Web-based supporting materials for “Concave likelihood-based regression with finite-support response variables” by Ekvall and Bottai

Karl Oskar Ekvall Matteo Bottai

Division of Biostatistics, Institute of Environmental Medicine, Karolinska Institutet

karl.oskar.ekvall@ki.se matteo.bottai@ki.se

A Technical details

A.1 Concavity of the log-likelihood

Lemma A.1. *For any log-concave Lebesgue density r on $\mathbb{R}$ the function $D : \{t = (t_1, t_2) \in \mathbb{R}^2 : t_1 < t_2\} \to [0, 1]$ defined by $D(t) = \int_{t_1}^{t_2} r(w)\,\mathrm{d}w$ is log-concave. Moreover, if r is strictly positive and strictly log-concave on an interval (a, b), $-\infty \leq a < b \leq \infty$, then D is strictly log-concave on $\{t \in \mathbb{R}^2 : a < t_1 < t_2 < b\}$.*

Remark. The first assertion of the lemma implies the cumulative distribution function R and the survival function $1 - R$ are both log-concave. The second assertion implies $t_1 \mapsto D(t_1, t_2)$ is strictly log-concave on (a, t_2) and $t_2 \mapsto D(t_1, t_2)$ is strictly log-concave on (t_1, b).

Proof. The first assertion follows from (i) the maps $(t_1, t_2, w) \mapsto \mathbb{I}(t_1 \leq w \leq t_2)$ and $(t_1, t_2, w) \mapsto r(w)$ are log-concave, (ii) the product of log-concave functions is log-concave, and (iii) integrating out one variable from a log-concave function on $\mathbb{R}^3$ gives a log-concave function on $\mathbb{R}^2$ (Prékopa, 1973, Theorem 6).

To prove the strict log-concavity we will use Theorem 4 of Prékopa (1973) in a way similar to the proof of their Theorem 5. Denote $\mathcal{T} = \{t \in \mathbb{R}^2 : a < t_1 < t_2 < b\}$, where $a < b$ and r is strictly positive and strictly log-concave on (a, b). Note $\mathcal{T}$ non-empty, convex, and open. Pick $u, v \in \mathcal{T}$, $u \neq v$. If $u_1 \neq v_1$, define the intervals $U = [u_1, u_2)$ and $V = [v_1, v_2)$. Define also $U_1 = [u_1, u_1 + \epsilon]$ and $V_1 = [v_1, v_1 + \delta]$, where $\epsilon > 0$ and $\delta > 0$ are small enough that $U_2 = U \setminus U_1$ and $V_2 = V \setminus V_1$ are non-empty.

We will omit the arguments for the case $u_1 = v_1, u_2 \neq v_2$ since they are very similar but with the definitions $U = (u_1, u_2]$, $V = (v_1, v_2]$, $U_1 = [u_2 - \epsilon, u_2]$, and $V_1 = [v_2 - \delta, v_2]$.

Now, with R denoting the distribution with cumulative distribution function R, we have $D(u) = \mathsf{R}(U)$, $D(v) = \mathsf{R}(V)$, and

$$D(su + (1-s)v) = \mathsf{R}(sU + (1-s)V),$$

where for sets addition is in the Minkowski sense and scalar multiplication is elementwise. Thus, we need to show $\mathsf{R}(sU + (1-s)V) > \mathsf{R}(U)^s \mathsf{R}(V)^{1-s}$. By construction, U_1 and U_2 are convex and partition U, U_1 is closed and bounded, and both U_1 and U_2 have positive R-measure. Similar statements apply to V_1, V_2, and V. This verifies condition a) and Equation (3.5) of Theorem 4 by Prékopa (1973). Condition d) holds because the convex hull of $U_1 \cup V_1$ is a closed interval contained in (a, b). It remains to verify their condition (b) and Equation (3.6).

Observe

$$\begin{aligned} sU_1 + (1-s)V_1 &= [su_1, s(u_1 + \epsilon)] + [(1-s)v_1, (1-s)(v_1 + \delta)] \\ &= [su_1 + (1-s)v_1, su_1 + (1-s)v_1 + s\epsilon + (1-s)\delta] \end{aligned}$$

and

$$\begin{aligned} sU_2 + (1-s)V_2 &= (s(u_1 + \epsilon), su_2) + ((1-s)(v_1 + \delta), (1-s)v_2) \\ &= (su_1 + (1-s)v_1 + s\epsilon + (1-s)\delta, su_2 + (1-s)v_2). \end{aligned}$$

Thus, Equations (3.1) and (3.2) in Prékopa (1973) hold by inspection. To ensure their Equations (3.3) and 3.4 also hold, note that as $\epsilon, \delta \to 0$, $sU_1 + (1-s)V_1$ shrinks towards the point $su_1 + (1-s)v_1$ and U_1 shrinks towards the point u_1. Thus, they are disjoint for small enough ϵ and δ because $u_1 \neq v_1$ and $s \in (0, 1)$. Similarly, V_1 and $sU_1 + (1-s)V_1$ are disjoint for small enough ϵ and δ and that verifies Equations (3.3) and (3.4). As argued in the proof of Theorem 5 in Prékopa (1973), their Equation (3.6), which says $\mathsf{R}(U_2)/\mathsf{R}(U_1) = \mathsf{R}(V_2)/\mathsf{R}(V_1)$, can be made to hold because the left-hand side does not depend on δ and tends to infinity as $\epsilon \to 0$, and the right-hand side does not depend on ϵ and tents to infinity as $\delta \to 0$. We conclude all the sufficient conditions hold, and hence $\mathsf{R}(sU + (1-s)V) > \mathsf{R}(U)^s \mathsf{R}(V)^{1-s}$ as desired.

$\square$

Proof of Theorem 1. The non-strict log-concavity follows from the non-strict log-concavity given by Lemma A.1 and the fact that the composition of a concave and affine function is concave. To prove the strict part, note continuous differentiability of r implies $\nabla^2 \ell_n(\theta; y^n, x^n)$ exists and is equal to

$$\sum_{i=1}^{n} Z_i^\mathsf{T} H_i Z_i,$$

where $H_i = H(y_i, x_i, \theta)$ is the Hessian of the function D in lemma A.1 evaluated at $t_1 = a(y_i, x_i, \theta)$ and $t_2 = b(y_i, x_i, \theta)$. To be precise, when $a(y_i, x_i, \theta) = -\infty$ we set the first row and column of H_i to zero, and similarly with the second row and column when $b(y_i, x_i, \theta) = \infty$. Recall, in the former case we also defined the first row of Z_i to be zero, so for such y_i we have

$$Z_i^\mathsf{T} H_i Z_i = H_{22}(y_i, x_i, \theta) z^b(y_i, x_i) z^b(y_i, x_i)^\mathsf{T} = H_{22}(y_i, x_i, \theta) Z_i^\mathsf{T} Z_i.$$

Similarly, for y_i such that $b(y_i, x_i, \theta) = \infty$ we have

$$Z_i^\mathsf{T} Z_i = H_{11}(y_i, x_i, \theta) z^a(y_i, x_i) z^a(y_i, x_i)^\mathsf{T} = H_{11}(y_i, x_i, \theta) Z_i^\mathsf{T} Z_i.$$

By Lemma A.1, H_{11} and H_{22} are strictly positive. And, since r is continuously differentiable, they are continuous functions in θ. Thus, they attain non-zero infima on compact subsets of Θ. Similarly, for y_i such that $a(y_i, x_i, \theta)$ and $b(y_i, x_i, \theta)$ are both finite, $H(y_i, x_i, \theta)$ is positive definite by Lemma A.1 and its smallest eigenvalue is bounded away from zero on compact subsets of Θ. Thus, for every $i \in \{1, \ldots, n\}$ and any compact subset A of Θ, we can find find $c_i = c_i(A) > 0$ such that $Z_i^\mathsf{T} H_i Z_i \succeq c_i Z_i^\mathsf{T} Z_i$, and hence

$$\sum_{i=1}^{n} Z_i^\mathsf{T} H_i Z_i \succeq \sum_{i=1}^{n} c_i Z_i^\mathsf{T} Z_i \succeq \min_{i \in \{1, \ldots, n\}} c_i \sum_{i=1}^{n} Z_i^\mathsf{T} Z_i,$$

which is positive definite. From this the conclusion follows since any two points $\theta_1, \theta_2 \in \mathbb{R}^d$ are contained in a large enough compact ball. $\square$

A.2 Asymptotics with fixed number of parameters

We state two ancillary lemmas before before proving Theorem 3.1.

Lemma A.2 (Bartlett identities). *If r is strictly positive and continuous, then $\mathbb{E}_\theta\{\nabla \ell(\theta; Y, x)\} = 0$ and $-\mathbb{E}_\theta\{\nabla^2 \ell(\theta; Y, xY)\} = \mathrm{cov}_\theta\{\nabla \ell(\theta; Y, x)\}$ for every $x \in \mathbb{R}^p$ and $\theta \in \Theta$.*

Proof. By a classical argument, it suffices to show we can differentiate twice under the integral in the identity $\int f_\theta(y \mid x)\,\mathrm{d}y$, where $\mathrm{d}y$ indicates integration with respect to the measure against which Y has density $f_\theta(y \mid x)$. We show it can be done once by showing that for every $\theta' \in \Theta$ there exists a neighborhood on which $\|\nabla f_\theta(y \mid x)\|_1$ is bounded by an integrable function of y not depending on θ (Folland, 2007, Theorem 2.27). We have

$$\begin{aligned}\|\nabla f_\theta(y \mid x)\|_1 &= \|r(b(y,x,\theta))z^b(y,x) - r(a(\theta,y,x))z^a(y,x)\|_1 \\ &\leq |r(b(y,x,\theta))|\|z^b(y,x)\|_1 + |r(a(\theta,y,x))|\|z^a(y,x)\|_1\end{aligned}$$

with z^a and z^b replaced by a vector of zeros if, respectively, $a(y,x,\theta) = -\infty$ or $b(y,x,\theta) = \infty$. Now, since the one-norms of z^a and z^b are continuous in y and x, they bounded on compact sets, and hence both bounded by some $c < \infty$ on $\mathcal{Y} \times \mathcal{X}$, where $\mathcal{X} = \{x \in \mathbb{R}^p : \|x\|_\infty \leq M\}$. Similarly, for every $y \in \mathcal{Y}$ such that $a(y,x,\theta) > -\infty$, around any $\theta' \in \Theta$ there is a compact ball B such that

$$c_a(y) := \sup_{(x,\theta)\in\mathcal{X}\times B} r(a(y,x,\theta)) < \infty.$$

Define also $c_a(y) = 0$ for y at which a is $-\infty$. By similar arguments,

$$c_b(y) := \sup_{(x,\theta)\in\mathcal{X}\times B} r(b(y,x,\theta)) < \infty,$$

and we take $c_b(y) = 0$ for y at which b is ∞. We thus have the dominating function $c\{c_a(y) + c_b(y)\}$ which has finite expectation because it is finite at every $y \in \mathcal{Y}$. We omit the arguments for second-order derivatives since they are very similar. $\square$

Lemma A.3. *If the density r is strictly positive and continuously differentiable and, for every $t \in \mathbb{R}^d$, as $n \to \infty$,*

$$\frac{1}{n}\sum_{i=1}^n \mathbb{E}\left[\int_0^1 \{\nabla^2\ell(\theta_* + st/\sqrt{n}; Y_i, x_i) - \nabla^2\ell(\theta_*; Y_i, x_i)\}s\,\mathrm{d}s\right] \to 0, \tag{1}$$

$$\frac{1}{n^2}\sum_{i=1}^n \operatorname{cov}\left\{\int_0^1 \nabla^2\ell(\theta_* + st/\sqrt{n}; Y_i, x_i)s\,\mathrm{d}s\right\} \to 0, \tag{2}$$

and

$$\frac{1}{n}\sum_{i=1}^n \operatorname{cov}\{\nabla\ell(\theta_*; Y_i, x_i)\} = n^{-1}\mathcal{I}_n(\theta_*; x^n) \to \mathcal{I}(\theta_*) \tag{3}$$

for some positive definite $\mathcal{I}(\theta_*)$; then

$$\sqrt{n}(\hat{\theta}_n - \theta_*) = \mathcal{I}(\theta_*)^{-1} n^{-1/2} \sum_{i=1}^{n} \nabla \ell(\theta_*; Y_i, x_i) + o_{\mathsf{P}}(1).$$

Proof. We verify the conditions of Theorem 2.2 in Hjort and Pollard (2011) from which the conclusion follows. Define the remainder D_i in a linear approximation of ℓ around the true parameter θ_* by

$$D_i = D_i(Y_i, x_i, t) = \ell(\theta_* + t; Y_i, x_i) - \ell(\theta_*; Y_i, x_i) - \nabla \ell(\theta_*; Y_i, x_i)^{\mathsf{T}} t.$$

The likelihood has continuous second order partial derivatives because r is continuously differentiable and hence we can use the mean value theorem with integral-form remainder to write

$$D_i = t^{\mathsf{T}} \left[\int_0^1 \nabla^2 \ell(\theta_* + st; Y_i, x_i) s \, \mathrm{d}s \right] t.$$

With that, some algebra shows (1) and (2) are equivalent to, respectively, $\sum_{i=1}^n v_{i,0}(t/\sqrt{n}) \to 0$, where $v_{i,0}(t) = \mathbb{E}(D_i) - t^{\mathsf{T}} \mathbb{E}\{\nabla^2 \ell(\theta_*; Y_i, x_i)\} t$, and $\sum_{i=1}^n v_i(t/\sqrt{n}) \to 0$, where $v_i(t) = \mathrm{var}(D_i)$. □

Remark. By specializing the remarks following Theorem 2.2 in Hjort and Pollard (2011) to the present setting one sees that condition (3) can be weakened to

$$0 < \liminf_{n \to \infty} n^{-1} \lambda_{\min}\{\mathcal{I}_n(\theta_*; x^n)\} \leq \limsup_{n \to \infty} n^{-1} \lambda_{\max}\{\mathcal{I}_n(\theta_*; x^n)\} < \infty,$$

in which case the conclusion is

$$\mathcal{I}_n(\theta_*; x^n)^{1/2}(\hat{\theta} - \theta_*) = \mathcal{I}_n(\theta_*; x^n)^{-1/2} \nabla \ell_n(\theta_*; Y^n, x^n) + o_{\mathsf{P}}(1).$$

Proof of Theorem 2. We start by verifying (1)–(2) in Lemma A.3. For the former, consider

$$\nabla^2 \ell(\theta; y_i, x_i) = -Z_i^{\mathsf{T}} H\{a(y_i, x_i, \theta), b(y_i, x_i, \theta)\} Z_i$$

and fix arbitrary y_i. Because r is continuously differentiable, $\nabla^2 \ell(\theta; y_i, x_i)$ is continuous in (θ, x_i). Thus, it is uniformly continuous on the compact set $B \times \{x \in \mathbb{R}^p : \|x\|_1 \leq M\}$, where B is a compact neighborhood of θ_*. Thus, for every $\epsilon > 0$ and $y \in \mathcal{Y}$ there is an $N(y) < \infty$

such that

$$\sup_{x\in\mathbb{R}^p:\|x\|_1\le M} |\nabla^2\ell(\theta_* + st/\sqrt{n}; y, x) - \nabla^2\ell(\theta_*; y_i, x_i)| \le \epsilon$$

for all $n \ge N(y)$. Because $\mathcal{Y}$ is finite, the last display holds uniformly in y for all n greater than $\max_{y\in\mathcal{Y}} N(y) < \infty$, and (1) follows upon sending $\epsilon \to 0$.

By a similar argument, the elements of $\nabla^2\ell(\theta + st/\sqrt{n}; y, x)$ are bounded uniformly in $(y, x) \in \mathcal{Y} \times \{x \in \mathbb{R}^p : \|x\|_1 \le M\}$ for all large enough n, and hence (2) follows. From this it also follows that the eigenvalues of $n^{-1}\mathcal{I}_n(\theta; x^n) = -n^{-1}\sum_{i=1}^n \mathbb{E}\{\nabla^2\ell(\theta_*; Y_i, x_i)\}$ are bounded. Thus, by remarks follow Lemma A.3, we are done if we can verify that

$$\mathcal{I}_n(\theta_*; x^n)^{-1/2}\ell_n(\theta_*; Y^n, x^n) \rightsquigarrow \mathcal{N}(0, I_d).$$

This is straightforwardly done in two steps: first, for any $t \in \mathbb{R}^d$

$$\sum_{i=1}^n t^\mathsf{T}\nabla\ell(\theta_*; Y_i, x_i)/\sqrt{t^\mathsf{T}\mathcal{I}_n(\theta_*; Y_i, x_i)t} \rightsquigarrow \mathcal{N}(0, 1)$$

by Lyapunov's central limit theorem, using that the elements of $\nabla\ell(\theta_*; Y_i, x_i)$ are uniformly bounded and hence have uniformly bounded third (say) moment. Then, the conclusion follows from the Cramér-Wold theorem and $0 < \liminf_{n\to\infty} n^{-1}\lambda_{\min}\{\mathcal{I}_n(\theta_*; x^n)\} \le \limsup_{n\to\infty} n^{-1}\lambda_{\max}\{\mathcal{I}_n(\theta_*; x^n)\} < \infty$ (Biscio et al., 2018). □

A.3 Asymptotics with diverging number of parameters

We will use the framework of Negahban et al. (2012) to prove Theorem 4.1. To that end we establish a concentration inequality for the gradient of the objective function and a restricted strong convexity.

Lemma A.4. *If (i) r is continuous and strictly positive on $\mathbb{R}$, (ii) θ_* is s-sparse and $\|\theta_*\|_\infty \le C_1 < \infty$, and (iii) $|x_{ij}| \le C_2 < \infty$; then there exists a $c_1 \in (0, \infty)$ such that, for any $t \ge 0$,*

$$\mathsf{P}\left(\|\nabla G_n(\theta_*; Y^n, x^n)\|_\infty > t\right) \le 2d\exp\left(-c_1 n t^2\right)$$

Proof. Throughout the proof $M < \infty$ is a generic constant which can change between statements. The gradient at θ_* is

$$\nabla G_n(\theta_*; y^n, x^n) = -n^{-1}\sum_{i=1}^n \frac{z^b(y_i, x_i)r(b(y_i, x_i, \theta_*)) - z^a(y_i, x_i)r(a(y_i, x_i, \theta_*))}{R(b(y_i, x_i, \theta_*)) - R(a(y_i, x_i, \theta_*))}.$$

First note, by (ii), $\|Z_i\theta_*\|_1 \le \|Z_i\|_\infty sM$. Moreover, for each $y \in \mathcal{Y}$, because each element of Z_i is a continuous function on the compact $\mathcal{X} = \{x \in \mathbb{R}^p : \|x\|_\infty \le M\}$, it is bounded uniformly in $x \in \mathcal{X}$. Thus, because $\mathcal{Y}$ is finite, it is also bounded uniformly $(y, x) \in \mathcal{Y} \times \mathcal{X}$. Thus, using continuity of R, $R(b(y, x, \theta_*)) - R(a(y, x, \theta_*))$ is bounded away from zero uniformly in y, x, n, and p. By similar arguments, using continuity of r, every element of the vectors in the numerator in the last display is bounded in absolute value by an M not depending on y, x, n, or p. It follows that each element in each summand in the last display is uniformly bounded in absolute value by an $M < \infty$. We then get by Hoeffding's inequality, using that the gradient has mean zero by Lemma A.2, for $j \in \{1, \dots, d\}$,

$$\mathsf{P}\left(|[\nabla G_n(\theta_*; Y^n, x^n)]_j| > t\right) \le 2\exp\left(-\frac{2nt^2}{M^2}\right).$$

Thus, by a union bound (sub-additivity of measures),

$$\mathsf{P}\left(\|\nabla G_n(\theta_*; Y^n, x^n)\|_\infty > t\right) \le 2d\exp\left(-\frac{2nt^2}{M^2}\right),$$

which finishes the proof. □

Lemma A.5. *Let $\bar{B}_M$ denote the closed ball of radius M centered at the origin in $\mathbb{R}^d$. For every $M > 0$, there exists a $\kappa_M > 0$ not depending on n or p such that, for $\Delta \in \mathbb{C}(S) \cap \bar{B}_M$ and realizations (y^n, x^n) in the event $\mathcal{C}_{\kappa,n,p}$ in Theorem 4.1,*

$$G_n(\theta_* + \Delta; y^n, x^n) - G_n(\theta_*; y^n, x^n) - \nabla G_n(\theta_*; y^n, x^n)^\mathsf{T}\Delta \ge \kappa_M\|\Delta\|^2.$$

Proof. Continuity of the derivative of r ensures $\nabla^2 G_n(\cdot; y^n, x^n)$ is continuous, and hence, by the mean-value theorem, the left-hand side in the inequality to be established is equal to

$$\Delta^\mathsf{T}\nabla^2 G_n(\theta_* + \tilde{\Delta}; y^n, x^n)\Delta/2$$

for some $\tilde{\Delta}$ on the line connecting 0 and Δ. Let $\tilde{\theta} = \theta_* + \tilde{\Delta}$. Following the same arguments as in the proof of Theorem 2, we see it now suffices to show that $H_{11}(y, x, \tilde{\theta})$, $H_{22}(y, x, \tilde{\theta})$, and $\lambda_{\min}\{H(y, x, \tilde{\theta})\}$ are bounded away from zero when, respectively, $b(y, x, \tilde{\theta}) = \infty$, $a(y, x, \tilde{\theta}) = -\infty$, and both are finite. To that end, observe $\|\tilde{\theta}\|_1 \le \|\theta_*\|_1 + \|\tilde{\Delta}\|_1 \le \sqrt{s}M + 4\|\Delta_S\|_1 \le 5\sqrt{s}M$. Thus, for example, $\|z^a(y, x)^\mathsf{T}\tilde{\theta}\| \le \|z^a(y, x)\|_\infty\|\tilde{\theta}\|_1 \le \|z^a(y, x)\|_\infty\|5sM$, and $\|z^a(y, x)\|_\infty$ is bounded uniformly in y and x (and hence, d). Thus, for y such that $b(y, x, \theta) = \infty$, we have that $H_{11}(y, x, \tilde{\theta})$ is a strictly positive, continuous function of $z^a(y, x)^\mathsf{T}\tilde{\theta}$

in a compact set not depending on d. Thus, because continuous function attain their infimum on compact sets, we have established the required uniform lower bound of $H_{11}(y, x, \tilde{\theta})$. The bounds for H_{22} and $\lambda_{\min}(H)$ follow by very similar arguments which we omit. $\square$

Before giving a proof of Theorem 3 we recall Corollary 1 of Negahban et al. (2012) and specialize it to our setting in the following lemma.

Lemma A.6. *If $\lambda_n > 2\|\nabla G_n(\theta_*)\|_\infty$ and conditions (a)–(d) of Theorem 3 hold, then there exists $c_1, c_2 < \infty$ such that for large enough n, d, and every outcome in the set $\mathcal{C}_{\kappa,n,d}$, $\|\hat{\theta}_n^\lambda - \theta_*\| \leq c_1 \lambda_n^2$ and $\|\hat{\theta}_n^\lambda - \theta_*\|_1 \leq c_2 \lambda_n$.*

Proof. Condition (G1) in (Negahban et al., 2012) is satisfied because $\|\cdot\|_1$ is a norm. Theorem 1 and Lemma A.5 show their condition (G2) is satisfied on any compact ball centered at the origin, which is sufficient (Negahban et al., 2012, p. 9). More specifically, pick an $M \in (0, \infty)$ and note that for any $\Delta \in \mathbb{C}(S) \cap B_M$ it holds that (Negahban et al., 2012, Supplementary Material, p.29)

$$G_n^\lambda(\theta_* + \Delta) \geq G_n^\lambda(\theta_*) + \kappa_M \|\Delta\|^2 - 3\sqrt{s}\lambda_n \|\Delta\|/2.$$

Thus for all large enough n and d and $\Delta \in \mathbb{C}(S) \cap \{\|\Delta\| = M\}$, since $\lambda_n = o(1)$, $G_n^\lambda(\theta_* + \Delta) > G_n^\lambda(\theta_*)$. Hence, by convexity, $\|\hat{\theta}_n^\lambda - \theta_*\| \leq M$. Because $\|\theta_*\|$ is bounded as d varies, this shows $\hat{\theta}_n^\lambda$ is in a compact ball of fixed radius for all large enough n and d. The proof of Theorem 1 in Negahban et al. (2012) now applies almost verbatim, with the "global" $\kappa_{\mathcal{L}}$ (their notation) replaced by the κ_M given by Lemma A.5, for some large enough M. $\square$

Proof of Theorem 3. By Lemma A.4 we can pick $\lambda_n^2 = c_3 \log(d)/n$ and have that the probability of the event $\mathcal{A}_{n,d} = \{(Y^n, x^n) : \|\nabla G_n(\theta_*; Y^n, x^n)\|_\infty > 2\lambda_n\}$ is upper bounded by $2\exp\{-c_7 n \lambda_n + \log(d)\}$ for all n and d and some $c_7 > 0$. Thus, by picking large enough c_3 we get that $\mathcal{A}_{n,d}$ happens with probability at most d^{-c_4} for some $c_4 > 0$. Thus, the result follows from Lemma A.6 and noting that $\mathsf{P}(\mathcal{C}_{\kappa,n,d} \cap \mathcal{A}_{n,d}^c) = \mathsf{P}(\mathcal{C}_{\kappa,n,d}) - \mathsf{P}(\mathcal{C}_{\kappa,n,d} \cap \mathcal{A}_{n,d}) \geq \mathsf{P}(\mathcal{C}_{\kappa,n,d}) - d^{-c_4}$, and that completes the proof. $\square$

A.4 Convergence of algorithm

Proof of Theorem 4. We check the conditions of Theorem 2 by Byrd et al. (2016). The termination criteria ensure the descent property $G_n^\lambda(\theta^k) \geq G_n^\lambda(\theta^{k+1})$ (Byrd et al., 2016, p.5). Moreover, under either of conditions (a) and (b) G_n^λ is strictly convex by Theorem 2.1. Thus, for any starting value θ^0, the sequence of iterates are contained in a large enough compact ball B centered at $\hat{\theta}$, which under condition (b) exists because G_n^λ is strongly convex. We have by continuity and strict convexity that $\sup_{\theta \in B} \|\nabla G_n(\theta) + \lambda_2 \theta\| < \infty$, $\inf_{\theta \in B} \lambda_{\min}\{\nabla^2 G_n(\theta) + \lambda_2 I_d\} > 0$, and $\sup_{\theta \in B} \lambda_{\max}\{\nabla^2 G_n(\theta) + \lambda_2 I_d\} < \infty$. The bound on the gradient implies the required Lipschitz-continuity and the eigenvalue bounds imply the eigenvalues of $\nabla^2 G_n(\theta^k) + \lambda_2 I_d$ are bounded away from zero and from above for all $k \in \{1, 2, \dots\}$, which completes the proof. □

References

Biscio, C. A. N., Poinas, A. and Waagepetersen, R. (2018) A note on gaps in proofs of central limit theorems. *Statistics & Probability Letters*, **135**, 7–10.

Byrd, R. H., Nocedal, J. and Oztoprak, F. (2016) An inexact successive quadratic approximation method for L-1 regularized optimization. *Mathematical Programming*, **157**, 375–396.

Folland, G. B. (2007) *Real Analysis: Modern Techniques and Their Applications*. New York: Wiley, second edn.

Hjort, N. L. and Pollard, D. (2011) Asymptotics for minimisers of convex processes.

Negahban, S. N., Ravikumar, P., Wainwright, M. J. and Yu, B. (2012) A unified framework for high-dimensional analysis of M-estimators with decomposable regularizers. *Statistical Science*, **27**.

Prékopa, A. (1973) On logarithmic concave measures and functions. *Acta Scientiarum Mathematicarum*, **34**, 335–343.