# 1. Variational Inference

## 1.1. Matrix Inequality for Concave Function

For a positie semidefinite (PSD) matrix $\boldsymbol{V}$, we can say

$$\log|\boldsymbol{V}| \leq \log|\boldsymbol{\Omega}| + \mathrm{tr}(\boldsymbol{\Omega}^{-1}\boldsymbol{V}) - M, \tag{1}$$

where $\boldsymbol{\Omega}$ is an arbitrary PSD matrix. Letting the partial derivative of the right-hand side of inequality (1) with respect to $\boldsymbol{\Omega}$ to be zero, we have

$$\boldsymbol{\Omega} = \boldsymbol{V}. \tag{2}$$

Substituting $\boldsymbol{\Omega} = \boldsymbol{V}$ back in the right-hand side of inequality (1), we have

$$\log|\boldsymbol{\Omega}| + \mathrm{tr}(\boldsymbol{\Omega}^{-1}\boldsymbol{V}) - M = \log|\boldsymbol{V}| + \mathrm{tr}(\boldsymbol{V}^{-1}\boldsymbol{V}) - M = \log|\boldsymbol{V}|. \tag{3}$$

Since this is the minimum of the right-hand side, we have inequality (1) for arbitrary $\boldsymbol{\Omega}$.

## 1.2. Matrix Inequality for Convex Function

For a PSD matrix $\boldsymbol{Z}$ and a set of PSD matrices $\{\boldsymbol{V}_k\}_{k=1}^K$ such that $\sum_{k=1}^K \boldsymbol{V}_k$ is invertible, we can say

$$\mathrm{tr}\left(\left(\sum_{k=1}^K \boldsymbol{V}_k\right)^{-1}\boldsymbol{Z}\right) \leq \sum_{k=1}^K \mathrm{tr}\left(\boldsymbol{\Phi}_k^T\boldsymbol{V}_k^{-1}\boldsymbol{\Phi}_k\boldsymbol{Z}\right) \quad \text{s.t.} \quad \sum_{k=1}^K \boldsymbol{\Phi}_k = \boldsymbol{I}, \tag{4}$$

where $\{\boldsymbol{\Phi}_k\}_{k=1}^K$ is a set of atbitrary matrices that sum to the identity matrix. Define the Lagrangian of the right-hand side of inequality (4) as follows:

$$L\left(\{\boldsymbol{\Phi}_k\}_{k=1}^K, \boldsymbol{A}_k\right) = \frac{1}{2}\sum_{k=1}^K \mathrm{tr}\left(\boldsymbol{\Phi}_k^T\boldsymbol{V}_k^{-1}\boldsymbol{\Phi}_k\boldsymbol{Z}\right) + \mathrm{tr}\left(\boldsymbol{A}_k^T\left(\boldsymbol{I} - \sum_{k=1}^K \boldsymbol{\Phi}_k\right)\right), \tag{5}$$

where $\boldsymbol{A}_k$ is a Lagrangian multiplier that corresponds to the equality constraint in inequality (4). Letting the partial derivative with respect to $\boldsymbol{\Phi}_k$ to be zero, we have

$$\boldsymbol{\Phi}_k = \boldsymbol{V}_k\boldsymbol{A}_k. \tag{6}$$

Summing both sides, we have $\sum_{k=1}^K \boldsymbol{\Phi}_k = \sum_{k=1}^K \boldsymbol{V}_k\boldsymbol{A}_k = \boldsymbol{I}$, from which we have the optimal Lagrangian multiplier $\boldsymbol{A}_k = (\sum_{k=1}^K \boldsymbol{V}_k)^{-1}$. Substituting $\boldsymbol{\Phi}_k = \boldsymbol{V}_k(\sum_{k=1}^K \boldsymbol{V}_k)^{-1}$ back in the right-hand side of inequality (4), we have

$$\begin{aligned}\sum_{k=1}^K \mathrm{tr}\left(\boldsymbol{\Phi}_k^T\boldsymbol{V}_k^{-1}\boldsymbol{\Phi}_k\boldsymbol{Z}\right) &= \sum_{k=1}^K \mathrm{tr}\left(\left(\sum_{k=1}^K \boldsymbol{V}_k\right)^{-1}\boldsymbol{V}_k\left(\sum_{k=1}^K \boldsymbol{V}_k\right)^{-1}\boldsymbol{Z}\right)\\ &= \sum_{k=1}^K \mathrm{tr}\left(\left(\sum_{k=1}^K \boldsymbol{V}_k\right)^{-1}\boldsymbol{Z}\right).\end{aligned} \tag{7}$$

Since this is the minimum of the right-hand side, we have inequality (4) for arbitrary $\{\boldsymbol{\Phi}_k\}_{k=1}^K$ that sum to the identity matrix.

# 2. MGIG Distribution

As to matrix variable $\boldsymbol{V}_k$, we propose to use the matrix GIG (MGIG) distribution (Barndorff-Nielsen et al., 1982) as the functional form of $q(\boldsymbol{V}_k)$. The MGIG distribution over PSD matrix $\boldsymbol{X}$ is defined as

$$\mathrm{MGIG}(\boldsymbol{X}|\gamma, \boldsymbol{R}, \boldsymbol{T}) = \frac{2^{\gamma M}}{|\boldsymbol{T}|^{\gamma}B_\gamma(\boldsymbol{R}\boldsymbol{T}/4)}|\boldsymbol{X}|^{\gamma - \frac{M+1}{2}}\mathrm{etr}\left(-\frac{1}{2}\left(\boldsymbol{R}\boldsymbol{X} + \boldsymbol{T}\boldsymbol{X}^{-1}\right)\right), \tag{8}$$

where $\gamma$ is a real number and $\boldsymbol{R}, \boldsymbol{T} > \boldsymbol{0}$ are PSD matrices, $M$ is the size of $\boldsymbol{X}$, $B_\gamma$ is the matrix Bessel function of the second kind (Herz, 1955), and $\mathrm{etr}(z)$ indicates $\exp(\mathrm{tr}(z))$. It includes the Wishart and inverse-Wishart distributions as special cases (Butler, 1998), and its sufficient statistics are $\log|\boldsymbol{X}|$, $\boldsymbol{X}$, and $\boldsymbol{X}^{-1}$. It is, however, difficult to analytically calculate the expectations $\mathbb{E}[\boldsymbol{X}]$ and $\mathbb{E}[\boldsymbol{X}^{-1}]$. We therefore need to simulate those values by using a Monte Carlo method. More specifically, $\mathbb{E}[\boldsymbol{X}]$ is given by the following integral:

$$\mathbb{E}[\boldsymbol{X}] \propto \int \boldsymbol{X}|\boldsymbol{X}|^{\gamma-\frac{M+1}{2}} \mathrm{etr}\left(-\frac{1}{2}\left(\boldsymbol{R}\boldsymbol{X}+\boldsymbol{T}\boldsymbol{X}^{-1}\right)\right) d\boldsymbol{X} = \int \boldsymbol{X} \mathrm{etr}\left(-\frac{1}{2}\boldsymbol{T}\boldsymbol{X}^{-1}\right) d\mu(\boldsymbol{X}), \tag{9}$$

where $d\mu(\boldsymbol{X}) = |\boldsymbol{X}|^{\gamma-\frac{M+1}{2}} \mathrm{etr}\left(-\frac{1}{2}\boldsymbol{R}\boldsymbol{X}\right) d\boldsymbol{X}$. Thus if $\boldsymbol{U}$ is a random matrix drawn from a Wishart distribution as $\boldsymbol{U} \sim \mathcal{W}(2\gamma, \boldsymbol{R}^{-1})$, we can say

$$\mathbb{E}[\boldsymbol{X}] = \frac{C_{\mathcal{W}}(2\gamma, \boldsymbol{R}^{-1})}{C_{\mathrm{MGIG}}(\gamma, \boldsymbol{R}, \boldsymbol{T})} \mathbb{E}\left[\boldsymbol{U} \mathrm{etr}\left(-\frac{1}{2}\boldsymbol{T}\boldsymbol{U}^{-1}\right)\right], \tag{10}$$

where $C_{\mathcal{W}}(2\gamma, \boldsymbol{R}^{-1})$ and $C_{\mathrm{MGIG}}(\gamma, \boldsymbol{R}, \boldsymbol{T})$ are normalizing constants of the Wishart and MGIG distributions. Similarly, $\mathbb{E}[\boldsymbol{X}^{-1}]$ is given by

$$\mathbb{E}[\boldsymbol{X}^{-1}] = \frac{C_{\mathcal{W}}(2\gamma, \boldsymbol{R}^{-1})}{C_{\mathrm{MGIG}}(\gamma, \boldsymbol{R}, \boldsymbol{T})} \mathbb{E}\left[\boldsymbol{U}^{-1} \mathrm{etr}\left(-\frac{1}{2}\boldsymbol{T}\boldsymbol{U}^{-1}\right)\right], \tag{11}$$

The matrix Bessel function of $C_{\mathrm{MGIG}}(\gamma, \boldsymbol{R}, \boldsymbol{T})$ can be calculated by using Laplace approximation or Monte Carlo simulation (Butler & Wood, 2003).

## 3. Music Analysis

LD-PSDTF is useful for source separation of music audio signals. In general, source separation has been done on the frequency domain. In KL- or IS-NMF, for example, a given amplitude or power spectrogram can be split into a set of $K$ source spectrograms by using a Wiener-filtering technique. However, it is difficult to recover natural source signals from those spectrograms having no phase information. If the phase of the observed spectrogram is directly attached to the source spectrograms, the resulting signals have some unpleasant artifacts.

An advantage of time-domain LD-PSDTF is that real-valued source signals can be directly estimated in a probabilistic framework without tackling a difficult problem of phase reconstruction. This is achieved by generalized Wiener filtering (Eq. (15) of the paper) that assumes source signals to follow full-covariance Gaussians.

We tested LD-PSDTF on an audio signal synthesized by concatenating seven piano sounds (C, E, G, C+E, C+G, E+G, and C+E+G) with a MIDI synthesizer. The total length was 8.4 s (1.2 s * 7). The task was to separate the observed signal into three source signals having different pitches (C, E, and G). The signal was sampeld at 16 kHz and split into short overlapping frames by using a Gaussian window with a width of 512 samples ($M = 512$) and a shifting interval of 160 samples ($N = 840$). The PSD basis matrices and their activations were estimated by using the multiplicative update (MU) algorithm with $K = 3$. For comparison, we tested KL-NMF with $K = 3$ for amplitude-spectrogram decomposition and IS-NMF with $K = 3$ for power-spectrogram decomposition.

The experimental results showed overwhelming superiority of LD-PSDTF for source separation. The average SDR, SIR, and SAR were 16.7dB, 21.1dB, and 18.7dB for KL-NMF, 18.9dB, 24.1dB, and 20.5dB for IS-NMF, and 26.7dB, 33.2dB, and 27.8dB for LD-PSDTF. We found it practically effective to initialize LD-PSDTF by using basis vectors and their activations obtained by IS-NMF for reducing the computational cost and avoiding the local optima. This means that LD-PSDTF can be used as a high-quality sound generator for IS-NMF.

## References

Barndorff-Nielsen, O., Blæsild, P., Jensen, J. L., and Jørgensen, B. Exponential transformation models. *Royal Society of London*, 379(1776):41–65, 1982.

Butler, R. W. Generalized inverse Gaussian distributions and their Wishart connections. *Scandinavian Journal of Statistics*, 25(1):69–75, 1998.

Butler, R. W. and Wood, A. Laplace approximation for Bessel functions of matrix argument. *Journal of Computational and Applied Mathematics*, 155(2):359–382, 2003.

Herz, C. S. Bessel functions of matrix argument. *Annals of Mathematics*, 61(3):474–523, 1955.

# Source Separation of Audio Signals

## Mixture signal

| | Source signals of pitch C | Source signals of pitch E | Source signals of pitch G |
|---|---|---|---|
| Original | | | |
| KL-NMF | SDR 17.4dB<br>SIR 21.9dB<br>SAR 19.4dB | SDR 15.5dB<br>SIR 21.0dB<br>SAR 18.5dB | SDR 16.2dB<br>SIR 20.6dB<br>SAR 18.2dB |
| IS-NMF | SDR 18.3dB<br>SIR 23.9dB<br>SAR 19.7dB | SDR 20.5dB<br>SIR 26.2dB<br>SAR 21.9dB | SDR 17.9dB<br>SIR 22.4dB<br>SAR 19.8dB |
| LD-PSDTF | SDR 25.5dB<br>SIR 33.7dB<br>SAR 26.2dB | SDR 30.2dB<br>SIR 36.4dB<br>SAR 31.4dB | SDR 24.2dB<br>SIR 29.4dB<br>SAR 25.8dB |

Adobe Acrobat Reader is required
for listening to these audio samples