Supplement: Solving Ridge Regression using Sketched Preconditioned SVRG

Alon Gonen ALONGNN@CS.HUJI.AC.IL
The Hebrew University

Francesco Orabona FRANCESCO@ORABONA.COM
Yahoo Research, 229 West 43rd Street, 10036 New York, NY, USA

Shai Shalev-Shwartz SHAIS@CS.HUJI.AC.IL
The Hebrew University

1. Omitted Proofs

Proof. **(of Theorem 4)** We first show that the average smoothness of $\tilde{L}$ is bounded by

$$\frac{1}{n+d}\sum_{i=1}^{n+d}\tilde{\beta}_i \le \operatorname{tr}\left(P^{-1/2}\left(C+\lambda I\right)P^{-1/2}\right) . \qquad (1)$$

Note that for any w,

$$\nabla^2\tilde{\ell}_i(w) = \begin{cases} \frac{n+d}{n}\tilde{x}_i\tilde{x}_i^\top & 1 \le i \le n, \\ \lambda(n+d)b_{i-n}b_{i-n}^\top & n < i \le n+d\,. \end{cases}$$

Therefore, using the fact that the spectral norm of a rank-1 psd matrix is equal to its trace, we obtain

$$\begin{aligned}
\frac{1}{n+d}\sum_{i=1}^{n}\tilde{\beta}_i &= \frac{1}{n+d}\frac{n+d}{n}\sum_{i=1}^{n}\|\tilde{x}_i\tilde{x}_i^\top\| + \frac{1}{n+d}\lambda(n+d)\sum_{j=1}^{d}\|b_jb_j^\top\| \\
&= \frac{1}{n}\sum_{i=1}^{n}\operatorname{tr}(\tilde{x}_i\tilde{x}_i^\top) + \lambda\sum_{j=1}^{d}\operatorname{tr}(b_ib_i^\top) \\
&= \frac{1}{n}\operatorname{tr}(\sum_{i=1}^{n}P^{-1/2}x_ix_i^\top P^{-1/2}) + \lambda\operatorname{tr}\sum_{j=1}^{d}(P^{-1/2}e_ie_i^\top P^{-1/2}) \\
&= \operatorname{tr}(P^{-1/2}(C+\lambda I)P^{-1/2})\,.
\end{aligned}$$

Hence, we deduce (1).

We will conclude the theorem by showing that $\tilde{L}$ is $\lambda_d(P^{-1/2}(C+\lambda I)P^{-1/2})$-strongly convex. Indeed, a similar calculation shows that the Hessian of L at any point w is given by

$$\nabla^2\tilde{L}(w) = P^{-1/2}(C+\lambda I)P^{-1/2}\,.$$

Hence, we conclude the claimed bound. □