

Appendix

Proof of Theorem 1 Let us proceed with some algebraic manipulations. As the regularized loss function is assumed to be of class C^3 in a neighborhood of the solution, invoking Taylor's theorem notice that

$$\nabla_{\theta} w_{n-1}(z_j; \widehat{\theta}_{\lambda}(z^n), \boldsymbol{\lambda}) = \nabla_{\theta} w_{n-1}(z_j; \widehat{\theta}_{\lambda}(z^{n \setminus i}), \boldsymbol{\lambda}) \quad (17a)$$

$$+ \nabla_{\theta}^2 w_{n-1}(z_j; \widehat{\theta}_{\lambda}(z^{n \setminus i}), \boldsymbol{\lambda})(\widehat{\theta}_{\lambda}(z^n) - \widehat{\theta}_{\lambda}(z^{n \setminus i})) \quad (17b)$$

$$+ \frac{1}{2} \sum_{\kappa \in [k]} (\widehat{\theta}_{\lambda}(z^n) - \widehat{\theta}_{\lambda}(z^{n \setminus i}))^{\top} \left(\frac{\partial}{\partial \theta_{\kappa}} \nabla_{\theta}^2 w_{n-1}(z_j; \boldsymbol{\zeta}_{\lambda, \kappa}^{i, j, 1}(z^n), \boldsymbol{\lambda}) \right) (\widehat{\theta}_{\lambda}(z^n) - \widehat{\theta}_{\lambda}(z^{n \setminus i})) \widehat{e}_{\kappa} \quad (17c)$$

where $\boldsymbol{\zeta}_{\lambda, \kappa}^{i, j, 1}(z^n) = \alpha_{\kappa}^{i, j, 1} \widehat{\theta}_{\lambda}(z^n) + (1 - \alpha_{\kappa}^{i, j, 1}) \widehat{\theta}_{\lambda}(z^{n \setminus i})$ for some $0 \leq \alpha_{\kappa}^{i, j, 1} \leq 1$. Hence, we can sum both sides up over $j \in [n] \setminus i$ to get

$$\sum_{j \in [n] \setminus i} \nabla_{\theta} w_{n-1}(z_j; \widehat{\theta}_{\lambda}(z^n), \boldsymbol{\lambda}) = \sum_{j \in [n] \setminus i} \nabla_{\theta} w_{n-1}(z_j; \widehat{\theta}_{\lambda}(z^{n \setminus i}), \boldsymbol{\lambda}) \quad (18a)$$

$$+ \sum_{j \in [n] \setminus i} \nabla_{\theta}^2 w_{n-1}(z_j; \widehat{\theta}_{\lambda}(z^{n \setminus i}), \boldsymbol{\lambda})(\widehat{\theta}_{\lambda}(z^n) - \widehat{\theta}_{\lambda}(z^{n \setminus i})) \quad (18b)$$

$$+ \boldsymbol{\varepsilon}_{\boldsymbol{\lambda}, n}^{(i), 1}, \quad (18c)$$

where $\boldsymbol{\varepsilon}_{\boldsymbol{\lambda}, n}^{(i), 1}$ is defined in (7). Notice that by definition of $\widehat{\theta}_{\lambda}(z^n)$ and $\widehat{\theta}_{\lambda}(z^{n \setminus i})$, the left hand side term in (18a) is equal to $-\nabla_{\theta} \ell(z_i; \widehat{\theta}_{\lambda}(z^n))$ and the right hand side term is zero. Then,

$$-\nabla_{\theta} \ell(z_i; \widehat{\theta}_{\lambda}(z^n)) = \sum_{j \in [n] \setminus i} \nabla_{\theta}^2 w_{n-1}(z_j; \widehat{\theta}_{\lambda}(z^{n \setminus i}), \boldsymbol{\lambda})(\widehat{\theta}_{\lambda}(z^n) - \widehat{\theta}_{\lambda}(z^{n \setminus i})) + \boldsymbol{\varepsilon}_{\boldsymbol{\lambda}, n}^{(i), 1}. \quad (19)$$

Applying Taylor's theorem on $\nabla_{\theta}^2 w_{n-1}(z_j; \widehat{\theta}_{\lambda}(z^{n \setminus i}), \boldsymbol{\lambda})$ we get:

$$\begin{aligned} \nabla_{\theta}^2 w_{n-1}(z_j; \widehat{\theta}_{\lambda}(z^{n \setminus i}), \boldsymbol{\lambda}) &= \nabla_{\theta}^2 w_{n-1}(z_j; \widehat{\theta}_{\lambda}(z^n), \boldsymbol{\lambda}) \\ &+ \sum_{\kappa, \nu \in [k]} (\widehat{\theta}_{\lambda}(z^{n \setminus i}) - \widehat{\theta}_{\lambda}(z^n))^{\top} \left(\frac{\partial^2}{\partial \theta_{\kappa} \partial \theta_{\nu}} \nabla_{\theta} w_{n-1}(z_j; \boldsymbol{\zeta}_{\lambda, \kappa, \nu}^{i, j, 2}(z^n), \boldsymbol{\lambda}) \right) \widehat{e}_{\kappa} \widehat{e}_{\nu}^{\top}. \end{aligned} \quad (20)$$

By substituting (20) in (19), using some algebraic manipulations, and noting the definition of $\boldsymbol{\varepsilon}_{\boldsymbol{\lambda}, n}^{i, j, 2}$ in (8), we can get

$$-\nabla_{\theta} \ell(z_i; \widehat{\theta}_{\lambda}(z^n)) = \sum_{j \in [n] \setminus i} \nabla_{\theta}^2 w_{n-1}(z_j; \widehat{\theta}_{\lambda}(z^n), \boldsymbol{\lambda})(\widehat{\theta}_{\lambda}(z^n) - \widehat{\theta}_{\lambda}(z^{n \setminus i})) + \boldsymbol{\varepsilon}_{\boldsymbol{\lambda}, n}^{(i)}, \quad (21)$$

where $\boldsymbol{\varepsilon}_{\boldsymbol{\lambda}, n}^{(i)}$ is defined in (6). Consequently,

$$\widehat{\theta}_{\lambda}(z^{n \setminus i}) - \widehat{\theta}_{\lambda}(z^n) = \left(\sum_{j \in [n] \setminus i} \nabla_{\theta}^2 w_{n-1}(z_j; \widehat{\theta}_{\lambda}(z^n), \boldsymbol{\lambda}) \right)^{-1} \nabla_{\theta} \ell(z_i; \widehat{\theta}_{\lambda}(z^n)) + \boldsymbol{\xi}_{\boldsymbol{\lambda}, n}^{(i)} \quad (22)$$

$$= \frac{1}{n-1} \left(\widehat{\mathcal{H}}_{z^{n \setminus i}}(\widehat{\theta}_{\lambda}(z^n), \boldsymbol{\lambda}) \right)^{-1} \nabla_{\theta} \ell(z_i; \widehat{\theta}_{\lambda}(z^n)) + \boldsymbol{\xi}_{\boldsymbol{\lambda}, n}^{(i)}, \quad (23)$$

where

$$\boldsymbol{\xi}_{\boldsymbol{\lambda}, n}^{(i)} \triangleq \frac{1}{n-1} \left(\widehat{\mathcal{H}}_{z^{n \setminus i}}(\widehat{\theta}_{\lambda}(z^n), \boldsymbol{\lambda}) \right)^{-1} \boldsymbol{\varepsilon}_{\boldsymbol{\lambda}, n}^{(i)}. \quad (24)$$

Note that Assumption 1 implies that as n grows, the above inverse with high probability exists and converges to $\mathcal{H}(\boldsymbol{\theta}^*)^{-1}$ in probability. Further, the inverse is bounded in probability.

Finally, it is deduced from Assumption 1 that $\|\widehat{\boldsymbol{\theta}}_{\lambda}(z^n) - \widehat{\boldsymbol{\theta}}_{\lambda}(z^{n \setminus i})\|_{\infty} = o_p(1)$. On the other hand, by noticing the definition of $\boldsymbol{\varepsilon}_{\lambda,n}^{(i)}$, we can see that

$$\|\boldsymbol{\varepsilon}_{\lambda,n}^{(i)}\|_{\infty} = O_p\left(n\|\widehat{\boldsymbol{\theta}}_{\lambda}(z^n) - \widehat{\boldsymbol{\theta}}_{\lambda}(z^{n \setminus i})\|_{\infty}^2\right) = o_p\left(n\|\widehat{\boldsymbol{\theta}}_{\lambda}(z^n) - \widehat{\boldsymbol{\theta}}_{\lambda}(z^{n \setminus i})\|_{\infty}\right). \quad (25)$$

Hence, considering (24),

$$\|\boldsymbol{\xi}_{\lambda,n}^{(i)}\|_{\infty} = o_p\left(\|\widehat{\boldsymbol{\theta}}_{\lambda}(z^n) - \widehat{\boldsymbol{\theta}}_{\lambda}(z^{n \setminus i})\|_{\infty}\right). \quad (26)$$

Now, considering (23), we deduce that

$$\|\widehat{\boldsymbol{\theta}}_{\lambda}(z^n) - \widehat{\boldsymbol{\theta}}_{\lambda}(z^{n \setminus i})\|_{\infty} = O_p\left(\frac{1}{n}\right). \quad (27)$$

Hence, $\|\boldsymbol{\varepsilon}_{\lambda,n}^{(i)}\|_{\infty} = O_p(1/n)$ in (6). Thus, the error term is

$$\widehat{\boldsymbol{\theta}}_{\lambda}(z^{n \setminus i}) - \widetilde{\boldsymbol{\theta}}_{\lambda}^{(i)}(z^n) = \boldsymbol{\xi}_{\lambda,n}^{(i)}, \quad (28)$$

and

$$\|\widehat{\boldsymbol{\theta}}_{\lambda}(z^{n \setminus i}) - \widetilde{\boldsymbol{\theta}}_{\lambda}^{(i)}(z^n)\|_{\infty} = O_p\left(\frac{1}{n^2}\right), \quad (29)$$

completing the proof. \blacksquare

Proof of Lemma 3 Notice that

$$\begin{aligned} \ell(z_i; \widetilde{\boldsymbol{\theta}}_{\lambda}^{(i)}(z^n)) &= \ell(z_i; \widehat{\boldsymbol{\theta}}_{\lambda}(z^n)) \\ &\quad + \nabla_{\boldsymbol{\theta}}^{\top} \ell(z_i; \widehat{\boldsymbol{\theta}}(z^n)) (\widetilde{\boldsymbol{\theta}}_{\lambda}^{(i)}(z^n) - \widehat{\boldsymbol{\theta}}_{\lambda}(z^n)) \\ &\quad + O(\|\widetilde{\boldsymbol{\theta}}_{\lambda}^{(i)}(z^n) - \widehat{\boldsymbol{\theta}}_{\lambda}(z^n)\|_{\infty}^2) \end{aligned} \quad (30)$$

$$\begin{aligned} &= \ell(z_i; \widehat{\boldsymbol{\theta}}_{\lambda}(z^n)) \\ &\quad + \frac{1}{n-1} \nabla_{\boldsymbol{\theta}}^{\top} \ell(z_i; \widehat{\boldsymbol{\theta}}_{\lambda}(z^n)) \left[\widehat{\mathcal{H}}_{z^{n \setminus i}}(\widehat{\boldsymbol{\theta}}_{\lambda}(z^n), \boldsymbol{\lambda}) \right]^{-1} \nabla_{\boldsymbol{\theta}} \ell(z_i; \widehat{\boldsymbol{\theta}}_{\lambda}(z^n)) \\ &\quad + O_p\left(\frac{1}{n^2}\right), \end{aligned} \quad (31)$$

where (30) follows from Assumption 1, and (31) follows from the definition of $\widetilde{\boldsymbol{\theta}}_{\lambda}^{(i)}(z^n)$, and the fact that

$$\|\widetilde{\boldsymbol{\theta}}_{\lambda}^{(i)}(z^n) - \widehat{\boldsymbol{\theta}}_{\lambda}(z^n)\|_{\infty} \leq \|\widehat{\boldsymbol{\theta}}_{\lambda}(z^{n \setminus i}) - \widetilde{\boldsymbol{\theta}}_{\lambda}^{(i)}(z^n)\|_{\infty} + \|\widehat{\boldsymbol{\theta}}_{\lambda}(z^n) - \widehat{\boldsymbol{\theta}}_{\lambda}(z^{n \setminus i})\|_{\infty} \quad (32)$$

$$= O_p\left(\frac{1}{n^2}\right) + O_p\left(\frac{1}{n}\right) = O_p\left(\frac{1}{n}\right), \quad (33)$$

which implies $\|\widehat{\boldsymbol{\theta}}_{\lambda}(z^n) - \widetilde{\boldsymbol{\theta}}_{\lambda}^{(i)}(z^n)\|_{\infty}^2 = O_p(1/n^2)$. The proof is completed by noticing the definition of $\widehat{\text{ACV}}_{\lambda}(z^n)$ in (11). \blacksquare

Proof of Lemma 4 By definition,

$$\nabla_{\boldsymbol{\theta}} \widehat{\mathcal{L}}_{z^n}(\widehat{\boldsymbol{\theta}}_{\lambda}(z^n)) + \frac{1}{n} \nabla_{\boldsymbol{\theta}} \mathbf{r}(\widehat{\boldsymbol{\theta}}_{\lambda}(z^n)) \boldsymbol{\lambda} = \nabla_{\boldsymbol{\theta}} \widehat{\mathcal{W}}_{z^n}(\widehat{\boldsymbol{\theta}}_{\lambda}(z^n)) = 0. \quad (34)$$

Using the implicit function theorem, we can further differentiate the left-hand-side with respect to $\boldsymbol{\lambda}$ to get:

$$\nabla_{\boldsymbol{\theta}}^2 \widehat{\mathcal{L}}_{z^n}(\widehat{\boldsymbol{\theta}}_{\lambda}(z^n)) \nabla_{\boldsymbol{\lambda}} \widehat{\boldsymbol{\theta}}_{\lambda}(z^n) + \frac{1}{n} \nabla_{\boldsymbol{\theta}} \mathbf{r}(\widehat{\boldsymbol{\theta}}_{\lambda}(z^n)) + \frac{1}{n} \sum_{m \in [M]} \lambda_m \nabla_{\boldsymbol{\theta}}^2 r_m(\widehat{\boldsymbol{\theta}}_{\lambda}(z^n)) \nabla_{\boldsymbol{\lambda}} \widehat{\boldsymbol{\theta}}_{\lambda}(z^n) = 0. \quad (35)$$

Thus,

$$\widehat{\mathcal{H}}_{z^n}(\widehat{\boldsymbol{\theta}}_{\lambda}(z^n), \boldsymbol{\lambda}) \nabla_{\boldsymbol{\lambda}} \widehat{\boldsymbol{\theta}}_{\lambda}(z^n) = -\frac{1}{n} \nabla_{\boldsymbol{\theta}} \mathbf{r}(\widehat{\boldsymbol{\theta}}_{\lambda}(z^n)), \quad (36)$$

which completes the proof. \blacksquare

Proof of Corollary 5 This directly follows from Lemma 4. \blacksquare