Shashank Singh[1]

3 Binary sequences of piecewise constant expectation [30 points]

(a) Let $m = n - 1$, and, for each $j \in \{1, \dots, m\}$ let $d_j := (0, \dots, 0, 1, -1, 0, \dots, 0) \in \mathbb{R}^n$, with $d_{j,j} = 1$ and $d_{j,j+1} = -1$. Then, for $x_i = e_i$, (4) is clearly equivalent to

$$\min_{\beta \in \mathbb{R}^p} \sum_{i=1}^n -t_i \cdot x_i^T \beta + \log\left(1 + \exp(x_i^T \beta)\right) + \lambda \sum_{j=1}^m |d_j \beta|,$$

which is of the form considered in Homework 3. $D \in \mathbb{R}^{m \times n}$ is the matrix with j^{th} row d_j. ■

(b) For notational convenience, we note that, if $H : [0,1] \to [0, \infty)$ is the entropy function

$$H(p) = -p \log p - (1-p)\log(1-p),$$

then $g(u) = \sum_{t=1}^n H(y_t(D^T u)_t)$. Since $\frac{d}{dp} H(p) = -\log\left(\frac{p}{1-p}\right)$, by the Chain Rule,

$$\nabla g(u) = -\sum_{t=1}^n \log\left(\frac{y_t(D^T u)_t}{1 - y_t(D^T u)_t}\right) y_t \nabla (D^T u)_t = -\sum_{t=1}^n \log\left(\frac{y_t(D^T u)_t}{1 - y_t(D^T u)_t}\right) y_t d_t^T = Dc(u),$$

where, for $t \in \{1, \dots, n\}$,

$$(c(u))_t = -\log\left(\frac{y_t(D^T u)_t}{1 - y_t(D^T u)_t}\right) y_t.$$

Since $\frac{d}{dp} \log\left(\frac{p}{1-p}\right) = \frac{1}{p(1-p)}$,

$$\nabla (c(u))_t = -\frac{y_t d_t^T}{y_t(D^T u)_t(1 - y_t(D^T u)_t)} = -\frac{d_t^T}{(D^T u)_t(1 - y_t(D^T u)_t)},$$

so $\nabla c(u) = W(u) D^T$, where $W(u)$ is the diagonal matrix with

$$W_{t,t}(u) = -\frac{1}{(D^T u)_t(1 - y_t(D^T u)_t)},$$

and hence $\nabla^2 g(u) = D \nabla c(u) = D W(u) D^T$.

The log barrier function is

$$\phi(u) = \sum_{t=1}^n \log(y_t(D^T u)_t) + \log(1 - y_t(D^T u)_t) + \sum_{i=1}^m \log(\lambda - u_i) + \log(u_i + \lambda).$$

[1] sss1@andrew.cmu.edu

Thus, for each $i \in \{1, \dots, m\}$,

$$
\begin{aligned}
(\nabla(\phi(u))_i &= \frac{1}{u_i - \lambda} + \frac{1}{u_i + \lambda} + \sum_{t=1}^{n} \frac{2y_t(D^T u)_t - 1}{y_t(D^T u)_t(y_t(D^T u)_t - 1)}(y_t d_t^T) \\
&= \frac{2u_i}{u_i^2 - \lambda^2} + \sum_{t=1}^{n} \frac{2y_t(D^T u)_t - 1}{(D^T u)_t(y_t(D^T u)_t - 1)} d_t^T
\end{aligned}
$$

Thus, $\nabla\phi(u) = a(u) + Db(u)$, where for each $i \in \{1, \dots, m\}$,

$$
(a(u))_i = \frac{2u_i}{u_i^2 - \lambda^2} \quad \text{and} \quad (b(u))_i = \sum_{t=1}^{n} \frac{2y_t(D^T u)_t - 1}{(D^T u)_t(y_t(D^T u)_t - 1)}.
$$

Since $(\nabla(a(u))_i)_i = -\dfrac{2(u_i^2 + 1)}{(u_i^2 - 1)^2}$,

$$
\nabla^2\phi(u) = U(u) + DV(u)D^T,
$$

where $U(u)$ and $W(u)$ are diagonal matrices with

$$
U_{i,i}(u) = -\frac{2(u_i^2 + 1)}{(u_i^2 - 1)^2} \quad \text{and} \quad V_{t,t}(u) = \frac{-2(y_t(D^T)_t)^2 + 2y_t(D^T)_t - 1}{(D^T u)_t^2(y_t(D^T u)_t - 1)^2}.
$$

The Newton step is

$$
\begin{aligned}
&\left[\nabla^2(\tau g(u) + \phi(u))\right]^{-1} \nabla(\tau g(u) + \phi(u)) \\
&\qquad = \left[D(\tau W(u) + V(u))D^T + U(u)\right]^{-1} \nabla(\tau D(c(u) + b(u)) + a(u)).
\end{aligned}
$$

(c) See attached code.

(d)