David Gold April 22, 2017
STAT 572 — Questions for mentor (Adrian Dobra)

The following questions concern the motivation and examples that [1] give for their condition ASM on page 3-7. We recall their set-up. Suppose for $i \in 1, \ldots, n$ that

$$y_i \;=\; f(\boldsymbol{z}_i) + \epsilon_i,$$

where $y_i \in \mathbb{R}, \boldsymbol{z}_i \in \mathbb{R}^{p_{\boldsymbol{z}}}$, and $\epsilon_i \sim_{iid} \mathrm{N}(0, \sigma^2)$. Let $\boldsymbol{x}_i = P(\boldsymbol{z}_i) = \big(P_1(\boldsymbol{z}_i), \ldots, P_{p_{\boldsymbol{x}}}(\boldsymbol{z}_i)\big) \in \mathbb{R}^{p_{\boldsymbol{x}}}$ be a vector-valued transformation of $\boldsymbol{z}_i$. Say that $f(\boldsymbol{z})$ is *approximately sparse* if there exists a $\boldsymbol{\beta} \in \mathbb{R}^{p_{\boldsymbol{x}}}$ so that $s = \|\boldsymbol{\beta}\|_0$ (that is, s is the number of non-zero elements of $\boldsymbol{\beta}$) and

$$r_i \;:=\; f(\boldsymbol{z}_i) - \langle \boldsymbol{x}_i, \boldsymbol{\beta} \rangle$$

satisfy

1. $c_s := \big[n^{-1} \sum_i (r_i)^2\big]^{1/2} \leq K\sigma\sqrt{s/n}$ for a universal constant K;
2. $s \equiv s(n) = o(n/\log p_{\boldsymbol{x}})$.

(All sums taken over the index i are evaluated over $i \in 1, \ldots, n$ unless noted otherwise.)

1. The authors mention one can identify such a $\boldsymbol{\beta}$ as the solution to the optimization problem

$$\boldsymbol{\beta} \;\in\; \underset{\boldsymbol{b} \in \mathbb{R}^{p_{\boldsymbol{x}}}}{\arg\min} \Bigg\{ \underbrace{\frac{1}{n} \sum_i \big(\underbrace{f(\boldsymbol{z}_i) - \langle \boldsymbol{x}_i, \boldsymbol{b} \rangle}_{r_i} \big)^2}_{c_s^2} + \sigma^2 \|\boldsymbol{b}\|_0 / n \Bigg\}. \tag{0.1}$$

They mention that, in "common nonparametric problems", the optimal solution $\boldsymbol{\beta}$ balances the order of "approximation error" c_s^2 and the order of the "estimation error" $\sigma^2\|\boldsymbol{b}\|_0/n$. Why is this the case? Is it actually a condition that the solution must satisfy this "balancing", or is this a restriction imposed by the authors? If the former, where can one find a reference for this phenomenon?

2. One example of a "nonparametric problem" that [1] study (page 6-7) is that of an f for which a Fourier expansion $f(\boldsymbol{z}) = \sum_{j=1}^{\infty} \delta_j P_j(\boldsymbol{z}_i)$, where the $P_j(\boldsymbol{z})$, $j \in \mathbb{N}$ are orthonormal basis functions on $\mathcal{Z} = [0,1]^{p_{\boldsymbol{z}}}$ and the $\boldsymbol{z}_i$ are assumed distributed uniformly on $\mathcal{Z}$. If one assumes that the Fourier coefficients decay as $\delta_j \propto j^{-\nu}$, where ν is a smoothness measure of f, and one considers $\boldsymbol{\beta} = (\delta_j)_{j=1}^d$ and $\boldsymbol{x}(\boldsymbol{z}) = (P_1(\boldsymbol{z}), \ldots, P_d(\boldsymbol{z}))$, then c_s as defined above satisfies

$$c_s \;=\; O_P\big(\sqrt{\mathrm{E}[f(\boldsymbol{z}) - \langle \boldsymbol{x}, \boldsymbol{\beta} \rangle]}\big), \qquad\qquad \sqrt{\mathrm{E}[f(\boldsymbol{z}) - \langle \boldsymbol{x}, \boldsymbol{\beta} \rangle]} \;\lesssim\; d^{1/2-\nu}.$$

(I have their K to a d since the authors use K previously for the different, conflicting purpose identified above.) The authors state that "Balancing the order $d^{1/2-\nu}$ of approximation error with the order $\sqrt{d/n}$ of the estimation error gives the oracle-rate-optimal number of series terms $s = d \propto n^{1/(2\nu)}$. My first question is: is this an instance where the authors are citing the fact that the optimal solution to (0.1) must balance the orders of the two terms, or are they imposing that restriction themselves and then deriving conditions on the relevant quantities from that self-imposed restriction? My second question is: is it even true that $\boldsymbol{\beta}$ in this instance satisfies (0.1)? If so, how does one show this? Finally, why do the authors consider random $\boldsymbol{z}_i$ in this example but explicitly consider the $\boldsymbol{z}_i$ (and hence the $\boldsymbol{x}_i$) fixed in their definition of condition ASM (essentially the definition of approximate sparsity above) on page 3? Is this purposeful, or sloppiness?

References

[1] Alexandre Belloni, Victor Chernozhukov, and Christian Hansen. Inference for high-dimensional sparse econometric models. *Advances in Economics and Econometrics*, 2011.