A APPENDIX

In this supplementary file, we provide technical proofs of the theoretical results in Section 4.3, and present extra empirical experiments regarding our kernel diffusion approach with symmetric and asymmetric Gaussian kernels applied to DBSCAN. All the numerical experiments are carried out on a standard work station with a Intel 64-cores CPU and two Nvidia P100 GPUs.

A.1 PROOFS OF THEORETICAL RESULT.

Proof of Theorem 1. Since $\{D_1, \cdots, D_m\}$ are disjoint, we have $p(x, y) = 0$ if x and y belong to different clusters. By the definition of matrix P, for each $x \in D_j$, we have

$$\int_D p(x, y) dF_n(y) = 1,$$

which implies that

$$\int_{x \in D_j} \int_{y \in D_j} p(x, y) dF_n(x) dF_n(y) = |D_j|.$$

Therefore,

$$\bar{\rho}_j |D_j| = \int_{x \in D_j} \int_{y \in D_j} p(x, y) dF_n(x) dF_n(y) = |D_j|,$$

which implies that $\bar{\rho}_j = 1$ for any $j = 1, , \ldots, m$. □

Before proceeding to the proof of Theorem 2, we need following auxiliary lemma that relates the stationary distribution of a Markov chain to an arbitrary vector g.

Lemma A.1. *Let P be transition probability matrix of a finite inreducible discrete time Markov chain with n states, which admits a stationary distribution, denoted by vector π. We write $e = (1, \ldots, 1)^T \in \mathbb{R}^n$ as the a column vector of ones. The following holds for any vector g such that $g^T e \neq 0$:*

(1) $(I - P + eg^T)$ is non-singular.

(2) Let $H = (I - P + eg^T)^{-1}$, then $\pi^T = g^T H$.

Proof. Since π is the stationary distribution, we have $\pi^T e = 1$. Applying Theorem 3.3 in (Hunter, 1982) yields that matrix $(I - P + eg^T)$ is non-singular.

Next recall that $\pi^T P = \pi^T$, therefore we have

$$\begin{aligned} \pi^T (I - P + eg^T) &= \pi^T - \pi^T P + \pi^T eg^T \\ &= \pi^T eg^T \\ &= g^T, \end{aligned}$$

which implies $\pi^T = g^T H$. □

Proof of Theorem 2. Note that for for each $x \in D$, the linear reference function $\rho_{\text{FKD}}(x) = \int_D p(y, x) dF_n(y)$ is the corresponding column average of the transition matrix P. We write the i-th column vector of P as

$$p_i = \big(p(x_1, x_i), \ldots, p(x_n, x_i)\big)^T.$$

Therefore $\rho_{\text{FKD}}(x_i) = e^T p_i / n$

Since the Markov chain induced by the kernel $k(x, y)$ is ergodic, the density $\rho(x, t)$ of the diffusion process X_t, will converge to the limiting stationary distribution of the Markov chain, denoted by π.

We can write the n-vectors of g and π in the following form:

$$g = (g_1, \ldots, g_n)^T = n\big(\rho_{\text{FKD}}(x_1), \ldots, \rho_{\text{FKD}}(x_n)\big)^T \quad \text{and} \quad \pi = \big(\rho_{\text{KD}}(x_1), \ldots, \rho_{\text{KD}}(x_n)\big)^T,$$

where $g_i = e^T p_i$ is the i-th column sums of matrix P. As a result, we have

$$\int_D \hat{g}(x) dF_n(x) = \frac{1}{n} e^T g = 1, \quad \text{and} \quad n \int_D \hat{\rho}(x) dF_n(x) = e^T \pi = 1.$$

By the definition of g, we know

$$(eg^T)^2 = neg^T \quad \text{and} \quad e^T P = g^T.$$

It follows from Lemma A.1 that $(I - P + eg^T)$ is non-singular and $\pi^T = g^T H$, where $H = (I - P + eg^T)^{-1}$.

We define $D = I + eg^T$. By simple algebra calculation, we can find D is non-singular with

$$D^{-1} = I - \frac{eg^T}{n+1}.$$

As a result, it is easy to see that that $g^T D^{-1} = \frac{g^T}{n+1}$ and

$$H^{-1} = D - P = (I - PD^{-1})D.$$

Use the Neumann series, we have

$$H = D^{-1}(I - PD^{-1})^{-1} = D^{-1} \sum_{i=0}^{\infty} (PD^{-1})^i.$$

Thus

$$\pi^T - g^T/n = g^T \left(H - \frac{I}{n} \right) = g^T \left[D^{-1} \sum_{i=0}^{\infty} (PD^{-1})^i - \frac{I}{n} \right].$$

Since we assume for any $x \in D$, $\hat{g}(x) < c \quad$ for some $0 < c < 1$. This leads to

$$g^T p_j \leq nce^T p_j = ncg_j.$$

Therefore, let κ_j be the j-th compoenent of $g^T PD^{-1}$, it is straightforward

$$\kappa_j \leq \frac{nc}{n+1} g_j \leq cg_j.$$

This implies for every $x \in D$,

$$\begin{aligned} |\rho_{\text{KD}}(x) - \rho_{\text{FKD}}(x)| &\leq \rho_{\text{FKD}}(x) \left| \frac{1}{n+1} \sum_{i=0}^{\infty} c^i - \frac{1}{n} \right| \\ &\leq \rho_{\text{FKD}}(x) \left| \frac{1}{(n+1)(1-c)} - \frac{1}{n} \right|. \end{aligned}$$

Hence we have $\lim_{n \to \infty} |\frac{\rho_{\text{KD}}(x)}{\rho_{\text{FKD}}(x)} - 1| = 0$, which completes the proof. □

A.2 Additional Experiment Results

Metadata of benchmark datasets. The number of samples n, the number of clusters c, and feature dimension d for each benchmark dataset are listed in Table 4 below.

Benckmark datasets with DBSCAN. We provide the performance of the conventional density functions, ρ_{naive} and ρ_{LC}, and the proposed kernel diffusion density functions with symmetric and asymmetric Gaussian kernels, ρ^*_{KD} and ρ^*_{FKD} ($* \in \{\text{sym}, \text{asym}\}$), applied to DBSCAN on 13 benchmark datasets. The results are summarised in Table 5. Similar to DPC, we see that both $\rho^{\text{sym}}_{\text{KD}}$ and $\rho^{\text{asym}}_{\text{KD}}$ uniformly outperform ρ_{naive} and ρ_{LC} in terms of clustering quality. $\rho^{\text{asym}}_{\text{KD}}$, which has better local adaptivity analytically, achieves the best results on most datasets and outperforms others by a significant margin in Breast-o, Control, Haberma and Seeds.

Table 4: Metadata of benchmark datasets, includes sample size (n), the number of clusters (c), and feature dimension d.

Dataset	n	c	d
Banknote	1372	2	4
Breast-d	569	2	30
Breast-o	699	2	9
Control	600	6	60
Glass	214	7	9
Haberman	306	2	3
Ionosphere	351	2	34
Iris	150	3	4
Libras	360	15	90
Pageblocks	5473	5	10
Seeds	210	3	7
Segment	210	7	19
Wine	178	3	13

Table 5: Clustering performance on benchmark datasets with different density functions applied to DBSCAN. Pairwise F-score (F_P) and BCube F-score (F_B) under optimal parameter tuning are given. The best and second-best results in each dataset are bolded and underlined, respectively.

Dataset	F_P						F_B					
	ρ_{naive}	ρ_{LC}	$\rho_{\text{KD}}^{\text{sym}}$	$\rho_{\text{KD}}^{\text{asym}}$	$\rho_{\text{FKD}}^{\text{sym}}$	$\rho_{\text{FKD}}^{\text{asym}}$	ρ_{naive}	ρ_{LC}	$\rho_{\text{KD}}^{\text{sym}}$	$\rho_{\text{KD}}^{\text{asym}}$	$\rho_{\text{FKD}}^{\text{sym}}$	$\rho_{\text{FKD}}^{\text{asym}}$
Banknote	26.8	60.7	62.0	**66.4**	65.4	**66.4**	26.5	65.1	60.7	**67.4**	65.7	**67.4**
Breast-d	56.7	63.0	65.0	66.6	**67.2**	66.6	60.9	64.7	66.0	67.4	**67.2**	67.4
Breast-o	18.2	55.3	59.2	**70.6**	70.5	**70.6**	15.9	50.7	52.3	**71.3**	70.6	71.2
Control	32.5	37.1	51.0	**60.3**	48.9	59.1	34.2	50.1	53.7	**66.9**	51.6	65.7
Glass	22.0	29.8	29.8	**42.5**	42.0	**42.5**	25.8	36.9	36.9	**45.2**	43.5	**45.2**
Haberman	68.6	72.2	68.6	**75.6**	68.9	75.3	69.2	73.1	69.2	**75.8**	68.3	75.7
Ionosphere	25.9	68.4	68.0	**74.2**	**74.2**	**74.2**	23.8	64.1	63.7	**72.1**	**72.1**	**72.1**
Iris	66.2	69.8	66.2	57.2	**73.7**	73.3	67.2	76.6	67.2	67.0	**79.4**	79.0
Libras	18.1	12.0	15.6	13.8	**20.2**	13.5	31.1	16.5	42.1	**45.5**	32.9	37.7
Pageblocks	48.4	89.2	90.0	**90.1**	89.9	**90.1**	45.2	85.5	**89.7**	89.5	**89.7**	89.5
Seeds	57.8	47.6	57.8	**63.2**	22.4	62.4	59.2	53.0	59.2	**70.0**	24.4	69.2
Segment	18.5	47.9	**54.8**	30.8	41.4	30.8	22.5	55.2	**66.6**	53.6	59.9	53.6
Wine	40.5	40.5	40.5	49.5	**50.0**	49.5	45.7	45.7	45.7	**52.3**	51.3	**52.3**