Supplementary Materials for Robust Subspace Clustering with Independent and Piecewise Identically Distributed Noise Modeling

Yuanman Li, Jiantao Zhou, Xianwei Zheng, Jinyu Tian, Yuan Yan Tang
Department of Computer and Information Science, University of Macau, Macau, China
{yuanmanli, jtzhou, yb47430, yb77405, yytang}@um.edu.mo

1. Proof of Theorem 3.1 in Section 3.3

Proof. Let $\hat{f}(e)$ be the *pdf* estimated using the Parzen window estimation, i.e.,

$$\hat{f}(e) = \frac{1}{N}\sum_{i=1}^{N}\kappa_\sigma(e - e_i).$$

In order to prove $H_2(\mathbf{e}) = \bar{H}_2(\mathbf{e})$ under the *i.i.d.* assumption, we first show that $\hat{f}(e)$ is equivalent to $\hat{f}_{Iq}(e)$ ($I_q \in [1, N]$) in terms of the mean integrated squared error ($MISE$) [5]. Since $\hat{f}_{Iq}(e)$ and $\hat{f}(e)$ are density estimators over finite samples independently sampled from the same distribution, we prove their equivalence by showing that

$$MISE(\hat{f}_{Iq}(e), \hat{f}(e)) = 0,$$

where

$$\begin{aligned} MISE(\hat{f}_I(e), \hat{f}(e)) &= E\int \left(\hat{f}_{Iq}(e) - \hat{f}(e)\right)^2 de \\ &= \int E\left(\left(\hat{f}_{Iq}(e) - \hat{f}(e)\right)^2\right) de \end{aligned} \tag{1}$$

Note that $E(\cdot)$ takes the expected value over all possible sequences $\mathbf{e}$. By the definitions of $\hat{f}_{Iq}(e)$ and $\hat{f}(e)$, we have

$$\begin{aligned} & E\left(\left(\hat{f}_{Iq}(e) - \hat{f}(e)\right)^2\right) = E\left(\hat{f}_{Iq}(e)^2 + \hat{f}(e)^2 - 2\hat{f}_{Iq}(e)\hat{f}(e)\right) \\ &= E\left(\sum_{i=1}^{N} c(D_{q,i})\kappa_\sigma(e - \hat{e}_i)\right)^2 + E\left(\sum_{i=1}^{N}\frac{1}{N}\kappa_\sigma(e - \hat{e}_i)\right)^2 \\ &- 2E\left(\sum_{i,j=1}^{N}\frac{1}{N}c(D_{q,i})\kappa_\sigma(e - \hat{e}_i)\kappa_\sigma(e - \hat{e}_j)\right) \\ &= E\left(\sum_{i,j=1}^{N} U_{i,j}\kappa_\sigma(e - \hat{e}_i)\kappa_\sigma(e - \hat{e}_j)\right), \end{aligned}$$

where

$$U_{i,j} = c(D_{q,i})c(D_{q,j}) + \left(\frac{1}{N}\right)^2 - \frac{2}{N}c(D_{q,i})$$

Since $\mathbf{e}$ is generated by an *i.i.d.* source, we have

$$E\left(\kappa_\sigma(e - \hat{e}_i)\kappa_\sigma(e - \hat{e}_j)\right) = E\left([\kappa_\sigma(e - \hat{e})]^2\right).$$

Then

$$E\left(\left(\hat{f}_{Iq}(e)-\hat{f}(e)\right)^2\right)=\sum_{i,j=1}^{N}U_{i,j}E\left([\kappa_\sigma(e-\hat{e})]^2\right)$$

Since $c(\cdot)\geq 0$ and $\sum_{i=1}^{N}c(D_{q,i})=1$ for each I_q, it can be readily proved that

$$\sum_{i,j=1}^{N}U_{i,j}=0.$$

So $MISE(\hat{f}_{Iq}(e),\hat{f}(e))=0$ for all $I_q\in[1,N]$. Finally we have

$$\begin{aligned}\bar{H}_2(\mathbf{e})&=-\frac{1}{N}\sum_{I_q}\log\int\left(\hat{f}_{I_q}(e)\right)^2de\\&=-\frac{1}{N}\sum_{I_q}\log\int\left(\hat{f}(e)\right)^2de\qquad(2)\\&=-\int\left(\hat{f}(e)\right)^2de=H_2(\mathbf{e})\qquad(3)\end{aligned}$$

This completes the proof. □

2. Proof of $\sum_i w_{i,j}=1$ in (31)

Proof. Since $\sum_{i=1}^{N}c(D_{q,i})=1$ for any q, and $c(x)$ is an even function, we can directly have $\sum_q c(D_{q,j})=1$ for any I_j. Then

$$\begin{aligned}\sum_i w_{i,j}&=\sum_i\sum_q c_{i,j}^q=\sum_i\sum_q c(D_{q,i})c(D_{q,j})\\&=\sum_q\sum_i c(D_{q,i})c(D_{q,j})\\&=\sum_q c(D_{q,j})=1.\qquad(4)\end{aligned}$$

This completes the proof. □

3. Algorithm for the Problem Shown in (32)

For simplicity, we replace the notation $\sqrt{2}\sigma$ by σ in (32). Define the objective function in (32) as $J(\mathbf{z})$, i.e.,

$$J(\mathbf{z})=-\sum_{i=1}^{N}\kappa_\sigma(\tilde{y}_i-\tilde{\mathbf{x}}_i\mathbf{z})+\lambda||\mathbf{z}||_1.\qquad(5)$$

Since the first term of $J(\mathbf{z})$ is highly nonlinear, making the problem (32) difficult to directly optimize. Fortunately, we show that (32) can be efficiently solved by applying the half-quadratic theory [4], which is widely used for ITL-based optimization problems [2]. According to the property of convex conjugate function [1], we have:

Proposition 1. *For the function $\kappa_\sigma(x)$, there exists a convex conjugate function $\varphi(\cdot)$ of $\kappa_\sigma(x)$, such that*

$$\kappa_\sigma(x)=\sup_s\left(\frac{sx^2}{\sigma^2}-\varphi(s)\right).\qquad(6)$$

Given x, the supremum is reached at $s=-\kappa_\sigma(x)$.

Letting $u = -\frac{s}{\sigma^2}$ and defining a function $\psi(u) = \varphi(-\sigma^2 u)$, (6) can be equivalently written as

$$-\kappa_\sigma(x) = \inf_u \Big(ux^2 + \psi(u) \Big). \tag{7}$$

The infimum is then achieved at $u = \frac{1}{\sigma^2}\kappa_\sigma(x)$. Now we rewrite (5) as an augmented cost function

$$J(\mathbf{z}, \mathbf{u}) = \sum_{i=1}^{N} \Big(u_i(\tilde{y}_i - \tilde{\mathbf{x}}_i \mathbf{z})^2 + \psi(u_i) \Big) + \lambda ||\mathbf{z}||_1, \tag{8}$$

where $\mathbf{u} = [u_1, ..., u_N]^T$ are the auxiliary variables introduced by half-quadratic optimization. According to (7), for a fixed $\mathbf{z}$, we have

$$J(\mathbf{z}) = \min_{\mathbf{u}} J(\mathbf{z}, \mathbf{u}).$$

Then the original problem (29) is identical to minimizing the augmented cost function, i.e.,

$$\min_{\mathbf{z}} J(\mathbf{z}) = \min_{\mathbf{z}, \mathbf{u}} J(\mathbf{z}, \mathbf{u}). \tag{9}$$

According to the half-quadratic optimization [4], $J(\mathbf{z}, \mathbf{u})$ can be minimized in the following alternate steps:

$$u_i^{t+1} = \frac{1}{\sigma^2}\kappa_\sigma(\tilde{y}_i - \tilde{\mathbf{x}}_i \mathbf{z}^t), \ i = 1, 2, \cdots, N, \tag{10}$$

$$\mathbf{z}^{t+1} = \operatorname*{argmin}_{\mathbf{z}} \, (\tilde{\mathbf{y}} - \tilde{\mathbf{X}}\mathbf{z})^T \text{diag}(\mathbf{u}^{t+1})(\tilde{\mathbf{y}} - \tilde{\mathbf{X}}\mathbf{z}) + \lambda ||\mathbf{z}||_1, \tag{11}$$

where t means the t-th iteration, $\tilde{\mathbf{y}} = [\tilde{y}_1, ..., \tilde{y}_N]^T$, $\tilde{\mathbf{X}} = [\tilde{\mathbf{x}}_1^T, ..., \tilde{\mathbf{x}}_N^T]^T$ and diag$(\cdot)$ is an operator to convert a vector to a diagonal matrix. The convergence of the algorithm was proved in [4]. The subproblem (11) can be efficiently solved using the sparse coding method proposed in [3].

References

[1] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004. 2

[2] R. He, B. Hu, X. Yuan, L. Wang, et al. *Robust recognition via information theoretic learning*. Springer, Amsterdam, The Netherlands, 2014. 2

[3] H. Lee, A. Battle, R. Raina, and A. Y. Ng. Efficient sparse coding algorithms. In B. Schölkopf, J. C. Platt, and T. Hoffman, editors, *Advances in Neural Information Processing Systems 19*, pages 801–808. MIT Press, 2007. 3

[4] M. Nikolova and M. K. Ng. Analysis of half-quadratic minimization methods for signal and image recovery. *SIAM J. Sci. Comput.*, 27(3):937–966, 2005. 2, 3

[5] M. P. Wand and M. C. Jones. *Kernel smoothing*. Chapman and Hall/CRC, 1994. 1