Erratum: Local Identification of Overcomplete Dictionaries

Karin Schnass karin.schnass@uibk.ac.at
University of Innsbruck
Technikerstraße 13
6020 Innsbruck, Austria

Editor: Francis Bach

The proof of Proposition 7 is incorrect, due to the wrong assumption that the set Σ_p is symmetric in the sense that either $(\sigma_1, \ldots, \pm\sigma_i, \ldots, \sigma_K) \in \Sigma_p$ or $(\sigma_1, \ldots, \pm\sigma_i, \ldots, \sigma_K) \notin \Sigma_p$. As Sebastian Kaiser and Felix Krahmer pointed out, if Σ_p had the described property, by iteratively flipping the signs we could reach all sign sequences so either Σ_p would be the set of all sign sequences or the empty set. A corrected proof can be found below:

Proof To prove the proposition we digress from the conventional scheme of first calculating the expectation of our objective function for both the original and a perturbed dictionary and then comparing and instead bound the difference of the expectations directly.

$$\begin{aligned}
&\mathbb{E}_y\left(\max_{|I|=S}\|\Phi_I^\star y\|_1\right) - \mathbb{E}_y\left(\max_{|I|=S}\|\Psi_I^\star y\|_1\right)\\
&\quad= \mathbb{E}_{p,\sigma,r}\left(\max_{|I|=S}\left\|\frac{\Phi_I^\star(\Phi c_{p,\sigma}+r)}{\sqrt{1+\|r\|_2^2}}\right\|_1 - \max_{|I|=S}\left\|\frac{\Psi_I^\star(\Phi c_{p,\sigma}+r)}{\sqrt{1+\|r\|_2^2}}\right\|_1\right)\\
&\quad= \mathbb{E}_{p,\sigma,r}\left(\frac{\max_{|I|=S}\|\Phi_I^\star(\Phi c_{p,\sigma}+r)\|_1 - \max_{|I|=S}\|\Psi_I^\star(\Phi c_{p,\sigma}+r)\|_1}{\sqrt{1+\|r\|_2^2}}\right) := \mathbb{E}_{p,\sigma,r}(\Delta_{p,\sigma,r})
\end{aligned}$$

Again our strategy is to show that for a fixed p for most σ and r the maximal response of both the original dictionary and the perturbation is attained at I_p. The expressions we therefore need to lower (upper) bound for $i \in I_p$ ($i \notin I_p$) are

$$\begin{aligned}
|\langle\phi_i, \Phi c_{p,\sigma}+r\rangle| &= \Big|\sigma_i c_{p(i)} + \sum_{j\neq i}\sigma_j c_{p(j)}\langle\phi_i,\phi_j\rangle + \langle\phi_i, r\rangle\Big|,\\
&= \Big|c_{p(i)} + \sigma_i\sum_{j\neq i}\sigma_j c_{p(j)}\langle\phi_i,\phi_j\rangle + \sigma_i\langle\phi_i, r\rangle\Big|,\\
|\langle\psi_i, \Phi c_{p,\sigma}+r\rangle| &= \Big|\alpha_i\sigma_i c_{p(i)} + \alpha_i\sum_{j\neq i}\sigma_j c_{p(j)}\langle\phi_i,\phi_j\rangle + \omega_i\langle z_i, \Phi c_{p,\sigma}\rangle + \langle\psi_i, r\rangle\Big|\\
&= \Big|\alpha_i c_{p(i)} + \sigma_i\alpha_i\sum_{j\neq i}\sigma_j c_{p(j)}\langle\phi_i,\phi_j\rangle + \sigma_i\omega_i\langle z_i, \Phi c_{p,\sigma}\rangle + \sigma_i\langle\psi_i, r\rangle\Big|.
\end{aligned}$$

However, instead of using a worst case estimate for the gap between the responses of the original dictionary within and without I_p, we now make use of the fact that for most sign sequences we have a gap size of order $c_S - c_{S+1}$. This means that as soon as $|\sum_{j\neq i}\sigma_j c_{p(j)}\langle\phi_i,\phi_j\rangle|$, $\omega_i|\langle z_i,\Phi c_{p,\sigma}\rangle|$ and the noise related terms $|\langle\phi_i,r\rangle|$ and $|\langle\psi_i,r\rangle|$ are of order (c_S-c_{S+1}) the maximal response of both the original dictionary and the perturbation is attained at I_p. In particular, setting $\delta_p(i)=-1$ for $i\in I_p$ and $\delta_p(i)=1$ for $i\notin I_p$ defining the sets,

$$\Sigma_p := \bigcup_i \left\{ \sigma \text{ s.t. } \sigma_i\delta_p(i)\sum_{j\neq i}\sigma_j c_{p(j)}\langle\phi_i,\phi_j\rangle \geq \frac{c_S-c_{S+1}}{6} \right.$$
$$\left. \text{or } \sigma_i\delta_p(i)\omega_i\langle z_i,\Phi c_{p,\sigma}\rangle \geq \frac{c_S-c_{S+1}-\frac{3\varepsilon^2}{2}}{6}\right\},$$

for a fixed permutation p and

$$R := \bigcup_i \left\{ r \text{ s.t. } |\langle\phi_i,r\rangle| \geq \frac{c_S-c_{S+1}}{3} \text{ or } |\langle\psi_i,r\rangle| \geq \frac{c_S-c_{S+1}}{6}\right\},$$

we see that both maxima are attained at I_p as long as $\sigma\notin\Sigma_p$ and $r\notin R$. Using Hoeffding's inequality we get that

$$\mathbb{P}\left(\sigma_i\delta_p(i)\sum_{j\neq i}\sigma_j c_{p(j)}\langle\phi_i,\phi_j\rangle > t\right) \leq \exp\left(\frac{-t^2}{2\sum_{j\neq i}c_{p(j)}^2|\langle\phi_i,\phi_j\rangle|^2}\right) \leq \exp\left(\frac{-t^2}{2\mu^2}\right),$$

and similarly (compare also Proposition 6) we get that for $\varepsilon_i \neq 0$

$$\mathbb{P}(\sigma_i\delta_p(i)\omega_i\langle z_i,\Phi c_{p,\sigma}\rangle \geq s) \leq \exp\left(\frac{-s^2}{2\varepsilon_i^2}\right).$$

Setting $t=(c_S-c_{S+1})/6$, $s=(c_S-c_{S+1}-\frac{3\varepsilon^2}{2})/6$ and using a union bound then leads to

$$\mathbb{P}(\Sigma_p) \leq K\exp\left(-\frac{\left(c_S-c_{S+1}-\frac{3\varepsilon^2}{2}\right)^2}{72\varepsilon^2}\right) + K\exp\left(\frac{-(c_S-c_{S+1})^2}{72\mu^2}\right). \tag{1}$$

Since the $r(i)$ are subgaussian with parameter ρ we have for any $v=(v_1\dots v_d)$ and $t\geq 0$, $\mathbb{P}(|\langle v,r\rangle|\geq t)\leq \exp\left(-\frac{t^2}{2\rho^2\|v\|_2^2}\right)$, see e.g. [31]. Taking a union bound over all ϕ_i,ψ_i with the corresponding choice for t then leads to the estimate

$$\mathbb{P}(R) \leq 2K\exp\left(-\frac{(c_S-c_{S+1})^2}{72\rho^2}\right) + 2K\exp\left(-\frac{(c_S-c_{S+1})^2}{18\rho^2}\right). \tag{2}$$

We now split the expectations over the sign and noise patterns for a fixed p to get

$$\begin{aligned}\mathbb{E}_{\sigma,r}(\Delta_{p,\sigma,r}) &= \int_{r\notin R}\mathbb{E}_\sigma\left(\Delta_{p,\sigma,r}\right)d\nu_r + \int_{r\in R}\mathbb{E}_\sigma\left(\Delta_{p,\sigma,r}\right)d\nu_r \\ &= \int_{r\notin R}\Big(\sum_{\sigma\notin\Sigma_p}\mathbb{P}(\sigma)\Delta_{p,\sigma,r}\Big)d\nu_r + \int_{r\notin R}\Big(\sum_{\sigma\in\Sigma_p}\mathbb{P}(\sigma)\Delta_{p,\sigma,r}\Big)d\nu_r \\ &\qquad + \mathbb{E}_\sigma\left(\int_{r\in R}\Delta_{p,\sigma,r}d\nu_r\right).\end{aligned} \tag{3}$$

We start by bounding $\sum_{\sigma\notin\Sigma_p}\Delta_{p,\sigma,r}$ for a fixed $r\notin R$. We have

$$\begin{aligned}
\sum_{\sigma\notin\Sigma_p}\Delta_{p,\sigma,r} &= \sum_{\sigma\notin\Sigma_p}\mathbb{P}(\sigma)\left(\frac{\left\|\Phi^\star_{I_p}(\Phi c_{p,\sigma}+r)\right\|_1-\left\|\Psi^\star_{I_p}(\Phi c_{p,\sigma}+r)\right\|_1}{\sqrt{1+\|r\|_2^2}}\right)\\
&=(1+\|r\|_2^2)^{-\frac12}\sum_{\sigma\notin\Sigma_p}\sum_{i\in I_p}|\langle\phi_i,\Phi c_{p,\sigma}+r\rangle|-|\langle\psi_i,\Phi c_{p,\sigma}+r\rangle|\\
&=(1+\|r\|_2^2)^{-1/2}\sum_{i\in I_p}\sum_{\sigma\notin\Sigma_p}\sigma_i\langle\phi_i,\Phi c_{p,\sigma}+r\rangle-\sigma_i\langle\psi_i,\Phi c_{p,\sigma}+r\rangle\\
&=(1+\|r\|_2^2)^{-\frac12}\sum_{i\in I_p}\sum_{\sigma\notin\Sigma_p}(1-\alpha_i)|\langle\phi_i,\Phi c_{p,\sigma}+r\rangle|-\sigma_i\omega_i\langle z_i,\Phi c_{p,\sigma}+r\rangle.
\end{aligned}$$

Define σ^i via setting $\sigma^i_i=-\sigma_i$ and $\sigma^i_j=\sigma_j$ for $j\neq i$. For every index $i\in I_p$ there are two types of sequences $\sigma\notin\Sigma_p$, those were $\sigma^i\notin\Sigma_p$ and those were $\sigma^i\in\Sigma_p$. Since $\omega_i\langle z_i,\Phi c_{p,\sigma}+r\rangle$ is independent of σ_i, in the sum over the first type of sequences the term is scaled once with σ_i and once with $\sigma^i_i=-\sigma_i$ and therefore cancels out. This leads to

$$\begin{aligned}
\sum_{\sigma\notin\Sigma_p}\Delta_{p,\sigma,r} &= (1+\|r\|_2^2)^{-\frac12}\sum_{i\in I_p}\left(\sum_{\sigma\notin\Sigma_p}\frac{\varepsilon_i^2}{2}|\langle\phi_i,\Phi c_{p,\sigma}+r\rangle|-\sum_{\sigma\notin\Sigma_p:\sigma^i\in\Sigma_p}\sigma_i\omega_i\langle z_i,\Phi c_{p,\sigma}+r\rangle\right)\\
&\geq(1+\|r\|_2^2)^{-\frac12}\sum_{i\in I_p}\left(\sum_{\sigma\notin\Sigma_p}\frac{\varepsilon_i^2}{2}|\langle\phi_i,\Phi c_{p,\sigma}+r\rangle|-\sum_{\sigma\notin\Sigma_p:\sigma^i\in\Sigma_p}\varepsilon_i\|\Phi c_{p,\sigma}+r\|_2\right)\\
&\geq(1+\|r\|_2^2)^{-\frac12}\sum_{i\in I_p}\left(\sum_{\sigma\notin\Sigma_p}\frac{\varepsilon_i^2}{2}|\langle\phi_i,\Phi c_{p,\sigma}+r\rangle|-\sum_{\sigma\in\Sigma_p}\varepsilon(\sqrt{B}+\|r\|_2)\right)\\
&\geq(1+\|r\|_2^2)^{-\frac12}\left(\sum_{\sigma\notin\Sigma_p}\sum_{i\in I_p}\frac{\varepsilon_i^2}{2}|\langle\phi_i,\Phi c_{p,\sigma}+r\rangle|-\sum_{\sigma\in\Sigma_p}\varepsilon S(\sqrt{B}+\|r\|_2)\right).
\end{aligned}\tag{4}$$

Before substituting this bound into (3) we develop a general bound on $\Delta_{p,\sigma,r}$. We have

$$\begin{aligned}
\max_{|I|=S}\|\Psi^\star_I(\Phi c_{p,\sigma}+r)\|_1 &= \max_{|I|=S}\sum_{i\in I}|\langle\alpha_i\phi_i+\omega_i z_i,\Phi c_{p,\sigma}+r)\rangle|\\
&\leq\max_{|I|=S}\sum_{i\in I}\big(1-\frac{\varepsilon_i^2}{2}\big)|\langle\phi_i,\Phi c_{p,\sigma}+r)\rangle|+\varepsilon_i\|\Phi c_{p,\sigma}+r\|_2\\
&\leq\max_{|I|=S}\sum_{i\in I}\big(1-\frac{\varepsilon^2}{2}\big)|\langle\phi_i,\Phi c_{p,\sigma}+r)\rangle|+\varepsilon\big(\sqrt{B}+\|r\|_2\big)\\
&=\big(1-\frac{\varepsilon^2}{2}\big)\max_{|I|=S}\|\Phi^\star_I(\Phi c_{p,\sigma}+r)\|_1+\varepsilon S\big(\sqrt{B}+\|r\|_2\big),
\end{aligned}$$

which immediately leads to the lower bound

$$\begin{aligned}\Delta_{p,\sigma,r} &\geq (1+\|r\|_2^2)^{-\frac{1}{2}} \left(\max_{|I|=S} \|\Phi_I^\star(\Phi c_{p,\sigma}+r)\|_1 \frac{\varepsilon^2}{2} - \varepsilon S\big(\sqrt{B}+\|r\|_2\big)\right)\\ &\geq (1+\|r\|_2^2)^{-\frac{1}{2}} \left(\|\Phi_{I_p}^\star(\Phi c_{p,\sigma}+r)\|_1 \frac{\varepsilon^2}{2} - \varepsilon S\big(\sqrt{B}+\|r\|_2\big)\right)\\ &\geq (1+\|r\|_2^2)^{-\frac{1}{2}} \left(\sum_{i\in I_p} \frac{\varepsilon_i^2}{2} |\langle \phi_i, \Phi c_{p,\sigma}+r\rangle| - \varepsilon S\big(\sqrt{B}+\|r\|_2\big)\right).\end{aligned}$$

Substituting the estimate above together with (4) into (3) we get

$$\begin{aligned}\mathbb{E}_{\sigma,r}(\Delta_{p,\sigma,r}) \geq &\int_{r\notin R} \frac{\mathbb{P}(\sigma)}{\sqrt{1+\|r\|_2^2}} \left(\sum_{\sigma\notin\Sigma_p}\sum_{i\in I_p} \frac{\varepsilon_i^2}{2} |\langle \phi_i, \Phi c_{p,\sigma}+r\rangle| - \sum_{\sigma\in\Sigma_p} \varepsilon S(\sqrt{B}+\|r\|_2)\right) d\nu_r\\ &+ \int_{r\notin R} \frac{\mathbb{P}(\sigma)}{\sqrt{1+\|r\|_2^2}} \sum_{\sigma\in\Sigma_p} \left(\sum_{i\in I_p} \frac{\varepsilon_i^2}{2} |\langle \phi_i, \Phi c_{p,\sigma}+r\rangle| - \varepsilon S\big(\sqrt{B}+\|r\|_2\big)\right) d\nu_r\\ &+ \mathbb{E}_\sigma \left(\int_{r\in R} \frac{1}{\sqrt{1+\|r\|_2^2}} \left(\sum_{i\in I_p} \frac{\varepsilon_i^2}{2} |\langle \phi_i, \Phi c_{p,\sigma}+r\rangle| - \varepsilon S\big(\sqrt{B}+\|r\|_2\big)\right) d\nu_r \right),\end{aligned}$$

which after collecting all the terms, we can further bound as

$$\begin{aligned}\mathbb{E}_{\sigma,r}(\Delta_{p,\sigma,r}) &\geq \mathbb{E}_{\sigma,r} \left(\frac{\sum_{i\in I_p} \frac{\varepsilon_i^2}{2} |\langle \phi_i, \Phi c_{p,\sigma}+r\rangle|}{\sqrt{1+\|r\|_2^2}} \right)\\ &\qquad - 2\mathbb{P}(\Sigma_p) \int_{r\notin R} \frac{\varepsilon S\big(\sqrt{B}+\|r\|_2\big)}{\sqrt{1+\|r\|_2^2}} d\nu_r - \int_{r\in R} \frac{\varepsilon S\big(\sqrt{B}+\|r\|_2\big)}{\sqrt{1+\|r\|_2^2}} d\nu_r\\ &\geq \mathbb{E}_r \left(\frac{\sum_{i\in I_p} c_{p(i)} \frac{\varepsilon_i^2}{2}}{\sqrt{1+\|r\|_2^2}} \right) - \varepsilon S\big(\sqrt{B}+1\big) \cdot \big(2\mathbb{P}(\Sigma_p)+\mathbb{P}(R)\big). \qquad (5)\end{aligned}$$

Taking the expectation over the permutations then yields

$$\begin{aligned}\mathbb{E}_{p,\sigma,r}(\Delta_{p,\sigma,r}) &\geq \mathbb{E}_r\mathbb{E}_p \left(\frac{\sum_{i\in I_p} c_{p(i)} \frac{\varepsilon_i^2}{2}}{\sqrt{1+\|r\|_2^2}} \right) - \varepsilon S\big(\sqrt{B}+1\big) \cdot \big(2\mathbb{E}_p\mathbb{P}(\Sigma_p)+\mathbb{P}(R)\big)\\ &\geq \mathbb{E}_r \left(\frac{1}{\sqrt{1+\|r\|_2^2}} \right) \frac{c_1+\ldots+c_S}{2K} \sum_i \varepsilon_i^2 - \varepsilon S\big(\sqrt{B}+1\big) \cdot \big(2\mathbb{E}_p\mathbb{P}(\Sigma_p)+\mathbb{P}(R)\big).\end{aligned}$$

Using the probability estimates from (1)/(2) we see that $\mathbb{E}_{p,\sigma,r}(\Delta_{p,\sigma,r}) > 0$ is implied by

$$\varepsilon \geq \frac{4SK^2\big(\sqrt{B}+1\big)}{C_r\gamma} \left(\exp\left(\frac{-\big(\beta - \frac{3\varepsilon^2}{2}\big)^2}{72\varepsilon^2} \right) + \exp\left(\frac{-\beta^2}{72\mu^2}\right) + \exp\left(\frac{-\beta^2}{72\rho^2}\right) + \exp\left(\frac{-\beta^2}{18\rho^2}\right) \right),$$

where we have used the abbreviations $\gamma = c_1 + \ldots + c_S$, $\beta = c_S - c_{S+1}$ and $C_r = \mathbb{E}_r\left((1+\|r\|_2^2)^{-1/2}\right)$. We now proceed by splitting the above condition. We define $\varepsilon_{\min}$ by asking that

$$\frac{\varepsilon}{3} \geq \frac{4SK^2\left(\sqrt{B}+1\right)}{C_r\gamma}\exp\left(-\frac{\beta^2}{72\max\{\mu^2,\rho^2\}}\right) := \frac{\varepsilon_{\min}}{3}$$

and $\varepsilon_{\max}$ implicitely by asking that

$$\frac{\varepsilon}{3} - \frac{\varepsilon^4}{81} \geq \frac{4SK^2\left(\sqrt{B}+1\right)}{C_r\gamma}\exp\left(-\frac{\left(\beta - \frac{3\varepsilon^2}{2}\right)^2}{72\varepsilon^2}\right).$$

Following the line of argument in the proof of Proposition 6, we see that the above condition is guaranteed as soon as

$$\varepsilon \leq \frac{\beta}{\frac{5}{2} + 9\sqrt{\log\left(\frac{112K^2S(\sqrt{B}+1)}{C_r\beta\gamma}\right)}} := \varepsilon_{\max}.$$

The statement follows from making sure that $\varepsilon_{\min} < \varepsilon_{\max}$. ∎

Acknowledgments

I would like to thank Sebastian Kaiser and Felix Krahmer for pointing out the error.