Random-Projection Based Singular Value Decomposition (RSVD)

December 2019

Random Projections (RP)

There are multiple methods for dimensionality reduction, including SVD but RP is the only method that's data-independent and consequently, is capable of solving the redundancy problem while keeping the computational costs low. RP involves choosing a random subspace for the projection of data from a significantly high-dimensional space. The mapping from high-dimensional space to lower-dimensional space is achieved by simply multiplying the data with a random matrix.

Theoretical Origin of RP

RP is based on the Johnson-Lindenstrauss lemma [2] which guarantees that in Euclidean space, there always exists a projection from high-dimensional data to low-dimensional data that satisfies distortion conditions with high probability.

J-L lemma Definition:

For any dataset of n points, $\mathcal{V} = \{x_1, x_2, ..., x_n\}$ in $\mathbb{R}^d$, any $0 < \epsilon < 1$, let k be a positive integer such that

$$k \geq \frac{4logn}{\epsilon^2/2 - \epsilon^3/3}$$

then, there exists a linear map $f : \mathbb{R}^d \rightarrow \mathbb{R}^k$ such that $\forall$ $\mathrm{x}_i, x_j \in \mathcal{V}$, with very high probability we have

$$(1 - \epsilon)\|\mathrm{x}_i - x_j\|^2 \leq \|\mathrm{f}(\mathrm{x}_i) - f(x_j)\|^2 \leq (1{+}\epsilon)\|\mathrm{x}_i - x_j\|^2$$

RP of GW-Templates

The reason RP is a great dimensionality reduction method is because it preserves embedding. This means that the Euclidean distance between two data points in Euclidean space is preserved even when it is mapped onto a lower-dimensional subspace. By Euclidean distance, we literally mean the distance dS^2 between two data points. The GW template banks are Euclidean in nature and hence, RP can be effectively applied since the data structure will be preserved under RP of the template bank to a lower-dimensional space.

Generating a Random Matrix

The J-L lemma above describes a linear map f which describes a mapping between two Euclidean spaces of different dimensions. The main question here is - what is f and how can it be defined?

This proves to be a very complicated mathematical procedure. However, in our case, f can also be understood as a random matrix Ω which is a function that maps a higher-dimensional space to a lower-dimensional space. This "embedding function" f or Ω is simply a Gaussian matrix.

Randomized-SVD

A Summary of SVD and RSVD

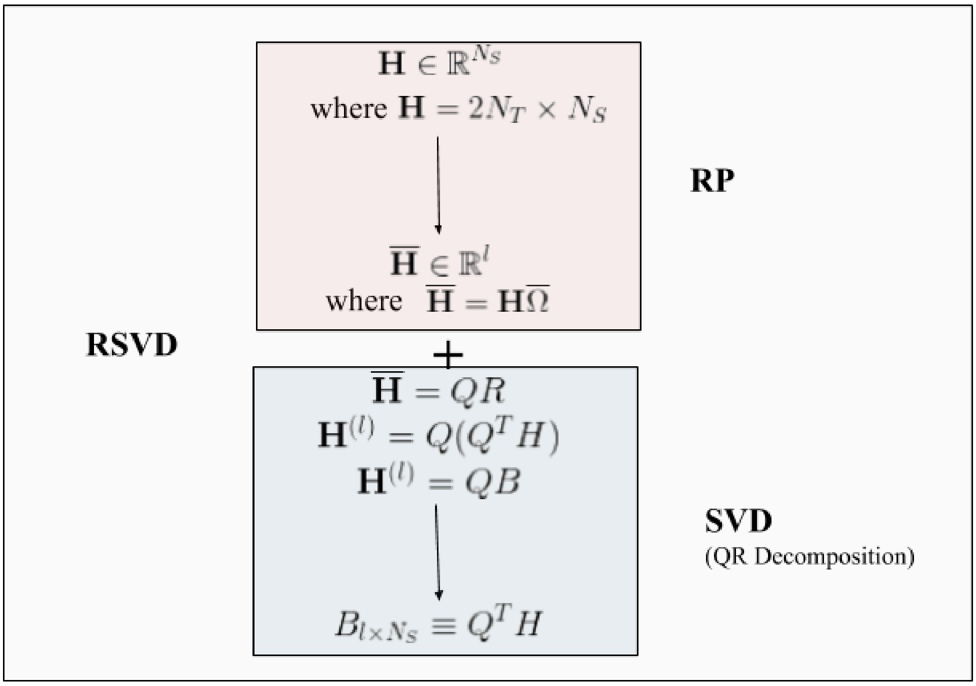
SVD

$\#1$ Define a vector $\vec{a}$ and transpose it to get $\vec{a}^T$.
$\#2$ Define projection direction unit vectors $\vec{v}_1, \vec{v}_2$ and projection lengths $\vec{S}_{a1}, \vec{S}_{a2}$.
$\#3$ Write in the form of $a^T.V = S$
$\#4$ Extend it to more vectors i.e. more number of points and dimensions to basically get $A = SV^T$ (A -vector of $n \times d$ dimensions, V - projection direction unit vectors of dimension $d \times d$ and S - projection lengths vector of $n \times d$ dimensions)
$\#5$ The SVD formula is given by $A = U\Sigma V^T$ which implies that $S = U\Sigma$.
(Σ is a vector of the magnitude values of the S matrix columns)
$\#6$ Matrix V is used to correlate the data and S is used for reconstruction.

RSVD

$\#1$ We have $2N_T$ vectors in $\mathbb{R}^{N_S}$, describing the template matrix $\mathbf{H}$ where $\mathbf{H} = 2N_T \times \mathrm{N}_S$
$\#2$ The individual row-vectors of $\mathbf{H}$ are mapped into a randomly generated subspace $\mathbb{R}^l$ using $\Omega_{\vec{N_S \times l}}$ ϵ $\mathcal{N}(0,\ 1/l)$
$\#3$ We get $\vec{\mathbf{H}}_{2N_T \times l} = \mathbf{H}\vec{\Omega}$ where Ω is the random matrix.
$\#4$ $\vec{\mathbf{H}}$ can now be used for SVD-like operations i.e. QR decomposition. [3]
$\#5$ We obtain a matrix B after decomposition, the rows of this matrix can be treated as filters and be used to correlate the data.

Summary Page of RSVD



l	$\sim$ O(log(2N$_T$)/ϵ^2
ϵ	$0 < \epsilon < 1$
H	$= 2N_T \times$N$_S$

H	template matrix
N_T	number of templates
N_S	sample points
$\mathbb{R}^{N_S}$	N_S-dimensional Euclidean space
$\mathbb{R}^l$	l-dimensional randomly generated subspace
ϵ	mismatch/distortion factor after projection
$\overline{\mathbf{H}}$	template matrix in lower-dimensional Euclidean space
$\overline{\Omega}$	random matrix
Q	orthonormal matrix with $2N_T \times l$ dimensions
B	orthonormal projection of template waveforms into the compressed subspace

What do we do with the B matrix?

$$B_{l\times N_S} \equiv \mathrm{Q}^T H$$

The matrix B defines the orthonormal projection of the template waveforms into the compressed subspace. We can use the l rows of B as surrogate templates which can be used to correlate against the detector data $\vec{S}$. We can simply understand the l rows as filters. And hence, the next thing we do with the B matrix is proceed to perform matched filtering.

Reconstruction of SNR

Introduction

The template bank consists of templates of alternating polarizations. This implies that every alternate template corresponds to plus-polarization and the others to cross-polarization. Hence, the phase shift between any two consecutive templates in the template bank is always 90 degrees. The phase shift can then be maximized over all space to get a complex matched filter output.

Template banks can be made up of either intrinsic or extrinsic parameter spaces (templates):
Extrinsic Space: If there is an overlap between two templates when maximized over extrinsic parameters, it produces a **match**.
Instrinsic Space: If there is an overlap between two templates when maximized over intrinsic parameters, it produces **rank deficiency.** GW searches are usually done over intrinsic parameter space.
Typically, CBC searches involve calculating the cross-correlation between $\vec{S}$ and every row of $\mathbf{H}$ for a series of relative time-shifts, generating a time-series of ρ_α values, for every α.

But why do we need reconstruction?

The original $\mathbf{H}$ exists in a large parameter space, described by intrinsic parameters like mass and spin. (In our case, only masses m_1 and m_2). Using SVD or RSVD, we reduce this large parameter space into a smaller subspace $\overline{\mathbf{H}}$ or $\mathbf{H}^{(l)}$ which has no physical attributes or parameters. So after we have finished our work within the smaller subspace, we need to reconstruct it back into the original space to correlate the data with the physical parameters. However, since there is always a loss of data while using SVD or RSVD, it is nearly impossible to get back our original matrix exactly. However, the reconstructed matrix will be approximately equal to the original matrix.

Reconstruction of SNR in SVD

The rank deficiency of $\mathbf{H}$ is exploited in the truncated SVD approach - which means that singular values can be truncated and the template bank can be made smaller overall. To be precise, $(2N_T - l)$ singular values are discarded from the original template matrix. And hence, the reconstructed fractional SNR loss can be measured as a function of these discarded values.

The SNR of the α-th row of the template bank is given by

$$\rho_\alpha = (H_{(2\alpha-1)} \text{ - } i\ H_{(2\alpha)})\ \vec{S}^T$$

where $H_{(2\alpha-1)}$ and $H_{(2\alpha)}$ are rows (slices) of the template matrix $\mathbf{H}$; $\vec{S}^T$ is the transpose of the whitened signal data.

The average fractional SNR loss (or the truncation) is given by:

$$\left\langle \frac{\delta\rho}{\rho} \right\rangle \leq \frac{\|\mathbf{H}\|_F^2 - \|\mathbf{H}^{(l)}\|_F^2}{\|\mathbf{H}\|_F^2}$$

Reconstruction of SNR in RSVD

The rank-l matrix factorization of $\mathbf{H}$ using RSVD is given by $\mathbf{H}^{(l)} = QB$. Hence, the SNR $\rho_\alpha{}'$ for any given time t_0 can be reconstructed in $\mathbb{R}^{N_S}$ as

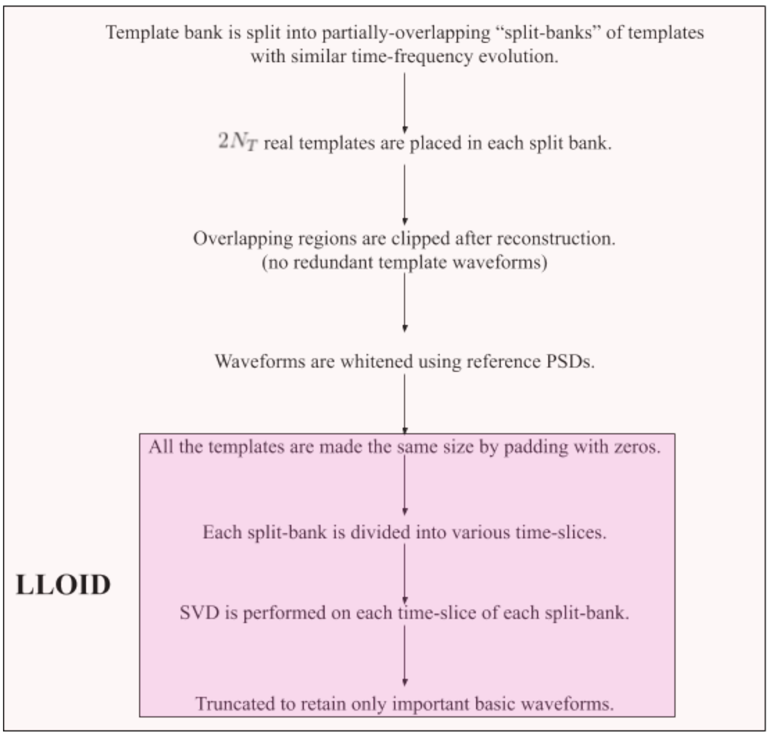
$$\rho_\alpha = (H_{(2\alpha-1)} \text{ - } i\ H_{(2\alpha)})\ \vec{S}^T = \sum_{\nu=1}^{l} \left(Q_{(2\alpha-1)\nu} - iQ_{(2\alpha)\nu}\right) \left(B_\nu \vec{S}^T\right)$$

Using both Pythagoras theorem and the relation $\|\mathbf{H}\|_F^2 - 2\mathrm{N}_T$, the average fractional loss of SNR is given by

$$\left\langle \frac{\delta\rho}{\rho} \right\rangle \leq \frac{\|\mathbf{H}\|_F^2 - \|\mathbf{H}^{(l)}\|_F^2}{\|\mathbf{H}\|_F^2} = 1 - \frac{\sum_{\mu=1}^{l} \sigma_\mu^2}{2N_T}$$

Final Thoughts - Comparison between SVD and RSVD

The current gstLAL pipeline performs SVD within the LLOID algorithm, as seen in the figure below. In simpler terms, the template bank is divided into smaller "split-banks". Each split-bank contains templates of similar time-frequency evolution. And then, SVD is performed individually on each split-bank.



However, RSVD reduces the entire template matrix space to a subspace and performs the decomposition operation only once. This could potentially cure the redundancy problem to a large extent.

In conclusion, RSVD will potentially perform better when it is introduced to larger template banks. This will help in conducting more complex searches, such as including the spin parameter during CBC searches.

Glossary

1. **CBC:** Compact Binary Coalescence;
2. **LLOID:** Low-Latency Inspiral Online Detection
3. **Dimension of Data:** number of variables in the data;
4. **Structure of the Data:** in the context of CBC searches, represented by the Euclidean distance between different points of a given database;
5. **Intrinsic Parameters:** parameters which are directly related to the GW event; e.g. mass, spin, etc.
6. **Extrinsic Parameters:** parameters which are not directly related to the GW event but are to do with the detector; e.g. time of arrival, phase at arrival, etc.
7. **Rank deficiency:** when one row of a matrix can be represented as a linear combination of the other rows; a higher-rank matrix can then be represented as a lower-rank matrix;
8. **Frobenius Norm:** mathematical term; defined as a matrix norm defined as the square root of the sum of the absolute squares of its elements;
9. **Embedding:** the metric dS^2 distance between two N_S sample-points;

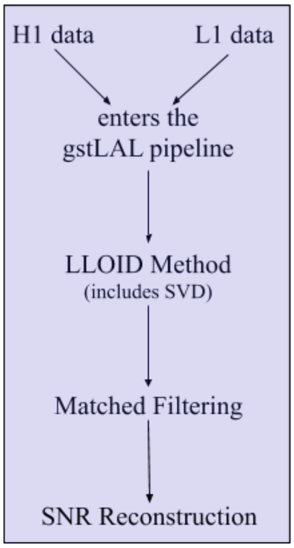
References

[1] Phukon et al. *Random projections in gravitational wave searches of compact binaries.* Phys. Rev. D 99, 101503(R), 2019.

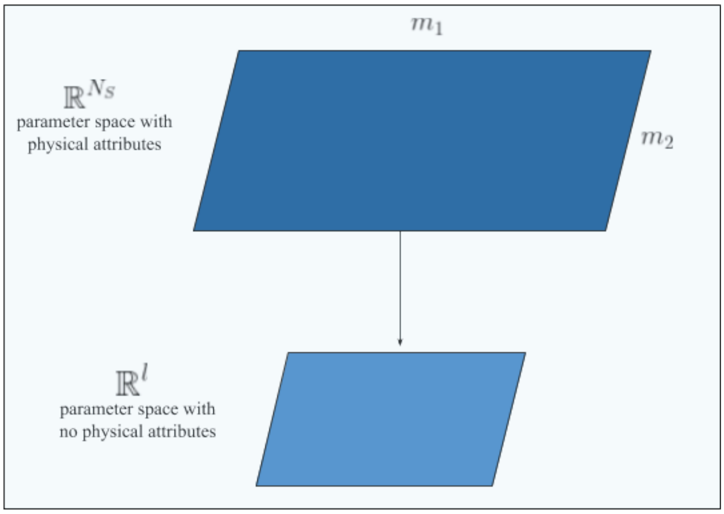
[2] W. B. Johnson and J. Lindenstrauss *Conference in modern analysis and probability (New Haven, Conn.,1982)* Contemp. Math., Vol. 26 (Amer. Math. Soc., Providence, RI, 1984) pp. 189–206.

[3] G. H. Golub and C. F. Van Loan *Matrix Computations (3rd Ed.)* (Johns Hopkins University Press, Baltimore, MD, USA, 1996).

Miscellaneous



Basic Structure of the gstLAL pipeline



SNR Reconstruction