Compressed Least-Squares Regression (Supplementary material)

Odalric-Ambrym Maillard and **Rémi Munos**
SequeL Project, INRIA Lille - Nord Europe, France
{odalric.maillard, remi.munos}@inria.fr

1 Proof of Proposition 1

Proposition 1 *Let $(u_k)_{1\leq k\leq K}$ and v be vectors of $\mathbb{R}^N$. Let A be a $M \times N$ matrix of i.i.d. elements drawn from one of the previously defined distributions. For any $\varepsilon > 0$, $\delta > 0$, for $M \geq \frac{1}{\frac{\varepsilon^2}{4}-\frac{\varepsilon^3}{6}} \log \frac{4K}{\delta}$, we have, with probability at least $1-\delta$, for all $k \leq K$,*

$$|Au_k \cdot Av - u_k \cdot v| \leq \varepsilon ||u_k||\,||v||.$$

Proof: We make use of the following lemma, which states that the random (with respect to the choice of the matrix A) variable $||Au||^2$ concentrates around its expectation $||u||^2$ when M is large. The proof uses concentration inequalities (Cramer's large deviation Theorem) and may be found e.g. in [1].

Lemma 1 *For any vector u in $\mathbb{R}^N$ and any $\varepsilon \in (0,1)$, we have*

$$\begin{aligned}
\mathbb{P}\Big(||Au||^2 \geq (1+\varepsilon)||u||^2\Big) &\leq e^{-M(\varepsilon^2/4-\varepsilon^3/6)}\\
\mathbb{P}\Big(||Au||^2 \leq (1-\varepsilon)||u||^2\Big) &\leq e^{-M(\varepsilon^2/4-\varepsilon^3/6)}
\end{aligned}$$

To prove the proposition, we apply the lemma to any couple of vectors $u+w$ and $u-w$, where u and w are vectors of norm 1. From the parallelogram law, we have that

$$\begin{aligned}
4Au\cdot Aw &= ||Au+Aw||^2 - ||Au-Aw||^2\\
&\leq (1+\varepsilon)||u+w||^2 - (1-\varepsilon)||u-w||^2\\
&= 4u\cdot w + \varepsilon(||u+w||^2 + ||u-w||^2)\\
&= 4u\cdot w + 2\varepsilon(||u||^2+||w||^2) = 4u\cdot w + 4\varepsilon.
\end{aligned}$$

fails with probability $2e^{-M(\varepsilon^2/4-\varepsilon^3/6)}$ (we applied the previous lemma twice at line 2).

Thus for each $k \leq K$, we have with same probability:

$$Au_k \cdot Av \leq u_k \cdot v + \varepsilon||u_k||\,||v||.$$

Now the symmetric inequality holds with the same probability, and using a union bound for considering all $(u_k)_{k\leq K}$, we have that

$$|Au_k \cdot Av - u_k \cdot v| \leq \varepsilon||u_k||\,||v||,$$

holds for all $k \leq K$, with probability $1-4Ke^{-M(\varepsilon^2/4-\varepsilon^3/6)}$, and the proposition follows. □

2 Proof of Proposition 2

Proposition 2 *Assume that f^* is (L,γ)-Lipschitz (i.e. for all $v \in \mathcal{X}$ there exists a polynomial p_v of degree $\lfloor\gamma\rfloor$ such that for all $u \in \mathcal{X}$, $|f(u) - p_v(u)| \leq L|u-v|^{\gamma}$) with $1/2 < \gamma \leq p$. Then setting $c_h = 2^{h(1-2\gamma)/4}$, we have $||\alpha^+|| \sup_x ||\varphi(x)|| \leq L\frac{2^{\gamma}}{1-2^{1/2-\gamma}}\int_0^1 |\varphi_0|$, which is independent of N.*

Proof: f^+ decomposes as $f^+ = \sum_{1\leq h\leq H}\sum_l \alpha^0_{h,l}\varphi^0_{h,l} = \sum_{1\leq h\leq H}\sum_l (\alpha^0_{h,l}2^{h/2}c_h^{-1})\varphi_{h,l}$. By Theorem 6.3 of [2], since f^* is (L,γ)-Lipschitz and φ^0 has at least $p \geq \gamma$ vanishing moments, we have $|\alpha^0_{h,l}| \leq L2^{\gamma}\int_0^1|\varphi_0|2^{-h(\gamma+1/2)}$. Thus we deduce:

$$||\alpha^+||^2 = \sum_{1\leq h\leq H}\sum_l \alpha^2_{h,l} \leq \big(L2^{\gamma}\int_0^1|\varphi_0|\big)^2 \sum_{1\leq h\leq H} 2^h 2^{-2h\gamma}c_h^{-2}$$

and

$$||\varphi(x)||^2 = \sum_{1\leq h\leq H}\sum_l c_h^2 2^{-h}(\varphi^0_{h,l}(x))^2 \leq \sum_{1\leq h\leq H} c_h^2$$

Thus, setting $c_h = 2^{(1-2\gamma)h/4}$, we deduce

$$||\alpha^+||^2 \sup_x ||\varphi(x)||^2 \leq \big(L2^{\gamma}\int_0^1|\varphi_0|\big)^2\big(1-2^{1/2-\gamma}\big)^{-2}$$

since $\gamma > 1/2$. □

References

[1] Dimitris Achlioptas. Database-friendly random projections: Johnson-Lindenstrauss with binary coins. *Journal of Computer and System Sciences*, 66(4):671–687, June 2003.

[2] Stephane Mallat. *A Wavelet Tour of Signal Processing*. Academic Press, 1999.