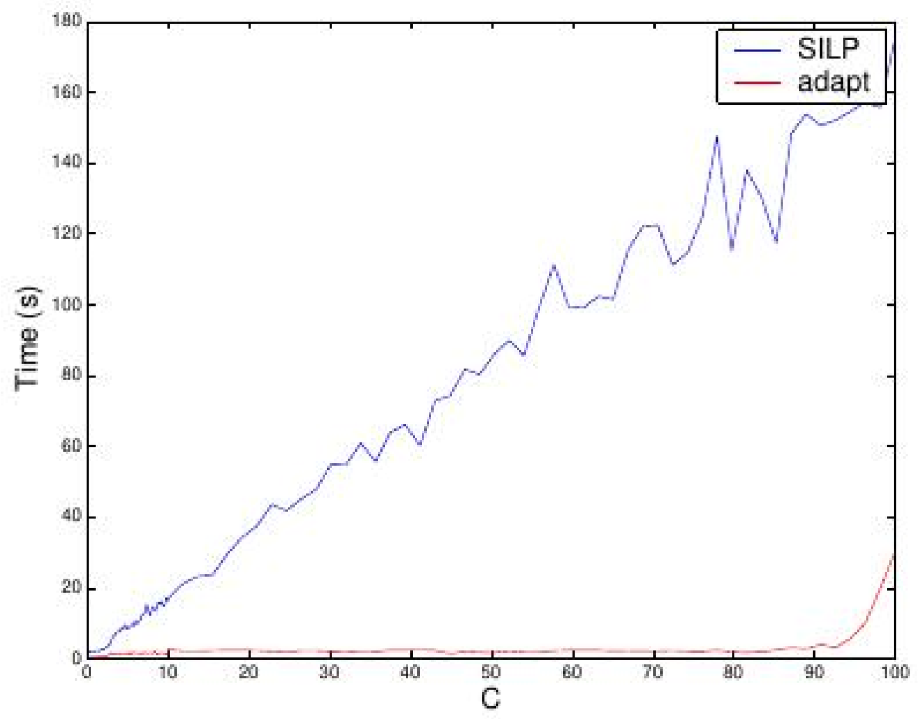

Figure 3. Averaged over 5 runs of each step computation time for both algorithms for the Ionosphere dataset. For $C = 100$, we have the largest computation time since in such case there is no warm-start.

Table 2. Averaged time computation (in seconds) of an approximated regularization path with 100 samples of C varying from 0.1 to 100. We compare the performance of the MKL SILP algorithm and our algorithm.

data	SILP	Adaptive	Ratio
Liver	612 ± 180	116 ± 49	5.3
Pima	4496 ± 736	754 ± 44	6.0
Credit	4830 ± 754	753 ± 84	6.4
Ionosphere	5000 ± 788	221 ± 22	22.6
Sonar	11234 ± 2928	170 ± 25	66.1

ing more kernels than the SILP MKL algorithm. However, we have empirically showed that our algorithm is more stable in the kernel weights and thus it needs fewer iterations to converge towards a reasonable solution, making it globally faster than the SILP algorithm.

Future works aim at improving both speed and sparsity in kernels of the algorithm (for instance by initializing the SILP framework with the results of our algorithm) and by extending this algorithm to other SVM algorithm such as the one-class or the regression SVM.

Acknowledgments

We wish to thank reviewers for their useful comments. This work was supported by grants from the IST programme of the European Community under the PASCAL Network of excellence, IST-2002-506778.

References

Argyriou, A., Evgeniou, T., & Pontil, M. (2007). *Convex multi-task feature learning* (Technical Report).

Bach, F., Lanckriet, G., & Jordan, M. (2004). Multiple kernel learning, conic duality, and the smo algorithm. *Proceedings of the 21st International Conference on Machine Learning* (pp. 41–48).

Bonnans, J., Gilbert, J., Lemaréchal, C., & Sagastizbal, C. (2003). *Numerical optimization theoretical and practical aspects*. Springer.

Bonnans, J., & Shapiro, A. (1998). Optimization problems with pertubation : A guided tour. *SIAM Review*, *40*, 202–227.

Chapelle, O., Vapnik, V., Bousquet, O., & Mukerjhee, S. (2002). Choosing multiple parameters for SVM. *Machine Learning*, *46*, 131–159.

DeCoste, D., & Wagstaff., K. (2000). Alpha seeding for support vector machines. *International Conference on Knowledge Discovery and Data Mining*.

Grandvalet, Y. (1998). Least absolute shrinkage is equivalent to quadratic penalization. *ICANN'98* (pp. 201–206). Springer.

Grandvalet, Y., & Canu, S. (2003). Adaptive scaling for feature selection in svms. *Advances in Neural Information Processing Systems*. MIT Press.

Lanckriet, G., Bie, T. D., Cristianini, N., Jordan, M., & Noble, W. (2004a). A statistical framework for genomic data fusion. *Bioinformatics*, *20*, 2626–2635.

Lanckriet, G., Cristianini, N., Ghaoui, L. E., Bartlett, P., , & Jordan, M. (2004b). Learning the kernel matrix with semi-definite programming. *Journal of Machine Learning Research*, *5*, 27–72.

Lemaréchal, C., & Sagastizabal, C. (1997). Practical aspects of moreau-yosida regularization : theoretical preliminaries. *SIAM Journal of Optimization*, *7*, 867–895.

Micchelli, C., & Pontil, M. (2005). Learning the kernel function via regularization. *Journal of Machine Learning Research*, *6*, 1099–1125.

Scholkopf, B., & Smola, A. (2001). *Learning with kernels*. MIT Press.

Sonnenburg, S., Raetsch, G., Schaefer, C., & Scholkopf, B. (2006). Large scale multiple kernel learning. *Journal of Machine Learning Research*, *7*, 1531–1565.

Wahba, G. (1990). *Spline models for observational data*. Series in Applied Mathematics, Vol. 59, SIAM.