A Verification of Iterative Algorithms for Computing G_t

In this section, we verify that the iterative algorithm for computing G_t is going to converge in the binary case. The proof for the multiclass case follows immediately as a simple extension. We only need to verify that $\tilde{a}_{(k)}$ converges to the corresponding $\tilde{a}$ of a such that the value of G_t normalizes the sum.

First of all, given a, since $t > 1$ and $Z(\tilde{a}) > 1$, it is clear that $0 < \tilde{a} < a$. On the domain of $0 < u < a$, it is easy to verify that $Z(u)^{1-t}a - u$ is a monotonically decreasing function and it crosses at 0 only at $\tilde{a}$. Therefore, when $\tilde{a}_{(k)} > \tilde{a}$, $\tilde{a}_{(k+1)} < \tilde{a}_{(k)}$; when $\tilde{a}_{(k)} < \tilde{a}$, $\tilde{a}_{(k+1)} > \tilde{a}_{(k)}$.

We then prove that $\tilde{a}_{(k)}$ is a monotonically decreasing sequence. We prove this by mathematical induction. Since $\tilde{a}_{(0)} = \hat{a}$, $\tilde{a}_{(1)} < a = \tilde{a}_{(0)}$. Next assume that in the k-th iteration, $\tilde{a}_{(k)} < \tilde{a}_{(k-1)}$. Since $Z(\tilde{a}_{(k)}) > Z(\tilde{a}_{(k-1)})$, we have $\tilde{a}_{(k+1)} < \tilde{a}_{(k)}$. Therefore, it follows that $\tilde{a}_{(k)}$ is monotonically decreasing and it is lower bounded by $\tilde{a}$. Furthermore, $\lim_{k\to+\infty} \tilde{a}_{(k)}$ exists.

Finally,

$$\begin{aligned}\lim_{k\to+\infty} \tilde{a}_{(k)} &= \lim_{k\to+\infty} \tilde{a}_{(k+1)} \\ &= \lim_{k\to+\infty} Z(\tilde{a}_{(k)})^{1-t}a \\ &= Z(\lim_{k\to+\infty} \tilde{a}_{(k)})^{1-t}a, \qquad \text{(A.1)}\end{aligned}$$

where (A.1) holds because $Z(u)^{1-t}$ is continuous in u. Therefore, it follows that $\lim_{k\to+\infty} \tilde{a}_{(k)} = \tilde{a}$.

For the binary case when $t = 2$, note that

$$\exp_t(x) = (1-x)^{-1} \quad \text{and} \quad \log_t(x) = 1 - x^{-1}\,.$$

The value $G_t(a)$ needs to satisfy

$$\begin{aligned}1 &= \exp_t(\frac{a}{2} - G_t(a)) + \exp_t(-\frac{a}{2} - G_t(a)) \\ &= \frac{1}{1 + {}^{a}\!/_{2} + G_t(a)} + \frac{1}{1 - {}^{a}\!/_{2} + G_t(a)} \\ &= \frac{2\,(1+G_t(a))}{(1+G_t(a))^2 - {}^{a^2}\!/_{4}}\,,\end{aligned}$$

which yields

$$(1+G_t(a))^2 - \frac{a^2}{4} = 2\,(1+G_t(a))\,.$$

By cancelling the terms from both sides, we have

$$G_t(a)^2 = \frac{a^2}{4} + 1\,.$$

Since $G_t(a) \geq 0$, we have $G_t(a) = \sqrt{{}^{a^2}\!/_{4} + 1}$.

B Proof of Remark 1

For the surrogate loss

$$\xi_{t_1}^{t_2}(a) = -\log_{t_1} \exp_{t_2}(a/2 - G_{t_2}(a)),$$

we have

$$\begin{aligned}\frac{\partial \xi_{t_1}^{t_2}(a)}{\partial a} &= -\hat{p}_{t_2}(a)^{t_2-t_1}\left(\frac{1}{2} - \partial G_{t_2}(a)\right), \\ \frac{\partial^2 \xi_{t_1}^{t_2}(a)}{\partial a^2} &= \hat{p}_{t_2}(a)^{t_2-t_1} \times \qquad \text{(B.1)} \\ &\left[\partial^2 G_{t_2}(a) - (t_2 - t_1)\,\hat{p}_{t_2}(a)^{t_2-1}\left(\frac{1}{2} - G_{t_2}(a)\right)^2\right],\end{aligned}$$

where we define $\hat{p}_{t_2}(a) := \exp_{t_2}(a/2 - G_{t_2}(a))$ and $\partial G_{t_2}(a)$ and $\partial^2 G_{t_2}(a)$ are given as follows.

$$\partial G_{t_2}(a) = \frac{1}{2}\frac{\sum_c c\,\exp_{t_2}(\frac{c}{2}a - G_{t_2}(a))^{t_2}}{\sum_c \exp_{t_2}(\frac{c}{2}a - G_{t_2}(a))^{t_2}}\,, \qquad \text{(B.2)}$$

$$\partial^2 G_{t_2}(a) = \frac{t_2 \sum_c \exp_{t_2}(\frac{c}{2}a - G_{t_2}(a))^{2t_2-1}\left[\frac{c}{2} - \partial G_{t_2}(a)\right]^2}{\sum_c \exp_{t_2}(\frac{c}{2}a - G_{t_2}(a))^{t_2}}. \qquad \text{(B.3)}$$

For $t_2 = t_1 \geq 1$, we have

$$\frac{\partial^2 \xi_{t_1}^{t_2}(a)}{\partial a^2} = \partial^2 G_{t_2}(a) \geq 0\,,$$

which can be verified from (B.3). Moreover, for $t_1 \geq 1$ and $t_1 \geq t_2$, we have

$$\begin{aligned}\frac{\partial^2 \xi_{t_1}^{t_2}(a)}{\partial a^2} &= \frac{1}{\hat{p}_{t_2}(a)^{t_1-t_2}} \times \\ &\quad \left[\partial^2 G_{t_2}(a) + (t_1 - t_2)\,\hat{p}_{t_2}(a)^{t_2-1}\left(\frac{1}{2} - G_{t_2}(a)\right)^2\right] \\ &\geq \partial^2 G_{t_2}(a) + (t_1 - t_2)\,\hat{p}_{t_2}(a)^{t_2-1}\left(\frac{1}{2} - G_{t_2}(a)\right)^2 \\ &\geq \partial^2 G_{t_2}(a) \geq 0\,. \qquad \text{(B.4)}\end{aligned}$$

Thus, the loss is convex, similar to the latter case.

Now, consider the case $t_2 \geq t_1$. Suppose $\hat{p}_{t_2}(-a) = (1 - \hat{p}_{t_2}(a)) = \lambda\,\hat{p}_{t_2}(a)$ for some $\lambda \geq 0$. Substituting for $\hat{p}_{t_2}(-a)$ in (B.2) and (B.3), we can write (B.1) as

$$\begin{aligned}\frac{\partial^2 \xi_{t_1}^{t_2}(a)}{\partial a^2} &= \hat{p}_{t_2}(a)^{t_2-1}\frac{1}{(1+\lambda^{t_2})^2} \\ &\quad \times \left[t_2\left(\frac{1+\frac{1}{\lambda}}{1+\lambda^{t_2}}\right) - (t_2 - t_1)\right].\end{aligned}$$

For sufficiently small (respectively, large) value of λ, we have $\frac{\partial^2 \xi_{t_1}^{t_2}(a)}{\partial a^2} > 0$ (respectively, $\frac{\partial^2 \xi_{t_1}^{t_2}(a)}{\partial a^2} < 0$). The inflection point happens when $t_2(1+\frac{1}{\lambda}) = (t_2 - t_1)(1+\lambda^{t_2})$, i.e. $\frac{\partial^2 \xi_{t_1}^{t_2}(a)}{\partial a^2} = 0$.

Finally, we show the case $t_1 < 1$. We only need to consider the case $t_2 \leq t_1 < 1$. Note that for the binary case,

$$\exp_{t_2}(a/2 - G_{t_2}(a)) + \exp_{t_2}(-a/2 - G_{t_2}(a)) = 1\,. \tag{B.5}$$

Using the definition of $\exp_{t_2}$, we can write (B.5) as

$$[1 + (1-t_2)\,(a/2 - G_{t_2}(a))]_+^{1/(1-t_2)} + [1 + (1-t_2)\,(-a/2 - G_{t_2}(a))]_+^{1/(1-t_2)} = 1\,. \tag{B.6}$$

For $a = 0$, (B.6) yields

$$[1 + (1-t_2)\,(-G_{t_2}(0))]_+^{1/(1-t_2)} = \frac{1}{2}\,.$$

From $t_2 < 1$, we have $(1-t_2) > 0$ and therefore, $G_{t_2}(0) > 0$. From convexity and symmetry $(G_{t_2}(a) = G_{t_2}(-a))$ conditions, we conclude $G_{t_2}(a) \geq G_{t_2}(0) \geq 0, \forall a$. Consequently, for values of $a \leq -\frac{1}{(1-t_2)}$, $G_{t_2}(a) = -\frac{a}{2}$ satisfies (B.5). This implies that for $a \leq -\frac{1}{(1-t_2)}$, we have $\hat{p}_{t_2}(a) = 0$ and thus, $\xi_{t_1}^{t_2}(a) = -\log_{t_1}(0) = -\frac{1}{1-t_1}$ is a constant. From (B.4), we conclude that the loss is convex for $a > -\frac{1}{(1-t_2)}$ and is a constant for $a \leq -\frac{1}{(1-t_2)}$ Thus, it is quasi-convex.