ERCIM "ALAIN BENSOUSSAN"
FELLOWSHIP PROGRAMME

ERCIM
European Research Consortium
for Informatics and Mathematics

# Scientific Report

| | |
|---|---|
| First name / Family name | Dmitrii Ostrovskii |
| Nationality | Russia |
| Name of the *Host Organisation* | INRIA |
| First Name / family name of the *Scientific Coordinator* | Francis Bach |
| Period of the fellowship | 01/02/2018 to 31/01/2019 |

## I – SCIENTIFIC ACTIVITY DURING YOUR FELLOWSHIP

My research activities during the period of the fellowship are related to four projects in statistical learning theory and numerical optimization. Below I give their brief overview.

**Fast learning rates through self-concordance.** In [J1], I investigated the question of how the *generalized self-concordance* of the loss can be exploited to obtain fast learning rates in M-estimation. Self-concordance was introduced by Nesterov and Nemirovski (1994) in the context of interior-point algorithms; a convex, and sufficiently smooth, loss is called self-concordant if its third derivative is upper-bounded with the 3/2 power of the second. I demonstrated that self-concordance and its extension introduced in Bach (2010) in the context of logistic regression, allows to quantify the generalization properties of the associated M-estimators with random design in a non-asymptotic way, and results in fast (and asymptotically optimal) rates of convergence for the excess risk. In the follow-up work [C1], the framework has been extended to the regularized M-estimation setting.

**Efficient algorithms for large-scale multiclass learning.** Together with coauthors, I study finite-sum optimization problems arising as the empirical risk objective in linear classification with very large number of classes and dimensionality of the feature space. Our focus is on so-called Fenchel-Young losses (Blondel, 2018) that can be represented as the maximum of a finite number of affine functions. This leads to well-structured bilinear saddle-point problems, which can be efficiently solved with certain stochastic primal-dual algorithms (Nesterov and A. Nemirovski, 2013) equipped with ad-hoc variance reduction techniques. Using this approach, in [C2] we propose a sublinear algorithm to train multiclass support vector machines – to our best knowledge, the first algorithm of this kind for multiclass linear classification. Extending this result to other losses is a potential direction for further research.

**Covariance estimation for heavy-tailed distributions.** Recently, Wei and Minsker (2017) proposed an estimator of the covariance matrix with a remarkable property: its deviations from the target, measured in the spectral norm, are sub-Gaussian under extremely weak moment assumptions on the underlying distribution. For the chosen criterion, this result is near-optimal. On the other hand, in some applications one is interested in relative error bounds, i.e., approximating the *regularized* covariance matrix, up to a multiplicative factor close to one, in the positive-semidefinite sense. Such guarantees for the estimator of Wei and Minsker do not hold, the reason being its non-linearity in observations. In the recent work [C3], we proposed and studied an estimator that admits such guarantees, while having essentially the same computational cost as the sample covariance one. We then considered its applications to noisy principal component analysis and random-design linear regression.

## II – PUBLICATION(S) DURING YOUR FELLOWSHIP

My work during the fellowship period resulted in four publications, all of which are currently under review: one in the leading journal on statistical theory, and three others submitted at two highly competitive machine learning conferences (acceptance rates around 25% and 40%, the revision process is double-blind and blind in the second case).

[J1] D. M. Ostrovskii and F. Bach. Finite-Sample Analysis of M-estimators using Self-Concordance. *In review at Annals of Statistics*. arXiv:1810.06838, Oct. 2018.

[C1] U. Marteau-Ferey, D. M. Ostrovskii, A. Rudi, and F. Bach. Beyond Least-Squares: Fast Rates for Regularized Empirical Risk Minimization through Self-Concordance. *In review*. arXiv:1902.03046, Feb. 2019.

[C2] D. Babichev, D. M. Ostrovskii, and F. Bach. Efficient Primal-Dual Algorithms for Large-Scale Multiclass Classification. *In review*. arXiv:1902.03755, Feb. 2019.

[C3] D. M. Ostrovskii and A. Rudi. Affine Invariant Covariance Estimation for Heavy-Tailed Distributions. *In review*. arXiv:1902.03086, Feb 2019.

## III – ATTENDED SEMINARS, WORKHOPS, CONFERENCES

International Conference on Machine Learning (ICML) 2018, Stockholm. Presented a paper (short oral).

CWI-INRIA Workshop, Paris. Presented my work during the fellowship related to the self-concordance project (30 min. talk).

## IV – RESEARCH EXCHANGE PROGRAMME (REP)

**Centrum Wiskunde & Informatica (CWI), Amsterdam**
Dates: 09.02.2018-22.02.2018
Peter Grünwald

While visiting CWI, I was introduced to the members of the theoretical machine learning group led by Prof. Grünwald. Our research interests have a significant overlap, which allowed us to discuss the ongoing projects, and impacted my progress in a positive way. I was also happy to present the results of my PhD thesis at a seminar talk at CWI, and those on the self-concordant project, on which I worked during the fellowship period, at the mutual CWI-INRIA Workshop at INRIA, Paris.

**References**

Y. Nesterov and A. Nemirovski. *Interior-point Polynomial Algorithms in Convex Programming*. Society of Industrial and Applied Mathematics, 1994.

F. Bach. Self-concordant analysis for logistic regression. *Electronic Journal of Statistics,* 4:384–414, 2010.

M. Blondel, A. F. Martins, and V. Niculae. Learning classifiers with Fenchel-Young losses: Generalized entropies, margins, and algorithms. *arXiv:1805.09717*, 2018.

Y. Nesterov and A. Nemirovski. On first-order algorithms for l1/nuclear norm minimization. *Acta Numerica*, 22:509–575, 5 2013.

X. Wei and S. Minsker. Estimation of the covariance structure of heavy-tailed distributions. In *Advances in Neural Information Processing Systems*, pages 2859–2868, 2017.