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
{jiangk, 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

$$N(\boldsymbol{x})^2 = \|P_{\hat{V}_k}(\Phi(\boldsymbol{x}))\|^2 = \|\Phi(\boldsymbol{x})\|^2 - \|P^{\perp}_{\hat{V}_k}(\Phi(\boldsymbol{x}))\|^2. \quad (1)$$

The residual $P^{\perp}_{\hat{V}_k}(\Phi(\boldsymbol{x}))$ can be further decomposed into

$$P^{\perp}_{\hat{V}_k}(\Phi(\boldsymbol{x})) = P^{\perp}_{V_k}(\Phi(\boldsymbol{x})) + \left(P^{\perp}_{\hat{V}_k}(\Phi(\boldsymbol{x})) - P^{\perp}_{V_k}(\Phi(\boldsymbol{x}))\right). \quad (2)$$

For the first term, we have

$$\|P^{\perp}_{V_k}(\Phi(\boldsymbol{x}))\| \leq \sqrt{\lambda_k}. \quad (3)$$

Then applying Theorem 4 in [4], with probability at least $1 - e^{-\xi}$, we can also bound the second part of (2):

$$\left\|P^{\perp}_{\hat{V}_k}(\Phi(\boldsymbol{x})) - P^{\perp}_{V_k}(\Phi(\boldsymbol{x}))\right\| \leq \frac{2M}{\delta_k\sqrt{m}}\left(1 + \sqrt{\frac{\xi}{2}}\right), \quad (4)$$

where $\delta_k = \frac{\lambda_k - \lambda_{k+1}}{2}$ and $M = \sup_{\boldsymbol{x}} \kappa(\boldsymbol{x}, \boldsymbol{x}) = 1$. Thus,

$$\|P^{\perp}_{\hat{V}_k}(\Phi(\boldsymbol{x}))\| \leq \sqrt{\lambda_k} + \frac{2}{\delta_k\sqrt{m}}\left(1 + \sqrt{\frac{\xi}{2}}\right). \quad (5)$$

Putting these pieces together, with probability at least $1 - e^{-\xi}$, we have

$$N(\boldsymbol{x}) = \sqrt{1 - \|P^{\perp}_{\hat{V}_k}(\Phi(\boldsymbol{x}))\|^2} \geq 1 - \|P^{\perp}_{\hat{V}_k}(\Phi(\boldsymbol{x}))\| \quad (6)$$

$$\geq 1 - \sqrt{\lambda_k} - \frac{2}{\delta_k\sqrt{m}}\left(1 + \sqrt{\frac{\xi}{2}}\right). \quad (7)$$

□

Our proof of Theorem 3, which is given below, requires a few prerequisite results, which we briefly summarize now. The first is regarding the upper bound of inner product of complement of projections onto the subspace from kernel principal component analysis.

Lemma 1. *Consider a feature map $\Phi \in \mathcal{H}$ defined by a normalized kernel function $\kappa(\cdot,\cdot)$ in $\mathcal{X}$ with a probability measure p. Let $S_m = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_m\}$ be m i.i.d. samples drawn from p and C be the covariance operator of p with decreasing eigenvalues $\lambda_1 \geq \lambda_2 \geq \ldots$. Let V_k and $\hat{V}_k$ be the eigen-spaces corresponding the covariance operator C and its empirical counterpart C_{S_m}. Then, with probability at least $1 - e^{-\xi}$ over the selection of S_m, we have*

$$\langle P^{\perp}_{\hat{V}_k}(\Phi(\boldsymbol{x})), P^{\perp}_{\hat{V}_k}(\Phi(\boldsymbol{y}))\rangle \leq \left(\sqrt{\lambda_k} + \frac{2}{\delta_k\sqrt{m}}\left(1 + \sqrt{\frac{\xi}{2}}\right)\right)^2. \quad (8)$$

Proof. From Cauchy-Schwarz inequality, we have

$$\langle P^{\perp}_{\hat{V}_k}(\Phi(\boldsymbol{x})), P^{\perp}_{\hat{V}_k}(\Phi(\boldsymbol{y}))\rangle \leq \|P^{\perp}_{\hat{V}_k}(\Phi(\boldsymbol{x}))\|\|P^{\perp}_{\hat{V}_k}(\Phi(\boldsymbol{y}))\|. \quad (9)$$

Then, for any $x \in \mathcal{X}$,

$$\|P^{\perp}_{\hat{V}_k}(\Phi(\boldsymbol{x}))\|$$
$$\leq \|P^{\perp}_{V_k}(\Phi(\boldsymbol{x}))\| + \|P^{\perp}_{\hat{V}_k}(\Phi(\boldsymbol{x})) - P^{\perp}_{V_k}(\Phi(\boldsymbol{x}))\| \quad (10)$$
$$\leq \|P^{\perp}_{V_k}(\Phi(\boldsymbol{x}))\| + \|P^{\perp}_{\hat{V}_k} - P^{\perp}_{V_k}\|\|\Phi(\boldsymbol{x})\|. \quad (11)$$

By the definition of operator norms, we have $\|P^{\perp}_{V_k}(\Phi(\boldsymbol{x}))\| \leq \sqrt{\lambda_k}\|\Phi(\boldsymbol{x})\|$. Moreover, as stated in Theorem 4 in [4], with probability at least $1 - e^{-\xi}$, we have that

$$\|P^{\perp}_{\hat{V}_k} - P^{\perp}_{V_k}\| \leq \frac{2M}{\delta_k\sqrt{m}}\left(1 + \sqrt{\frac{\xi}{2}}\right) \quad (12)$$

where $\delta_k = \frac{\lambda_k - \lambda_{k+1}}{2}$ and $M = \sup_x \kappa(\boldsymbol{x}, \boldsymbol{x}) = 1$.

Hence, with probability at least $1 - e^{-\xi}$ over S_m, we have

$$\langle P^{\perp}_{\hat{V}_k}(\Phi(\boldsymbol{x})), P^{\perp}_{\hat{V}_k}(\Phi(\boldsymbol{y}))\rangle \quad (13)$$
$$\leq \left(\sqrt{\lambda_k} + \frac{2}{\delta_k\sqrt{m}}\left(1 + \sqrt{\frac{\xi}{2}}\right)\right)^2 \|\Phi(\boldsymbol{x})\|\|\Phi(\boldsymbol{y})\|.$$

978-1-4673-6964-0/15/$31.00 ©2015 IEEE

Since $\|\Phi(\boldsymbol{x})\| = 1$ for any $\boldsymbol{x}$, 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 $O(dn + n^{1+\frac{1}{1+\epsilon}})$ space, with query time dominated by $O(n^{\frac{1}{1+\epsilon}})$ distance computations.*

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

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

$$\kappa(\boldsymbol{q}, \hat{\boldsymbol{y}}_{q,k}) = \hat{\kappa}(\boldsymbol{q}, \hat{\boldsymbol{y}}_{q,k}) + \langle P^{\perp}_{\hat{V}_k}(\Phi(\boldsymbol{q})), P^{\perp}_{\hat{V}_k}(\Phi(\hat{\boldsymbol{y}}_{q,k}))\rangle. \tag{14}$$

Thus, by Lemma 1, we have

$$\begin{aligned}
&\kappa(\boldsymbol{q}, \hat{\boldsymbol{y}}_{q,k}) \\
\geq&\hat{\kappa}(\boldsymbol{q}, \hat{\boldsymbol{y}}_{q,k}) - \left|\langle P^{\perp}_{\hat{V}_k}(\Phi(\boldsymbol{q})), P^{\perp}_{\hat{V}_k}(\Phi(\hat{\boldsymbol{y}}_{q,k}))\rangle\right| \qquad (15)\\
\geq&\hat{\kappa}(\boldsymbol{q}, \hat{\boldsymbol{y}}_{q,k}) - \left(\sqrt{\lambda_k} + \frac{2}{\delta_k\sqrt{m}}\left(1 + \sqrt{\frac{\xi}{2}}\right)\right)^2. \qquad (16)
\end{aligned}$$

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

$$\hat{\kappa}_n(\boldsymbol{q}, \hat{\boldsymbol{y}}_{q,k}) = \frac{\hat{\kappa}(\boldsymbol{q}, \hat{\boldsymbol{y}}_{q,k})}{N(\boldsymbol{q})N(\hat{\boldsymbol{y}}_{q,k})}. \tag{17}$$

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

$$\hat{d}(\boldsymbol{q}, \hat{\boldsymbol{y}}_{q,k}) \leq (1+\epsilon)\hat{d}(\boldsymbol{q}, \boldsymbol{y}^*_{q,k}), \tag{18}$$

which is equivalent to

$$\hat{\kappa}_n(\boldsymbol{q}, \hat{\boldsymbol{y}}_{q,k}) \geq (1+\epsilon)\hat{\kappa}_n(\boldsymbol{q}, \boldsymbol{y}^*_{q,k}) - \epsilon, \tag{19}$$

where $\boldsymbol{y}^*_{q,k} = \text{argmax}_{\boldsymbol{x}\in S}\hat{\kappa}_n(\boldsymbol{q}, \boldsymbol{x})$. Applying Lemma 2, with probability $1 - e^{-\xi}$, the true optimal $\boldsymbol{y}^*_q$ with respect to κ is not eliminated for LSH, thus we have $\hat{\kappa}_n(\boldsymbol{q}, \boldsymbol{y}^*_{q,k}) \geq \hat{\kappa}_n(\boldsymbol{q}, \boldsymbol{y}^*_q)$ due to the optimality of $\boldsymbol{y}^*_{q,k}$ with respect to $\hat{\kappa}_n$, and with probability $0.5 \times (1 - e^{-\xi})$

$$\hat{\kappa}_n(\boldsymbol{q}, \hat{\boldsymbol{y}}_{q,k}) \geq (1+\epsilon)\hat{\kappa}_n(\boldsymbol{q}, \boldsymbol{y}^*_q) - \epsilon. \tag{20}$$

Expanding $\hat{\kappa}_n$, we get

$$\frac{\hat{\kappa}(\boldsymbol{q}, \hat{\boldsymbol{y}}_{q,k})}{N(\boldsymbol{q})N(\hat{\boldsymbol{y}}_{q,k})} \geq (1+\epsilon)\frac{\hat{\kappa}(\boldsymbol{q}, \boldsymbol{y}^*_q)}{N(\boldsymbol{q})N(\boldsymbol{y}^*_q)} - \epsilon. \tag{21}$$

which can be reduced to

$$\hat{\kappa}(\boldsymbol{q}, \hat{\boldsymbol{y}}_{q,k}) \geq (1+\epsilon)(1 - \sqrt{\lambda_k} - \eta)\hat{\kappa}(\boldsymbol{q}, \boldsymbol{y}^*_q) - \epsilon, \tag{22}$$

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

$$\hat{\kappa}(\boldsymbol{q}, \boldsymbol{y}^*_q) = \kappa(\boldsymbol{q}, \boldsymbol{y}^*_q) - \langle P^{\perp}_{\hat{V}_k}(\Phi(\boldsymbol{q})), P^{\perp}_{\hat{V}_k}(\Phi(\boldsymbol{y}^*_q))\rangle. \tag{23}$$

Thus,

$$\begin{aligned}
\hat{\kappa}(\boldsymbol{q}, \hat{\boldsymbol{y}}_{q,k}) \geq& (1+\epsilon)(1 - \sqrt{\lambda_k} - \eta)\kappa(\boldsymbol{q}, \boldsymbol{y}^*_q) \qquad (24)\\
&- (1+\epsilon)\left|\langle P^{\perp}_{\hat{V}_k}(\Phi(\boldsymbol{q})), P^{\perp}_{\hat{V}_k}(\Phi(\boldsymbol{y}^*_q))\rangle\right| - \epsilon.
\end{aligned}$$

Combining results in Equation 14 and 24,

$$\begin{aligned}
&\kappa(\boldsymbol{q}, \hat{\boldsymbol{y}}_{q,k}) \\
\geq&\hat{\kappa}(\boldsymbol{q}, \hat{\boldsymbol{y}}_{q,k}) - \left|\langle P^{\perp}_{\hat{V}_k}(\Phi(\boldsymbol{q})), P^{\perp}_{\hat{V}_k}(\Phi(\hat{\boldsymbol{y}}_{q,k}))\rangle\right| \qquad (25)\\
\geq&(1+\epsilon)(1 - \sqrt{\lambda_k} - \eta)\kappa(\boldsymbol{q}, \boldsymbol{y}^*_q) - \epsilon \\
&- (1+\epsilon)\left|\langle P^{\perp}_{\hat{V}_k}(\Phi(\boldsymbol{q})), P^{\perp}_{\hat{V}_k}(\Phi(\boldsymbol{y}^*_q))\rangle\right| \\
&- \left|\langle P^{\perp}_{\hat{V}_k}(\Phi(\boldsymbol{q})), P^{\perp}_{\hat{V}_k}(\Phi(\hat{\boldsymbol{y}}_{q,k}))\rangle\right| \qquad (26)
\end{aligned}$$

Now we can apply Lemma 1:

$$\begin{aligned}
\kappa(\boldsymbol{q}, \hat{\boldsymbol{y}}_{q,k}) \geq&(1+\epsilon)(1 - \sqrt{\lambda_k} - \eta)\kappa(\boldsymbol{q}, \boldsymbol{y}^*_q) - \epsilon \\
&- (2+\epsilon)\left(\sqrt{\lambda_k} + \eta\right)^2. \qquad (27)
\end{aligned}$$

$\square$

References

[1] M. Charikar. Similarity estimation techniques for rounding algorithms. In *ACM Symposium on Theory of Computing*, 2002. 2

[2] M. Datar, N. Immorlica, P. Indyk, and V. Mirrokni. Locality-sensitive hashing scheme based on p-stable distributions. In *Proceedings of the Symposium on Computational Geometry*, 2004. 2

[3] A. Gionis, P. Indyk, and R. Motwani. Similarity search in high dimension via hashing. In *Proceedings of the International Conference on Very Large DataBases*, 1999. 2

[4] Z. Laurent and G. Blanchard. On the convergence of eigenspaces in kernel principal component analysis. In *Advances in Neural Information Processing Systems*, 2005. 1