Supplementary Materials for "Geometric Analysis of Non-Convex Optimization Landscape for Robust M-Estimation of Location"

Hongyuan Yang, Ziping Zhao, and Ying Sun

Appendix A
Proof of Proposition 3

Considering the first-order optimality condition of problem (2), we have

$$\begin{aligned}\nabla \mathcal{L}(\hat{\boldsymbol{\mu}}) &= \frac{1}{n}\sum_{i=1}^{n} \ell'\left(d_i^2(\hat{\boldsymbol{\mu}})\right)\nabla d_i^2(\boldsymbol{\mu}) \\ &= \frac{2}{n}\sum_{i=1}^{n} \ell'\left(d_i^2(\hat{\boldsymbol{\mu}})\right)(\hat{\boldsymbol{\mu}}-\boldsymbol{x}_i) = \mathbf{0},\end{aligned} \tag{6}$$

where $\hat{\boldsymbol{\mu}}$ denotes any stationary point. Then we have

$$\hat{\boldsymbol{\mu}} = \sum_{i=1}^{n} \frac{\ell'\left(d_i^2(\hat{\boldsymbol{\mu}})\right)}{\sum_{i=1}^{n}\ell'\left(d_i^2(\hat{\boldsymbol{\mu}})\right)}\boldsymbol{x}_i. \tag{7}$$

Note that $\sum_{i=1}^{n}\ell'\left(d_i^2(\hat{\boldsymbol{\mu}})\right) \neq 0$, since ℓ' is strictly positive by Assumption 1. The right hand side (RHS) of (7) can be interpreted as a weighted average of the sample $\{\boldsymbol{x}_i\}_{i=1}^n$. (For robust functions, ℓ' is a non-increasing function for large values, so outlying samples will receive smaller weights, which brings the robustness.)

Substitute $\boldsymbol{x}_i = \boldsymbol{\mu}^\star + \boldsymbol{\varepsilon}_i$ into (7), and we obtain

$$\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}^\star = \sum_{i=1}^{n} \frac{\ell'\left(d_i^2(\hat{\boldsymbol{\mu}})\right)}{\sum_{i=1}^{n}\ell'\left(d_i^2(\hat{\boldsymbol{\mu}})\right)}\boldsymbol{\varepsilon}_i. \tag{8}$$

Taking Euclidean norm to both sides of (8) leads to

$$\begin{aligned}\|\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}^\star\| &= \left\| \sum_{i=1}^{n} \frac{\ell'\left(d_i^2(\hat{\boldsymbol{\mu}})\right)}{\sum_{i=1}^{n}\ell'\left(d_i^2(\hat{\boldsymbol{\mu}})\right)}\boldsymbol{\varepsilon}_i \right\| \\ &\leq \frac{\sum_{i=1}^{n}\ell'\left(d_i^2(\hat{\boldsymbol{\mu}})\right)}{\sum_{i=1}^{n}\ell'\left(d_i^2(\hat{\boldsymbol{\mu}})\right)}\|\boldsymbol{\varepsilon}_i\| \leq \kappa,\end{aligned}$$

where the second inequality follows from Assumption 2.

Appendix B
Proof of Theorem 5

Our goal is to prove $\mathcal{L}$ is strongly convex and smooth in the interior of $\mathcal{B}(\hat{\boldsymbol{\mu}}, R)$. We first prove the strong convexity of $\mathcal{L}$. Since ℓ is twice differentiable according to Assumption 1, we can compute the Hessian of $\mathcal{L}$ as follows:

$$\begin{aligned}\nabla^2 \mathcal{L}(\boldsymbol{\mu}) = \frac{4}{n}\sum_{i=1}^{n} \ell''\left(d_i^2(\boldsymbol{\mu})\right)(\boldsymbol{\mu}-\boldsymbol{x}_i)(\boldsymbol{\mu}-\boldsymbol{x}_i)^\top \\ + \frac{2}{n}\sum_{i=1}^{n} \ell'\left(d_i^2(\boldsymbol{\mu})\right)\boldsymbol{I}.\end{aligned}$$

Then, it is equivalent to prove $\nabla^2\mathcal{L}(\hat{\boldsymbol{\mu}}+\boldsymbol{r}) \succ \mathbf{0}$ for all $\boldsymbol{r}$ within some ball.

We have

$$\begin{aligned}&\lambda_{\min}\left(\nabla^2\mathcal{L}(\hat{\boldsymbol{\mu}}+\boldsymbol{r})\right) \\ =&\lambda_{\min}\Bigg(\frac{4}{n}\sum_{i=1}^{n}\ell''\left(d_i^2(\hat{\boldsymbol{\mu}}+\boldsymbol{r})\right)(\hat{\boldsymbol{\mu}}+\boldsymbol{r}-\boldsymbol{x}_i)(\hat{\boldsymbol{\mu}}+\boldsymbol{r}-\boldsymbol{x}_i)^\top \\ &\qquad + \frac{2}{n}\sum_{i=1}^{n}\ell'\left(d_i^2(\hat{\boldsymbol{\mu}}+\boldsymbol{r})\right)\boldsymbol{I}\Bigg).\end{aligned}$$

Using Weyl's inequality, we get

$$\begin{aligned}&\lambda_{\min}\left(\nabla^2\mathcal{L}(\hat{\boldsymbol{\mu}}+\boldsymbol{r})\right) \\ \geq&\lambda_{\min}\left(\frac{4}{n}\sum_{i=1}^{n}\ell''\left(d_i^2(\hat{\boldsymbol{\mu}}+\boldsymbol{r})\right)(\hat{\boldsymbol{\mu}}+\boldsymbol{r}-\boldsymbol{x}_i)(\hat{\boldsymbol{\mu}}+\boldsymbol{r}-\boldsymbol{x}_i)^\top\right) \\ &\qquad + \frac{2}{n}\sum_{i=1}^{n}\ell'\left(d_i^2(\hat{\boldsymbol{\mu}}+\boldsymbol{r})\right),\end{aligned} \tag{9}$$

Consider the first term in (9). We have

$$\begin{aligned}&\lambda_{\min}\left(\frac{4}{n}\sum_{i=1}^{n}\ell''\left(d_i^2(\hat{\boldsymbol{\mu}}+\boldsymbol{r})\right)(\hat{\boldsymbol{\mu}}+\boldsymbol{r}-\boldsymbol{x}_i)(\hat{\boldsymbol{\mu}}+\boldsymbol{r}-\boldsymbol{x}_i)^\top\right) \\ \geq&\frac{4}{n}\sum_{i=1}^{n}\lambda_{\min}\left(\ell''\left(d_i^2(\hat{\boldsymbol{\mu}}+\boldsymbol{r})\right)(\hat{\boldsymbol{\mu}}+\boldsymbol{r}-\boldsymbol{x}_i)(\hat{\boldsymbol{\mu}}+\boldsymbol{r}-\boldsymbol{x}_i)^\top\right),\end{aligned}$$

where the inequality follows from the Weyl's inequality. We further have

$$\begin{aligned}&\frac{4}{n}\sum_{i=1}^{n}\lambda_{\min}\left(\ell''\left(d_i^2(\hat{\boldsymbol{\mu}}+\boldsymbol{r})\right)(\hat{\boldsymbol{\mu}}+\boldsymbol{r}-\boldsymbol{x}_i)(\hat{\boldsymbol{\mu}}+\boldsymbol{r}-\boldsymbol{x}_i)^\top\right) \\ \geq& -\frac{4}{n}\sum_{i=1}^{n}\left|\ell''\left(d_i^2(\hat{\boldsymbol{\mu}}+\boldsymbol{r})\right)\right|\|\hat{\boldsymbol{\mu}}+\boldsymbol{r}-\boldsymbol{x}_i\|^2 \\ \geq& -\frac{4}{n}\sum_{i=1}^{n}\left|\ell''\left(d_i^2(\hat{\boldsymbol{\mu}}+\boldsymbol{r})\right)\right|\left(\|\boldsymbol{r}\|+\|\hat{\boldsymbol{\mu}}-\boldsymbol{\mu}^\star\|+\|\boldsymbol{\varepsilon}_i\|\right)^2 \\ \geq& -\frac{4}{n}\sum_{i=1}^{n}\left|\ell''\left(d_i^2(\hat{\boldsymbol{\mu}}+\boldsymbol{r})\right)\right|\left(\|\boldsymbol{r}\|+2\kappa\right)^2,\end{aligned} \tag{10}$$

where the first inequality is due to $\lambda_{\min}\left(\omega\boldsymbol{a}\boldsymbol{a}^\top\right) = -|\omega|\,\|\boldsymbol{a}\|^2$ for any scalar ω, and the last inequality follows from $\|\boldsymbol{\varepsilon}_i\| \leq \kappa$ by Assumption 2 and $\|\hat{\boldsymbol{\mu}}-\boldsymbol{\mu}^\star\| \leq \kappa$ by Proposition 3. Upon substituting (10) into (9), we get

$$\begin{aligned}&\lambda_{\min}\left(\nabla^2\mathcal{L}(\hat{\boldsymbol{\mu}}+\boldsymbol{r})\right) \\ \geq& -\frac{4}{n}\sum_{i=1}^{n}\left|\ell''\left(d_i^2(\hat{\boldsymbol{\mu}}+\boldsymbol{r})\right)\right|\left(\|\boldsymbol{r}\|+2\kappa\right)^2 \\ &\qquad + \frac{2}{n}\sum_{i=1}^{n}\ell'\left(d_i^2(\hat{\boldsymbol{\mu}}+\boldsymbol{r})\right) \\ \geq& -4U(r+2\kappa)^2 + 2L,\end{aligned} \tag{11}$$

where we define $\|\boldsymbol{r}\| = r$ and the second inequality follows from Lemma 6. The RHS of (11) is positive whenever

$$r < \sqrt{\frac{L}{2U}} - 2\kappa. \tag{12}$$

Denote the supremum of the solution of r to (12) as R. (Note that there always exists a solution r satisfying (12). Since $\sqrt{\frac{L}{2U}} > 0$ always holds in (12). Thus, (12) can always be satisfied when κ and r are sufficiently small.) Then we have that in the interior of the ball $\mathcal{B}(\hat{\boldsymbol{\mu}}, R)$, $\mathcal{L}$ α-strongly convex with $\alpha = -4U(r+2\kappa)^2 + 2L > 0$.

Next, we prove the β-smoothness of $\mathcal{L}$. It is equivalent to show $\nabla^2\mathcal{L}(\hat{\boldsymbol{\mu}}+\boldsymbol{r}) \preceq \beta\boldsymbol{I}$, i.e., $\lambda_{\min}\left(\beta\boldsymbol{I} - \nabla^2\mathcal{L}(\hat{\boldsymbol{\mu}}+\boldsymbol{r})\right) \geq 0$ for a constant $\beta > 0$. We have

$$\begin{aligned}
&\lambda_{\min}\left(\beta\boldsymbol{I} - \nabla^2\mathcal{L}(\hat{\boldsymbol{\mu}}+\boldsymbol{r})\right)\\
=&\lambda_{\min}\Bigg(\left(\beta - \frac{2}{n}\sum_{i=1}^{n}\ell'\left(d_i^2(\hat{\boldsymbol{\mu}}+\boldsymbol{r})\right)\right)\boldsymbol{I}\\
&\qquad -\frac{4}{n}\sum_{i=1}^{n}\ell''\left(d_i^2(\hat{\boldsymbol{\mu}}+\boldsymbol{r})\right)(\hat{\boldsymbol{\mu}}+\boldsymbol{r}-\boldsymbol{x}_i)(\hat{\boldsymbol{\mu}}+\boldsymbol{r}-\boldsymbol{x}_i)^\top\Bigg)\\
\geq& -\frac{4}{n}\sum_{i=1}^{n}\lambda_{\min}\left(\ell''\left(d_i^2(\hat{\boldsymbol{\mu}}+\boldsymbol{r})\right)(\hat{\boldsymbol{\mu}}+\boldsymbol{r}-\boldsymbol{x}_i)(\hat{\boldsymbol{\mu}}+\boldsymbol{r}-\boldsymbol{x}_i)^\top\right)\\
&\qquad + \beta - \frac{2}{n}\sum_{i=1}^{n}\ell'\left(d_i^2(\hat{\boldsymbol{\mu}}+\boldsymbol{r})\right), \quad (13)
\end{aligned}$$

where the inequality follows from Weyl's inequality. For the first term in (13), we have

$$\begin{aligned}
&-\frac{4}{n}\sum_{i=1}^{n}\lambda_{\min}\left(\ell''\left(d_i^2(\hat{\boldsymbol{\mu}}+\boldsymbol{r})\right)(\hat{\boldsymbol{\mu}}+\boldsymbol{r}-\boldsymbol{x}_i)(\hat{\boldsymbol{\mu}}+\boldsymbol{r}-\boldsymbol{x}_i)^\top\right)\\
\geq& -\frac{4}{n}\sum_{i=1}^{n}\left|\ell''\left(d_i^2(\hat{\boldsymbol{\mu}}+\boldsymbol{r})\right)\right|\left\|\hat{\boldsymbol{\mu}}+\boldsymbol{r}-\boldsymbol{x}_i\right\|^2, \quad (14)
\end{aligned}$$

where the first inequality is due to $\lambda_{\min}\left(\omega\boldsymbol{a}\boldsymbol{a}^\top\right) = -|\omega|\left\|\boldsymbol{a}\right\|^2$ for any scalar ω. Upon substituting (14) into (13), we get

$$\begin{aligned}
&\lambda_{\min}\left(\beta\boldsymbol{I} - \nabla^2\mathcal{L}(\hat{\boldsymbol{\mu}}+\boldsymbol{r})\right)\\
\geq&\beta - \frac{4}{n}\sum_{i=1}^{n}\left|\ell''\left(d_i^2(\hat{\boldsymbol{\mu}}+\boldsymbol{r})\right)\right|\left\|\hat{\boldsymbol{\mu}}+\boldsymbol{r}-\boldsymbol{x}_i\right\|^2\\
&\qquad - \frac{2}{n}\sum_{i=1}^{n}\ell'\left(d_i^2(\hat{\boldsymbol{\mu}}+\boldsymbol{r})\right)\\
\geq&\beta - \frac{4}{n}\sum_{i=1}^{n}\left|\ell''\left(d_i^2(\hat{\boldsymbol{\mu}}+\boldsymbol{r})\right)\right|\left(\|\boldsymbol{r}\| + \|\hat{\boldsymbol{\mu}}-\boldsymbol{\mu}^\star\| + \|\boldsymbol{\varepsilon}_i\|\right)^2.\\
&\qquad - \frac{2}{n}\sum_{i=1}^{n}\ell'\left(d_i^2(\hat{\boldsymbol{\mu}}+\boldsymbol{r})\right)\\
\geq&\beta - \frac{4}{n}\sum_{i=1}^{n}\left|\ell''\left(d_i^2(\hat{\boldsymbol{\mu}}+\boldsymbol{r})\right)\right|\left(\|\boldsymbol{r}\| + 2\kappa\right)^2\\
&\qquad - \frac{2}{n}\sum_{i=1}^{n}\ell'\left(d_i^2(\hat{\boldsymbol{\mu}}+\boldsymbol{r})\right),
\end{aligned}$$

where the second inequality follows from triangular inequality and the third inequality follows from Proposition 3 and Assumption 2. By using Lemma 5, we then have

$$\begin{aligned}
&\lambda_{\min}\left(\beta\boldsymbol{I} - \nabla^2\mathcal{L}(\hat{\boldsymbol{\mu}}+\boldsymbol{r})\right)\\
\geq&\beta - 4U(r+2\kappa)^2 - 2\overline{G}(\kappa, r)\\
\geq&\beta - 2L - 2U \quad (15)
\end{aligned}$$

The RHS of (15) is positive whenever

$$\beta \geq 2(L+U).$$

And it is easy to see $\alpha \leq \beta$.

Appendix C
Proof of Theorem 7

The gradient of $\mathcal{L}$ is given by

$$\nabla\mathcal{L}(\boldsymbol{\mu}) = \frac{2}{n}\sum_{i=1}^{n}\ell'\left(d_i^2(\boldsymbol{\mu})\right)(\boldsymbol{\mu}-\boldsymbol{x}_i).$$

Then we have

$$\begin{aligned}
&\nabla\mathcal{L}(\boldsymbol{\mu})^\top(\boldsymbol{\mu}-\hat{\boldsymbol{\mu}})\\
=&\frac{2}{n}\sum_{i=1}^{n}\ell'\left(d_i^2(\boldsymbol{\mu})\right)(\boldsymbol{\mu}-\boldsymbol{x}_i)^\top(\boldsymbol{\mu}-\hat{\boldsymbol{\mu}})\\
=&\frac{2}{n}\sum_{i=1}^{n}\ell'\left(d_i^2(\boldsymbol{\mu})\right)(\boldsymbol{\mu}-\hat{\boldsymbol{\mu}}+\hat{\boldsymbol{\mu}}-\boldsymbol{\mu}^\star-\boldsymbol{\varepsilon}_i)^\top(\boldsymbol{\mu}-\hat{\boldsymbol{\mu}})\\
=&\frac{2}{n}\sum_{i=1}^{n}\ell'\left(d_i^2(\boldsymbol{\mu})\right)\|\boldsymbol{\mu}-\hat{\boldsymbol{\mu}}\|^2\\
&\qquad + \frac{2}{n}\sum_{i=1}^{n}\ell'\left(d_i^2(\boldsymbol{\mu})\right)(\hat{\boldsymbol{\mu}}-\boldsymbol{\mu}^\star-\boldsymbol{\varepsilon}_i)^\top(\boldsymbol{\mu}-\hat{\boldsymbol{\mu}}). \quad (16)
\end{aligned}$$

For the second term in (16), by using the Cauchy-Schwarz inequality and the triangle inequality, we have

$$\begin{aligned}
&\frac{2}{n}\sum_{i=1}^{n}\ell'\left(d_i^2(\boldsymbol{\mu})\right)(\hat{\boldsymbol{\mu}}-\boldsymbol{\mu}^\star-\boldsymbol{\varepsilon}_i)^\top(\boldsymbol{\mu}-\hat{\boldsymbol{\mu}})\\
\geq& -\frac{2}{n}\sum_{i=1}^{n}\ell'\left(d_i^2(\boldsymbol{\mu})\right)\|\hat{\boldsymbol{\mu}}-\boldsymbol{\mu}^\star-\boldsymbol{\varepsilon}_i\|\|\boldsymbol{\mu}-\hat{\boldsymbol{\mu}}\|\\
\geq& -\frac{2}{n}\sum_{i=1}^{n}\ell'\left(d_i^2(\boldsymbol{\mu})\right)\left(\|\hat{\boldsymbol{\mu}}-\boldsymbol{\mu}^\star\| + \|\boldsymbol{\varepsilon}_i\|\right)\|\boldsymbol{\mu}-\hat{\boldsymbol{\mu}}\|. \quad (17)
\end{aligned}$$

Upon substituting (17) into (16), we have

$$\begin{aligned}
&\nabla\mathcal{L}(\boldsymbol{\mu})^\top(\boldsymbol{\mu}-\hat{\boldsymbol{\mu}})\\
\geq&\frac{2}{n}\sum_{i=1}^{n}\ell'\left(d_i^2(\boldsymbol{\mu})\right)\left(\|\boldsymbol{\mu}-\hat{\boldsymbol{\mu}}\| - \|\hat{\boldsymbol{\mu}}-\boldsymbol{\mu}^\star\| - \|\boldsymbol{\varepsilon}_i\|\right)\|\boldsymbol{\mu}-\hat{\boldsymbol{\mu}}\|\\
\geq&\frac{2}{n}\sum_{i=1}^{n}\ell'\left(d_i^2(\boldsymbol{\mu})\right)\left(\|\boldsymbol{\mu}-\hat{\boldsymbol{\mu}}\| - 2\kappa\right)\|\boldsymbol{\mu}-\hat{\boldsymbol{\mu}}\|,
\end{aligned}$$

where the last inequality follows from Proposition 3 and Assumption 2.

When $\kappa = 0$, we have

$$\begin{aligned}
\nabla\mathcal{L}(\boldsymbol{\mu})^\top(\boldsymbol{\mu}-\hat{\boldsymbol{\mu}}) \geq&\frac{2}{n}\sum_{i=1}^{n}\ell'\left(d_i^2(\boldsymbol{\mu})\right)\|\boldsymbol{\mu}-\hat{\boldsymbol{\mu}}\|^2\\
=&\gamma_1\|\boldsymbol{\mu}-\hat{\boldsymbol{\mu}}\|^2,
\end{aligned}$$

where $\gamma_1 = \frac{2}{n}\sum_{i=1}^{n} \ell'\left(d_i^2\left(\boldsymbol{\mu}\right)\right) > 0$ is a positive constant.

When $\kappa > 0$, if $\|\boldsymbol{\mu} - \hat{\boldsymbol{\mu}}\| \geq 4\kappa$, we have $\|\boldsymbol{\mu} - \hat{\boldsymbol{\mu}}\| - 2\kappa \geq 2\kappa$. Then there exists a positive constant $C_1 = \frac{2\kappa}{\|\boldsymbol{\mu} - \hat{\boldsymbol{\mu}}\|}$ such that $\|\boldsymbol{\mu} - \hat{\boldsymbol{\mu}}\| - 2\kappa \geq C_1 \|\boldsymbol{\mu} - \hat{\boldsymbol{\mu}}\|$. For the lower bound of C_1, we have

$$C_1 = \frac{2\kappa}{\|\boldsymbol{\mu} - \hat{\boldsymbol{\mu}}\|} \geq \frac{2\kappa}{\|\boldsymbol{\mu} - \boldsymbol{\mu}^\star\| + \|\boldsymbol{\mu}^\star - \hat{\boldsymbol{\mu}}\|}. \tag{18}$$

Consider term $\|\boldsymbol{\mu} - \boldsymbol{\mu}^\star\|$, we have

$$\begin{aligned} \|\boldsymbol{\mu} - \boldsymbol{\mu}^\star\| &\leq \|\boldsymbol{\mu} - \boldsymbol{x}_i\| + \|\boldsymbol{x}_i - \boldsymbol{\mu}^\star\| \\ &= \|\boldsymbol{\mu} - \boldsymbol{x}_i\| + \|\boldsymbol{\varepsilon}_i\| \\ &\leq \max_{i=1,\ldots,n} \|\boldsymbol{\mu} - \boldsymbol{x}_i\| + \kappa, \end{aligned} \tag{19}$$

where the first inequality follows from the triangle inequality and the second inequality follows from $\|\boldsymbol{\mu}^\star - \hat{\boldsymbol{\mu}}\| \leq \kappa$ by Proposition 3. Upon substituting (19) into (18), we have

$$C_1 \geq \frac{2\kappa}{\max_{i=1,\ldots,n} \|\boldsymbol{\mu} - \boldsymbol{x}_i\| + 2\kappa}.$$

Then we have

$$\begin{aligned} \nabla \mathcal{L}\left(\boldsymbol{\mu}\right)^\top \left(\boldsymbol{\mu} - \hat{\boldsymbol{\mu}}\right) &\geq \frac{2C_1}{n}\sum_{i=1}^{n} \ell'\left(d_i^2\left(\boldsymbol{\mu}\right)\right) \|\boldsymbol{\mu} - \hat{\boldsymbol{\mu}}\|^2 \\ &\geq \gamma_2 \|\boldsymbol{\mu} - \hat{\boldsymbol{\mu}}\|^2, \end{aligned}$$

where $\gamma_2 = \frac{4\kappa}{\max_{i=1,\ldots,n}\|\boldsymbol{\mu} - \boldsymbol{x}_i\| + 2\kappa} L > 0$ when $\kappa > 0$. Then we have a positive constant

$$\gamma = \min\left\{\gamma_1, \gamma_2\right\}$$

such that

$$\nabla \mathcal{L}\left(\boldsymbol{\mu}\right)^\top \left(\boldsymbol{\mu} - \hat{\boldsymbol{\mu}}\right) \geq \gamma \|\boldsymbol{\mu} - \hat{\boldsymbol{\mu}}\|^2.$$

Since we have $\gamma > 0$, we have $\mathcal{L}$ is one-point strongly convex in $\mathcal{B}\left(\hat{\boldsymbol{\mu}}, r\right)$, where $r \geq 4\kappa$.

APPENDIX D
PROOF OF COROLLARY 8

From Theorem 6, $\mathcal{L}$ is locally strong convex and smooth in the interior of the ball $\mathcal{B}\left(\hat{\boldsymbol{\mu}}, R\right)$, so there is only one local minimum in $\mathcal{B}\left(\hat{\boldsymbol{\mu}}, R\right)$. From Theorem 7, $\mathcal{L}$ is one-point strong convex in the interior of the ball $\mathcal{B}\left(\hat{\boldsymbol{\mu}}, r\right)^c$, so there is no local minimum in $\mathcal{B}\left(\hat{\boldsymbol{\mu}}, r\right)^c$. If $R > r$, there is an overlap between $\mathcal{B}\left(\hat{\boldsymbol{\mu}}, R\right)$ and $\mathcal{B}\left(\hat{\boldsymbol{\mu}}, r\right)^c$. Then we have $\mathcal{L}$ has a unique local minimum, which is the global minimum.

Furthermore, we demonstrate that $R > r$ can not have no solution. If $R > r$, it indicates that there exists an R satisfying $R < \sqrt{\frac{L}{2U}} - 2\kappa$ and $r \geq 4\kappa$. Combining these two condition, we need $\kappa < \frac{1}{6}\sqrt{\frac{L}{2U}}$, which can be satisfied by a small enough κ since $\sqrt{\frac{L}{2U}} > 0$.