Submitted to Annals of Statistics
arXiv: arXiv:0000.0000

SUPPLEMENT TO "GLOBAL SOLUTIONS TO FOLDED CONCAVE PENALIZED NONCONVEX LEARNING"

By Hongcheng Liu*, Tao Yao* and Runze Li†

The Pennsylvania State University

S1. Proof of Proposition 2.1. Notice that (2.1) admits a closed-form representation given as:

$$P_\lambda(\theta) = \begin{cases} \lambda\theta & \text{if } 0 \le \theta \le \lambda \\ \lambda^2 + \frac{1}{a-1}[\frac{\lambda^2}{2}(1-2a) + a\lambda\theta - \frac{1}{2}\theta^2] & \text{if } \lambda < \theta \le a\lambda \\ \lambda^2 + \frac{1}{a-1}[\frac{\lambda^2}{2}(1-2a) + \frac{1}{2}a^2\lambda^2] & \text{if } \theta > a\lambda \end{cases}$$

$$(\text{S1.1}) \qquad = \begin{cases} \lambda\theta & \text{if } 0 \le \theta \le \lambda \\ \frac{1}{a-1}[-\frac{\lambda^2}{2} + a\lambda\theta - \frac{1}{2}\theta^2] & \text{if } \lambda < \theta \le a\lambda \\ \frac{1}{2}(a+1)\lambda^2 & \text{if } \theta > a\lambda \end{cases}$$

Let $F(\cdot) : \mathbb{R}_+ \to \mathbb{R}$ be defined as

$$F(\theta) := \min_{\kappa\in[0,\,\lambda]} \left\{ (\theta - a\lambda)\kappa + \frac{1}{2}(a-1)\kappa^2 \right\} + \frac{a+1}{2}\lambda^2.$$

We next show that $F(\theta) = P_\lambda(\theta)$ for any $\theta \ge 0$. Observe that

$$(\text{S1.2}) \qquad F(\theta) = \min_{\kappa\in[0,\,\lambda]} \left\{ \frac{1}{2}(a-1)\left(\frac{(\theta - a\lambda)\kappa}{\frac{1}{2}(a-1)} + \kappa^2 \right) \right\} + \frac{a+1}{2}\lambda^2$$

$$= \min_{\kappa\in[0,\,\lambda]} \left\{ \frac{1}{2}(a-1)\left(\frac{\theta - a\lambda}{a-1} + \kappa \right)^2 - \frac{\theta^2 - 2a\lambda\theta + \lambda^2}{2(a-1)} \right\}$$

Further notice that the unique minimizer κ^* of $\min_{\kappa\in[0,\,\lambda]} \left\{ (\theta - a\lambda)\kappa + \frac{1}{2}(a-1)\kappa^2 \right\}$ is

$$\kappa^* = \begin{cases} -\frac{\theta - a\lambda}{a-1} & \text{if } 0 \le -\frac{\theta - a\lambda}{a-1} \le \lambda \\ 0 & \text{if } -\frac{\theta - a\lambda}{a-1} \le 0 \\ \lambda & \text{if } -\frac{\theta - a\lambda}{a-1} \ge \lambda \end{cases}$$

*Support by Penn State GraceWoodward Collaborative Engineering/Medicine Research Grant.

†Support by National Science Foundation DMS 1512422 and National Institute of Health grants P50 DA036107 and P50 DA039838.

Equivalently,

$$\kappa^* = \begin{cases} -\frac{\theta - a\lambda}{a-1} & \text{if } \lambda \leq \theta \leq a\lambda \\ 0 & \text{if } \theta \geq a\lambda \\ \lambda & \text{if } \theta \leq \lambda \end{cases} \tag{S1.3}$$

Eq. (S1.3) provides a closed-form solution to the single variable minimization problem involved in (S1.2). Therefore, we can equivalently rewrite $F(\theta)$ into

$$F(\theta) = (\theta - a\lambda)\kappa^* + \frac{1}{2}(a-1)(\kappa^*)^2 + \frac{a+1}{2}\lambda^2. \tag{S1.4}$$

Combining (S1.1), (S1.3), and (S1.4) immediately gives us the equality, $F(\theta) = P_\lambda(\theta)$. Therefore, Model (1.1) is equivalent to

$$\min_{\beta \in \Lambda} \frac{1}{2}\beta^\top Q\beta + q^\top \beta + n\sum_{i=1}^{d} \min_{g_i \in [0,\, \lambda]} \left\{ (|\beta_i| - a\lambda) \cdot g_i + \frac{1}{2}(a-1) \cdot g_i^2 \right\} + n\frac{a+1}{2}\lambda^2 d.$$

which immediately provides the desired result in Part (a).

As to Part (b), comparing (S1.3) and (2.1), we immediately have $\kappa^* = P'_\lambda(\theta)$, which is the desired result.

S2. Proof for Proposition 2.3. This proof follows a closely similar argument to the proof for Proposition 2.1 in Appendix S1. Notice that, per (2.2),

$$P_\lambda(\theta) := \begin{cases} \lambda\theta - \frac{\theta^2}{2a} & \text{if } 0 \leq \theta \leq a\lambda \\ \frac{1}{2}a\lambda^2 & \text{if } \theta > a\lambda \end{cases} \tag{S2.1}$$

Define

$$F(\theta) := \min_{\kappa \in [0,\, a\lambda]} \frac{1}{2a}\kappa^2 - \frac{1}{a}\theta\kappa + \theta\lambda$$

and we want to show that $F(\theta) = P_\lambda(\theta)$. Observe that, for any $\theta \geq 0$,

$$F(\theta) = \min_{\kappa \in [0,\, a\lambda]} \frac{1}{2a}\kappa^2 - \frac{1}{a}\theta\kappa + \theta\lambda = \min_{\kappa \in [0,\, a\lambda]} \frac{1}{2a}(\kappa - \theta)^2 - \frac{\theta^2}{2a} + \theta\lambda \tag{S2.2}$$

whose closed form solution κ^* is given as

$$\kappa^* := \begin{cases} \theta & \text{if } 0 \leq \theta \leq a\lambda \\ a\lambda & \text{if } \theta > a\lambda \end{cases}$$

Replacing κ by the minimizer κ^* in (S2.2), we can drop the minimization operator "min". Then this easily leads to the equality, $F(\theta) = P_\lambda(\theta)$. This immediately results in both (a) and (b).

S3. Proof for Lemma 4.1.

PROOF. Note that $\frac{64k\tau\log d}{n}+\mu\le\alpha$. In view of (4.2), we have for all $\|\beta'-\beta''\|_2\le 3$:

$$\begin{aligned}&L(\beta')-L(\beta'')-\langle\nabla_\beta L(\beta''),\beta'-\beta''\rangle\\ \ge&\alpha_1\|\beta'-\beta''\|_2^2-\tau_1\frac{\log d}{n}|\beta'-\beta''|^2\\ \ge&\left(\frac{64k\tau\log d}{n}+\mu\right)\|\beta'-\beta''\|_2^2-\tau\frac{\log d}{n}|\beta'-\beta''|^2.\end{aligned}$$

Observe that for all $(\beta',\beta'')\in\{(\beta',\beta''):\|\beta'-\beta''\|_0\le 64k-1\}$, we have

$$\tau\frac{\log d}{n}|\beta'-\beta''|^2\le(64k-1)\tau\frac{\log d}{n}\|\beta'-\beta''\|_2^2.$$

Thus for all $(\beta',\beta'')\in\{(\beta',\beta''):\|\beta'-\beta''\|_2\le 3,\ \|\beta'-\beta''\|_0\le 64k-1\}$

$$\begin{aligned}L(\beta')-L(\beta'')\ge&\langle\nabla L(\beta''),\beta'-\beta''\rangle+\left(\frac{64k\tau\log d}{n}+\mu\right)\|\beta'-\beta''\|_2^2\\ &-\frac{\tau(64k-1)\log d}{n}\|\beta'-\beta''\|_2^2\\ \ge&\langle\nabla L(\beta''),\beta'-\beta''\rangle+\left(\frac{\tau\log d}{n}+\mu\right)\|\beta'-\beta''\|_2^2\end{aligned}\tag{S3.1}$$

Notice that per Taylor expansion we have that for some $\chi\in[\beta',\ \beta'']$

$$L(\beta')-L(\beta'')=\langle\nabla L(\beta''),\ \beta'-\beta''\rangle+\frac{1}{2}\langle\beta'-\beta'',\ [\nabla^2L(\chi)]^\top(\beta'-\beta'')\rangle\tag{S3.2}$$

Thus, combining (S3.1) and (S3.2), we obtain that for all $(\beta',\beta'')\in\{(\beta',\beta''):\|\beta'-\beta''\|_2\le 3,\ \|\beta'-\beta''\|_0\le 64k-1\}$,

$$\frac{1}{2}\langle\beta'-\beta'',\ [\nabla^2L(\chi)]^\top(\beta'-\beta'')\rangle\ge\left(\frac{\tau\log d}{n}+\mu\right)\|\beta'-\beta''\|_2^2.$$

Further notice $\nabla^2L=\frac{X^\top X}{n}$. Then it follows that for all $c\in\mathbb{R}_+,\forall(\beta',\beta'')\in\{(\beta',\beta''):\|\beta'-\beta''\|_2\le 3,\ \|\beta'-\beta''\|_0\le 64k-1\}$,

$$\frac{1}{2n}\langle c(\beta'-\beta''),\ X^\top X[c(\beta'-\beta'')]\rangle\ge\left(\frac{\tau\log d}{n}+\mu\right)\|c(\beta'-\beta'')\|_2^2.$$

This implies that for all $(\beta', \beta'') \in \{(\beta', \beta'') : \|\beta' - \beta''\|_0 \leq 64k - 1\}$, and $\chi \in [\beta',\, \beta'']$,

$$\text{(S3.3)}\quad \frac{1}{2}\langle \beta' - \beta'',\ [\nabla^2 L(\chi)]^\top (\beta' - \beta'')\rangle = \frac{1}{2n}\langle \beta' - \beta'',\ X^\top X(\beta' - \beta'')\rangle \geq \left(\frac{\tau \log d}{n} + \mu\right)\|\beta' - \beta''\|_2^2,$$

which, combining with (S3.2), gives that

$$L(\beta') - L(\beta'') \geq \langle \nabla L(\beta''),\ \beta' - \beta''\rangle + \left(\frac{\tau \log d}{n} + \mu\right)\|\beta' - \beta''\|_2^2$$

for all $(\beta', \beta'') \in \{(\beta', \beta'') : \|\beta' - \beta''\|_0 \leq 64k - 1\}$.

As an immediate result,

$$\begin{aligned}\text{(S3.4)}\quad & L(\beta') - \mu\|\beta'\|_2^2 - L(\beta'') + \mu\|\beta''\|_2^2 \\ & \geq \langle \nabla L(\beta''),\ \beta' - \beta''\rangle + \left(\frac{\tau \log d}{n} + \mu\right)\|\beta' - \beta''\|_2^2 - \mu\|\beta'\|_2^2 + \mu\|\beta''\|_2^2 \\ & = \langle \nabla L(\beta'') - 2\mu\beta'',\ \beta' - \beta''\rangle + \left(\frac{\tau \log d}{n} + \mu\right)\|\beta' - \beta''\|_2^2 - \mu\|\beta'\|_2^2 \\ & \quad + \mu\|\beta''\|_2^2 + 2\mu\langle \beta'',\, \beta' - \beta''\rangle.\end{aligned}$$

Notice that

$$\text{(S3.5)}\qquad \mu\|\beta' - \beta''\|_2^2 - \mu\|\beta'\|_2^2 + \mu\|\beta''\|_2^2 + 2\mu\langle \beta'',\, \beta' - \beta''\rangle = 0.$$

Combining (S3.4) and (S3.5), we obtain

$$\begin{aligned}\text{(S3.6)}\qquad & L(\beta') - \mu\|\beta'\|_2^2 - L(\beta'') + \mu\|\beta''\|_2^2 \\ & \geq \langle \nabla L(\beta'') - 2\mu\beta'',\ \beta' - \beta''\rangle + \frac{\tau \log d}{n}\|\beta' - \beta''\|_2^2.\end{aligned}$$

for all $(\beta', \beta'') \in \{(\beta', \beta'') : \|\beta' - \beta''\|_0 \leq 64k - 1\}$. It is easy to check that $\sum_{i=1}^d P_\lambda(|\beta_i|) - \lambda|\beta|$ is continuously differentiable in β. Denote that $\mathcal{G}(\beta) := \frac{\partial[\sum_{i=1}^d P_\lambda(|\beta_i|) - \lambda|\beta|]}{\partial\beta}$. Invoking the assumption that $\mu\|\beta\|_2^2 + \sum_{i=1}^d P_\lambda(|\beta_i|)$ is convex, we know that, for all $(\beta', \beta'') \in \{(\beta', \beta'') : \|\beta' - \beta''\|_0 \leq 64k - 1\}$:

$$\begin{aligned}\text{(S3.7)}\qquad & \mu\|\beta'\|_2^2 + \sum_{i=1}^d P_\lambda(|\beta_i'|) - \mu\|\beta''\|_2^2 - \sum_{i=1}^d P_\lambda(|\beta_i''|) \\ & \geq \left\langle 2\mu\beta'' + \mathcal{G}(\beta'') + \mathcal{D}(\beta''),\ \beta' - \beta''\right\rangle\end{aligned}$$

where $\mathcal{D}(\beta'') \in \lambda\left[\partial|\beta|\right]_{\beta=\beta''}$ is a subgradient of $\lambda|\cdot|$ at β'' and $\partial|\cdot|$ is the subddiferential of $\lambda|\cdot|$. Combining (S3.6) and (S3.7), we have

$$\begin{aligned}
&\frac{\mathcal{L}(\beta')}{n} - \frac{\mathcal{L}(\beta'')}{n} \\
&= L(\beta') + \sum_{i=1}^{d} P_\lambda(|\beta_i'|) - L(\beta'') - \sum_{i=1}^{d} P_\lambda(|\beta_i''|) \\
&\geq \left\langle \nabla L(\beta'') + \mathcal{G}(\beta'') + \mathcal{D}(\beta''),\ \beta' - \beta'' \right\rangle + \frac{\tau \log d}{n}\|\beta' - \beta''\|_2^2.
\end{aligned}$$

Observe that, by definition, $\nabla L(\beta'') + \mathcal{G}(\beta'') = \nabla \tilde{L}(\beta'')$, where $\tilde{L}$ is defined as in (4.3). Therefore, we have the desired result. □

S4. Extended Numerical Results for Comparison in Optimization Quality. This section of the appendix presents extended numerical results that complete Table 4. More specifically, Table S1 extends the LR-SCAD case (in the upper panel) of Table 4 and Table S2 extends the LR-MCP case (in the lower panel) of Table 4.

Department of Industrial
and Manufacturing Engineering
Pennsylvania State University
State College, PA 16802 USA
E-mail: hql5143@psu.edu
taoyao@psu.edu

Department of Statistics
Pennsylvania State University
State College, PA 16802 USA
E-mail: rzli@psu.edu

Table S1

Extended results for the upper panel of Table 4 on comparison between LLA and the proposed MIPGO to solve LR-SCAD

		MIPGO	LLA_r	$\mathrm{gap}_{(\%)}$	LLA_0	$\mathrm{gap}_{(\%)}$	LLA_1	$\mathrm{gap}_{(\%)}$
TS 1	Min.	67.01	67.01	**0.00**	83.24	19.50	88.84	24.57
$d = 10$	Ave.	67.01	84.44	20.64	83.24	19.50	88.84	24.57
$n = 10$	Max.	67.01	120.30	44.30	83.24	19.50	88.84	24.57
TS 2	Min.	76.15	76.15	**0.00**	76.15	**0.00**	106.02	28.16
$d = 10$	Ave.	76.15	93.58	18.63	76.15	**0.00**	106.02	28.16
$n = 10$	Max.	76.15	113.86	33.12	76.15	**0.00**	106.02	28.16
TS 4	Min.	90.58	92.88	2.48	106.05	14.59	90.58	**0.00**
$d = 20$	Ave.	90.58	110.35	17.91	106.05	14.59	90.58	**0.00**
$n = 10$	Max.	90.58	149.19	39.29	106.05	14.59	90.58	**0.00**
TS 5	Min.	64.98	64.98	**0.00**	64.98	**0.00**	107.19	39.38
$d = 20$	Ave.	64.98	86.10	24.53	64.98	**0.00**	107.19	39.38
$n = 10$	Max.	64.98	167.61	61.23	64.98	**0.00**	107.19	39.38
TS 7	Min.	132.04	132.04	**0.00**	162.05	18.52	168.77	21.76
$d = 40$	Ave.	132.04	181.12	27.10	162.05	18.52	190.76	21.76
$n = 15$	Max.	132.04	284.86	53.65	162.05	18.52	190.76	21.76
TS 8	Min.	146.79	146.79	**0.00**	165.00	11.03	146.79	**0.00**
$d = 40$	Ave.	146.79	196.85	25.43	165.00	11.03	146.79	**0.00**
$n = 15$	Max.	146.79	291.78	49.69	165.00	11.03	146.79	**0.00**
TS 10	Min.	702.54	702.54	**0.00**	830.80	15.44	830.80	15.44
$d = 200$	Ave.	702.54	788.15	10.86	830.80	15.44	830.80	15.44
$n = 60$	Max.	702.54	866.43	18.92	830.80	15.44	830.80	15.44
TS 11	Min.	782.59	905.47	13.57	914.25	14.40	782.59	**0.00**
$d = 200$	Ave.	782.59	1026.62	23.77	914.25	14.40	782.59	**0.00**
$n = 60$	Max.	782.59	1332.02	41.25	914.25	14.40	782.59	**0.00**
TS 13	Min.	962.10	1032.77	6.84	1225.51	21.49	1135.80	15.29
$d = 500$	Ave.	962.10	1187.43	18.98	1225.51	21.49	1135.80	15.29
$n = 80$	Max.	962.10	1252.43	23.18	1225.51	21.49	1135.80	15.29
TS 14	Min.	703.79	703.79	**0.00**	921.93	23.66	892.53	21.15
$d = 500$	Ave.	703.79	914.86	23.07	921.93	23.66	892.53	21.15
$n = 80$	Max.	703.79	1321.40	46.74	921.93	23.66	892.53	21.15
TS 16	Min.	1037.19	1037.19	**0.00**	1246.49	16.79	1246.49	16.79
$d = 1000$	Ave.	1037.19	1200.53	13.61	1246.49	16.79	1246.49	16.79
$n = 100$	Max.	1037.19	1400.49	25.94	1246.49	16.79	1246.49	16.79
TS 17	Min.	1036.06	1036.06	**0.00**	1413.73	27.55	1413.73	26.71
$d = 1000$	Ave.	1036.06	1276.52	18.84	1413.73	27.55	1413.73	26.71
$n = 100$	Max.	1036.06	1429.95	27.55	1413.73	27.55	1413.73	26.71

Table S2
Extended results for the lower panel of Table 4 on comparison between LLA and the proposed MIPGO to solve LR-MCP

		MIPGO	LLA_r	$gap_{(\%)}$	LLA_0	$gap_{(\%)}$	LLA_1	$gap_{(\%)}$
TS 1	Min.	10.23	20.48	32.03	13.46	24.04	15.04	32.03
$d = 10$	Ave.	10.23	23.03	37.99	13.46	24.04	15.04	32.03
$n = 10$	Max.	10.23	30.63	55.02	13.46	24.04	15.04	32.03
TS 2	Min.	19.21	20.48	6.19	23.52	18.31	22.50	14.61
$d = 10$	Ave.	19.21	23.03	16.58	23.52	18.31	22.50	14.61
$n = 10$	Max.	19.21	30.63	37.27	23.52	18.31	22.50	14.61
TS 4	Min.	14.18	17.52	19.10	17.88	20.71	22.51	37.02
$d = 20$	Ave.	14.18	26.05	45.57	17.88	20.71	22.51	37.02
$n = 10$	Max.	14.18	35.00	59.50	17.88	20.71	22.51	37.02
TS 5	Min.	13.54	17.61	23.09	15.79	14.21	25.00	45.83
$d = 20$	Ave.	13.54	26.44	48.78	15.79	14.21	25.00	45.83
$n = 10$	Max.	13.54	42.50	68.14	15.79	14.21	25.00	45.83
TS 7	Min.	32.05	39.03	17.88	32.05	**0.00**	48.76	34.28
$d = 40$	Ave.	32.05	48.73	34.24	32.05	**0.00**	48.76	34.28
$n = 15$	Max.	32.05	71.25	55.02	32.05	**0.00**	48.76	34.28
TS 8	Min.	29.28	32.97	11.18	31.87	8.11	41.49	29.41
$d = 40$	Ave.	29.28	43.23	32.27	31.87	8.11	41.49	29.41
$n = 15$	Max.	29.28	56.25	47.94	31.87	8.11	41.49	29.41
TS 10	Min.	119.06	130.46	8.73	122.72	2.98	130.46	8.73
$d = 200$	Ave.	119.06	156.16	24.76	122.72	2.98	130.46	8.73
$n = 60$	Max.	119.06	550.37	78.37	122.72	2.98	130.46	8.73
TS 11	Min.	118.83	124.04	4.20	157.43	24.52	124.04	4.20
$d = 200$	Ave.	118.83	136.78	13.13	157.43	24.52	124.04	4.20
$n = 60$	Max.	118.83	201.01	40.89	157.43	24.52	124.04	4.20
TS 13	Min.	155.77	171.67	9.26	155.77	**0.00**	176.95	11.97
$d = 500$	Ave.	155.77	184.67	15.65	155.77	**0.00**	176.95	11.97
$n = 80$	Max.	155.77	355.23	56.15	155.77	**0.00**	176.95	11.97
TS 14	Min.	167.06	171.33	2.50	194.65	14.18	186.32	10.34
$d = 500$	Ave.	167.06	198.97	16.04	194.65	14.18	186.32	10.34
$n = 80$	Max.	167.06	250.78	33.39	194.65	14.18	186.32	10.34
TS 16	Min.	201.65	212.05	4.91	212.05	4.91	265.49	24.05
$d = 1000$	Ave.	201.65	223.29	9.69	212.05	4.91	265.49	24.05
$n = 100$	Max.	201.65	265.49	24.05	212.05	4.91	265.49	24.05
TS 17	Min.	139.16	139.16	**0.00**	139.16	**0.00**	162.25	14.23
$d = 1000$	Ave.	139.16	156.37	11.00	139.16	**0.00**	162.25	14.23
$n = 100$	Max.	139.16	283.85	50.98	139.16	**0.00**	162.25	14.23