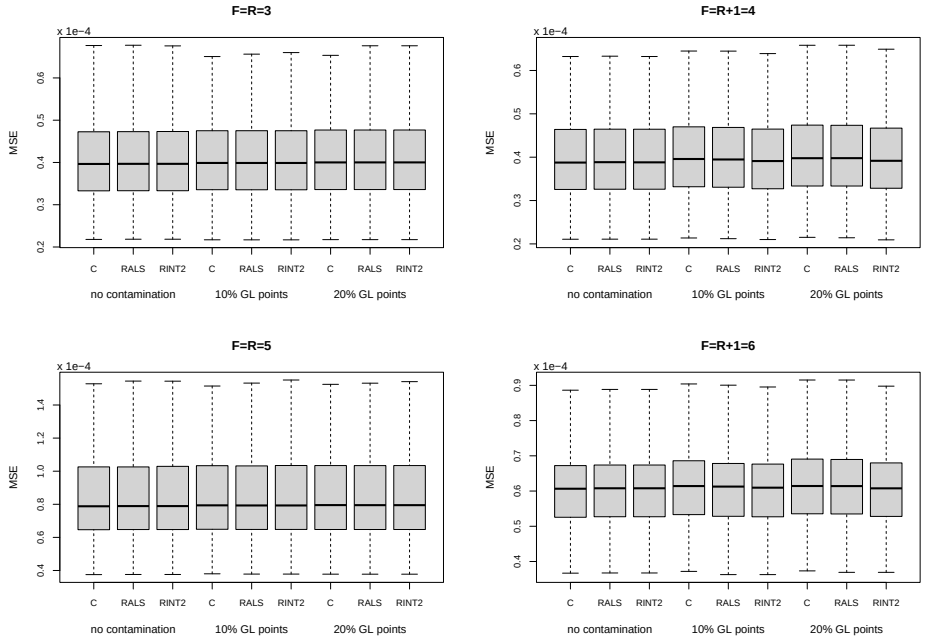

Figure 3: MSE values for classical CP (C), robust CP with ALS estimation (R-ALS) and robust CP with INT-2 estimation (R-INT2) on data sets without contamination, and with 10% and 20% **good leverage (GL) points** respectively. The results are aggregated over all considered CONG and Noise levels.

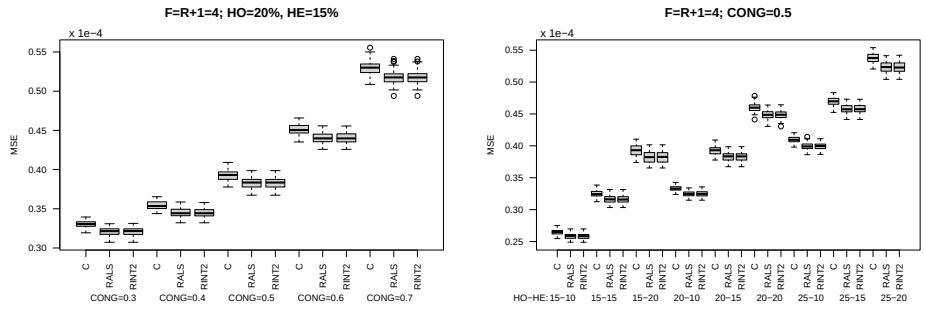

Figure 4: MSE values for classical CP (C), robust CP with ALS estimation (R-ALS) and robust CP with INT-2 estimation (R-INT2) on data sets with **20% bad leverage points** aggregated by CONG (left) and by Noise (right).

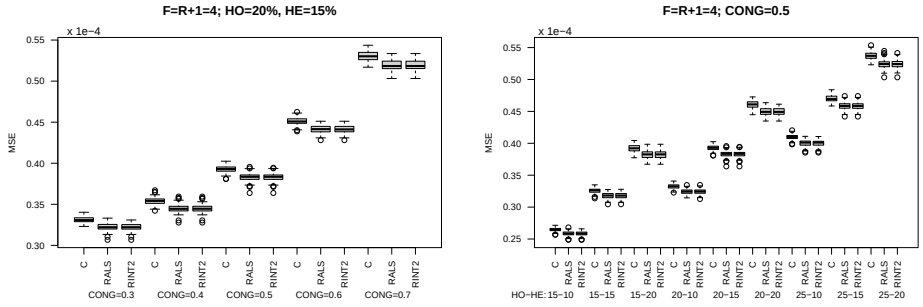

Figure 5: MSE values for classical CP (C), robust CP with ALS estimation (R-ALS) and robust CP with INT-2 estimation (R-INT2) on data sets with **10% bad leverage points** aggregated by CONG (left) and by Noise (right).

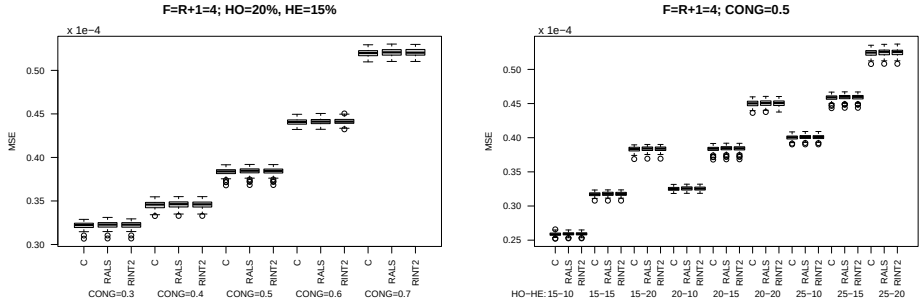

Figure 6: MSE values for classical CP (C), robust CP with ALS estimation (R-ALS) and robust CP with INT-2 estimation (R-INT2) on data sets with **no contamination** aggregated by CONG (left) and by Noise (right).

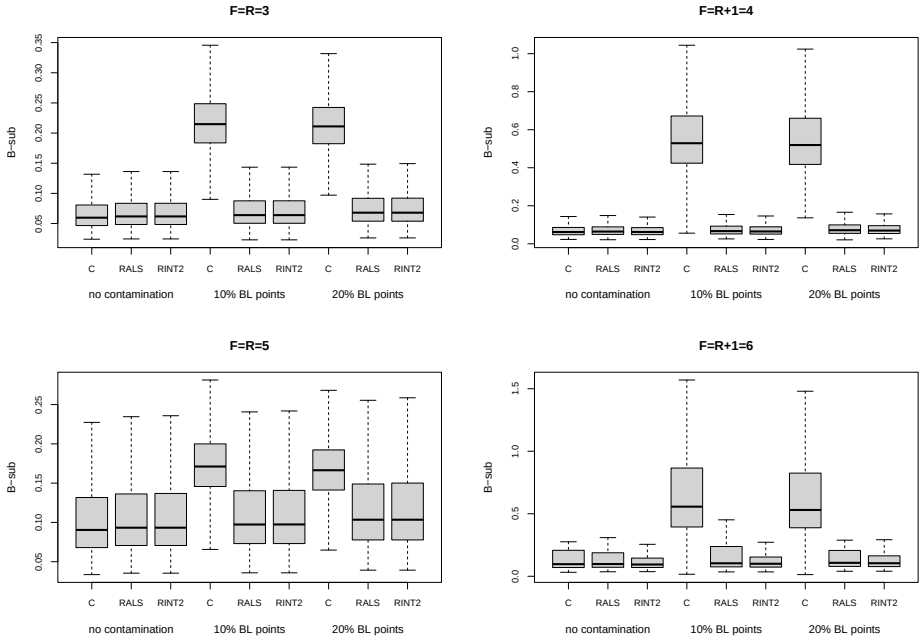

Figure 7: Angle of B-loadings of classical CP (C), robust CP with ALS estimation (R-ALS) and robust CP with INT-2 estimation (R-INT2) on data sets without contamination, and with 10% and 20% **bad leverage (BL)** points respectively. The results are aggregated over all considered CONG and Noise levels.

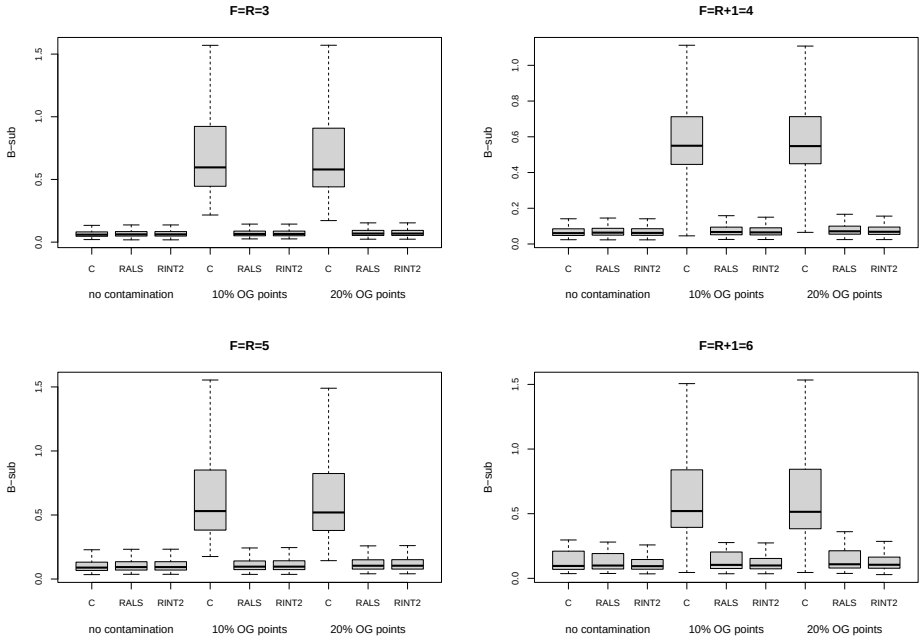

Figure 8: Angle of B-loadings of classical CP (C), robust CP with ALS estimation (R-ALS) and robust CP with INT-2 estimation (R-INT2) on data sets without contamination, and with 10% and 20% **residual outliers** points respectively. The results are aggregated over all considered CONG and Noise levels.

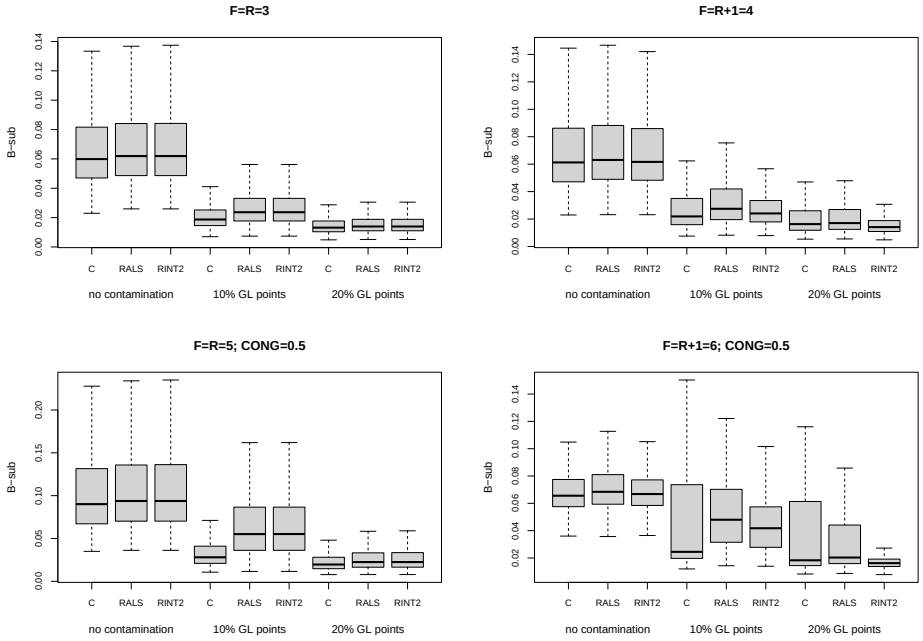

Figure 9: Angle of B-loadings of classical CP (C), robust CP with ALS estimation (R-ALS) and robust CP with INT-2 estimation (R-INT2) on data sets without contamination, and with 10% and 20% **good leverage (GL)** points respectively. The results are aggregated over all considered CONG and Noise levels.

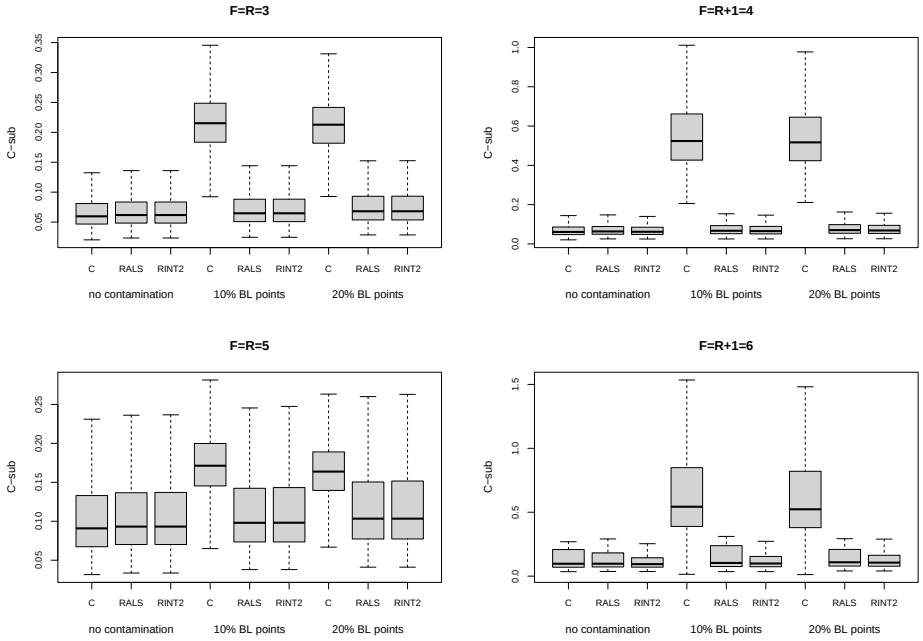

Figure 10: Angle of C-loadings of classical CP (C), robust CP with ALS estimation (R-ALS) and robust CP with INT-2 estimation (R-INT2) on data sets without contamination, and with 10% and 20% **bad leverage (BL)** points respectively. The results are aggregated over all considered CONG and Noise levels.

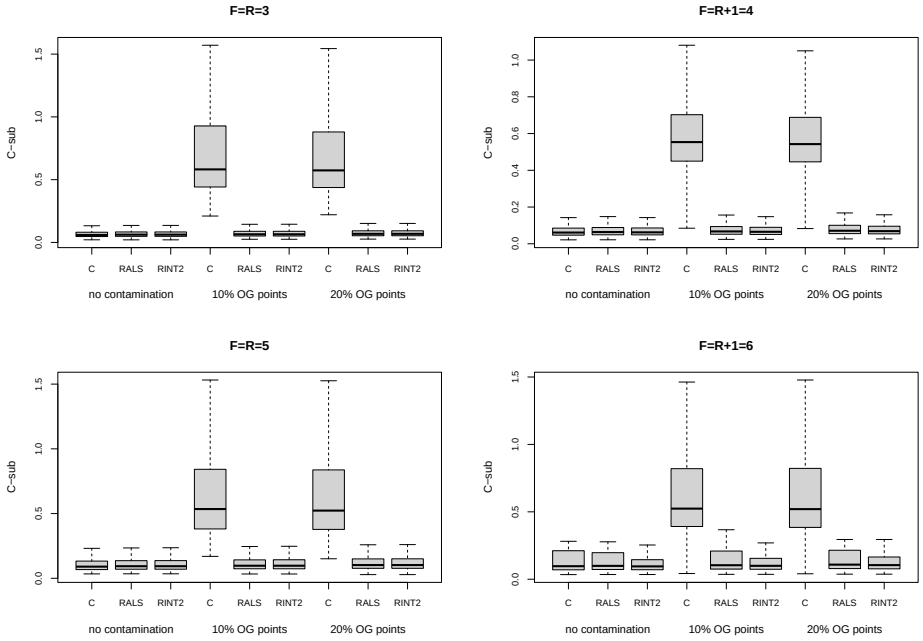

Figure 11: Angle of C-loadings of classical CP (C), robust CP with ALS estimation (R-ALS) and robust CP with INT-2 estimation (R-INT2) on data sets without contamination, and with 10% and 20% **residual outliers** points respectively. The results are aggregated over all considered CONG and Noise levels.

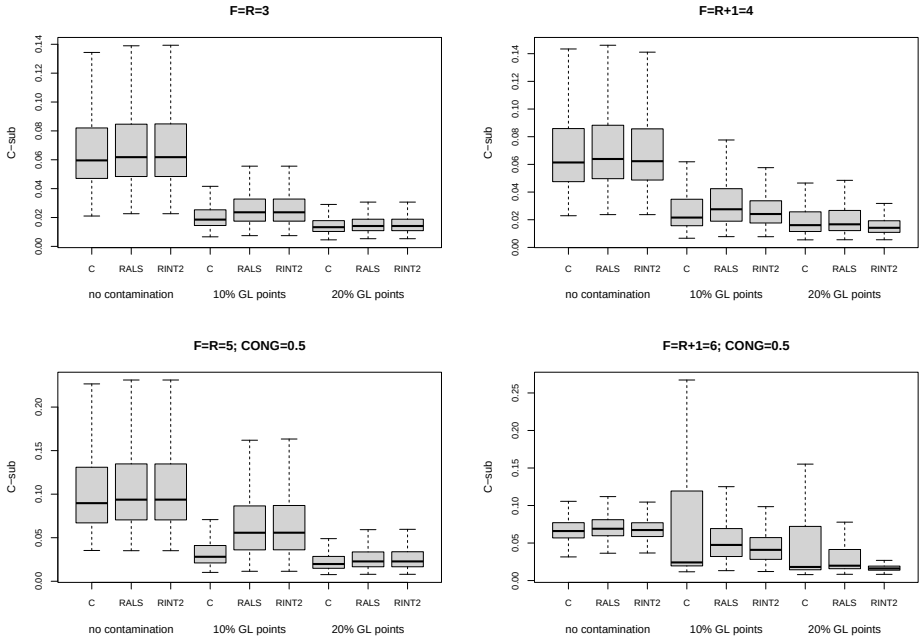

Figure 12: Angle of C-loadings of classical CP (C), robust CP with ALS estimation (R-ALS) and robust CP with INT-2 estimation (R-INT2) on data sets without contamination, and with 10% and 20% **good leverage (GL)** points respectively. The results are aggregated over all considered CONG and Noise levels.

1.3. Percentage of fault recoveries (FR) and number of swamps (SWAMPS)

		$F = R = 3$			$F = R + 1 = 4$		
		C	RALS	RINT2	C	RALS	RINT2
	0%	0.4	0.2	0.6	0.6	1.0	0.1
FR	10%	98.2	0.2	0.6	0.0	0.7	0.1
	20%	98.8	0.2	0.8	0.0	0.4	0.1
	0%	0	0	0	13	7	0
SWAMPS	10%	0	0	0	1	14	1
	20%	0	0	0	0	11	5
		$F = R = 5$			$F = R + 1 = 6$		
		C	RALS	RINT2	C	RALS	RINT2
	0%	4.3	4.1	4.8	1.4	1.4	1.6
FR	10%	95.1	4.6	5.7	1.6	1.7	1.8
	20%	95.0	5.0	6.6	0.1	2.2	2.5
	0%	1	2	3	8	10	3
SWAMPS	10%	0	2	0	4	10	7
	20%	0	2	2	4	14	14

Table 2: Total percentages of FR and number of swamps (out of 4500 repetitions) by rank and number of factors for different levels of contamination with **bad leverage points**.

		$F = R = 3$			$F = R + 1 = 4$		
		C	RALS	RINT2	C	RALS	RINT2
	0%	0.3	0.2	0.8	0.5	0.5	0.1
FR	10%	100.0	0.5	0.8	1.8	0.6	0.1
	20%	100.0	0.4	0.8	2.3	0.6	0.0
	0%	0	0	0	8	7	2
SWAMPS	10%	0	0	0	2	8	2
	20%	0	0	1	0	19	3
		$F = R = 5$			$F = R + 1 = 6$		
		C	RALS	RINT2	C	RALS	RINT2
	0%	4.2	4.3	4.6	1.1	1.4	1.6
FR	10%	100.0	4.9	5.6	27.5	1.9	2.0
	20%	100.0	5.0	6.4	29.5	2.1	2.5
	0%	1	3	1	18	5	16
SWAMPS	10%	0	1	0	7	17	17
	20%	0	1	2	4	14	16

Table 3: Total percentages of FR and number of swamps (out of 4500 repetitions) by rank and number of factors for different levels of contamination with **residual outliers**.

		$F = R = 3$			$F = R + 1 = 4$		
		C	RALS	RINT2	C	RALS	RINT2
	0%	0.4	0.4	0.7	0.4	0.6	0.0
FR	10%	0.0	0.1	0.1	52.8	36.4	0.1
	20%	0.0	0.1	0.2	75.0	70.3	0.2
	0%	0	0	0	11	9	1
SWAMPS	10%	0	0	0	95	90	1
	20%	0	0	0	56	65	1
		$F = R = 5$			$F = R + 1 = 6$		
		C	RALS	RINT2	C	RALS	RINT2
	0%	3.6	3.9	4.4	1.3	1.2	1.6
FR	10%	0.0	2.6	3.4	32.2	13.1	1.0
	20%	0.0	2.2	2.6	46.3	43.3	0.7
	0%	0	0	3	11	9	8
SWAMPS	10%	0	2	2	112	69	6
	20%	0	0	1	88	105	6

Table 4: Total percentages of FR and number of swamps (out of 4500 repetitions) by rank and number of factors for different levels of contamination with **good leverage points**.

1.4. CPU time and number of iterations

	$F = R = 3$			$F = R + 1 = 4$		
	C	RALS	RINT2	C	RALS	RINT2
0%	0.08	0.58	0.52	0.15	0.87	0.80
10%	0.21	0.57	0.52	0.77	0.86	0.78
20%	0.20	0.57	0.51	0.88	0.86	0.77
	$F = R = 5$			$F = R + 1 = 6$		
	C	RALS	RINT2	C	RALS	RINT2
0%	0.10	0.77	0.70	0.19	1.13	1.02
10%	0.27	0.75	0.68	0.90	1.12	1.00
20%	0.26	0.75	0.66	0.97	1.05	0.97

Table 5: Median CPU time in seconds (TIME), classical CP (C), robust CP with ALS estimation (R-ALS) and robust CP with INT-2 estimation (R-INT2) on data sets with different level of **bad leverage points** contamination: correct factor estimation (left) and over-factoring (right).

	$F = R = 3$			$F = R + 1 = 4$		
	C	RALS	RINT2	C	RALS	RINT2
0%	0.07	0.63	0.55	0.15	1.00	0.85
10%	0.08	0.63	0.54	0.22	0.98	0.83
20%	0.09	0.60	0.53	0.24	0.85	0.76
	$F = R = 5$			$F = R + 1 = 6$		
	C	RALS	RINT2	C	RALS	RINT2
0%	0.10	0.68	0.61	0.19	1.12	1.00
10%	0.11	0.67	0.60	0.27	1.09	0.97
20%	0.12	0.67	0.58	0.28	0.97	0.90

Table 6: Median CPU time in seconds (TIME), classical CP (C), robust CP with ALS estimation (R-ALS) and robust CP with INT-2 estimation (R-INT2) on data sets with different level of **residual outliers** contamination: correct factor estimation (left) and over-factoring (right).

	$F = R = 3$			$F = R + 1 = 4$		
	C	RALS	RINT2	C	RALS	RINT2
0%	0.08	0.62	0.54	0.15	0.94	0.81
10%	0.09	0.89	0.74	7.35	10.05	1.03
20%	0.09	0.96	0.80	12.88	33.79	1.13
	$F = R = 5$			$F = R + 1 = 6$		
	C	RALS	RINT2	C	RALS	RINT2
0%	0.11	0.70	0.64	0.20	1.14	1.02
10%	0.14	0.94	0.80	2.33	1.95	1.17
20%	0.13	1.27	1.01	5.15	26.54	1.50

Table 7: Median CPU time in seconds (TIME), classical CP (C), robust CP with ALS estimation (R-ALS) and robust CP with INT-2 estimation (R-INT2) on data sets with different level of **good leverage points** contamination: correct factor estimation (left) and over-factoring (right).

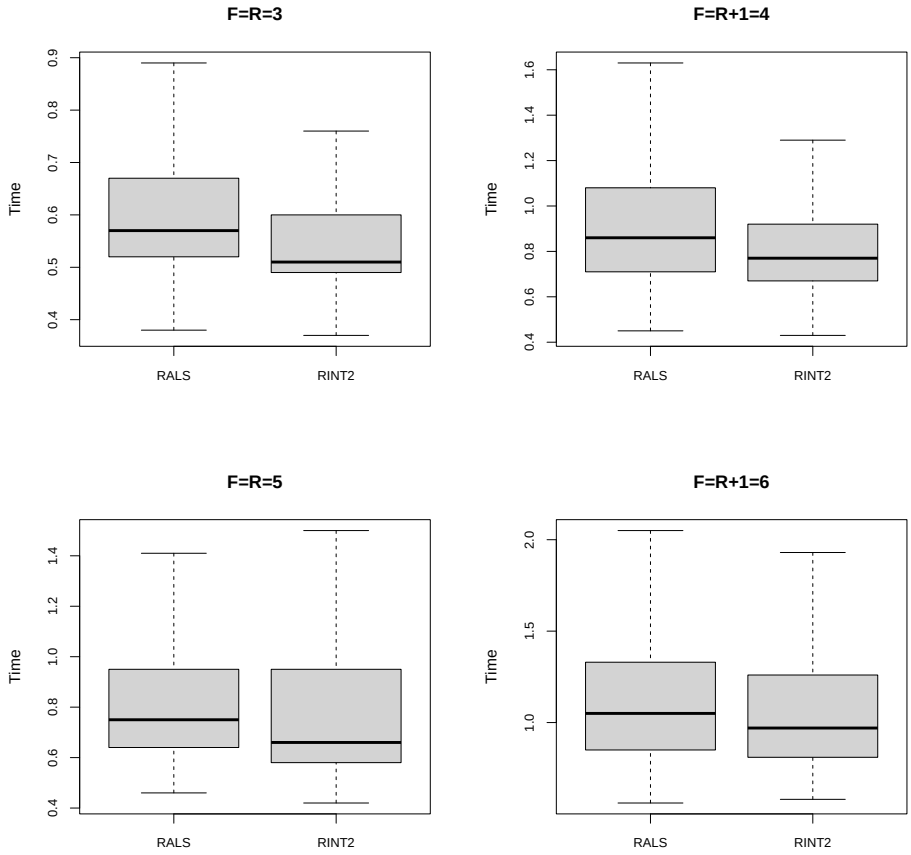

Figure 13: CPU time in seconds, robust CP with ALS estimation (R-ALS) and robust CP with INT-2 estimation (R-INT2) on data sets with 20% bad leverage points. The results are aggregated over all considered CONG and Noise levels.

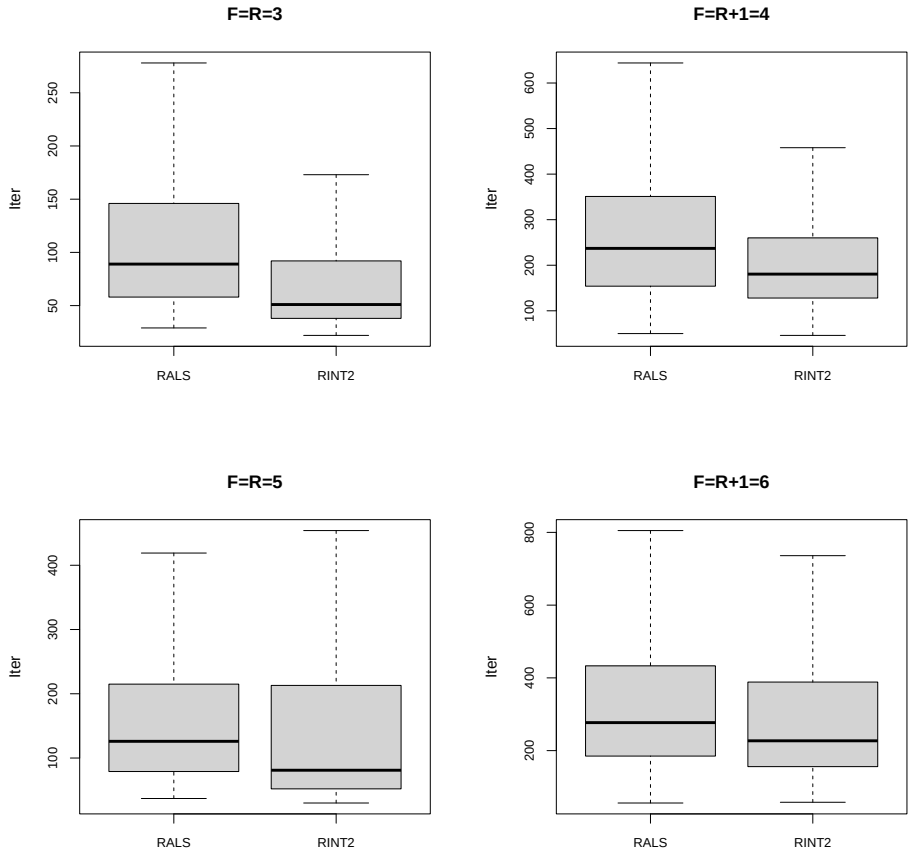

Figure 14: Number of iterations, robust CP with ALS estimation (R-ALS) and robust CP with INT-2 estimation (R-INT2) on data sets with 20% bad leverage points. The results are aggregated over all considered CONG and Noise levels.

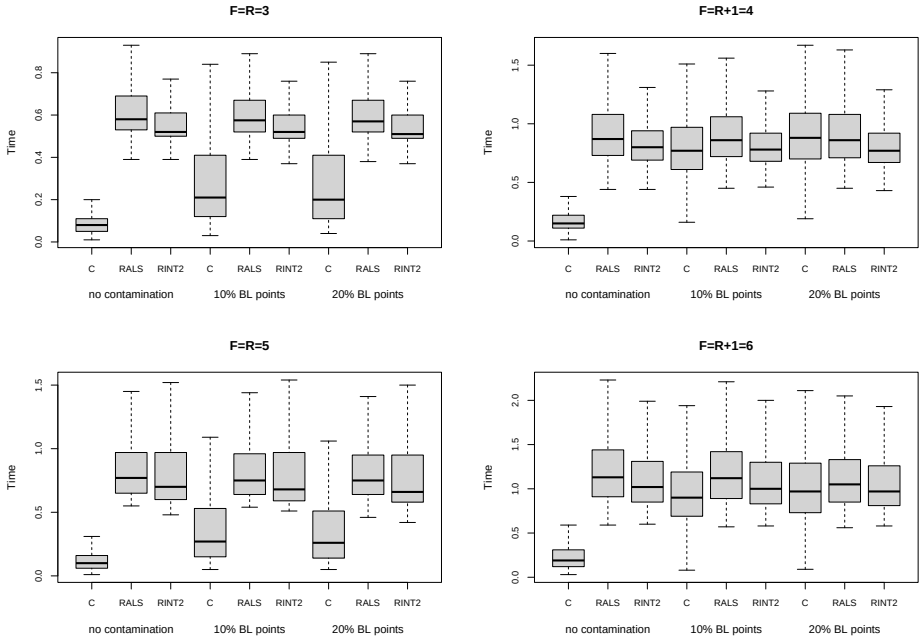

Figure 15: CPU time seconds, classical CP (C), robust CP with ALS estimation (R-ALS) and robust CP with INT-2 estimation (R-INT2) on data sets with different level of contamination with **bad leverage (BL) points**. The results are aggregated over all considered CONG and Noise levels.

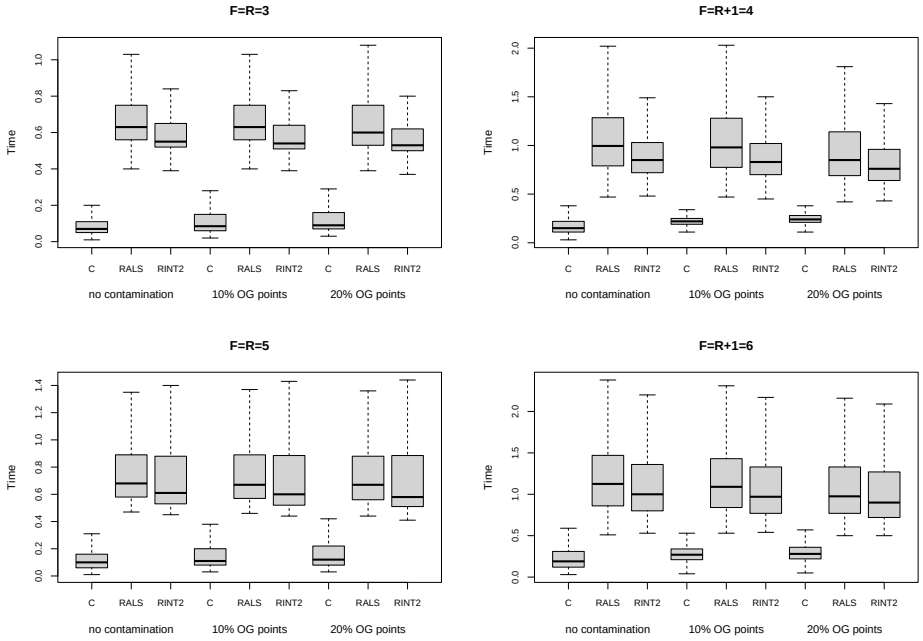

Figure 16: CPU time seconds, classical CP (C), robust CP with ALS estimation (R-ALS) and robust CP with INT-2 estimation (R-INT2) on data sets with different level of contamination with **residual outliers**. The results are aggregated over all considered CONG and Noise levels.

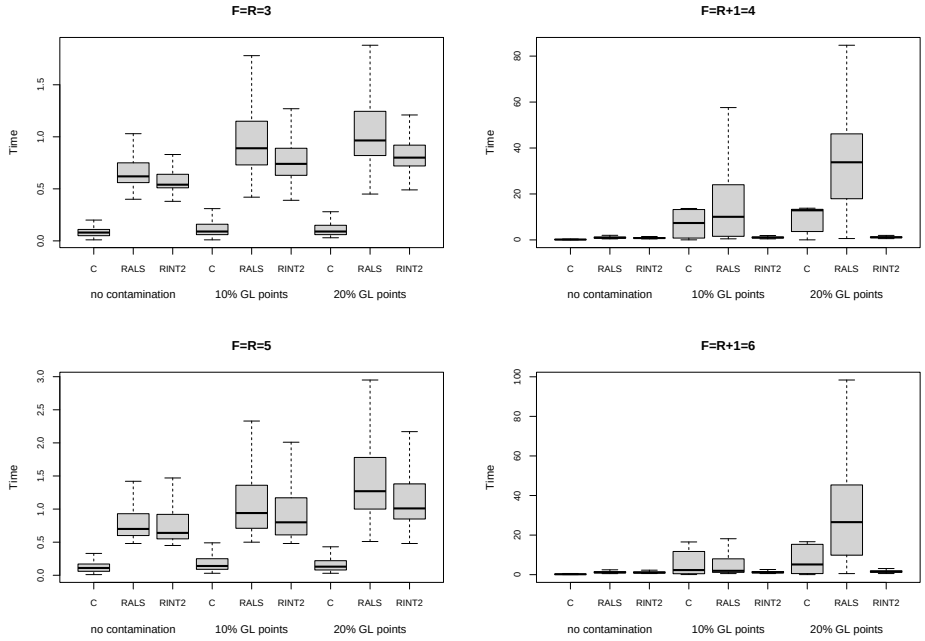

Figure 17: CPU time seconds, classical CP (C), robust CP with ALS estimation (R-ALS) and robust CP with INT-2 estimation (R-INT2) on data sets with different level of contamination with **good leverage (GL) points**. The results are aggregated over all considered CONG and Noise levels.

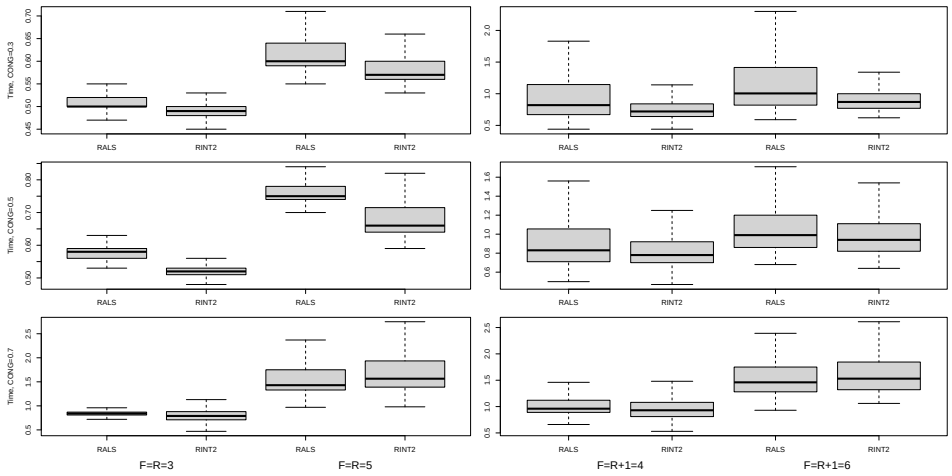

Figure 18: CPU time in seconds, robust CP with ALS estimation (R-ALS) and robust CP with INT-2 estimation (R-INT2) on data sets **without contamination**. The results are aggregated over all considered Noise levels and presented by selected levels of CONG (0.3, 0.5 and 0.7) for different ranks, with correct rank estimation and over-factoring.

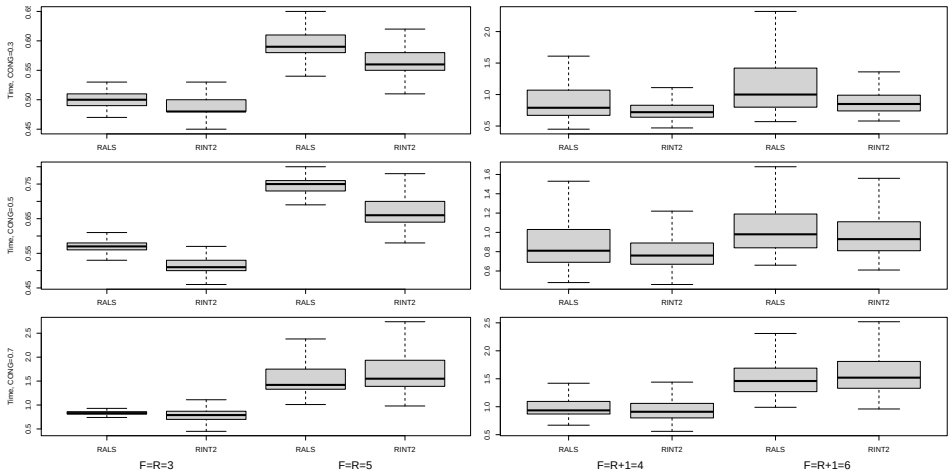

Figure 19: CPU time in seconds, robust CP with ALS estimation (R-ALS) and robust CP with INT-2 estimation (R-INT2) on data sets with **10% bad leverage points**. The results are aggregated over all considered Noise levels and presented by selected levels of CONG (0.3, 0.5 and 0.7) for different ranks, with correct rank estimation and over-factoring.

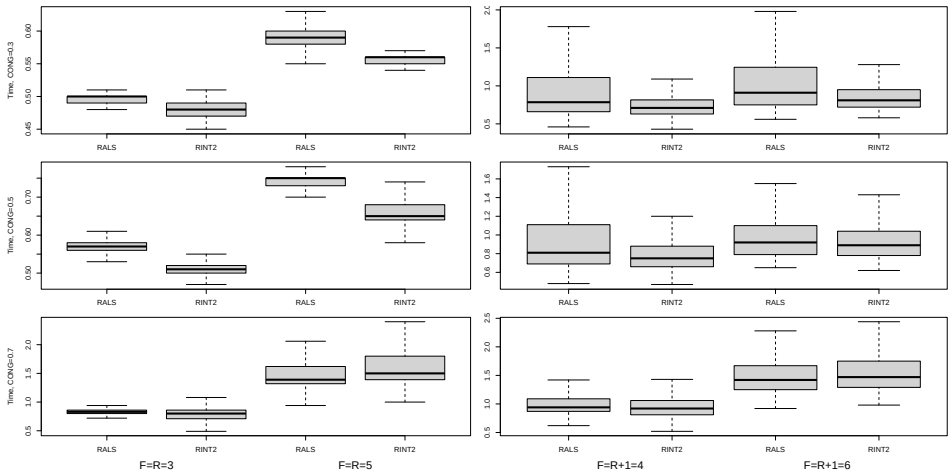

Figure 20: CPU time in seconds, robust CP with ALS estimation (R-ALS) and robust CP with INT-2 estimation (R-INT2) on data sets with **20% bad leverage points**. The results are aggregated over all considered Noise levels and presented by selected levels of CONG (0.3, 0.5 and 0.7) for different ranks, with correct rank estimation and over-factoring.

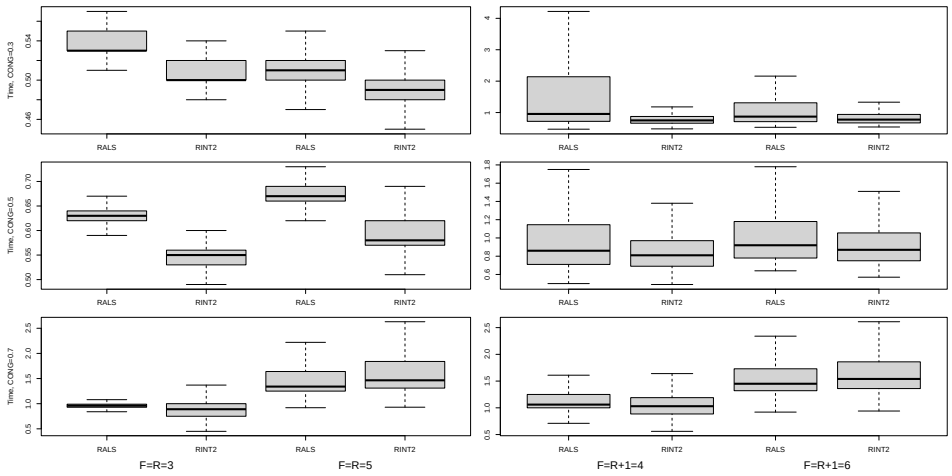

Figure 21: CPU time in seconds, robust CP with ALS estimation (R-ALS) and robust CP with INT-2 estimation (R-INT2) on data sets with **10% residual outliers**. The results are aggregated over all considered Noise levels and presented by selected levels of CONG (0.3, 0.5 and 0.7) for different ranks, with correct rank estimation and over-factoring.

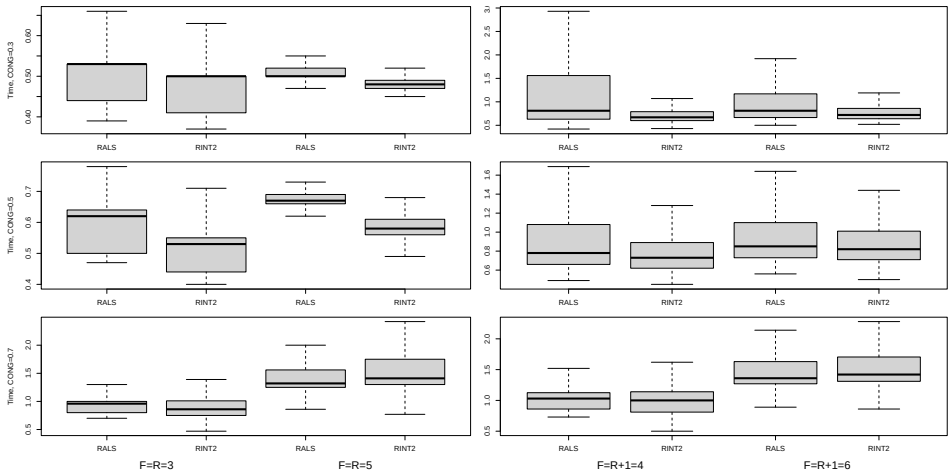

Figure 22: CPU time in seconds, robust CP with ALS estimation (R-ALS) and robust CP with INT-2 estimation (R-INT2) on data sets with **20% residual outliers**. The results are aggregated over all considered Noise levels and presented by selected levels of CONG (0.3, 0.5 and 0.7) for different ranks, with correct rank estimation and over-factoring.

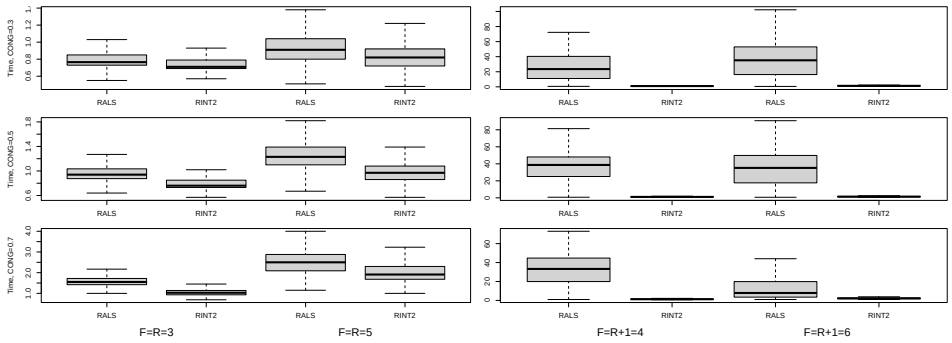

Figure 23: CPU time in seconds, robust CP with ALS estimation (R-ALS) and robust CP with INT-2 estimation (R-INT2) on data sets with **20% good leverage points**. The results are aggregated over all considered Noise levels and presented by selected levels of CONG (0.3, 0.5 and 0.7) for different ranks, with correct rank estimation and over-factoring.

2. Simulation results for larger tensors: $100 \times 100 \times 20$

2.1. $I = 100, J = 100, K = 20$*: Comparing the FIT diagnostic of R-INT2 and R-ALS.*

	$F = R = 3$		$F = R + 1 = 4$	
	A	B	A	B
0%	0.1	59.8	38.4	52.5
10%	0.0	60.2	24.8	53.9
20%	0.0	60.7	17.9	53.6

Table 8: Compare the FIT diagnostic and verify that the variability explained by R-INT2 is not significantly different from that explained by R-ALS. **Column A:** Percent cases in which the difference of R-ALS FIT and R-INT2 FIT is higher than $1e^{-4}$. **Column B:** Percent cases in which the fit of R-INT2 is better than that of R-ALS which demonstrates that R-INT2 is capable of identifying the best low rank approximation as good as R-ALS.

2.2. $I = 100, J = 100, k = 20$*: Mean squared error (MSE) and subspaces* B *and* C

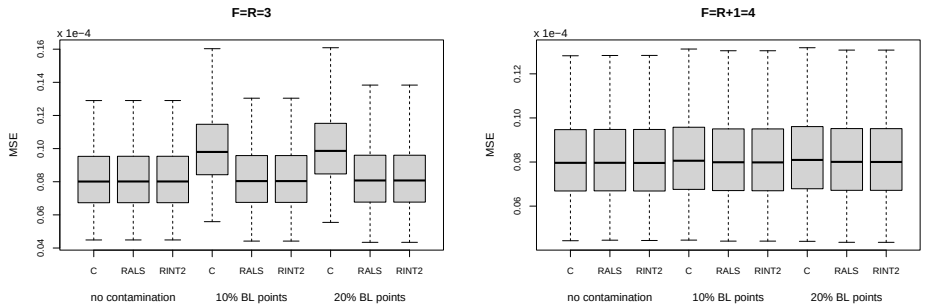

Figure 24: MSE values for classical CP (C), robust CP with ALS estimation (R-ALS) and robust CP with INT-2 estimation (R-INT2) on data sets without contamination, and with 10% and 20% **bad leverage (BL) points** respectively. The results are aggregated over all considered CONG and Noise levels.

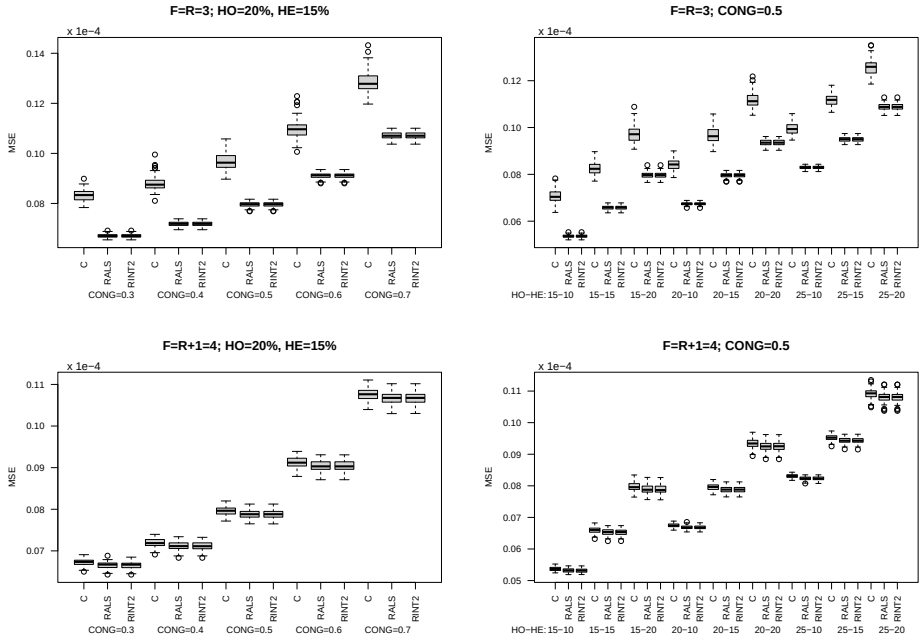

Figure 25: MSE values for classical CP (C), robust CP with ALS estimation (R-ALS) and robust CP with INT-2 estimation (R-INT2) on data sets with 20% bad leverage points aggregated by CONG (left) and by Noise (right).

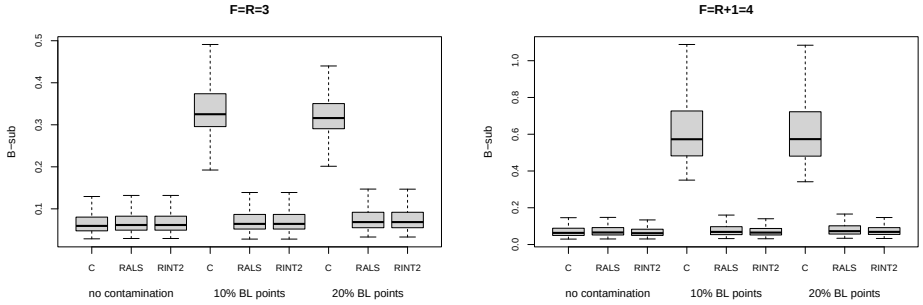

Figure 26: Angle of B-loadings of classical CP (C), robust CP with ALS estimation (R-ALS) and robust CP with INT-2 estimation (R-INT2) on data sets without contamination, and with 10% and 20% **bad leverage (BL)** points respectively. The results are aggregated over all considered CONG and Noise levels.

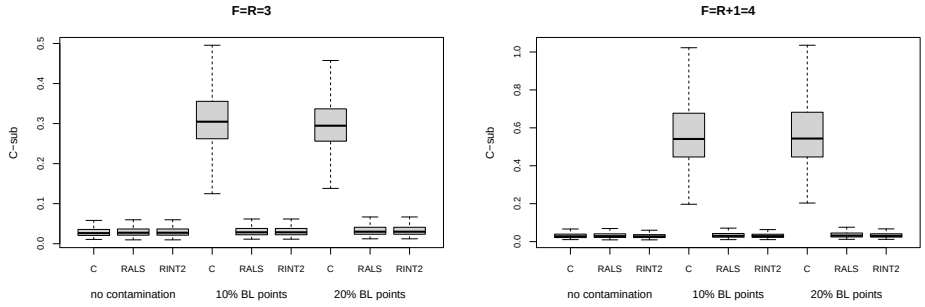

Figure 27: Angle of C-loadings of classical CP (C), robust CP with ALS estimation (R-ALS) and robust CP with INT-2 estimation (R-INT2) on data sets without contamination, and with 10% and 20% **bad leverage (BL)** points respectively. The results are aggregated over all considered CONG and Noise levels.

2.3. $I = 100, J = 100, K = 20$*: Percentage of fault recoveries (FR) and number of swamps (SWAMPS)*

		$F = R = 3$			$F = R + 1 = 4$		
		C	RALS	RINT2	C	RALS	RINT2
	0%	0.0	0.0	0.1	6.4	8.0	0.2
FR	10%	100.0	0.0	0.0	0.0	7.3	0.1
	20%	100.0	0.0	0.0	0.0	7.1	0.0
	0%	0	0	0	17	30	2
SWAMPS	10%	0	0	0	0	30	2
	20%	0	0	0	0	34	1

Table 9: Total percentages of FR and number of swamps (out of 4500 repetitions) by rank and number of factors for different levels of contamination with **bad leverage points**.

2.4. $I = 100, J = 100, K = 20$*: CPU time and number of iterations*

	$F = R = 3$			$F = R + 1 = 4$		
	C	RALS	RINT2	C	RALS	RINT2
0%	0.39	1.67	1.06	0.99	4.94	3.08
10%	1.33	1.65	1.03	3.22	4.69	2.89
20%	1.31	1.61	1.00	3.70	4.06	2.63

Table 10: Median CPU time in seconds (TIME), classical CP (C), robust CP with ALS estimation (R-ALS) and robust CP with INT-2 estimation (R-INT2) on data sets with different level of **bad leverage points** contamination: correct factor estimation (left) and over-factoring (right).

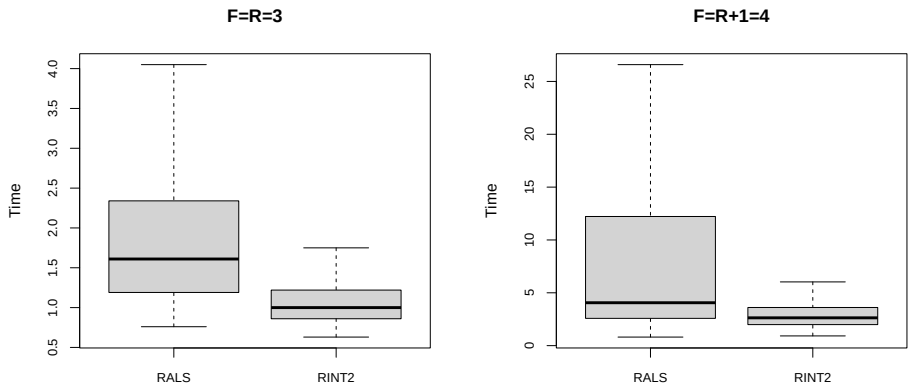

Figure 28: CPU time in seconds, robust CP with ALS estimation (R-ALS) and robust CP with INT-2 estimation (R-INT2) on data sets with 20% bad leverage points. The results are aggregated over all considered CONG and Noise levels.

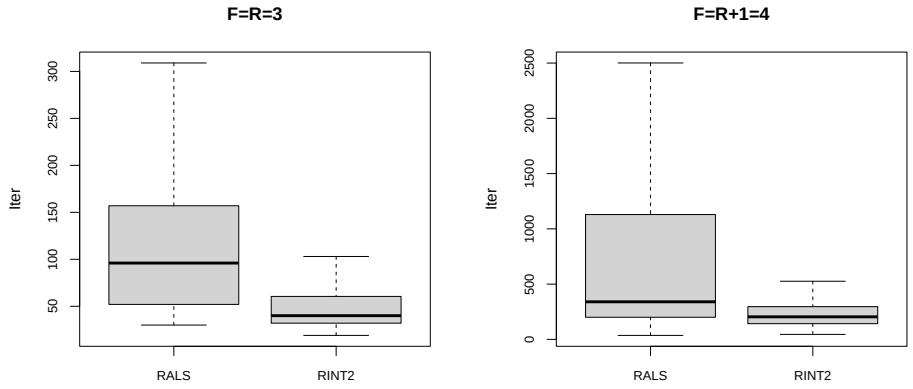

Figure 29: Number of iterations, robust CP with ALS estimation (R-ALS) and robust CP with INT-2 estimation (R-INT2) on data sets with 20% bad leverage points. The results are aggregated over all considered CONG and Noise levels.

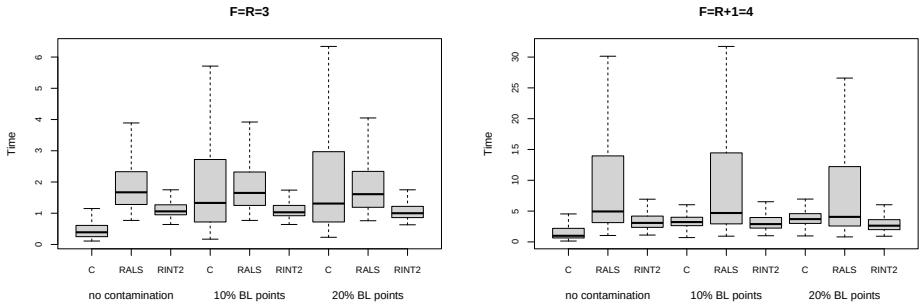

Figure 30: CPU time seconds, classical CP (C), robust CP with ALS estimation (R-ALS) and robust CP with INT-2 estimation (R-INT2) on data sets with different level of contamination with **bad leverage (BL) points**. The results are aggregated over all considered CONG and Noise levels.