Release Note of LIBLINEAR 2.40

Leonardo Galli* Chih-Jen Lin†

September 11, 2020

1 Introduction

Following the work in Galli and Lin (2020), in version 2.40 we update the Newton-based solver in LIBLINEAR (Fan et al., 2008). In this release note, through several aspects we compare the performance between the existing implementation and some new settings.

Currently in LIBLINEAR, the truncated Newton method considers a trust region setting and applies the following inner stopping condition in the CG procedure (Lin et al., 2008; Hsia et al., 2018)

$$\frac{\|\boldsymbol{g}_k + H_k \boldsymbol{s}_k^j\|_{M^{-1}}}{\|\boldsymbol{g}_k\|_{M^{-1}}} \leq \epsilon_{\text{CG}}, \tag{1}$$

where $\epsilon_{\text{CG}} = 0.1$, j is the CG-step index, M is the current preconditioner and $\boldsymbol{s}_k^j$ is the current iterate of the CG procedure. In Galli and Lin (2020), the authors consider

- a line search instead of a trust region setting,

and found that the following "quadratic" inner stopping condition works well.

$$\frac{(Q_j - Q_{j-1})}{Q_j / j} \leq \min\{\epsilon_{\text{CG}}; \sqrt{\|\boldsymbol{g}_k\|_{M^{-1}}}\}, \tag{2}$$

where $Q_j := Q(\boldsymbol{s}_k^j)$, $Q_{j-1} := Q(\boldsymbol{s}_k^{j-1})$, and

$$Q(\boldsymbol{s}) \equiv \nabla f(\boldsymbol{w}_k)^T \boldsymbol{s} + \frac{1}{2} \boldsymbol{s}^T \nabla^2 f(\boldsymbol{w}_k) \boldsymbol{s}. \tag{3}$$

In this note we focus on comparing the existing implementation with the following new settings:

- The same trust region framework of LIBLINEAR is used, but the CG stopping tolerance ϵ_{CG} is slightly tuned;
- The same trust region framework of LIBLINEAR is used, but the CG stopping criterion is replaced with (2);
- A line search framework considered in Galli and Lin (2020) is used and the CG stopping criterion is (2).

*Department of Information Engineering, University of Florence, Via Santa Marta 3, Firenze, Italia e-mail: leonardo.galli@unifi.it.

†Department of Computer Science, National Taiwan University, Taipei, Taiwan, 106 e-mail: cjlin@csie.ntu.edu.tw.

2 Comparison of Convergence Speed

In Figures 1–4, we compare the following approaches:

- quadada_p: line search and preconditioning are applied. The stopping condition is (2) with $\epsilon_{\text{CG}} = 0.5$;
- tr_resadaZtr_p: trust region and preconditioning are applied. The inner stopping condition is

$$\frac{\|\boldsymbol{g}_k + H_k \boldsymbol{s}_k^j\|_{M^{-1}}}{\|\boldsymbol{g}_k\|_{M^{-1}}} \leq \min\{\epsilon_{\text{CG}}; \sqrt{\|\boldsymbol{g}_k\|_{M^{-1}}}\},$$

  where $\epsilon_{\text{CG}} = 0.5$;
- tr_resadaZtr01_p: the same as tr_resadaZtr_p, but $\epsilon_{\text{CG}} = 0.1$;
- tr_rescons_p: this is the one implemented in LIBLINEAR now; see (1).

Note that tr_resadaZtr_p and tr_resadaZtr01_p differ only in without/with the adaptive setting:

$$\epsilon_{\text{CG}} \text{ and } \min\{\epsilon_{\text{CG}}; \sqrt{\|\boldsymbol{g}_k\|_{M^{-1}}}\}. \tag{4}$$

The aim of the comparison is to check the overall convergence speed. In each figure, the y-axis is

$$\frac{f(\boldsymbol{w}_k) - f(\boldsymbol{w}^*)}{f(\boldsymbol{w}^*)},$$

where $\boldsymbol{w}^*$ is an approximate optimal solution obtained by running enough iterations, and the x-axis is the cumulative number of CG steps. The four horizontal lines in each figure indicate places where the following stopping condition is met respectively with $\epsilon = \{10^{-1}, 10^{-2}, 10^{-3}, 10^{-4}\}$

$$\|\boldsymbol{g}_k\| \leq \epsilon \frac{\min\{\#\text{pos}, \#\text{neg}\}}{l} \cdot \|\nabla f(\mathbf{0})\|, \tag{5}$$

where #pos, #neg are the numbers of positive- and negative-labeled instances respectively, and $\epsilon > 0$ is the thresholding constant that controls the precision of the training procedure. Note that (5) with $\epsilon = 10^{-2}$ is the outer stopping condition employed in LIBLINEAR. Thus in general the behavior around $\epsilon = 10^{-2}$ and 10^{-3} is the most crucial.

We have the following observations:

- From tr_rescons_p and tr_resadaZtr01_p, the use of $\sqrt{\|\boldsymbol{g}_k\|_{M^{-1}}}$ in (4) is not effective.
- From tr_resadaZtr_p and tr_resadaZtr01_p, where the difference is only on ϵ_{CG}, we see that the convergence is affected by this parameter. Usually $\epsilon_{\text{CG}} = 0.1$ is too tight, but sometimes the opposite situation occurs. For example, $\epsilon_{\text{CG}} = 0.5$ is worse for kddb in Figure 3 and kdd12-svm in Figure 4.
- For trust region, sometimes it needs many iterations because of failed directions (i.e., the function value does not decrease and trust region must be adjusted).
- For L2-loss SVM, the final convergence of using trust region is slower. The reason might be that L2-loss SVM is not twice differentiable. Thus the estimation of actual versus predicted function-value reduction is not accurate. However, this issue may not be serious in practice because users do not need to strictly solve the optimization problem and encounter the slow convergence.
- The approach quadada_p, while not the best in all situations, is generally superior and robust

To further investigate trust region and line search, in Figures 5–8, we compare the following settings:

Table 1: Data sets and $\log_2(C_{\text{best}})$

Data set	LR	L2 SVM
astro-ph_62369.dat	3	-1
australian	-15	-18
breast-cancer	-57	-57
cod-rna	-13	-17
covtype.libsvm.binary	-20	-24
gisette_scale	-3	-10
kdda	-3	-7
kddb	-1	-4
leisure.scale	6	1
news20.binary	9	5
rcv1_test.binary	3	-1
real-sim	5	-1
train308.scale	2	-1
url_combined	-9	-9

- quadada_p: the same line search implementation considered earlier;
- tr_quadada_p: the same as quadada_p but we use trust region.

We have the following observations:

- Focusing on LR-loss in Figures 5 and 6 we can see that the behavior of the two settings is pretty similar. The setting tr_quadada_p is better on yahookr, while quadada_p is better on kdda, but a part from these detachments, the two curves are almost everywhere overlapping.

  Earlier in analyzing Figures 1-4 we mentioned that the trust region setting may need more iterations because of failed iterations. This is observed on problems yahookr, kdda, and kddb in Figure 6. For highly sparse data, more outer iterations imply higher cost on function/gradient evaluation and may cause longer running time For example, for yahookr (leisure_scale) in Table **??**, while quadada_p needs more CG steps, with fewer outer iterations, its total running time is competitive. On the other hand, a trust-region setting may take fewer CG steps per iteration because of the trust-region constraint. In the practical use the code is stopped after outer iterations reach a stopping condition. Thus the line-search setting may take too many CG steps in the last iteration and over-solve the problem; see the result of kddb in Table 5.

- Focusing instead on L2-loss in Figures 7 and 8 we can see a remarkable difference between the two settings. In fact, especially with $C = 100C_{\text{best}}$, quadada_p is faster than tr_quadada_p in the final part of the optimization procedure. When this detachment starts after the third horizontal line, as on rcv1 and real-sim, this issue is not very troublesome. It is instead problematic when this different behavior starts around the second horizontal line, as it happens on kdda, kddb and kdd12-svm. As conjectured above, a possible hypothesis is the lack of accordance between the predicted and the actual reduction in non-twice differentiable function.

3 Running Time Experiments

Figures in Section 2 show the overall convergence of each setting, but in practice a user specifies a stopping tolerance and the program terminates after reaching the stopping condition in (5). To see the running time in the practical use, in Tables 2–9 we conduct a comparison by using two tolerances

$$\epsilon = 0.01 \text{ and } 0.001.$$

In this and the next sections, we consider a different set of problems. For each problem, we search for the regularization C_{best} that achieves the best cross validation accuracy. Details are given in Table 1. Note that for the same problem the C_{best} value used here may be different from that in Section 2. In Section 2, C_{best} values follow from Hsia et al. (2017), which are obtained by an earlier version of LIBLINEAR, but here we obtain them using LIBLINEAR 2.30.

For the experimental environment, we use a laptop equipped with an Intel i7-10510U CPU. Note that we omit "_p" in the name of each approach as by default preconditioning is always applied.
We have the following observations:

- In general, quadada and tr_quadada are better than tr_rescons;
- quadada is the best for most cases. Even if it is slower, it reaches a smaller function value (e.g., kddb in Table 5).

4 Warm-start for Parameter Selection

The Newton method is the back-end of the parameter selection procedure in LIBLINEAR (Chu et al., 2015). The search procedure involves sweeping the following sequence of parameters

$$C_{\min}, \Delta C_{\min}, \Delta^2 C_{\min}, \ldots, C_{\max}, \tag{6}$$

where Δ is a given factor. Currently LIBLINEAR adjust the ϵ_{CG} value in the following way

$$\epsilon_{\text{CG}} = \begin{cases} 0.1 & \text{for the first } C, \\ 0.5 & \text{for subsequent } C \text{ values.} \end{cases} \tag{7}$$

The rationale of (7) is that if warm start is applied, the initial $\boldsymbol{w}$ at ΔC may be close to optimal because it is an approximate solution at C. Thus an $\epsilon_{\text{CG}} = 0.1$ may be too strict; see Section III.1.4 of the supplementary materials of Chu et al. (2015). Now for the new quadratic inner CG stopping condition, we hope that the rule (7) may not be necessary. To this end, we compare the following settings:

- tr_rescons: the one implemented in LIBLINEAR, where the rule (7) is applied;
- tr_rescons_fixed: we use tr_rescons for each individual problem but ϵ_{CG} is always kept to be 0.1;
- quadada_fixed: we use quadada (with line search) for each individual problem, and ϵ_{CG} is always kept to be 0.5.

Results are in Tables 10–11 and we have the following observations:

- The approach tr_rescons_fixed does not take a lot more time that tr_rescons. This observation is different from that in Section III.1.4 of the supplementary materials of Chu et al. (2015). One possible reason might be that significant improvements on our Newton method were made after the work by Chu et al. (2015); see, for example, Hsia et al. (2017, 2018).
- The approach tr_rescons_fixed usually takes more total # CG steps because of a smaller ϵ_{CG}. However, the number of total iterations is smaller.
- The approach quadada_fixed is in general competitive in terms of the total running time.

5 Conclusions

Based on experimental results, we decide to implement quadada in LIBLINEAR 2.40. An important change is that we move from trust region to line search. The adaptive setting in (2) is not very effective in practice because in almost all cases

$$\epsilon_{\text{CG}} < \sqrt{\|\boldsymbol{g}_k\|_{M^{-1}}}.$$

However, we keep the term $\sqrt{\|\boldsymbol{g}_k\|_{M^{-1}}}$ because with the adaptive setting, the theoretical local Q-super-linear convergence can be obtained (Galli and Lin, 2020).

References

B.-Y. Chu, C.-H. Ho, C.-H. Tsai, C.-Y. Lin, and C.-J. Lin. Warm start for parameter selection of linear classifiers. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*, 2015. URL `http://www.csie.ntu.edu.tw/~cjlin/libsvmtools/warm-start/warm-start.pdf`.

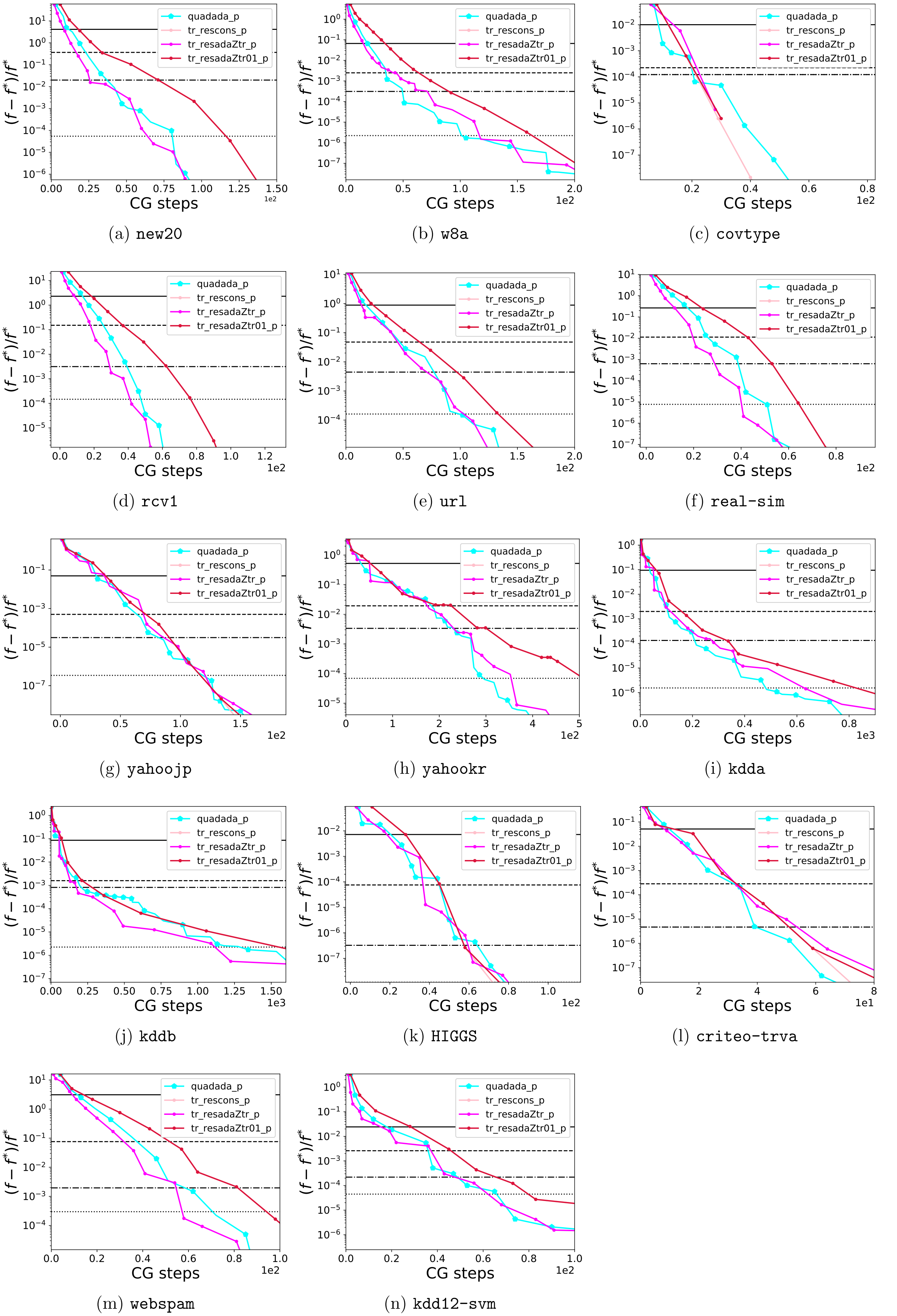
R.-E. Fan, K.-W. Chang, C.-J. Hsieh, X.-R. Wang, and C.-J. Lin. LIBLINEAR: a library for large linear classification. *Journal of Machine Learning Research*, 9:1871–1874, 2008. URL `http://www.csie.ntu.edu.tw/~cjlin/papers/liblinear.pdf`.

L. Galli and C.-J. Lin. Truncated Newton methods for linear classification. Technical report, National Taiwan University, 2020. URL `https://www.csie.ntu.edu.tw/~cjlin/papers/tncg/tncg.pdf`.

C.-Y. Hsia, Y. Zhu, and C.-J. Lin. A study on trust region update rules in Newton methods for large-scale linear classification. In *Proceedings of the Asian Conference on Machine Learning (ACML)*, 2017. URL `http://www.csie.ntu.edu.tw/~cjlin/papers/newtron/newtron.pdf`.

C.-Y. Hsia, W.-L. Chiang, and C.-J. Lin. Preconditioned conjugate gradient methods in truncated Newton frameworks for large-scale linear classification. In *Proceedings of the Asian Conference on Machine Learning (ACML)*, 2018. URL `http://www.csie.ntu.edu.tw/~cjlin/papers/tron_pcg/precondition.pdf`.

C.-J. Lin, R. C. Weng, and S. S. Keerthi. Trust region Newton method for large-scale logistic regression. *Journal of Machine Learning Research*, 9:627–650, 2008. URL `http://www.csie.ntu.edu.tw/~cjlin/papers/logistic.pdf`.



(a) `new20` (b) `w8a` (c) `covtype` (d) `rcv1` (e) `url` (f) `real-sim` (g) `yahoojp` (h) `yahookr` (i) `kdda` (j) `kddb` (k) `HIGGS` (l) `criteo-trva` (m) `webspam` (n) `kdd12-svm`

Figure 1: Loss=LR and $C = C_{\text{best}}$

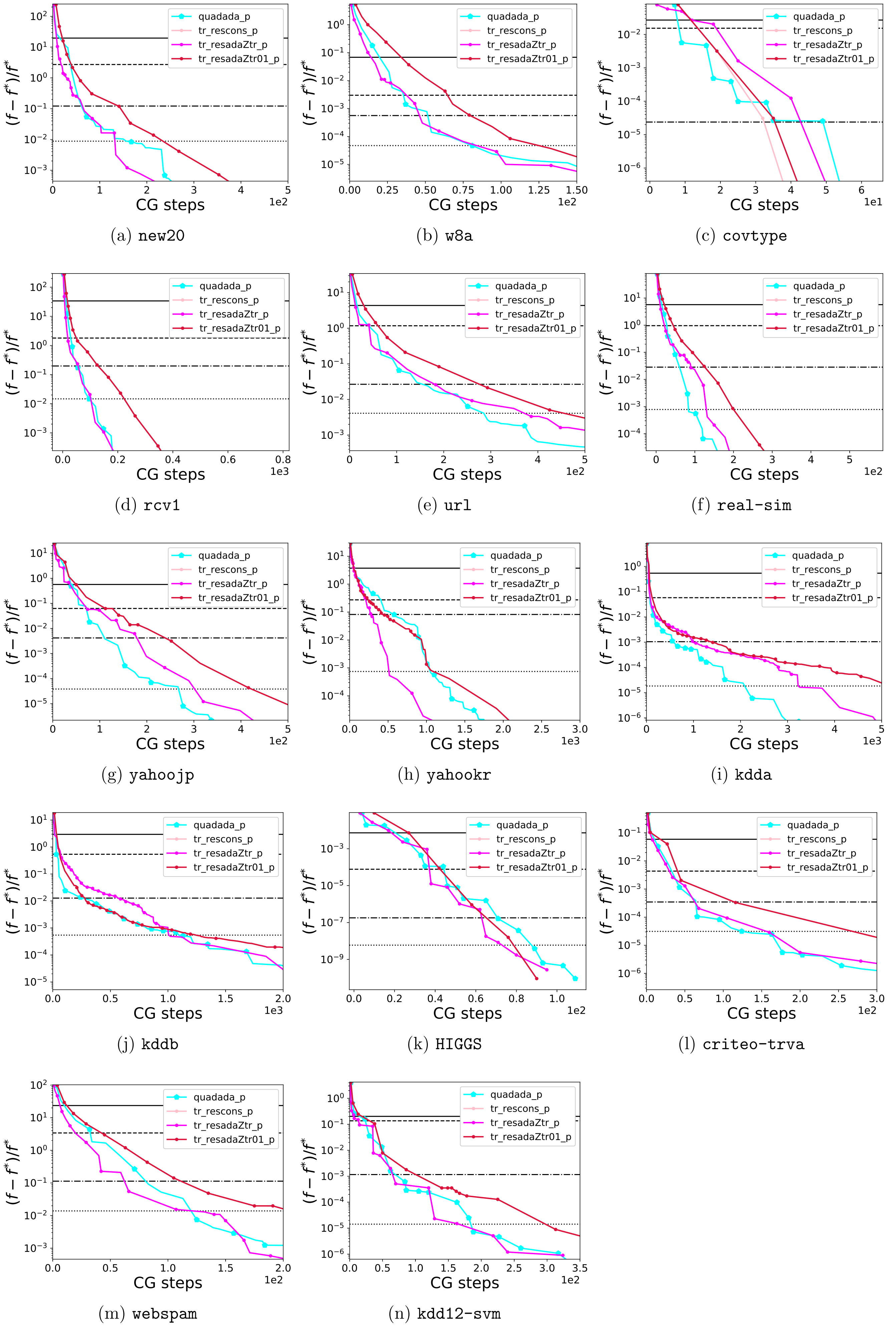
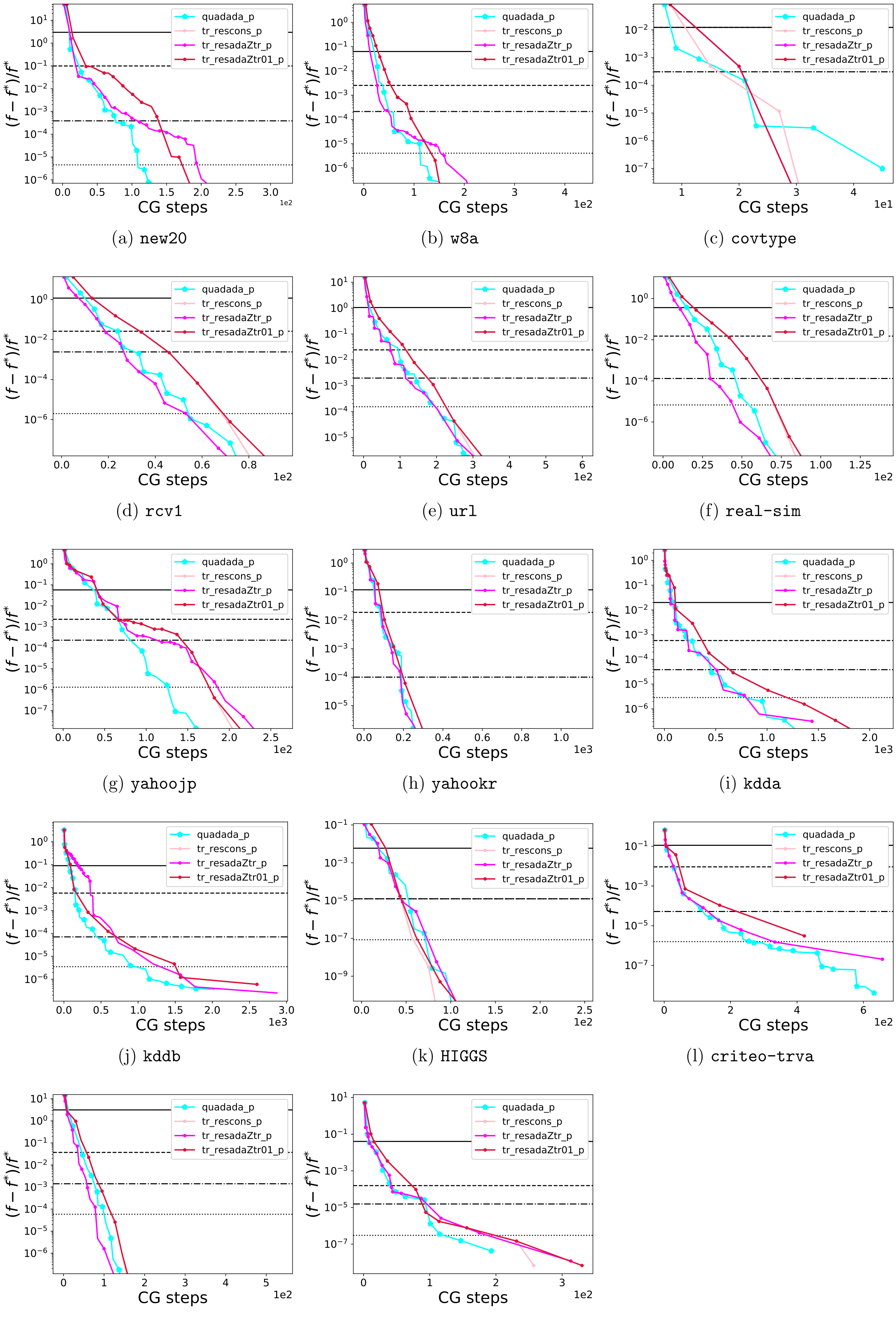


Figure 2: Loss=LR and $C = 100C_{\text{best}}$



(a) new20 (b) w8a (c) covtype

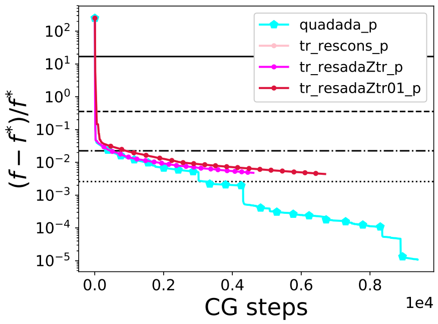
(d) rcv1 (e) url (f) real-sim

(g) yahoojp (h) yahookr (i) kdda

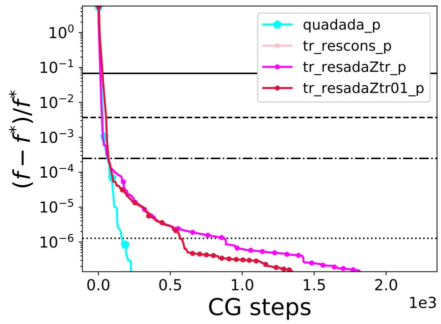
(j) kddb (k) HIGGS (l) criteo-trva

(m) webspam (n) kdd12-svm

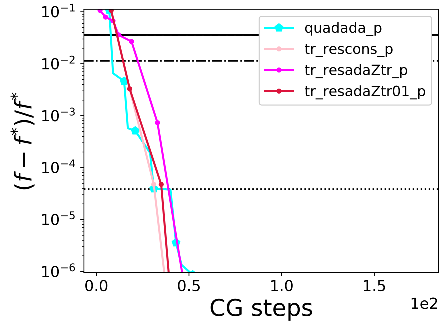
Figure 3: Loss=L2 and $C = C_{\text{best}}$



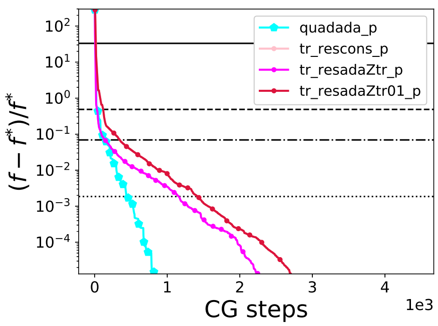
(a) new20



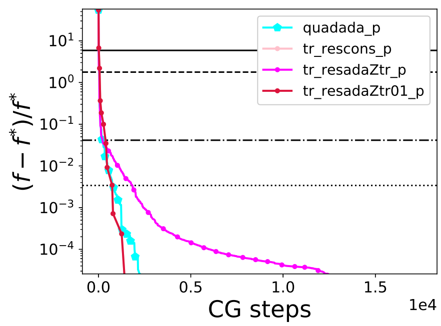
(b) w8a



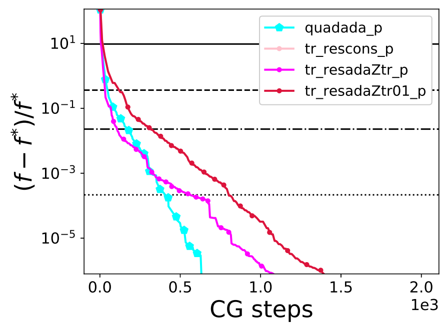
(c) covtype



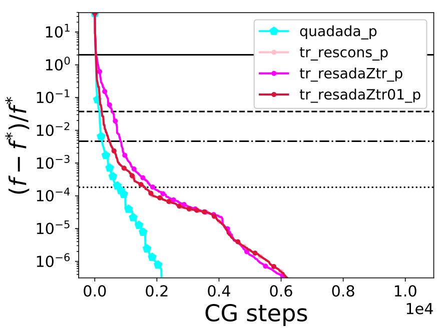
(d) rcv1



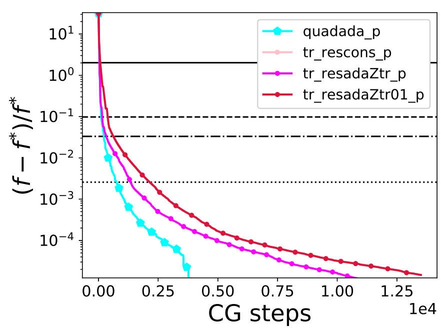
(e) url



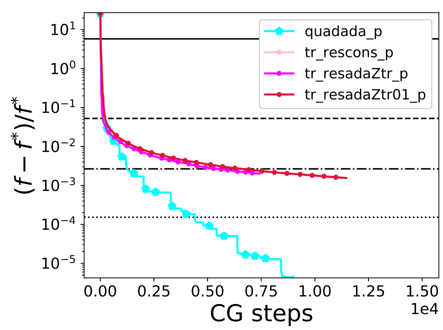
(f) real-sim



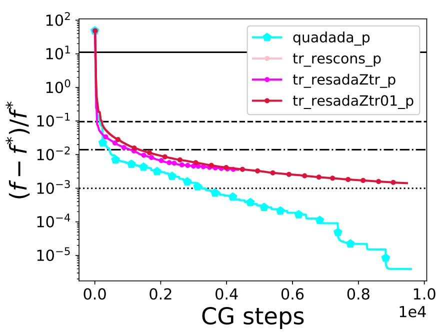
(g) yahoojp



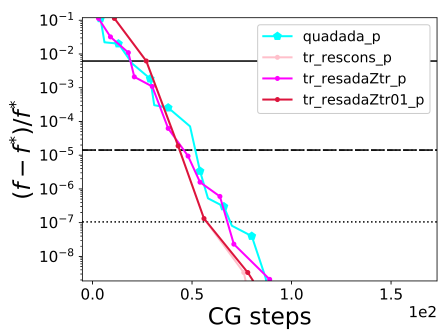
(h) yahookr



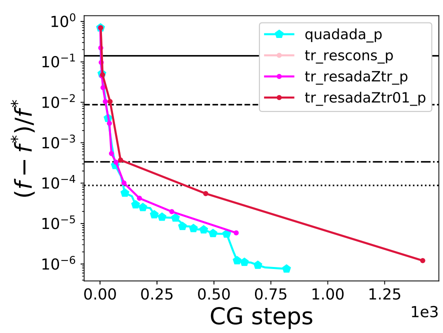
(i) kdda



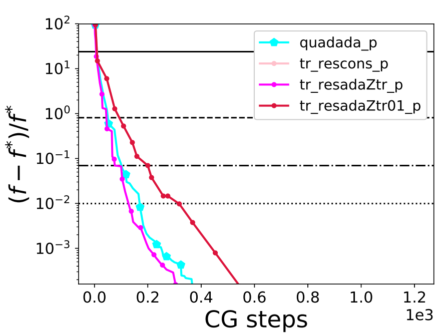
(j) kddb



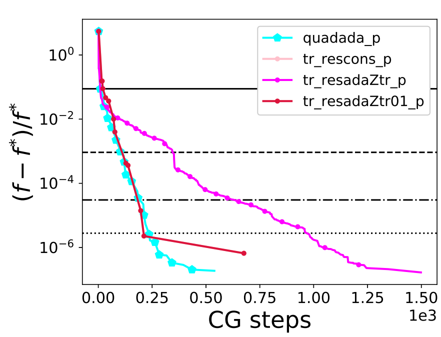
(k) HIGGS



(l) criteo-trva

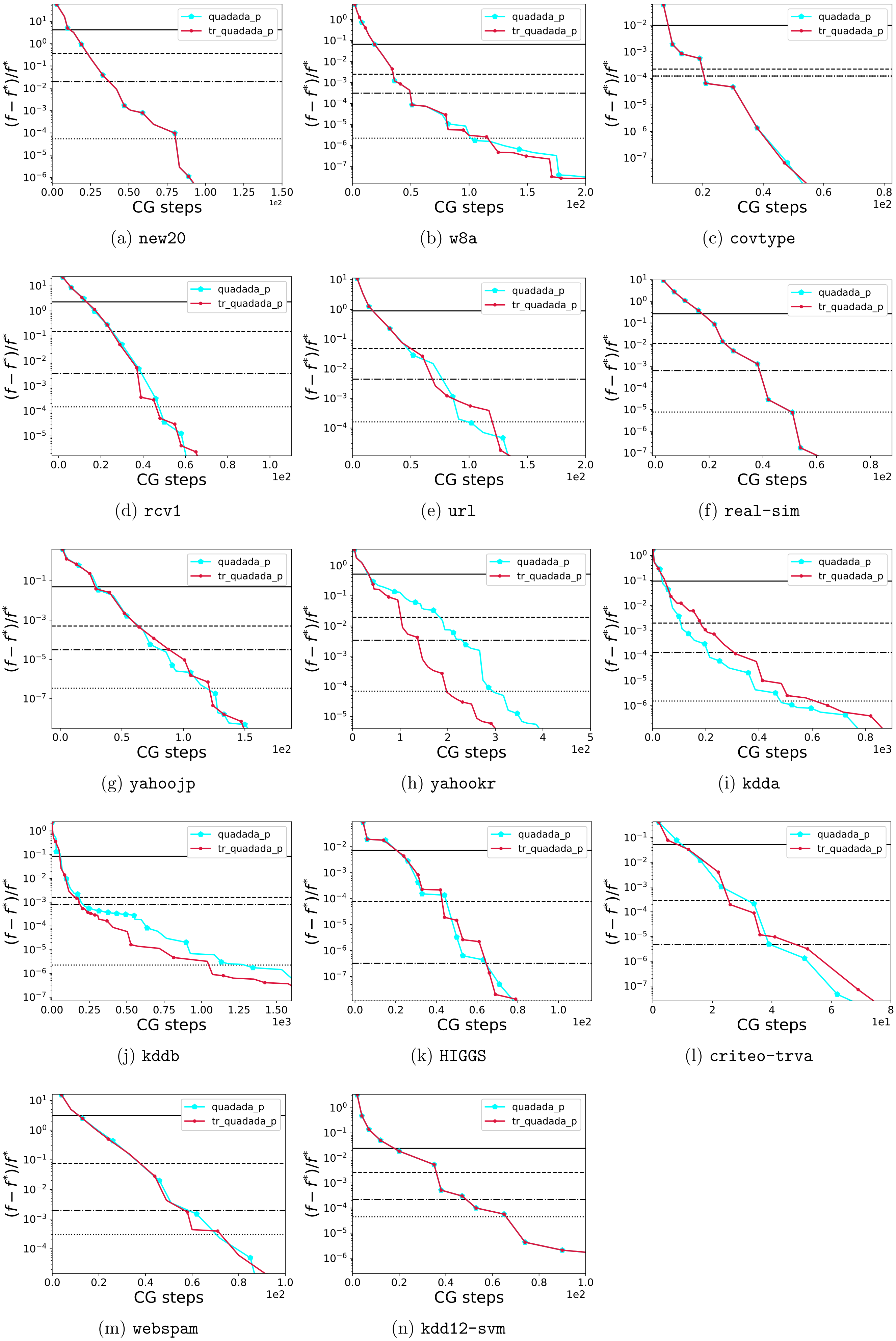


(m) webspam



(n) kdd12-svm

Figure 4: Loss=L2 and $C = 100C_{\text{best}}$



(a) `new20` (b) `w8a` (c) `covtype`

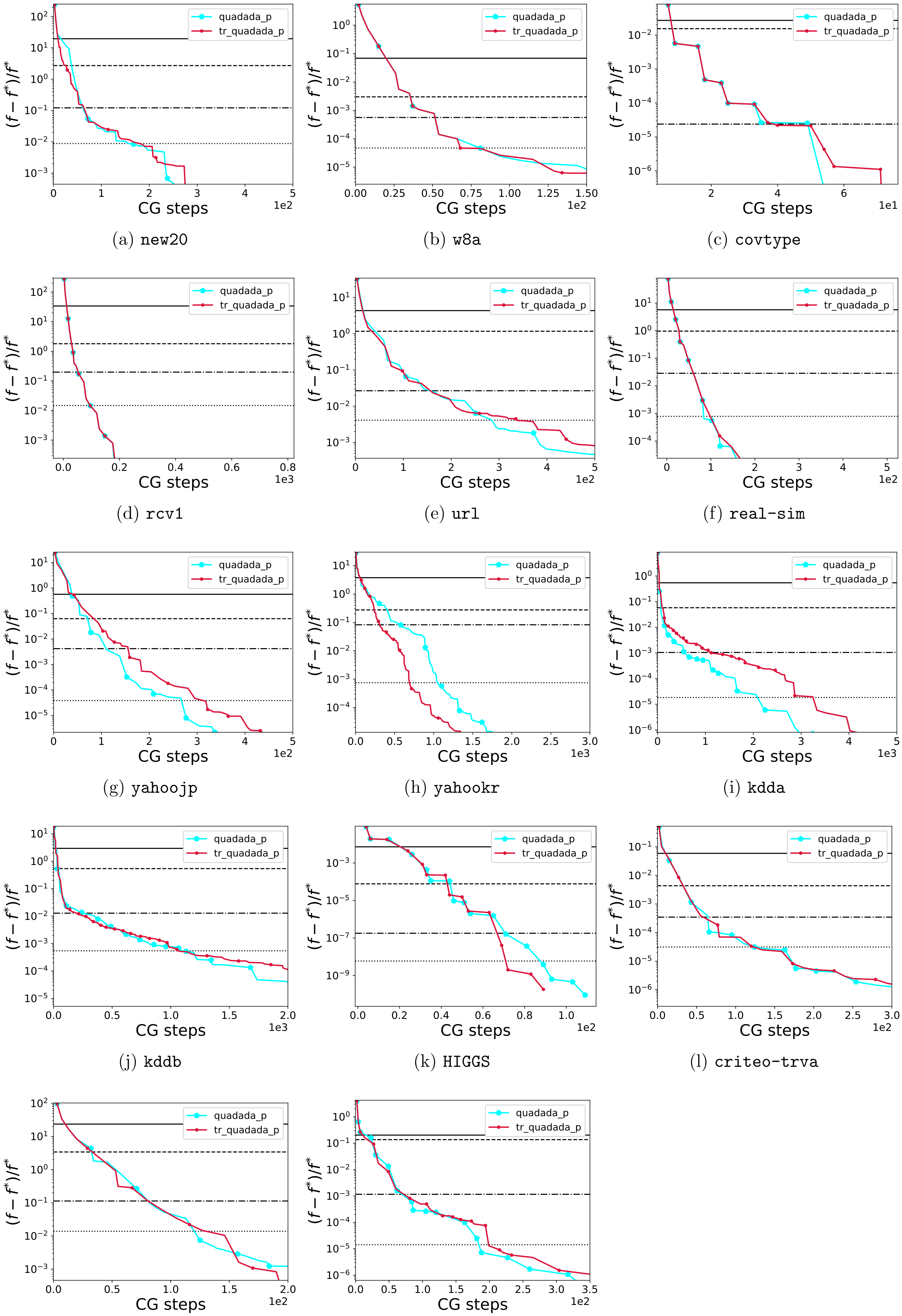
(d) `rcv1` (e) `url` (f) `real-sim`

(g) `yahoojp` (h) `yahookr` (i) `kdda`

(j) `kddb` (k) `HIGGS` (l) `criteo-trva`

(m) `webspam` (n) `kdd12-svm`

Figure 5: Loss=LR and $C = C_{\text{best}}$



(a) `new20` (b) `w8a` (c) `covtype`

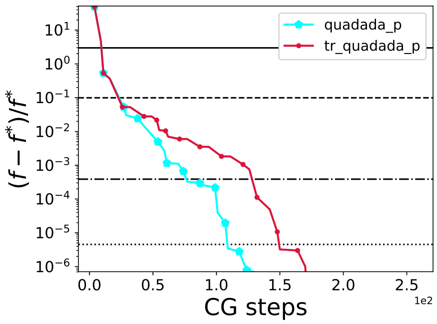
(d) `rcv1` (e) `url` (f) `real-sim`

(g) `yahoojp` (h) `yahookr` (i) `kdda`

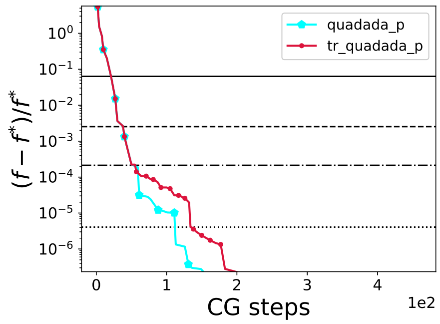
(j) `kddb` (k) `HIGGS` (l) `criteo-trva`

(m) `webspam` (n) `kdd12-svm`

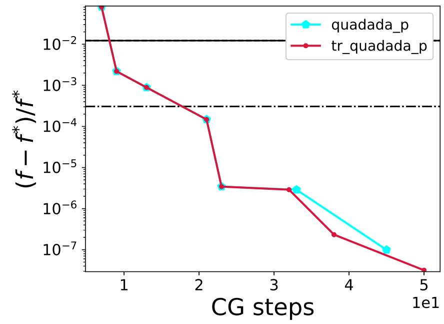
Figure 6: Loss=LR and $C = 100C_{\text{best}}$



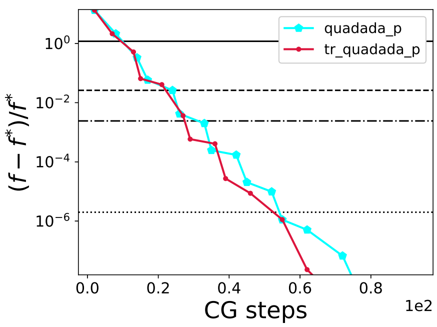
(a) `new20`



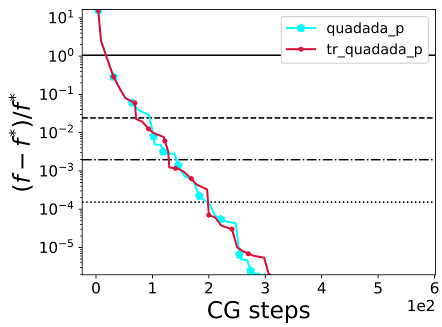
(b) `w8a`



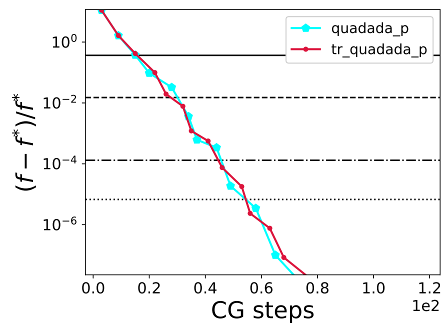
(c) `covtype`



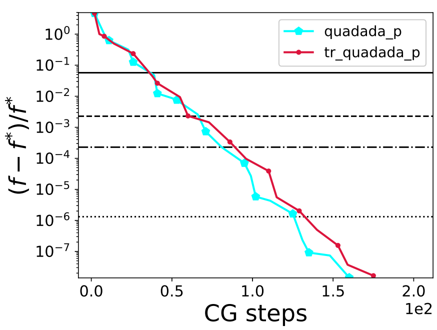
(d) `rcv1`



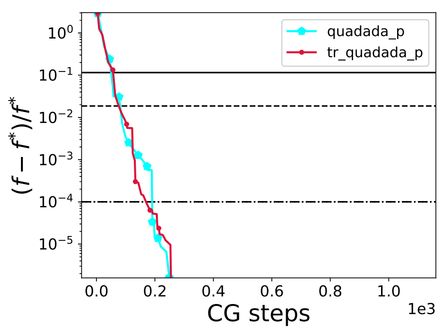
(e) `url`



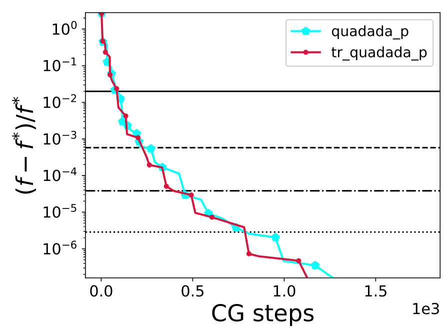
(f) `real-sim`



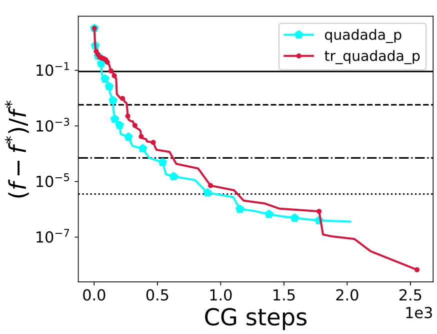
(g) `yahoojp`



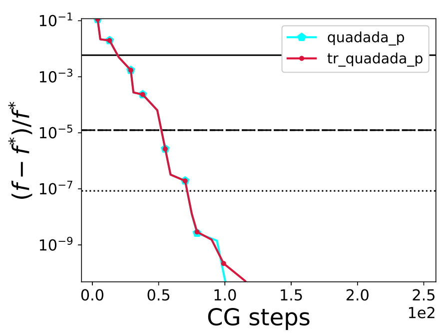
(h) `yahookr`



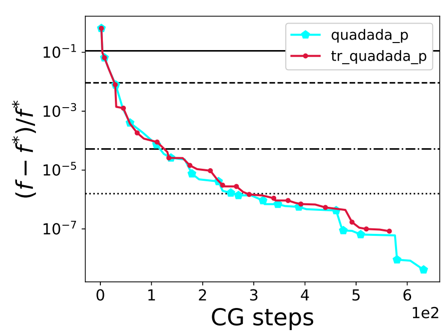
(i) `kdda`



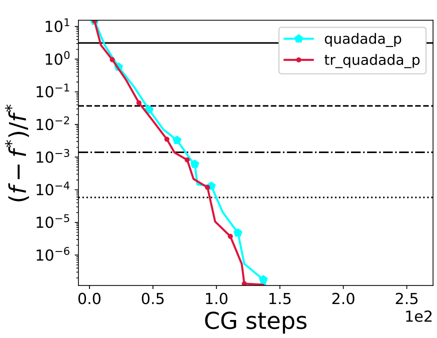
(j) `kddb`



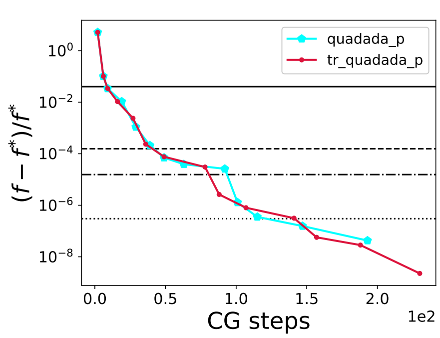
(k) `HIGGS`



(l) `criteo-trva`

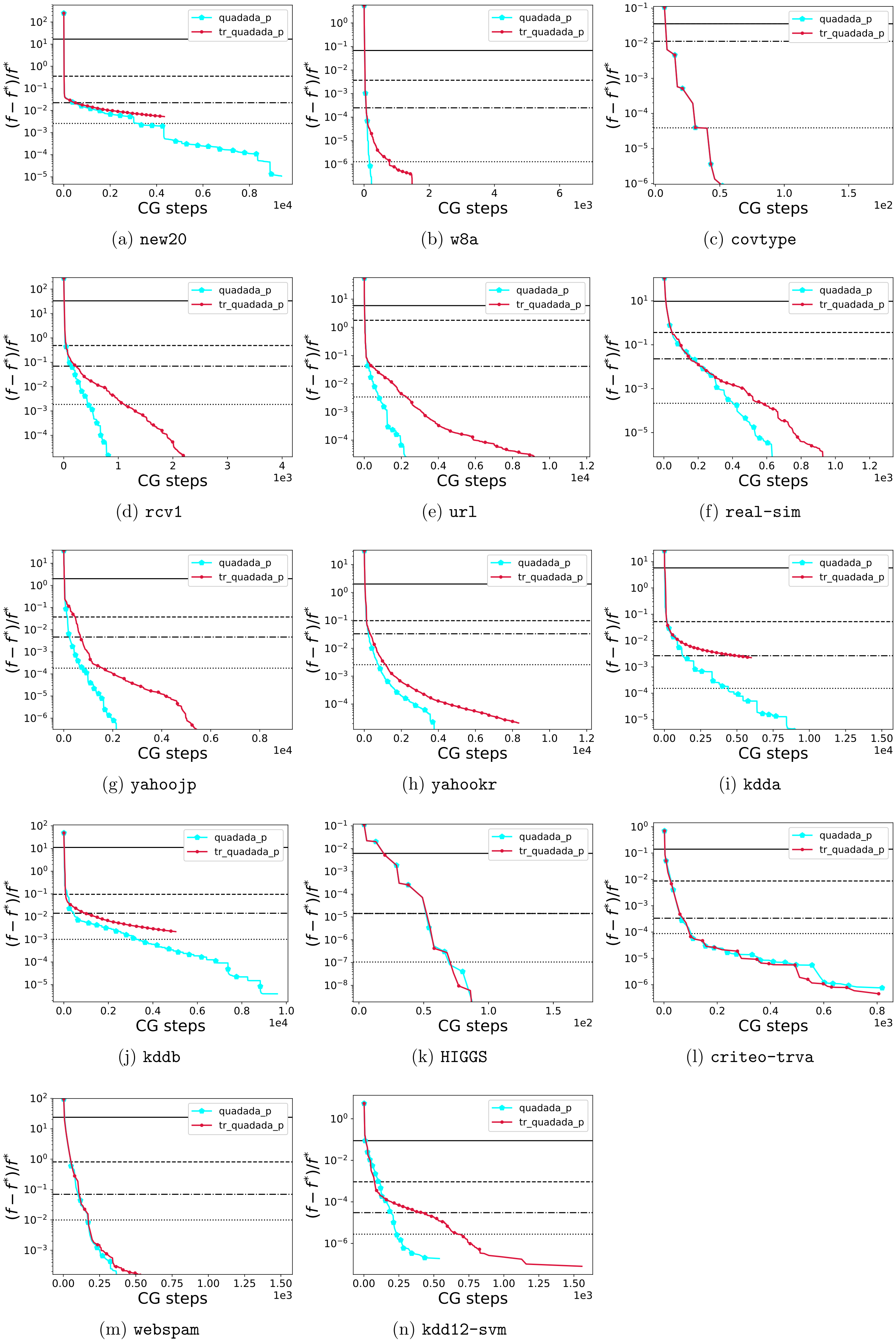


(m) `webspam`



(n) `kdd12-svm`

Figure 7: Loss=L2 and $C = C_{\text{best}}$



(a) new20 (b) w8a (c) covtype (d) rcv1 (e) url (f) real-sim (g) yahoojp (h) yahookr (i) kdda (j) kddb (k) HIGGS (l) criteo-trva (m) webspam (n) kdd12-svm

Figure 8: Loss=L2 and $C = 100C_{\text{best}}$

Table 2: Logistic Regression. Stopping tolerance is 0.01. $C = C_{\text{best}}$

Data set and approaches	# Iter.	# CG	Time	Function value
astro-ph_62369.dat.quadada05	6	25	1.60	37230.8
astro-ph_62369.dat.tr_quadada05	7	28	1.75	37193.1
astro-ph_62369.dat.tr_rescons01	6	31	1.77	37209.9
australian.quadada05	6	12	0.00	0.00653429
australian.tr_quadada05	7	14	0.00	0.00653429
australian.tr_rescons01	6	12	0.00	0.00653431
breast-cancer.quadada05	2	2	0.00	3.28462e-15
breast-cancer.tr_quadada05	2	2	0.00	3.28462e-15
breast-cancer.tr_rescons01	2	1	0.00	3.28462e-15
cod-rna.quadada05	5	12	0.11	2.33952
cod-rna.tr_quadada05	5	12	0.10	2.33952
cod-rna.tr_rescons01	5	8	0.14	2.33961
covtype.libsvm.binary.quadada05	5	25	1.56	0.361434
covtype.libsvm.binary.tr_quadada05	7	31	1.80	0.361434
covtype.libsvm.binary.tr_rescons01	4	27	1.65	0.361434
gisette_scale.quadada05	7	40	7.06	12.4518
gisette_scale.tr_quadada05	7	39	7.19	12.4792
gisette_scale.tr_rescons01	7	65	9.44	12.3755
kdda.quadada05	9	111	220.66	269916
kdda.tr_quadada05	11	184	331.26	270027
kdda.tr_rescons01	7	169	290.07	269814
kddb.quadada05	12	216	690.24	2.01502e+06
kddb.tr_quadada05	11	178	629.89	2.01556e+06
kddb.tr_rescons01	7	213	646.82	2.01449e+06
leisure.scale.quadada05	8	80	99.99	4.77221e+06
leisure.scale.tr_quadada05	14	138	149.45	4.73468e+06
leisure.scale.tr_rescons01	7	209	195.62	4.82295e+06
news20.binary.quadada05	7	25	4.45	127369
news20.binary.tr_quadada05	7	25	4.44	127369
news20.binary.tr_rescons01	6	34	4.94	135363
rcv1_test.binary.quadada05	8	34	19.08	328302
rcv1_test.binary.tr_quadada05	8	34	19.70	328302
rcv1_test.binary.tr_rescons01	7	64	25.10	325214
real-sim.quadada05	8	37	1.32	86892
real-sim.tr_quadada05	8	37	1.36	86892
real-sim.tr_rescons01	7	47	1.49	87674.9
train308.scale.quadada05	8	52	11.23	115274
train308.scale.tr_quadada05	7	45	10.14	115320
train308.scale.tr_rescons01	7	56	11.37	115270
url_combined.quadada05	6	34	98.09	358.956
url_combined.tr_quadada05	6	34	99.06	358.938
url_combined.tr_rescons01	6	38	104.65	357.982

Table 3: Logistic Regression. Stopping tolerance is 0.01. $C = 100C_{\text{best}}$

Data set and approaches	# Iter.	# CG	Time	Function value
astro-ph_62369.dat_100.quadada05	8	36	1.87	413187
astro-ph_62369.dat_100.tr_quadada05	8	36	1.90	413187
astro-ph_62369.dat_100.tr_rescons01	8	56	2.23	381079
australian_100.quadada05	6	19	0.00	0.55557
australian_100.tr_quadada05	6	19	0.00	0.55557
australian_100.tr_rescons01	6	19	0.00	0.555578
breast-cancer_100.quadada05	2	2	0.00	3.25189e-13
breast-cancer_100.tr_quadada05	2	2	0.00	3.25189e-13
breast-cancer_100.tr_rescons01	2	1	0.00	3.25189e-13
cod-rna_100.quadada05	5	16	0.10	214.455
cod-rna_100.tr_quadada05	5	16	0.11	214.455
cod-rna_100.tr_rescons01	5	12	0.14	214.455
covtype.libsvm.binary_100.quadada05	10	66	2.52	33.6756
covtype.libsvm.binary_100.tr_quadada05	6	30	1.74	33.7001
covtype.libsvm.binary_100.tr_rescons01	4	40	1.83	33.675
gisette_scale_100.quadada05	8	49	7.81	145.282
gisette_scale_100.tr_quadada05	9	57	8.75	77.8878
gisette_scale_100.tr_rescons01	7	70	9.61	178.955
kdda_100.quadada05	8	74	168.43	8.22089e+06
kdda_100.tr_quadada05	13	149	287.16	8.14424e+06
kdda_100.tr_rescons01	7	195	320.11	8.07774e+06
kddb_100.quadada05	9	75	314.06	3.6007e+07
kddb_100.tr_quadada05	9	77	325.03	3.58259e+07
kddb_100.tr_rescons01	7	169	543.46	3.6474e+07
leisure.scale_100.quadada05	14	318	316.30	9.63739e+07
leisure.scale_100.tr_quadada05	32	148	189.97	1.38588e+08
leisure.scale_100.tr_rescons01	13	145	154.76	1.07347e+08
news20.binary_100.quadada05	8	41	5.69	4.71473e+06
news20.binary_100.tr_quadada05	15	39	6.44	4.44918e+06
news20.binary_100.tr_rescons01	6	42	5.46	5.35747e+06
rcv1_test.binary_100.quadada05	7	28	17.71	2.3323e+07
rcv1_test.binary_100.tr_quadada05	7	28	17.99	2.3323e+07
rcv1_test.binary_100.tr_rescons01	7	73	26.76	2.11581e+07
real-sim_100.quadada05	8	35	1.28	1.74773e+06
real-sim_100.tr_quadada05	8	35	1.36	1.74773e+06
real-sim_100.tr_rescons01	7	53	1.60	1.96544e+06
train308.scale_100.quadada05	8	59	11.77	2.64626e+06
train308.scale_100.tr_quadada05	9	68	13.21	2.60408e+06
train308.scale_100.tr_rescons01	9	195	26.73	2.57713e+06
url_combined_100.quadada05	8	70	154.88	15443.3
url_combined_100.tr_quadada05	7	38	106.67	19623.7
url_combined_100.tr_rescons01	7	109	208.14	15185.1

Table 4: Logistic Regression. Stopping tolerance is 0.001. $C = C_{\text{best}}$

Data set and approaches	# Iter.	# CG	Time	Function value
astro-ph_62369.dat.quadada05	9	40	1.98	37163
astro-ph_62369.dat.tr_quadada05	9	39	1.92	37163
astro-ph_62369.dat.tr_rescons01	7	40	1.94	37163.5
australian.quadada05	7	17	0.00	0.00653428
australian.tr_quadada05	7	14	0.00	0.00653429
australian.tr_rescons01	7	15	0.00	0.00653429
breast-cancer.quadada05	2	2	0.00	3.28462e-15
breast-cancer.tr_quadada05	2	2	0.00	3.28462e-15
breast-cancer.tr_rescons01	2	1	0.00	3.28462e-15
cod-rna.quadada05	6	15	0.11	2.33952
cod-rna.tr_quadada05	6	15	0.10	2.33952
cod-rna.tr_rescons01	6	11	0.16	2.33952
covtype.libsvm.binary.quadada05	6	35	1.78	0.361434
covtype.libsvm.binary.tr_quadada05	8	41	2.01	0.361434
covtype.libsvm.binary.tr_rescons01	4	27	1.60	0.361434
gisette_scale.quadada05	9	57	9.21	12.0616
gisette_scale.tr_quadada05	11	70	9.78	12.0527
gisette_scale.tr_rescons01	9	95	11.86	12.0514
kdda.quadada05	15	299	484.50	269719
kdda.tr_quadada05	19	392	651.80	269717
kdda.tr_rescons01	10	414	619.29	269718
kddb.quadada05	21	596	1675.35	2.0138e+06
kddb.tr_quadada05	22	540	1589.59	2.01379e+06
kddb.tr_rescons01	8	366	1057.56	2.01389e+06
leisure.scale.quadada05	15	181	183.71	4.72727e+06
leisure.scale.tr_quadada05	22	233	225.11	4.72712e+06
leisure.scale.tr_rescons01	10	350	309.45	4.72887e+06
news20.binary.quadada05	9	42	5.68	122781
news20.binary.tr_quadada05	9	42	6.07	122781
news20.binary.tr_rescons01	8	71	7.71	122839
rcv1_test.binary.quadada05	11	52	23.26	324891
rcv1_test.binary.tr_quadada05	11	52	24.08	324891
rcv1_test.binary.tr_rescons01	8	83	29.33	324645
real-sim.quadada05	9	48	1.46	86652.5
real-sim.tr_quadada05	9	48	1.46	86652.5
real-sim.tr_rescons01	9	76	1.88	86639.5
train308.scale.quadada05	12	70	13.67	115266
train308.scale.tr_quadada05	10	67	13.03	115266
train308.scale.tr_rescons01	8	76	13.96	115266
url_combined.quadada05	8	52	129.58	355.785
url_combined.tr_quadada05	10	69	158.60	355.653
url_combined.tr_rescons01	7	54	129.91	355.757

Table 5: Logistic Regression. Stopping tolerance is 0.001. $C = 100C_{\text{best}}$

Data set and approaches	# Iter.	# CG	Time	Function value
astro-ph_62369.dat_100.quadada05	10	55	2.29	364297
astro-ph_62369.dat_100.tr_quadada05	10	55	2.28	364297
astro-ph_62369.dat_100.tr_rescons01	10	99	3.03	361586
australian_100.quadada05	7	25	0.00	0.555569
australian_100.tr_quadada05	7	25	0.00	0.555569
australian_100.tr_rescons01	7	25	0.00	0.555569
breast-cancer_100.quadada05	3	4	0.00	3.25189e-13
breast-cancer_100.tr_quadada05	3	4	0.00	3.25189e-13
breast-cancer_100.tr_rescons01	3	2	0.00	3.25189e-13
cod-rna_100.quadada05	6	20	0.11	214.454
cod-rna_100.tr_quadada05	6	20	0.11	214.454
cod-rna_100.tr_rescons01	6	15	0.15	214.454
covtype.libsvm.binary_100.quadada05	11	82	2.83	33.675
covtype.libsvm.binary_100.tr_quadada05	14	97	3.21	33.675
covtype.libsvm.binary_100.tr_rescons01	5	51	2.07	33.675
gisette_scale_100.quadada05	11	73	10.37	44.9448
gisette_scale_100.tr_quadada05	11	73	9.93	44.8718
gisette_scale_100.tr_rescons01	9	94	11.72	56.1174
kdda_100.quadada05	22	487	774.40	8.03658e+06
kdda_100.tr_quadada05	57	612	1052.77	8.05269e+06
kdda_100.tr_rescons01	16	551	817.21	8.04512e+06
kddb_100.quadada05	29	667	1928.05	3.52761e+07
kddb_100.tr_quadada05	24	278	935.04	3.54942e+07
kddb_100.tr_rescons01	24	311	988.91	3.55692e+07
leisure.scale_100.quadada05	21	484	408.46	7.20816e+07
leisure.scale_100.tr_quadada05	62	396	412.70	7.2868e+07
leisure.scale_100.tr_rescons01	44	408	396.69	7.7787e+07
news20.binary_100.quadada05	11	63	7.25	3.1211e+06
news20.binary_100.tr_quadada05	19	61	8.27	3.16241e+06
news20.binary_100.tr_rescons01	9	140	12.62	3.06152e+06
rcv1_test.binary_100.quadada05	13	70	27.32	2.02834e+07
rcv1_test.binary_100.tr_quadada05	13	72	28.53	2.02702e+07
rcv1_test.binary_100.tr_rescons01	9	146	41.41	2.02156e+07
real-sim_100.quadada05	12	75	1.78	1.31317e+06
real-sim_100.tr_quadada05	12	75	1.84	1.31317e+06
real-sim_100.tr_rescons01	10	140	2.55	1.3419e+06
train308.scale_100.quadada05	13	127	19.50	2.55838e+06
train308.scale_100.tr_quadada05	19	160	24.13	2.55897e+06
train308.scale_100.tr_rescons01	11	226	30.76	2.56151e+06
url_combined_100.quadada05	12	138	264.17	14734.6
url_combined_100.tr_quadada05	18	209	388.39	14614.9
url_combined_100.tr_rescons01	8	183	315.30	14669.3

Table 6: L2-loss SVM. Stopping tolerance is 0.01. $C = C_{\text{best}}$.

Data set and approaches	# Iter.	# CG	Time	Function value
astro-ph_62369.dat.quadada05	6	32	1.48	2700.12
astro-ph_62369.dat.tr_quadada05	6	28	1.46	2701.52
astro-ph_62369.dat.tr_rescons01	5	29	1.45	2704.73
australian.quadada05	6	13	0.00	0.00115195
australian.tr_quadada05	6	13	0.00	0.00115195
australian.tr_rescons01	6	12	0.00	0.00115195
breast-cancer.quadada05	2	2	0.00	2.36808e-15
breast-cancer.tr_quadada05	2	2	0.00	2.36808e-15
breast-cancer.tr_rescons01	2	1	0.00	2.36808e-15
cod-rna.quadada05	4	9	0.10	0.0941001
cod-rna.tr_quadada05	4	9	0.10	0.0941001
cod-rna.tr_rescons01	5	8	0.14	0.0940987
covtype.libsvm.binary.quadada05	3	15	1.27	0.0319538
covtype.libsvm.binary.tr_quadada05	3	15	1.29	0.0319538
covtype.libsvm.binary.tr_rescons01	3	15	1.40	0.0319538
gisette_scale.quadada05	7	44	5.56	0.362546
gisette_scale.tr_quadada05	6	33	5.33	0.36335
gisette_scale.tr_rescons01	5	53	6.14	0.364112
kdda.quadada05	15	138	250.68	21143.8
kdda.tr_quadada05	36	211	399.85	21146.4
kdda.tr_rescons01	9	220	333.72	21132.3
kddb.quadada05	18	201	610.04	283532
kddb.tr_quadada05	49	271	978.47	283801
kddb.tr_rescons01	6	136	415.95	283625
leisure.scale.quadada05	9	88	84.66	237919
leisure.scale.tr_quadada05	10	86	88.57	239173
leisure.scale.tr_rescons01	6	103	92.00	237367
news20.binary.quadada05	5	16	3.01	5557.42
news20.binary.tr_quadada05	5	16	2.99	5557.42
news20.binary.tr_rescons01	4	56	4.61	5604.69
rcv1_test.binary.quadada05	6	30	14.42	24901.6
rcv1_test.binary.tr_quadada05	7	34	15.00	24740
rcv1_test.binary.tr_rescons01	6	63	17.69	24488.9
real-sim.quadada05	6	28	1.03	2995.8
real-sim.tr_quadada05	7	32	1.08	2988.68
real-sim.tr_rescons01	6	42	1.18	2988.71
train308.scale.quadada05	12	71	9.61	15452.9
train308.scale.tr_quadada05	12	75	9.72	15453.3
train308.scale.tr_rescons01	8	78	10.04	15474.4
url_combined.quadada05	7	49	77.58	273.037
url_combined.tr_quadada05	8	63	80.58	258.27
url_combined.tr_rescons01	6	124	119.96	252.758

Table 7: L2-loss SVM. Stopping tolerance is 0.01. $C = 100C_{\text{best}}$.

Data set and approaches	# Iter.	# CG	Time	Function value
astro-ph_62369.dat_100.quadada05	7	48	1.48	21325.8
astro-ph_62369.dat_100.tr_quadada05	7	53	1.58	21251.9
astro-ph_62369.dat_100.tr_rescons01	8	119	1.99	20171.1
australian_100.quadada05	5	17	0.00	0.0930242
australian_100.tr_quadada05	5	17	0.00	0.0930242
australian_100.tr_rescons01	6	19	0.00	0.0930253
breast-cancer_100.quadada05	3	4	0.00	2.27875e-13
breast-cancer_100.tr_quadada05	3	4	0.00	2.27875e-13
breast-cancer_100.tr_rescons01	3	2	0.00	2.27875e-13
cod-rna_100.quadada05	4	9	0.09	8.96774
cod-rna_100.tr_quadada05	4	9	0.10	8.96774
cod-rna_100.tr_rescons01	5	10	0.17	8.96754
covtype.libsvm.binary_100.quadada05	5	31	1.58	2.99748
covtype.libsvm.binary_100.tr_quadada05	5	31	1.63	2.99748
covtype.libsvm.binary_100.tr_rescons01	3	27	1.59	2.99747
gisette_scale_100.quadada05	9	60	5.53	0.740472
gisette_scale_100.tr_quadada05	7	43	5.29	0.922069
gisette_scale_100.tr_rescons01	5	64	6.21	0.929273
kdda_100.quadada05	15	217	305.04	594749
kdda_100.tr_quadada05	84	411	702.94	598116
kdda_100.tr_rescons01	7	149	220.12	597654
kddb_100.quadada05	15	243	620.45	2.48172e+06
kddb_100.tr_quadada05	16	117	389.80	2.5675e+06
kddb_100.tr_rescons01	7	200	529.80	2.6709e+06
leisure.scale_100.quadada05	7	61	62.57	4.04802e+06
leisure.scale_100.tr_quadada05	13	146	87.94	3.02813e+06
leisure.scale_100.tr_rescons01	6	352	153.52	3.17828e+06
news20.binary_100.quadada05	5	14	2.76	285636
news20.binary_100.tr_quadada05	5	14	2.75	285636
news20.binary_100.tr_rescons01	3	18	3.27	282204
rcv1_test.binary_100.quadada05	8	41	14.94	1.62701e+06
rcv1_test.binary_100.tr_quadada05	8	41	15.07	1.62701e+06
rcv1_test.binary_100.tr_rescons01	6	77	18.13	1.64005e+06
real-sim_100.quadada05	8	55	1.02	36996.4
real-sim_100.tr_quadada05	7	59	1.08	38892.1
real-sim_100.tr_rescons01	6	92	1.28	43383.2
train308.scale_100.quadada05	11	72	8.75	255891
train308.scale_100.tr_quadada05	16	86	9.62	274161
train308.scale_100.tr_rescons01	8	126	10.65	243964
url_combined_100.quadada05	7	61	83.13	7393.68
url_combined_100.tr_quadada05	9	50	76.83	8218.4
url_combined_100.tr_rescons01	6	151	124.29	6551.31

Table 8: L2-loss SVM. Stopping tolerance is 0.001. $C = C_{\text{best}}$.

Data set and approaches	# Iter.	# CG	Time	Function value
astro-ph_62369.dat.quadada05	7	38	1.59	2699.68
astro-ph_62369.dat.tr_quadada05	8	38	1.56	2699.63
astro-ph_62369.dat.tr_rescons01	7	52	1.80	2699.6
australian.quadada05	6	13	0.00	0.00115195
australian.tr_quadada05	6	13	0.00	0.00115195
australian.tr_rescons01	6	12	0.00	0.00115195
breast-cancer.quadada05	2	2	0.00	2.36808e-15
breast-cancer.tr_quadada05	2	2	0.00	2.36808e-15
breast-cancer.tr_rescons01	2	1	0.00	2.36808e-15
cod-rna.quadada05	5	12	0.10	0.0940986
cod-rna.tr_quadada05	5	12	0.10	0.0940986
cod-rna.tr_rescons01	5	8	0.13	0.0940987
covtype.libsvm.binary.quadada05	4	28	1.53	0.0319537
covtype.libsvm.binary.tr_quadada05	4	28	1.52	0.0319537
covtype.libsvm.binary.tr_rescons01	4	27	1.60	0.0319537
gisette_scale.quadada05	10	68	6.20	0.362156
gisette_scale.tr_quadada05	13	84	6.67	0.362143
gisette_scale.tr_rescons01	7	84	6.86	0.362144
kdda.quadada05	20	287	437.62	21131.7
kdda.tr_quadada05	41	430	676.62	21131.4
kdda.tr_rescons01	10	402	543.26	21131.3
kddb.quadada05	23	434	1144.17	283401
kddb.tr_quadada05	64	596	1732.14	283400
kddb.tr_rescons01	8	594	1421.59	283393
leisure.scale.quadada05	19	189	143.38	237100
leisure.scale.tr_quadada05	20	160	128.82	237113
leisure.scale.tr_rescons01	7	150	114.23	237106
news20.binary.quadada05	12	75	5.08	5226.7
news20.binary.tr_quadada05	26	142	8.53	5226.1
news20.binary.tr_rescons01	24	213	11.89	5241.33
rcv1_test.binary.quadada05	8	53	15.93	24434.4
rcv1_test.binary.tr_quadada05	14	87	19.06	24423.5
rcv1_test.binary.tr_rescons01	7	82	19.05	24430
real-sim.quadada05	9	44	1.10	2985.14
real-sim.tr_quadada05	10	46	1.15	2985.14
real-sim.tr_rescons01	7	53	1.26	2985.21
train308.scale.quadada05	15	104	11.63	15449.8
train308.scale.tr_quadada05	15	111	11.88	15449.8
train308.scale.tr_rescons01	13	155	14.18	15449.7
url_combined.quadada05	15	113	105.92	253.945
url_combined.tr_quadada05	15	125	107.53	251.049
url_combined.tr_rescons01	7	171	135.30	249.267

Table 9: L2-loss SVM. Stopping tolerance is 0.001. $C = 100C_{\text{best}}$.

Data set and approaches	# Iter.	# CG	Time	Function value
astro-ph_62369.dat_100.quadada05	16	139	2.06	16739.4
astro-ph_62369.dat_100.tr_quadada05	25	164	2.35	17006
astro-ph_62369.dat_100.tr_rescons01	27	284	3.19	16977.8
australian_100.quadada05	6	23	0.00	0.093024
australian_100.tr_quadada05	6	23	0.00	0.093024
australian_100.tr_rescons01	6	19	0.00	0.0930253
breast-cancer_100.quadada05	3	4	0.00	2.27875e-13
breast-cancer_100.tr_quadada05	3	4	0.00	2.27875e-13
breast-cancer_100.tr_rescons01	3	2	0.00	2.27875e-13
cod-rna_100.quadada05	5	12	0.10	8.96754
cod-rna_100.tr_quadada05	5	12	0.10	8.96754
cod-rna_100.tr_rescons01	6	12	0.14	8.96754
covtype.libsvm.binary_100.quadada05	5	31	1.60	2.99748
covtype.libsvm.binary_100.tr_quadada05	5	31	1.63	2.99748
covtype.libsvm.binary_100.tr_rescons01	4	32	1.68	2.99747
gisette_scale_100.quadada05	12	81	5.86	0.711897
gisette_scale_100.tr_quadada05	12	79	5.78	0.703561
gisette_scale_100.tr_rescons01	13	260	8.64	0.773443
kdda_100.quadada05	33	977	1128.81	591283
kdda_100.tr_quadada05	106	710	1093.05	593364
kdda_100.tr_rescons01	25	988	1114.80	593067
kddb_100.quadada05	37	946	2131.26	2.44601e+06
kddb_100.tr_quadada05	50	285	860.76	2.51482e+06
kddb_100.tr_rescons01	48	710	1747.25	2.49899e+06
leisure.scale_100.quadada05	24	255	123.66	2.91532e+06
leisure.scale_100.tr_quadada05	17	228	113.63	2.97383e+06
leisure.scale_100.tr_rescons01	18	566	221.33	2.95069e+06
news20.binary_100.quadada05	9	32	3.25	252856
news20.binary_100.tr_quadada05	9	32	3.29	252856
news20.binary_100.tr_rescons01	10	245	10.95	250554
rcv1_test.binary_100.quadada05	15	103	18.31	1.55304e+06
rcv1_test.binary_100.tr_quadada05	19	157	21.56	1.55861e+06
rcv1_test.binary_100.tr_rescons01	9	167	21.94	1.55081e+06
real-sim_100.quadada05	18	180	1.28	33422.2
real-sim_100.tr_quadada05	19	129	1.24	34325.7
real-sim_100.tr_rescons01	24	278	1.64	33951.2
train308.scale_100.quadada05	29	289	17.25	230929
train308.scale_100.tr_quadada05	130	722	43.05	231087
train308.scale_100.tr_rescons01	49	676	33.98	230856
url_combined_100.quadada05	11	121	101.79	6259.39
url_combined_100.tr_quadada05	15	129	99.17	6224.76
url_combined_100.tr_rescons01	9	573	221.14	5780.31

Table 10: Parameter selection for logistic regression. # Iter. and # CG are respectively the total number of Newton iterations and CG steps in the entire parameter-selection procedure.

Data set and approaches	C_{best}	Best Acc.	# Iter.	# CG	Time
astro-ph_62369.dat.quadada05_fixed	8	96.82%	357	690	34.23
astro-ph_62369.dat.tr_rescons01	8	96.83%	434	619	35.79
astro-ph_62369.dat.tr_rescons01_fixed	8	96.84%	330	834	35.29
australian.quadada05_fixed	1.53e-05	68.99%	337	433	0.03
australian.tr_rescons01	1.53e-05	68.12%	364	242	0.02
australian.tr_rescons01_fixed	1.53e-05	68.55%	340	266	0.02
breast-cancer.quadada05_fixed	6.94e-18	65.01%	235	225	0.01
breast-cancer.tr_rescons01	6.94e-18	65.01%	235	110	0.01
breast-cancer.tr_rescons01_fixed	6.94e-18	65.01%	235	110	0.01
cod-rna.quadada05_fixed	0.000122	87.58%	305	425	3.56
cod-rna.tr_rescons01	0.000122	87.58%	375	325	3.85
cod-rna.tr_rescons01_fixed	0.000122	87.58%	325	300	3.68
covtype.libsvm.binary.quadada05_fixed	1.91e-06	61.36%	307	667	135.16
covtype.libsvm.binary.tr_rescons01	9.54e-07	61.30%	425	717	180.20
covtype.libsvm.binary.tr_rescons01_fixed	9.54e-07	61.15%	309	500	137.18
gisette_scale.quadada05_fixed	0.125	97.25%	362	637	86.89
gisette_scale.tr_rescons01	0.125	97.25%	409	601	92.56
gisette_scale.tr_rescons01_fixed	0.125	97.22%	332	807	96.65
kdda.quadada05_fixed	0.125	88.23%	543	2953	19224.99
kdda.tr_rescons01	0.125	88.23%	602	2508	18418.84
kdda.tr_rescons01_fixed	0.125	88.24%	400	3257	19611.02
kddb.quadada05_fixed	0.25	88.78%	578	2861	44900.96
kddb.tr_rescons01	0.5	88.83%	629	2510	43606.88
kddb.tr_rescons01_fixed	0.0625	88.65%	405	2883	41949.05
leisure.scale.quadada05_fixed	32	87.33%	435	1802	1916.34
leisure.scale.tr_rescons01	64	87.29%	450	1390	1773.29
leisure.scale.tr_rescons01_fixed	32	87.31%	353	2251	2237.48
news20.binary.quadada05_fixed	512	96.54%	346	604	80.60
news20.binary.tr_rescons01	512	96.49%	400	548	88.26
news20.binary.tr_rescons01_fixed	512	96.51%	319	659	90.05
rcv1_test.binary.quadada05_fixed	4	97.73%	325	507	351.58
rcv1_test.binary.tr_rescons01	8	97.77%	404	504	431.96
rcv1_test.binary.tr_rescons01_fixed	16	97.75%	330	743	443.28
real-sim.quadada05_fixed	8	97.53%	349	605	28.70
real-sim.tr_rescons01	32	97.53%	416	538	30.53
real-sim.tr_rescons01_fixed	16	97.54%	335	758	30.82
train308.scale.quadada05_fixed	4	92.66%	410	1399	287.35
train308.scale.tr_rescons01	4	92.65%	516	1304	307.93
train308.scale.tr_rescons01_fixed	4	92.66%	385	1634	318.46
url_combined.quadada05_fixed	0.00781	97.94%	329	606	2251.23
url_combined.tr_rescons01	0.00195	97.71%	396	557	2535.49
url_combined.tr_rescons01_fixed	0.00195	97.80%	325	670	2462.06

Table 11: Parameter selection for L2-loss SVM. # Iter. and # CG are respectively the total number of Newton iterations and CG steps in the entire parameter-selection procedure.

Data set and approaches	C_{best}	Best Acc.	# Iter.	# CG	Time
astro-ph_62369.dat.quadada05_fixed	0.5	96.94%	358	786	25.37
astro-ph_62369.dat.tr_rescons01	0.5	96.94%	426	693	27.97
astro-ph_62369.dat.tr_rescons01_fixed	0.5	96.93%	310	859	24.38
australian.quadada05_fixed	1.91e-06	67.83%	303	392	0.02
australian.tr_rescons01	1.91e-06	67.68%	331	222	0.02
australian.tr_rescons01_fixed	1.91e-06	67.83%	318	248	0.02
breast-cancer.quadada05_fixed	3.47e-18	65.01%	208	194	0.01
breast-cancer.tr_rescons01	3.47e-18	65.01%	208	93	0.01
breast-cancer.tr_rescons01_fixed	3.47e-18	65.01%	208	93	0.00
cod-rna.quadada05_fixed	3.81e-06	87.57%	265	378	2.59
cod-rna.tr_rescons01	3.81e-06	87.57%	344	309	3.02
cod-rna.tr_rescons01_fixed	3.81e-06	87.57%	285	270	2.76
covtype.libsvm.binary.quadada05_fixed	2.98e-08	60.93%	266	636	116.48
covtype.libsvm.binary.tr_rescons01	5.96e-08	61.33%	407	685	159.64
covtype.libsvm.binary.tr_rescons01_fixed	1.49e-08	61.10%	275	479	114.49
gisette_scale.quadada05_fixed	0.000977	97.42%	343	813	70.55
gisette_scale.tr_rescons01	0.000977	97.37%	405	697	78.40
gisette_scale.tr_rescons01_fixed	0.000977	97.32%	287	888	70.66
kdda.quadada05_fixed	0.00781	88.25%	526	3009	16486.75
kdda.tr_rescons01	0.00781	88.25%	570	2568	16019.09
kdda.tr_rescons01_fixed	0.00781	88.25%	366	2771	15109.19
kddb.quadada05_fixed	0.0312	88.92%	542	3034	39075.86
kddb.tr_rescons01	0.0625	88.91%	586	2747	40066.64
kddb.tr_rescons01_fixed	0.00781	88.68%	380	2782	36330.79
leisure.scale.quadada05_fixed	4	87.40%	413	1843	1635.57
leisure.scale.tr_rescons01	2	87.37%	423	1529	1610.22
leisure.scale.tr_rescons01_fixed	4	87.35%	310	2185	1788.81
news20.binary.quadada05_fixed	64	96.80%	357	698	72.23
news20.binary.tr_rescons01	32	96.79%	394	633	79.79
news20.binary.tr_rescons01_fixed	64	96.77%	301	734	77.10
rcv1_test.binary.quadada05_fixed	0.5	97.79%	315	538	275.09
rcv1_test.binary.tr_rescons01	0.5	97.80%	377	550	330.99
rcv1_test.binary.tr_rescons01_fixed	0.5	97.81%	295	763	316.38
real-sim.quadada05_fixed	0.5	97.56%	353	648	20.80
real-sim.tr_rescons01	0.5	97.56%	414	643	23.60
real-sim.tr_rescons01_fixed	0.5	97.56%	310	778	21.01
train308.scale.quadada05_fixed	0.5	92.80%	482	2061	235.02
train308.scale.tr_rescons01	0.5	92.79%	536	1822	245.32
train308.scale.tr_rescons01_fixed	0.5	92.78%	351	1976	222.56
url_combined.quadada05_fixed	0.00195	98.06%	288	586	1668.38
url_combined.tr_rescons01	0.00195	98.15%	381	613	2103.13
url_combined.tr_rescons01_fixed	0.00391	98.52%	305	1078	2124.37