Supplementary material for: The Variational Nyström Method for Large-Scale Spectral Problems

Max Vladymyrov*
Google, Inc.

Miguel Á. Carreira-Perpiñán
EECS, UC Merced

May 24, 2016

Abstract

In this supplementary material we prove the propositions from the main paper (Vladymyrov and Carreira-Perpiñán, 2016) regarding the choice of normalization for the graph Laplacian kernel. We also show the relation between Variational Nyström and modified Nyström (Wang and Zhang, 2013) as well as between LLL (Vladymyrov and Carreira-Perpiñán, 2013) and Randomized Projections (Halko et al., 2011).

1 Normalization for out-of-sample kernel for graph Laplacians

First we prove the normalization properties of the out-of-sample kernel for the Nyström and Column Sampling methods.

Proposition (3.1 from the main paper)**.** *Given a subsample* $\mathbf{L}_\mathbf{A}^{Nys} = \mathbf{D}_\mathbf{A}^{-1/2}\mathbf{A}\mathbf{D}_\mathbf{A}^{-1/2}$ *for the Nyström method, or* $\mathbf{L}_\mathbf{A}^{CS} = (\mathbf{D}_1\mathbf{C}\mathbf{D}_2)^T(\mathbf{D}_1\mathbf{C}\mathbf{D}_2)$ *for the Column Sampling, and* $\mathbf{Z} = \mathbf{D}_1\mathbf{C}\mathbf{D}_2$ *as an out-of-sample kernel, the exact eigenvectors of graph Laplacian* $\mathbf{L}_\mathbf{W}$ *for* $L = N$ *are recovered when* $\mathbf{D}_1 = \mathbf{D}_2 = \mathbf{D}_\mathbf{W}^{-1/2}$.

Proof. First, let us start with the Nyström method. The Nyström approximation to the eigenvectors of $\mathbf{L}_\mathbf{W}$ is $\widetilde{\mathbf{U}} = \mathbf{Z}\mathbf{U}_\mathbf{A}\mathbf{\Lambda}_\mathbf{A}^{-1}$, where $\mathbf{U}_\mathbf{A}$ and $\mathbf{\Lambda}_\mathbf{A}$ are given by the eigendecomposition $\mathbf{D}_\mathbf{A}^{-1/2}\mathbf{A}\mathbf{D}_\mathbf{A}^{-1/2}\mathbf{U}_\mathbf{A} = \mathbf{U}_\mathbf{A}\mathbf{\Lambda}_\mathbf{A}$ (see table 1 from the main paper). If $\mathbf{A}$ is positive definite, we have

$$\begin{aligned}\widetilde{\mathbf{U}} &= \mathbf{Z}\mathbf{U}_\mathbf{A}\mathbf{\Lambda}_\mathbf{A}^{-1} = \mathbf{Z}(\mathbf{D}_\mathbf{A}^{-1/2}\mathbf{A}\mathbf{D}_\mathbf{A}^{-1/2})^{-1}(\mathbf{D}_\mathbf{A}^{-1/2}\mathbf{A}\mathbf{D}_\mathbf{A}^{-1/2})\mathbf{U}_\mathbf{A}\mathbf{\Lambda}_\mathbf{A}^{-1}\\ &= \mathbf{Z}\mathbf{D}_\mathbf{A}^{1/2}\mathbf{A}^{-1}\mathbf{D}_\mathbf{A}^{1/2}\mathbf{U}_\mathbf{A}\mathbf{\Lambda}_\mathbf{A}\mathbf{\Lambda}_\mathbf{A}^{-1} = \mathbf{Z}\mathbf{D}_\mathbf{A}^{1/2}\mathbf{A}^{-1}\mathbf{D}_\mathbf{A}^{1/2}\mathbf{U}_\mathbf{A}.\end{aligned}$$

In the case $N = L$ we have $\mathbf{C} = \mathbf{A} = \mathbf{W} = \mathbf{W}^T$, $\mathbf{U}_\mathbf{A} = \mathbf{U}$, $\mathbf{D}_\mathbf{A} = \mathbf{D}_\mathbf{W}$ and so the approximation becomes $\mathbf{U} = \mathbf{D}_1\mathbf{W}\mathbf{D}_2\mathbf{D}_\mathbf{W}^{1/2}\mathbf{W}^{-1}\mathbf{D}_\mathbf{W}^{1/2}\mathbf{U}$. Thus, for the approximation to be exact, we need to have $\mathbf{D}_1\mathbf{W}\mathbf{D}_2 = \mathbf{D}_\mathbf{W}^{-1/2}\mathbf{W}\mathbf{D}_\mathbf{W}^{-1/2}$. One way to achieve this is to take $\mathbf{D}_1 = \mathbf{D}_2 = \mathbf{D}_\mathbf{W}^{-1/2}$.

Now for the column sampling method. The eigenvectors of the subproblem are given by

$$(\mathbf{D}_1\mathbf{C}\mathbf{D}_2)^T(\mathbf{D}_1\mathbf{C}\mathbf{D}_2)\hat{\mathbf{U}} = \hat{\mathbf{U}}\hat{\mathbf{\Lambda}}.$$

From there the eigendecomposition of $\mathbf{D}_1\mathbf{C}\mathbf{D}_2$ is given by $(\mathbf{D}_1\mathbf{C}\mathbf{D}_2)\widetilde{\mathbf{U}} = \widetilde{\mathbf{U}}\widetilde{\mathbf{\Lambda}}^{1/2}$. The approximation of the eigenvectors of $\mathbf{L}_\mathbf{W}$ is $\widetilde{\mathbf{U}} = \mathbf{Z}\mathbf{U}_\mathbf{C}\mathbf{\Lambda}_\mathbf{C}^{-1/2}$. When $L = N$ we get:

$$\widetilde{\mathbf{U}} = \mathbf{D}_1\mathbf{W}\mathbf{D}_2(\mathbf{D}_1\mathbf{W}\mathbf{D}_2)^{-1}(\mathbf{D}_1\mathbf{W}\mathbf{D}_2)\widetilde{\mathbf{U}}\widetilde{\mathbf{\Lambda}}^{-1/2} = \mathbf{D}_1\mathbf{W}\mathbf{D}_2\mathbf{D}_2^{-1}\mathbf{W}^{-1}\mathbf{D}_1^{-1}\widetilde{\mathbf{U}}\widetilde{\mathbf{\Lambda}}^{1/2}\widetilde{\mathbf{\Lambda}}^{-1/2} = \widetilde{\mathbf{U}}.$$

Thus, the solution is exact when $\mathbf{D}_1\mathbf{W}\mathbf{D}_2 = \mathbf{D}^{-1/2}\mathbf{W}\mathbf{D}^{-1/2}$. One way to satisfy this is to take $\mathbf{D}_1 = \mathbf{D}_2 = \mathbf{D}_\mathbf{W}^{-1/2}$. □

*Part of this work was performed while the author was a PhD student at UC Merced.

Now for the properties of the Variational Nyström method.

Proposition (3.2 from the main paper)**.** *Given a subsample* $\mathbf{L}_{\mathbf{A}}^{VN} = \mathbf{D}_{\mathbf{A}}^{-1/2}\mathbf{A}\mathbf{D}_{\mathbf{A}}^{-1/2}$ *and* $\mathbf{Z} = \mathbf{D}_1\mathbf{C}\mathbf{D}_2$ *as an out-of-sample kernel, the exact eigenvectors of graph Laplacian* $\mathbf{L}_{\mathbf{W}}$ *are recovered for any* $L \leq N$ *using any arbitrary symmetrical matrix* $\mathbf{D}_1$*. When* $L = N$ *the eigenvectors are recovered using any symmetrical* $\mathbf{D}_2$*.*

Proof. Let us prove first that premultiplying $\mathbf{Z}$ by any symmetrical matrix $\mathbf{D}_1$ does not change the approximation if $L \leq N$. Using the out-of-sample kernel $\mathbf{D}_1\mathbf{Z}$, the generalized eigenproblem becomes $\mathbf{D}_1\mathbf{Z}\mathbf{A}\mathbf{Z}^T\mathbf{D}_1\widetilde{\mathbf{U}} = \mathbf{D}_1\mathbf{Z}\mathbf{Z}^T\mathbf{D}_1\widetilde{\mathbf{U}}\widetilde{\mathbf{\Lambda}}$, which is identical to $\mathbf{Z}\mathbf{A}\mathbf{Z}^T\hat{\mathbf{U}} = \mathbf{Z}\mathbf{Z}^T\hat{\mathbf{U}}\hat{\mathbf{\Lambda}}$ if we take $\hat{\mathbf{U}} = \mathbf{D}_1\widetilde{\mathbf{U}}$ and $\hat{\mathbf{\Lambda}} = \widetilde{\mathbf{\Lambda}}$. For the out-of-sample extension we use $\mathbf{Z}\widetilde{\mathbf{U}} = \mathbf{Z}^T\mathbf{D}_1\widetilde{\mathbf{U}}$, which is equal to $\mathbf{Z}^T\hat{\mathbf{U}}$.

Now, let us show that for $L = N$ post-multiplying $\mathbf{Z}$ by any symmetrical $\mathbf{D}_2$ does not change the results. Let $\mathbf{W}\mathbf{U} = \mathbf{D}_{\mathbf{W}}\mathbf{U}\mathbf{\Lambda}$ be the solution of the exact problem. For $L = N$ we have $\mathbf{Z} = \mathbf{W}$ and the approximation to the eigenproblem becomes $\mathbf{W}\mathbf{D}_2\mathbf{W}\mathbf{D}_2\mathbf{W}\widetilde{\mathbf{U}} = \mathbf{W}\mathbf{D}_2\mathbf{D}_{\mathbf{W}}\mathbf{D}_2\mathbf{W}\widetilde{\mathbf{U}}\widetilde{\mathbf{\Lambda}}$ with the final approximation $\mathbf{U} = \mathbf{D}_2\mathbf{W}\widetilde{\mathbf{U}}$. Plugging this approximation in the exact solution gives $\mathbf{W}\mathbf{D}_2\mathbf{W}\widetilde{\mathbf{U}} = \mathbf{D}_{\mathbf{W}}\mathbf{D}_2\mathbf{W}\widetilde{\mathbf{U}}\mathbf{\Lambda}$, which is identical to the approximate solution when $\mathbf{W}\mathbf{D}_2$ is invertible. □

2 Relation between the variational and modified Nyström methods

Here we show the connection between the variational Nyström (Vladymyrov and Carreira-Perpiñán, 2016) and modified Nyström (Wang and Zhang, 2013) methods. Let $\mathbf{M}$ be an $N \times N$ symmetric matrix and $\mathbf{Z} = \mathbf{C}$ a subset of its L columns. Then, the modified Nyström method approximates $\mathbf{M}$ with a low-rank approximation $\widetilde{\mathbf{M}}^{\text{mod}} = \mathbf{Z}\mathbf{Z}^{+}\mathbf{M}(\mathbf{Z}^{+})^T\mathbf{Z}^T$. Variational Nyström, on the contrary, starts with the optimization problem (P) from the main paper and results in the approximation of the eigenvectors of $\mathbf{M}$.

Proposition 2.1. *If* $\mathbf{Z}$ *has full rank and the eigenvalues of* $\mathbf{M}$ *are approximated as* $\mathbf{\Lambda}^{VN} = \mathbf{\Lambda}_{\mathbf{Z}}$*, where* $\mathbf{\Lambda}_{\mathbf{Z}}$ *are the eigenvectors of the generalized eigenproblem* $\mathbf{Z}^T\mathbf{M}\mathbf{Z}\mathbf{U}_{\mathbf{Z}} = \mathbf{Z}^T\mathbf{Z}\mathbf{U}_{\mathbf{Z}}\mathbf{\Lambda}_{\mathbf{Z}}$*, then the spectral approximation* $\widetilde{\mathbf{M}}^{VN} = \mathbf{U}^{VN}\mathbf{\Lambda}^{VN}(\mathbf{U}^{VN})^T$ *becomes identical to the modified Nyström approximation* $\widetilde{\mathbf{M}}^{mod} = \mathbf{Z}\mathbf{Z}^{+}\mathbf{M}(\mathbf{Z}^{+})^T\mathbf{Z}^T$*.*

Proof. Variational Nyström approximates the eigenvectors of $\mathbf{M}$ as $\mathbf{U}^{\text{VN}} = \mathbf{Z}\mathbf{U}_{\mathbf{Z}}$, where $\mathbf{U}_{\mathbf{Z}}$ are eigenvectors of the generalized eigenproblem $\mathbf{Z}^T\mathbf{M}\mathbf{Z}\mathbf{U}_{\mathbf{Z}} = \mathbf{Z}^T\mathbf{Z}\mathbf{U}_{\mathbf{Z}}\mathbf{\Lambda}_{\mathbf{Z}}$. Therefore, $\widetilde{\mathbf{M}}^{\text{VN}} = \mathbf{Z}\mathbf{U}_{\mathbf{Z}}\mathbf{\Lambda}_{\mathbf{Z}}\mathbf{U}_{\mathbf{Z}}^T\mathbf{Z}^T$. We can alternatively prove the proposition by showing that if $\mathbf{\Lambda}^{\text{VN}} = \mathbf{\Lambda}_{\mathbf{Z}}$ the intersection matrices of $\widetilde{\mathbf{M}}^{\text{VN}}$ and $\widetilde{\mathbf{M}}^{\text{mod}}$ are equal, i.e. $\mathbf{U}_{\mathbf{Z}}\mathbf{\Lambda}_{\mathbf{Z}}\mathbf{U}_{\mathbf{Z}}^T = \mathbf{Z}^{+}\mathbf{M}(\mathbf{Z}^{+})^T$.

Converting the generalized eigenvalue into a standard one using a substitution $\mathbf{U}_{\mathbf{R}} = (\mathbf{Z}^T\mathbf{Z})^{1/2}\mathbf{U}_{\mathbf{Z}}$ (assuming $\mathbf{Z}$ is full-rank) we get:

$$(\mathbf{Z}^T\mathbf{Z})^{-1/2}\mathbf{Z}^T\mathbf{M}\mathbf{Z}(\mathbf{Z}^T\mathbf{Z})^{-1/2} = \mathbf{U}_{\mathbf{R}}\mathbf{\Lambda}_{\mathbf{Z}}\mathbf{U}_{\mathbf{R}}^T. \tag{1}$$

Now the intersection matrix of $\widetilde{\mathbf{M}}^{\text{VN}}$ becomes

$$\begin{aligned}
\mathbf{U}_{\mathbf{Z}}\mathbf{\Lambda}_{\mathbf{Z}}\mathbf{U}_{\mathbf{Z}}^T &= (\mathbf{Z}^T\mathbf{Z})^{-1/2}\mathbf{U}_{\mathbf{R}}\mathbf{\Lambda}_{\mathbf{Z}}\mathbf{U}_{\mathbf{R}}^T(\mathbf{Z}^T\mathbf{Z})^{-1/2} \\
&= (\mathbf{Z}^T\mathbf{Z})^{-1/2}(\mathbf{Z}^T\mathbf{Z})^{-1/2}\mathbf{Z}^T\mathbf{M}\mathbf{Z}(\mathbf{Z}^T\mathbf{Z})^{-1/2}(\mathbf{Z}^T\mathbf{Z})^{-1/2} \\
&= (\mathbf{Z}^T\mathbf{Z})^{-1}\mathbf{Z}^T\mathbf{M}\mathbf{Z}(\mathbf{Z}^T\mathbf{Z})^{-1} = \mathbf{Z}^{+}\mathbf{M}(\mathbf{Z}^{+})^T.
\end{aligned}$$

□

3 Relation between the LLL and randomized projection methods

Here we show the relation between the LLL (Vladymyrov and Carreira-Perpiñán, 2013) and random projection (RP) (Halko et al., 2011) methods. Given the input matrix $\mathbf{M}$, the RP method first computes the sample matrix $\mathbf{M}_{\mathbf{S}} = \mathbf{M}\mathbf{S}$, where $\mathbf{S}$ is an $N \times L$ uniformly distributed random matrix. It then finds an orthonormal basis $\mathbf{Q}$ of the column space of $\mathbf{M}_{\mathbf{S}}$ (e.g. using a QR decomposition); projects the data matrix $\mathbf{M}$ onto this basis, $\mathbf{M}_{\mathbf{S}} = \mathbf{Q}^T\mathbf{M}$; and then computes a reduced SVD of the projection, $\mathbf{M}_{\mathbf{S}} = \mathbf{U}_{\mathbf{S}}\mathbf{\Sigma}_{\mathbf{S}}\mathbf{V}_{\mathbf{S}}$. The approximate eigenvectors of $\mathbf{A}$ are then reconstructed using a re-projection of the left singular vectors as $\mathbf{Q}\mathbf{U}_{\mathbf{S}}$.

The RP method can be viewed as an instance of LLL when the orthonormal basis $\mathbf{Q}$ of RP is used as projection matrix $\mathbf{Z}$ of LLL. The left singular vectors of $\mathbf{Q}^T\mathbf{M}$ are equal to the eigenvectors of $\mathbf{Q}^T\mathbf{M}\mathbf{Q}$ (left-multiplying by the orthonormal matrix just changes the right singular vectors). Since $\mathbf{Q}^T\mathbf{Q} = \mathbf{I}$, the eigenproblem becomes identical to $\mathbf{Z}^T\mathbf{M}\mathbf{Z}\widetilde{\mathbf{U}} = \mathbf{Z}^T\mathbf{Z}\widetilde{\mathbf{U}}\widetilde{\mathbf{\Lambda}}$, which is the solution to the LLL optimization problem.

References

N. Halko, P. G. Martinsson, and J. A. Tropp. Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions. *SIAM Review*, 53(2):217–288, 2011.

M. Vladymyrov and M. Á. Carreira-Perpiñán. Locally Linear Landmarks for large-scale manifold learning. In H. Blockeel, K. Kersting, S. Nijssen, and F. Zelezný, editors, *Proc. of the 24th European Conf. Machine Learning (ECML–13)*, pages 256–271, Prague, Czech Republic, Sept. 23–27 2013.

M. Vladymyrov and M. Á. Carreira-Perpiñán. The Variational Nyström method for large-scale spectral problems. In M.-F. Balcan and K. Weinberger, editors, *Proc. of the 33rd Int. Conf. Machine Learning (ICML 2016)*, New York, NY, June 19–24 2016.

S. Wang and Z. Zhang. Improving CUR matrix decomposition and the Nyström approximation via adaptive sampling. *J. Machine Learning Research*, 14:2729–2769, Sept. 2013.