Appendix to: On the Relation Between Low Density Separation, Spectral Clustering and Graph Cuts

Hariharan Narayanan
Department of Computer Science
University of Chicago
Chicago IL 60637
hari@cs.uchicago.edu

Mikhail Belkin
Department of Computer Science and Engineering
The Ohio State University
Columbus, OH 43210
mbelkin@cse.ohio-state.edu

Partha Niyogi
Department of Computer Science
University of Chicago
Chicago IL 60637
niyogi@cs.uchicago.edu

A Regularity conditions on p and S

We make the following assumptions about p:

1. p can be extended to a function p' that is $L-$Lipshitz and which is bounded above by p_{max}.
2. For $0 < t < t_0$,
$$\min(p(x), \int K_t(x,y)p(y)dy) \geq p_{min}.$$
Note that this is a property of both of the boundary ∂M and p.

We note that since p' is $L-$Lipshitz over $\mathbb{R}^d$, so is $\int_M K_t(x,z)p'(z)dz$.

We assume that S has condition number $1/\tau$. We also make the following assumption about S:-
The volume of the set of points whose distance to both S and ∂M is $\leq R$, is $O(R^2)$ as $R \to 0$. This is reasonable, and is true if $S \cap \partial M$ is a manifold of codimension 2.

B Proof of Theorem 1

This follows from Theorem 4 (which is proved in a later section), by setting μ to be equal to $\frac{1-2\epsilon}{2d+2}$.

C Proof of Theorem 2

In the proof we will use a generalization of McDiarmid's inequality from [7, 8]. We start with the with the following

Definition 1 *Let $\Omega_1, \dots, \Omega_m$ be probability spaces. Let $\Omega = \prod_1^m \Omega_k$ and let Y be a random variable on Ω. We say that Y is strongly difference-bounded by (b, c, δ) if the following holds: there is a "bad" subset $B \subset \Omega$, where $\delta = Pr(\omega \in B)$. If $\omega, \omega' \in \Omega$ differ only in the kth coordinate, and $\omega \notin B$, then*
$$|Y(\omega) - Y(\omega')| \leq c.$$

Furthermore, for any ω and ω' differing only in the kth coordinate,

$$|Y(\omega) - Y(\omega')| \leq b.$$

Theorem 1 *([7, 8]) Let $\Omega_1, \ldots, \Omega_m$ be probability spaces. Let $\Omega = \prod_1^m \Omega_k$ and let Y be a random variable on Ω which is strongly difference-bounded by (b, c, δ). Assume $b \geq c > 0$. Let $\mu = E(Y)$. Then for any $r > 0$,*

$$Pr(|Y - \mu| \geq r) \leq 2\left(\exp\left(\frac{-r^2}{8mc^2}\right) + \frac{mb\delta}{c}\right).$$

By Hoeffding's inequality

$$\begin{aligned} P[|\frac{\sum_{z\neq x} K_t(x,z)}{N-1} - E(K_t(x,z))| > \epsilon_1 E(K_t(x,z))] &< \mathrm{e}^{-\frac{2(N-1)E(K_t(x,z))^2\epsilon_1^2}{M_t^2}} \\ &\leq \mathrm{e}^{-\frac{2(N-1)p_{min}^2\epsilon_1^2}{M_t^2}}. \end{aligned}$$

We set ϵ_1 to be $M_t/N^{\frac{1-\mu}{2}}$. Let $\mathrm{e}^{-\frac{2(N-1)p_{min}^2\epsilon_1^2}{M_t^2}}$ be δ/N. By the union bound, the probability that the above event happens for some $x \in X$ is $\leq \delta$. The set of all $\omega \in \Omega$ for which this occurs shall be denoted by B. Also, for any X, the largest possible value that

$$1/N\sqrt{\pi/t}\sum_{x\in X_1}\sum_{y\in X_2}\frac{K_t(x,y)}{\{(\sum_{z\neq x}K_t(x,z))(\sum_{z\neq y}K_t(y,z))\}^{1/2}}$$

could take is $\sqrt{\pi/t}(N-1)$. Then,

$$|E[\beta] - \alpha| < |1 - (1-\epsilon_1)^{-1}|\alpha + \delta\sqrt{\pi/t}(N-1). \tag{1}$$

Let $q = (p_{min}/M_t)^2$. β is *strongly difference-bounded* by (b, c, δ) where $c = O((qN\sqrt{t})^{-1})$, $b = O(N/\sqrt{t})$. We now apply the generalization of McDiarmid's inequality in Theorem 1. Using the notation of Theorem 1,

$$\Pr[|\beta - E[\beta]| > r] \leq 2\left(\exp\left(\frac{-r^2}{8mc^2}\right) + \frac{Nb\delta}{c}\right) \leq 2\left(\exp\left(-O(Nr^2q^2t)\right) + O\left(N^3 q \exp\left(-O(Nq\epsilon_1^2)\right)\right)\right). \tag{2}$$

Putting this together with the relation between $E[\beta]$ and α in (1), the theorem is proved. We note that in (1), the rate of convergence of $E[\beta]$ to α is controlled by ϵ_1, which is $M_t/N^{\frac{1-\mu}{2}}$, and in (2), the rate of convergence of β to $E[\beta]$ depends on r, which we set to be

$$M_t^2/\sqrt{tN^{1-\mu}}.$$

We note that in (2), the dependence on r of the *probability* is exponential. Since we have assumed that $u = M_t^2/\sqrt{(tN^{1-\mu})} = o(1)$, $M_t/N^{\frac{1-\mu}{2}} = O(t^{\frac{d+1}{2}}u)$. Thus the result follows.

□

D Proof of Theorem 3

We shall prove theorem 3 through a sequence of lemmas.

Without a loss of generality we can assume that $\tau = 1$ by rescaling, if necessary.

Let $R = \sqrt{2dt\ln(1/t)}$ and $\epsilon = \int_{\|z\|>R} K_t(0,z)dx$. Using the inequality

$$\int_{\|z\|>R} K_t(0,z)dx \leq \left(\frac{2td}{R^2}\right)^{-d/2} \mathrm{e}^{-\frac{R^2}{4t}+\frac{d}{2}} = (\mathrm{e}t\ln(1/t))^{d/2} \tag{3}$$

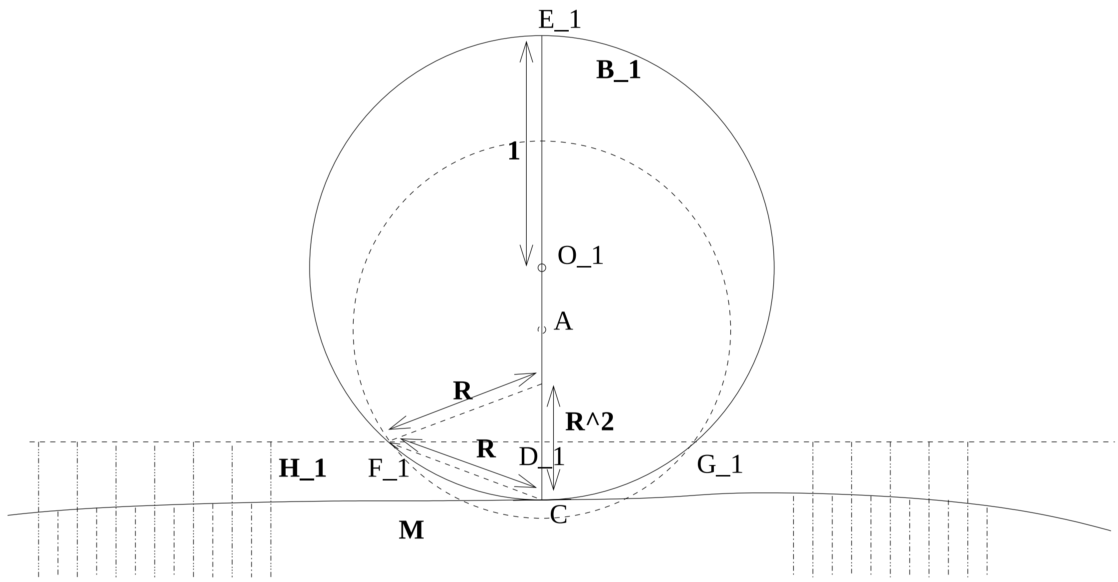

Figure 1: A sphere of radius 1 outside S_1 that is tangent to S

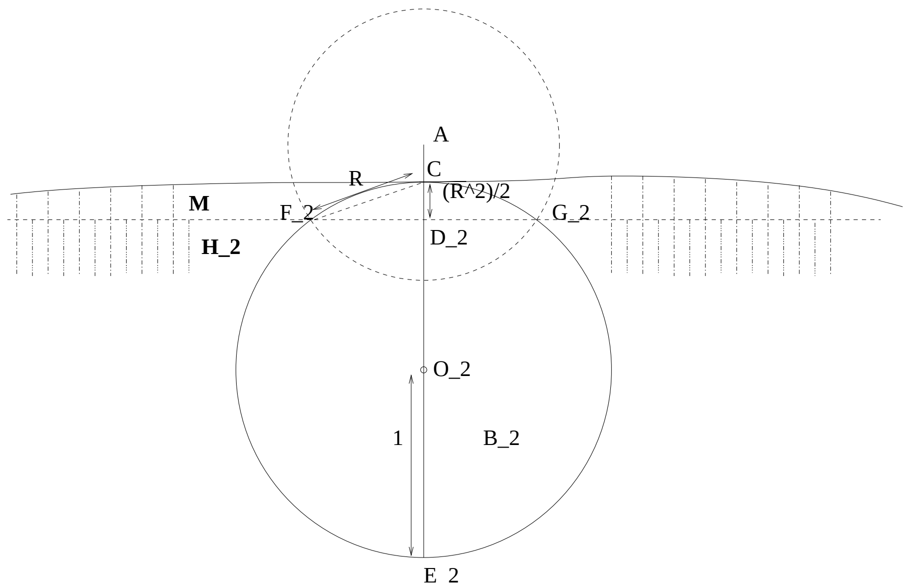

Figure 2: A sphere of radius 1 inside S_1 that is tangent to S

we know that $\epsilon \leq (\mathrm{e}t\ln(1/t))^{d/2}$. For any positive real t,

$$\ln(1/t) \leq t^{\frac{-1}{\mathrm{e}}}.$$

Therefore the assumption that

$$\frac{t}{\tau} \in \left(0, \frac{1}{(2d)^{\frac{\mathrm{e}}{\mathrm{e}-1}}}\right)$$

implies that $R \leq \sqrt{2dt^{1-1/\mathrm{e}}} < 1$.

Let the point y (represented as A in figures D and D) be at a distance $r < R$ from M. Let us choose a coordinate system where $y = (r, 0, \ldots, 0)$ and the point nearest to it on M is the origin. There is a unique such point since $r < R < 1$. Let this point be C. Let D_1 lie on the segment AC, at a distance $R^2/2$ from C. Let D_2 lie on the extended segment AC, at a distance $R^2/2$ from C. Thus C is the midpoint of D_1D_2.

Definition 2 *1. Denote the ball of radius* 1 *tangent to* ∂M *at* C *that is outside* M *by* B_1.

2. *Denote the ball of radius* 1 *tangent to* ∂M *at* C *which is inside* M *by* B_2.
3. *Let* H_1 *be the halfspace containing* C *bounded by the hyperplane perpendicular to* AC *and passing through* D_1.
4. *Let* H_2 *be the halfspace not containing* C *bounded by the hyperplane perpendicular to* AC *and passing through* D_2.
5. *Let* H_3 *be the halfspace not containing* A*, bounded by the hyperplane tangent to* ∂M *at* C.
6. *Let* B_1' *be the ball with center* $y = A$*, whose boundary contains the intersection of* H_1 *and* B_1.
7. *Let* B_2' *be the ball with center* $y = A$*, whose boundary contains the intersection of* H_2 *and* B_2.

Definition 3 *1.* $h(r) := \int_{H_3} K_t(x, y)dx$.

2. $f(r) := \int_{H_2 \cap B_2'} K_t(x, y)dx$.
3. $g(r) := \int_{H_1 \cap B_1'} K_t(x, y)dx$.

It follows that

$$\int_{H_1} K_t(x, y)dx = h(r - R^2/2)$$

and

$$\int_{H_2} K_t(x, y)dx = h(r + R^2/2).$$

Observation 1 *Although* $h(r)$ *is defined by an d-dimensional integral, this can be simplified to*

$$h(r) = \int_{x_1 < 0} \frac{e^{-(r-x_1)^2/4t}}{\sqrt{4\pi t}} dx_1,$$

by integrating out the coordinates $x_2, \ldots, x_d$.

Lemma 1 *If* $r > R^2$*, the radius of* B_1' *is* $\geq R$.

Proof: By the similarity of triangles CF_1D_1 and CE_1F_1 in figure D, it follows that $\frac{CF_1}{CE_1} = \frac{CD_1}{CF_1}$. $|CE_1| = 2$ and $|CD_1| = R^2/2$. Therefore $CF_1 = R$. Since CD_1F_1 is right angled at D_1, and $|CD_1| = R^2/2$, this proves the claim. □

Lemma 2 *The radius of* B_2' *is* $\geq R$.

Proof: By the similarity of triangles CF_2E_2 and CD_2F_2 in figure D $|CF_2| = R$. However, the distance of point $y := A$ from F_2 is $\geq |CF_2|$. Therefore, the radius of B_2' is $\geq R$. □

Definition 4 *Let the set of points* x *such that* $B(x, 10R) \subseteq M$ *be denoted by* M^0*. Let* $S_1 \cap M^0$ *be* S_1^0 *and* $S_2 \cap M^0$ *be* S_2^0*. Let* $M - M^0 = M^1$*,* $S_1 \cap M^1$ *be* S_1^1 *and* $S_2 \cap M^1$ *be* S_2^1*. We shall denote* $(1 + L/p_{min})R)$ *by* ℓ.

Consider a point $x \in M^0$, Then,

$$\begin{aligned}\int_M K_t(x,y)p(y)dy &\geq \int_{\|y-x\|<R} K_t(x,y)p(y)dy \\ &\geq (1-\epsilon)(p(x)-LR) \\ &\geq (1-\epsilon)p(x)(1-LR/p_{min}) \\ &= p(x)(1-O(\ell))\end{aligned}$$

On the other hand,

$$\begin{aligned}\int_M K_t(x,y)p(y)dy &\leq \int_{\|y-x\|\leq 2R} K_t(x,y)p(y)dy + \int_{\|y-x\|>2R} K_t(x,y)p(y)dy \\ &\leq p(x)(1+2\ell) + K_t(0,2R) \\ &= p(x)(1+O((1+L/p_{min})R))\end{aligned}$$

Therefore, $\psi_t(x) = \sqrt{p(x)}(1 \pm O((1+\frac{L}{p_{min}})R))$.

Lemma 3 *$B(x,5R) \subseteq M$ implies that $\frac{d}{dx}\int K_t(x,z)p(z)dy = O(L)$.*

Proof: Consider the function p', which is equal to p on M, but which has a larger support and is $L-$Lipshitz as a function on $\mathbb{R}^d$. $\int K_t(x,z)p'(z)dy$ is $L-$Lipshitz and on points x where $B(x,5R) \subseteq M$, the contribution of points z outside M, is $o(1)$. Therefore $\frac{d}{dx}\int K_t(x,z)p(z)dy = O(L)$. □

This implies that on the set of points x such that $B(x,5R) \subseteq M$, $\psi_t(X)$ is $O(L)-$Lipshitz.

We now estimate $\int_{S_1} K_t(y,z)p(z)dz$ for $y \in S_2^0$.

Definition 5 *For a point $y \in S_2^0$, such that $d(y,S_1) < R < \tau = 1$ let $\pi(y)$ be the nearest point to y in S.*

Note that by the assumption that the condition number of S is 1, since R is smaller than 1, there is a unique candidate for $\pi(y)$. Let y be as in Definition 5.

Lemma 4

$$h(r+R^2/2) - \epsilon < f(r) \leq \int_{S_1} K_t(y,z)dz.$$

Proof:

$$\begin{aligned}\int_{S_1} K_t(x,y)dx &\geq \int_{H_2\cap B_2'} K_t(x,y)dx (\text{since } H_2\cap B_2' \subseteq S_1) \\ &> \int_{H_2} K_t(x,y)dx - \int_{B_2'^c} K_t(x,y)dx \\ &> h(r+R^2/2) - \epsilon\end{aligned}$$

The last inequality follows from lemma 2. □

Lemma 5 *$\int_{S_1} K_t(x,y)\psi_t(x)\psi_t(y)dx > p(\pi(y))(1-O(\ell))(h(r+R^2/2)-\epsilon)$.*

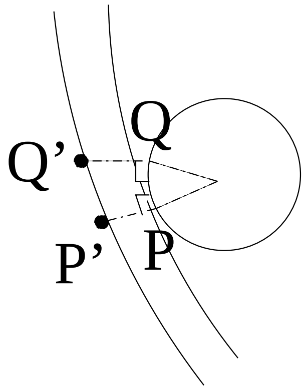

Figure 3: The correspondence between points on ∂S_1 and $\partial [S_1]_r$

Proof:

$$\begin{aligned}\int_{S_1} K_t(x,y)\psi_t(x)\psi_t(y)dx &\geq \int_{H_2\cap B_2'} K_t(x,y)\psi_t(x)\psi_t(y)dx (\text{since } H_2 \cap B_2' \subseteq S_1)\\ &> p(\pi(y))(1-O(\ell))(h(r+R^2/2)-\epsilon).\end{aligned}$$

□

Lemma 6 *Let $r > R^2$. Then,*

$$\int_{S_1} K_t(x,y)\psi_t(x)\psi_t(y)dx < (1+O(\ell))(h(r-R^2/2)p(\pi(y)) + \epsilon p_{max}).$$

Proof:

$$\begin{aligned}\int_{S_1} K_t(x,y)\psi_t(x)\psi_t(y)dx &\leq \int_{\mathbb{R}^d - B_1} K_t(x,y)\psi_t(x)\psi_t(y)dx\\ &\leq \int_{H_1\cup\mathbb{R}^d - B_1'} K_t(x,y)\psi_t(x)\psi_t(y)dx\\ &< \int_{H_1\cap B_1'} K_t(x,y)\psi_t(x)\psi_t(y)dx + \int_{B_1'^c} K_t(x,y)\psi_t(x)\psi_t(y)dx\\ &< h(r-R^2/2)p(\pi(y))(1+O(\ell)) + \epsilon p_{max}(1+O(\ell))\\ &< (1+O(\ell))(h(r-R^2/2)p(\pi(y)) + \epsilon p_{max})\end{aligned}$$

The last inequality follows from lemma 1. □

Definition 6 *Let $[S_1]_r$ denote the set of points at a distance of $\leq r$ to $[S_1]$. Let π_r be map from $\partial[S_1]_r$ to $\partial[S_1]$ that takes a point P on $\partial[S_1]_r$ to the foot of the perpendicular from P to ∂S_1. (This map is well-defined since $r < \tau = 1$.)*

Lemma 7 *Let $y \in \partial[S_1]_r$. Let the Jacobian of a map f be denoted by Df.*

$$(1-r)^{d-1} \leq |D\pi_r(y)| \leq (1+r)^{d-1}.$$

Proof: Let $\widehat{PQ}$ be a geodesic arc of infinitesimal length ds on ∂S_1 joining P and Q. Let $\pi_r^{-1}(P) = P'$ and $\pi_r^{-1}(Q) = Q'$ (see Figure D.) The radius of curvature of $\widehat{PQ}$ is ≥ 1. Therefore the distance between P' and Q' is in the interval $[ds(1-r), ds(1+r)]$. This implies that the Jacobian of the map π_r has a magnitude that is always in the interval $[(1+r)^{1-d}, (1-r)^{1-d}]$. □

Lemma 8

$$\int_{\mathbb{R}^d - [S_1]_R} \int_{S_1} K_t(x,y)\psi_t(x)\psi_t(y)dxdy \leq \epsilon vol\, S_1 p_{max}(1 + O(\ell)).$$

Proof:

$$\begin{aligned}
\int_{\mathbb{R}^d - [S_1]_R} \int_{S_1} K_t(x,y)\psi_t(x)\psi_t(y)dxdy &= \int_{S_1} \int_{\mathbb{R}^d - [S_1]_R} K_t(x,y)\psi_t(x)\psi_t(y)dydx \\
&\leq \int_{S_1} \int_{\|z\| > R} K_t(0,z)p_{max}(1 + O(\ell))dzdx \\
&< \text{vol } S_1 p_{max}(1 + O(\ell)).
\end{aligned}$$

$\leq$ in line 2 holds because the distance between x and y in the double integral is always $\geq R$. □

Lemma 9

$$(1 - \mathrm{e}^{-\alpha^2/4t})\sqrt{\pi/t} \leq \int_0^\alpha h(r)dr \leq \sqrt{\pi/t}.$$

Proof: Using observation 1,

$$\int_\alpha^\infty h(r)dr = \int_\alpha^\infty \int_{-\infty}^0 \frac{\mathrm{e}^{-(x_1-y_1)^2/4t}}{\sqrt{4\pi t}} dx_1 dy_1.$$

Setting $y_1 - x_1 := r$, this becomes

$$\int_\alpha^\infty \int_\alpha^r \frac{\mathrm{e}^{-r^2/4t}}{\sqrt{4\pi t}} dy_1 dr = \int_\alpha^\infty \frac{\mathrm{e}^{-r^2/4t}}{\sqrt{4\pi t}}(r - \alpha)dr.$$

Making the substitution $r - \alpha := z$, we have

$$\begin{aligned}
\int_0^\infty \frac{\mathrm{e}^{-(z+\alpha)^2/4t}}{\sqrt{4\pi t}} zdz &\leq \int_0^\infty \frac{\mathrm{e}^{-\alpha^2/4t}\mathrm{e}^{-z^2/4t} zdz}{\sqrt{4\pi t}} \\
&= \sqrt{\frac{t}{\pi}}\mathrm{e}^{-\alpha^2/4t}
\end{aligned}$$

Equality holds in the above calculation if and only if $\alpha = 0$. Hence the proof is complete. □

Definition 7 *Let $[S_2]^0 \cap \partial[S_1]_r$ be ∂M_r. Let $[S_2]^1 \cap \partial[S_1]_r$ be ∂M_r^1 and $[S_2]^1 \cap [S_1]_r$ be M_r^1.*

We assume that vol$(M_R^1 - S_1) < C'R^2$ for some absolute constant C'. Since the thickness of $(M_R^1 - S_1)$ is $O(R)$ in two dimensions, this is a reasonable assumption to make. The assumption that ∂M has a $d-1-$dimensional volume implies that vol$S_2^1 = O(R)$.

Putting these together to prove Theorem 3:

$$
\begin{aligned}
\int_{S_2^1}\int_{S_1} K_t(x,y)\psi_t(x)\psi_t(y)dxdy &= \int_0^\infty \int_{\partial M_r^1}\int_{S_1} K_t(x,y)\psi_t(x)\psi_t(y)dxdydr \\
&\leq O(p_{max}^2/p_{min}) \int_0^\infty \int_{\partial M_r^1}\int_{S_1} K_t(x,y)dxdydr \\
&\leq O(p_{max}^2/p_{min}) \left(\int_0^R \int_{\partial M_r^1}\int_{S_1} K_t(x,y)dxdydr + \mathrm{vol} S_2^1 \epsilon\right) \\
&\leq O(p_{max}^2/p_{min}) \left(\mathrm{vol}\ (M_R^1 - S_1) + \epsilon \mathrm{vol} S_2^1\right). \\
&\leq O(p_{max}^2/p_{min}) \left(C' t^{1-\mu} + O(t^{1-\mu}\mathrm{vol}\ \partial M)\right)
\end{aligned}
$$

$$
\begin{aligned}
\int_{S_2^0}\int_{S_1} K_t(x,y)\psi_t(x)\psi_t(y)dxdy &= \int_0^\infty \int_{\partial M_r}\int_{S_1} K_t(x,y)\psi_t(x)\psi_t(y)dxdydr \\
&= (\int_0^R \int_{\partial M_r}\int_{S_1} K_t(x,y)\psi_t(x)\psi_t(y)dxdydr) \\
&+ (\int_R^\infty \int_{\partial M_r}\int_{S_1} K_t(x,y)\psi_t(x)\psi_t(y)dxdydr) \\
&\leq (\int_0^R \int_{\partial M_r}\int_{S_1} K_t(x,y)\psi_t(x)\psi_t(y)dxdydr) \\
&+ \underbrace{\epsilon \mathrm{vol}\ S_1 p_{max}(1+O(\ell))}_{E}.
\end{aligned}
$$

(from lemma 8)

$$
\begin{aligned}
&\leq \int_0^{R^2}\int_{\partial M_r} p_{max}(1+O(\ell))dydr \\
&+ \int_{R^2}^R \int_{\partial M_r} (1+O(\ell))(h(r-R^2/2)p(\pi(y)+\epsilon p_{max})) + E
\end{aligned}
$$

The last line follows from Lemma 6.

$$
\begin{aligned}
&\leq E + R^2(1+R^2)^{d-1}p_{max}(1+O(\ell))\mathrm{vol}\ (\partial M_0)dr (\text{from lemma 7}) \\
&+ (1+R)^{d-1}(1+O(\ell))(\int_0^R h(r)dr \int_{\partial M_0} p(y)dy + \epsilon p_{max} R). \\
&\leq (1+O(\ell))(\sqrt{t/\pi}\int_{\partial M_0} p(y)dy + p_{max}((R^2+\epsilon R)\mathrm{vol}\ (\partial M_0) + \epsilon \mathrm{vol}\ S_1)). \\
&\leq (1+O(\ell))(\int_{\partial M_0} p(y)dy(\sqrt{t/\pi} + \frac{p_{max}}{p_{min}} o(t^{1-\mu})) + p_{max}\epsilon \mathrm{vol}\ S_1)
\end{aligned}
$$

Similarly, we see that

$$
\int_{S_2^0}\int_{S_1} K_t(x,y)\psi_t(x)\psi_t(y)dxdy
$$

$$
\begin{aligned}
&= \int_0^\infty \int_{\partial M_r} \int_{S_1} K_t(x,y)\psi_t(x)\psi_t(y)dxdydr \\
&> (\int_0^R \int_{\partial M_r} \int_{S_1} K_t(x,y)\psi_t(x)\psi_t(y)dxdydr) \\
&> \int_0^R \int_{\partial M_r} p(\pi(y))(1+O(\ell))f(r)dxdr \\
&> \int_0^R (1-R)^{d-1}(1-O(\ell))(\int_{\partial M_0} p(y)dy)(h(r+R^2/2)-\epsilon)dr \\
&> (1-R)^{d-1}(1-O((1-L/p_{min})R))(\int_{\partial M_0} p(y)dy)((\int_0^R (h(r)dr)-\epsilon R - R^2/2) \\
&\geq (1-O(\ell))(\int_{\partial M_0} p(y)dy)((1-\mathrm{e}^{-R^2/4t})\sqrt{t/\pi} - \epsilon R - R^2/2) \\
&\geq (1-O(\ell))(\int_{\partial M_0} p(y)dy)(\sqrt{t/\pi} - o(t^{1-\mu})).
\end{aligned}
$$

Noting only the dependence of the rate on t, and introducing the condition number τ,

$$
\sqrt{\frac{\pi}{t}} \int_{S_2} \int_{S_1} K_t(x,y)\psi_t(x)\psi_t(y)dxdy = \left(1 + o((t/\tau^2))^{\frac{1-\mu}{2}}\right) \int_S p(s)ds.
$$

□

Proof of Theorem 4: This follows directly from Theorem 2 and Theorem 3. The only change made was that the $t^{\frac{d+1}{2}}$ term was eliminated since it is dominated by t^ϵ when t is small. □

References

[1] M. Belkin and P. Niyogi (2004)."Semi-supervised Learning on Riemannian Manifolds." In *Machine Learning* 56, Special Issue on Clustering, 209-239.

[2] M. Belkin and P. Niyogi. "Toward a theoretical foundation for Laplacian-based manifold methods." COLT 2005.

[3] A. Blum and S. Chawla, "Learning from labeled and unlabeled data using graph mincuts", ICML 2001.

[4] O.Chapelle, J. Weston, B. Scholkopf, "Cluster kernels for semi-supervised learning", NIPS 2002.

[5] O. Chapelle and A. Zien, "Semi-supervised Classification by Low Density Separation", AISTATS 2005.

[6] Chris Ding, Spectral Clustering, ICML 2004 Tutorial.

[7] Samuel Kutin, Partha Niyogi, "Almost-everywhere Algorithmic Stability and Generalization Error.", UAI 2002, 275-282

[8] S. Kutin, TR-2002-04, "Extensions to McDiarmid's inequality when differences are bounded with high probability." Technical report TR-2002-04 at the Department of Computer Science, University of Chicago.

[9] S. Lafon, *Diffusion Maps and Geodesic Harmonics*, Ph. D. Thesis, Yale University, 2004.

[10] M. Hein, J.-Y. Audibert, U. von Luxburg, *From Graphs to Manifolds – Weak and Strong Pointwise Consistency of Graph Laplacians*, COLT 2005.

[11] J. Shi and J. Malik. "Normalized cuts and image segmentation."

[12] U. von Luxburg, M. Belkin, O. Bousquet, *Consistency of Spectral Clustering*, Max Planck Institute for Biological Cybernetics Technical Report TR 134, 2004.

[13] X. Zhu, J. Lafferty and Z. Ghahramani, *Semi-supervised learning using Gaussian fields and harmonic functions*, ICML 2003.