Adaptive Stochastic Approximation Algorithm - Additional Results

Milena Kresoja* Zorana Lužanin* Irena Stojkovska †

December 19, 2016

Additional numerical results of the Mean-Sigma Stochastic Approximation Algorithm are presented. Problems and testing procedure are the same as in the in the original manuscript, [1]. Each test consists of $N = 50$ independent runs starting from the same initial point. Run is considered successful if a method stops due to gradient tolerance. The number of successful runs is denoted by $Nconv$. If $||G_k|| > 200\sqrt{n}$, run is unsuccessful i.e. divergence is declared. The number of divergent runs is denoted by $Ndiv$. Finally, the runs that stop due to exhausting the maximal number of allowed function evaluations are considered partially successful and their number is denoted by $Npar$.

We analyze the performance and sensitivity of the Algorithm with respect to the width $2 \cdot \hat{\sigma}$ of interval

$$J_k = (\frac{1}{m(k)} \sum_{j=1}^{m(k)} F_{k-j} - \hat{\sigma}, \frac{1}{m(k)} \sum_{j=1}^{m(k)} F_{k-j} + \hat{\sigma}).$$

For the parameter $\hat{\sigma}$, three different values are chosen: $\hat{\sigma} = 0.1, 0.3, 0.6$. Other parameters used in the step size scheme are the following: $\theta = 0.999$, $m = 10$. Three different values for the noise level σ are tested, $\sigma = 0.01, 0.4, 1$.

The following notation is used:

- MSGD - Mean-Sigma Stochastic Approximation Algorithm with $d_k = -G_k$
- MSDD - Mean-Sigma Stochastic Approximation Algorithm with BFGS direction

Table 1a and Table 1b show the number of successful, partially successful and divergent runs for all levels of noise and all widths of the interval J_k.

*Department of Mathematics and Informatics, Faculty of Science, University of Novi Sad, Trg Dositeja Obradovića 4, 21000 Novi Sad, Serbia, e-mail: {milena.kresoja@dmi.uns.ac.rs, zorana@dmi.uns.ac.rs}

†Department of Mathematics, Faculty of Natural Sciences and Mathematics, Ss. Cyril and Methodius University, Arhimedova 3, 1000 Skopje, Macedonia, e-mail: {irenatra@pmf.ukim.mk}

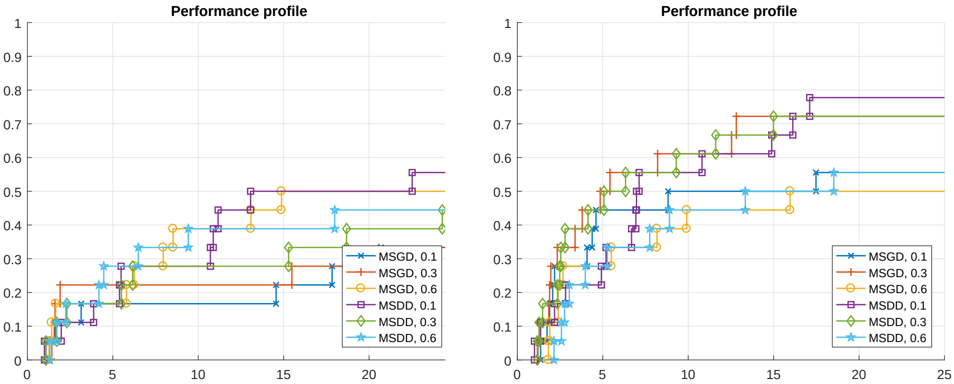
$m = 10, \theta = 0.99$	$\sigma = 0.01$			$\sigma = 0.4$		
Solver	$Nconv$	$Npar$	$Ndiv$	$Nconv$	$Npar$	$Ndiv$
MSGD, $\widehat{\sigma} = 0.1$	328	322	250	585	64	251
MSGD, $\widehat{\sigma} = 0.3$	294	406	200	628	32	240
MSGD, $\widehat{\sigma} = 0.6$	488	12	400	473	8	419
MSDD, $\widehat{\sigma} = 0.1$	421	373	106	562	207	131
MSDD, $\widehat{\sigma} = 0.3$	279	410	211	413	224	263
MSDD, $\widehat{\sigma} = 0.6$	352	438	110	384	363	153

Table **1a**: Number of successful, partially successful and divergent runs, $\sigma = 0.01, \sigma = 0.4$

$m = 10, \theta = 0.99$	$\sigma = 1$		
Solver	$Nconv$	$Npar$	$Ndiv$
MSGD, $\widehat{\sigma} = 0.1$	386	234	280
MSGD, $\widehat{\sigma} = 0.3$	485	167	248
MSGD, $\widehat{\sigma} = 0.6$	445	165	306
MSDD, $\widehat{\sigma} = 0.1$	439	271	190
MSDD, $\widehat{\sigma} = 0.3$	258	337	305
MSDD, $\widehat{\sigma} = 0.6$	275	369	256

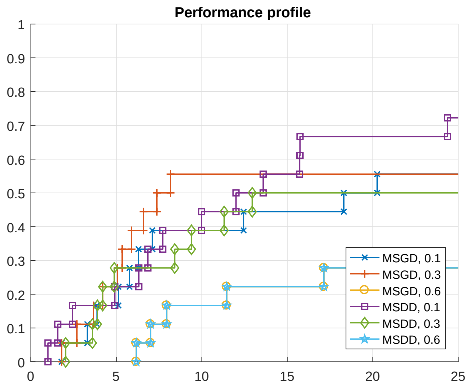
Table **1b**: Number of successful, partially successful and divergent runs, $\sigma = 1$

Figure 1a and Figure 1b show performance profiles for all levels of noise and all widths of the interval J_k. As a performance measure we use the number of function evaluations needed in successful and partially successful runs.



a. $\sigma = 0.01$ b. $\sigma = 0.4$

Figure **1a**: Performance profiles, $\sigma = 0.01, \sigma = 0.4$



$\sigma = 1$

Figure **1b**: Performance profiles, $\sigma = 1$

References

[1] Kresoja, M., Lužanin, Z., Stojkovska, I.: Adaptive stochastic approximation algorithm, Tech. Rep. (2016)