1 Supplementary Figures

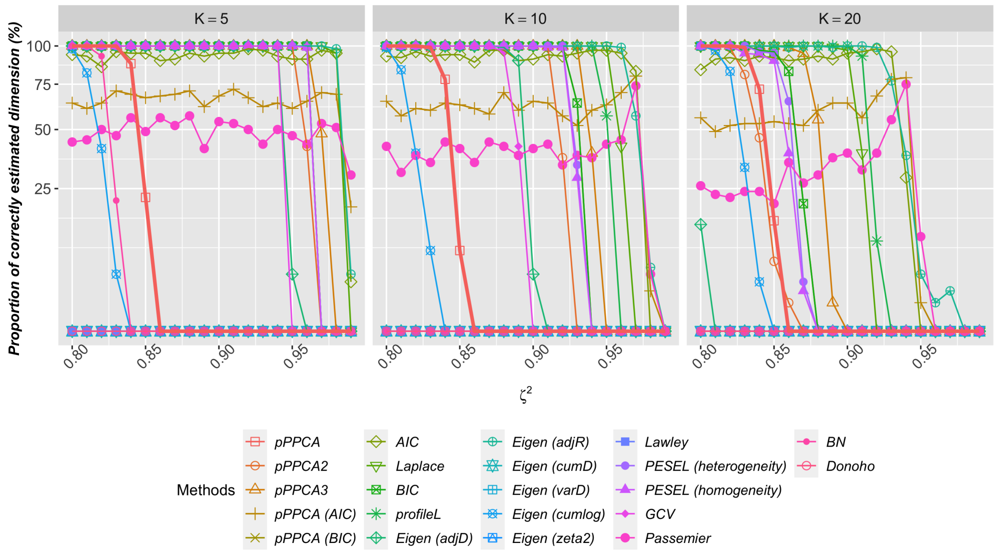

Figure 1: Proportion of correctly estimated k over 100 replicates as a function of ζ^2 assuming the first k^* squared singular values are equal. The colored lines indicate the proportion of correctly estimated values for all methods.

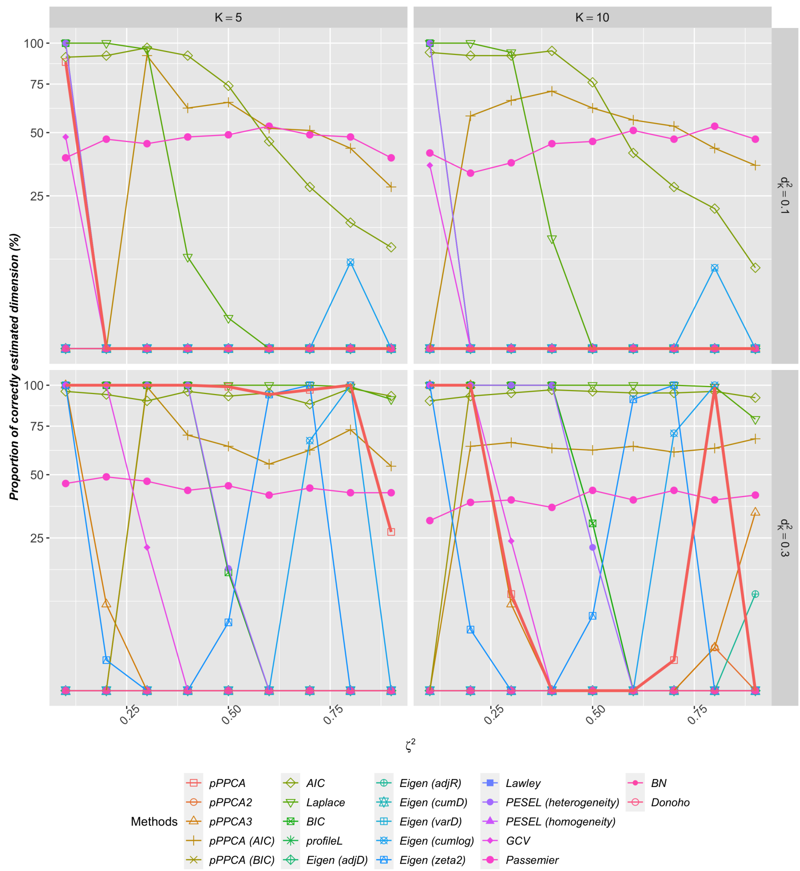

Figure 2: Proportion of correctly estimated k over 100 replicates as a function of ζ^2 assuming a linear decay in the first k^* squared singular values.

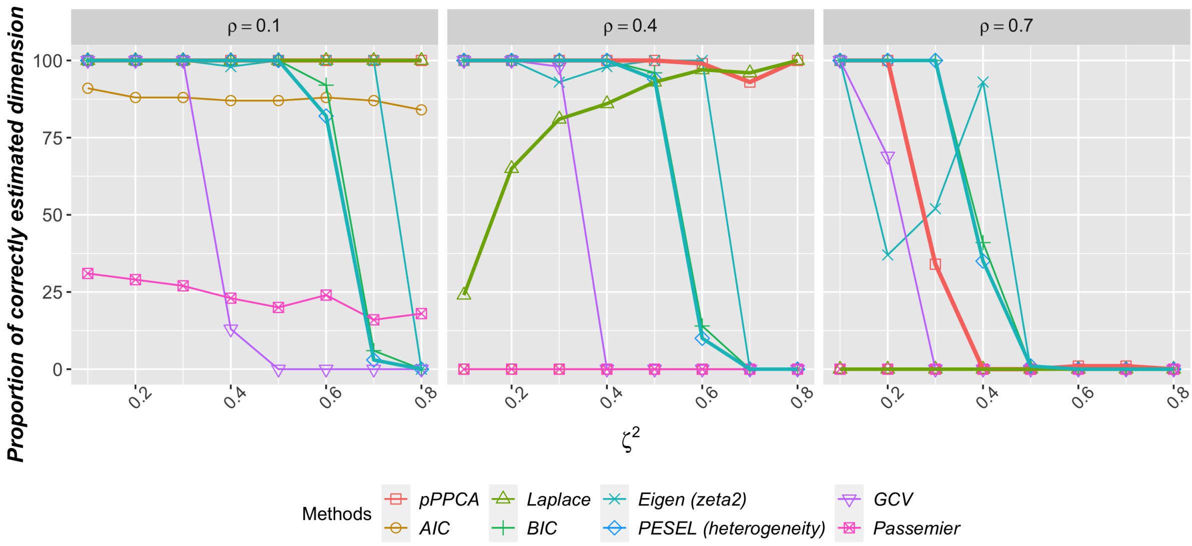

Figure 3: Proportion of correctly estimated dimension over 100 replicates as a function of ζ^2 assuming an exponential decay in the first k^* squared singular values when error terms are correlated.

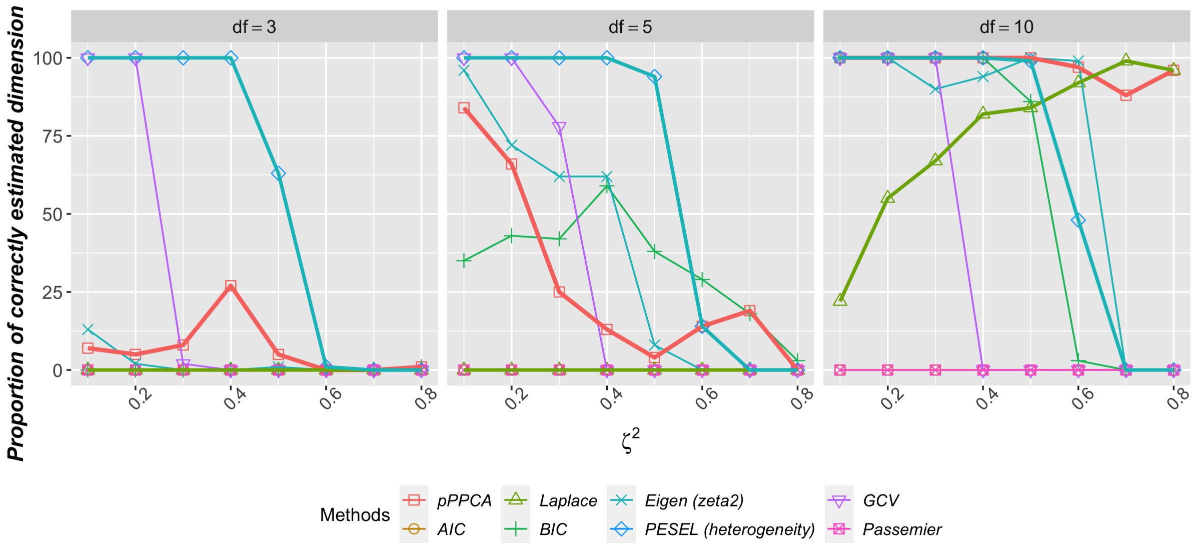

Figure 4: Proportion of correctly estimated dimension over 100 replicates as a function of ζ^2 assuming an exponential decay in the first k^* squared singular values when error follows a student's t distribution.

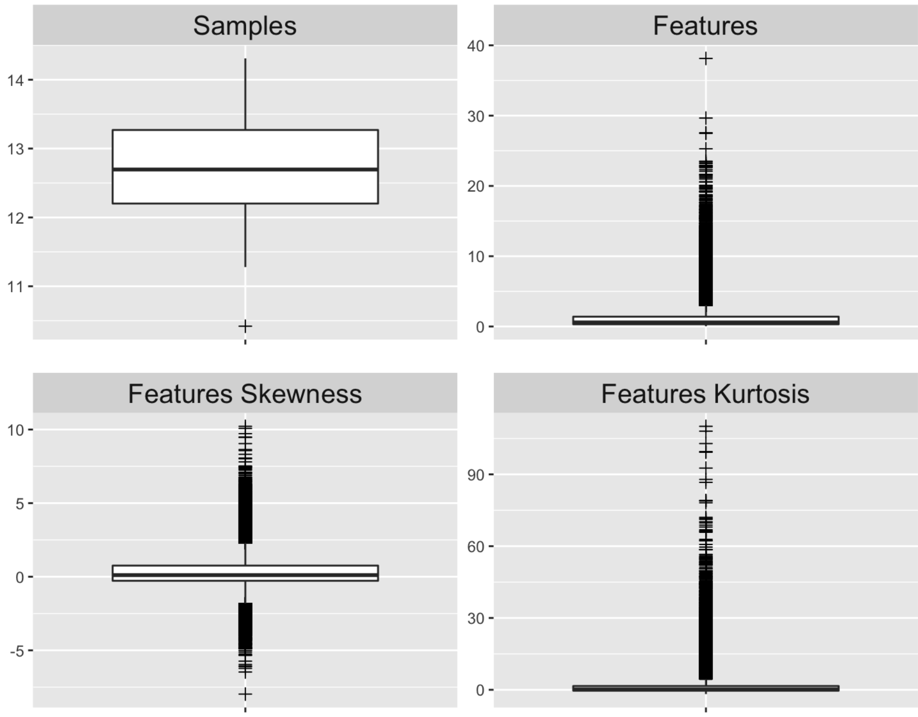

Figure 5: A summary of the sample variance and gene feature variance, skewness and kurtosis for the NCI60 gene expression data.

2 Supplementary Materials

2.1 An illustration of the voting strategy

I demonstrate the utility of theoretical results on two simulated scenarios, assuming $k^* = 10$ and fixing the number of observations at $m = 10,000$ and number of samples $n = 100$. The covariance structure was generated according to a singular value decomposition using random orthogonal matrices and specified squared singular values d^2 equal to $(15, 12, 11.5, 8, 7, 6, 5, 3, 2, 0.5)$ (Figure 6) and $(6, 5, 4, 3.5, 3.5, 3.25, 2, 1.5, 0.75, 0.5)$ (Figure 7), corresponding to $\zeta_{k^*}^2 = 0.3$ and $\zeta_{k^*}^2 = 0.7$, respectively. The signal-to-noise ratio (SNR) is defined as the ratio of the last non-zero squared singular value and the noise variance, evaluated at $0.5/0.3 = 1.6$ and $0.5/0.7 = 0.7$ for these two scenarios, respectively.

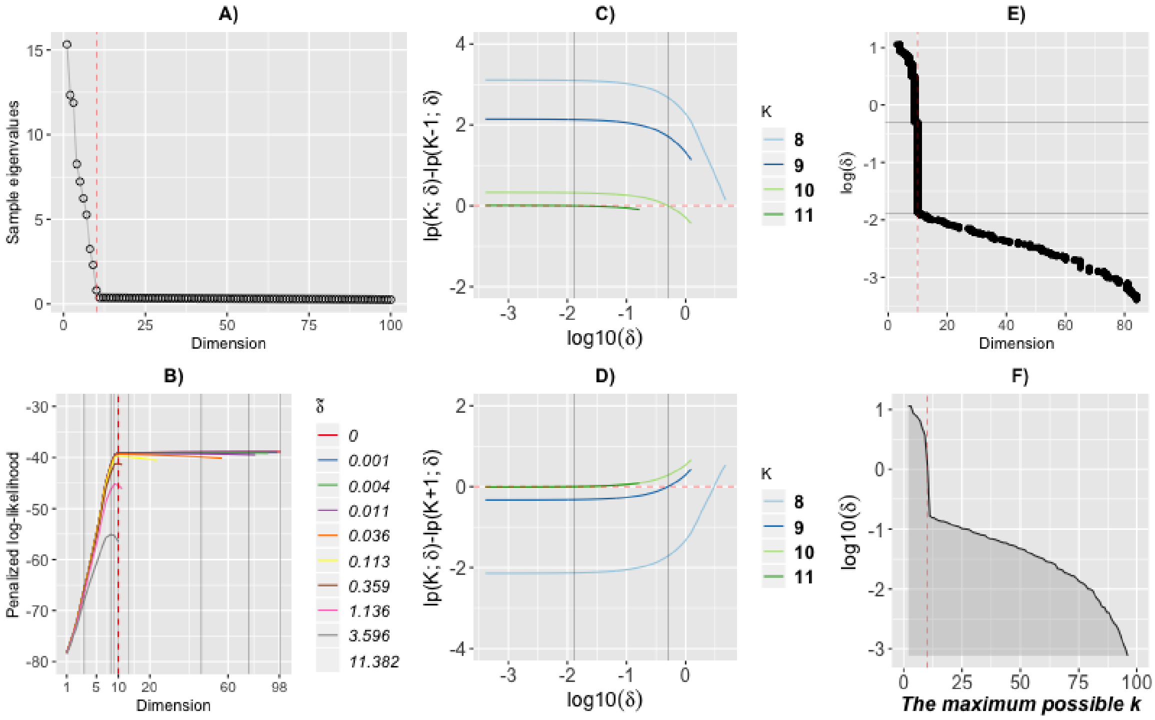

Figure 6: **An illustration of the voting procedure under scenario 1**.
The true dimension $k^* = 10$ is marked by the vertical dashed line in red and the maximum of penalized log-likelihood at each penalty tuning parameter is identified by the black vertical line. In C) and D), the black vertical line on the left corresponds to the $\log(\delta)$ value such that $l_p(k, \tilde{\delta}) - l_p(k+1, \tilde{\delta})$ becomes positive for each k, while the vertical line on the right corresponds to the value such that $l_p(k, \tilde{\delta}) - l_p(k-1, \tilde{\delta})$ becomes negative. In E), the horizontal lines mark the values of $\log(\tilde{\delta})$ for which the correct value ($k^* = 10$) maximizes the penalized log-likelihood.

Figures 6A and 7A show the sample eigenvalue with respect to each of the principal dimension with $k^* = 10$ marked by the vertical dashed line in red. Figures 6B and 7B on bottom left show $l_p(k, \tilde{\delta})$ at various $\tilde{\delta}$-values as a function of k, and clearly as the penalty tuning parameter value increases, a maximum emerges (identified by the black vertical lines) and some choices of $\tilde{\delta}$ identifies the correct dimension $k^* = 10$ as marked by the vertical dashed line in red.

Since $k = \tilde{k}$ maximizes the penalized profile log-likelihood for a particular $\tilde{\delta}$ value only when

- $l_p(k, \tilde{\delta}) - l_p(k+1, \tilde{\delta}) > 0$ and
- $l_p(k, \tilde{\delta}) - l_p(k-1, \tilde{\delta}) > 0$,

simultaneously, the range of $\tilde{\delta}$ where these hold can be visualized for a few possible k's including k^* (Figures 6-C,D and 7-C,D). Visibly, $l_p(k, \delta) - l_p(k-1, \delta)$ is a smooth function of δ monotonic as proven in Lemma 2. Similar observations can be made for $l_p(k, \delta) - l_p(k+1, \delta)$. With little or no penalty, $l_p(k, \tilde{\delta}) - l_p(k+1, \tilde{\delta})$ is negative and $l_p(k, \tilde{\delta}) - l_p(k-1, \tilde{\delta})$ positive for all k's. As $\tilde{\delta}$ increases, the difference in penalized log-likelihood between k and $k+1$ increases and eventually becomes positive while the difference between k and $k-1$ decreases as $(\tilde{\delta})$ reaches $(1/k - 1/n)$.

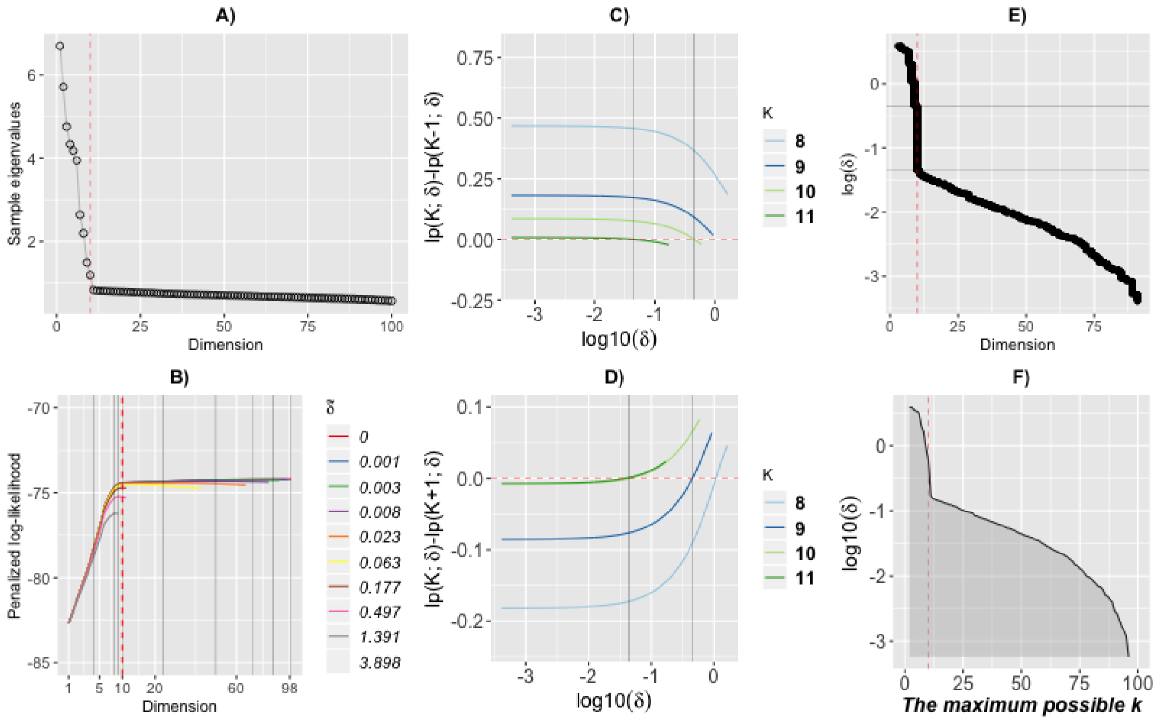

Figure 7: **An illustration of the voting procedure under scenario 2**. The true dimension $k^* = 10$ is marked by the vertical dashed line in red and the maximum of penalized log-likelihood at each penalty tuning parameter is identified by the black vertical line. In C) and D), the black vertical line on the left corresponds to the $\log(\delta)$ value such that $l_p(k, \tilde{\delta}) - l_p(k+1, \tilde{\delta})$ becomes positive for each k, while the vertical line on the right corresponds to the value such that $l_p(k, \tilde{\delta}) - l_p(k-1, \tilde{\delta})$ becomes negative. In E), the horizontal lines mark the values of $\log(\tilde{\delta})$ for which the correct value $(k^* = 10)$ maximizes the penalized log-likelihood.

The differences $l_p(k, \tilde{\delta}) - l_p(k-1, \tilde{\delta})$ and $l_p(k, \tilde{\delta}) - l_p(k+1, \tilde{\delta})$ for $k = 9, 10, 11$ as functions of $\log(\delta)$ are illustrated in Figures 6-C,D and Figures 7-C,D, where the black vertical line on the left corresponds to the $\log(\delta)$ value such that $l_p(k, \tilde{\delta}) - l_p(k+1, \tilde{\delta})$ becomes positive for $k = 10$, while the vertical line on the right corresponds to the value such that $l_p(k, \tilde{\delta}) - l_p(k-1, \tilde{\delta})$ becomes negative for $k = 10$. When both differences are positive, the numerically approximated $(\tilde{a}_k, \tilde{b}_k)$ are formed and the vertical distance represents the amount of evidence for each possible k (Figure 6-E and Figure 7-E). The horizontal lines mark the values of $\log(\tilde{\delta})$ for which the correct value $(k^* = 10)$ maximizes the penalized log-likelihood. The shorter distance between the two lines suggest recovery of the correct k in scenario 2 is more difficult than scenario 1.

It is easy to confirm visually that the numerically evaluated sets $(\tilde{a}_k, \tilde{b}_k)$ are non-overlapping and k is indeed monotonically decreasing in $\tilde{\delta}$. Notice that the log difference in $\tilde{\delta}$ is the largest for $k = k^* = 10$ as compared to $k = 9$ or $k = 11$, suggesting that for a grid-set of equidistant values constructed on *log* scale, the majority would support k^*. Consequently, with the search grid constructed as proposed, the procedure would estimate the dimension by a majority vote, which is expected given the relationship between the log distance $\log(\tilde{b}_k) - \log(\tilde{a}_k)$ and the number of votes for a particular k (Figures 6 and 7).

$$\tilde{\zeta}_k^2 = \frac{\sum_{i=k+1}^{n} \hat{\lambda}_i}{n - k - \delta k} = \frac{n-k}{n-k-\delta k} \hat{\zeta}_k^2. \tag{1}$$

$$\begin{cases} l_p(k; \tilde{\delta}_o) - l_p(k-1; \tilde{\delta}_o) > 0 \\ l_p(k; \tilde{\delta}_o) - l_p(k+1; \tilde{\delta}_o) > 0. \end{cases} \tag{2}$$

Finally, the restriction embedded in (1) and (2) imposes a relationship whereby the maximum possible k is non-increasing in Δ_k, shown by the area in gray (Figures 6-F and 7-F).

2.2 Proofs of Lemmas

Lemma 1. *Consider a sample $X \in \mathbb{R}^{n\times m}$ with each column following a multivariate Gaussian distribution $\mathcal{N}(0, WW^T + \zeta^2 I)$. Suppose W has rank k^* and further, the sample covariance matrix of X^T is positive semi-definite. Then, the penalized maximum log-likelihood at each fixed $k \in \{1, \ldots, n-1\}$ is a smooth function of $\tilde{\delta}$ on the interval $(0, 1/k - 1/n)$ and is monotonically decreasing on*

$$(0, (1/k - 1/n)[1 - \hat{\zeta}_k^2]),$$

where $\hat{\zeta}_k^2 = (\sum_{i=k+1}^{n} \hat{\lambda}_i)/(n-k)$.

Proof (Lemma 1)**.** *The penalized profile log-likelihood is a smooth function of the scaled tuning parameter $\tilde{\delta}$ for each k, provided that it is differentiable with respect to $\tilde{\delta}$ and all derivatives of $l_p(k, \tilde{\delta})$ exist on $\tilde{\delta} \in (0, 1/k - 1/n)$.*

The penalized profile log-likelihood at $k = k'$ is

$$l_p(k'; \tilde{\delta}) = l_p(k') - \frac{m}{2}\Big\{(n-k')\log\Big(1 - \frac{k'}{n}\Big) - (n-k')\log\Big(1 - \frac{k'}{n} - k'\tilde{\delta}\Big) - nk'\tilde{\delta}\Big[1 + \log\hat{\zeta}^2(k') + \log\Big(1 - \frac{k'}{n}\Big) - \log\Big(1 - \frac{k'}{n} - k\tilde{\delta}\Big)\Big]\Big\}$$

and the first order derivative with respect to $\tilde{\delta}$ is given by:

$$\frac{\partial l}{\partial \tilde{\delta}} = -\frac{mnk'}{2}\log\Big[\frac{n - k' - nk'\tilde{\delta}}{(n-k')\hat{\zeta}^2(k')}\Big], \tag{3}$$

the second order derivative with respect to $\tilde{\delta}$ is:

$$\frac{\partial^2 l}{\partial \tilde{\delta}^2} = \frac{mnk'}{2}\frac{nk'}{n - k' - nk'\tilde{\delta}}.$$

Consequently, the tth order derivative is

$$\frac{\partial^t l}{\partial \tilde{\delta}^t} = \frac{mnk'}{2}(-1)^t\frac{(t-2)!(nk')^{t-1}}{(n - k' - nk'\tilde{\delta})^{t-1}}. \tag{4}$$

Since (4) is a rational function of $\tilde{\delta}$ on $(0, 1/k' - 1/n)$ and all derivatives of $l_p(k', \tilde{\delta})$ exist, $l_p(k, \tilde{\delta})$ is a smooth function of $\tilde{\delta}$ on the same range.

Observe the first order derivative (3) is negative for any $k \in \{1, 2, \ldots, n-1\}$ whenever:

$$\log\Big(1 - \frac{k}{n} - k\tilde{\delta}\Big) - \log\Big(1 - \frac{k}{n}\Big) - \log\hat{\zeta}_k^2 > 0,$$
$$\tilde{\delta} < \Big(\frac{1}{k} - \frac{1}{n}\Big)[1 - \hat{\zeta}_k^2], \tag{5}$$

and equals to zero when $\tilde{\delta} = (1/k - 1/n)[1 - \hat{\zeta}_k^2]$. Thus, $l_p(k; \tilde{\delta})$ is a monotonically decreasing function of $\tilde{\delta}$ for any $k \in \{1, 2, \ldots, n-1\}$ on $(0, (1/k - 1/n)[1 - \hat{\zeta}_k^2])$, and notice that on this range we also have $\tilde{\zeta}^2(k) < 1$.

Lemma 2. *Consider $\tilde{\delta} \in G(k+1)$, where*

$$G(k+1) = \Big(0, \frac{1}{n}\frac{(n-k-1)(\hat{\lambda}_{k+1} - \hat{\zeta}_{k+1}^2)}{(k+1)\hat{\lambda}_{k+1} + (n-k-1)\hat{\zeta}_{k+1}^2}\Big).$$

Then, $G(k+1) \subset G(k)$ and for any fixed $k \in \{2, 3, \ldots, n-2\}$, $l_p(k; \tilde{\delta}) - l_p(k+1; \tilde{\delta})$ is a monotonically increasing and concave function of $\tilde{\delta} \in G(k+1)$ and $l_p(k; \tilde{\delta}) - l_p(k-1; \tilde{\delta})$ is a monotonically decreasing and convex function of $\tilde{\delta} \in G(k+1)$.

Proof (Lemma 2). *Lemma 1 proves that $l_p(k;\tilde{\delta})$, $l_p(k-1;\tilde{\delta})$, and $l_p(k+1;\tilde{\delta})$ are smooth and monotonically decreasing on $\tilde{\delta} \in (0, (1/(k+1)-1/n)[1-\hat{\zeta}^2_{k+1}])$ for any fixed $k \in \{2,\ldots,n-1\}$. Since*

$$\Big(\frac{1}{k+1}-\frac{1}{n}\Big)[1-\hat{\zeta}^2_{k+1}] > \frac{1}{n}\frac{(n-k-1)[\hat{\lambda}_{k+1}-\hat{\zeta}^2_{k+1}]}{(k+1)\hat{\lambda}_{k+1}+(n-k-1)\hat{\zeta}^2_{k+1}},$$

and

$$\frac{1}{n}\frac{(n-k)[\hat{\lambda}_k-\hat{\zeta}^2_k]}{k\hat{\lambda}_k+(n-k)\hat{\zeta}^2_k} > \frac{1}{n}\frac{(n-k-1)[\hat{\lambda}_{k+1}-\hat{\zeta}^2_{k+1}]}{(k+1)\hat{\lambda}_{k+1}+(n-k-1)\hat{\zeta}^2_{k+1}},$$

$G(k+1) \subset G(k)$ holds.

The difference between the penalized profile log-likelihood at k and $k-1$ or $k+1$ is a simple function. Thus $l_p(k;\tilde{\delta})-l_p(k-1;\tilde{\delta})$ is smooth on $G(k+1) \subset G(k)$. Similarly, $l_p(k;\tilde{\delta})-l_p(k+1;\tilde{\delta})$ is also smooth on $G(k+1)$.

Next, I investigate the behaviours of these two functions separately:

A: Show that $l_p(k;\tilde{\delta})-l_p(k-1;\tilde{\delta})$ is convex on $G(k)$

The first order derivative with respect to $\tilde{\delta}$, which is the difference between the first order derivative of $l_p(k;\tilde{\delta})$ and $l_p(k-1;\tilde{\delta})$, is

$$\begin{aligned}\frac{\partial l_p(k;\tilde{\delta})-l_p(k-1;\tilde{\delta})}{\partial\tilde{\delta}} &= -\frac{mnk}{2}\log\frac{1-\frac{k}{n}-k\tilde{\delta}}{(1-\frac{k}{n})\hat{\zeta}^2(k)} \\ &+ \frac{mn(k-1)}{2}\log\frac{1-\frac{k-1}{n}-(k-1)\tilde{\delta}}{(1-\frac{k-1}{n})\hat{\zeta}^2(k-1)} \\ &= \frac{mnk}{2}\log\frac{\tilde{\zeta}^2(k)}{\tilde{\zeta}^2(k-1)} + \frac{mn}{2}\log\tilde{\zeta}^2(k-1). \qquad (6)\end{aligned}$$

Notice that $\frac{\tilde{\zeta}^2(k)}{\tilde{\zeta}^2(k-1)}$ can be greater or less than 1 depending on $\tilde{\delta}$:

$$\frac{\tilde{\zeta}^2(k)}{\tilde{\zeta}^2(k-1)} = \frac{n-k+1-n(k-1)\tilde{\delta}}{n-k-nk\tilde{\delta}}\frac{n-k}{n-k+1}\frac{\hat{\zeta}^2_k}{\hat{\zeta}^2_{k-1}}. \qquad (7)$$

However, its first order derivative

$$\frac{n^2}{(n-k-nk\tilde{\delta})^2}\frac{(n-k)}{(n-k+1)}\frac{\hat{\zeta}^2_k}{\hat{\zeta}^2_{k-1}} > 0,$$

which indicates that (7) is increasing in $\tilde{\delta}$. In order for (7) to be less than 1, solve for the maximum $\tilde{\delta}$ that satisfies $\frac{\tilde{\zeta}^2(k)}{\tilde{\zeta}^2(k-1)} < 1$ gives:

$$\tilde{\delta} < \frac{n-k}{n}\frac{\hat{\lambda}_k-\hat{\zeta}^2_k}{k\hat{\lambda}_k+(n-k)\hat{\zeta}^2_k}$$

Thus, $\tilde{\delta} \in G(k)$ such that (7) is less than 1, while $\tilde{\zeta}^2(k-1) < 1$ due to $max_{\tilde{\delta}}\{G(k)\} < (1/k-1/n)[1-\hat{\zeta}^2_k]$. Thus, I conclude the first order derivative is negative on $\tilde{\delta} \in G(k)$. The second order derivative can be shown to be positive,

$$\begin{aligned}\frac{\partial^2 l_p(k;\tilde{\delta})-l_p(k-1;\tilde{\delta})}{\partial\tilde{\delta}^2} &= \frac{mnk}{2}\frac{k}{1-\frac{k}{n}-k\tilde{\delta}} - \frac{mn(k-1)}{2}\frac{k-1}{1-\frac{k-1}{n}-(k-1)\tilde{\delta}} \\ &= \frac{mn}{2}\Big[\frac{nk^2}{n-k-nk\tilde{\delta}} - \frac{n(k-1)^2}{n-(k-1)-n(k-1)\tilde{\delta}}\Big] > 0, \qquad (8)\end{aligned}$$

and I conclude that $l_p(k;\tilde{\delta})-l_p(k-1;\tilde{\delta})$ is a convex function of $\tilde{\delta} \in G(k)$ for any fixed k.

B: Show that $l_p(k;\tilde\delta)-l_p(k+1;\tilde\delta)$ is concave on $G(k+1)$

The first order derivative with respect to $\tilde\delta$:

$$\begin{aligned}\frac{\partial l_p(k;\tilde\delta)-l_p(k+1;\tilde\delta)}{\partial\tilde\delta} &= -\frac{mnk}{2}\log\frac{n-k-nk\tilde\delta}{(n-k)\hat\zeta^2(k)} \\ &+\frac{mn(k+1)}{2}\Big[-\log\frac{n-(k+1)}{n-(k+1)-n(k+1)\tilde\delta}-\log\hat\zeta^2_{k+1}\Big] \\ &= \frac{mn(k+1)}{2}\log\frac{\tilde\zeta^2(k)}{\tilde\zeta^2(k+1)}-\frac{mn}{2}\log\tilde\zeta^2(k). \end{aligned} \tag{9}$$

It can be shown similarly to the first case by taking $k'=k+1$ and taking the opposite sign, we can conclude $\frac{\partial l_p(k;\tilde\delta)-l_p(k+1;\tilde\delta)}{\partial\tilde\delta}>0$ on $\tilde\delta\in G(k+1)$. The second order derivative can be shown to be negative,

$$\begin{aligned}\frac{\partial^2 l_p(k;\tilde\delta)-l_p(k+1;\tilde\delta)}{\partial\tilde\delta^2} &= \frac{mnk}{2}\frac{k}{1-\frac{k}{n}-k\tilde\delta}-\frac{mn(k+1)}{2}\frac{k+1}{1-\frac{k+1}{n}-(k+1)\tilde\delta} \\ &= \frac{mn}{2}\Big[\frac{nk^2}{n-k-nk\tilde\delta}-\frac{n(k+1)^2}{n-k-1-n(k+1)\tilde\delta}\Big]<0, \end{aligned} \tag{10}$$

and that $l_p(k;\tilde\delta)-l_p(k+1;\tilde\delta)$ is a concave function of $\tilde\delta\in G(k+1)$ for any fixed k.

Finally, I conclude that for any fixed $k\in\{2,3,\ldots,n-2\}$, $l_p(k;\tilde\delta)-l_p(k-1;\tilde\delta)$ is a monotonically decreasing and convex function of $\tilde\delta\in G(k+1)\subset G(k)$, taking positive value when $\tilde\delta=0$; $l_p(k;\tilde\delta)-l_p(k+1;\tilde\delta)$ is a monotonically increasing and concave function of $\tilde\delta\in G(k+1)$, taking negative value when $\tilde\delta=0$.

Lemma 3. *Assume the same notation from Lemma 2. For some $k\in\{2,\ldots,n-2\}$, there exists $\tilde\delta_o\in\cup_k G(k+1)$ such that $k=\operatorname{argmax}_{k'} l_p(k';\tilde\delta_o)$ if and only if*

$$\begin{cases} l_p(k;\tilde\delta_o)-l_p(k-1;\tilde\delta_o)>0 \\ l_p(k;\tilde\delta_o)-l_p(k+1;\tilde\delta_o)>0. \end{cases} \tag{11}$$

Proof (Lemma 3)**.** *For a given $\tilde\delta_o$ value, there is a restricted subset of possible $k\in\{1,2,3,\ldots,k(\tilde\delta_o)\}$. Following Lemma 2, $l_p(k;\tilde\delta)>l_p(k-1;\tilde\delta)$ and $l_p(k;\tilde\delta)-l(k+1;\tilde\delta)$ are smooth and monotone functions of $\tilde\delta_o\in\cup_k G(k+1)$.*

For some $k\in\{2\le k\le n-2\}$, define

$$\Delta_k=(a_k,b_k)\subset\cup_k G(k+1)$$

where

$$a_k=min\Big\{\tilde\delta\in\cup_k G(k+1);\ l(k;\tilde\delta)-l_p(k+1;\tilde\delta)>0\Big\}$$

and

$$b_k=max\Big\{\tilde\delta\in\cup_k G(k+1);\ l(k;\tilde\delta)-l_p(k-1;\tilde\delta)>0\Big\}.$$

The existence of b_k is clear as $l_p(k;\tilde\delta)-l_p(k-1;\tilde\delta)$ is a smooth and decreasing function on $G(k+1)$ and $\{l_p(k;\tilde\delta)-l_p(k-1;\tilde\delta)\}_{\tilde\delta=0}>0$. Thus, b_k is bounded between $\tilde\delta=0$ and the solution to $l_p(k;\tilde\delta)-l_p(k-1;\tilde\delta)=0$.

The existence of $a_k\in G(k+1)$ is unclear depending on whether $\hat\lambda_{k+1}<1$ as shown below:

$$\begin{aligned} l_p(k;\tilde\delta)-l_p(k+1;\tilde\delta) &> \{l_p(k;\tilde\delta)-l_p(k+1;\tilde\delta)\}_{\tilde\delta=maxG(k+1)} \\ &= -\frac{m}{2}\Big\{\log\frac{\tilde\zeta^2(k+1)}{\hat\lambda_{k+1}}+\tilde\delta n[\log\tilde\zeta^2(k+1)+1]\Big\} \\ &> -\frac{m}{2}\Big\{\frac{n-k-1}{n}[\hat\zeta^2_{k+1}-\hat\lambda_{k+1}](1/\hat\lambda_{k+1}-1)\Big\} \end{aligned} \tag{12}$$

Consider $k \in \{i : \hat{\lambda}_{i+1} < 1\}$ and suppose for any $\tilde{\delta}_o \in \Delta_{k'} = (a_{k'}, b_{k'}) \subset G(k')$, $l_p(k;\tilde{\delta})$ takes its maximum at $k = k'$. Then clearly $l_p(k';\tilde{\delta}) > l_p(k'-1;\tilde{\delta})$ and $l_p(k';\tilde{\delta}) > l(k'+1;\tilde{\delta})$. In other words, if $\hat{\lambda}_{k+1} < 1$, then $l_p(k;\tilde{\delta}) - l_p(k+1;\tilde{\delta}) > 0$ on $G(k+1)$, otherwise it is undetermined.

Thus, it might not always be that $a_k < b_k \in G(k+1)$ for k such that $\hat{\lambda}_{k+1} > 1$. But for k such that $\hat{\lambda}_{k+1} < 1$, according to this definition, $a_{k-1} = b_k$ whenever a_{k-1} exists since $G(k+1) \in G(k)$ and combined with the fact that the first order derivative of $l_p(k;\tilde{\delta}) - l(k+1;\tilde{\delta})$ is smaller than that of $-\{l_p(k;\tilde{\delta}) - l(k-1;\tilde{\delta})\}$, implying that $a_k < a_{k-1} = b_k$. Thus, the interval $\Delta_k = (a_k, b_k) \subset G(k+1)$ is not empty and gives the range of $\tilde{\delta}$-values in $G(k+1)$ such that conditions (2) hold.

For k's such that $\hat{\lambda}_{k+1} > 1$, they could still be the maximizer of the penalized profile log-likelihood for some $\tilde{\delta}$ such that $l_p(k;\tilde{\delta}) - l_p(k+1;\tilde{\delta}) > 0$.

On the other hand, if conditions (2) hold for some $\tilde{\delta}_o \in \Delta_{k'} = (a_{k'}, b_{k'})$, we need to show $l_p(k';\tilde{\delta}) - l_p(k'';\tilde{\delta}) > 0$ for any $k'' < k'$ by showing the following sum is greater than zero:

$$\begin{aligned} l_p(k';\tilde{\delta}) - l_p(k'';\tilde{\delta}) &= [l_p(k';\tilde{\delta}) - l_p(k'-1;\tilde{\delta})] + [l_p(k'-1;\tilde{\delta}) - l_p(k'-2;\tilde{\delta})] \\ &+ \cdots + [l_p(k''+1;\tilde{\delta}) - l_p(k'';\tilde{\delta})] > 0, \end{aligned} \tag{13}$$

Since $l_p(k';\tilde{\delta}) - l_p(k'-1;\tilde{\delta})$ is positive on (a'_k, b'_k), we only need to show $l_p(k'-1;\tilde{\delta}) - l_p(k'-2;\tilde{\delta})$ and all other telescoping terms are positive on (a'_k, b'_k). This is straightforward to show according to the definitions of $(a_{k'-1}, b_{k'-1})$ and $a_{k'-1} = b_{k'}$. This implies for each $k'' < k'$, $l_p(k'';\tilde{\delta}) - l_p(k''-1;\tilde{\delta})$ is positive on $(0, b_{k''}) \subset (0, b_{k'})$.

Similarly, $l_p(k';\tilde{\delta}_o) - l_p(k'';\tilde{\delta}_o) > 0$ is equivalent to the telescoping sum being greater than zero:

$$\begin{aligned} l_p(k';\tilde{\delta}_o) - l_p(k'';\tilde{\delta}_o) &= [l_p(k';\tilde{\delta}_o) - l_p(k'+1;\tilde{\delta}_o)] + [l_p(k'+1;\tilde{\delta}_o) - l_p(k'+2;\tilde{\delta}_o)] \\ &+ \cdots + [l_p(k''-1;\tilde{\delta}_o) - l_p(k'';\tilde{\delta}_o)] > 0, \end{aligned} \tag{14}$$

where we must have $l_p(k';\tilde{\delta}_o) - l_p(k'+1;\tilde{\delta}_o) > 0$ on $\Delta_k = (a'_k, b'_k)$.

It is sufficient to show $l_p(k'+1;\tilde{\delta}) - l_p(k'+2;\tilde{\delta})$ and all other telescoping terms are positive on (a'_k, b'_k). According to the definitions of $(a_{k'+1}, b_{k'+1})$ and the result that $a_{k'} = b_{k'+1}$. This implies for each $k'' > k'$, $l_p(k'';\tilde{\delta}) - l_p(k''+1;\tilde{\delta})$ is positive when $\tilde{\delta} > a_{k''}$, and $(a_{k''}, b_{k''}) \subset (a'_k, b'_k)$.

I conclude the proof with a comment on the choice of $k(\tilde{\delta}_o)$. Observe that $n - k - nk\tilde{\delta} > 0$ poses a restriction on both k and $\tilde{\delta}$, for any given $\tilde{\delta}_o$, the maximum k searchable is $k_{max'} = (\tilde{\delta}_o + n^{-1})^{-1}$. However, for this $k = k_{max'}$, the penalized profile log-likelihoods $l_p(k_{max'};\tilde{\delta})$ and $l_p(k_{max'}+1;\tilde{\delta})$ are smooth, but not monotonically decreasing on $(0, \tilde{\delta}_o)$. However, these choices of k are ruled out by condition (2).

Lemma 4. *Consider*

$$\Delta_k = \left\{\tilde{\delta} \in G(k+1);\ \textit{conditions (2) are satisfied}\right\} = (a_k, b_k),$$

whenever a_k exists. Then Δ_k can be approximated by $(u_a(k), u_b(k)) \subset \Delta_k \subset G(k+1)$, where $u_a(k)$ denote an upper bound for a_k, and $u_b(k)$ a lower bound for b_k, such that $b_k/a_k > u_b(k)/u_a(k)$.

Proof (Lemma 4)**.** *Since a_k and b_k are not analytically tractable, we approximate the differences in penalized log-likelihoods using Taylor series, and then find roots to the Taylor approximations.*
A: $l_p(k;\tilde{\delta}) - l_p(k-1;\tilde{\delta}) > 0$

Expanding the difference in penalized profile log-likelihoods in terms of $\hat{\zeta}_{k-1}^2$, $\hat{\zeta}_k^2$, $\hat{\lambda}_k$ *and* k *:*

$$\begin{aligned}
l_p(k;\tilde{\delta}) - l_p(k-1;\tilde{\delta}) = -\frac{m}{2}\Big\{ & \log\frac{\hat{\lambda}_k}{\hat{\zeta}_{k-1}^2} + (n-k)\log\frac{\hat{\zeta}_k^2}{\hat{\zeta}_{k-1}^2} \\
& + (n-k)\log\Big(1-\frac{k}{n}\Big) - (n-k+1)\log\Big(1-\frac{k-1}{n}\Big) \\
& - n\tilde{\delta}\Big[k\log\frac{\hat{\zeta}_k^2}{\hat{\zeta}_{k-1}^2} + k\log\frac{n-k}{n-(k-1)} + \\
& \log(\hat{\zeta}_{k-1}^2) + \log\Big(1-\frac{k-1}{n}\Big) + 1\Big] \\
& + (n-k+1)\log\Big[1-\frac{k-1}{n} - (k-1)\tilde{\delta}\Big] - (n-k)\log\Big(1-\frac{k}{n} - k\tilde{\delta}\Big) \\
& + nk\tilde{\delta}\log\Big(1-\frac{k}{n}-k\tilde{\delta}\Big) - n(k-1)\tilde{\delta}\log\Big[1-\frac{k-1}{n}-(k-1)\tilde{\delta}\Big]\Big\}
\end{aligned}$$

The following approximations can be obtained for any k *by Taylor expansion at* $\tilde{\delta} = 0$*:*

$$\log(1 - \tfrac{k}{n} - k\tilde{\delta}) = \log(1 - \tfrac{k}{n}) - \tfrac{nk\tilde{\delta}}{n-k} + O(\tilde{\delta}^2), \tag{15a}$$

$$\log[1 - \tfrac{k-1}{n} - (k-1)\tilde{\delta}] = \log(1 - \tfrac{k-1}{n}) - \tfrac{n(k-1)\tilde{\delta}}{n-k+1} + O(\tilde{\delta}^2), \tag{15b}$$

and they respectively converge if $\tilde{\delta} < 1/k - 1/n$ *and* $\tilde{\delta} < 1/(k-1) - 1/n$*. The difference is approximated by*

$$\begin{aligned}
l_p(k;\tilde{\delta}) - l_p(k-1;\tilde{\delta}) &= -\frac{m}{2}\Big[\zeta(\tilde{\delta}) + O(\tilde{\delta}^2)\Big] \\
&= -\frac{m}{2}\Big\{\log\frac{\hat{\lambda}_k}{\hat{\zeta}_{k-1}^2} + (n-k)\log\frac{\hat{\zeta}_k^2}{\hat{\zeta}_{k-1}^2} \\
&\quad - n\tilde{\delta}\Big[k\log\frac{\hat{\zeta}_k^2}{\hat{\zeta}_{k-1}^2} + \log\hat{\zeta}_{k-1}^2\Big] \\
&\quad + \tilde{\delta}^2\Big[\frac{n^2(k-1)^2}{n-k+1} - \frac{n^2k^2}{n-k}\Big] + O(\tilde{\delta}^3)\Big\}
\end{aligned} \tag{16}$$

Now we need to solve the inequality and find the smallest $\tilde{\delta}$ *such that* $\zeta(\tilde{\delta}) \leq 0$*. Clearly,* $\zeta(\tilde{\delta})$ *is a quadratic function of* $\tilde{\delta}$ *for fixed* k *and when* $\tilde{\delta} = 0$, $\zeta(\tilde{\delta}) < 0$*. In quadratic equation representation, we can rewrite* $\zeta(\tilde{\delta}) = c_2\tilde{\delta}^2 + c_1\tilde{\delta} + c_0$*, where*

$$c_2 = \frac{n^2(k-1)^2}{n-k+1} - \frac{n^2k^2}{n-k} = \frac{n(1-2k) - k(1-k)}{(n-k+1)(n-k)} < 0 \tag{17}$$

$$c_1 = -n\Big[k\log\frac{\hat{\zeta}_k^2}{\hat{\zeta}_{k-1}^2} + \log\hat{\zeta}_{k-1}^2\Big] > 0 \tag{18}$$

$$c_0 = \log\frac{\hat{\lambda}_k}{\hat{\zeta}_{k-1}^2} + (n-k)\log\frac{\hat{\zeta}_k^2}{\hat{\zeta}_{k-1}^2} < 0 \tag{19}$$

Clearly, the discriminant ${c_1}^2 - 4c_2c_0 > 0$ *is positive and* $c_2 < 0$*, then there are two positive roots,* $r_1(k)$ *and* $r_2(k)$*, where*

$$\begin{cases} \zeta(\tilde{\delta}) < 0 \text{ for } \tilde{\delta} \in (0, r_1(k)) \cup (r_2(k), max(G(k+1))), \\ \zeta(\tilde{\delta}) > 0 \text{ for } \tilde{\delta} \in (r_1(k), r_2(k)). \end{cases} \tag{20}$$

B: $l_p(k;\tilde{\delta}) - l_p(k+1;\tilde{\delta}) > 0$

Similarly, expanding the difference:

$$l_p(k;\tilde{\delta}) - l_p(k+1;\tilde{\delta}) = -\frac{m}{2}\Bigg\{ -\log\frac{\hat{\lambda}_{k+1}}{\hat{\zeta}_k^2} + (n-k-1)\log\frac{\hat{\zeta}_k^2}{\hat{\zeta}_{k+1}^2} \tag{21}$$

$$+ n\tilde{\delta} + [(n-k)\log\frac{n-k}{n-k-nk\tilde{\delta}} - (n-k-1)\log\frac{n-k-1}{n-k-1-n(k+1)\tilde{\delta}}] \tag{22}$$

$$- n\tilde{\delta}\big[k\log\frac{\hat{\zeta}_k^2}{\hat{\zeta}_{k+1}^2} + k\log n - kn - k - nk\tilde{\delta} \tag{23}$$

$$- \log\hat{\zeta}_{k+1}^2 - (k+1)\log\frac{n-k-1}{n-k-1-n(k+1)\tilde{\delta}}\big]\Bigg\} \tag{24}$$

The following approximation is obtained for any k by Taylor expansion at $\tilde{\delta} = 0$:

$$\log\Big[1 - \frac{k+1}{n} - (k+1)\tilde{\delta}\Big] = \log\Big(1 - \frac{k'+1}{n}\Big) - \frac{n(k+1)\tilde{\delta}}{n-k-1} + O(\tilde{\delta}^2), \tag{25}$$

and it converges if $\tilde{\delta} < 1/(k+1) - 1/n$.

Again, define $\zeta'(\tilde{\delta})$ and the approximated difference is:

$$\begin{aligned} l_p(k;\tilde{\delta}) - l_p(k+1;\tilde{\delta}) &= -\frac{m}{2}\Big[\zeta'(\tilde{\delta}) + O(\tilde{\delta}^2)\Big] \\ &= -\frac{m}{2}\Big\{ -\log\frac{\hat{\lambda}_{k+1}}{\hat{\zeta}^2(k')} + (n-k'-1)\log\frac{\hat{\zeta}^2(k')}{\hat{\zeta}^2(k'+1)} \\ &\quad - n\tilde{\delta}\Big[k\log\frac{\hat{\zeta}_k^2}{\hat{\zeta}_{k+1}^2} - \log\hat{\zeta}_{k+1}^2\Big] \\ &\quad + \tilde{\delta}^2\Big[\frac{n^2(k+1)^2}{n-k-1} - \frac{n^2k^2}{n-k}\Big] + O(\tilde{\delta}^3)\Big\} \end{aligned} \tag{26}$$

Again, rewrite $\zeta'(\tilde{\delta}) = c_2'\tilde{\delta}^2 + c_1'\tilde{\delta} + c_0'$, where

$$\begin{aligned} c_2' &= \frac{n^2(k+1)^2}{n-k-1} - \frac{n^2k^2}{n-k} = \frac{n-k+(2n-k)k}{(n-k-1)(n-k)} > 0 \\ c_1' &= n\Big[(k+1)\log\frac{\hat{\zeta}_{k+1}^2}{\hat{\zeta}_k^2} + \log\hat{\zeta}_k^2\Big] < 0 \\ c_0' &= -\log\frac{\hat{\lambda}_{k+1}}{\hat{\zeta}_k^2} + (n-k-1)\log\frac{\hat{\zeta}_k^2}{\hat{\zeta}_{k+1}^2} > 0 \end{aligned}$$

Now solve the inequality and find the smallest $\tilde{\delta}$ such that $\zeta'(\tilde{\delta}) \leq 0$. Clearly, when $\tilde{\delta} = 0$, $\zeta'(\tilde{\delta}) > 0$. Since the discriminant ${c_1'}^2 - 4c_2'c_0' > 0$ is positive and $c_2' > 0$, there are two positive roots, $r_3(k)$ and $r_4(k)$, where

$$\begin{cases} \zeta'(\tilde{\delta}) < 0 \ \textit{for}\ \delta \in (r_3(k), r_4(k)) \\ \zeta'(\tilde{\delta}) > 0 \ \textit{for}\ \delta \in, (0, r_3(k)) \cup (r_4(k), max(G(k+1))). \end{cases}$$

Since $\tilde{\delta} \in (0, 1/(k+1) - 1/n)$, the root near 0 can be approximated by the Vieta's solution $-\frac{c_0}{c_1}$ and $-\frac{c_0'}{c_1'}$. So finally we have

$$u_b(k) = -\frac{\log\frac{\hat{\lambda}_k}{\hat{\zeta}_{k-1}^2} + (n-k)\log\frac{\hat{\zeta}_k^2}{\hat{\zeta}_{k-1}^2}}{n[k\log\hat{\zeta}_k^2 - (k-1)\log\hat{\zeta}_{k-1}^2]}$$

and

$$u_a(k) = -\frac{\log\frac{\hat{\lambda}_{k+1}}{\hat{\zeta}_k^2} + (n-k-1)\log\frac{\hat{\zeta}_k^2}{\hat{\zeta}_{k+1}^2}}{n[(k+1)\log\hat{\zeta}_{k+1}^2 - k\log\hat{\zeta}_k^2]}.$$

Since $u_b(k) = u_a(k-1)$, these approximated intervals are also non-overlapping.

Lemma 5. *Suppose k^* is the true rank of W, and denote $\zeta^2 = \zeta^2_{k^*}$, then as $m \to \infty$,*

- $u_b(k^*)/u_a(k^*) \to \infty$ *in probability*
- $|u_b(k) - u_a(k)| \to 0$ *in probability for* $k > k^*$.

Proof (Lemma 5)**.** *It is useful to have the following:*

$$\frac{\hat{\lambda}_{k+1}}{\hat{\zeta}^2_k} = \frac{n-k}{\sum_{i=k+2}^{n} \hat{\lambda}_i/\hat{\lambda}_{k+1} + 1} > 1,$$

$$\frac{\hat{\lambda}_k}{\hat{\zeta}^2_{k-1}} = \frac{n-k-1}{\sum_{i=k+1}^{n} \hat{\lambda}_i/\hat{\lambda}_k + 1} > 1,$$

$$\frac{\hat{\zeta}^2_k}{\hat{\zeta}^2_{k+1}} = \frac{\hat{\lambda}_{k+1}}{(n-k)\hat{\zeta}^2_{k+1}} + 1 - \frac{1}{n-k} > 1,$$

and

$$\frac{\hat{\zeta}^2_k}{\hat{\zeta}^2_{k-1}} = \frac{(n-k+1)\hat{\zeta}^2_k}{\hat{\lambda}_k + (n-k)\hat{\zeta}^2_k} < 1.$$

To show $u_b(k^*)/u_a(k^*) \to \infty$ in probability, it suffices to show that $u_a(k^*)/u_b(k^*) \to 0$ in probability. Suppose $u_a(k^*)$ and $u_b(k^*)$ converges in probability to 0 and a constant $cc > 0$, respectively, then every subsequence of these two sequences converges almost surely to 0 and cc, respectively. Thus, every subsequence of $u_a(k^*)/u_b(k^*)$ now has a further subsequence converging almost surely to 0. This completes the proof that $u_b(k^*)/u_a(k^*) \to \infty$ in probability.

On the other hand, to show $|u_b(k) - u_a(k)| \to 0$ in probability for $k > k^*$, it suffices to show $u_a(k) \to 0$ and $u_b(k) \to 0$ in probability as both $u_b(k)$ and $u_a(k)$ are positive.

Now, I will show both $u_a(k^*)$ and $u_a(k^*+1)$ converge in probability to 0 and $u_b(k^*)$ converges in probability to some constant $a > 0$. For these to hold, it suffices to show $c_0'(k^*) \to 0$, $c_0'(k^*+1) \to 0$ while $c_1'(k^*) \to \zeta^2 > 0$, $c_1'(k^*+1) \to \zeta^2 > 0$ and $c_0(k^*) \to aa > 0$, $c_1(k^*) \to aa' > 0$, where aa and aa' are non-zero constants.

Note that the denominators c_1, c_1' and numerators c_0, c_0' of $u_b(k)$ and $u_a(k)$ are smooth functions of $\hat{\zeta}^2_k$, $\hat{\zeta}^2_{k-1}$ and $\hat{\zeta}^2_{k+1}$ on $(0,1)$ as both first derivatives with respect to these random variables exist.

$$\frac{\partial c_0}{\partial \hat{\zeta}^2_k} = (\hat{\lambda}_k - \hat{\zeta}^2_k)\frac{n-k}{(\hat{\zeta}^2_k)^2(n-k+\frac{\hat{\lambda}_k}{\hat{\zeta}^2})}$$

$$\frac{\partial c_0'}{\partial \hat{\zeta}^2_k} = (\hat{\lambda}_{k+1} - \hat{\zeta}^2_k)\frac{n-k}{(\hat{\zeta}^2_k)^2(n-k-\frac{\hat{\lambda}_{k+1}}{\hat{\zeta}^2})}$$

$$\frac{\partial c_1}{\partial \hat{\zeta}^2_k} = -\frac{k\hat{\lambda}_k + \hat{\zeta}^2_k(n-k)}{\hat{\zeta}^2_k[(n-k)\hat{\zeta}^2_k - \hat{\lambda}_k]}$$

$$\frac{\partial c_1'}{\partial \hat{\zeta}^2_k} = -\frac{k\hat{\lambda}_{k+1} + \hat{\zeta}^2_k(n-k)}{\hat{\zeta}^2_k[(n-k)\hat{\zeta}^2_k - \hat{\lambda}_{k+1}]}$$

*Since the population value of the $n-k$ last eigenvalues are equal to $\zeta^2(k)$, Theorem 8.3.2 of [**?**] implies that for normal data, the maximum likelihood estimator $\hat{\zeta}^2_k = (\sum_{i=k+1}^{n} \hat{\lambda}_i)/(n-k)$ converges to ζ^2_k almost surely. Similarly, the maximum likelihood estimator of $\zeta^2(k+1)$ is $\hat{\zeta}^2_{k+1} = (\sum_{i=k+2}^{n} \hat{\lambda}_i)/(n-k-1)$ and also converges to ζ^2_k, combined with the fact that:*

$$\frac{1}{\hat{\zeta}^2_k}1_{\hat{\zeta}^2_k \neq 0} \to \frac{1}{\zeta^2_k}1_{\zeta^2_k \neq 0}$$

and log *is a continuous function on* $(0,\infty)$*, the continuous mapping theorem suggests the following result:*

$$\log\frac{\hat{\zeta}^2_{k+1}}{\hat{\zeta}^2_k}\to 0.$$

These together proves $c_1'(k^*)\to\zeta^2>0$ *in probability and similarly* $c_1'(k^*)\to\zeta^2$ *in probability.*

On the other hand, for $k>k^*$*, the* k*th sample eigenvalue* $\hat{\lambda}_k\to\zeta^2$ *almost surely combined with* (2.2) *via continuous mapping theorem suggests:*

$$\log\frac{\hat{\lambda}^2_{k+1}}{\hat{\zeta}^2_k}\to 0.$$

These together then proves $c_0'(k^*)\to 0$ *and similarly* $c_0'(k^*+1)\to\zeta^2$ *in probability.*

Lastly, since the k^**th sample eigenvalue* $\hat{\lambda}_{k^*}\to\zeta^2+d_k^2$ *almost surely and thus:*

$$\hat{\zeta}^2_{k-1}\to\frac{d_k^2+(n-k+1)\zeta^2}{n-k+1}\quad\textit{in probability},$$

and implying

$$\frac{\hat{\zeta}^2_k}{\hat{\zeta}^2_{k-1}}\to\frac{(n-k+1)\zeta^2}{d_k^2+(n-k)\zeta^2}<1\quad\textit{in probability},$$

Similarly,

$$\frac{\hat{\lambda}_k}{\hat{\zeta}^2_{k-1}}\to\frac{(n-k+1)(\zeta^2+d_k^2)}{d_k^2+(n-k+1)\zeta^2}>1\quad\textit{in probability.}$$

These together then proves $c_0(k^*)\to aa$*, where* $aa=\log\frac{(n-k+1)(\zeta^2+d_k^2)}{d_k^2+(n-k+1)}$ *and* $c_1(k^*)\to aa'$ *in probability, where* $aa'=(k-1)\log\frac{(n-k+1)\zeta^2}{d_k^2+(n-k)\zeta^2}+\zeta^2$.

2.3 Proofs

Proof (of Proposition 1)**.** *Following [?], we have:*

$$\begin{aligned}\hat{\lambda}_n &= \hat{\zeta}^2_{n-1}\\ \hat{\lambda}_{n-1} &= 2\hat{\zeta}^2_{n-2}-\hat{\lambda}_n=2\hat{\zeta}^2_{n-2}-\hat{\zeta}^2_{n-1}\\ \vdots &\ \vdots\ \vdots\\ \hat{\lambda}_k &= (n-k+1)\hat{\zeta}^2_{k-1}-(n-k)\hat{\zeta}^2_k\end{aligned}$$

Notice that the sample eigenvalues are decreasing $\hat{\lambda}_1>\cdots>\hat{\lambda}_n>0$*, thus implying* $\hat{\zeta}^2_1>\cdots>\hat{\zeta}^2_{n-1}$*. Therefore,*

$$\hat{\lambda}_k=(n-k+1)\hat{\zeta}^2_{k-1}-(n-k)\hat{\zeta}^2_k>\hat{\zeta}^2_{k-1}.$$

For $k\in\{2,\ldots,n-1\}$*, based on the inequality* $\frac{x}{1+x}<\log(1+x)<x$*, for any* $x>-1$*, we obtain*

$$\begin{aligned}l_p(k)-l_p(k-1) &= -\frac{m}{2}[\log\hat{\lambda}_k+(n-k)\log\hat{\zeta}^2_k-(n-k+1)\log\hat{\zeta}^2_{k-1}]\\ &> -\frac{m}{2}\Big[(n-k)(1-\frac{\hat{\zeta}^2_k}{\hat{\zeta}^2_{k-1}})+(n-k)\Big(\frac{\hat{\zeta}^2_k}{\hat{\zeta}^2_{k-1}}-1\Big)\Big]=0\end{aligned}$$

Proof (of Proposition 2)**.** *Though in theory,* $\tilde{\delta}\in(0,\infty)$*, the practical consideration that the penalized MLE of* ζ^2 *needs to be between 0 and 1 restricts* $\tilde{\delta}$ *to be in* $(0,1-1/n)$*.*

(27)

As shown in Lemma 1 and Lemma 2, when $\tilde{\delta} \in G(k+1)$*, where*

$$G(k+1) = \Big(0, \frac{n(n-k-1)(\hat{\lambda}_{k+1} - \hat{\zeta}^2_{k+1})}{(k+1)\hat{\lambda}_{k+1} + (n-k-1)\hat{\zeta}^2_{k+1}}\Big), \tag{28}$$

and $\cup_k G(k+1) \subset (0, 1-1/n)$*, the difference* $l_p(k;\tilde{\delta}) - l_p(k+1;\tilde{\delta})$ *is a smooth and concave function of* $\tilde{\delta}$ *and is increasing. Similarly,* $l_p(k;\tilde{\delta}) - l_p(k-1;\tilde{\delta})$ *is a smooth and convex function of* $\tilde{\delta}$ *and is decreasing. Naturally, an interval can be defined:*

$$\Delta_k = \Big\{\tilde{\delta} \in G(k);\ \textit{conditions (2) are satisfied}\Big\} = (a_k, b_k) \tag{29}$$

for some $k \in \{2, \ldots, n-2\}$*.*

Lemma 3 implies that on this interval, there exists $\tilde{\delta}_o \in G(k)$ *such that* $k^* = \operatorname{argmax}_{k'} l_p(k';\tilde{\delta}_o)$*, where* k^* *denote the rank of the parameter* W*.*

Following Lemmas 3 and 4, since $a_{k^*} < u(a_{k^*}) < u(b_{k^*}) < b_{k^*}$ *or* $\Delta_{k^*} \neq \varnothing$*, there must exist* $\tilde{\delta}_o \in \Delta_{k^*}$*, such that* $l_p(k^*;\tilde{\delta}_o) - l_p(k^*+1;\tilde{\delta}_o) > 0$ *and* $l_p(k^*;\tilde{\delta}_o) - l_p(k^*-1;\tilde{\delta}_o) > 0$*.*

Since the probability that two sample eigenvalues are identical is 0, then $u_a(k^*) < u_b(k^*)$ *implies the existence of some*

$$\tilde{\delta}_o \in (u(a_{k^*}), u(b_{k^*})) \subset (a_{k^*}, b_{k^*}) \subset G(k^*+1) \subset (0, 1-1/n), \tag{30}$$

which concludes the proof.

A final note, relating to the use of the data averaging heuristic, is that as $m \to \infty$, we have $\frac{u_b(k^*)}{u_a(k^*)} \to \infty$ and $|u_b(k) - u_a(k)| \to 0$ for $k > k^*$ such that $k^* = \operatorname{argmax}_{k'} l_p(k';\tilde{\delta}_o)$. In other words, the consistency of the data averaging procedure hinges on the convergence of sample eigenvalues to the true population values. In finite samples, the correct estimate will depend on both the rate of convergence, the signal-to-noise ratio, and additional factors that impact the behaviours of the sample eigenvalues.