Supplementary Material to
Revisiting Kernelized Locality-Sensitive Hashing
for Improved Large-Scale Image Retrieval

Ke Jiang, Qichao Que, Brian Kulis
Department of Computer Science and Engineering
The Ohio State University
{jiongk, que, kulis}@cse.ohio-state.edu

We first present a proof of Lemma 2 from the main text.

**Proof.** By the Pythagorean theorem, we have
\[
N(x)^2 = \|P_{V_k} (\Phi(x))\|^2 = \|\Phi(x)\|^2 - \|P_{\perp V_k} (\Phi(x))\|^2.
\]
The residual \(P_{\perp V_k} (\Phi(x))\) can be further decomposed into
\[
P_{\perp V_k} (\Phi(x)) = P_{\perp V_k} (\Phi(x)) + \left( P_{\perp V_k} (\Phi(x)) - P_{\perp V_k} (\Phi(x)) \right).
\]
For the first term, we have
\[
\|P_{\perp V_k} (\Phi(x))\| \leq \sqrt{\lambda_k}.
\]
Then applying Theorem 4 in [4], with probability at least \(1 - e^{-\xi}\), we can also bound the second part of (2):
\[
\left\|P_{\perp V_k} (\Phi(x)) - P_{\perp V_k} (\Phi(x))\right\| \leq \frac{2M}{\delta_k \sqrt{m}} \left( 1 + \sqrt{\frac{\xi}{2}} \right),
\]
where \(\delta_k = \frac{\lambda_{k-1} - \lambda_k}{2}\) and \(M = \sup_x \kappa(x, x) = 1\). Thus,
\[
\left\|P_{\perp V_k} (\Phi(x))\right\| \leq \sqrt{\lambda_k} + \frac{2}{\delta_k \sqrt{m}} \left( 1 + \sqrt{\frac{\xi}{2}} \right).
\]
Putting these pieces together, with probability at least \(1 - e^{-\xi}\), we have
\[
N(x)^2 = \sqrt{1 - \left\|P_{\perp V_k} (\Phi(x))\right\|^2} \geq 1 - \left\|P_{\perp V_k} (\Phi(x))\right\| \geq 1 - \sqrt{\lambda_k} - \frac{2}{\delta_k \sqrt{m}} \left( 1 + \sqrt{\frac{\xi}{2}} \right).
\]
\[
\text{Lemma 1.} \text{ Consider a feature map } \Phi \in H \text{ defined by a normalized kernel function } \kappa(\cdot, \cdot) \text{ in } \mathcal{X} \text{ with a probability measure } p. \text{ Let } S_m = \{x_1, \ldots, x_m\} \text{ be } m \text{ i.i.d. samples drawn from } p \text{ and } C \text{ be the covariance operator of } p \text{ with decreasing eigenvalues } \lambda_1 \geq \lambda_2 \geq \ldots. \text{ Let } V_k \text{ and } \tilde{V}_k \text{ be the eigen-spaces corresponding the covariance operator } C \text{ and its empirical counterpart } C_{S_m}. \text{ Then, with probability at least } 1 - e^{-\xi} \text{ over the selection of } S_m, \text{ we have}
\]
\[
\left\langle P_{\perp V_k} (\Phi(x)), P_{\tilde{V}_k} (\Phi(y)) \right\rangle \leq \left( \sqrt{\lambda_k} + \frac{2}{\delta_k \sqrt{m}} \left( 1 + \sqrt{\frac{\xi}{2}} \right) \right)^2.
\]

**Proof.** From Cauchy-Schwarz inequality, we have
\[
\left\langle P_{\perp V_k} (\Phi(x)), P_{\perp V_k} (\Phi(y)) \right\rangle \leq \left\| P_{\perp V_k} (\Phi(x)) \right\| \left\| P_{\perp V_k} (\Phi(y)) \right\|.
\]
Then, for any \(x \in \mathcal{X}\),
\[
\left\| P_{\perp V_k} (\Phi(x)) \right\|
\leq \left\| P_{\perp V_k} (\Phi(x)) \right\| + \left\| P_{\perp V_k} (\Phi(x)) - P_{\perp V_k} (\Phi(x)) \right\|
\leq \left\| P_{\perp V_k} (\Phi(x)) \right\| + \left\| P_{\perp V_k} (\Phi(x)) - P_{\perp V_k} (\Phi(x)) \right\|.
\]
By the definition of operator norms, we have
\[
\left\| P_{\tilde{V}_k} (\Phi(x)) \right\| \leq \left\| \lambda_k \right\| \leq \sqrt{\lambda_k} \left\| \Phi(x) \right\|.
\]
Moreover, as stated in Theorem 4 in [4], with probability at least \(1 - e^{-\xi}\), we have that
\[
\left\| P_{\tilde{V}_k} (\Phi(x)) \right\| \leq \frac{2M}{\delta_k \sqrt{m}} \left( 1 + \sqrt{\frac{\xi}{2}} \right).
\]
where \(\delta_k = \frac{\lambda_k - \lambda_{k+1}}{2}\) and \(M = \sup_x \kappa(x, x) = 1\).
Hence, with probability at least \(1 - e^{-\xi}\) over \(S_m\), we have
\[
\left\langle P_{\perp V_k} (\Phi(x)), P_{\tilde{V}_k} (\Phi(y)) \right\rangle \leq \left( \sqrt{\lambda_k} + \frac{2}{\delta_k \sqrt{m}} \left( 1 + \sqrt{\frac{\xi}{2}} \right) \right)^2 \left\| \Phi(x) \right\| \left\| \Phi(y) \right\|.
\]
Since \( \| \Phi(x) \| = 1 \) for any \( x \), we have proved the lemma. \( \square \)

For completeness, we also state the performance bound of standard LSH:

**Theorem 2.** [3, 2, 1]. Let \((X, d_X)\) be a metric space on a subset of \( \mathbb{R}^d \). Suppose that \((X, d_X)\) admits a similarity hashing family. Then for any \( \epsilon > 0 \), there exists a randomized algorithm for \((1 + \epsilon)\)-near neighbor on \( n \)-point database with success probability larger than 0.5, which uses \( O(dn + n^{1+\frac{\epsilon}{\sqrt{\epsilon}}} \) \) space, with query time dominated by \( O(n^{1+\frac{\epsilon}{\sqrt{\epsilon}}} \) distance computations.

Using the above results, we are ready to prove our main result in Theorem 3.

**Proof.** By the definition of \( P_{\tilde{V}_k} \), we can decompose \( \kappa(q, \hat{y}_{\theta,k}) \) into two parts,

\[
\kappa(q, \hat{y}_{\theta,k}) = \hat{\kappa}(q, \hat{y}_{\theta,k}) + \langle P_{\tilde{V}_k}^\perp(\Phi(q)), P_{\tilde{V}_k}^\perp(\Phi(\hat{y}_{\theta,k})) \rangle.
\]

Thus, by Lemma 1, we have

\[
\kappa(q, \hat{y}_{\theta,k}) 
\geq \hat{\kappa}(q, \hat{y}_{\theta,k}) - \left| \langle P_{\tilde{V}_k}^\perp(\Phi(q)), P_{\tilde{V}_k}^\perp(\Phi(\hat{y}_{\theta,k})) \rangle \right|.
\]

\[
\geq \hat{\kappa}(q, \hat{y}_{\theta,k}) - \left( \sqrt{\lambda_k} + \frac{2}{\delta_k \sqrt{m}} \left( 1 + \sqrt{\frac{\xi}{2}} \right) \right)^2.
\]

To lower-bound \( \hat{\kappa}(q, \hat{y}_{\theta,k}) \), we need to use the result for LSH, which asks for normalized kernels. Thus, we consider the normalized version,

\[
\hat{\kappa}_n(q, \hat{y}_{\theta,k}) = \frac{\hat{\kappa}(q, \hat{y}_{\theta,k})}{N(q)N(\hat{y}_{\theta,k})}.
\]

Then we can relate a distance function via \( \hat{d}(q, \hat{y}_{\theta,k}) = 1 - \hat{\kappa}_n(q, \hat{y}_{\theta,k}) \) \[1\]. By the LSH guarantee in Theorem 2, with probability larger than 0.5, we have

\[
\hat{d}(q, \hat{y}_{\theta,k}) \leq (1 + \epsilon)\hat{d}(q, y^*_k),
\]

which is equivalent to

\[
\hat{\kappa}_n(q, \hat{y}_{\theta,k}) \geq (1 + \epsilon)\hat{\kappa}_n(q, y^*_k) - \epsilon.
\]

Expanding \( \hat{\kappa}_n \), we get

\[
\frac{\hat{\kappa}(q, \hat{y}_{\theta,k})}{N(q)N(\hat{y}_{\theta,k})} \geq (1 + \epsilon)\frac{\hat{\kappa}(q, y^*_k)}{N(q)N(y^*_k)} - \epsilon.
\]

which can be reduced to

\[
\hat{\kappa}(q, \hat{y}_{\theta,k}) \geq (1 + \epsilon)(1 - \sqrt{\lambda_k - \eta})\hat{\kappa}(q, y^*_k) - \epsilon,
\]

since \((1 - \sqrt{\lambda_k - \eta}) \leq 1\). Decompose \( \hat{\kappa}(q, y^*_k) \) on right hand-side above as

\[
\hat{\kappa}(q, y^*_k) = \kappa(q, y^*_k) - \langle P_{\tilde{V}_k}^\perp(\Phi(q)), P_{\tilde{V}_k}^\perp(\Phi(y^*_k)) \rangle.
\]

Thus,

\[
\hat{\kappa}(q, \hat{y}_{\theta,k}) \geq (1 + \epsilon)(1 - \sqrt{\lambda_k - \eta})\kappa(q, y^*_k)
\]

\[
- (1 + \epsilon) \left| \langle P_{\tilde{V}_k}^\perp(\Phi(q)), P_{\tilde{V}_k}^\perp(\Phi(y^*_k)) \rangle \right| - \epsilon.
\]

Combining results in Equation 14 and 24,

\[
\kappa(q, \hat{y}_{\theta,k}) \geq \kappa(q, y^*_k) - \left| \langle P_{\tilde{V}_k}^\perp(\Phi(q)), P_{\tilde{V}_k}^\perp(\Phi(y^*_k)) \rangle \right| - (2 + \epsilon) \left( \sqrt{\lambda_k + \eta} \right)^2.
\]

References


