Discussion of *Sure independence screening for ultrahigh dimensional feature space* by Fan and Lv

Richard Samworth†
University of Cambridge, UK

I would like to congratulate the authors for a very interesting and timely contribution to an important problem. The power of the methodology is well demonstrated and there are many opportunities to explore its possible extensions beyond the linear model. Most of the technical conditions required for their theoretical results are natural and interpretable. An exception is the concentration property imposed on the $n \times p$ matrix $\mathbf{Z}$ in (16), namely that there exist $c, c_1 > 1$ and $C_1 > 0$ such that

$$P\{\lambda_{\max}(\tilde{p}^{-1}\tilde{\mathbf{Z}}\tilde{\mathbf{Z}}^T) > c_1 \text{ or } \lambda_{\min}(\tilde{p}^{-1}\tilde{\mathbf{Z}}\tilde{\mathbf{Z}}^T) < 1/c_1\} \leq \exp(-C_1 n)$$

for any $n \times \tilde{p}$ submatrix $\tilde{\mathbf{Z}}$ of $\mathbf{Z}$ with $cn \leq \tilde{p} \leq p$. The authors prove this condition is satisfied when the entries of $\mathbf{Z}$ are independent standard normal random variables and we now study their conjecture that it holds for a wide class of spherically symmetric distributions.

The condition is expected to be most restrictive when $\tilde{p}$ is at the lower end of the range, so in the simulations below I took $\tilde{p} = cn$ with $c = 2$. To compare with the case of independent Gaussian entries, the rows of $\tilde{\mathbf{Z}}$ were taken to be independent and each row was generated as $a\mathbf{Y}$, where the distribution of $\mathbf{Y}$ was multivariate-t with $\nu = 10$ and $\nu = 20$ degrees of freedom, and a was such that each component had unit variance. Figures 1(a) and (b) plot estimated densities of the condition number of $\tilde{p}^{-1}\tilde{\mathbf{Z}}\tilde{\mathbf{Z}}^T$ (i.e. the ratio of the largest and smallest eigenvalues) when $n = 100$ and $n = 1000$ respectively. Notice that in the Gaussian case, the condition number density becomes more concentrated as n increases. For the multivariate-t case with $\nu = 20$, the distribution is rather similar at both sample sizes, whereas when $\nu = 10$ it is more dispersed with a long right tail at the larger sample size, suggesting that the concentration property may fail to hold there.

We can generate the rows of $\tilde{\mathbf{Z}}$ having a spherically symmetric distribution as $aR\mathbf{U}$, where the direction $\mathbf{U}$ is uniform on the unit sphere in $\mathbb{R}^{\tilde{p}}$, where R is independent of $\mathbf{U}$ and is supported on $[0, \infty)$, and a is chosen so that each component has unit variance. However, it is important to remember (cf. Hall *et al.* (2005), for instance) that in high dimensions, the distribution of the radial component aR tends to be quite highly concentrated – see Figure 1(c). Even when R has a distribution traditionally thought of as light-tailed in a univariate context, such as a Weibull distribution with shape parameter 2, the distribution of aR is much more dispersed (Figure 1(c)), and Figure 1(d) shows that in such a context the corresponding matrix $\tilde{\mathbf{Z}}$ can become badly ill-conditioned as n becomes large.

†*Address for correspondence*: Richard Samworth, Statistical Laboratory, Centre for Mathematical Sciences, Wilberforce Road, Cambridge, UK. CB3 0WB.
E-mail: r.j.samworth@statslab.cam.ac.uk.

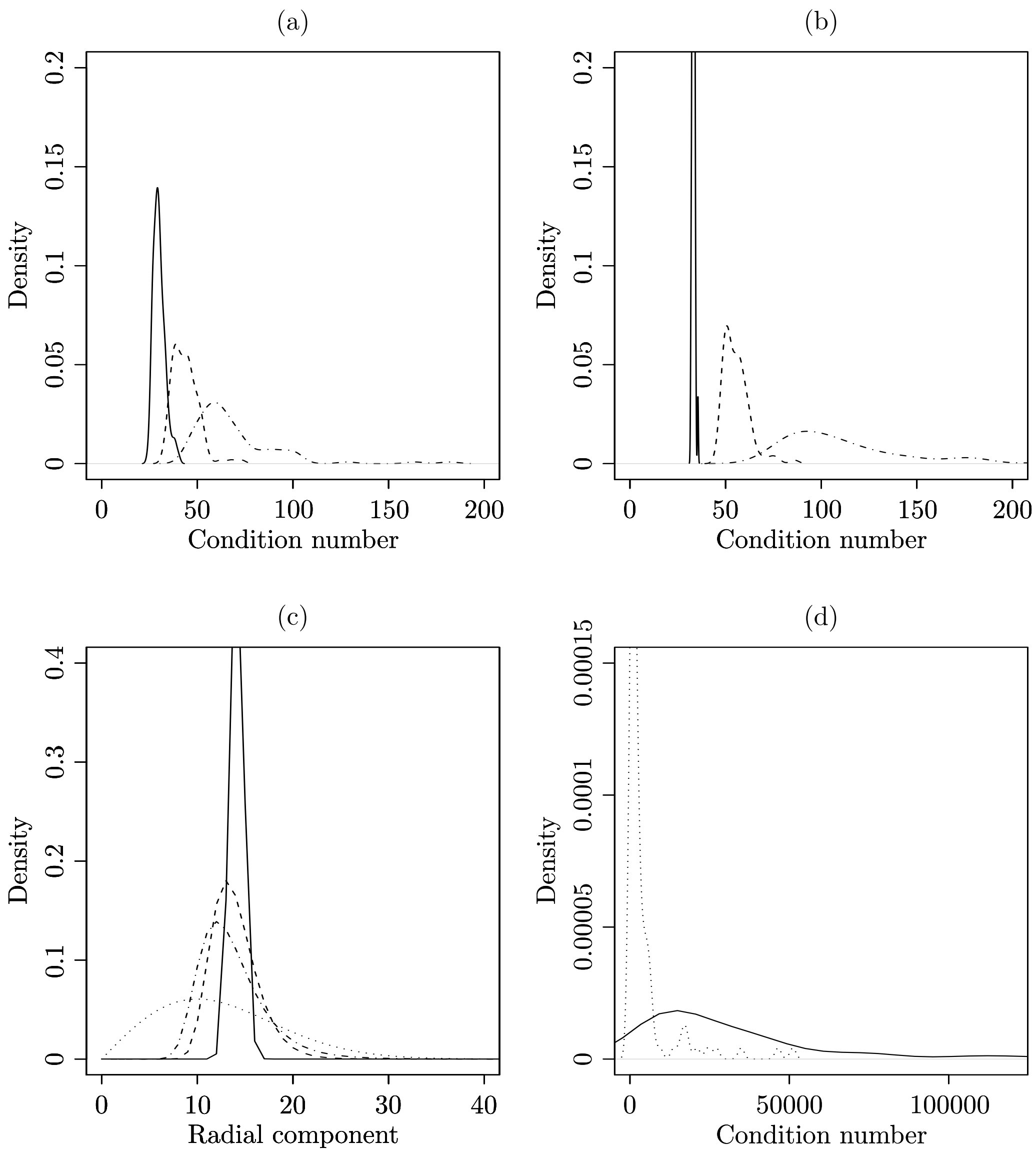

Fig. 1. Panels (a) and (b) give estimated densities of the condition number of $\tilde{p}^{-1}\tilde{\mathbf{Z}}\tilde{\mathbf{Z}}^T$, based on 100 simulations, when $\mathbf{Y}$ is Gaussian (solid) and multivariate-t with $\nu = 20$ (dashed) and $\nu = 10$ (dot-dashed) degrees of freedom; in (a) $n = 100$, $\tilde{p} = 200$, in (b) $n = 1000$, $\tilde{p} = 2000$. Panel (c) gives the corresponding densities of the radial components of these spherically symmetric densities when $\tilde{p} = 200$, as well as the density of a scaled Weibull distribution with shape parameter 2 (dotted). Panel (d) shows the estimated condition number density of $\tilde{p}^{-1}\tilde{\mathbf{Z}}\tilde{\mathbf{Z}}^T$ in the Weibull case, with $n = 100$, $\tilde{p} = 200$ (dotted) and $n = 1000$, $\tilde{p} = 2000$ (solid).

Overall then, we conclude that while condition (16) appears reasonable for the authors' purposes, it remains of interest to describe the theoretical properties of independence screening when the rows of $\mathbf{Z}$ have heavier tails and condition (16) may fail.

References

Hall, P., Marron, J. S. and Neeman, A. (2005) Geometric representation of high dimension, low sample size data. *J. R. Statist. Soc. Ser. B.*, **67**, 427–444.