ERRATUM: HYBRID DETERMINISTIC-STOCHASTIC METHODS FOR DATA FITTING

MICHAEL P. FRIEDLANDER† AND MARK SCHMIDT‡


Key words. optimization, data fitting, incremental gradient, gradient descent

AMS subject classifications. 47N10, 90C06, 90C25, 94A20

DOI. 10.1137/130908257

The proof of Theorem 3.1 in Friedlander and Schmidt [1] is incorrect, though the theorem statement is correct. Here is a corrected proof. (Equation numbers in this erratum refer to those of the original paper.)

THEOREM 3.1 (weak linear rate with deterministic sampling). Suppose that (1.8) holds and that the sample size $|B_k|$ increases geometrically toward $M$, i.e.,

$$\frac{M - |B_k|}{M} = O(\gamma^{k/2})$$

for some $\gamma < 1$. Then at each iteration of algorithm (1.2) with $\alpha_k \equiv 1/L$, for any $\epsilon > 0$ we have

$$f(x_k) - f(x^*) = [f(x_0) - f(x^*)]O([1 - \mu/L + \epsilon]^k) + O(\sigma^k),$$

where $\sigma = \max\{\gamma, 1 - \mu/L\} + \epsilon$.

Proof. Let $\rho_k = \left(\frac{M - |B_k|}{M}\right)^2$. Using (3.2) and Lemma 2.1, we obtain the bound

$$f(x_{k+1}) - f(x^*) \leq (1 - \mu/L)[f(x_k) - f(x^*)] + \frac{2\rho_k}{L}(\beta_1 + 2\beta_2 L)[f(x_k) - f(x^*)]$$

$$= (1 - \mu/L + 4\beta_2 \rho_k)[f(x_k) - f(x^*)] + \frac{2\beta_1}{L}\rho_k$$

$$= \omega_k[f(x_k) - f(x^*)] + \frac{2\beta_1}{L}\rho_k,$$

where $\omega_k := 1 - \mu/L + 4\beta_2 \rho_k$. Apply this recursively and use $\rho_k = O(\gamma^k)$ to obtain

$$f(x_k) - f(x^*) \leq [f(x_0) - f(x^*)] \prod_{i=0}^{k-1} \omega_i + \sum_{i=0}^{k-1} O\left(\gamma^i \prod_{j=i+1}^{k-1} \omega_j\right).$$

*Submitted to the journal’s Computational Methods in Science and Engineering section February 1, 2013; accepted for publication March 25, 2013; published electronically August 30, 2013.

http://www.siam.org/journals/sisc/35-4/90825.html

†Department of Computer Science, University of British Columbia, Vancouver V6T 1Z4, BC, Canada (mpf@cs.ubc.ca).

‡INRIA-SIERRA team, Laboratoire d’Informatique de l’Ecole Normale Superieure (INRIA/ENS/CNRS UMR 8548), 23 avenue d’Italie, 75214 Paris CEDEX 13, France (mark.schmidt@inria.fr).
Take $\delta_k := \max\{\gamma, \omega_k\}$. Because $\gamma^i = \prod_{j=1}^i \gamma \leq \prod_{j=1}^i \delta_j$ for all $i$,
\[
f(x_k) - f(x_*) \leq [f(x_0) - f(x_*)] \prod_{i=0}^{k-1} \omega_i + \sum_{i=0}^{k-1} O \left( \prod_{j=1}^{k-1} \delta_j \right) = [f(x_0) - f(x_*)] \prod_{i=0}^{k-1} \omega_i + O \left( k \prod_{j=1}^{k-1} \delta_j \right).
\]
Because $\rho_k \to 0$, it follows that $\omega_k \to 1 - \mu/L$ and thus $\prod_{i=0}^{k} \omega_i = O(1 - \mu/L + \epsilon)^k$ for any $\epsilon > 0$, which bounds the first term in the right-hand side above. Furthermore, we now use the fact that $\delta_k \searrow \bar{\delta} := \max\{\gamma, 1 - \mu/L\}$ to show that the second term is in $O(\bar{\delta} + \epsilon)^k$ for any $\epsilon > 0$. In particular, choose $N$ large enough that $(\delta_k + \delta_k/k) \leq \bar{\delta} + \epsilon$ for all $k \geq N$ and choose the constant $\xi$ so that
\[
k \prod_{j=1}^{k-1} \delta_j \leq \xi(\bar{\delta} + \epsilon)^k
\]
for all $k < N$. Then by induction, $k \prod_{j=1}^{k-1} \delta_j = O(\sigma^k)$ for all $k$ because
\[
(k+1) \prod_{j=1}^{k} \delta_j = (\delta_k + \delta_k/k)k \prod_{j=1}^{k-1} \delta_j \leq (\delta_k + \delta_k/k)\xi(\bar{\delta} + \epsilon)^k \leq (\bar{\delta} + \epsilon)\xi(\bar{\delta} + \epsilon)^k = \xi(\bar{\delta} + \epsilon)^{k+1}.
\]

Acknowledgment. We thank Ming Yan for highlighting the mistake in the original proof.

REFERENCE